



 **STAINS**
Seminar Nasional Teknologi & Sains



E-ISSN : 2828-299X

PROSIDING

EMINAR NASIONAL TEKNOLOGI & SAINS

STAINS

"War : Gen-Z VS AI"
"Ketika perkembangan AI
akan menggantikan peran SDM"

Vol.3 No.1 (2024)
13/01/2024

PROGRAM STUDI TEKNIK INFORMATIKA
UNIVERSITAS NUSANTARA PGRI KEDIRI

PROSIDING

Seminar Nasional Teknologi dan Sains (STAINS)

WAR : GEN -Z VS AI Ketika Perkembangan AI Akan Menggantikan Peran SDM

Diselenggarakan oleh :

**Prodi Teknik Informatika
Fakultas Teknik dan Ilmu Komputer
Universitas Nusantara PGRI Kediri
Kediri, 13 Januari 2024**

KATA PENGANTAR

Assalamu’alaikum warohmatullahi wabarokatuh

Puji syukur senantiasa kita panjatkan kepada Allah SWT, atas segala limpahan rahmat dan hidayah-Nya, sehingga kita semua dapat bertemu pada kegiatan ilmiah Seminar Nasional Teknologi dan Sains tahun 2024. STAINS merupakan kegiatan seminar berskala nasional yang rutin diselenggarakan oleh Program Studi Teknik Informatika Universitas Nusantara PGRI Kediri. Seminar ini dimaksudkan sebagai forum untuk mempublikasikan hasil penelitian dan pemikiran tentang penerapan teknologi informasi dan teknologi terapan diberbagai bidang.

STAINS 2024 menerima 61 makalah dari berbagai provinsi di Indonesia. Pemakalah STAINS 2024 berasal dari 11 perguruan tinggi di Indonesia. Makalah STAINS 2024 direview setidaknya dua blind review. Prosiding STAINS 2024 di indexed oleh google scholar. STAINS 2024 ini bertemakan “**WAR : GEN -Z VS AI (Ketika Perkembangan AI Akan Menggantikan Peran SDM)**”. Melalui seminar ini diharapkan dapat memberikan sumbangan pemikiran baik secara empiris maupun teoritis tentang pengembangan teknologi informasi dan komunikasi sebagai sarana pendukung untuk mencapai kemandirian bangsa.

Pada kesempatan ini panitia menyampaikan ucapan terima kasih kepada semua pihak yang telah berpartisipasi aktif dalam STAINS 2024, khususnya kepada pemakalah yang telah mengirimkan makalahnya untuk di review dan dipresentasikan pada STAINS 2024. Ucapan terimakasih dan apresiasi juga kami sampaikan kepada para reviewer STAINS 2024 yang telah mengevaluasi makalah untuk menjadi lebih baik. Panitia juga menyampaikan terimakasih kepada Universitas Nusantara PGRI Kediri yang telah memberikan bantuan untuk menyelenggarakan kegiatan STAINS 2024. Akhir kata, panitia menyampaikan terima kasih kepad aberbagai pihak yang terlibat dalam penyelenggaraan STAINS 2024.

Wassalamu’alaikum Warohmatullahi wabarokatuh

Danar Putra Pamungkas,M.Kom.

Ketua Pelaksana
Seminar Nasional Teknologi dan Sains
STAINS 2024

SUSUNAN PANITIA

Pelaksanaan Seminar Nasional Call Paper STAINS 2024
Seminar Nasional Teknologi dan Sains
Program Studi Teknik Informatika Universitas Nusantara PGRI Kediri

WAR : GEN -Z VS AI
"KETIKA PERKEMBANGAN AI AKAN MENGGANTIKAN PERAN SDM"
Pada Tanggal : Sabtu, 13 Januari 2024

Penasehat	: Dr. Sulistiono, M.Si Dr. Risky Aswi Ramadhani, M.Kom
Ketua Umum	: Risa Helilintar, M.Kom
Ketua Pelaksana	: Danar Putra Pamungkas, M.Kom
Sekretaris	: Ahmad Bagus Setiawan, S.T., M.Kom., M.M
Bendahara	: Patmi Kasih, M.Kom
Sie Acara	: Daniel Swanjaya, M.Kom Made Ayu Dusea Widyadara, M.Kom Muh. Aris Saputra, M.Kom Ahmad Baihaqi, M.Kom
Sie Prosiding	: Wahyu Cahyo Utomo, S.Kom., M.Cs Bahrul Subkhi, M.Kom
Sie Reviewer	: Arif Mudi Priyatno, S.T., M.Kom (Universitas Pahlawan Tuanku Tambusai) Ilham Saifudin, S.Pd., M.Si (Universitas Muhammadiyah Jember) Fajar Rohman Hariri, M.Kom (Universitas Islam Negeri Maulana Malik Ibrahim Malang) Ratih Kumalasari Niswatin, S.ST., M.Kom Intan Nur Farida, M.Kom. Siti Rochana, M.Pd Lilia Sinta Wahyuniar, M.Pd Candra Ratna Hariyanti, M.Kom Resty Wulanningrum, M.Kom
Sie Kesekretariatan	: Umi Mahdiyah, S.Pd., M.Si Juli Sulaksono, Ir., M.M., M.Kom. Andi Prasetyo
Sie Perlengkapan Dokumentasi, & Humas	: Ardi Sanjaya, M.Kom. Danang Wahyu Widodo, S.P., M.Kom Julian Sahertian, S.Pd., M.T Rony Heri Irawan, M.Kom.
Keynote	: Dr. Indah Agustien Siradjuddin, S.Kom., M.Kom Umi Mahdiyah, S.Pd., M.Si
Invited Speaker	: Intan Nur Farida, M.Kom (UNP KEDIRI) Fitri Bimantoro, M.Kom (UNRAM)

DAFTAR ISI

HALAMAN JUDUL	i
KATA PENGANTAR	ii
SUSUNAN PANITIA	iii
DAFTAR ISI	iv

DAFTAR ISI ARTIKEL

Klasifikasi Jenis Daun Jambu Air dengan <i>Backpropagation</i> <i>Erry Anggraini, Nur Nafiyah</i>	1
Implementasi Metode k-Nearest Neighbor (k-NN) untuk Memprediksi Penjualan Buah di Indonesia berbasis Website <i>Siti Mujilahwati, Lilis Devita Windasari</i>	7
Implementasi Algoritma Naive Bayes pada Sistem Pendukung Keputusan Rekomendasi Wisata di Lamongan <i>Indra Dwi Rizqi, Retno Wardhani, Moh. Rosidi Zamroni</i>	15
Sistem Pakar Diagnosa Gangguan Pencernaan Balita Dengan Metode Naive Bayes <i>Emi Milda Prawestina, M.Ghofar Rohman, Moh. Rosidi Zamroni</i>	23
Case Based Reasoning Untuk Diagnosa Penyakit Gigi Berbasis Web <i>Chervin Fradyacsyah, Miftahus Sholihin, Danang Bagus Reknadi</i>	29
Identifikasi Tingkat Kesadaran Pengemudi dari Data Video dengan Menggunakan Convolutional Long Short Term Memory <i>Fajar Fatha Romadhan, Andrian Dwi Baitur Rizky, Muhammad Aulia Faqihuddin, Indah Agustien Siradjuddin</i>	37
Penerapan Jaringan Syaraf Tiruan Backpropagation untuk Klasifikasi Akreditasi Sekolah Menengah Pertama <i>Rachel Elisa Utama, Eko Hari Parmadi</i>	45
Korelasi antara Kekuatan Militer dan Faktor-Faktor Multidimensi: Analisis Tahun 2021 <i>Fathul Am, Riky Ananda Setyanto, Imanuel Puspa Wardaya, Ari Hendrawan</i>	53
Otomatisasi Pemisah Minyak dan Air Laut Menggunakan Separator pada Kapal Berbasis IoT <i>Hana Fitri Fikriyah, Anggit Yustyawan, Muhammad Farizul Hadi, Faisal Budiman, Brahmantya Aji Pramudita</i>	61
Perancangan Sistem Informasi Jdih Berbasis Web Dengan Metode <i>Prototype</i> <i>Ronaldo Arsad, Muhamad Son Muare</i>	67
Optimasi Model Prediksi Kesuksesan Startup Menggunakan StandartScaler Transform <i>Wulan Sri Lestari</i>	76
Review Komprehensif: Ekstraksi Fitur GLCM, GLRLM, dan LBP untuk Pendeteksian Korosi <i>Mizanul Ridho Aohana, Ratu Nisful Laily Hidhayah, Melki Jonathan Andara, Nadya Amara, Fitri Bimantoro</i>	82

Deteksi Parasit Malaria Menggunakan Metode <i>Gray Level Co-Occurance Matrix</i> (GLCM) <i>Ahmad Arsyad Surgi Mukti, Ghina Briliana Fatin Octariana, Krisna Dian Sukmana, Fitri Bimantoro</i>	91
Klasifikasi Jeruk Segar dan Busuk Melalui GLCM dan HSV dengan Menggunakan Metode ANN <i>Alisyia Kornelia Ulandari, Ghina Kamilah Ramdhani, Wahyuningsih, M. Naufal Arwansyuri, Fitri Bimantoro</i>	97
Klasifikasi Tumor Otak Menggunakan CNN Dengan Arsitektur Resnet50 <i>Mohammad Liyananta S., Muhammad Shata' Hibrizi, Nurun Latifah, Rosalina, Fitri Bimantoro</i>	103
Pendeteksian Kecurangan Ujian Melalui CCTV Menggunakan Algoritma YOLOv5 <i>Fitri Bimantoro, I Gede Pasek Suta Wijaya, Mizanul Ridho Aohana</i>	109
Analisis Sentimen di Twitter: Mengungkap Persepsi dan Emosi Publik Seputar Konflik Palestina-Israel <i>Rizki Azhar, Muhammad Farid Wijayanto</i>	118
Klasifikasi Dan Pengenalan Pola Penyakit Cabai Dengan Metode CNN (Convolution Neural Network) <i>Bella Nurbuana Tri Cahya Ningrum, Erlina Nasrinatun Ni'mah, Miranda Putri Arifin, Made Ayu Dusea Widya Dara</i>	125
Implementasi Convolutional Neural Network (CNN) dalam Pengenalan Pola Penulisan Tangan <i>Salis Nilam Amartama, Alvi Nurul Hidayah, Putri Kartika Sari, Risky Aswi Ramadhani</i>	133
Pengembangan Media <i>Game</i> Bertema <i>Serious Game</i> Sebagai Pembelajaran Berbasis <i>Ice Breaking</i> <i>Candra Bagus Pratama</i>	139
Implementasi NLP(Natural Language Processing) Dasar pada Analisis Sentiment Review Spotify <i>Marsha Auriel Prasetya, Miftakhul Wulandari, Siti Alvi Nikmah</i>	145
Perancangan Desain Sistem Informasi Pariwisata Kota Kediri Menggunakan Aplikasi <i>MampirSek</i> <i>Muhamad Jallu Alfatih, Teddy Restyono, Muhamad Saiful Adi Ibrahim</i>	154
Perancangan Sistem Presensi Siswa Menggunakan Kartu Tanda Pelajar Dengan Penerapan Teknologi QR Code <i>Prima Sanubari, Budi Darmawan, M. Husain D.M</i>	162
Smart City, Konsep Kota Pintar Deteksi Objek Pada CCTV Lalu Lintas di Kota Nganjuk <i>Dicky Candra Zulkarnain, Ramadhan Bayu Aji, Burhanudin</i>	169
Aplikasi Pengontrolan Energi Berbasis Smart Plug, Stop Kontak Otomatis Pada Rumah Pintar (Smart Home) <i>M.Galihleo Yafan Dolar Febriliant, Muhammad Krishna Luthfi, Sadam Hidayatullah</i>	175
Perancangan Sistem Elektronik Pajak Bumi dan Bangunan <i>Ahmad Mudhofar Yusuf, Ahmad Fatkhur Rozi, Mohammad Aqil Muhaimin</i>	183
Perancangan Aplikasi <i>Smart Transportation Tracking</i> Bus Sekolah di Daerah Kabupaten Nganjuk Berbasis Android <i>Muhammad Setyo Budi, Nailusofa Al Mukhtari, Bayu Dwi Cahyono</i>	189

Perancangan E-arsip Disposisi surat Di Kecamatan Pace <i>Avif Bayu Saputra, Ahmad Ilham Aldiansyah, Farouk Ryan Hidayat</i>	199
Perancangan Sistem E-Arsip Pengolahan Data Klinik Nusa Medika Lestari <i>Rizky Nurwahyudi, Yoga Setya Adi Pradana, Muhammad Ikhbal Rokhmad</i>	205
Perancangan Sistem Informasi E-Surat Pengantar Desa Jatirejo Kecamatan Nganjuk <i>Dewi Zulaikah, Dela Karmeylia Putri, Tata Jeniarta</i>	212
Rancangan Sistem Tata Laksana Kependudukan Pada Kelurahan Warujayeng <i>Ainun Usnaini, Yuniswatin Nahdiyah, Duwita Yuli Harsasi</i>	221
Klasifikasi Penyakit Pneumonia Citra Digital X-Ray Menggunakan Metode <i>Convolutional Neural Network</i> dan <i>RGB Equalization</i> <i>Muhammad Attiqi Alghozali, Johan Rizky Triosaputra, Arwienda Kayan</i>	229
Deteksi Ketersediaan Lahan Parkir Dengan Menggunakan OpenCV <i>Anwar Muzaki, Tsalina Tsaniatul Mabruroh, Rahmad Ibrahim, Resty Wulaningrum</i>	237
Implementasi <i>Local Binary Pattern Histogram</i> Dalam <i>Multiple Face Recognition</i> <i>Rechtifano Microsofania R, M. Mukhlis Nurrahman S.A, Alvin Ardiansyah</i>	245
Penerapan Metode <i>Convolutional Neural Network</i> (CNN) Dalam Klasifikasi Penyakit Tanaman Jagung <i>Ary Yogyanto, Arfan Maulana, Diky Tri Cahyo A</i>	251
Klasifikasi Tingkat Kerusakan Kayu Menggunakan Metode <i>Convolutional Neural Network</i> (CNN) <i>Ricky Agung Sumiranto, Ika Maria Daniati, Anas Tasia</i>	257
Presensi Karyawan Menggunakan <i>Face Recognition</i> Untuk Meningkatkan Kinerja Karyawan <i>Raul Byntana Ardianto, Reza Naim Zakaria, Nandito Pramudya Arsyad</i>	263
Rancang Bangun Aplikasi Mobile Untuk Klasifikasi Jenis Ikan Koi Menggunakan Algoritma <i>Convolutional Neural Network</i> <i>Shandy Sadewa Asmoro, M. Farij Amrulloh, Moch. Anas Toybah, Muh Aris Saputra</i>	270
Klasifikasi Batik Menggunakan Algoritma CNN (<i>Convolutional Neural Network</i>) <i>Dias Nur Ramadhan, Rio Aldi Erwanto, Rony Tan Enwan</i>	278
Klasifikasi Tingkat Roasting Biji Kopi Dengan Metode CNN <i>Ilham Alfiantama, Michael Ilham Kresnawan, Andres Putra Handoko</i>	285
Implementasi Algoritma <i>Convolutional Neural Network</i> Arsitektur <i>Mobilenetv2</i> Untuk Klasifikasi Ekspresi Wajah Pada Dataset FER <i>Devfris Dhimas Permana Putra, Galang Kurnia Anaga, Wahyu Tia Fitriyana</i>	291
Implementasi CNN Arsitektur <i>Mobilenetv2</i> Untuk Klasifikasi Tulisan Aksara Jawa <i>Nindo Syafi'al Arief, Wahyu Anggara Putra, Dieky Septhian Rastra Pratama</i>	298
Implementasi <i>NLIDB</i> Pada <i>Chatbot</i> CV Owlsoft Media <i>Ilham Ainur Rohman, Daniel Swanjaya, Ardi Sanjaya</i>	304

SeBats : Sistem Deteksi Penyakit Diabetes Berbasis Android <i>Dhavis Alvi Chandra, Nando Rahmat Prasetyo Mulyo, Donny Firdani, Wahyu Cahyo Utomo</i>	310
Penerapan Metode Yolo V5 Dalam Mendeteksi Penyakit Tanaman Buah Naga <i>Qutrido antoko mohti, Rendy wahyudi, M habib Mustofa</i>	318
Sistem Pencatatan Barang Pada Toko Hesti Menggunakan Algoritma FIFO Berbasis Web <i>Refan Pahatsyah Iswitama, Patmi Kasih, Rony Heri Irawan</i>	324
Implementasi Deteksi Bahasa Isyarat Tangan Menggunakan OpenCV dan MediaPipe <i>Achmad Hasyim Nur'azizan, Abdul Riqza Ardiansyah, Rasio Fernandis</i>	331
Pemanfaatan Data Mining Untuk Memprediksi Kelulusan Mata Kuliah dan Referensi Strategi Pembelajaran <i>Andry Firdiansyah, Ibnu Al Ikrom, Moh. Khamdanni, Wahyu Cahyo Utomo</i>	338
Sistem Pakar Diagnosa Penyakit Kambing Menggunakan Metode <i>Case Based Reasoning</i> Untuk Kesehatan Ternak <i>Hikmah Tiar Alamsyah, Intan Nur Farida, Danang Wahyu Widodo</i>	345
Penerapan Algoritma <i>Binary Search</i> dan MD5 Pada Sistem Informasi Manajemen Klinik <i>Yulius Christanto, Rizki Saputro Mu'alim</i>	353
Deteksi Bahasa Isyarat Berdasarkan SIBI (Sistem Bahasa Isyarat) menggunakan <i>Transfer Learning</i> <i>M. Bahrul Subkhi, Mochamad Yuda Trinurais, Ridho Kuncoro Adji Wibowo, Bryan Rizqi Prakosa</i>	361
Sistem Bantu Penentuan Konsentrasi Mahasiswa Menggunakan Metode K-Nearest Neighbor Classification <i>Achmad Mukti Wibowo, Patmi Kasih, Intan Nur Farida</i>	370
Pembuatan Game 3D “HOLO-PAC” Dengan Menerapkan Algoritma A* <i>Hani Hermanto, Julian Sahertian, Ratih Kumalasari Niswatin</i>	380
Klasifikasi Penyakit Daun Sawi Hijau Dengan Metode CNN <i>Errin Dwi Ratnasari, Dhira Ananta Rudira, Anom Surya Buana</i>	388
Klasifikasi Tanaman Anggrek Menggunakan Metode CNN Berbasis Web Django <i>Mochammad Fedro Firdaus, Yodhi Pratama Iswoyo, Yansyah Nurullah Ahmadi</i>	394
Rancang Bangun Sistem Pakar Penentuan Penyakit Kambing Menggunakan Metode <i>Certainty Factor</i> <i>Septa Ardian Prayogi, Risa Helilintar, Intan Nur Farida</i>	404
Sistem Pendukung Keputusan Penentuan Kelayakan Pemberian Pinjaman Menggunakan Metode Smart <i>Defa Hangga Prawiratama, Risa Helilintar, Patmi Kasih</i>	412
Perancangan Sistem Pendukung Gaya Hidup Smart Lifestyle Untuk Meningkatkan Kualitas Hidup Di Masa Depan <i>Tito Pangestu, Nugroho Wisma Nurpanto, Azriel Akbar Firman Syah</i>	422

Perancangan Sistem Informasi Menejemen Surat Pada Prumda Tirta Musi Palembang Dengan Metode Waterfall <i>Ihsan sahidin, Muhamad son muare</i>	428
Analisis Hasil Perbaikan Citra Menggunakan Median Filter dan 2D Median Filter <i>Juli Sulaksono, Danang Wahyu Widodo, Ratih Kumalasari Niswatin</i>	438
Perancangan Sistem <i>E-Office</i> Pada Dinas Pendidikan Kota Kediri Untuk Peningkatan Pelayanan Administrasi <i>Al Dian, Moh farih fauzi</i>	444

Klasifikasi Jenis Daun Jambu Air dengan *Backpropagation*

Erry Anggraini¹, Nur Nafiyah²

^{1,2}Program Studi Teknik Informatika, Fakultas Teknik, Universitas Islam Lamongan, Indonesia
E-mail: ¹eryanggraini17@gmail.com, ²mynaff@unisla.ac.id

Abstrak – *Syzygium Aqueum* merupakan anggota keluarga jambu-jambuan (*Myrtaceae*) yang disebut sebagai jambu air. Akar, batang, daun, bunga, buah merupakan ciri bagian dari tanaman. Namun masih banyak masyarakat yang belum terlalu mengenali ciri jenis tanaman dari bagian daun. Kami membuat sistem untuk klasifikasi jenis daun jambu dengan metode *Backpropagation*. Fitur yang digunakan adalah ciri warna RGB (*meanR, meanG, meanB, stdR, stdG, stdB, varR, varG, varB, skewnessR, skewnessG, skewnessB*) dan ciri tekstur Gray Level Co-Occurrence Matrix (GLCM) (*contrast, energy, correlation, homogeneity*). Data yang digunakan sebanyak 367 citra yang dibagi menjadi 2 bagian yaitu 332 data training dan 35 data testing. Kami membuat 2 model arsitektur *Backpropagation*, yaitu (16-16-3) dan (16-32-3), fitur input sebanyak 16, dan hidden sesuai usulan, dan output ada tiga (Camplong, Cincalo Merah, Madu Deli Hijau). Dengan melakukan pengujian sebanyak 8 kali, setiap pengujian arsitektur *Backpropagation* dilakukan 4 kali pengujian. Hasil terbaik yang didapat menggunakan arsitektur 16-32-3 dengan tingkat akurasi sebesar 80%.

Kata Kunci — klasifikasi, *Backpropagation*, daun jambu air

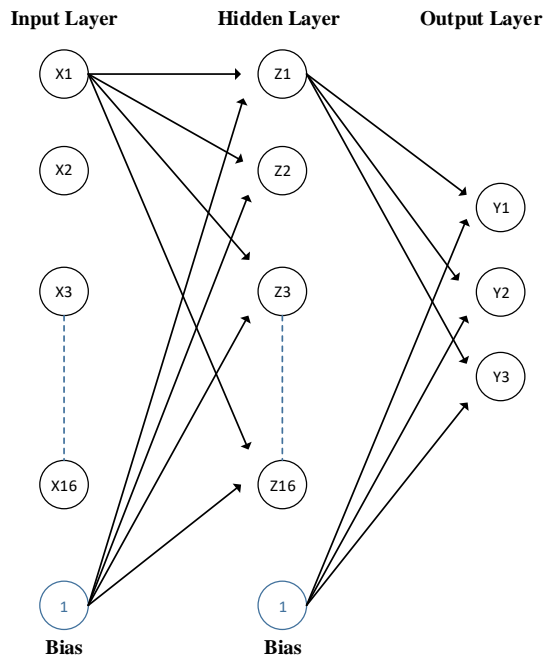
1. PENDAHULUAN

Di Indonesia, *Syzygium Aqueum* merupakan anggota keluarga jambu-jambuan (*Myrtaceae*) yang disebut sebagai jambu air. Banyak pembudidaya jambu air dengan berbagai macam jenis jambu air dari yang memiliki rasa sepat sampai rasa manis. Ada beberapa jenis jambu air seperti jambu air camplong, jambu air cincalo merah, dan jambu air madu deli hijau. Jenis tanaman ini termasuk spesies pohon favorit untuk ditanam sebagai peneduh rumah, biasanya ditanam di pekarangan rumah [1].

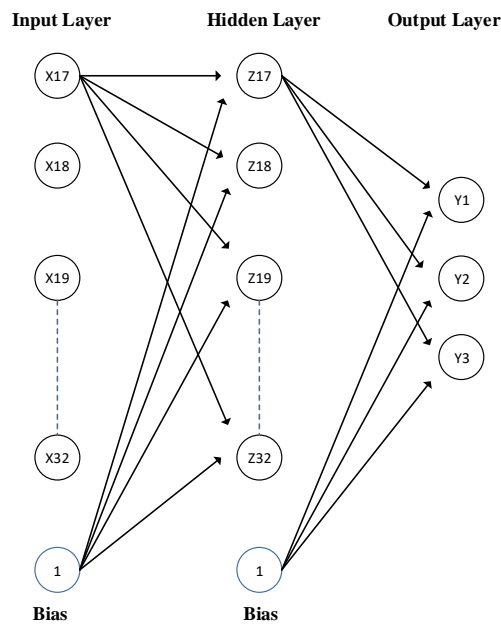
Dalam melakukan proses identifikasi tanaman memerlukan pemahaman untuk mengenali ciri karakteristik tanaman jambu air. Ciri karakteristik tanaman dapat dilihat pada bagian akar, batang, daun, bunga, buah. Namun masih banyak masyarakat yang belum terlalu mengenali ciri jenis tanaman dari bagian daun. Daun jambu air memiliki banyak ragam mulai dari bentuk, ukuran, tekstur, warna dan memiliki banyak kesamaan dari jenis jambu air lainnya. Sistem otomatis diperlukan untuk mengenali jenis jambu air berdasarkan fitur daun agar dapat memudahkan bagi masyarakat umum. Penelitian yang melibatkan klasifikasi atau identifikasi dengan menggunakan citra daun telah banyak dilakukan, diantaranya klasifikasi penyakit daun alpukat dengan metode CNN (Convolutional Neural Network) dengan hasil yang baik [2]. Beberapa penelitian lainnya terkait jambu air baik, yaitu identifikasi penyakit daun jambu untuk deteksi dini supaya menghasilkan buah yang berkualitas baik [3]. Penelitian terkait identifikasi jambu yang berulat dengan baik dengan metod deep learning, hasilnya mempunyai akurasi sampai 88% [4]. Penelitian klasifikasi mutu jambu baik dengan fitur tekstur ataupun warna [5], [6]. Penelitian terkait klasifikasi jenis jambu diantaranya berdasarkan tulang daun jambu [7], berdasarkan bentuk daun [8], berdasarkan fitur morfologi dari citra biner [9], [10], berdasarkan fitur tekstur GLCM (Gray Level Co-Occurrence Matrix) [11], fitur filter Gabor dengan metode k-NN (K-Nearest Neighbor) [12] dan ANFIS (Adaptive Neuro Fuzzy Inference System) [13]. Berdasarkan penelitian sebelumnya banyak yang melakukan klasifikasi jenis jambu dengan fitur GLCM dan metode klasifikasi *Neural Network* [11] akan tetapi mempunyai akurasi 78,89%. Sehingga Kami mengusulkan sistem klasifikasi jenis jambu agar bisa meningkatkan akurasi. Perbedaan penelitian Kami dengan sebelumnya [11] adalah jenis fitur dan arsitektur *Backpropagation*.

2. METODE PENELITIAN

Arsitektur jaringan adalah jenis arsitektur yang mengontrol bagaimana *neuron* diatur. Input layer, hidden layer dan output layer merupakan tiga jenis layer yang menyusun jaringan syaraf tiruan, di mana neuron dikumpulkan dalam layer yang dikenal sebagai layer neuron. Kami mengusulkan klasifikasi jenis jambu dengan metode *Backpropagation*. Kami membuat 2 model arsitektur *Backpropagation*. Model arsitektur *Backpropagation* yang digunakan 16 *hidden layer* pada Gambar 1 dan 32 *hidden layer* pada Gambar 2.



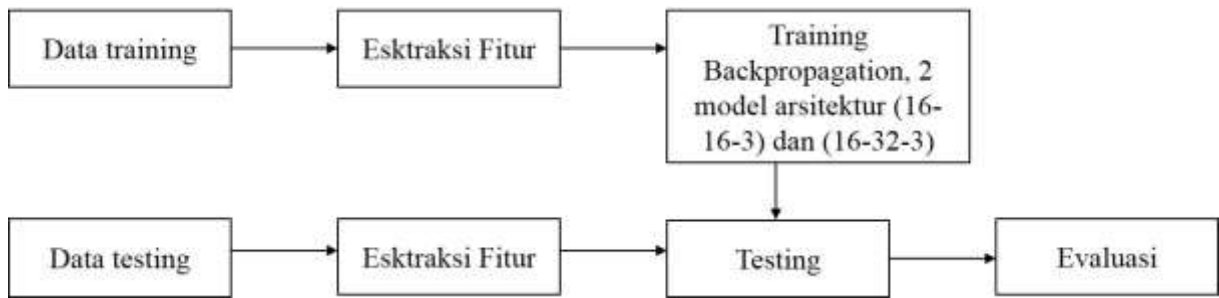
Gambar 1. Arsitektur 16-16-3



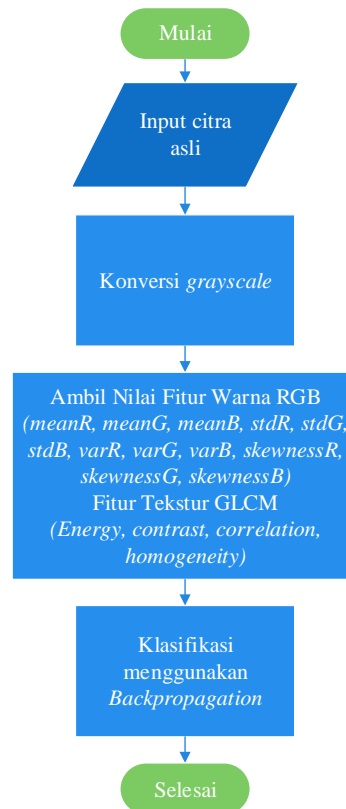
Gambar 2. Arsitektur 16-32-3

Pada Gambar 1 menunjukkan model arsitektur Backpropagation menggunakan 16-16-3. Pada Gambar 2 Menunjukkan model arsitektur Backpropagation menggunakan 16-32-3. Dengan kedua model arsitektur Backpropagation guna untuk memaksimalkan tingkat akurasi dalam melakukan klasifikasi. Tahapan seluruh proses klasifikasi jenis jambu terdapat ekstraksi fitur data training, data testing, training data metode Backpropagation dengan menggunakan 2 model arsitektur, kemudian uji coba data tes dan evaluasi. Gambar 3 merupakan proses keseluruhan penelitian dan Gambar 4 proses testing klasifikasi.

Gambar 3 menjelaskan data dibagi menjadi training dan testing, data training dan testing diekstraksi fitur *meanR*, *meanG*, *meanB*, *stdR*, *stdG*, *stdB*, *varR*, *varG*, *varB*, *skewnessR*, *skewnessG*, *skewnessB*, *energy*, *contrast*, *correlation*, *homogeneity*. Data fitur kemudian dilakukan pembuatan model *Backpropagation*, Kami menggunakan dua arsitektur *Backpropagation*. Kedua arsitektur dievaluasi akurasinya dalam mengklasifikasi jenis daun jambu.



Gambar 3. Tahapan Penelitian






Gambar 4. Tahap Testing

Gambar 4 menunjukkan tahap klasifikasi atau testing dengan melakukan penginputan citra yang akan dikonversi ke *grayscale* kemudian melakukan pengambilan nilai fitur warna RGB meliputi (*meanR, meanG, meanB, stdR, stdG, stdB, varR, varG, varB, skewnessR, skewnessG, skewnessB*) dan fitur tekstur GLCM meliputi (*energy, contrast, correlation, homogeneity*) kemudian proses klasifikasi jenis jambu. Dari Tabel 1 pada penelitian data citra yang digunakan sebagai sampel sejumlah 367 citra data yang masing-masing dibagi menjadi 2 data yaitu 332 data latih dan 35 data uji. Dengan 3 jenis kelas citra daun jambu air dengan contoh data daun jambu air seperti dalam Tabel 2.

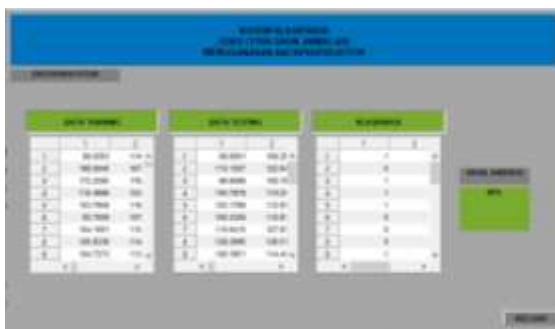
Tabel 1. Jumlah Data

No.	Jenis kelas citra	Data Training	Data Testing	Jumlah Data
1.	Camplong	110	10	120
2.	Cincalo Merah	102	15	117
3.	Madu Deli Hijau	120	10	130
	Total Data	332	35	367

Tabel 2. Citra Daun

Citra Daun	Jenis kelas Citra
	Camplong
	Cincalo Merah
	Madu Deli Hijau

3. HASIL DAN PEMBAHASAN



Gambar 5. Hasil Ekstraksi Fitur

Pada Gambar 5 merupakan ekstraksi fitur pengambilan nilai dari data *training* dan data *testing*. Fitur yang diambil nilai citra sebanyak 16 ciri dari ciri statistik warna RGB dan ciri warna tekstur GLCM kemudian melakukan klasifikasi untuk jenis kelas citra dan melakukan evaluasi agar menghasilkan akurasi.



Gambar 6. Hasil Klasifikasi

Pada Gambar 6 merupakan proses uji coba/testing dilakukan penginputan citra untuk dikonversi ke grayscale dari grayscale dilakukan pengambilan nilai citra yang diambil 16 ciri statistik warna RGB dan ciri statistik tekstur. Kemudian melakukan klasifikasi untuk menghasilkan jenis kelas daun jambu air. Berdasarkan Gambar 5 dan Gambar 6 menunjukkan proses klasifikasi jenis citra daun jambu air menggunakan *Backpropagation*, yang dimulai dalam proses ekstraksi fitur sampai dengan proses klasifikasi menghasilkan jenis jambu air, dan hasil evaluasi dari sistem klasifikasi.

Dari Tabel 3 proses uji coba sangat bervariasi, dan kualitas citra sangat mempengaruhi hasil klasifikasi. Kami melakukan pengujian sebanyak 8 kali dan di setiap masing-masing arsitektur *Backpropagation*. Hasil pengujian tingkat akurasi tertinggi dengan menggunakan arsitektur 16-32-3 bernilai sebesar 80%.

Tabel 3. Citra Daun

Uji ke-	Arsitektur	Akurasi
1	Hidden 16	68,57%
2	Hidden 16	65,71%
3	Hidden 16	62,86%
4	Hidden 16	71,43%
5	Hidden 32	60 %
6	Hidden 32	80%
7	Hidden 32	74,29%
8	Hidden 32	77,14%

4. SIMPULAN

Sistem klasifikasi jenis daun jambu air menggunakan *Backpropagation* ini menggunakan ciri statistik warna RGB yaitu (*meanR*, *meanG*, *meanB*, *stdR*, *stdG*, *stdB*, *varR*, *varG*, *varB*, *skewnessR*, *skewnessG*, *skewnessB*) dan ciri tekstur *Gray Level Co-Occurrence Matrix* (GLCM) (*contrast*, *energy*, *correlation*, *homogeneity*). Menggunakan 2 model arsitektur *Backpropagation* (16-16-3) dan (16-32-3). Daun jambu air sebanyak 367 yang dibagi menjadi 2 data yaitu 322 data latih dan 35 data uji yang dengan 3 jenis kelas citra yaitu camplong, cincalo merah dan madu deli hijau. Pengujian dilakukan sebanyak 8 kali, setiap pengujian arsitektur dilakukan 4 kali pengujian. Hasil tingkat akurasi tertinggi yang didapat sebesar 80% dari arsitektur 16-32-3.

5. SARAN

Melakukan klasifikasi beberapa jenis jambu air dengan berbagai macam lainnya dengan metode klasifikasi dan ekstraksi fitur yang lebih baik.

DAFTAR PUSTAKA

- [1] S. Wahyuni, "Sosialisasi Pemanfaatan Jambu Air Menjadi Nata De Syzigium," *Din. J. Pengabd. Kpd. Masy.*, vol. 4, no. 1, 2020, doi: 10.31849/dinamisia.v4i1.3285.
- [2] J. V. P. Putra, F. Ayu, and B. Julianto, "Implementasi Pendeteksi Penyakit pada Daun Alpukat Menggunakan Metode CNN," *Stain. (Seminar Nas. Teknol. Sains)*, vol. 2, no. 1, 2023.
- [3] W. F. Rischa *et al.*, "Pengenalan Pola Penyakit Daun Jambu Air," vol. 7, no. 2, pp. 158–163, 2023.
- [4] R. S. Pambudi and M. Akbar, "Klasifikasi Jambu Air Berulat menggunakan Convolutional Neural Network," vol. 1, no. 1, pp. 136–139, 2023.
- [5] T. Y. Prahudaya and A. Harjoko, "Metode Klasifikasi Mutu Jambu Biji Menggunakan k-NN Berdasarkan Fitur Warna dan Tekstur," *J. Teknosains*, vol. 6, no. 2, 2017, doi: 10.22146/teknosains.26972.
- [6] I. G. Wirayudhana, "Klasifikasi Mutu Buah Jambu Biji Getas Merah Berdasarkan Tekstur Menggunakan Grey Level Co-Occurrence Matrix (GLCM) dengan Klasifikasi KNN," *J. Indones. Sos. Teknol.*, vol. 2, no. 6, 2021, doi: 10.36418/jist.v2i6.166.
- [7] Y. Nursita and F. Rhozman, "Klasifikasi Jenis Jambu Air Berdasarkan Tulang Daun Menggunakan Metode K-Means," ... *Pgri Kediri.(online) simki. unpkediri. ac ...* 2016.
- [8] D. Wijaya, "Identifikasi Jenis Jambu Air Berdasarkan Bentuk Daun Menggunakan Jaringan Syaraf Tiruan Backpropagation dengan Fitur Histogram of Oriented Gradient," *J. Pendidik. dan Konseling*, vol. 4, no. 5, 2022.
- [9] I. Afrianty, R. Al Hafiz, F. Yanto, and E. P. Cynthia, "Klasifikasi Daun Jambu Air Menggunakan Ekstraksi Ciri Morfologi dan Backpropagation," *Semin. Nas. Teknol. Informasi, Komun. dan Ind. 12, Fak. Sains dan Teknol. UIN Sultan Syarif Kasim Riau*, no. SNTIKI, 2020.
- [10] R. Al Hafiz, "Penerapan Ekstraksi Ciri Morfologi dan Backpropagation Neural Network pada Klasifikasi Jenis Tumbuhan Jambu Air," Riau, 2020. [Online]. Available: <http://repository.uin-suska.ac.id/29295/>.

- [11] S. Suhendri and P. Rahayu, “Metode Grayscale Co-occurrence Matrix (GLCM) Untuk Klasifikasi Jenis Daun Jambu Air Menggunakan Algoritma Neural Network,” *J. Inf. Technol.*, vol. 1, no. 1, 2019, doi: 10.47292/joint.v1i1.4.
- [12] Y. R. Prayogi, “Klasifikasi Jenis Jambu Air Berdasarkan Citra Daun Menggunakan Metode Filter Gabor Dan K-NN (K-Nearest Neighbors),” *J. Pelita Teknol.*, vol. 17, no. 2, pp. 18–23, 2023.
- [13] F. Fitriani, “Implementasi Pengolahan Citra Digital dan Adaptive Neuro Fuzzy Inference System untuk Klasifikasi Tanaman Jambu Air,” Riau, 2022. [Online]. Available: <http://repository.uin-suska.ac.id/58504/>.

Implementasi Metode k-Nearest Neighbor (k-NN) untuk Memprediksi Penjualan Buah di Indonesia berbasis Website

Siti Mujilahwati ¹, Lilis Devita Windasari ²

^{1,2}Teknik Informatika, Fakultas Teknik, Universitas Islam Lamongan
E-mail: ¹moedjee@unisla.ac.id, ²lilisdevitawindasari1901@gmail.com

Abstrak – Badan Pusat Statistik (BPS) merekam data penjualan buah di Indonesia setiap tahun sebagai laporan dan sumber informasi bagi masyarakat. Sayangnya, meskipun data ini disediakan, masih banyak pelaku bisnis yang tidak memanfaatkannya secara optimal. Data penjualan buah memiliki nilai sebagai indikator untuk mengarahkan produksi suatu produk; semakin tinggi penjualan, semakin penting persiapan stok yang matang. Untuk mengatasi tantangan ini, perlu adanya sistem yang mampu melakukan prediksi penjualan buah untuk membantu mengelola stok dan produksi dengan efektif. Dalam studi ini, telah diimplementasikan metode K-Nearest Neighbors (KNN) sebagai alat untuk memprediksi penjualan buah. Metode ini terbukti mampu memberikan hasil yang memuaskan, dengan tingkat akurasi mencapai 70.93%. Prediksi yang dihasilkan menunjukkan peningkatan penjualan pada tahun 2023. Implementasi metode ini memberikan kontribusi positif dalam membantu pelaku bisnis untuk mengambil keputusan yang lebih tepat dalam manajemen stok dan produksi buah, sehingga dapat mengoptimalkan kinerja bisnis mereka.

Kata Kunci — Data mining, Prediksi penjualan buah, K-Nearest Neighbors (KNN)

1. PENDAHULUAN

Prediksi adalah proses meramalkan nilai atau kejadian di masa depan berdasarkan informasi yang sudah ada [1][2]. Ini melibatkan penggunaan data historis atau pola untuk membuat perkiraan tentang apa yang mungkin terjadi selanjutnya. Prediksi dapat digunakan dalam berbagai bidang, seperti ilmu pengetahuan data, analisis keuangan, peramalan cuaca, dan banyak lagi [3]–[5].

Metode KNN (K-Nearest Neighbors) adalah salah satu metode prediksi yang umum digunakan dalam pembelajaran mesin, khususnya pada tugas klasifikasi dan regresi. KNN bekerja dengan cara membandingkan suatu data poin dengan data poin terdekat dalam ruang fitur. Ide dasar di balik KNN adalah bahwa entitas yang serupa cenderung berada dekat satu sama lain [6]. Diberbagai bidang KNN telah dibuktikan untuk melakukan klasifikasi dan prediksi, menunjukkan kinerja yang baik [7], [8].

Berdasarkan laporan Badan Pusat Statistik (BPS) hasil produksi buah di Indonesia setiap tahun mengalami naik turun sedangkan di Indonesia memiliki banyak varian jenis buah. Hasil penjualan dari buah ini akan sangat berpengaruh pada kebutuhan produksi buah itu sendiri. Akan tetapi informasi jumlah penjualan buah dari data tersebut masih kurang informatif, khususnya untuk informasi pada penjualan buah di tahun berikutnya. Hal ini mengakibatkan tidak efisiensi pada produksi buah yang dibutuhkan pada tahun berikutnya. Untuk itu guna membantu meningkatkan informasi dari laporan BPS tersebut, dari laporan naik atau turun dari berbagai jenis buah di Indonesia, maka dibutuhkan sebuah metode yang mampu melakukan prediksi penjualan sebagai kesimpulan laporan untuk mengetahui produksi yang harus terpenuhi pada periode berikutnya. Oleh dari itu pada penelitian ini bertujuan mengimplementasikan metode KNN untuk meningkatkan informasi melalui prediksi penjualan buah di Indonesia.

2. METODE PENELITIAN

Penelitian ini menggunakan beberapa tahap untuk mendapatkan hasil yang maksimal, dimulai dari analisis kebutuhan dengan identifikasi masalah, tahap berikutnya pengumpulan data dan selanjutnya Analisa Sistem, terakhir melakukan implementasi metode pada sistem prediksi dan validasi hasil prediksi. Kebutuhan analisis sistem di gambarkan seperti pada Gambar 1 berikut.



Gambar 1. Alur Penelitian

Pada tahap analisis kebutuhan dilakukan analisis fitur yang meliputi pencatatan data pengguna, fitur untuk mengelola data prediksi penjualan buah, data buah, data training, data uji, laporan dan user. Kebutuhan fungsional sistem prediksi ini adalah melihat data *forecasting* atau ramalan penjualan buah kemudian bisa melakukan CRUD (*createl, relad update, delete*) penjualan. Selain itu terdapat fitur laporan prediksi dan *forecasting* yang dapat di export ke dalam format excel. Selain kebutuhan fungsional tersebut terdapat pula fitur *login* dan *logout* untuk membatasi pengguna yang bisa akses ke dalam sistem.

Tahap pengumpulan data, data yang digunakan pada penelitian ini dari Badan Pusat Statistik (BPS) dari tahun 2019 sampai tahun 2022. Tabel 1 berikut adalah data dari tahun 2019 sampai 2022 yang digunakan sebagai data training atau data pelatihan.

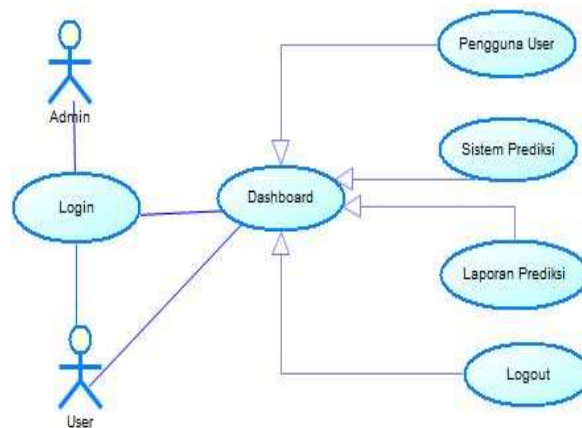
Tabel 1. Data Penjualan Buah di Indonesia

No	Alpukat/Provinsi	2019	2020	2021
1	Aceh	11576	20609	44484
2	Sumatera Utara	18525	32012	35378
3	Sumatera Barat	54204	69787	84083
4	Riau	965	1179	1714
5	Jambi	4892	9817	10489
6	Sumatera Selatan	38266	36343	15927
7	Bengkulu	6470	6282	6877
8	Lampung	18502	17304	24008
9	Kep. Bangka Belitung	491	839	1509
10	Kep. Riau	114	230	93
11	Dki Jakarta	998	1491	1698
12	Jawa Barat	104084	104665	110046
13	Jawa Tengah	60145	73933	75707
14	Di Yogyakarta	2385	2304	10931
15	Jawa Timur	101310	175735	158581
16	Banten	2047	2019	1444
17	Bali	2452	2815	5123
18	Nusa Tenggara Barat	4189	19260	8889
19	Nusa Tenggara Timur	11682	13870	34638
20	Kalimantan Barat	504	687	1667
21	Kalimantan Tengah	194	176	895
22	Kalimantan Selatan	42	29	58
23	Kalimantan Timur	1719	1148	674
24	Kalimantan Utara	105	119	85
25	Sulawesi Utara	4017	3317	8697
26	Sulawesi Tengah	2945	2131	10478
27	Sulawesi Selatan	5159	6795	8972
28	Sulawesi Tenggara	760	1175	1093
29	Gorontalo	21	24	24

No	Alpukat/Provinsi	2019	2020	2021
30	Sulawesi Barat	637	503	312
31	Maluku	806	1044	2143
32	Maluku Utara	931	878	1061
33	Papua Barat	339	400	575

Kemudian dari data tersebut akan dimasukkan pada rumus jarak dimana variable tahun akan menjadi variable X, untuk data satu jenis buah (misal : Alpukat) dari tahun 2019 sampai 2021 sebanyak 96 record, pada penelitian ini memprediksi 10 jenis buah yaitu Alpukat, Belimbing, Mangga, Semangka, Pisang, Manggis, Salak, Pepaya, Rambutan dan Melon.

Gambar 2 menunjukkan Use Case Diagram yang merupakan pemodelan untuk fungsi-fungsi apa saja yang bisa dilakukan dan siapa saja yang akan menggunakan sistem atau perangkat lunak yang dibuat. Secara garis besar, bisnis proses sistem yang dirancang digambarkan dengan use case diagram.



Gambar 2. Use Case Diagram

Aplikasi prediksi yang dibuat ditunjukkan pada gambar 2. Mulai dari login masuk kedalam aplikasi yang selanjutnya ditampilkan sebuah dashboard yang berisi menu informasi pengguna, form prediksi yang didalamnya terdapat isian jenis buah yang akan dilakukan prediksi, form pelaporan yang berisi hasil semua prediksi yang diproses dan selanjutnya akan digunakan sebagai informasi penjualan buah dan prediksi penjualan buah pada tahun berikutnya sebagai wacana kebutuhan produksi buah tersebut. Sistem prediksi yang dibangun menerapkan metode KNN sebagai berikut [7], [9], [10]:

1. Memilih Nilai K (Jumlah Tetangga Terdekat): Langkah pertama dalam menggunakan KNN adalah memilih nilai K, yaitu jumlah tetangga terdekat yang akan digunakan untuk membuat prediksi. Nilai K ini bisa menjadi faktor kritis dan dapat mempengaruhi kinerja model. Penelitian ini menggunakan nilai K= 15.
2. Mengukur Jarak: KNN menggunakan metrik jarak, seperti Euclidean distance atau Manhattan distance, untuk mengukur seberapa dekat atau jauh suatu data poin dengan data poin lainnya dalam ruang fitur.

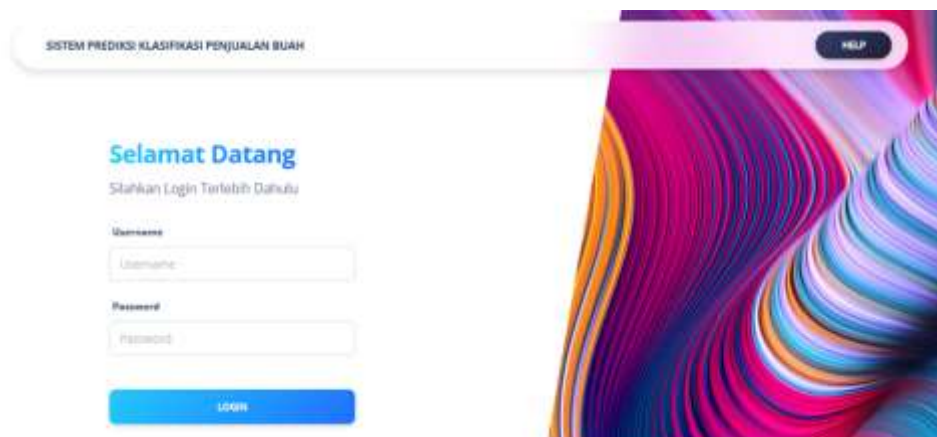
$$\text{Rumus distance : } d(a, b) = \sum_{i=0}^n (X_i - Y_i)^2 \dots\dots\dots 1)$$

3. Menentukan Tetangga Terdekat: Setelah mengukur jarak, KNN menemukan K tetangga terdekat dari data poin yang ingin diprediksi.
4. Menyusun Voting atau Menghitung Rata-rata: Untuk tugas klasifikasi, prediksi dilakukan dengan melakukan voting mayoritas dari kelas tetangga terdekat. Sedangkan untuk tugas regresi, prediksi dapat dilakukan dengan mengambil rata-rata nilai dari tetangga terdekat.

3. HASIL DAN PEMBAHASAN

Penelitian ini menghasilkan sebuah aplikasi prediksi yang berbasis website. Sistem ini dirancang untuk memberikan penggunaan yang mudah dan intuitif. Berikut adalah gambaran visual dari antarmuka sistem prediksi yang telah dibuat, mencerminkan keseluruhan tata letak dan desain aplikasi. Melalui tampilan visual ini, pengguna dapat dengan jelas melihat bagaimana aplikasi ini disusun dan diorganisir, memberikan gambaran yang komprehensif tentang fitur-fitur yang tersedia dan cara navigasi dalam sistem prediksi tersebut. Dengan

demikian, hasil penelitian ini dapat diakses melalui antarmuka website yang user-friendly, memberikan kemudahan bagi pengguna untuk menggunakan dan memanfaatkan aplikasi prediksi ini dalam berbagai konteks.



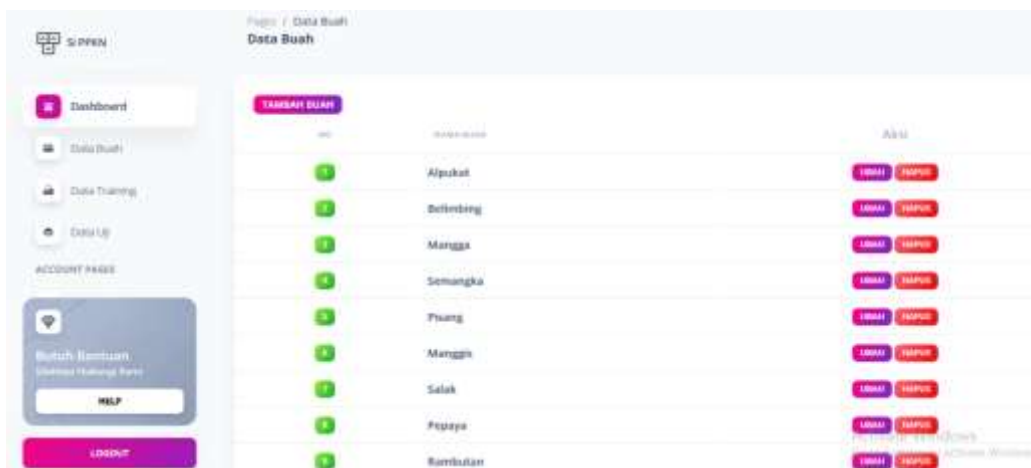
Gambar 3. Halaman Login

Gambar 3 menggambarkan sebuah halaman login yang dapat diakses baik oleh administrator maupun pengguna dalam konteks sistem prediksi. Di halaman ini, seorang administrator diwajibkan untuk mengisi kolom username, yang digunakan untuk mengidentifikasi dan memasukkan pengguna admin yang telah terdaftar dalam database. Selain itu, terdapat pula kolom password yang harus diisi dengan kata sandi yang sesuai. Selanjutnya, terdapat menu tombol login yang berfungsi sebagai akses untuk memasuki halaman utama dari sistem prediksi. Dengan demikian, Gambar 3 menyajikan tampilan antarmuka yang memungkinkan pengguna untuk melakukan otentikasi, melalui kombinasi username dan password, guna mengakses fungsionalitas lengkap sistem prediksi tersebut.



Gambar 4. Dashboard

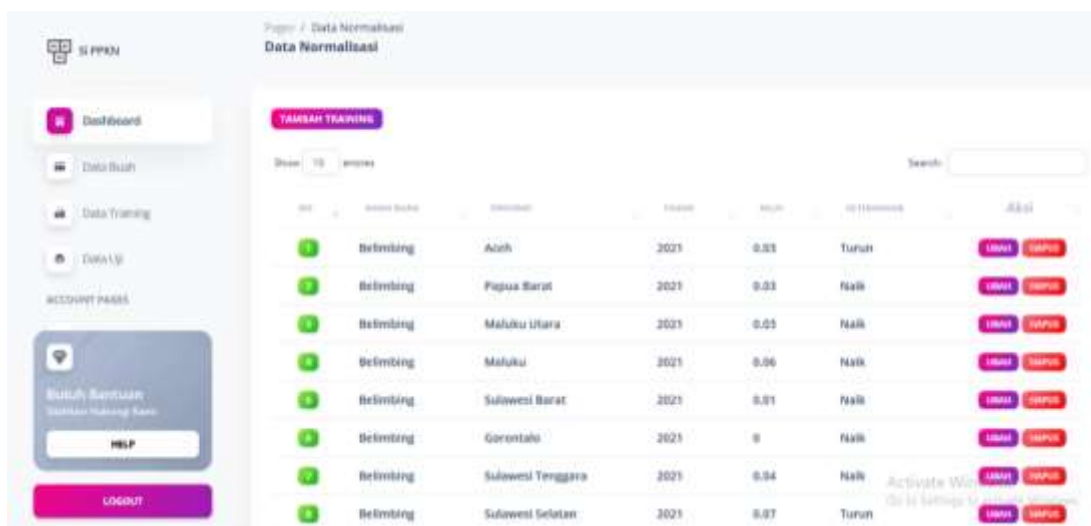
Gambar 4 memberikan penjelasan terkait tampilan halaman utama atau dashboard pada sistem. Halaman ini akan memperlihatkan gambaran umum, termasuk informasi tentang jumlah buah dan jumlah pengguna yang terdaftar dalam sistem. Dengan adanya halaman dashboard, pengguna dapat dengan cepat melihat dan memantau statistik terkait jumlah buah yang tersedia dan jumlah pengguna yang aktif. Ini menciptakan pengalaman pengguna yang informatif dan mempermudah akses cepat ke data penting mengenai inventaris buah dan jumlah pengguna dalam lingkungan sistem.



ID	Nama Buah	Aksi
1	Alpukat	UBAH HAPUS
2	Belimbing	UBAH HAPUS
3	Mangga	UBAH HAPUS
4	Semangka	UBAH HAPUS
5	Puang	UBAH HAPUS
6	Mangga	UBAH HAPUS
7	Selak	UBAH HAPUS
8	Pepaya	UBAH HAPUS
9	Bambutan	UBAH HAPUS

Gambar 5. Data Buah

Pada Gambar 5, terlihat tampilan data buah yang mencakup informasi mengenai berbagai jenis buah. Antarmuka ini dirancang untuk memberikan pemahaman yang jelas terkait dengan berbagai jenis buah yang ada, disertai dengan fitur tombol yang memfasilitasi manajemen data buah tersebut. Terdapat opsi untuk menambahkan data buah baru, mengubah informasi yang sudah ada, serta menghapus data buah dari sistem. Dengan adanya tombol-tombol tersebut, pengguna dapat dengan mudah melakukan penambahan, pengeditan, atau penghapusan data buah sesuai kebutuhan. Ini memberikan fleksibilitas dan kontrol penuh kepada pengguna dalam mengelola dan memelihara informasi mengenai jenis-jenis buah yang terdapat dalam sistem. Dengan demikian, tampilan data buah pada Gambar 5 memberikan aksesibilitas yang optimal dan kemudahan dalam pengelolaan data terkait buah-buahan dalam aplikasi prediksi ini.



ID	Nama Buah	Lokasi	Tahun	Nilai	Aksi
1	Belimbing	Aceh	2021	0.81	Turun
2	Belimbing	Papua Barat	2021	0.83	Naik
3	Belimbing	Maluku Utara	2021	0.85	Naik
4	Belimbing	Maluku	2021	0.86	Naik
5	Belimbing	Sulawesi Barat	2021	0.81	Naik
6	Belimbing	Gorontalo	2021	0	Naik
7	Belimbing	Sulawesi Tenggara	2021	0.84	Naik
8	Belimbing	Sulawesi Selatan	2021	0.87	Turun

Gambar 6. Data Training

Gambar 6 menampilkan tampilan dari data training yang digunakan dalam sistem. Data training ini memiliki karakteristik khusus karena telah dilabeli sebelumnya dengan kelas "Naik" dan "Turun" sesuai dengan jumlah penjualan yang diprediksi untuk tahun berikutnya. Dengan demikian, data training ini berfungsi sebagai set pelatihan untuk mengajarkan model dalam sistem untuk mengenali pola dan korelasi antara berbagai faktor yang mempengaruhi perubahan jumlah penjualan. Pentingnya data training dengan label "Naik" dan "Turun" adalah untuk memungkinkan model melakukan prediksi dengan menghubungkan karakteristik tertentu dalam data dengan perubahan arah penjualan. Dengan memanfaatkan data training ini, sistem dapat belajar dan mengoptimalkan prediksinya, menghasilkan hasil yang lebih akurat ketika dihadapkan pada data penjualan yang belum terlihat sebelumnya. Dengan demikian, Gambar 6 memberikan pandangan terinci mengenai kualitas dan struktur data training yang merupakan dasar dari kemampuan prediktif sistem ini.



Gambar 7. Data Uji

Gambar 7 menunjukkan antarmuka yang berkaitan dengan data uji buah dalam sistem, yang nantinya akan menjadi objek prediksi. Proses ini melibatkan pemilihan data uji tertentu, di mana hasil prediksi terkait dengan apakah penjualan buah tersebut diperkirakan akan "Naik" atau "Turun". Setelah mendapatkan hasil prediksi, pengguna diberikan opsi untuk memasukkan data uji beserta hasil prediksinya ke dalam database. Proses klik pada tombol "Masukkan ke Database" memungkinkan integrasi hasil prediksi ke dalam basis data, sehingga data tersebut dapat diakses dan dikelola secara terpusat. Ini menciptakan catatan historis yang berguna untuk analisis lebih lanjut, serta memfasilitasi pemantauan dan evaluasi performa sistem prediksi terhadap data aktual. Dengan demikian, Gambar 7 memberikan gambaran visual tentang bagaimana sistem ini tidak hanya melakukan prediksi terhadap data uji buah, tetapi juga memberikan fungsionalitas untuk menyimpan dan merekam hasil prediksi tersebut ke dalam database, menciptakan keseluruhan siklus operasional yang terintegrasi.



Gambar 8. Laporan

Gambar 8 menampilkan halaman laporan yang berfungsi sebagai wadah untuk menampilkan hasil prediksi tahun 2022 dari sistem. Pada halaman laporan ini, terdapat informasi terinci mengenai 10 buah yang telah diprediksi, yang disertai dengan nilai "Naik" dan "Turun" yang berbeda-beda untuk setiap buah tersebut. Hasil prediksi ini mencerminkan evaluasi sistem terhadap data masukan dan kemampuannya dalam mengantisipasi perubahan tren penjualan. Dengan memberikan nilai "Naik" dan "Turun" untuk setiap buah, halaman laporan ini memberikan gambaran komprehensif tentang proyeksi pertumbuhan atau penurunan dalam penjualan buah-buahan pada tahun 2022. Informasi yang disajikan di halaman laporan ini dapat menjadi dasar untuk pengambilan keputusan dan perencanaan strategi dalam konteks bisnis atau manajemen persediaan.

Dengan demikian, Gambar 8 memberikan insight visual yang penting tentang hasil prediksi yang diperoleh melalui sistem prediksi yang telah dijelaskan sebelumnya, dan hal ini dapat digunakan sebagai dasar untuk analisis dan pengambilan keputusan lebih lanjut.

Tabel 2. Hasil Prediksi Buah

Buah	Faktual		Prediksi	
	Naik	Turun	Naik	Turun
Alpukat	16	18	13	21
Belimbing	24	10	23	11
Mangga	9	25	18	16

Buah	Faktual		Prediksi	
	Naik	Turun	Naik	Turun
Semangka	25	9	17	17
Pisang	7	27	20	14
Manggis	13	21	27	7
Salak	17	17	30	4
Pepaya	20	14	33	1
Rambutan	18	16	27	7
Melon	19	15	19	15

Tabel 2 di atas menunjukkan hasil prediksi penjualan buah menggunakan metode K-Nearest Neighbors (K-NN). Dengan nilai ketersediaan tetangga (K) sebesar 15, hasil prediksi penjualan buah di Indonesia dapat diuraikan sebagai berikut:

1. Buah Alpukat diprediksi memiliki nilai "Naik" sebanyak 13 dan "Turun" sebanyak 21.
2. Buah Belimbing diprediksi memiliki nilai "Naik" sebanyak 23 dan "Turun" sebanyak 11.
3. Buah Mangga diprediksi memiliki nilai "Naik" sebanyak 18 dan "Turun" sebanyak 16.
4. Buah Semangka diprediksi memiliki nilai "Naik" sebanyak 17 dan "Turun" sebanyak 17.
5. Buah Pisang diprediksi memiliki nilai "Naik" sebanyak 20 dan "Turun" sebanyak 14.
6. Buah Manggis diprediksi memiliki nilai "Naik" sebanyak 27 dan "Turun" sebanyak 7.
7. Buah Salak diprediksi memiliki nilai "Naik" sebanyak 30 dan "Turun" sebanyak 4.
8. Buah Pepaya diprediksi memiliki nilai "Naik" sebanyak 33 dan "Turun" sebanyak 1.
9. Buah Rambutan diprediksi memiliki nilai "Naik" sebanyak 27 dan "Turun" sebanyak 7.
10. Buah Melon diprediksi memiliki nilai "Naik" sebanyak 19 dan "Turun" sebanyak 15.

Analisis terhadap hasil prediksi menunjukkan bahwa jumlah nilai faktual dan prediksi tidak terlalu berbeda jauh. Oleh karena itu, dapat disimpulkan bahwa beberapa buah seperti Manggis, Salak, Pepaya, dan Rambutan memiliki kecenderungan peningkatan penjualan berdasarkan hasil prediksi. Analisis ini memberikan wawasan yang berharga untuk perencanaan strategis dalam manajemen persediaan atau kebijakan penjualan di masa mendatang.

4. SIMPULAN

Sistem prediksi penjualan buah di Indonesia yang telah berhasil dikembangkan merupakan sebuah platform berbasis website yang melibatkan dua aktor utama, yaitu admin dan user. Kedua aktor tersebut memiliki akses untuk melakukan tindakan login, input dan penghapusan data, serta melibatkan diri dalam proses prediksi hasil penjualan. Proses prediksi dilaksanakan setelah admin menginput data training, yang nantinya akan menghasilkan proyeksi naik dan turun dalam penjualan. Metode yang digunakan untuk prediksi ini adalah K-Nearest Neighbor (K-NN). Hasil uji coba sistem menunjukkan keberhasilan dalam mencapai hasil prediksi yang sesuai dengan perhitungan manual, dengan menggunakan nilai ketetanggaan (K) sebesar 15. Dengan rincian prediksi untuk beberapa buah, seperti Alpukat (naik=13, turun=21), Belimbing (naik=23, turun=11), Mangga (naik=18, turun=16), Semangka (naik=17, turun=17), Pisang (naik=20, turun=14), Manggis (naik=27, turun=7), Salak (naik=30, turun=4), Pepaya (naik=33, turun=1), Rambutan (naik=27, turun=7), dan Melon (naik=19, turun=15). Analisis terhadap hasil prediksi menunjukkan kesesuaian yang baik antara nilai faktual dan prediksi, dengan perbedaan yang tidak signifikan. Beberapa buah menunjukkan kecenderungan peningkatan dan penurunan penjualan sesuai dengan hasil prediksi. Ini memberikan kontribusi positif terhadap pengambilan keputusan di tingkat manajemen, terutama dalam perencanaan dan strategi penjualan di masa mendatang. Dengan demikian, sistem prediksi ini berhasil memenuhi tujuan analisis dan desain sistem yang telah ditetapkan sebelumnya.

5. SARAN

Berdasarkan hasil penelitian dan pengujian aplikasi yang terdokumentasikan dalam penyusunan skripsi ini, terdapat beberapa saran yang dapat dijadikan pertimbangan untuk penelitian selanjutnya:

1. Diversifikasi Bahasa Pemrograman dan Platform
Meningat beragamnya teknologi dan platform yang tersedia, disarankan untuk melakukan penelitian selanjutnya dengan mempertimbangkan penggunaan bahasa pemrograman dan platform yang berbeda. Selain PHP dan website, penelitian dapat memperluas cakupan ke platform mobile seperti Android atau menggunakan bahasa pemrograman yang lebih spesifik untuk kebutuhan tertentu.
2. Penggunaan Metode Prediksi Alternatif

Penelitian selanjutnya dapat menjelajahi berbagai metode prediksi selain K-Nearest Neighbor (K-NN). Penggunaan metode-metode lain seperti regresi, jaringan saraf (neural networks), atau algoritma machine learning yang lebih kompleks dapat memberikan wawasan yang lebih mendalam terhadap prediksi penjualan.

3. Pengembangan dan Pemeliharaan Sistem

Agar sistem prediksi dapat tetap relevan dan efektif, diperlukan pengembangan dan pemeliharaan rutin. Pengembangan dapat mencakup penambahan fitur baru, peningkatan performa, serta adaptasi terhadap perubahan kebutuhan bisnis atau teknologi. Selain itu, perawatan sistem secara rutin juga penting untuk memastikan keberlanjutan fungsionalitasnya.

4. Pertimbangkan Aspek Keamanan

Dalam penelitian mendatang, perhatikan keamanan sistem dengan lebih cermat. Pastikan bahwa data yang dimasukkan dan diakses oleh sistem terlindungi dengan baik dari potensi ancaman keamanan, termasuk perlindungan terhadap akses tidak sah.

5. Melibatkan Pengguna Akhir

Sebagai tambahan, melibatkan pengguna akhir (admin dan user) dalam tahap perancangan dan pengembangan dapat membantu memastikan bahwa sistem yang dibangun memenuhi kebutuhan dan ekspektasi mereka dengan lebih baik

Dengan menerapkan saran-saran ini, diharapkan penelitian selanjutnya dapat memberikan kontribusi yang lebih besar dalam pengembangan sistem prediksi penjualan buah dan memenuhi kebutuhan yang berkembang dalam dunia teknologi informasi.

DAFTAR PUSTAKA

- [1] L. Suryadi, N. E. Pratiwi, F. Ardhy, and ..., "PENERAPAN DATA MINING PREDIKSI PENJUALAN MEBEL TERLARIS MENGGUNAKAN METODE K-NEAREST NEIGHBOR (K-NN)(STUDI KASUS: TOKO ZERITA ...," *JUSIM (Jurnal Sistem ...*, 2022.
- [2] S. Mujilawati, N. Qomariyah Nawafilah, and M. Aliyudin, "ANALISIS HASIL PREDIKSI DENGAN METODE PROMETHEE," *Jurnal Mnemonic*, vol. 2, no. 1, 2019, doi: 10.36040/mnemonic.v2i1.49.
- [3] P. Studi, T. Informatika, and U. A. Yogyakarta, "Optimasi Parameter Artificial Neural Network Menggunakan Algoritma Genetika Untuk Prediksi Kelulusan Mahasiswa," vol. 18, no. 1, pp. 54–59, 2019.
- [4] M. Yunus and N. K. A. Pratiwi, "Prediksi Status Gizi Balita Dengan Algoritma K-Nearest Neighbor (KNN) di Puskesmas Cakranegara," *JTIM: Jurnal Teknologi Informasi dan Multimedia*, vol. 4, no. 4, 2023, doi: 10.35746/jtim.v4i4.328.
- [5] N. Arifin, R. Heri Irawan, I. Nur Farida, K. Kunci -Penumpukan Stok Bahan Baku, P. Stok Bahan Baku, and A. K-Means, "Algoritma K-Means Untuk Memprediksi Stok Bahan Baku Produksi," *Universitas Nusantara PGRI Kediri. Kediri*, vol. 1, 2022.
- [6] M. Kafil, "Penerapan Metode K-Nearest Neighbors," *Jurnal Mahasiswa Teknik Informatika (JATI)*, vol. 3, no. 2, 2019.
- [7] R. D. Probo, B. Irawan, and R. Rumani M, "Analisis dan Implementasi Perbandingan Algoritma KNN (K-Nearest Neighbor) Dengan SVM (Support Vector Machine) untuk Prediksi Penawaran Produk," *eProceedings of Engineering*, vol. 3, no. 3, 2016.
- [8] J. Maulani and M. Sari, "Komparasi Metode K-Nearest Neighbor (Knn) Dengan Support Vector Machine (Svm) Terhadap Tingkat Akurasi Klasifikasi Kualitas Air," *Smart Comp: Jurnalnya Orang Pintar Komputer*, vol. 12, no. 2, 2023, doi: 10.30591/smartcomp.v12i2.4205.
- [9] M. Ruswandi, D. I. Mulyana, and A. Awaludin, "Optimasi Klasifikasi Kematangan Buah Alpukat Menggunakan KNN dan Fitur Statistik," *Smart Comp :Jurnalnya Orang Pintar Komputer*, vol. 11, no. 2, 2022.
- [10] J. Han and M. Kamber, *Data Mining Concepts And Techniq*

Implementasi Algoritma Naive Bayes pada Sistem Pendukung Keputusan Rekomendasi Wisata di Lamongan

Indra Dwi Rizqi¹, Retno Wardhani², Moh. Rosidi Zamroni³

^{1,2,3}Teknik Informatika, Fakultas Teknik, Universitas Islam Lamongan

E-mail: ¹indradwirizqi99@gmail.com, ²retnowardhani@unisla.ac.id, ³rosidizamroni@unisla.ac.id

Abstrak – Kabupaten Lamongan di Jawa Timur memiliki potensi wisata yang beragam, mencakup alam, buatan, budaya, dan religi. Total wisatawan dari 22 destinasi di Kabupaten Lamongan dari tahun 2015-2021 mencapai 19.167.971 jiwa, dengan mayoritas berasal dari wisatawan nusantara. Dalam menghadapi beragam pilihan destinasi, calon wisatawan seringkali menghadapi dilema dalam memilih tempat wisata yang sesuai. Oleh karena itu, artikel ini mengusulkan solusi berupa pemanfaatan sistem pendukung keputusan. Sistem ini menggunakan algoritma Naive Bayes, sebuah metode probabilistik yang dapat mengatasi data dimensi tinggi dan efisien dalam mengklasifikasikan data besar. Metode penelitian melibatkan pengumpulan data dari Dinas Pariwisata Kabupaten Lamongan dan Google Map, mencakup informasi seperti lokasi, jam buka, harga tiket, fasilitas, ulasan Google Map, jarak, dan kebersihan. Kriteria yang digunakan dalam sistem pendukung keputusan mencakup waktu buka, fasilitas, harga masuk, jarak, kebersihan, dan bintang wisata. Dalam pengujian, sistem ini memiliki akurasi sebesar 94%.

Kata Kunci — Sistem Pendukung Keputusan, Metode Naive Bayes, Rekomendasi Wisata Lamongan

1. PENDAHULUAN

Setelah periode sulit yang disebabkan oleh pandemi COVID-19, berwisata atau traveling kembali muncul sebagai kebutuhan utama bagi banyak individu. Namun, perjalanan wisata saat ini melibatkan perencanaan matang dan penyesuaian baru dalam gaya perjalanan. Terdapat beberapa alasan yang mendorong orang-orang untuk merencanakan perjalanan ini, termasuk hasrat untuk mengembalikan kehidupan normal, eksplorasi destinasi yang telah lama ditunggu-tunggu, dan mengambil kembali kendali atas waktu luang mereka. Setelah berbulan-bulan pembatasan perjalanan dan isolasi, banyak orang merindukan pengalaman berlibur yang dapat memulihkan semangat mereka. Perjalanan ini juga menjadi peluang untuk menjelajahi budaya orang lain, menjalin kembali hubungan dengan keluarga dan teman-teman, dan merayakan kembali momen-momen berharga dalam kehidupan.

Kabupaten Lamongan merupakan salah satu daerah tujuan wisata di Jawa Timur, dengan potensi wisata (alam, buatan, budaya dan religi) yang ditunjang oleh topografis, geografis, keadaan sosial, iklim, fauna dan kekayaan alamnya. Tercatat total wisatawan dari 22 destinasi wisata yang ada di Kabupaten Lamongan dari tahun 2015-2021 adalah 19.167.971 jiwa, di mana 99,99% berasal dari wisatawan nusantara dan 0,01% wisatawan mancanegara [1]. Terdapat 4 kategori utama wisata di Kabupaten Lamongan yang dapat diidentifikasi dengan total destinasi yang berbeda-beda. Pertama, terdapat jenis wisata alam dengan sebanyak 16 destinasi menawarkan keindahan alam yang memukau. Selanjutnya, wisata buatan memiliki 17 tempat yang menciptakan pengalaman berlibur yang unik. Sementara itu, wisata religi menyediakan 8 tujuan yang berkaitan dengan keagamaan dan spiritualitas. Kemudian, ada juga wisata kuliner dengan 6 pilihan makanan khas yang menggugah selera, dan terakhir, wisata budaya yang menawarkan 9 tempat untuk memahami budaya dan sejarah setempat dengan lebih mendalam. Penentuan destinasi wisata di Lamongan bukanlah tugas yang mudah, mengingat beragamnya pilihan yang tersedia. Calon wisatawan sering kali menghadapi dilema dalam memilih tempat wisata yang sesuai dengan preferensi mereka. Dalam proses pengambilan keputusan ini, ada beberapa faktor yang perlu dipertimbangkan dengan seksama. Pertama-tama, biaya masuk menjadi faktor kunci, karena setiap destinasi memiliki biaya yang berbeda-beda. Selanjutnya, jarak tempuh ke destinasi juga perlu diperhitungkan, mengingat Lamongan memiliki berbagai destinasi yang tersebar di berbagai wilayah. Fasilitas yang tersedia di destinasi tersebut juga memainkan peran penting, karena kenyamanan dan keamanan pengunjung menjadi prioritas. Selain itu, peringkat atau bintang destinasi wisata juga menjadi acuan dalam pemilihan, karena menunjukkan kualitas dan reputasi tempat tersebut. Menghadapi kerumitan ini, salah satu solusi yang dapat membantu calon wisatawan adalah pemanfaatan sistem pendukung keputusan.

Sistem pendukung keputusan adalah sebuah sistem komputer yang mengubah data menjadi informasi untuk mengambil keputusan semistruktur dan spesifik [2]. Sistem pendukung keputusan dapat membantu seseorang

untuk pengambilan sebuah keputusan dalam meningkatkan pekerjaannya [3]. Sistem pendukung keputusan ini diharapkan dapat digunakan wisatawan untuk mendapatkan informasi rekomendasi pengambilan keputusan dalam pemilihan objek wisata. Sistem pendukung keputusan dibuat menggunakan algoritma Naive Bayes. Algoritma Naive Bayes merupakan Teori Probabilitas, yaitu suatu pendekatan dalam bidang probabilitas yang digunakan untuk menyatakan pengetahuan atau keyakinan tentang terjadinya atau telah terjadinya suatu peristiwa [4]. Metode ini menjadi bagian dari teknik probabilitas yang sangat relevan dengan kehidupan sehari-hari, dengan fokus pada konsep probabilitas peristiwa dan bukti yang mendukungnya [5]. Kelebihan dari algoritma Naive Bayes termasuk kemampuannya mengatasi data dimensi yang tinggi (banyak fitur), efisien dalam mengklasifikasikan data besar, dan kemampuannya dalam menangani data kategorikal.

Sehingga dalam penelitian ini dibuatlah aplikasi berbasis web yang berjudul “Implementasi Algoritma Naive Bayes pada Sistem Pendukung Keputusan Rekomendasi Wisata di Kabupaten Lamongan”. Dengan menerapkan algoritma Naive Bayes diharapkan dapat membantu para calon pengunjung dalam menentukan rekomendasi pemilihan wisata dengan cepat, tepat dan pastinya kompetitif serta sesuai ekspektasi.

2. METODE PENELITIAN

Data yang digunakan dalam penelitian ini berasal dari dua sumber utama, yaitu Dinas Pariwisata kabupaten Lamongan dan Google Map. Data tersebut nantinya digunakan sebagai data training dan testing. Data tersebut mencakup:

- Data tentang alternatif tempat wisata yang mencakup lokasi wisata berbeda.
- Data mengenai waktu jam buka wisata.
- Data mengenai harga tiket dari masing-masing objek wisata.
- Informasi tentang fasilitas-fasilitas yang tersedia di setiap objek wisata.
- Data berupa ulasan bintang Google Map terkait tempat-tempat wisata tersebut.
- Data jarak yang harus ditempuh dari pusat kota titik nol (0) menuju masing-masing objek wisata.
- Informasi tentang kerapihan kebersihan objek wisata.

Algoritma yang digunakan dalam penelitian ini adalah Naïve bayes dengan persamaan sebagai berikut:

$$P(Y|X) = \frac{P(X|Y) \cdot P(Y)}{P(X)} \dots\dots\dots(1)$$

Keterangan :

- X : merujuk pada data yang memiliki kelas yang belum teridentifikasi,
- Y : merupakan gagasan bahwa data X mungkin termasuk dalam suatu kelas tertentu
- P(Y|X) : probabilitas hipotesis H berdasar kondisi X (Posteriori Probability)
- P(Y) : mewakili peluang bahwa hipotesis H benar (Probabilitas Priori)
- P(X|Y) : menggambarkan probabilitas X diberikan kondisi hipotesis H
- P(X) : probabilitas X secara keseluruhan.

Densitas.Gauss :

$$P(X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi}\sigma_{ij}} e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}} \dots\dots\dots(2)$$

Keterangan :

- P : melambangkan kemungkinan,
- X_i : mewakili atribut ke-i,
- x_i : mengacu pada nilai atribut ke-i,
- Y : merupakan kelas yang sedang dicari,
- y_i : subkelas Y yang sedang dicari,
- μ : mewakili rata-rata, yang mengindikasikan nilai tengah dari semua atribut
- σ : adalah deviasi standar, mencerminkan variasi dari seluruh atribut.

Standar Deviasi :

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n-1}} \dots\dots\dots(3)$$

Keterangan :

- σ : adalah simbol untuk menyatakan standar deviasi,
- x_i : merujuk pada nilai x pada posisi ke-i,

μ : adalah rata-rata aritmatika,
 n : adalah total jumlah sampel yang ada.
 Mean :

$$\mu = \frac{x_1+x_2+x_3+\dots+x_n}{n} \dots\dots\dots(4)$$

Keterangan :

μ : nilai tengah rata-rata (mean),
 x_i : nilai dari sampel ke-i,
 n : total jumlah sampel yang ada.

Kriteria yang digunakan pada sistem ini adalah: waktu buka, fasilitas wisata, harga masuk, jarak, kebersihan, bintang wisata. Masing-masing kriteria mempunyai nilai ordinal seperti di Tabel 1 sampai Tabel 6. Tabel 1 menjelaskan nilai ordinal waktu buka wisata. Waktu jam buka wisata dapat bervariasi tergantung pada kebijakan pengelola dan karakteristik wisata itu sendiri. Biasanya, jam buka wisata ditentukan agar dapat memaksimalkan pengalaman pengunjung dan memastikan keamanan serta pemeliharaan fasilitas [1].

Tabel 1. Kriteria Waktu Buka (C1)

Nama Sub Kriteria	Nilai
09:00 sampai 16:00	1
08:00 sampai 16:00	2
08:00 sampai 18:00	3
Tidak ada batas	4

Tabel 2. Kriteria Fasilitas Wisata (C2)

Nama Sub Kriteria	Nilai
Tidak Ada	1
Kurang Lengkap	2
Sedang	3
Lengkap	4

Tabel 2 nilai ordinal dari kriteria fasilitas, fasilitas yang diperlukan calon wisatawan misalnya tempat parkir, toilet, tempat ibadah, dan lain-lain. Tabel 3 menampilkan nilai ordinal dari kriteria harga masuk. Harga masuk Setiap tempat wisata menawarkan berbagai keunikan faktor yang disediakan untuk memanjakan wisatawan, dan hal yang menyebabkan harga tiket masuk wisata bisa berbeda-beda tergantung pada beberapa faktor [4].

Tabel 3. Kriteria Harga Masuk (C3)

Nama Sub Kriteria	Nilai
Lebih Dari Rp. 50.000	1
Rp. 30.000 - Rp. 50.000	2
Rp. 10.000 - Rp. 30.000	3
Kurang Dari Rp. 10.000	4

Tabel 4. Kriteria Jarak (C4)

Nama Sub Kriteria	Nilai
lebih dari 20 KM	1
15 – 20 KM	2
10 – 15 KM	3
5 – 10 KM	4

Tabel 5. Kriteria Kebersihan (C5)

Nama Sub Kriteria	Nilai
Kotor	1
Kurang Bersih	2
Cukup Bersih	3
Bersih	4

Tabel 6. Kriteria Bintang Wisata (C6)

Nama Sub Kriteria	Nilai
1 - 2 Bintang	1
3 Bintang	2
4 Bintang	3

Nama Sub Kriteria	Nilai
5 Bintang	4

Tabel 4 merupakan nilai ordinal dari variabel jarak. Jarak tempuh didasarkan pada jarak dari alun-alun Lamongan ke lokasi objek wisata, yang diambil dari Google Maps dengan menggunakan kendaraan mobil. Semakin pendek jaraknya ke tempat wisata, semakin tinggi bobot yang diberikan [6]. Tabel 5 merupakan nilai ordinal dari variabel kebersihan. Kebersihan wisata merujuk pada kondisi kebersihan dan sanitasi yang terjaga di area wisata atau tempat pariwisata. Hal ini mencakup kebersihan lingkungan, fasilitas umum, toilet, area makan, dan tempat-tempat umum lainnya di dalam wisata [3]. Tabel 6 merupakan nilai ordinal dari bintang wisata. Bintang wisata merupakan penilaian dari banyaknya ulasan google map yang diambil sampai tanggal 23 Juli 2023, pembobotan dinilai berdasarkan bintang pada tempat wisata yang semakin banyak ulasan maka bobot akan semakin tinggi pembobotan bintang wisata.

3. HASIL DAN PEMBAHASAN

Dalam pembahasa ini data yang telah diperoleh dari dinas kebudayaan dan pariwisata Lamongan sebanyak 47 data training yang ditunjukkan pada tabel 7 berikut:

Tabel 7. Data Training

No	C1	C2	C3	C4	C5	C6	Hasil
1	4	4	4	4	4	4	Direkomendasi
2	4	4	4	4	4	3	Direkomendasi
3	4	1	1	1	1	1	Tidak
4	3	3	2	3	3	3	Direkomendasi
5	4	3	4	4	3	3	Direkomendasi
6	1	4	4	3	4	3	Direkomendasi
7	1	3	4	4	4	4	Direkomendasi
8	1	4	4	2	4	4	Direkomendasi
9	3	2	4	3	4	2	Direkomendasi
10	1	1	1	1	1	1	Tidak
11	2	4	4	4	4	2	Direkomendasi
12	3	2	2	2	2	2	Direkomendasi
13	4	4	4	2	3	1	Direkomendasi
14	2	2	2	2	2	1	Tidak
15	4	3	4	2	3	4	Direkomendasi
16	4	4	4	1	4	4	Direkomendasi
17	3	3	3	3	3	3	Direkomendasi
18	2	2	1	2	2	2	Tidak
19	1	2	1	2	1	2	Tidak
20	2	2	2	2	1	2	Tidak
21	3	1	4	1	1	1	Tidak
22	3	2	3	2	3	2	Direkomendasi
23	1	2	4	3	1	3	Direkomendasi
24	1	3	4	3	3	4	Direkomendasi
25	2	1	4	2	1	3	Direkomendasi
26	2	3	4	3	3	4	Direkomendasi
27	2	2	3	3	2	1	Direkomendasi
28	1	2	1	2	1	1	Tidak
29	4	3	4	2	4	3	Direkomendasi
30	3	1	4	4	3	3	Direkomendasi
31	1	3	4	3	3	3	Direkomendasi
32	2	2	3	2	4	4	Direkomendasi
33	4	1	3	1	1	1	Tidak
34	1	4	1	1	3	4	Direkomendasi
35	3	2	4	4	4	3	Direkomendasi
36	1	1	4	4	2	2	Direkomendasi
37	3	2	3	2	3	2	Direkomendasi
38	3	2	2	3	2	4	Direkomendasi
39	4	4	1	1	1	3	Direkomendasi
40	4	1	1	1	1	1	Direkomendasi
41	1	1	4	1	1	2	Tidak
42	3	1	2	1	2	2	Tidak
43	1	4	3	4	3	4	Direkomendasi
44	3	1	1	1	1	2	Tidak
45	3	4	4	4	4	4	Direkomendas
46	2	4	1	1	4	3	Direkomendasi

No	C1	C2	C3	C4	C5	C6	Hasil
47	1	3	3	3	3	3	Direkomendasi

Pada tabel 7 merupakan bagian dataset yang akan dilatih untuk menjalankan algoritma atau melakukan rekomendasi. Pemerolehan data training didapatkan melalui kuisioner yang sebarakan kepada beberapa calon responden.

Tabel 8. Data Testing

No	C1	C2	C3	C4	C5	C6
1	4	4	4	4	4	4
2	4	4	4	4	4	3
3	4	1	1	1	1	1
4	3	3	2	3	3	3
5	4	3	4	4	3	3
6	1	4	4	3	4	3
7	1	3	4	4	4	4
8	1	4	4	2	4	4
9	3	2	4	3	4	2
10	1	1	1	1	1	1
11	2	4	4	4	4	2
12	3	2	2	2	2	2
13	4	4	4	2	3	1
14	2	2	2	2	2	1
15	4	3	4	2	3	4
16	4	4	4	1	4	4
17	3	3	3	3	3	3
18	2	2	1	2	2	2
19	1	2	1	2	1	2
20	2	2	2	2	1	2
21	3	1	4	1	1	1
22	3	2	3	2	3	2
23	1	2	4	3	1	3
24	1	3	4	3	3	4
25	2	1	4	2	1	3
26	2	3	4	3	3	4
27	2	2	3	3	2	1
28	1	2	1	2	1	1
29	4	3	4	2	4	3
30	3	1	4	4	3	3
31	1	3	4	3	3	3
32	2	2	3	2	4	4
33	4	1	3	1	1	1
34	1	4	1	1	3	4
35	3	2	4	4	4	3
36	1	1	4	4	2	2
37	3	2	3	2	3	2
38	3	2	2	3	2	4
39	4	4	1	1	1	3
40	4	1	1	1	1	1
41	1	1	4	1	1	2
42	3	1	2	1	2	2
43	1	4	3	4	3	4
44	3	1	1	1	1	2
45	3	4	4	4	4	4
46	2	4	1	1	4	3
47	1	3	3	3	3	3

Data testing yang digunakan berjumlah 47 data wisata yang ada di Lamongan, data yang diperoleh berasal dari dua sumber utama yaitu Dinas Pariwisata Kabupaten Lamongan dan Google Map. Untuk menguji keakuratan dalam perhitungan yang dilakukan, pengujian menggunakan perhitungan algoritma Naïve Bayes.

Berikut adalah contoh perhitungan yang dilakukan:

1. Mencari nilai mean yang direkomendasikan dengan persamaan (4) :

$$\begin{aligned} \mu (C1 \text{ Rekom}) &= \frac{4+4+3+4+1+1+1+1+3+2+3+4+4+4+3+3+3+1+1+2+2+2+2+4+3+1+2+1+3+1+3}{28} = \frac{70}{28} = 2.5 \\ \mu (C2 \text{ Rekom}) &= \frac{4+4+3+3+4+3+4+2+4+2+4+3+4+3+2+2+3+1+3+2+4+2+1+2}{28} = \frac{78}{28} = 2.785714 \\ \mu (C3 \text{ Rekom}) &= \frac{4+4+2+4+4+4+4+4+4+2+4+4+4+3+3+4+4+4+4+3+4+4+3+1+4+4+3}{28} = \frac{100}{28} = 3.571429 \\ \mu (C4 \text{ Rekom}) &= \frac{4+4+3+4+3+4+2+3+4+2+2+2+1+3+2+3+3+2+3+3+2+4+3+2+1+4+4+2}{28} = \frac{79}{28} = 2.821429 \\ \mu (C5 \text{ Rekom}) &= \frac{4+4+3+3+4+4+4+4+4+2+3+3+4+3+3+1+3+1+3+2+4+3+3+4+3+4+2+3}{28} = \frac{88}{28} = 3.142857 \\ \mu (C6 \text{ Rekom}) &= \frac{4+3+3+3+3+4+4+2+2+2+1+4+4+3+2+3+4+3+4+1+4+3+2+2}{28} = \frac{83}{28} = 2.964286 \end{aligned}$$

2. Mencari nilai mean yang tidak direkomendasikan:

$$\begin{aligned} \mu (C1 \text{ Tidak}) &= \frac{4+1+2+2+1+2+3+1+4}{7} = \frac{20}{7} = 2.222222 \\ \mu (C2 \text{ Tidak}) &= \frac{1+1+2+2+2+2+1+2+1}{7} = \frac{14}{7} = 1.555555 \\ \mu (C3 \text{ Tidak}) &= \frac{1+1+2+1+1+2+4+1+3}{7} = \frac{16}{7} = 1.777778 \\ \mu (C4 \text{ Tidak}) &= \frac{1+1+2+2+2+2+1+2+1}{7} = \frac{14}{7} = 1.555555 \\ \mu (C5 \text{ Tidak}) &= \frac{1+1+2+2+1+1+1+1+1}{7} = \frac{11}{7} = 1.222222 \\ \mu (C6 \text{ Tidak}) &= \frac{1+1+1+2+2+2+1+1+1}{7} = \frac{12}{7} = 1.333333 \end{aligned}$$

Tabel 9. Hasil Perhitungan Mean

Kelas	C1	C2	C3	C4	C5	C6
Direkomendasi	1.170628	0.994695	0.790151	0.944911	0.890870	0.922241
Tidak Direkomendasi	1.201850	0.527046	1.092906	0.527046	0.440958	0.5

Tabel 9. merupakan hasil perhitungan dalam menentukan mean kategori direkomendasikan dan tidak direkomendasikan.

3. Selanjutnya melakukan perhitungan pencarian nilai deviasi dengan menggunakan persamaan (3) sebagai berikut:

- Deviasi direkomendasikan:

$$\begin{aligned} \sigma (C1) &= \frac{(4 - 2.5)^2 + (4 - 2.5)^2 + (3 - 2.5)^2 + (4 - 2.5)^2 + (1 - 2.5)^2 + (1 - 2.5)^2 + (1 - 2.5)^2 + (3 - 2.5)^2 + (2 - 2.5)^2 + (3 - 2.5)^2 + (4 - 2.5)^2 + (4 - 2.5)^2 + (4 - 2.5)^2 + (3 - 2.5)^2 + (3 - 2.5)^2 + (1 - 2.5)^2 + (1 - 2.5)^2 + (2 - 2.5)^2 + (2 - 2.5)^2 + (2 - 2.5)^2 + (4 - 2.5)^2 + (3 - 2.5)^2 + (1 - 2.5)^2 + (2 - 2.5)^2 + (1 - 2.5)^2 + (3 - 2.5)^2 + (1 - 2.5)^2 + (3 - 2.5)^2}{(28 - 1)} = 1.170628 \end{aligned}$$

- Deviasi tidak direkomendasikan:

$$\sigma (C1) = \frac{(4 - 2.2222222)^2 + (1 - 2.2222222)^2 + (2 - 2.2222222)^2 + (2 - 2.2222222)^2 + (1 - 2.2222222)^2 + (2 - 2.2222222)^2 + (3 - 2.2222222)^2 + (1 - 2.2222222)^2 + (4 - 2.2222222)^2}{(9 - 1)} = 1.201850$$

Tabel 9. Hasil Perhitungan Deviasi

Kelas	C1	C2	C3	C4	C5	C6
Direkomendasikan	1.170628	0.994695	0.790151	0.944911	0.890870	0.922241
Tidak Direkomendasikan	1.201850	0.527046	1.092906	0.527046	0.440958	0.5

4. Mencari probabilitas Kelas Direkomendasikan dan Tidak Direkomendasikan menggunakan Persamaan (1)

$$\text{Direkomendasikan} = \frac{28}{37} = 0.756756757$$

$$\text{Tidak Direkomendasikan} = \frac{9}{37} = 0.243243243$$

Untuk klasifikasi dengan data kontinyu digunakan persamaan Densitas Gauss (2):

Tabel 10. Mencari Rekomendasi Perhitungan Hasil

Nama Wisata	C1	C2	C3	C4	C5	C6	Hasil
Wisata Bahari	08:00 - 16:00	Lengkap	> 50.000	20 KM	Bersih	Bintang 5	?
Lamongan	2	4	1	1	4	4	

Direkomendasikan

$$C1 = \frac{1}{\sqrt{2\pi \cdot 3,14 \cdot 1,17}} e^{-\frac{(2-2,5)^2}{2 \cdot (1,17)^2}} = 0.31116196$$

$$C2 = \frac{1}{\sqrt{2\pi \cdot 3,14 \cdot 0,99}} e^{-\frac{(4-2,57)^2}{2 \cdot (0,99)^2}} = 0.19042522$$

$$C3 = \frac{1}{\sqrt{2\pi \cdot 3,14 \cdot 0,79}} e^{-\frac{(1-3,57)^2}{2 \cdot (0,79)^2}} = 0.002532487$$

$$C4 = \frac{1}{\sqrt{2\pi \cdot 3,14 \cdot 0,94}} e^{-\frac{(1-2,82)^2}{2 \cdot (0,94)^2}} = 0.065882805$$

$$C5 = \frac{1}{\sqrt{2\pi \cdot 3,14 \cdot 0,89}} e^{-\frac{(4-3,14)^2}{2 \cdot (0,89)^2}} = 0.281961068$$

$$C6 = \frac{1}{\sqrt{2\pi \cdot 3,14 \cdot 0,92}} e^{-\frac{(4-2,96)^2}{2 \cdot (0,92)^2}} = 0.230305975$$

Tidak direkomendasikan

$$C1 = \frac{1}{\sqrt{2\pi \cdot 3,14 \cdot 1,20}} e^{-\frac{(2-2,22)^2}{2 \cdot (1,20)^2}} = 0.32639682$$

$$C2 = \frac{1}{\sqrt{2\pi \cdot 3,14 \cdot 0,52}} e^{-\frac{(4-1,55)^2}{2 \cdot (0,52)^2}} = 1.6147E - 05$$

$$C3 = \frac{1}{\sqrt{2\pi \cdot 3,14 \cdot 1,09}} e^{-\frac{(1-1,77)^2}{2 \cdot (1,09)^2}} = 0.283439792$$

$$C4 = \frac{1}{\sqrt{2\pi \cdot 3,14 \cdot 0,52}} e^{-\frac{(1-1,55)^2}{2 \cdot (0,52)^2}} = 0.434406901$$

$$C5 = \frac{1}{\sqrt{2\pi \cdot 3,14 \cdot 0,44}} e^{-\frac{(4-1,22)^2}{2 \cdot (0,44)^2}} = 2.18609E - 09$$

$$C6 = \frac{1}{\sqrt{2\pi \cdot 3,14 \cdot 0,5}} e^{-\frac{(4-1,33)^2}{2 \cdot (0,5)^2}} = 5.31395E - 07$$

$$\text{Direkomendasikan} = C1 * \text{Prob Rekom} * C2 * C3 * C4 * C5 * C6 = 4.85826E - 07$$

$$\text{Tidak DiRekomendasikan} = C1 * \text{Prob Tidak} * C2 * C3 * C4 * C5 * C6 = 1.83368E - 22$$

Dari hasil perhitungan di atas didapatkan kesimpulan bahwa hasil dari nilai direkomendasikan lebih besar dibandingkan dengan nilai tidak direkomendasikan, maka data uji Wisata Bahari Lamongan (WBL) di atas masuk ke dalam kelas sebagai wisata yang direkomendasikan.

Untuk menilai akurasi dari sistem yang bangun, ujicoba dilakukan dengan melakukan perbandingan antara perhitungan sistem dan perhitungan manual yang berjumlah 47 data. Hasil dari ujicoba disampaikan dalam tabel Confusion matrix pada tabel 11:

Tabel 11. Confusion Matrix

	Sesuai Aplikasi	Tidak Sesuai Aplikasi
Sesuai Perhitungan	47	0
Tidak Sesuai Perhitungan	3	0

Dari hasil confusion matrix dapat dilakukan perhitungan dengan persamaan (5) berikut:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \dots \dots \dots (5)$$

$$\text{Akurasi} = \frac{47+0}{47+3+0+0} \times 100$$

$$\text{Akurasi} = 94\%$$

Berdasarkan perhitungan yang dilakukan menggunakan algoritma Naïve Bayes dengan menggunakan 47 data training dan 47 data testing dilakukan pengujian tingkat akurasi dan menghasilkan akurasi sebesar 94%.

4. SIMPULAN

Sistem pendukung keputusan menggunakan algoritma Naïve Bayes dengan 47 data training dan 47 data testing menunjukkan tingkat akurasi sebesar 94%. Data uji yang mencakup 47 objek wisata di Lamongan dievaluasi berdasarkan parameter-parameter tersebut diatas, menunjukkan bahwa sebagian besar objek wisata direkomendasikan. Proses pengujian ini melibatkan perhitungan mean dan deviasi untuk setiap kelas, diikuti dengan perhitungan probabilitas menggunakan metode Densitas Gauss. Perbandingan hasil perhitungan dengan perhitungan manual menghasilkan akurasi sebesar 94%, menunjukkan bahwa algoritma Naïve Bayes dapat digunakan secara efektif untuk memberikan rekomendasi terhadap objek wisata di Lamongan berdasarkan data yang diberikan.

5. SARAN

Membuat sistem dengan dataset data testing lebih banyak dan lebih kompleks selanjutnya dibandingkan dengan metode yang lain, sehingga akurasi yang didapatkan bisa lebih baik.

DAFTAR PUSTAKA

- [1] D. H. Tanjung, 2015, *Pemilihan Objek Wisata di Sumatera Utara dengan Metode Analytical Hierarchy Process (AHP)*, Seminar Nasional Informatika.
- [2] Mirfan, 2020, *Sistem Pendukung Keputusan Pemilihan Destinasi Wisata Berbasis Web Dengan Algoritma K-Means Clustering Dan (Topsis)*, JINSTEK(Jurnal Informatika Sains dan Teknologi), vol. 5, no. 2.
- [3] H. S. Pakpahan, Y. Basani, and N. Shadrina, 2023, *Sistem Pendukung Keputusan Pemilihan Objek Wisata Menggunakan Metode Weighted Product dan Simple Additive Weighting*, *Informatika Mulawarman, Jurnal Ilmiah Ilmu Komputer*, vol. 18, no. 1, doi: 10.30872/jim.v18i1.8399.
- [4] P. A. Prasojo, M. Nurkamid, and M. I. Ghozali, 2020, *Sistem Pendukung Keputusan Pemilihan Objek Wisata Di Jepara Berbasis Android*, *Jurnal Dialektika Informatika (Detika)*, vol. 1, no. 1, doi: 10.24176/detika.v1i1.5531.
- [5] A. Nugroho and A. Safarudin, 2020, *Sistem Pendukung Keputusan Pemilihan Hotel Di Jakarta Dengan Menggunakan Algoritma Naïve Bayes*, *SIGMA - Jurnal Teknologi Pelita Bangsa* 167, vol. 10, no. September.
- [6] M. U. Nuhayati, D. Dedih, and J. Mulyana, 2021, *Sistem Pendukung Keputusan Untuk Menentukan Lokasi Usaha Kuliner Yang Strategis Menggunakan Metode Naive Bayes*, *Jurnal Interkom: Jurnal Publikasi Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 12, no. 1, doi: 10.35969/interkom.v12i1.22.

Sistem Pakar Diagnosa Gangguan Pencernaan Balita Dengan Metode Naive Bayes

Emi Milda Prawestina¹, M.Ghofar Rohman², Moh. Rosidi Zamroni³

^{1,2,3}Teknik Informatika, Fakultas Teknik, Universitas Islam Lamongan

E-mail: ¹emimildaprawestina@gmail.com, ²m.ghofarrohman@unisla.ac.id,

³rosidizamroni@unisla.ac.id

Abstrak – Gangguan pencernaan pada balita merupakan masalah umum yang sulit dideteksi karena bayi tidak dapat mengungkapkan ketidaknyamanan mereka. Artikel ini membahas pentingnya mendiagnosa gangguan pencernaan balita dengan cepat menggunakan sistem pakar berbasis metode Naive Bayes. Gangguan pencernaan pada balita dapat disebabkan oleh berbagai faktor seperti pola makan yang buruk dan infeksi saluran cerna. Sistem pakar merupakan solusi yang efektif untuk membantu mendiagnosa gangguan pencernaan pada balita. Dengan memanfaatkan metode Naive Bayes, sistem pakar dapat mengidentifikasi gejala-gejala yang mungkin menunjukkan adanya gangguan pencernaan. Artikel ini juga membahas hasil perhitungan dan pengujian sistem pakar menggunakan data gejala dan penyakit yang dikumpulkan dari Puskesmas Sugio Lamongan. Melalui uji coba dengan 100 pengguna, sistem pakar ini mencapai tingkat akurasi sebesar 87%, dengan 87 pasien mendapatkan diagnosis yang sesuai.

Kata Kunci — Sistem Pakar, Gangguan pencernaan, Balita, Naive Bayes

1. PENDAHULUAN

Gangguan pencernaan adalah masalah umum pada balita. pengkonsumsian makan dan perawatan yang tidak tepat sering menyebabkan gangguan pencernaan pada balita, yang sulit dideteksi karena bayi tidak dapat mengungkapkan ketidaknyamanan yang mereka rasakan. Gangguan pencernaan balita atau bayi dibawah usia 3 tahun biasanya ditandai dengan beberapa gejala seperti mudah tersinggung, kembung, mual, muntah, diare bahkan dehidrasi[1]. Kondisi ini dapat disebabkan oleh berbagai faktor pola makan yang buruk hingga infeksi saluran cerna hingga intoleransi laktosa. Gangguan pencernaan harus didiagnosa agar dapat diobati dengan cepat dan mengetahui gejala serta penyakit yang diderita [2]. Gangguan pada sistem pencernaan dapat disebabkan oleh gangguan pada saluran pencernaan, invasi bakteri dan pola makan yang buruk. Anak kecil sangat rentan terhadap infeksi bakteri patogen, dan rentan terhadap gejala penyakit merupakan ketakutan bagi orang tua. Kebanyakan orang tua tidak memahami gangguan pencernaan pada anak kecil [3].

Gangguan pencernaan merupakan masalah pada organ sistem pencernaan. Saluran cerna dimulai mulut, kerongkongan, lambung, usus halus, usus besar, anus. Sistem pencernaan adalah organ mencerna makanan sampai benar-benar tercerna bening dan cair serta terserap ke dalam darah untuk disalurkan ke seluruh bagian tubuh, yaitu perut besar dan lambung. Jika fungsi organ pencernaan terganggu, maka terjadilah gangguan pencernaan [4]. Sistem pencernaan adalah sistem fungsional untuk memproses makanan yang diserap secara fisik dan kimiawi yang digunakan oleh sel-sel tubuh yang terdiri dari mulut, kerongkongan, lambung, rektum, hati dan pankreas[5]. Penyakit pencernaan adalah suatu kondisi yang mempengaruhi saluran pencernaan dan disebabkan oleh pola makan yang tidak sehat, pola makan yang tidak seimbang, pola makan yang buruk, peradangan dan gangguan pada organ pencernaan [6]

Menurut Organisasi Kesehatan Dunia (WHO) 2019, menunjukkan bahwa sekitar 1,5 juta kebanyakan orang di dunia meninggal karena penyakit saluran cerna termasuk diare, sedangkan di Indonesia penyakit saluran cerna dan penyakit tidak menular lainnya secara bersama-sama menyumbang sekitar 30% kematian. Diare menyebabkan hilangnya air dan mineral dapat menyebabkan dehidrasi. Balita khususnya, dapat mengalami dehidrasi lebih cepat dari pada orang dewasa, jadi mengganti cairan sangatlah penting. Dehidrasi adalah suatu kondisi dimana tubuh kekurangan cairan yang terutama pada balita dapat berakibat fatal jika tidak segera ditangani [7].

Kendala utama yang biasanya dialami oleh pasien yaitu karena mahalnya biaya konsultasi ke dokter. Selain itu, terkadang pasien memiliki aktivitas yang sangat padat sehingga tidak ada waktu luang untuk konsultasi ke dokter. Untuk mempermudah pasien dalam mengenali penyakit gangguan pencernaan diperlukan sebuah aplikasi

yang dapat membantu mendiagnosa penyakit yaitu sistem pakar. Menurut Handoko[8], Sistem pakar dikembangkan pada pertengahan 1960-an, jadi mereka adalah bentuk kecerdasan buatan yang cukup tua. Sistem ini bertujuan buat mentransfer informasi manusia ke komputer yang menggabungkan informasi dasar untuk menggantikan ahli pemecahan masalah. Sistem pakar berasal dari istilah sistem pakar basis pengetahuan.

Sistem Pakar adalah aplikasi komputer yang dapat digunakan untuk memecahkan masalah yang disajikan oleh pakar[9]. Sistem pakar ini mentransfer pengetahuan manusia ke komputer dan memecahkan masalah yang biasanya dipecahkan oleh para pakar[10]. Sistem pakar membuat daftar gejala yang dimasukkan oleh pakar untuk mengidentifikasi objek dan menemukan jawaban yang dapat diterima[11]. Sistem pakar memiliki beberapa metode yang dapat digunakan, salah satunya adalah metode Naive Bayes.

Untuk membantu masyarakat dan pasien mengetahui jenis gangguan pencernaan balita Aplikasi sistem pakar dikembangkan dengan perawatan awal yang dapat dilakukan oleh pasien diagnosa gangguan pencernaan. Sebuah metode untuk menghitung probabilitas seseorang mengalami gangguan pencernaan menggunakan metode Naive Bayes.

Dari permasalahan diatas maka sangat dibutuhkan sebuah sistem pakar yang dapat membantu dalam pemecahan masalah, Sebuah aplikasi sistem pakar diagnosa gangguan pencernaan balita. Dan diharapkan dapat menjadi salah satu alternatif bantuan bagi pasien dalam memperoleh informasi penyakit gangguan pencernaan balita. Dan sekaligus dapat pula mengetahui pemecahan masalah atau solusi yang tepat untuk menangani penyakit tersebut. Sehingga penulis mencari solusi dalam pembangunan sistem pakar yang berjudul “Sistem Pakar Gangguan Pencernaan Balita Menggunakan Metode Naive Bayes”.

Penelitian terdahulu yang serupa yaitu menurut Sinaga [2], “Sistem Pakar Deteksi Gizi Buruk Pada Balita Menggunakan Metode Klasifikasi Naive Bayes” Tujuan dari penelitian ini adalah untuk mengidentifikasi gizi buruk pada anak usia 1-3 tahun dengan menggunakan klasifikasi Naive Bayes. Berdasarkan gejalanya, penelitian ini mengidentifikasi tiga penyakit dengan 24 gejala malnutrisi: Quartiorcor (P1), Marasmic Quartiorcor (P2), dan Marasmus (P3). Hasil penelitian menunjukkan bahwa hasil perkalian terbesar dari klasifikasi Naive Bayes adalah jenis malnutrisi yang diderita pasien. Pengamatan dapat digunakan sebagai informasi awal untuk mengidentifikasi malnutrisi.

Selanjutnya menurut Viviliani dan Radius Tanone [12] “Rancang Bangun Sistem Pakar Diagnosis Penyakit Anak Menggunakan Metode Forward Chain Berbasis Android” Tujuan dari penelitian ini adalah untuk mendiagnosa penyakit dengan cara yang tepat untuk membantu memberikan informasi tentang penyakit dan gejalanya, serta menunjukkan pengolahan data yang tepat dan jumlah responden yang puas dengan hasil aplikasi. Diagnosis penyakit bayi 40%, sedangkan 33,3% setuju sepenuhnya, sehingga dapat membantu masyarakat di Rumah Sakit Ibu dan Anak Mutiara Bunda Salatiga untuk mendiagnosa penyakit bayi.

2. METODE PENELITIAN

Dalam penelitian ini menggunakan beberapa metode yang digunakan, diantaranya adalah; Teknik pengumpulan data, analisis system, dan perancangan sistem.

1. Tahap pengumpulan data dilakukan dengan metode studi pustaka, pengamatan, dan wawancara.
2. Analisis system dilakukan dengan melakukan analisa kebutuhan dari sistem yang akan dibangun dalam penelitian ini.
3. Penelitian ini menggunakan algoritma naïve bayes yang didasarkan pada [13], dengan persamaan sebagai berikut:

$$P(a_1|v_j) = \frac{n_c + m_p}{n+m} \dots\dots\dots(1)$$

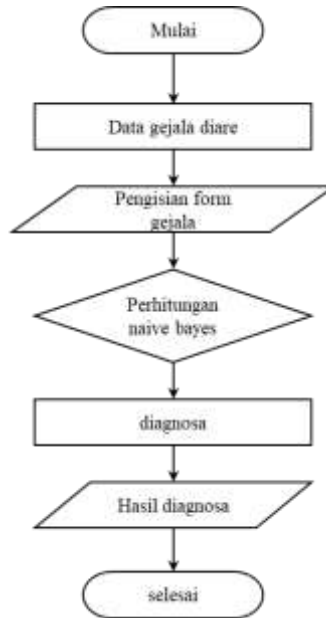
Keterangan:

- P ($a_i|v_j$) : peluang atribut-atribut (inputan) jika diketahui keadaan v_j
- a_i ($a_1, a_2, \dots a_n$) : atribut (inputan)
- v_j ($v_1, v_2, \dots v_m$) : penyaji ke j
- n_c : jumlah *record* pada data jika $v=v_j$ dan $a=a_i$
- m_p : 1/banyaknya penyakit
- m : jumlah peramalan penyakit

n : jumlah *record* pada data jika $v = v_j$ tiap kelas

Langkah-langkah menyelesaikan permasalahan berdasarkan persamaan (1) adalah sebagai berikut:

- a. Menentukan nilai n_c untuk setiap kelas
 - b. Menentukan nilai $P(a_i|v_j)$ dan menghitung $P(v_j)$
 - c. Menghitung $P(a_i|v_j) \times P(v_j)$ untuk setiap v
 - d. Menentukan hasil v
4. System yang akan dirancang dilakukan dengan proses seperti digambarkan dalam flowchart berikut:



Gambar 1. Flowchart Sistem

Pada gambar 1 di atas, ditunjukkan alur proses kerja sistem, dimana pada awal proses user akan melakukan proses input gejala, selanjutnya sistem akan melakukan perhitungan dengan *naïve bayes* untuk menentukan dan memberikan hasil diagnosa.

3. HASIL DAN PEMBAHASAN

Data penyakit ini diambil dari Puskesmas Sugio Lamongan. Proses perhitungan dilakukan dengan menggunakan algoritma *Naïve Bayes*. Data diambil pada tahun 2023 mulai dari bulan Januari sampai bulan Juli. Berikut adalah data gejala dan jenis penyakit, serta hasil wawancara dengan para pakar tentang aturan yang telah ditetapkan yang dirujuk pada Tabel 1, Tabel 2, dan Tabel 3.

Tabel 1. Data Penyakit

No	Kode	Nama Penyakit
1	P01	Diare
2	P02	Sembelit

Tabel 2. Data Gejala Gangguan Pencernaan Balita

No.	Kode	Gejala
1	G01	Frekuensi buang air besar (BAB) lebih sering dari biasanya
2	G02	Balita mengalami demam
3	G03	Dehidrasi
4	G04	Nafsu makan berkurang
5	G05	Balita kesakitan saat BAB
6	G06	Mual dan Muntah
7	G07	Perut kembung
8	G08	Sakit perut
9	G09	Ada bercak fases pada popok
10	G10	Mengalami diare tapi dalam jumlah sedikit
11	G11	Berat badan turun
12	G12	Frekuensi BAB kurang dari 3x seminggu

Tabel 3. Hubungan Penyakit Dan Gejala

Kode	Nama Penyakit	Gejala
P01	Diare	G01, G02,G03, G04,G11,G12
P02	Sembelit	G05,G06 G07, G08,G09, G10,

Tabel-tabel di atas (Tabel1, Tabel2, Tabel3) merupakan referensi untuk mengumpulkan data sebagai contoh saat menguji perhitungan naïve Bayes secara manual. Tujuan dari pengujian ini adalah untuk mengetahui akurasi perhitungan menggunakan metode Naive Bayes yang dihitung secara manual.

Contoh perhitungan dengan menggunakan persamaan (1) Naïve Bayes. Misal seseorang mengalami gejala dalam tabel 4 berikut:

Tabel 4. Contoh data

kode	Gejala
G01	Nafsu makan berkurang
G05	Balita kesakitan saat BAB
G06	Mual dan muntah
G08	Sakit perut

Melakukan perhitungan naïve bayes dengan tahapan berikut:

- a. Menentukan nilai n_c untuk setiap kelas

Diketahui :

$$n = 1$$

$$p=1/2=0,5$$

$$m= 12$$

Tabel nilai n_c dari setiap penyakit ditunjukkan pada tabel 5.

Tabel 5. Nilai n_c dari setiap penyakit

Kode	P1	P2
G05	0	1
G06	0	1
G04	1	0
G08	0	1

- b. Menghitung nilai $P(a_i|v_j)$ dan menghitung nilai $P(v_j)$

- 1) Penyakit 1 (Diare)

$$P(G05|P01) = \frac{0 + 12 \times 0,5}{1 + 12} = \frac{0 + 6}{13} = \frac{6}{13} = 0,4615$$

$$P(G06|P01) = \frac{0 + 12 \times 0,5}{1 + 12} = \frac{0 + 6}{13} = \frac{6}{13} = 0,4615$$

$$P(G04|P01) = \frac{1 + 12 \times 0,5}{1 + 12} = \frac{1 + 6}{13} = \frac{7}{13} = 0,5384$$

$$P(G08|P01) = \frac{0 + 12 \times 0,5}{1 + 12} = \frac{0 + 6}{13} = \frac{6}{13} = 0,4615$$

$$P(P01) = 1/2 = 0,5$$

- 2) Penyakit 2 (Sembelit)

$$P(G05|P01) = \frac{1 + 12 \times 0,5}{1 + 12} = \frac{1 + 6}{13} = \frac{7}{13} = 0,5384$$

$$P(G06|P01) = \frac{1 + 12 \times 0,5}{1 + 12} = \frac{1 + 6}{13} = \frac{7}{13} = 0,5384$$

$$P(G04|P01) = \frac{0 + 12 \times 0,5}{1 + 12} = \frac{0 + 6}{13} = \frac{6}{13} = 0,4615$$

$$P(G08|P01) = \frac{1 + 12 \times 0,5}{1 + 12} = \frac{1 + 6}{13} = \frac{7}{13} = 0,5384$$

$$P(P01) = 1/2 = 0,5$$

- c. Menghitung nilai $P(a_i|v_j) \times P(v_j)$ untuk tiap v

- 1) Penyakit 1 (Diare)

$$\begin{aligned} &P(P01) \times P(G05|P01) \times P(G06|P01) \times P(G04|P01) \times P(G08|P01) \\ &= 0,5 \times 0,4615 \times 0,4615 \times 0,5384 \times 0,4615 \\ &= 0,0264 \end{aligned}$$

- 2) Penyakit 2 (Sembelit)

$$\begin{aligned}
 &P(P01) \times P(G05|P01) \times P(G06|P01) \times P(G04|P01) \times P(G08|P01) \\
 &= 0,5 \times 0,5384 \times 0,5384 \times 0,4615 \times 0,5384 \\
 &= 0,0360
 \end{aligned}$$

d. Menentukan perangkungan nilai v yang dihasilkan

Hasil perangkungan dapat diketahui bahwa nilai v terbesar adalah nilai untuk penyakit ke 2 yaitu sembelit.

Tabel 6. Hasil diagnose

Penyakit	Hasil
Diare	0,0264
Sembelit	0,0360

Berdasarkan tabel diatas dapat disimpulkan bahwa seorang pasien dengan gejala nafsu makan berkurang (G4), balita kesakitan saat buang air besar (G5), mual dan muntah (G6), sakit perut (G8). Maka terdefinisi terkena penyakit Sembelit, karena dari hasil perhitungan nilai tertinggi 0,0360 adalah penyakit Sembelit.

Pada saat pengujian sistem perangkat lunak diambil beberapa sampel yaitu pasien dengan gangguan pencernaan pada balita. Tes sistematis dilakukan untuk menentukan kinerja sistem pakar khusus dalam diagnosis gangguan pencernaan pada balita. Data yang diolah terdiri dari 100 data pengguna yang dibandingkan dengan diagnosis dokter.

Tabel 5. Uji coba

No	Nama	Hasil Diagnosa		Ket
		Diagnosa Awal	Diagnosa Sistem	
1	G01, G02, G06,G12	Sembelit	Diare	Tidak sesuai
2	G05, G07, G08, G10	Sembelit	Sembelit	Sesuai
3	G08, G09, G10, G11	Sembelit	Sembelit	Sesuai
4	G01, G02, G03, G04	Diare	Diare	Sesuai
5	G03, G04, G08, G11	Diare	Diare	Sesuai
6	G02,G04.G06,G08,G10	Sembelit	Sembelit	Sesuai
7	G09,G10,G11,G12	Sembelit	Sembelit	Sesuai
8	G01,G04,G06,G09,G11	Diare	Diare	Sesuai
9	G04,G05,G06	Sembelit	Sembelit	Sesuai
10	G05,G06,G07,G08,G11	Diare	Sembelit	Tidak sesuai
11	G06,G08,G09	Sembelit	Sembelit	Sesuai
...
91	G03,G05,G06,G09,G10	Diare	Sembelit	Tidak sesuai
92	G01.G03,G06	Diare	Diare	Sesuai
93	G08,G11,G12	Diare	Diare	Sesuai
94	G03,G07,G09,G11	Sembelit	Diare	Tidak sesuai
95	G10,G12	Diare	Diare	Sesuai
96	G03,G08,G10	Sembelit	Sembelit	Sesuai
97	G04,G11	Diare	Diare	Sesuai
98	G03,G9	Diare	Diare	Sesuai
99	G09,G10	Sembelit	Sembelit	Sesuai

Dari Tabel 5 di atas, nilai akurasi diagnostik untuk sistem dapat diperoleh dengan menggunakan rumus berikut:

$$\text{Akurasi Benar} = \frac{\text{Jumlah Benar}}{\text{Jumlah Seluruh}} \times 100\% \dots\dots\dots(2)$$

$$\text{Akurasi Salah} = \frac{\text{Jumlah Salah}}{\text{Jumlah Seluruh}} \times 100\% \dots\dots\dots(3)$$

Dari rumus di atas, kita dapat melihat bahwa nilai akurasi sistem dihitung sebagai:

Berdasarkan penelitian yang telah dilakukan, ada 100 data pengguna atau pasien yang diuji cobakan pada sistem, hasil diagnosa mencapai 87% dengan jumlah ketepatan diagnosa sebanyak 87 pasien. Sedangkan jumlah ketidak tepatan diagnosa sebanyak 13 orang pasien dengan nilai persentase sejumlah 20%.

4. SIMPULAN

Penelitian ini menggunakan metode Naive Bayes untuk mengembangkan sistem pakar dalam diagnosa gangguan pencernaan balita. Pengumpulan data dilakukan melalui studi pustaka, pengamatan, dan wawancara

dengan para pakar. Sistem ini mampu mengidentifikasi gejala-gejala yang muncul dan memberikan hasil diagnosa berdasarkan probabilitas menggunakan algoritma Naive Bayes.

Dari hasil pengujian, sistem pakar ini memiliki tingkat akurasi sebesar 87%, yang berarti mampu memberikan diagnosa yang sesuai dengan diagnosa dokter pada sebagian besar kasus. Meskipun demikian, sistem ini tidak selalu sempurna dan masih ada beberapa kasus di mana diagnosa sistem tidak sesuai dengan diagnosa dokter.

Kelebihan sistem pakar ini terletak pada kemampuannya memberikan bantuan diagnosa secara cepat dan dapat diakses oleh masyarakat umum. Sistem ini dapat menjadi alternatif yang efektif terutama bagi orang tua yang mungkin menghadapi kendala biaya dan waktu untuk berkonsultasi ke dokter.

Secara keseluruhan, pengembangan sistem pakar dengan metode Naive Bayes untuk diagnosa gangguan pencernaan pada balita memiliki potensi besar dalam memberikan bantuan awal bagi orang tua dan pasien dalam mengenali serta mengatasi masalah kesehatan pada balita.

5. SARAN

Algoritma naïve bayes merupakan algoritma umum yang sudah sering dipergunakan dalam penelitian kepakaran, hasil akurasi akan meningkat jika algoritma dapat dikombinasikan dengan algoritma lain yang menjadikan akurasi menjadi lebih tinggi lagi.

DAFTAR PUSTAKA

- [1] M. Z. A. Putra and M. Ary, 2021, *SISTEM PAKAR DIAGNOSIS PENYAKIT GANGGUAN SALURAN PENCERNAAN MENGGUNAKAN METODE CERTAINTY FACTOR*, eProsiding Sistem Informasi (POTENSI), vol. 2, no. 1, Art. no. 1, Accessed: Nov. 21, 2023. [Online]. Available: <https://eprosiding.ars.ac.id/index.php/psi/article/view/375>
- [2] A. S. R. Sinaga and D. Simanjuntak, 2020, *SISTEM PAKAR DETEKSI GIZI BURUK BALITA DENGAN METODE NAÏVE BAYES CLASSIFIER*, Jurnal Inkofar, vol. 1, no. 2, Art. no. 2, doi: 10.46846/jurnalinkofar.v1i2.110.
- [3] R. H. Saputra, J. A. Baba, and M. el-Khaeri Kesuma, 2022, *SISTEM PAKAR BERBASIS ANDROID UNTUK DIAGNOSIS PENYAKIT BALITA PADA USIA NEONATAL*, SEAT: Journal Of Software Engineering and Technology, vol. 1, no. 2, Art. no. 2, Accessed: Nov. 21, 2023. [Online]. Available: <https://journal.institdla.ac.id/index.php/seat/article/view/46>
- [4] R. Hartono, N. D. Sofya, and Nawassyarif, 2021, *APLIKASI PENDAFTARAN SERTIFIKASI KOMPETENSI PADA CAREER DEVELOPMENT CENTER (CDC) UNIVERSITAS TEKNOLOGI SUMBAWA BERBASIS WEB*, Hxg, vol. 2, no. 2, pp. 32–42, doi: 10.36761/hexagon.v2i2.1085.
- [5] N. I. Santosa, 2022, *Sistem Pakar Diagnosa Gangguan Pencernaan Menggunakan Metode Forward Chaining Berbasis Android*, (JurTI) Jurnal Teknologi Informasi, vol. 6, no. 2, Art. no. 2, doi: 10.36294/jurti.v6i2.2889.
- [6] E. Rahmanita, W. Agustiono, and R. Juliyanti, 2019, *SISTEM PAKAR DIAGNOSA PENYAKIT SALURAN PENCERNAAN DENGAN PERBANDINGAN METODE FORWARD CHAINING DAN DEMPSTER SHAFER*, Jurnal Simantec, vol. 7, no. 2, Art. no. 2, doi: 10.21107/simantec.v7i2.6743.
- [7] H. B. Wijayanto, J. D. Irawan, and A. Faisol, 2020, *SISTEM PAKAR DIAGNOSIS PENYAKIT DEMAM BERDARAH MENGGUNAKAN METODE CASE BASED REASONING BERBASIS WEB*, JATI (Jurnal Mahasiswa Teknik Informatika), vol. 4, no. 1, Art. no. 1, doi: 10.36040/jati.v4i1.2317.
- [8] M. R. Handoko and N. Neneng, 2021, *SISTEM PAKAR DIAGNOSA PENYAKIT SELAMA KEHAMILAN MENGGUNAKAN METODE NAIVE BAYES BERBASIS WEB*, Jurnal Teknologi dan Sistem Informasi, vol. 2, no. 1, Art. no. 1, doi: 10.33365/jtsi.v2i1.739.
- [9] I. P. D. Suarnatha and I. M. A. O. Gunawan, 2022, *Implementasi Metode Certainty Factor dalam Sistem Pakar Deteksi Penyakit Pencernaan pada Manusia*, Jurnal CoSciTech (Computer Science and Information Technology), vol. 3, no. 2, Art. no. 2, doi: 10.37859/coscitech.v3i2.3872.
- [10] T. R. S. Hari and S. Sumijan, 2021, *Sistem Pakar dengan Menggunakan Metode Naive Bayes dalam Mengidentifikasi Penyakit Karies pada Gigi Manusia*, Jurnal Sistim Informasi dan Teknologi, pp. 233–238, doi: 10.37034/jsisfotek.v3i4.71.
- [11] W. C. Wahyun and A. S. Sitio, 2020, *Pest Detection Expert System And Method Using Bayes Rice Diseases*, Journal of Computer Networks, Architecture and High Performance Computing, vol. 2, no. 2, Art. no. 2, doi: 10.47709/cnipc.v2i2.411.
- [12] V. Viviliani and R. Tanone, 2019, *Perancangan Sistem Pakar Diagnosis Penyakit pada Bayi dengan Metode Forward Chaining Berbasis Android*, Jurnal Teknik Informatika dan Sistem Informasi, vol. 5, no. 1, Art. no. 1, doi: 10.28932/jutisi.v5i1.1577.
- [13] D. R. Anamisa, A. Rachmad, M. Yusuf, A. Jauhari, R. D. T. Erdiansa, and M. Y. Hariyawan, 2021, *Classification of diseases for rice plant based on Naive Bayes classifier with a combination of PROMETHEE*, Commun. Math. Biol. Neurosci., vol. 2021, no. 0, Art. no. 0, doi: 10.28919/cmbn/6674.

Case Based Reasoning Untuk Diagnosa Penyakit Gigi Berbasis Web

Chervin Fradyacsyah¹, Miftahus Sholihin², Danang Bagus Reknadi³

^{1,2,3}Teknik Informatika, Fakultas Teknik, Universitas Islam Lamongan

E-mail: ¹chervinfradyacsyah@gmail.com, ²miftahus.sholihin@unisla.ac.id, ³danz.0907@unisla.ac.id

Abstrak – Kesehatan sangat penting bagi setiap orang salah satu yang perlu dijaga adalah kesehatan gigi, gigi merupakan bagian tubuh yang memiliki fungsi penting yaitu untuk melancarkan makanan yang masuk ke dalam mulut karena fungsinya sangat penting, perlu dilakukan pemeriksaan gigi minimal 6 bulan sekali, namun karena ketidaktahuan akan bahaya penyakit gigi apabila dibiarkan terus menerus kemudian keterbatasan waktu yang dimiliki dan lamanya proses pengidentifikasian penyakit oleh dokter membuat orang seringkali malas memeriksakan giginya. Dengan adanya permasalahan tersebut mendorong peneliti untuk membuat sistem pakar diagnosa penyakit gigi dengan menggunakan metode Case Based Reasoning yang dimaksudkan dapat mendiagnosa penyakit gigi secara akurat dan dapat membantu dokter lebih cepat dalam proses identifikasi penyakit, sistem pakar juga disertai dengan informasi dan langsung dikelola oleh pakar, sistem pakar ini menggunakan pemrograman web php dan database mysql. Dengan tujuan dapat menghasilkan perangkat lunak dengan metode sistem pakar Case-Based Reasoning (CBR) berbasis sistem pakar untuk mendiagnosis Gigi dan memudahkan seseorang untuk menggunakan perangkat lunak yang diadopsi dari pakar tanpa menggantikan peran seorang pakar dengan biaya yang relatif kecil. System yang dibangun mampu memberikan hasil diagnosa berdasarkan kasus lama dan menjadikan kasus baru sebagai pengetahuan baru untuk memberikan keputusan dengan menggunakan tingkat kepercayaan berdasarkan presentasi dari hasil perhitungan similarity.

Kata Kunci — Gigi, Case Based Reasoning (CBR), Sistem Pakar, Similarity

1. PENDAHULUAN

Perkembangan zaman dapat memiliki dampak yang kompleks pada kesadaran masyarakat mengenai kesehatan. Sementara beberapa masyarakat mungkin semakin sadar akan pentingnya kesehatan, yang lain mungkin mengalami penurunan kesadaran. Memang benar bahwa banyaknya aktivitas atau kegiatan yang dilakukan oleh masyarakat dapat menyebabkan mereka mengabaikan kesehatan, termasuk kesehatan gigi. Beberapa faktor yang menyebabkan kurangnya perhatian terhadap kesehatan gigi adalah kesibukan, kurangnya kesadaran akan pentingnya perawatan gigi, dan kurangnya pengetahuan tentang bagaimana merawat gigi dengan benar. Menurut penelitian, penyakit gigi sering kali menjadi masalah kesehatan yang paling sering dikeluhkan oleh masyarakat Indonesia, berada di posisi teratas dalam daftar 10 penyakit yang paling umum. Persepsi dan tindakan masyarakat terhadap kesehatan gigi masih belum mencapai tingkat yang optimal. yang berkontribusi terhadap prevalensi penyakit gigi yang tinggi.

Diagnosis penyakit gigi dibuat oleh dokter gigi profesional dengan memperhatikan gejala yang dirasakan dan bentuk kerusakan gigi yang terlihat. Namun, seringkali masyarakat mengatasi masalah kesehatan gigi berdasarkan pengetahuan yang mereka miliki, contohnya dengan menggunakan obat pereda rasa sakit. Penyebabnya, masyarakat belum melakukan penyuluhan tentang penyakit yang dideritanya. Selain itu, keterbatasan tenaga medis dalam hal waktu dan jumlah kegiatan juga menjadi faktor kurangnya kegiatan penyuluhan bagi masyarakat. Agar dapat mengatasi permasalahan tersebut, diperlukan sebuah sistem yang ahli dalam mendiagnosa penyakit gigi berdasarkan informasi klinis dari pasien [1].

Sistem pakar merupakan cabang dalam bidang kecerdasan buatan yang berkonsentrasi pada penggunaan pengetahuan manusia dalam komputer untuk memecahkan permasalahan secara sistematis, mirip dengan cara manusia memecahkan masalah [2]. Dalam sistem pakar, terdapat teknik atau metode yang digunakan untuk menghasilkan suatu solusi melalui proses inferensi. Dari berbagai macam metode yang ada salah satu metode dalam sistem pakar adalah metode *Case Based Reasoning* (CBR) yang memungkinkan untuk memecahkan masalah yang ditimbulkan dengan menggunakan solusi dalam kasus-kasus tersebut yang mampu mendiagnosis penyakit dengan tingkat akurasi yang tinggi serta mengambil inspirasi dari kecerdasan manusia, yaitu mencari solusi untuk permasalahan baru dengan memanfaatkan jawaban atau pengalaman dari permasalahan sebelumnya melalui proses perbandingan [3].

Metode *Case Based Reasoning* merupakan suatu teknik yang digunakan oleh para peneliti dalam pengembangan sistem pakar, dimana mereka memanfaatkan informasi dari kasus-kasus sebelumnya dan

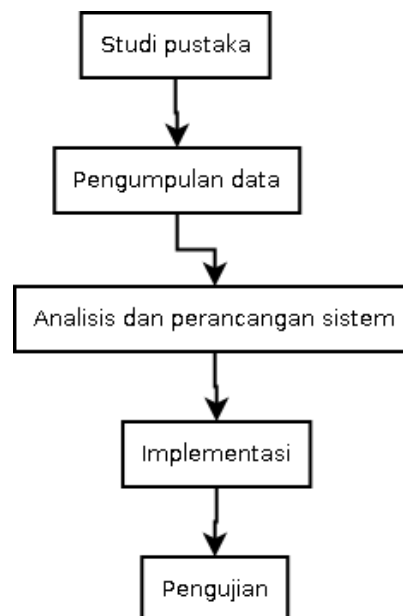
menerapkannya pada kasus baru berguna menemukan solusi yang tepat, kemudian mencari kasus dengan tingkat kemiripan tertinggi yang nantinya dapat mengambil keputusan dengan menggunakan *similarity* dengan *nearest neighbor retrieval* digunakan oleh penulis sebagai perhitungan nilai kesamaan antara kasus baru dengan kasus lama. Oleh karena itu diperlukan suatu sistem pakar yang mampu mendiagnosa penyakit gigi pada manusia dan memberikan solusi pencegahan dan pengobatan dini kepada pasien sebelum dibawa ke dokter gigi untuk evaluasi lebih lanjut.

Dari penelitian yang telah dibahas sebelumnya dimana masyarakat masih kurang kesadaran terhadap kesehatan gigi, maka pada penelitian ini peneliti membuat sebuah sistem yang mampu mendiagnosa penyakit gigi berbasis web dengan menggunakan metode *case based reasoning*. Dengan adanya sistem tersebut kesehatan gigi masyarakat bisa dilakukan deteksi dini secara mandiri.

2. METODE PENELITIAN

2.1 Tahapan Penelitian

Gambar 1 adalah tahapan-tahapan yang dilakukan pada penelitian ini.



Gambar 1 Tahapan Penelitian

- a) Studi Pustaka
Mengumpulkan informasi dengan cara mengumpulkan sumber-sumber tertulis, membaca, dan meneliti serta mencatat hal yang penting berkaitan dengan apa yang sedang diuji untuk mendapatkan gambaran teoritis.
- b) Pengumpulan Data
Pada tahap ini peneliti melakukan pengumpulan data kriteria penyakit gigi yang akan digunakan pada proses pembuatan aplikasi, yaitu melakukan wawancara terhadap pakar.
- c) Analisis dan Perancangan Sistem
Perancangan sistem merupakan tahap lanjutan setelah analisis sistem yang bertujuan untuk mengembangkan sistem yang lebih baik dan efektif.
- d) Implementasi
Pada tahapan ini peneliti mengimplementasikan hasil dari perancangan sistem kedalam bahasa pemrograman. Implementasi pada penelitian ini berbasis web.
- e) Pengujian
Pengujian ini adalah langkah paling akhir untuk menentukan data apakah sistem atau aplikasi dapat berfungsi normal seperti yang diharapkan.

2.2 Sistem Pakar

Sistem pakar yaitu program perilaku sebagai manusia ahli (*human expert*). Secara luas, sistem pakar, juga dikenal sebagai (*expert system*), adalah aplikasi komputer yang dibuat untuk mengeksekusi pengetahuan dan keterampilan seorang pakar dalam bidang tertentu[4–6].

Penggunaan sistem pakar dengan metode *Case Based Reasoning* lebih dipilih karena penggunaan metode ini mudah untuk mengambil kepastian berdasarkan kasus lama untuk mencari solusi dengan perbandingan kasus baru dalam database yang ada [7].

2.3 Case Based Reasoning

Metode pemecahan masalah yang disebut *Case Based Reasoning* (CBR) adalah penggunaan pengalaman dari kasus-kasus sebelumnya untuk memecahkan masalah yang baru muncul [8]. Konsep dasar CBR ialah bahwa dalam menghadapi suatu masalah baru, sistem mencari kasus serupa yang pernah terjadi sebelumnya, kemudian menggunakan solusi yang diterapkan pada kasus-kasus tersebut untuk menyelesaikan masalah baru [9,10].

2.4 Teknik Similarity

Untuk mencari kasus lama yang memiliki kemiripan dengan hasil kasus baru berdasarkan gejala-gejala yang dialami, dapat digunakan pendekatan berbasis kesamaan atau pemilihan berbasis gejala untuk menemukan kasus-kasus serupa. Persamaan 1 adalah untuk mencari *similarity* [11–14].

$$Similarity = \frac{s1*w1+s2*w2+\dots+sn*wn}{w1+w2+\dots+wn} \dots\dots\dots(1)$$

Dimana S : *similarity* (nilai kemiripan) yaitu 1 (sama) dan 0 (beda), sedangkan w adalah *weight* (bobot).

3. HASIL DAN PEMBAHASAN

3.1 Data Penelitian

Data pada penelitian ini adalah data penyakit gigi yang didapatkan dari wawancara secara langsung dengan pakar Drg. Sri Wulandari. Tabel 1 adalah tabel penyakit gigi, sedangkan Tabel 2 adalah tabel gejala penyakit.

Tabel 1. Jenis Penyakit Gigi

Kode Penyakit	Nama Penyakit
A01	Abses
A02	Gingivitis
A03	Periodontitis
A04	Retained Dental Root
A05	Pulpitis
A06	Neorsis Pulpa

Tabel 2. Gejala Penyakit

Kode Gejala	Nama Gejala	Bobot
G01	Gigi anda bermasalah	1
G02	Gigi anda mengalami sakit	5
G03	Gigi anda berlubang	5
G04	Gigi anda mengganggu penampilan	5
G05	Gigi anda ngilu rangsangan dominan dingin	3
G06	Berlubang pada mahkota	3
G07	Karies telah mengenai email	1
G08	Karies telah mengenai dentin	3
G09	Karies telah mengenai pulpa	5
G10	Berlubang pada pulpa	5
G11	Sensitif terhadap rangsangan	1
G12	Gigi anda bernanah	5
G13	Bernanah pada gusi	3
G14	Bernanah pada akar	5
G15	Gigi anda berdarah di daerah pulpa	5
G16	Gigi anda berdarah saat menggosok gigi (kalkulus menumpuk)	5
G17	Gigi anda berdarah akibat bakteri dari pulpa	5
G18	Gigi anda sakit akibat gigi patah	5

G19	Anda tidak bisa membuka mulut	5
G20	Gigi anda sakit di daerah gusi	3
G21	Gigi anda tampak terkikis (gigi menjadi rata serta mengkilap)	3
G22	Gigi anda tampak terkikis akibat menggosok gigi terlalu kuat	3
G23	Gigi anda tampak terkikis akibat bahan kimia atau obat	5
G24	Gigi anda terlihat kecil dari normalnya	3
G25	Gigi anda terlihat besar pada normalnya	3
G26	Gigi anda berubah warna karena kelebihan fluor (tampak bercak putih dan coklat)	1
G27	Gigi anda berubah warna akibat pengaruh obat(abu-abu, kuning, biru gelap)	5
G28	Gigi anda berubah warna pada email dan tampak kasar	5
G29	Terjadi bau mulut berasal dari pulpa	5
G30	Gigi anda berlebih pada normalnya	3
G31	Gigi anda tidak tumbuh semua / sebagian terdapat pada gigi susu maupun gigi permanen	5
G32	Tidak Berlubang Pada Pulpa	5
G33	Gusi Berdarah Saat Menggosok Gigi	5

Nilai bobot gejala didapatkan dari pakar seperti pada Tabel 3.

Tabel 3 Nilai Bobot Gejala Penyakit Gigi

Bobot parameter	Bobot
Gejala dominan	5
Gejala sedang	3
Gejala biasa	1

3.2 Pengujian Sistem

Program yang dibuat bertujuan untuk memudahkan pengguna dalam melakukan proses diagnose penyakit gigi berbasis web. Pengujian yang dilakukan pada penelitian ini adalah pengujian *Blackbox* dimana pengujian ini berfokus pada kebutuhan fungsional pada *software*, berdasarkan pada spesifikasi kebutuhan dari *software*. Tabel 4 adalah contoh hasil pengujian sistem yang dilakukan pada penelitian ini.

Tabel 4 Uji Coba Sistem User

Prosedur	Input	Output	Hasil
User masuk ke halaman utama website	<i>Login</i>	Menampilkan <i>login</i> untuk <i>user</i>	Berhasil membuka sistem dan menampilkan halaman <i>login</i>
Melakukan <i>login</i>	Memasukan <i>username</i> dan <i>password</i>	Menampilkan konsultasi	Berhasil melakukan <i>login</i>
Memilih menu konsultasi	Memilih data gejala-gejala	Menampilkan diagnosa	Berhasil melakukan konsultasi
Memilih menu <i>logout</i>	Menekan tombol <i>logout</i>	Dapat keluar dan kembali ke halaman utama	Berhasil melakukan <i>logout</i>

3.3 Uji Coba Aplikasi

Pada uji coba konsultasi ini peneliti akan memberikan hasil uji coba yang sudah dilakukan yakni dengan memilih beberapa gejala pada halaman konsultasi, seseorang melakukan konsultasi dan memilih gejala-gejala sebagai berikut:

1. [G33] Gusi berdarah saat menggosok gigi
2. [G02] Gigi anda mengalami sakit
3. [G29] Terjadi bau mulut berasal dari pulpa
4. [G20] Gigi anda sakit di daerah gusi
5. [G13] Bernanah pada gusi
6. [G14] Bernanah pada akar
7. [G15] Gigi anda berdarah di daerah pulpa

Selanjutnya, untuk mendapatkan hasil diagnosa, sistem membandingkan gejala baru yang dipilih selama konsultasi dengan gejala lama yang ada pada kasus yang disimpan sebelumnya. Perhitungan kesamaan sebagai metode untuk mengukur similarity diidentifikasi dan dilakukan dengan memasukkan kasus baru ke dalam sistem yang berisi gejala yang digunakan untuk mengidentifikasi atau mengklasifikasikan kasus baru berdasarkan kesamaan atau kemiripan dengan kasus-kasus yang telah diketahui sebelumnya. Sistem dibangun dengan 33 gejala untuk 6 jenis penyakit. Setiap gejala memiliki bobot yang berbeda dimana nilai ditentukan pakar, pengelompokan gejala dan bobot ada pada Tabel 5.

Tabel 5 Pengelompokan Gejala dan Bobot dari Uji Coba Aplikasi

No	Penyakit	Gejala Kasus (lama)	Bobot	Gejala Dipilih (baru)	Bobot
1	[A01] Abses	G01	1	G29	5
		G02	5	G20	3
		G04	5	G13	3
		G05	3	G02	5
		G06	3	G15	5
		G07	1		
		G08	3		
		G13	3		
		G15	5		
		G20	3		
2	[A03] Periodontitis	G01	1	G29	5
		G02	5	G20	3
		G13	3	G13	3
		G14	5	G14	5
		G20	3	G33	5
		G29	5	G02	5
		G33	5		
3	[A05] Pulpitis	G01	1	G29	5
		G02	5	G15	5
		G03	5	G02	5
		G05	3		
		G15	5		
		G17	5		
		G21	3		
		G22	3		
G29	5				

a. Proses reuse penyakit abses

Selama proses ini, perhitungan kemiripan kasus dilakukan antara kasus lama dan kasus baru yang sudah dikelompokkan pada Tabel 5, proses melibatkan langkah-langkah berikut:

Perhitungan *Similarity (problem, case)* penyakit [A01] Abses

$$= (1*5+1*3+1*3+1*5+1*5) / (1+5+5+3+3+1+3+3+5+3+5)$$

$$= 21 / 37 * 100$$

$$= 56,75$$

Dari hasil perhitungan penyakit Abses yang telah dilakukan di atas maka dapat diperoleh hasil dengan nilai sebesar $56,75 = 57\%$.

b. Proses reuse penyakit periodontitis

Selama proses ini, perhitungan kemiripan kasus dilakukan antara kasus lama dan kasus baru yang sudah dikelompokkan pada Tabel 5, proses melibatkan langkah-langkah berikut:

Perhitungan *Similarity (problem, case)* penyakit [A03] Periodontitis

$$= (1*5+1*3+1*3+1*5+1*5+1*5) / (1+5+3+5+3+5+5)$$

$$= 26 / 27 * 100 = 96,29$$

Dari hasil perhitungan penyakit *Periodontitis* yang telah dilakukan di atas maka dapat diperoleh hasil dengan nilai sebesar $96,29 = 96\%$.

c. Proses reuse penyakit *pulpitis*

Selama proses ini, perhitungan kemiripan kasus dilakukan antara kasus lama dan kasus baru yang sudah dikelompokkan pada Tabel 5, proses melibatkan langkah-langkah berikut:

Perhitungan *Similarity (problem, case)* penyakit [A05] *Pulpitis*

$$= (1*5+1*5+1*5) / (1+5+5+3+5+5+3+3+5)$$

$$= 15 / 35 * 100$$

$$= 42,85$$

Dari hasil perhitungan penyakit *Pulpitis* yang telah dilakukan di atas maka dapat diperoleh hasil dengan nilai sebesar $42,85 = 43\%$.

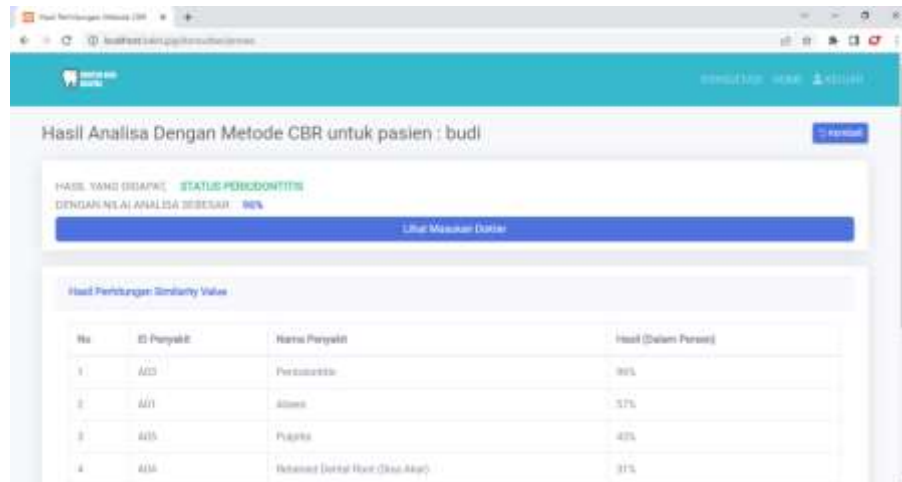
Dari hasil perhitungan *case based reasoning* di atas yakni hasil *user* sesudah memilih gejala gusi berdarah saat menggosok gigi, gigi anda mengalami sakit, terjadi bau mulut berasal dari pulpa, gigi anda sakit di daerah gusi, bernanah pada gusi bernanah pada akar, gigi anda berdarah di daerah pulpa. Sesudah dilakukan proses perhitungan kemiripan (*similarity*) dengan *nearest neighbor retrieval* maka didapatkan hasil dari persentase penyakit yang diderita yakni *Periodontitis* 96%, *Abses* 57%, dan *Pulpitis* 43%. Hasil yang paling mendominasi yakni *Periodontitis* sebesar 96%.

3.4 Implementasi Aplikasi

Gambar 2 adalah halaman yang ditampilkan pada saat *user* sesudah *login* dan memilih menu konsultasi. Dimenu ini, pengguna disajikan berbagai gejala. Selanjutnya, pengguna harus memilih dan sistem akan memproses diagnosa gejala yang dialami pasien saat ini.



Gambar 2 Halaman Konsultasi



Gambar 3 Hasil Diagnosa

Gambar 3 adalah halaman konsultasi setelah mengklik tombol konsultasi, dan kemudian halaman lain dengan hasil diagnosis dari kelompok gejala yang telah dipilih sebelumnya. Halaman ini memberi pengguna ringkasan skor gejala yang dipilih, yang kemudian digabungkan menjadi skor akhir yang mencerminkan kondisi yang paling sering dialami pasien.

Gambar 4 merupakan tampilan dari lanjutan hasil diagnosa yang mana halaman ini adalah untuk menginformasikan kepada penderita dengan memberikan solusi yang berisi definisi dari penyakit yang paling dominan diderita oleh pasien.



Gambar 4 Halaman Informasi & Solusi

4. SIMPULAN

Berdasarkan uji coba yang sudah dilakukan, *system* pakar diagnosa penyakit gigi dengan menggunakan metode *case based reasoning* ini bisa membantu dokter. Selain itu *system* yang dibangun ini mampu memberikan manfaat bagi masyarakat umum untuk melakukan diagnosa dini terhadap penyakit gigi. *System* yang dibangun mampu memberikan hasil diagnosa berdasarkan kasus lama dan menjadikan kasus baru sebagai pengetahuan baru untuk memberikan keputusan dengan menggunakan tingkat kepercayaan berdasarkan presentasi dari hasil perhitungan *similarity*.

5. SARAN

Aplikasi ini hanya untuk mendiagnosis penyakit gigi saja, diharapkan untuk pengembangan selanjutnya dapat digunakan untuk diagnosa penyakit lainnya dengan lebih banyak gejala untuk membawa lebih banyak manfaat bagi pengguna. Selain itu, penelitian selanjutnya diharapkan menambahkan beberapa metode algoritma lain yang dapat dipertimbangkan untuk membandingkan tingkat keakuratan dalam penelitian selanjutnya.

DAFTAR PUSTAKA

- [1] Nas C and Kusnadi K 2019 Sistem Pakar Diagnosa Penyakit Gigi Dan Mulut Menggunakan Metode Case-Based Reasoning *J. Digit Digit. Inf. Technol.* **9** 202–14
- [2] Asdar A, Saputra R A and Ningrum I P 2022 Sistem Pakar Penyakit Kulit Pada Manusia Menggunakan Metode Case Base Reasoning (CBR) Dengan Algoritma Sorensen Coefficient *JUMANJI J. Masy. Inform. Unjani* **6** 48–62
- [3] Gale R D S 2019 SISTEM PAKAR DIAGNOSA PENYAKIT TIDAK MENULAR MENGGUNAKAN METODE CASE BASED REASONING *Semin. Nas. Konf. Ilm. Sist. Inf. Inform. Komun.* 894–902
- [4] Afriansyah A, Annisa R and Mair Z R 2019 SISTEM PAKAR DETEKSI KERUSAKAN (TROUBLESHOOTING) PADA SMARTPHONE BERBASIS MOBILE *J. TIPS J. Teknol. Inf. Dan Komput. Politek. Sekayu* **9** 17–25
- [5] Zamroni M R, S Q C K N and Wahyudi A 2022 SISTEM PAKAR DIAGNOSA PENYAKIT SAPI SEBAGAI UPAYA PENCEGAHAN PENYEBARAN WABAH PMK DI LAMONGAN *J. Ilm. Inform.* **10** 145–52
- [6] Arista R, Sholihin M and Nafi'iyah N 2019 SISTEM PENENTUAN KUALITAS PISANG MENGGUNAKAN METODE FUZZY TSUKAMOTO *J. Inf. Syst. Manag. JOISM* **1** 27–30
- [7] Rajutidesli R 2020 Penerapan Metode Case Based Reasoning Dan Certainty Factor Dalam Perancangan Sistem Pakar Diagnosa Penyakit Flu Burung Berbasis Web *KOMIK Konf. Nas. Teknol. Inf. Dan Komput.* **4**
- [8] Minarni M and Warman I 2017 Sistem Pakar Identifikasi Penyakit Tanaman Padi Menggunakan Case-Based Reasoning *Semin. Nas. Apl. Teknol. Inf. SNATI*
- [9] Sihaloho T P, Tarigan W, Siallagan S and Simbolon F H 2022 MODEL CASE BASED REASONING DALAM MENDIAGNOSA PENYAKIT KELAPA SAWIT *J. Mnemon.* **5** 178–83
- [10] Telambanua N J, Nofriadi N and Dermawan A 2022 Sistem Pakar Untuk Mendeteksi Penyakit Mata Menerapkan Metode Case Based Reasoning *Build. Inform. Technol. Sci. BITS* **4** 570–80
- [11] Sitorus A J, Hutagalung J E and Dermawan A 2022 Sistem Pakar Diagnosa Penyakit Pencernaan Menggunakan Metode Case Based Reasoning (CBR) Berbasis Web *J. MEDIA Inform. BUDIDARMA* **6** 2214–21
- [12] Rosmini R, Syafiqoh U and Asmah A 2023 Implementasi Case Based Reasoning (CBR) Untuk Pengembangan Sistem Pakar Diagnosa Penyakit Gigi *J. MEDIA Inform. BUDIDARMA* **7** 1455–62
- [13] Faisal Z A 2019 SISTEM PAKAR DIAGNOSA PENYAKIT AYAM PETELUR MENGGUNAKAN METODE CASE BASED REASONING BERBASIS WEB *JATI J. Mhs. Tek. Inform.* **3** 126–32
- [14] Aldo D, Samosir K and Nengsih Y G 2021 Sistem Pakar Memilih Tipe Belajar Anak Retardasi Mental Menggunakan Metode Case Based Reasoning *J. Sci. Appl. Inform.* **4** 95–105

Identifikasi Tingkat Kesadaran Pengemudi dari Data Video dengan Menggunakan Convolutional Long Short Term Memory

Fajar Fatha Romadhan¹, Andrian Dwi Baitur Rizky², Muhammad Aulia Faqihuddin³, Indah Agustien Siradjuddin⁴

^{1,2,3,4}Teknik Informatika, Fakultas Teknik, Universitas Trunojoyo Madura

E-mail: *¹200411100047@student.trunojoyo.ac.id, ²200411100027@student.trunojoyo.ac.id,

³200411100210@student.trunojoyo.ac.id, ⁴indah.siradjuddin@trunojoyo.ac.id

Abstrak – Mengemudi dalam keadaan tidak sadar atau mengantuk merupakan salah satu penyebab utama terjadinya kecelakaan lalu lintas. Meskipun demikian, penumpang biasanya tidak akan menyadari jika pengemudi kendaraan yang mereka kendarai dalam keadaan mengantuk. Untuk mendeteksi kantuk pengemudi pada umumnya adalah dengan menggunakan model pembelajaran mesin. Akan tetapi model pembelajaran mesin biasanya hanya dapat mengenali pola dari sebuah data tanpa memperhatikan deret waktu pada data sekuensial atau time series. Metode yang digunakan adalah Convolutional Long Short Term Memory. Convolutional Long Short Term Memory adalah salah satu bentuk Recurrent Neural Network yang dapat digunakan untuk mengatasi masalah korelasi deret waktu baik dalam waktu singkat maupun lama pada data spatial time series. Data yang digunakan adalah data yang berupa video dari seseorang yang sedang mengemudi. Data video merupakan data berurutan yang terdiri atas banyak frame yang berupa citra. Pada penelitian ini setiap frame tersebut dilakukan deteksi wajah untuk mendapatkan citra wajah pengemudi, sedangkan Convolutional Long Short Term Memory digunakan untuk mempelajari pola perubahan ekspresi wajah pengemudi seiring waktu. Hasil menunjukkan model memiliki akurasi sebesar 75% yang menunjukkan bahwa model memiliki kemampuan yang signifikan dalam memprediksi kantuk pengemudi.

Kata Kunci — Convolutional, Data Video, Deteksi, Kantuk Pengemudi, Long Short Term Memory

1. PENDAHULUAN

Mengemudi dalam keadaan mengantuk merupakan salah satu penyebab utama terjadinya kecelakaan lalu lintas. Berdasarkan data dari World Health Organization (WHO), rasa kantuk pengemudi menyebabkan sekitar 20% kecelakaan lalu lintas di dunia. Menurut laporan dari National Highway Traffic Safety Administration (NHTSA) di Amerika Serikat lebih dari 1,25 juta kematian setiap tahunnya, dan 20-50 juta orang terluka atau cacat disebabkan oleh mengemudi dalam keadaan mengantuk[1]. Di Indonesia sendiri juga sangat banyak kecelakaan lalu lintas yang disebabkan oleh rasa kantuk pengemudi. Karena tingginya angka kecelakaan yang disebabkan oleh rasa kantuk pengemudi, maka perlu mengambil tindakan tepat waktu[2].

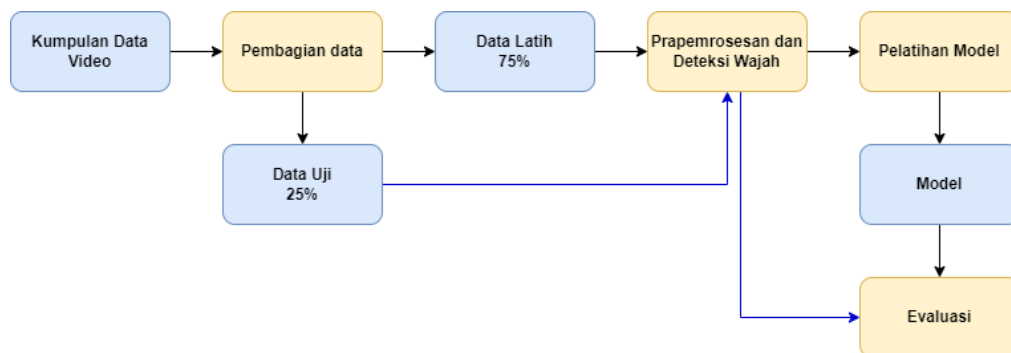
Penumpang biasanya tidak akan menyadari jika pengemudi dalam keadaan mengantuk. Mendeteksi rasa kantuk pengemudi dapat dilakukan menggunakan berbagai cara, termasuk dengan menggunakan teknologi komputer. Dengan memanfaatkan teknologi komputer, penumpang dapat mengetahui jika pengemudi dalam keadaan mengantuk sehingga dapat mengantisipasi terjadinya kecelakaan.

Metode yang umum digunakan untuk mendeteksi kantuk pengemudi adalah dengan menggunakan model pembelajaran mesin. Pembelajaran mesin adalah bagian dari bidang kecerdasan buatan (*Artificial Intelligence*) yang melibatkan pemrograman untuk mengaktifkan komputer agar dapat berperilaku cerdas seperti manusia, bahkan mereka dapat meningkatkan pemahaman mereka melalui pengalaman yang diperoleh secara otomatis selama pelatihan[3]. Akan tetapi model pembelajaran mesin biasanya hanya dapat mengenali pola dari sebuah data tanpa memperhatikan deret waktu pada data sekuensial atau time series. Pada penelitian [2] identifikasi kelelahan pengemudi dengan menggunakan pengenalan iris menghasilkan akurasi sekitar 80%. Pada penelitian [4] deteksi kelelahan dan kantuk menggunakan *deep convolutional network* menghasilkan rata-rata akurasi sebesar 68%, 81%, 91%, dan 93% pada skenario yang berbeda. Pada penelitian [5] deteksi kantuk pengemudi menggunakan analisis electrooculogram dan optimasi *grey wolf* menghasilkan akurasi 99,6%. Pada penelitian [6] deteksi kantuk pengemudi menggunakan optimasi *random forest* dan *single-channel electroencephalogram* menghasilkan akurasi sebesar 98,5%. Pada penelitian terkait sudah memiliki akurasi yang tinggi. Meskipun demikian, persamaan dari beberapa penelitian terkait tersebut adalah tidak memperhatikan pola perubahan pengemudi seiring dengan waktu. Padahal pola perubahan tersebut juga dapat digunakan untuk mengidentifikasi tingkat kesadaran atau kantuk pengemudi

Oleh karena itu, penelitian ini bertujuan untuk mengembangkan model yang dapat mendeteksi kantuk pengemudi dengan mempelajari pola perubahan ekspresi wajah pengemudi seiring dengan waktu dengan menggunakan metode pembelajaran mendalam. Metode pembelajaran mendalam digunakan karena popularitasnya dalam dunia ilmu komputer dan terbukti efektif dalam mengklasifikasikan dari kumpulan video. Metode pembelajaran mendalam menunjukkan kuat dan tepat untuk mengenali tindakan manusia[7]. Salah satu hal baru dari penelitian ini adalah penerapan langkah-langkah pra-pemrosesan pada video sebelum tahap pelatihan dan pengujian untuk meningkatkan performa model.

2. METODE PENELITIAN

Proses identifikasi tingkat kesadaran atau kantuk pengemudi ini dilakukan dengan melakukan klasifikasi biner dari kumpulan data video. Dari kumpulan data video tersebut akan diproses sedemikian rupa sehingga sesuai dengan arsitektur model yang dibangun. Kumpulan data video yang sudah diproses tersebut digunakan untuk melatih model hingga model tersebut dapat mengidentifikasi suatu video apakah orang dalam video tersebut dalam keadaan mengantuk atau tidak. Alur Proses identifikasi dapat dilihat pada diagram Gambar 1.



Gambar 1. Alur Proses Identifikasi

2.1 Data yang digunakan

Kumpulan data yang digunakan pada penelitian ini merupakan data yang berupa video dari seseorang yang sedang mengemudi. Video diambil dengan posisi kamera berada di *dashboard* kendaraan. Video diambil dengan pengemudi dalam dua kondisi yaitu pada saat mengantuk dan tidak mengantuk.

Kumpulan data didapatkan dari kaggle dengan nama SUST-DDD atau Sivas University of Science and Technology Driver Drowsiness Datasets. Kumpulan data ini terdiri dari dua kelas yaitu *drowsiness*, dan *not drowsiness*. Kumpulan data ini berupa video dengan jumlah video untuk kelas *drowsiness* 975 dan *not drowsiness* 1099. Kumpulan data ini dapat diakses pada <https://www.kaggle.com/datasets/esrakavalci/sust-ddd>. Tampilan data video dapat dilihat pada Gambar 2.



Gambar 2. Data Video

2.2 Praproses dan deteksi wajah

Pada praproses data, data yang berupa video akan diekstrak menjadi 20 *frame* per video. Setiap *frame* dilakukan deteksi wajah dengan menggunakan library *face_recognition* dari *dlib* python. Hal tersebut dilakukan

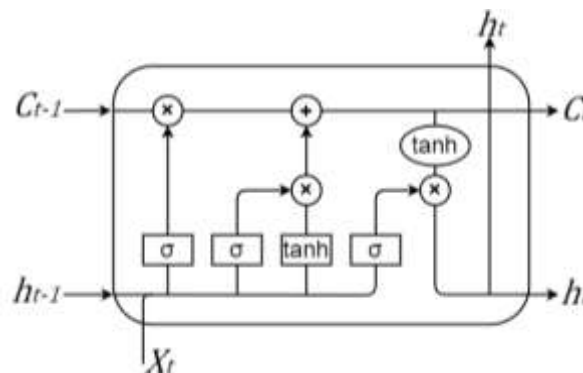
karena dalam pengenalan kantung yang dibutuhkan hanyalah ekspresi wajah dari pengemudi. Selain itu hasil dari proses klasifikasi pixel memiliki dampak yang besar pada akurasi model jika dengan mengambil bagian wajahnya saja atau tidak[8]. Setelah deteksi wajah *frame* tersebut akan di ubah ukurannya menjadi 200×200 *pixel* agar tidak terlalu besar. Setelah ukurannya di ubah, selanjutnya *frame* akan dinormalisasi.

2.3 Metode yang Digunakan

Konvolusi merupakan sebuah proses yang digunakan untuk mengekstrak fitur dari data spasial seperti pada citra. Konvolusi bekerja dengan melibatkan kernel atau filter konvolusi. Kernel akan dikalikan menggunakan perkalian titik dan digeser secara berurutan melintasi seluruh citra input. Hasil perkalian titik tersebut dijumlahkan sehingga menghasilkan nilai tunggal dan menghasilkan *feature map*. Pemanfaatan operasi konvolusi dalam *neural network* dapat dilihat pada CNN. CNN atau Convolutional Neural Network merupakan arsitektur jaringan syaraf tiruan yang memanfaatkan layer konvolusi. CNN banyak digunakan sebagai layer yang digunakan untuk mengekstrak fitur dari data inputan seperti citra. Layer konvolusi pada CNN dinilai dapat mengurangi beban komputasi dari *neural network*[9]. CNN memiliki performa yang bagus meskipun tidak memerlukan tahap ekstraksi fitur dikarenakan CNN memiliki ekstraksi fitur tersendiri[10]. CNN melakukan ekstraksi fitur menggunakan lapisan konvolusi, fungsi aktivasi, dan lapisan *pooling*[10].

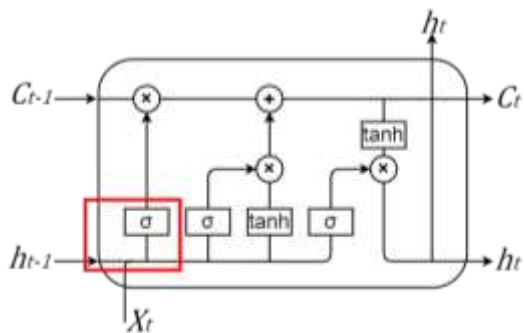
Recurrent Neural Network (RNN) merupakan suatu metode dalam *machine learning* yang digunakan untuk memproses data sekuensial atau *time series*. Pada data yang berbentuk sekuensial atau *time series*, nilai data $t+1$ akan dipengaruhi oleh data $t-0$ dan $t-1$. RNN memiliki kemampuan untuk mengingat input sebelumnya dengan menggunakan konsep memori internal atau *state* yang memungkinkan informasi sebelumnya dapat mempengaruhi pemrosesan informasi saat ini. RNN secara teori dapat melihat *long term dependency* jauh-jauh dari masa lampau, sehingga semua *input* dari masa lampau dapat memengaruhi *output*. Tapi pada implementasinya RNN memiliki masalah dengan *input* yang terlalu jauh di masa lampau. Untuk mengatasi hal tersebut arsitektur RNN dikembangkan sehingga tercipta arsitektur yang dikenal sebagai Long Short Term Memory (LSTM).

LSTM merupakan salah satu bentuk RNN yang dapat mengatasi masalah korelasi deret waktu baik dalam waktu singkat maupun lama dengan menggunakan lapisan tersembunyi sebagai sel memori. Ciri khas dari LSTM yaitu selain ada *hidden state* yang diberikan ke *time step* berikutnya terdapat juga *cell state*. *Cell state* ini seperti *state internal* pada LSTM. *Cell state* ini akan membawa informasi dari masa lampau. Informasi tersebut dapat dilupakan atau dibuang atau dapat ditambahkan pada *time step* berikutnya sehingga dapat diproses di *hidden state*. Dalam LSTM terdapat beberapa komponen penting, yaitu *Forget Gate*, *Input Gate*, dan *Output Gate*. LSTM memiliki arsitektur seperti pada Gambar 3:



Gambar 3. Diagram LSTM

Dari diagram pada Gambar 3 yang pertama kali dilakukan yaitu menghitung *forget gate*. *Forget gate* dalam LSTM berfungsi untuk menentukan seberapa banyak informasi dari *time step* sebelumnya yang akan dilupakan atau diabaikan dalam proses pengolahan informasi berikutnya. Nantinya *forget gate* ini akan dikalikan dengan *cell state time step* sebelumnya. Jika *forget gate* bernilai 0 maka *cell state time step* sebelumnya dilupakan atau dibuang. Jika *forget gate* bernilai 1 maka *cell state time step* sebelumnya dilanjutkan ke *time step* saat ini. Diagram *forget gate* dapat dilihat pada Gambar 4.



Gambar 4. Forget Gate

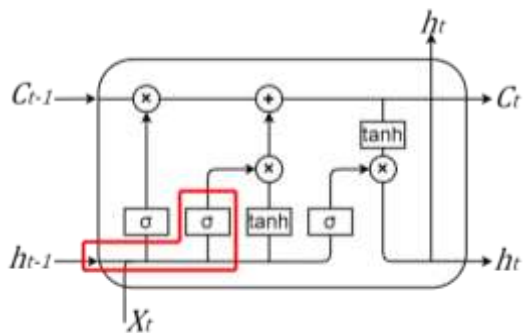
Forget gate dihitung dari gabungan *input* saat ini dan *hidden state time step* sebelumnya dengan *weight* dan *bias* yang kemudian dimasukkan ke dalam fungsi sigmoid. *Forget gate* dapat dihitung dengan Persamaan 1.

$$f_t = \sigma(W_f * [h_{t-1}, x_t] + b_f) \dots \dots \dots (1)$$

Dimana :

- f_t : *forget gate*
- σ : fungsi sigmoid
- W_f : *weight* atau bobot pada *forget gate*
- h_{t-1} : *hidden state time step* sebelumnya
- x_t : *input time step* saat ini
- b_f : *bias* pada *forget gate*

Setelah menghitung *forget gate*, berikutnya yaitu menghitung *input gate*. *Input gate* berfungsi untuk mengontrol seberapa banyak informasi baru dari *input* saat ini yang akan dimasukkan ke dalam memori jangka panjang sel LSTM. Diagram *input gate* dapat dilihat pada Gambar 5.



Gambar 5. Input Gate

Nilai dari *input gate* ini akan dikalikan per elemen dengan *inputnya* lalu ditambahkan ke *cell state*. *Input gate* melambangkan seberapa banyak *input* yang akan ditambahkan ke *cell state* baru. *Input gate* dapat dihitung menggunakan Persamaan 2.

$$i_t = \sigma(W_i * [h_{t-1}, x_t] + b_i) \dots \dots \dots (2)$$

Dimana :

- i_t : *input gate*
- σ : fungsi sigmoid
- W_i : *weight* atau bobot pada *input gate*
- h_{t-1} : *hidden state time step* sebelumnya
- x_t : *input time step* saat ini
- b_i : *bias* pada *input gate*

Setelah mendapatkan *input gate*, berikutnya yaitu menghitung *cell state candidate*. *Cell state candidate* ini nantinya akan digunakan untuk menghitung *cell state* baru. Untuk menghitung *cell state candidate* dapat menggunakan Persamaan 3

$$\tilde{C}_t = \tanh(W_c * [h_{t-1}, x_t] + b_c) \dots \dots \dots (3)$$

Dimana :

- \tilde{C}_t : cell state candidate
- \tanh : fungsi tangen hiperbolik
- W_c : weight atau bobot pada cell state
- h_{t-1} : hidden state time step sebelumnya
- x_t : input time step saat ini
- b_c : bias pada cell state

Hasil dari cell state candidate akan dikalikan dengan input gate pada perhitungan Cell state baru. Untuk menghitung cell state baru dapat menggunakan Persamaan 4.

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \dots \dots \dots (4)$$

Dimana :

- C_t : Cell state baru
- f_t : forget gate
- C_{t-1} : cell state time step sebelumnya
- i_t : input gate
- \tilde{C}_t : cell state candidate

Pada perhitungan cell state baru, hasil dari forget gate akan dikalikan dengan cell state time step sebelumnya menggunakan perkalian antar elemen. Sehingga jika forget gate menghasilkan nilai 0 maka cell state pada time step sebelumnya akan dilupakan atau dibuang, sedangkan jika forget gate menghasilkan nilai 1 maka cell state dari time step sebelumnya akan dilanjutkan ke cell state pada time step saat ini. Setelah mendapatkan cell state baru, cell state tersebut akan diproses untuk menjadi hidden state pada time step saat ini.

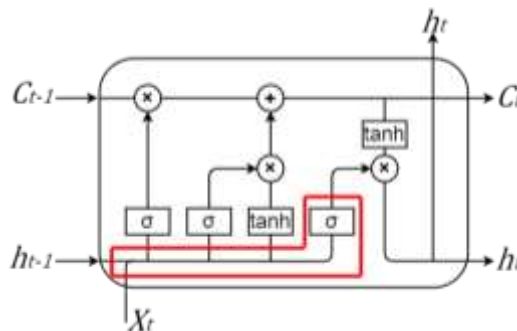
Jumlah output yang akan menjadi hidden state akan diatur oleh output gate. Untuk menghitung output gate dapat menggunakan Persamaan 5.

$$o_t = \sigma(W_o * [h_{t-1}, x_t] + b_o) \dots \dots \dots (5)$$

Dimana :

- o_t : output gate
- σ : fungsi sigmoid
- W_o : weight atau bobot pada output gate
- h_{t-1} : hidden state time step sebelumnya
- x_t : input time step saat ini
- b_o : bias pada input gate

Hasil dari output gate akan digunakan untuk menghitung output. Output ini akan digunakan sebagai hidden state time step saat ini. Diagram output gate dapat dilihat pada Gambar 6.



Gambar 6. Output Gate

Output dihitung dengan mengalikan output gate dan ditambahkan dengan hasil tangen hiperbolik dari cell state baru. Output dapat dihitung dengan Persamaan 6.

$$h_t = o_t * \tanh (C_t) \dots\dots\dots(6)$$

Dimana :

h_t : *output* atau *hidden state*

o_t : *output gate*

\tanh : fungsi tangen hiperbolik

C_t : *cell state time step* saat ini

Beberapa *gates* yang ada pada LSTM memberikan kontrol terhadap informasi yang dapat disimpan atau diabaikan, sehingga membuatnya menjadi alat yang lebih mumpuni untuk memahami hubungan kompleks dalam data [11].

LSTM sebagai pengembangan dari RNN telah terbukti stabil dan kuat dalam memodelkan ketergantungan jarak jauh dalam berbagai studi sebelumnya[12]. Meski demikian, LSTM tidak dapat menerima input data spasial seperti citra. Arsitektur LSTM dirancang untuk memproses data yang berurutan seperti suara, teks, atau data *time series*. Berbeda dengan citra yang pada dasarnya merupakan data spasial dengan dimensi yang lebih besar daripada data yang berurutan.

Data video merupakan data berurutan yang terdiri atas banyak frame yang berupa citra. Untuk memproses data video, arsitektur LSTM dapat dikombinasikan dengan CNN. Arsitektur ini dinamakan Convolutional LSTM. Berbeda dengan LSTM, Convolutional LSTM dapat menerima input berupa citra dan menyimpan informasi spasialnya sehingga memfasilitasi rekonstruksi data[13]. Dalam Convolutional LSTM, layer konvolusi digunakan untuk mengekstrak fitur spasial dari citra. Sedangkan LSTM digunakan untuk mengolah dan memahami urutan atau deret waktu pada video.

2.4 Arsitektur Model

Arsitektur model yang dibangun terdiri dari 5 layer, yaitu Convolutional LSTM 2D, maxpooling, timedistributed, Flatten, dan dense. Rincian layer yang digunakan dapat dilihat pada Tabel 1.

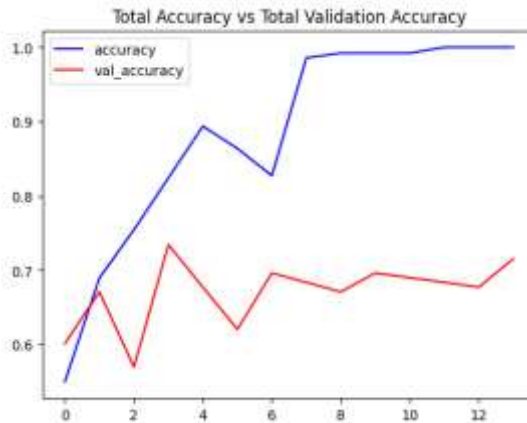
Layer	Parameter	Output Shape
Convolutonal LSTM 2D	filters = 16, kernel_size = (3, 3), activation = 'tanh', data_format = "channels_last", recurrent_dropout=0.2, return_sequences=True, input_shape = (20, 200, 200, 3)	(None, 20, 198, 198, 16)
Convolutional LSTM 2D	filters = 8, kernel_size = (3, 3), activation = 'tanh', recurrent_dropout=0.2, return_sequences=True,	(None, 20, 196, 196, 8)
Max Pooling 3D	pool_size=(1, 2, 2), padding='same', data_format='channels_last'	(None, 20, 98, 98, 8)
Time Distributed	Dropout(0.2)	(None, 20, 98, 98, 8)
Flatten	-	(None, 1536640)
Dense	Sesuai jumlah kelas (2), Activation="softmax"	(None, 2)

3. HASIL DAN PEMBAHASAN

Dari penelitian yang dilakukan telah mencapai hasil yang signifikan. Pada penelitian ini jumlah data yang digunakan hanya 900 video pada setiap kelas, sehingga total data yang digunakan adalah 1800 video. Pada proses deteksi wajah, dari 1800 video yang digunakan, hanya terdeteksi 1234 video dengan rincian 708 video kelas mengantuk dan 536 video tidak mengantuk. Jumlah data setiap kelas tersebut tidak sama atau tidak seimbang. Karena dikhawatirkan akan mempengaruhi hasil prediksi maka dari data tersebut dilakukan penyeimbangan data. Penyeimbangan data menggunakan undersampling sehingga kelas dengan jumlah data terbanyak akan dikurangi sehingga menjadi sama dengan jumlah data pada kelas dengan data paling sedikit. Pada kasus ini data dari kelas mengantuk akan dikurangi yang sebelumnya 708 video, menjadi 526 video.

Sehingga total data yang digunakan yaitu 1052 data video. Dari 1052 data video yang ada akan dibagi menjadi data untuk proses pelatihan dan data untuk proses evaluasi. Pembagian data untuk pelatihan sebanyak 75% dan data untuk evaluasi 25%. Sehingga jumlah data untuk pelatihan yaitu 789 video dan jumlah data untuk evaluasi 263 video.

Dari proses pelatihan dengan epoch sebanyak 50 dengan optimizer adam dan learning rate default sebesar 0,001 dan proses validasi, model yang telah dibangun menghasilkan akurasi yang signifikan. Grafik peningkatan akurasi pada proses pelatihan dapat dilihat pada Gambar 7.



Gambar 7. Grafik akurasi pelatihan

Dari Gambar 2 dapat diambil kesimpulan bahwa proses pelatihan berhenti pada epoch ke 14 dikarenakan tidak ada perubahan pada akurasi dan *loss*. Akurasi pelatihan mencapai 100% sedangkan akurasi validasinya sebesar 72%. Selain itu dilakukan juga proses evaluasi model menggunakan data evaluasi sebesar 25% dari total data yang digunakan. Sehingga data yang digunakan untuk proses evaluasi yaitu 263 video. Dari proses evaluasi yang dilakukan, menghasilkan confusion matriks seperti pada Tabel 2.

Tabel 2. Confusion Matriks

Class	True Positive	False Positive	False Negative	True Negative
Mengantuk	99	32	33	99
Tidak Mengantuk	99	33	32	99

Pada Tabel 2 disajikan tabel confusion matrix yang mencerminkan hasil evaluasi kinerja prediksi untuk huruf kelas mengantuk dan tidak mengantuk. Dari confusion matriks pada Tabel 2 dapat disimpulkan jika model yang dibuat memiliki hasil prediksi benar lebih banyak daripada prediksi salah. Dari Tabel 2 dapat dihitung nilai akurasi dengan Persamaan 7.

$$\begin{aligned}
 \text{Akurasi} &= \frac{\text{Banyak prediksi benar}}{\text{Total data yang diprediksi}} \dots\dots\dots(1) \\
 &= \frac{TP + TN}{TP + FP + TN + FN}
 \end{aligned}$$

Dimana :

TP : True Positive (Jumlah data yang diprediksi secara tepat sebagai kelas positif)

FP : False Positive (Jumlah data yang diprediksi secara tidak tepat sebagai kelas positif)

TN: True Negative (Jumlah data yang diprediksi secara tidak tepat sebagai kelas negatif)

FN: False Negative (Jumlah data yang diprediksi secara tidak tepat sebagai kelas negatif)

Dengan menggunakan Persamaan 7 nilai akurasi yang didapatkan yaitu 75%. Sehingga dapat disimpulkan bahwa model memiliki peluang 75% untuk memprediksi benar dan 25% untuk memprediksi salah.

4. SIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan dapat disimpulkan bahwa arsitektur Convolutional LSTM yang dibangun cukup baik digunakan untuk identifikasi tingkat kesadaran pengemudi dengan menggunakan data time series yang berupa video. Dari uji coba yang telah dilakukan, peneliti berhasil

mendapatkan nilai akurasi model sebesar 75%. Model yang dibangun dapat memprediksi data evaluasi dengan baik, namun terdapat kemungkinan 25% bagi model untuk salah memprediksi, sehingga model yang dibangun belum dapat dikatakan optimal.

5. SARAN

Model yang dibangun dalam penelitian ini belum dapat dikatakan optimal sehingga besar kemungkinan dapat dioptimalkan lagi. Pengoptimalan model dapat dilakukan dengan menambah jumlah layer pada arsitektur model, menambah jumlah epoch, mengganti nilai hyperparameter, maupun mengaplikasikan transfer learning yang telah ada.

DAFTAR PUSTAKA

- [1] S. S. Jasim and A. K. A. Hassan, "Modern drowsiness detection techniques: a review," *Int. J. Electr. Comput. Eng.*, vol. 12, no. 3, pp. 2986–2995, 2022, doi: 10.11591/ijece.v12i3.pp2986-2995.
- [2] K. Gopalakrishna and S. A. Hariprasad, "Real-time fatigue analysis of driver through iris recognition," *Int. J. Electr. Comput. Eng.*, vol. 7, no. 6, pp. 3306–3312, 2017, doi: 10.11591/ijece.v7i6.pp3306-3312.
- [3] F. Nur Fajri, A. Tholib, and W. Yuliana, "Application of Machine Learning Algorithm for Determining Elective Courses in Informatics Study Program," *J. Tek. Inform. dan Sist. Inf.*, vol. 8, no. 3, pp. 485–496, 2022, doi: 10.28932/jutisi.v8i3.3990.
- [4] V. P. Sharma, J. S. Yadav, and V. Sharma, "Deep convolutional network based real time fatigue detection and drowsiness alertness system," *Int. J. Electr. Comput. Eng.*, vol. 12, no. 5, pp. 5493–5500, 2022, doi: 10.11591/ijece.v12i5.pp5493-5500.
- [5] S. S. Jasim and A. K. A. Hassan, "Driving sleepiness detection using electrooculogram analysis and grey wolf optimizer," *Int. J. Electr. Comput. Eng.*, vol. 12, no. 6, pp. 6034–6044, 2022, doi: 10.11591/ijece.v12i6.pp6034-6044.
- [6] M. E. Elidrissi, E. Essoukaki, L. Ben Taleb, A. Mouhsen, and M. Harmouchi, "Drivers' drowsiness detection based on an optimized random forest classification and single-channel electroencephalogram," *Int. J. Electr. Comput. Eng.*, vol. 13, no. 3, pp. 3398–3406, 2023, doi: 10.11591/ijece.v13i3.pp3398-3406.
- [7] C. G. Pachón-Suescún, J. O. Pinzón-Arenas, and R. Jiménez-Moreno, "Abnormal gait detection by means of LSTM," *Int. J. Electr. Comput. Eng.*, vol. 10, no. 2, pp. 1495–1506, 2020, doi: 10.11591/ijece.v10i2.pp1495-1506.
- [8] B. Hardiansyah and A. Primasetya, "Sistem Deteksi Penggunaan masker (Face Mask Detection) Menggunakan Algoritma Deep Learning YOLOv4," *Stain. (SEMINAR Nas. Teknol. & SAINS)*, vol. 2, no. 1, pp. 313–318, 2023.
- [9] Yudi Widhiyana, Transmissia Semiawan, Ilham Gibran Achmad Mudzakir, and Muhammad Randi Noor, "Penerapan Convolutional Long Short-Term Memory untuk Klasifikasi Teks Berita Bahasa Indonesia," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 10, no. 4, pp. 354–361, 2021, doi: 10.22146/jnteti.v10i4.2438.
- [10] M. D. Darajat, Y. A. Sari, and R. C. Wihandika, "Convolutional Neural Network untuk Klasifikasi Citra Makanan Khas Indonesia," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 11, pp. 4764–4769, 2021, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [11] D. Yolanda, K. Gunadi, and E. Setyati, "Pengenalan Alfabet Bahasa Isyarat Tangan Secara Real-Time dengan Menggunakan Metode Convolutional Neural Network dan Recurrent Neural Network," *J. Infra*, vol. 8, no. 1, pp. 203–208, 2020, [Online]. Available: <https://publication.petra.ac.id/index.php/teknik-informatika/article/view/9791>
- [12] Z. Chao, F. Pu, Y. Yin, B. Han, and X. Chen, "Research on real-time local rainfall prediction based on MEMS sensors," *J. Sensors*, vol. 2018, pp. 1–9, 2018, doi: 10.1155/2018/6184713.
- [13] W. Luo, W. Liu, and S. Gao, "Remembering history with convolutional LSTM for anomaly detection," *Proc. - IEEE Int. Conf. Multimed. Expo*, no. July, pp. 439–444, 2017, doi: 10.1109/ICME.2017.8019325.

Penerapan Jaringan Syaraf Tiruan Backpropagation untuk Klasifikasi Akreditasi Sekolah Menengah Pertama

Rachel Elisa Utama¹, Eko Hari Parmadi²

^{1,2}Informatika, Fakultas Sains dan Teknologi, Universitas Sanata Dharma

E-mail: *¹rachelutama30@gmail.com, ²harimbi.parmadi@gmail.com

Abstrak – Akreditasi merupakan salah satu bagian yang penting dalam peningkatan mutu Pendidikan. Banyaknya instrumen akreditasi, membuat pengelola sekolah kesulitan dalam mengklasifikasi sekolah mereka termasuk terakreditasi A, B, C atau tidak terakreditasi. Melalui penerapan Jaringan Syaraf Tiruan Backpropagation proses klasifikasi akreditasi dapat ditentukan berdasarkan 12 atribut yaitu: Persentase Guru Ijazah Kurang Dari S1, Persentase Guru Ijazah S1 Atau Lebih, Persentase Guru Sertifikasi, Persentase Guru Belum Sertifikasi, Standar Sarana dan Prasarana, Standar Isi, Standar Penilaian, Standar Pengelolaan, Standar Kelulusan, Standar Tenaga Pendidik, Standar Pembiayaan, serta Standar Proses. Pengguna dapat memasukkan skor untuk 12 atribut tersebut dan sistem akan memberikan hasil berupa prediksi akreditasinya. Arsitektur optimal dari Jaringan Syaraf Tiruan ini menghasilkan akurasi sebesar 93,4925%. Adapun neuron yang digunakan dalam hidden layer pertama berjumlah 10 dan hidden layer dua berjumlah 20 dengan fungsi aktivasi tansig pada hidden layer pertama dan logsig pada hidden layer kedua serta menggunakan fungsi training trainlm. Sedangkan jumlah hidden layer yang digunakan adalah dua. Lanjutan dari penelitian ini dapat berupa penambahan atribut lain yang berpengaruh dalam akreditasi.

Kata Kunci — akreditasi, Jaringan Syaraf Tiruan Backpropagation, sekolah

1. PENDAHULUAN

Akreditasi menjadi salah satu bagian penting dalam meningkatkan mutu pendidikan di sekolah. Akreditasi tidak hanya sekedar pemberian sertifikat semata, tetapi juga merupakan proses evaluasi yang komprehensif terhadap lembaga pendidikan. Selain berperan dalam meningkatkan kualitas dan transparansi, akreditasi juga dapat meningkatkan kepercayaan orangtua dan masyarakat terhadap Lembaga Pendidikan [1].

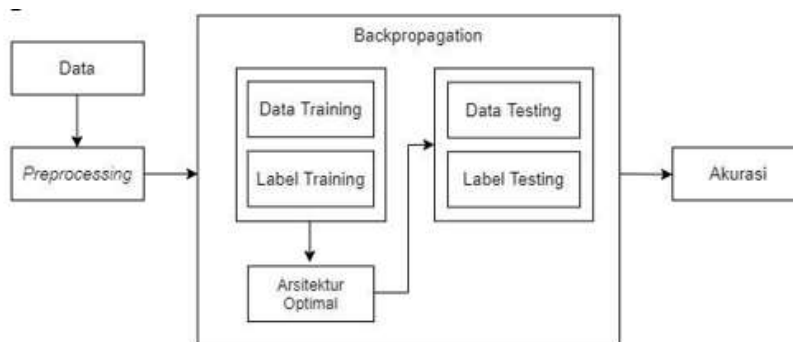
Sesuai dengan Permendikbud Nomor 59 Tahun 2012 pasal 9 ayat (2) huruf 6, dalam pelaksanaan akreditasi sekolah/madrasah, BAN-S/M merumuskan kriteria dan perangkat akreditasi sekolah/madrasah untuk diusulkan kepada Mendikbud. Selanjutnya pasal 16 ayat (1) menyebutkan bahwa Menteri menetapkan kriteria dan perangkat akreditasi dengan memperhatikan Standar Nasional Pendidikan. Penetapan kriteria dan perangkat akreditasi sebagaimana dimaksud pasal 16 ayat (1) didelegasikan kepada Kepala Badan Penelitian dan Pengembangan, setelah berkoordinasi dengan Direktorat Jenderal terkait. Perangkat akreditasi digunakan untuk mengukur sejauh mana sekolah/ madrasah telah memenuhi standar nasional pendidikan [2].

Saat ini sudah banyak sekolah yang melaksanakan akreditasi sekolah guna untuk meningkatkan mutu pendidikan dari sekolah tersebut. Namun, saat ini juga masih terdapat banyak sekolah yang belum memperoleh akreditasi yang baik karena beberapa faktor. Akreditasi sendiri memiliki 8 komponen standar yang harus dipenuhi untuk mendapatkan akreditasi yang baik, komponen standarnya yaitu Standar Isi, Standar Proses, Standar Kompetensi Lulusan, Standar Pendidik dan Tenaga Kependidikan, Standar Sarana dan Prasarana, Standar Pengelolaan, Standar Pembiayaan, dan Standar Penilaian Pendidikan. Akreditasi dikategorikan menjadi 4 kategori yaitu A (Unggul) dengan perolehan nilai 91 sampai dengan 100, B (Baik) dengan perolehan nilai 81 sampai dengan 90, C (Cukup Baik) dengan perolehan nilai 71 sampai dengan 80, dan TT (Tidak Terakreditasi) dengan perolehan nilai akhir sekurang-kurangnya 71, memperoleh nilai komponen sarana dan prasarana sekurang-kurangnya 61 dan tidak ada nilai komponen standar di bawah 50.

Merluarini, B., dkk (2014) telah melakukan penelitian tentang Perbandingan Analisis Klasifikasi Menggunakan Metode *K-Nearest Neighbor* (K-NN) dan *Multivariate Adaptive Regression Spline* (MARS) pada Data Akreditasi Sekolah Dasar Negeri di Kota Semarang. Hasilnya diperoleh nilai *APER* (*Apparent Error Rate*) 5,357 %, *specificity* 92,39%, dan *sensitivity* 97,37% lalu pada MARS didapat nilai *APER* (*Apparent Error Rate*) 2,381 %, *specificity* 94,74%, dan *sensitivity* 98,68% [3]. Penelitian ini bertujuan untuk mengetahui akurasi dari penerapan jaringan syaraf tiruan untuk klasifikasi sekolah menengah pertama. Adapun data yang digunakan pada penelitian ini adalah Data Pokok Pendidikan (DAPODIK) tahun ajaran 2017/2018.

2. METODE PENELITIAN

Secara umum gambaran penelitian meliputi tahap pengumpulan data, preprocessing, pemodelan backpropagation, serta perhitungan akurasi. Gambar 1 berikut ini merupakan gambaran umum proses penelitian.



Gambar 1. Gambaran Umum Penelitian

2.1 Data

Penelitian ini menggunakan Data Pokok Pendidikan (DAPODIK) tahun ajaran 2017/2018. Seluruh data berjumlah 22986 dengan 84 atribut. DAPODIK berisi data dari seluruh sekolah mulai dari tingkat Sekolah Dasar sampai Sekolah Menengah Atas/Sekolah Menengah Kejuruan serta mencakup data sekolah negeri maupun swasta dari provinsi yang ada di Indonesia. Namun pada penelitian ini hanya menggunakan DAPODIK yang meliputi Sekolah Menengah Pertama yang ada di pulau Sulawesi. Data yang digunakan sebanyak 1872 record.

2.2 Preprocessing

Tahap preprocessing pada penelitian ini, meliputi *data cleaning*, *data selection* dan *data transformation*. *Data cleaning* merupakan proses menghilangkan *noised* data yang tidak konsisten atau data tidak relevan. Pada penelitian ini dilakukan proses *data cleaning* karena ditemukan beberapa atribut yang memiliki nilai *missing value*. Atribut yang memiliki nilai *missing value* sebagai berikut :

1. Penyelenggaraan
2. Jumlah Guru Perempuan
3. Jumlah Guru Laki-laki

Nilai *missing value* diatasi dengan mengganti nilai *missing value* tersebut dengan nilai rata-rata setiap atribut. Nilai rata-rata atribut didapatkan dengan rumus berikut :

$$\bar{x} = \frac{x_1+x_2+\dots+x_n}{n} \dots\dots\dots(1)$$

Keterangan :

- \bar{x} = Rata – rata (*Mean*)
- x_1 = Nilai data ke-1
- x_2 = Nilai data ke-2
- x_n = Nilai data ke-n
- n = Banyak data

Proses *data selection* pada penelitian ini dilakukan dengan carai memilih data yang relevan dengan penelitian yang dilakukan, yaitu beberapa fitur / atribut yang dihilangkan diantaranya :

1. NPSN (Nomor Pokok Sekolah Nasional)
2. Nama Sekolah
3. Semester Data (Tahun ajaran saat pendataan akreditasi)
4. Kepala Sekolah (Nama Kepala Sekolah)
5. Operator
6. Email
7. Alamat
8. Kab/Kota
9. Provinsi
10. Kecamatan

Sesuai dengan panduan BAN-SM, akreditasi sekolah tidak dipengaruhi oleh atribut-atribut di atas. Kemudian, ada pula atribut yang tidak digunakan karena tidak ada nilai pembandingnya yaitu :

1. Manajemen Berbasis Sekolah
2. Jenjang
3. Rata-rata IPS

Transformasi data merupakan perubahan atau penggabungan ke dalam format yang sesuai untuk diproses dalam *data mining*. Pada penelitian ini menggunakan normalisasi *min-max* untuk mentransformasi data. Normalisasi data dilakukan dengan tujuan agar data pada setiap atribut memiliki bobot yang sama dengan batas atas dan batas bawah yang kita tentukan. Beberapa atribut dinormalisasi dengan transformasi *min-max*, seperti atribut-atribut: daya Listrik, luas tanah, total siswa Perempuan, total siswa laki-laki, total siswa, rata-rata UN dan sebagainya Berikut merupakan rumus normalisasi *min-max* dengan melakukan transformasi linear :

$$x_B = \frac{X - \text{Min}(X)}{\text{Max}(X) - \text{Min}(X)} (\text{Max}_{\text{newA}} - \text{Min}_{\text{newA}}) + \text{Min}_{\text{newA}} \dots\dots\dots(2)$$

dengan: $\text{Max}_{\text{newA}} = 1$ dan $\text{Min}_{\text{newA}} = 0$

Perangkingan Data, untuk menentukan atribut yang akan digunakan pada penelitian ini. Berdasarkan perangkingan atribut menggunakan information gain pada aplikasi Weka Tool 3.8, didapatkan hasil perangkingan seperti pada Tabel 1 berikut ini:

Tabel 1. Hasil Perangkingan Data

Rangking	Atribut
1	Nilai_akreditasi
2	Standar Kelulusan
3	Standar Proses
4	Standar Isi
5	Standar Sarana Prasarana
6	Standar Tenaga Pendidik
7	Standar Pengelolaan
8	Standar Penilaian
9	Standar Pembiayaan
10	Total Siswa
11	Total Siswa Perempuan
12	Total Rombongan Belajar
13	Total Siswa Laki-laki
14	Guru – Sertifikasi
15	Total Ruang Kelas
16	Jumlah Murid Islam
17	Guru Gol IV
18	Total Guru
19	Jumlah Guru Perempuan
20	Guru – Ijazah S1 atau lebih
21	Persentase Guru Sertifikasi
22	Umur Guru – 51-55 Tahun
23	Umur Guru – 46-50 Tahun
24	Umur Guru – Lebih dari 55 Tahun
25	Ratio Umur Guru – Kurang dari 30 Tahun
26	Akreditasi status
27	Ratio Umur Guru – 41-45 Tahun
28	Ratio Umur Guru – 31-35 Tahun
29	Daya Listrik
30	Ratio Umur Guru – Lebih dari 55 Tahun
31	Kurikulum
32	Persentase Guru PNS
33	Rasio Siswa Rombel
34	Guru Gol III
35	Jumlah Guru Laki-laki
36	Ratio Umur Guru – 36-40 Tahun
37	Ratio Umur Guru – 51-55 Tahun
38	Ratio Umur Guru – 46-50 Tahun
39	Umur Guru – 41-45 Tahun
40	Ratio Siswa – Guru
41	Persentase Guru Kualifikasi
42	Luas Tanah
43	Guru – Belum Sertifikasi
44	Guru – Ijazah Kurang dari S1
45	Jumlah Murid Kristen
46	Total Perpustakaan
47	Total Sanitasi Siswa
48	Jumlah Murid Katolik
49	Penyelenggaraan
50	Jumlah Murid Hindu

Rangking	Atribut
51	Lng
52	Lat
53	Umur Guru – 31-35 Tahun
54	Sumber Listrik
55	Akses Internet
56	Umur Guru – Kurang dari 30 Tahun
57	Rata-rata UN
58	Jenis Sekolah
59	Jumlah Murid Lainnya
60	Rata-rata IPA
61	Guru – Data Kosong
62	Jumlah Murid Konghucu
63	Jumlah Murid Budha
64	Persentase Siswa Mengulang
65	Persentase ruang kelas layak
66	Guru Gol I
67	Guru Gol II

Beberapa atribut bertipe kategorikal juga ditransformasi dalam bentuk numerik seperti:
 Akreditasi Status, ditransformasi menjadi :

A → 1	Penyelenggaraan	Kurikulum
B → 2	Pagi-6h → 1	K-13 → 1
C → 3	Siang-6h → 2	KTSP → 2
Belum → 4	Sore-6h → 3	Akreditasi
Tidak → 5	Malam-6h → 4	A → 11
Jenis Sekolah	Double Shift-6h → 5	B → 10
Negeri → 1	Sehari Penuh-5h → 6	C → 01
Swasta → 2	Sehari Penuh-6h → 7	TT → 00
	Lainnya → 8	

Setelah melakukan perangkingan atribut menggunakan information gain aplikasi Weka 3.8, dapat dilihat bahwa atribut Standar Sarana dan Prasarana, Standar Isi, Standar Penilaian, Standar Pengelolaan, Standar Kelulusan, Standar Tenaga Pendidik, Standar Pembiayaan, serta Standar Proses berada pada posisi rangking teratas. Hasil perangkingan atribut yang dilakukan, juga sesuai dengan 8 komponen standar akreditasi Sekolah Menengah Pertama. Penelitian ini menggunakan atribut dari nilai 8 komponen standar akreditasi sekolah yang berasal dari Sekolah Menengah Pertama yang berada di Pulau Sulawesi. Berikut contoh data setelah diseleksi sesuai perangkingan atribut yang digunakan untuk penelitian pada Tabel 2.

Tabel 2. Data Setelah Diseleksi

Standar Sarana Prasarana	Standar Isi	Standar Penilaian	Standar Pengelolaan	Standar Kelulusan	Standar Tenaga Pendidik	Standar Pembiayaan	Standar Proses
0,73	0,89	0,75	0,89	0,79	0,63	0,97	0,89
0,57	0,8	0,81	0,64	0,56	0,55	0,79	0,67
0,43	0,35	0,32	0,25	0,31	0,35	0,42	0,33
0,75	0,73	0,72	0,72	0,73	0,7	0,74	0,75
0,57	0,47	0,52	0,55	0,53	0,6	0,53	0,42

Pada penelitian ini juga dilakukan penambahan atribut pendukung yaitu atribut Total Guru, atribut Guru – Ijazah Kurang dari S1, Atribut Guru – Ijazah S1 atau lebih, Atribut Guru – Belum Sertifikasi, dan Atribut Guru – Sertifikasi. Penambahan atribut ini bertujuan untuk melihat pengaruh atribut lain terhadap akurasi sistem yang akan digunakan. Berikut merupakan contoh data yang telah ditambahkan dengan 5 atribut pada Tabel 3 berikut ini:

Tabel 3. Penambahan Atribut Pada Data yang telah Diseleksi

Total guru	Guru - Ijazah kurang dari S1	Guru - Ijazah S1 atau lebih	Guru - Sertifikasi	Guru - Belum Sertifikasi	Standar Sarana Prasarana	Standar Isi	Standar Penilaian	Standar Pengelolaan	Standar Kelulusan	Standar Tenaga Pendidik	Standar Pembiayaan	Standar Proses
4	0	4	0	4	0,73	0,89	0,75	0,89	0,79	0,63	0,97	0,89
7	0	7	1	6	0,57	0,8	0,81	0,64	0,56	0,55	0,79	0,67
1	0	0	0	1	0,43	0,35	0,32	0,25	0,31	0,35	0,42	0,33
6	0	6	2	4	0,75	0,73	0,72	0,72	0,73	0,7	0,74	0,75
6	0	5	0	6	0,57	0,47	0,52	0,55	0,53	0,6	0,53	0,42

Tabel 4 berikut ini merupakan data setelah dilakukan proses transformasi : Setelah proses perangkingan data dan penambahan atribut, dilakukan proses transformasi data kembali. Transformasi data ini dilakukan dengan tujuan untuk mengubah data dalam 4 atribut yang digunakan untuk penelitian ini yaitu atribut Guru – Ijazah kurang dari S1, Guru – Ijazah S1 atau lebih, Guru Sertifikasi, dan Guru Belum Sertifikasi. Atribut-atribut tersebut akan ditransformasi ke dalam nilai persentase, yaitu persentase Guru – Ijazah kurang dari S1, persentase Guru – Ijazah S1 atau lebih, persentase Guru Sertifikasi dan persentase Guru Belum Sertifikasi. Tabel 4 berikut ini merupakan data setelah dilakukan proses transformasi:

Tabel 4. Data Setelah Proses Transformasi

Persentase Guru - Ijazah kurang dari S1	Persentase Guru - Ijazah S1 atau lebih	Persentase Guru - Sertifikasi	Persentase Guru - Belum Sertifikasi	Standar Sarana Prasarana	Standar Isi	Standar Penilaian	Standar Pengelolaan	Standar Kelulusan	Standar Tenaga Pendidik	Standar Pembiayaan	Standar Proses
0	100	0	100	0,73	0,89	0,75	0,89	0,79	0,63	0,97	0,89
0	100	14,3	85,7	0,57	0,8	0,81	0,64	0,56	0,55	0,79	0,67
0	0	0	100	0,43	0,35	0,32	0,25	0,31	0,35	0,42	0,33
0	100	33,3	66,7	0,75	0,73	0,72	0,72	0,73	0,7	0,74	0,75
0	83,3	0	100	0,57	0,47	0,52	0,55	0,53	0,6	0,53	0,42

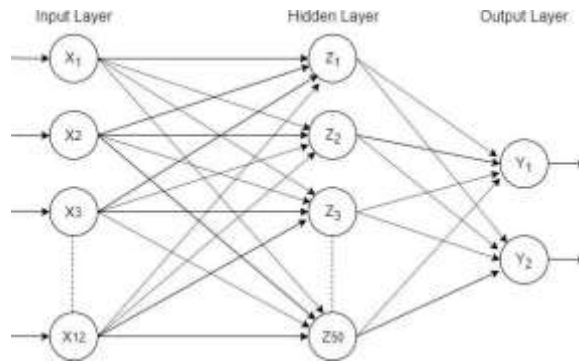
1.3 Backpropagation Neural Network (BPNN)

BPNN merupakan algoritma pembelajaran terawasi yang mempunyai banyak lapisan. BPNN menggunakan error output untuk mengubah nilai bobot-bobotnya dalam arah mundur (*backward*). Untuk mendapatkan error ini, tahap perambatan maju (*forward propagation*) harus dikerjakan terlebih dahulu.

Syarat fungsi aktivasi dalam BPNN adalah bersifat kontinu, terdiferensial dengan mudah, dan merupakan fungsi yang tidak turun. Fungsi aktivasi yang dapat memenuhi ketiga syarat tersebut adalah logsig, tansig, dan purelin. Metode pengenalan merupakan proses inialisasi data yang akan diolah selanjutnya oleh BPNN. Data yang akan dikenali disajikan dalam bentuk vektor. Masing-masing data mempunyai target yang disajikan juga dalam bentuk vektor. Target atau keluaran acuan merupakan suatu peta karakter yang menunjukkan lokasi dari vektor masukan.

Sedangkan metode pelatihan merupakan proses latihan mengenali data dan menyimpan pengetahuan atau informasi yang didapat ke dalam bobot-bobot [1].

Dalam arsitektur *backpropagation* terdapat n buah masukan dan layer tersembunyi yang terdiri satu atau lebih unit. Pembagian data dalam penelitian ini menggunakan *3-fold cross validation*, berikut ini arsitektur Jaringan Syaraf Tiruan *Backpropagation* yang digunakan dalam penelitian ini:



Gambar 2. Arsitektur *Backpropagation*

X1,X2,X3,..., X12 merupakan lapisan masukan dalam jaringan syaraf tiruan. Masukan ini berupa atribut yang digunakan dalam penelitian, pada penelitian kali ini terdapat 12 atribut. Z1,Z2,Z3..., 50 merupakan lapisan tersembunyi (*hidden layer*) yang akan digunakan untuk mengolah nilai masukan. Dalam *hidden layer* 2 ini akan dilakukan variasi pada jumlah neuron untuk meningkatkan tingkat akurasi pada sistem. Y1 dan Y2 merupakan lapisan luaran yang memiliki 2 neuron. Nilai luaran akan bernilai 11, 10, 01 atau 00. Berikut contoh representasi luaran dari sistem:

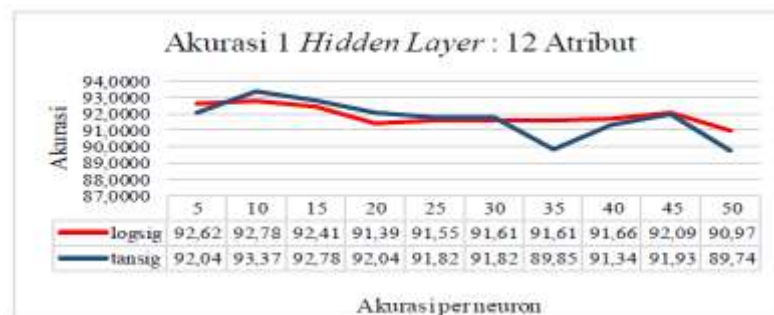
Tabel 5. Nilai Target

Kelas Akreditasi	Nilai Target	
	Y1	Y2
A	1	1
B	1	0
C	0	1
TT	0	0

Jaringan backpropagation akan menghasilkan luaran yang akan dicocokkan kembali dengan label asli untuk mengetahui akurasi. Akurasi dihitung dengan menggunakan *confusion matrix*. Kemudian pengujian yang dilakukan dengan berbagai variasi yaitu jenis data normalisasi/tidak, jumlah hidden layer, jumlah neuron (5 sampai 50 kelipatan 5), fungsi *training*, fungsi aktivasi dan jumlah epoch (50, 100, 150, 200, 250, 500, 1000).

3. HASIL DAN PEMBAHASAN

Proses klasifikasi yang dilakukan ini menggunakan data yang sudah di-preprocessing pada tahap sebelumnya. Data yang digunakan sudah dalam bentuk numerik dan berjumlah 1872 record data yang di dalamnya sudah memiliki label kategori akreditasi 1 1 (A), 1 0 (B), 0 1 (C), dan 0 0 (4). Proses klasifikasi ini akan menggunakan Jaringan Syaraf Tiruan dengan membandingkan dua fungsi aktivasi yaitu logsig dan tansig, kombinasi neuron (5, 10, 15, 20, 25, 30, 35, 40, 45, 50), fungsi training (trainlm, traingdx, traincgb, trainrp, trainscg, traincgp, dan trainoss), serta kombinasi epoch pada satu hidden layer dan dua hidden layer untuk mendapatkan arsitektur yang paling optimal.



Gambar 3. Grafik Akurasi 1 Layer Tersembunyi

Berdasarkan visualisasi pada Gambar 3 berikut ini dapat dilihat bahwa variasi neuron dengan satu *hidden layer* menghasilkan akurasi tertinggi pada 10 neuron, menggunakan fungsi aktivasi tansig dan fungsi training trainlm dengan akurasi sebesar 93.3761%.



Gambar 4. Grafik Variasi Fungsi Training

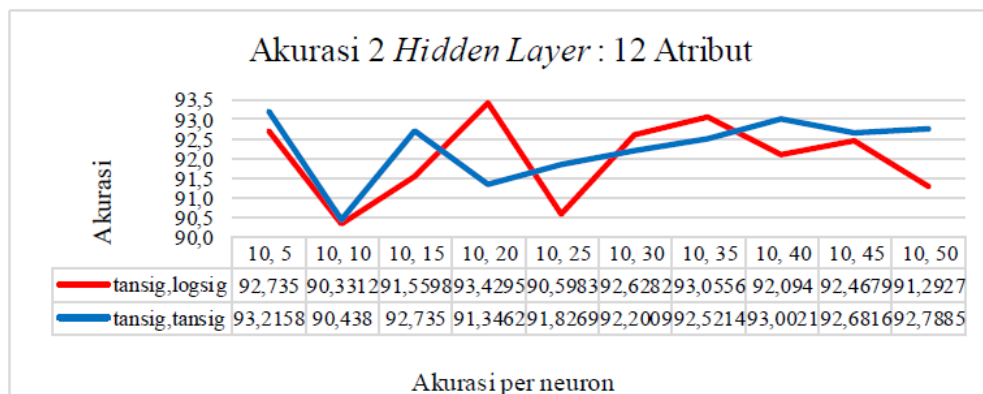
Dapat dilihat pada Gambar 4 bahwa hasil percobaan perbandingan akurasi terbaik dari fungsi *training* pada 10 neuron menggunakan fungsi aktivasi logsig berada pada fungsi training *trainlm* dengan akurasi terbesar yaitu 93,3761%.



Gambar 5. Grafik Variasi Epoch

Berdasarkan Gambar 5 dapat dilihat bahwa jumlah epoch yang digunakan sebagai perbandingan menghasilkan akurasi yang nilainya sama. Penambahan jumlah epoch tidak berpengaruh terhadap nilai akurasi dan sudah konvergen pada jumlah epoch yang terkecil.

Kombinasi neuron, fungsi aktivasi dan fungsi train terbaik menghasilkan hasil tertinggi pada neuron 10 di layer pertama dan neuron 20 di layer kedua dengan fungsi aktivasi tansig dan logsig serta fungsi training *trainlm* dengan hasil akurasi 93.4295 % seperti terlihat pada Gambar 6.



Gambar 6. Grafik Akurasi 2 Hidden Layer

Berdasarkan hasil percobaan klasifikasi menggunakan kombinasi hidden layer, variasi neuron, variasi fungsi aktivasi, dan fungsi training serta variasi epoch didapatkan arsitektur optimal dengan akurasi sebesar 93,4295% menggunakan 2 hidden layer dengan kombinasi 10 neuron pada hidden layer pertama dan 20 neuron pada hidden layer kedua. Pada arsitektur optimal ini fungsi aktivasi yang digunakan adalah tansig pada hidden layer pertama dan logsig pada hidden layer kedua serta menggunakan fungsi training *trainlm*. Berikut ini contoh masukan untuk 12 atribut pada penelitian ini, sedangkan hasil klasifikasi dari masukan data tersebut adalah terakreditasi B.

Uji Data Tunggal			
Total Guru	14	Standar Penilaian	77
Guru - Ijazah Kurang Dari S1	1	Standar Pengelolaan	75
Guru - Ijazah S1 Atau Lebih	13	Standar Kelulusan	60
Guru Sertifikasi	8	Standar Tenaga Pendidik	73
Guru Belum Sertifikasi	6	Standar Pembiayaan	97
Standar Sarana dan Prasarana	75	Standar Proses	88
Standar Isi	59	Persentase Guru Sertifikasi	57.1429
Persentase Guru - Ijazah Kurang Dari S1	7.1429	Persentase Guru Belum Sertifikasi	42.8571
Persentase Guru - Ijazah S1 Atau Lebih	92.8571		

Hasil Klasifikasi:

Gambar 7. Contoh Masukan Data Berdasarkan 12 Atribut Dan Hasil Klasifikasinya

4. SIMPULAN

Pengujian yang dilakukan dengan menggunakan 1872 record data menggunakan 3-fold cross validation menghasilkan keakuratan sebesar 93,4295% dengan struktur paling optimal pada dua hidden layer menggunakan variasi dari 10 dan 20 neuron, variasi fungsi aktivasi tansig dan logsig, serta fungsi training trainlm. Atribut yang digunakan pada pengujian yang dilakukan terdapat 12 atribut yaitu Persentase Guru Ijazah Kurang Dari S1, Persentase Guru Ijazah S1 Atau Lebih, Persentase Guru Sertifikasi, Persentase Guru Belum Sertifikasi, Standar Sarana dan Prasarana, Standar Isi, Standar Penilaian, Standar Pengelolaan, Standar Kelulusan, Standar Tenaga Pendidik, Standar Pembiayaan, serta Standar Proses. Label kategori akreditasi yang digunakan adalah Terakreditasi A, Terakreditasi B, Terakreditasi C, dan Tidak Terakreditasi.

5. SARAN

Penelitian ini dapat dikembangkan dengan menambahkan variasi balancing pada data yang tidak seimbang serta menghitung presisi, recall dan f1 score dari hasil klasifikasi.

DAFTAR PUSTAKA

- [1] Firda, M.R. 2023. Pentingnya Akreditasi Dalam Peningkatan Mutu Sekolah Dan Madrasah. <https://news.bsi.ac.id/2023/08/28/pentingnya-akreditasi-dalam-peningkatan-mutu-sekolah-dan-madrasah/#:~:text=Dengan%20adanya%20akreditasi%2C%20orang%20tua,dan%20memenuhi%20standar%20yang%20ditetapkan> diakses pada tanggal 1 Desember 2023
- [2] BANSNM. 2019 Badan Akreditasi Nasional Sekolah/Madrasah. Perangkat Akreditasi. <https://bansnm.kemdikbud.go.id/> diakses 18 Oktober 2019
- [3] Merluarini, B., dkk.2014. Perbandingan Analisis Klasifikasi Menggunakan Metode K-Nearest Neighbor (KNN) dan *Multivariate Adaptive Regression Spline* (MARS) pada Data Akreditasi Sekolah Dasar Negeri di Kota Semarang. *Jurnal Gaussian* Vol. 3, No. 3
- [4] Heaton, J. 2003. "Introduction to Neural Network with Java", <http://www.heatonresearch.com/articles/6/page2.html> diakses pada tanggal 2 November 2019.

Korelasi antara Kekuatan Militer dan Faktor-Faktor Multidimensi: Analisis Tahun 2021

Fathul Am¹, Riky Ananda Setyanto², Imanuel Puspa Wardaya³, Ari Hendrawan⁴

^{1,2,3,4}Informatika, Fakultas Sains dan Teknologi, Universitas Teknologi Yogyakarta

E-mail: *¹fathul.5200411277@student.uty.ac.id, ²riky.5200411278@student.uty.ac.id,

³imanuel.5200411349@student.uty.ac.id, ⁴ari.5200411429@student.uty.ac.id

Abstrak – Kekuatan militer adalah konsep yang kompleks dan multidimensi yang bergantung pada berbagai faktor, seperti ukuran, kualitas, kesiapan angkatan bersenjata, ketersediaan sumber daya, teknologi, geopolitik, dan dukungan domestik dan internasional. Memahami korelasi antara kekuatan militer dan faktor-faktor lain dapat membantu untuk menilai kekuatan relatif dan potensi negara-negara yang berbeda, serta untuk mengidentifikasi peluang dan tantangan yang mereka hadapi di arena global. Penelitian ini menganalisis setiap faktor dengan tujuan mengidentifikasi faktor yang paling signifikan dalam menentukan kekuatan militer suatu negara. Penelitian ini mengusulkan penggunaan metode korelasi Pearson untuk menganalisis faktor yang paling berpengaruh terhadap kekuatan militer. Hasil penelitian menunjukkan korelasi antar sektor yang paling mendekati nilai -1 adalah sektor *Land* yang bernilai -0.3076. Hasil rata-rata korelasi antara *Military Power index* dengan sektor *Navy Strength* bernilai -0.2714, hasil rata-rata korelasi antara *Military Power index* sektor *Airforce* bernilai -0.2640, hasil rata-rata korelasi antara *Military Power* sektor *Finance* bernilai -0.2140, hasil rata-rata korelasi antara *Military Power* sektor *Geography* bernilai -0.1780 dan hasil rata-rata korelasi antara *Military Power* sektor *Logistic* bernilai -0.1733. Sebagai kesimpulan, penelitian ini memberikan kontribusi dalam mengatasi kurangnya pengetahuan tentang faktor-faktor yang memengaruhi kekuatan militer. Secara khusus, penelitian menunjukkan bahwa sektor *Land* memiliki korelasi terbaik dengan *Military Power*, dengan nilai korelasi sebesar -0.3076.

Kata Kunci — data science, korelasi, militer, pearson

1. PENDAHULUAN

Militer adalah kekuatan yang digunakan untuk melumpuhkan musuh, mencapai tujuan, serta memerangi pihak lawan di medan perang [1]. Kekuatan militer adalah konsep yang kompleks dan multidimensi yang bergantung pada berbagai faktor, seperti ukuran, kualitas, dan kesiapan angkatan bersenjata, ketersediaan sumber daya dan teknologi, konteks geopolitik dan strategis, serta dukungan domestik dan internasional. Memahami korelasi antara kekuatan militer dan faktor-faktor lain dapat membantu untuk menilai kekuatan relatif dan potensi negara-negara yang berbeda, serta untuk mengidentifikasi peluang dan tantangan yang mereka hadapi di arena global. Oleh karena itu, penting untuk menjelaskan mengapa penelitian ini menggali korelasi antara kekuatan militer dan faktor-faktor lain yang memengaruhi kekuatan militer. Kekuatan militer diukur dari beberapa faktor [2]. Kurangnya pengetahuan tentang faktor yang paling berpengaruh menjadi salah satu hambatan untuk memperkuat kekuatan militer secara efektif. Dengan adanya teknologi, faktor yang paling berpengaruh dapat diketahui dengan menggunakan metode korelasi. Ini memungkinkan para pemimpin militer dan pengambil keputusan untuk lebih tepat dalam alokasi sumber daya dan strategi pertahanan, sehingga meningkatkan efisiensi dan efektivitas kekuatan militer suatu negara. Dengan pemahaman yang lebih mendalam tentang faktor-faktor kunci ini, sebuah negara dapat mengembangkan kemampuan militer yang lebih kuat dan adaptif dalam menghadapi berbagai tantangan keamanan.

Dalam upaya untuk mengatasi kurangnya pengetahuan yang telah dijelaskan, penelitian ini mencoba memberikan kontribusi berupa pemahaman yang lebih mendalam mengenai faktor-faktor yang memengaruhi kekuatan militer. Penelitian ini menggunakan metode korelasi dalam mengukur korelasi antara kekuatan militer dengan faktor lain. Metode korelasi adalah suatu alat statistik yang digunakan untuk menentukan hubungan atau keterkaitan antara dua variabel atau lebih [3]. Dengan menggunakan metode korelasi, analisis militer dapat menentukan sejauh mana faktor-faktor ini berkorelasi satu sama lain seperti aircraft strength, land strength, navy strength, finances, logistics, dan geography. Adapun Penelitian lain dilakukan oleh Susdarwono dan Sani meneliti korelasi antara *human development* terhadap anggaran pertahanan negara menurut GDP. Global Data Lab menyediakan data tentang *Human Development Index* (HDI), menurut data yang dihimpun pada penelitian ini nilai khai-kuadrat bagi taraf signifikan 2,50% dan derajat kebebasan 6 adalah 14,449. Berdasarkan hipotesis nihil

dengan hasil khai-kuadrat didapatkan sebesar 16,94355. Sehingga peneliti menyimpulkan bahwa terdapat korelasi antara *human development* dengan anggaran pertahanan suatu negara. Dengan nilai korelasi antara kedua variabel tersebut sebesar 36,30% [4]. Peneliti lainnya yang dilakukan oleh Maulidah, Budiono, dan Kurniawan meneliti tentang pengaruh budaya semi militer terhadap wawasan kebangsaan siswa di SMK Negeri 2 Turen. Metode korelasi yang digunakan dalam penelitian ini yaitu Pearson product moment correlation dengan bantuan program aplikasi SPSS versi 25.0. Hasil yang diperoleh terdapat pengaruh positif dari budaya semi militer terhadap wawasan kebangsaan siswa SMK Negeri 2 Turen dengan nilai korelasi 0.930 dan kontribusi sebesar 86.5% [5]. Penelitian terdahulu yang dilakukan oleh Yuanto, Sudardi dan Firdaus meneliti besarnya pengaruh gaya kepemimpinan partisipatif komandan satuan militer terhadap profesionalisme prajurit. Teknik analisis data yang dilakukan pada penelitian ini adalah uji normalitas, uji kesamaan dan uji Anova. Kemudian metode yang digunakan untuk melakukan uji hipotesis yaitu uji regresi sederhana dan uji koefisien. Hasil yang didapatkan bahwa variabel Gaya Kepemimpinan Partisipatif berpengaruh terhadap variabel Profesionalisme Prajurit dengan total pengaruh sebesar 64,2 % [6]. Penelitian lain yang dilakukan oleh Saputro, Rivai dan Meirinaldi membahas mengenai pengaruh anggaran pertahanan, import alutsista, ekspor alutsista, dan inflasi terhadap pertumbuhan ekonomi di Indonesia tahun 1980-2019. Penelitian ini menggunakan regresi linear berganda untuk mendapatkan hasil. Berdasarkan hasil yang didapat, anggaran pertahanan berpengaruh secara signifikan terhadap ekonomi dan impor alutsista berpengaruh signifikan terhadap pertumbuhan ekonomi [7]. Penelitian yang dilakukan oleh Nugroho, Johan, dan Mufidin meneliti tentang pemanfaatan data science untuk meningkatkan kualitas pembinaan SDM militer. Dalam penelitian ini menggunakan metode uji normalisasi data untuk membantu menghubungkan berbagai variabel yang diperlukan. Hasil yang diperoleh memperlihatkan bahwa pemanfaatan Data Science memberikan pengaruh pada Kualitas Pembinaan SDM Militer [8]. Penelitian yang dilakukan oleh Sari, W, dan Suhardianti membahas tentang pengaruh notam latihan militer terhadap pelayanan pemanduan lalu lintas penerbangan. Studi kasus di perum LPPNPI cabang pembantu curug dengan metode analisis yang digunakan yaitu uji instrumen, uji asumsi, dan analisis hubungan menggunakan korelasi Rank Spearman. Hasil yang diperoleh hubungan variabel NOTAM Latihan Militer dengan Pelayanan Pemanduan Lalu Lintas Penerbangan memiliki nilai korelasi $-0,69$ dengan pengaruh sebesar 47,61% [9].

Sejumlah penelitian terdahulu belum ada pembahasan yang secara khusus menentukan faktor yang paling berpengaruh terhadap kekuatan militer. Oleh karena itu, penelitian ini mengisi kekosongan tersebut dengan menganalisis setiap faktor dengan tujuan mengidentifikasi faktor yang paling signifikan dalam menentukan kekuatan militer suatu negara. Problem khusus yang dihadapi dalam penelitian ini adalah kompleksitas variabel-variabel yang terlibat dan perlunya mempertimbangkan interaksi antar faktor tersebut untuk memahami dampaknya secara menyeluruh. Penelitian ini melakukan analisis untuk dapat memberikan kontribusi yang berharga terhadap pemahaman kita tentang faktor yang paling berpengaruh untuk meningkatkan kekuatan militer.

Penelitian ini mengusulkan penggunaan metode korelasi Pearson untuk menganalisis faktor yang paling berpengaruh terhadap kekuatan militer. Korelasi Pearson digunakan karena dapat menunjukkan keeratan hubungan antara variabel [10]. Penelitian ini memiliki tujuan yang jelas, yaitu mengeksplorasi dan mengukur korelasi antara kekuatan militer dan faktor-faktor lain. Tujuan tersebut menjadi dasar dari seluruh penelitian ini. Selain itu, penelitian ini juga memiliki manfaat, yaitu memberikan pengetahuan dan rekomendasi berdasarkan analisis, yang dapat digunakan oleh pemangku kepentingan dalam pengambilan keputusan yang berkaitan dengan kekuatan militer.

2. METODE PENELITIAN

2.1 Pengumpulan Data

Dalam penelitian ini, pengumpulan data dilakukan untuk menganalisis korelasi antara kekuatan militer suatu negara dengan faktor-faktor multidimensi pada tahun 2021. Data kekuatan militer diperoleh dari sumber terpercaya, yaitu Word Military Power, yang diunduh dari kaggle.com. Pengumpulan data ini menjadi landasan utama dalam menjalankan analisis yang komprehensif untuk memahami hubungan antara kekuatan militer dan berbagai faktor yang memengaruhinya. Dengan mengekstraksi pola informasi dari data berguna untuk mengolah data menjadi informasi baru dan lebih berguna [11] [12].

2.2 Perancangan Sistem

Setelah berhasil mengumpulkan data, tahap selanjutnya dalam penelitian ini adalah perancangan sistem analisis. Peneliti akan merinci metode dan algoritma yang akan digunakan untuk mengolah data tersebut.

Perancangan sistem ini akan memastikan keakuratan dan kehandalan dalam menyusun analisis terperinci terkait korelasi antara kekuatan militer suatu negara dan faktor-faktor multidimensi yang diidentifikasi. Untuk tahapan perancangan sistem dapat dilihat pada Gambar 1.



Gambar 1. Perancangan Sistem

2.3 Pre-processing

Preprocessing adalah tahapan yang dilakukan untuk membersihkan data sehingga hasil perhitungan korelasi akan lebih optimal [13]. Proses preprocessing data terdiri dari pemisahan tabel menjadi perkolom, penggabungan perbagian, dan missing value. Pemisahan tabel data menjadi perkolom guna memfasilitasi analisis yang lebih spesifik. Selanjutnya, peneliti akan melakukan penggabungan data perbagian untuk menyusun hubungan antara variabel yang saling terkait. Proses preprocessing juga akan mencakup penanganan missing value guna memastikan integritas dan kualitas data yang diperoleh, sehingga analisis yang dilakukan dapat menghasilkan temuan yang akurat dan reliabel.

2.4 Korelasi Pearson

Korelasi Pearson adalah suatu metode korelasi yang sederhana, di mana hanya terlibat satu variabel terikat dan satu variabel bebas [14] [15]. Melalui metode ini, peneliti akan mengukur tingkat korelasi antara kekuatan militer dan faktor-faktor multidimensi secara kuantitatif. Hasil korelasi Pearson ini akan memberikan pemahaman lebih mendalam mengenai sejauh mana variabel-variabel tersebut saling berkaitan, serta arah dan kekuatan hubungannya. Proses ini menjadi kunci untuk mengungkap pola atau asosiasi yang signifikan dalam dataset.

$$r = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}} \dots\dots\dots(1)$$

Dimana:

r = koefisien korelasi pearson

x = variabel bebas

y = variabel terikat

n = jumlah sampel yang diteliti

2.5 Hipotesis

Hipotesis adalah jawaban sementara atas masalah yang dirumuskan dan searag dengan tujuan penelitian [16]. Berdasarkan penelitian ini, terdapat beberapa hipotesis yang dapat diuji untuk menjelajahi korelasi antara kekuatan militer dan faktor-faktor lain yang memengaruhinya.

Rumusan hipotesis nol (H_0) dan hipotesis alternatif (H_a) untuk menyatakan hubungan antara indeks kekuatan militer dengan faktor multidimensi adalah sebagai berikut:

1. Hipotesis Nol (H_0): Tidak terdapat hubungan yang signifikan antara indeks kekuatan militer dengan faktor multidimensi.

$$H_0 : \rho < 0,05 \dots\dots\dots(2)$$

2. Hipotesis Alternatif (H_a): Terdapat hubungan yang signifikan antara indeks kekuatan militer dengan faktor multidimensi.

$$H_a : \rho > 0,05 \dots\dots\dots(3)$$

Di sini, ρ adalah simbol korelasi yang mengukur sejauh mana indeks kekuatan militer berkorelasi dengan faktor multidimensi. Jika nilai ρ -value hasil uji statistik menunjukkan signifikansi yang cukup, hipotesis nol akan ditolak dan menerima hipotesis alternatif, menyatakan bahwa terdapat hubungan yang signifikan antara indeks kekuatan militer dan faktor multidimensi. Pada penelitian ini H_0 dan H_a yang digunakan terdapat pada Tabel 1.

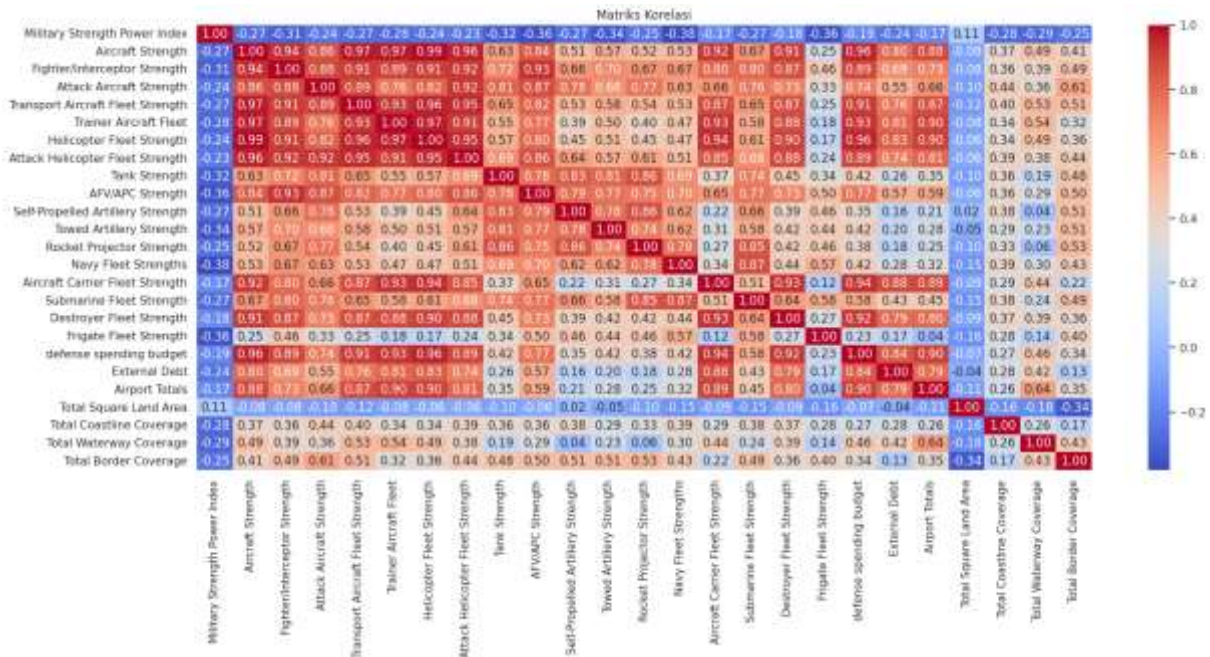
Tabel 1. Hipotesis

Hipotesis	Keterangan
H_{01}	Tidak terdapat hubungan yang signifikan antara <i>Military Strength Power</i> dan <i>Aircraft Strength</i>
H_{a1}	Terdapat hubungan yang signifikan antara <i>Military Strength Power</i> dan <i>Aircraft Strength</i>
H_{02}	Tidak terdapat hubungan yang signifikan antara <i>Military Strength Power</i> dan <i>Land Strength</i>
H_{a2}	Terdapat hubungan yang signifikan antara <i>Military Strength Power</i> dan <i>Land Strength</i>
H_{03}	Tidak terdapat hubungan yang signifikan antara <i>Military Strength Power</i> dan <i>Navy Strength</i>
H_{a3}	Terdapat hubungan yang signifikan antara <i>Military Strength Power</i> dan <i>Navy Strength</i>
H_{04}	Tidak terdapat hubungan yang signifikan antara <i>Military Strength Power</i> dan <i>Finance</i>
H_{a4}	Terdapat hubungan yang signifikan antara <i>Military Strength Power</i> dan <i>Finance</i>
H_{05}	Tidak terdapat hubungan yang signifikan antara <i>Military Strength Power</i> dan <i>Logistic</i>
H_{a5}	Terdapat hubungan yang signifikan antara <i>Military Strength Power</i> dan <i>Logistic</i>
H_{06}	Tidak terdapat hubungan yang signifikan antara <i>Military Strength Power</i> dan <i>Logistic</i>
H_{a6}	Terdapat hubungan yang signifikan antara <i>Military Strength Power</i> dan <i>Logistic</i>

3. HASIL DAN PEMBAHASAN

Dataset "World Military Power 2020" diambil dari Kaggle merupakan kumpulan data yang berisi informasi tentang kekuatan militer dari 140 negara di dunia. Data ini dikumpulkan oleh Global Firepower, sebuah organisasi non-profit yang berfokus pada analisis militer. Dataset ini terdiri dari data 140 baris, 42 kolom dengan tipe data numerik dan kategorik, yang masing-masing mewakili satu variabel. Analisis data World Military Power 2021 menunjukkan bahwa kekuatan militer suatu negara tidak hanya ditentukan oleh faktor-faktor tradisional, seperti PDB, pengeluaran militer, dan jumlah personel militer. Faktor-faktor multidimensi, seperti luas wilayah, kekuatan ekonomi, dan teknologi militer, juga memiliki korelasi yang signifikan dengan kekuatan militer. Kekuatan militer suatu negara merupakan hasil dari kombinasi dari berbagai faktor, baik tradisional maupun multidimensi. Faktor-faktor tersebut saling berinteraksi dan saling mempengaruhi untuk membentuk kekuatan militer suatu negara. Data yang digunakan dalam penelitian ini berupa data numerik dengan beberapa sektor. Sektor yang digunakan yaitu: Airforce, Land, Navy Strength, Finance, Logistic, Geography, bagian yang lebih detail dan korelasi setiap bagian dapat dilihat pada Gambar 2. Pada Gambar 2 dapat kita lihat bahwa

secara individual *Navy Fleet Strength*, *Tank Strength*, *AFV/APC Strength*, dan, *Towed Artillery Strength* menjadi bagian yang memiliki korelasi kuat dengan kekuatan militer.



Gambar 2. Matriks Korelasi

Hasil dari penelitian ini didapatkan p -value dari masing-masing korelasi yang dilakukan. Nilai korelasi setiap bagian bernilai sekitar -0.1 hingga -0.3 dimana seluruh bagian berkorelasi dengan kekuatan militer. Uji Hipotesis dilakukan dengan melihat hasil ρ dimana jika $\rho > 0.05$ maka H_0 merupakan hasil dari uji hipotesis, sedangkan jika $\rho < 0.05$ maka H_a merupakan hasil dari uji hipotesis. Hasil korelasi, nilai ρ , dan rata-rata ρ dapat dilihat pada Tabel 2. Pada penelitian ini p -value diambil dari rata-rata ρ setiap bagian. Hasil hipotesis pada penelitian ini adalah:

1. H_{a1} terpenuhi dimana terdapat korelasi antara *Military Strength Power* dan *Aircraft Strength* ($\rho = 0.0026174610916106316$)
2. H_{a2} terpenuhi dimana terdapat korelasi antara *Military Strength Power* dan *Land Strength* ($\rho = 0.0009972020212671187$)
3. H_{a3} terpenuhi dimana terdapat korelasi antara *Military Strength Power* dan *Navy Strength* ($\rho = 0.01843272969584182$)
4. H_{a4} terpenuhi dimana terdapat korelasi antara *Military Strength Power* dan *Finance* ($\rho = 0.01456669930604695$)
5. H_{a5} terpenuhi dimana terdapat korelasi antara *Military Strength Power* dan *Logistic* ($\rho = 0.04200812613546968$)
6. H_{06} terpenuhi dimana tidak terdapat korelasi antara *Military Strength Power* dan *Geography* ($\rho = 0.05601414179223369$)

Pada penelitian ini, nilai korelasi diperoleh melalui analisis korelasi antara kekuatan militer dan faktor-faktor lain yang diambil dari dataset. Pengukuran korelasi dilakukan menggunakan bahasa pemrograman Python dengan menggunakan fungsi *pearson* dari *library* yang sesuai. Sistem yang telah melakukan korelasi menghasilkan nilai hasil korelasi dan nilai ρ . Rata-rata ρ diperoleh dari rata-rata nilai ρ setiap bagian pada sektor.

Tabel 2. Nilai Korelasi

Sektor	Bagian	Hasil Korelasi	Nilai ρ	Rata-rata ρ
Airforce	Aircraft Strength value	-0.2673	0.0015	0.0026
	Fighter/Intercepto Strength value	-0.3109	0.0002	

Sektor	Bagian	Hasil Korelasi	Nilai ρ	Rata-rata ρ
Land	Attack Aircraft Strength value	-0.2419	0.0042	0.0009
	Transport Aircraft Fleet Strength value	-0.2671	0.0015	
	Trainer Aircraft Fleet value	-0.2848	0.0007	
	Helicopter Fleet Strength value	-0.2445	0.0038	
	Attack Helicopter Fleet Strength value	-0.2317	0.0062	
	Tank Strength value	-0.3243	0.0001	
	AFV/APC Strength value	-0.3618	0.0000	
	Self-Propelled Artillery Strength value	-0.2689	0.0014	
	Towed Artillery Strength value	-0.3354	0.0001	
	Rocket Projector Strength value	-0.2478	0.0034	
	Navy Fleet Strengths value	-0.3818	0.0000	
	Aircraft Carrier Fleet Strength value	-0.1659	0.0518	
	Navy Strength	Submarine Fleet Strength value	-0.2743	
Finance	Destroyer Fleet Strength value	-0.1758	0.0392	0.0145
	Frigate Fleet Strength value	-0.3594	0.0000	
	Defense spending budget value	-0.1924	0.0238	
Logistic	Extrenal Debt value	-0.2357	0.0054	0.0420
	Airport Total value	-0.1734	0.0420	
Geography	Total Square Land Area value	0.1053	0.2192	0.0560
	Total Coastline Coverage value	-0.2844	0.0007	
	Total Waterwat Coverage Value	-0.2861	0.0007	
	Total Border Coverage value	-0.2471	0.0035	

Tabel 3. Hasil Korelasi

Sektor	Rata-Rata
Airforce	-0.2640
Land	-0.3076
Navy Strength	-0.2714
Finance	-0.2140
Logistic	-0.1733
Geography	-0.1780

Berdasarkan pada Tabel 3, Nilai korelasi yang didapatkan pada penelitian ini menghasilkan nilai negatif yang berarti nilai yang paling mendekati -1 adalah yang terbaik [17]. Hasil rata-rata korelasi antar sektor yang paling mendekati nilai -1 adalah sektor *Land* yang bernilai -0.3076. Hasil rata-rata korelasi antara *Military Power index* dengan sektor *Navy Strength* bernilai -0.2714, hasil rata-rata korelasi antara *Military Power index* sektor *Airforce* bernilai -0.2640, hasil rata-rata korelasi antara *Military Power* sektor *Finance* bernilai -0.2140, hasil rata-rata korelasi antara *Military Power* sektor *Geography* bernilai -0.1780 dan hasil rata-rata korelasi antara *Military Power* sektor *Logistic* bernilai -0.1733. *Land* menjadi sektor yang paling berkorelasi dengan *Military Power* salah satunya karena tingginya ancaman teritorial terjadi di darat. Negara cenderung mengalami konflik darat lebih sering daripada konflik di laut atau udara, sehingga dapat dinyatakan hipotesis dari penelitian ini adalah bahwa sektor *Land* memiliki pengaruh parsial yang signifikan terhadap *Military Power*, yang ditunjukkan oleh nilai korelasi yang paling mendekati -1 (-0.3076). *Logistic* menjadi sektor dengan pengaruh parsial terkecil (-0.1733) karena *Logistic* hanya melibatkan aspek-aspek seperti pasokan dan transportasi, tidak berpengaruh secara signifikan terhadap kekuatan tempur langsung seperti yang terjadi pada sektor-sektor lain seperti darat, laut, atau udara. Sehingga *Logistic* bukanlah sektor yang paling berkorelasi dengan *Military Power*. Walaupun *Logistic* tetap penting, hubungannya dengan *Military Power* mungkin lebih terbatas dibandingkan dengan faktor-faktor lain yang secara langsung terlibat dalam kekuatan militer.

4. SIMPULAN

Berdasarkan dari hasil penelitian yang dilakukan, keseruhan sektor memiliki korelasi terhadap *Military Power* (H_a terpenuhi) kecuali *Geography* dimana nilai p dari *Geography* tidak lebih dari 0.05 (H_0 terpenuhi). Nilai korelasi terbaik didapatkan oleh sektor *Land* dimana korelasi dari *Military Power* dengan *Land* sebesar -0.3076, *Land* menjadi sektor yang paling berkorelasi dengan *Military Power* salah satunya karena tingginya ancaman teritorial terjadi di darat. Negara cenderung mengalami konflik darat lebih sering daripada konflik di laut atau udara, oleh karena itu *Land* menjadi sektor yang paling berkorelasi dengan *Military Power*, sehingga pemangku kepentingan dalam pengambilan keputusan yang berkaitan dengan kekuatan militer dapat meningkatkan kekuatan militer pada sektor darat untuk memperkuat kekuatan militer.

5. SARAN

Penelitian ini memiliki keterbatasan pengujian model karena hanya menggunakan korelasi pearson, sehingga untuk peneliti berikutnya dapat mencoba metode korelasi yang lain seperti spearman, phi-korelasi, biserial, dll untuk membandingkan hasil dari metode yang berbeda.

DAFTAR PUSTAKA

- [1] S. Hardiyanti, "Kebijakan Militer Pemerintah Amerika Serikat Dalam Memerangi Kelompok Taliban di Afghanistan Pada Kepemimpinan Barack Obama Periode 2009-2012," *Jurnal Ilmiah Mahasiswa FISIP Unsyiah*, vol. 3, no. 1, pp. 1-39, 2018.
- [2] R. B. Ramadhan, "Pengaruh Peningkatan Kekuatan Iran Terhadap Hegemoni Amerika Serikat di Timur Tengah," *Analisis: Jurnal Studi Keislaman*, vol. 19, no. 1, pp. 211-226, 2019.
- [3] A. R. Riadhi, M. K. Aidid and A. S. Ahmar, "Analisis Penyebaran Hunian dengan Menggunakan Metode Nearest Neighbor Analysis," *Journal of Statistics and Its Application on Teaching and Research*, vol. 2, no. 1, pp. 46-52, 2020.
- [4] E. T. Susdarwono and F. Sani, "Korelasi Antara Human Development Terhadap Anggaran Pertahanan Negara Menurut GDP," *Jurnal Wira Ekonomi Mikrosil*, vol. 13, no. 2, pp. 113-125, 2023.
- [5] S. N. Maulidah, Budiono and M. W. Kurniawan, "PENGARUH BUDAYA SEMI MILITER TERHADAP WAWASAN KEBANGSAAN SISWA DI SMK NEGERI 2 TUREN," *Jurnal Civic Hukum*, vol. 7, no. 2, pp. 174-182, 2022.
- [6] D. A. Yuanto and Firdaus, "Besarnya Pengaruh Penerapan Gaya Kepemimpinan Partisipatif Komandan Satuan Militer Terhadap Profesionalisme Prajurit," *Jurnal Ilmiah Manajemen*, vol. 16, no. 2, pp. 111 - 120, 2019.
- [7] G. E. Saputro, A. M. Rivai and Meirinaldi, "PENGARUH ANGGARAN PERTAHANAN, IMPOR ALUTSISTA, EKSPOR ALUTSISTA, DAN INFLASI TERHADAP PERTUMBUHAN EKONOMI DI INDONESIA TAHUN 1980-2019," *Jurnal Ekonomi*, vol. 23, no. 2, pp. 103 - 115, 2021.
- [8] D. G. Nugroho, F. Johan and Mufidin, "Pemanfaatan Data Sciencedalam Meningkatkan Kualitas Pembinaan SDM Militer," *Jurnal Pendidikan dan Pembelajaran*, vol. 4, no. 1, pp. 339-348, 2023.

- [9] F. Sari, R. B. K. W and S. Suhardianti, "PENGARUH NOTAM LATIHAN MILITER TERHADAP PELAYANAN PEMANDUAN LALU LINTAS PENERBANG DI PERUM LPPNPI CABANG PEMBANTU CURUG," *Jurnal Ilmiah Aviasi*, vol. 14, no. 03, pp. 23 - 30, 2021.
- [10] K. A. Fachrudin, D. L. Tarigan and M. F. Iman, "Analisis Rating dan Harga Kamar Hotel Bintang Lima di Indonesia (Analysis of Ratings and Room Rates for Five-Star Hotels in Indonesia)," *Jurnal Akuntansi, Keuangan, dan Manajemen (Jakman)*, vol. 3, no. 3, pp. 207-223, 2022.
- [11] N. Arifin, R. H. Irawan and I. N. Farida, "Algoritma K-Means Untuk Memprediksi Stok Bahan Baku Produksi," *Seminar Nasional Teknologi dan Sains*, vol. 1, no. 1, pp. 333-342, 2022.
- [12] F. F. Abdulloh and I. R. Pambudi, "Analisis Sentimen Pengguna Youtube Terhadap Program Vaksin COVID-19," *CSRID: Computer Science Research and ITS Development Journal*, vol. 13, no. 3, pp. 141-148, 2021.
- [13] R. Azhar, A. Surahman and C. Juliane, "Analisis Sentimen Terhadap Cryptocurrency Berbasis Python TextBlob Menggunakan Algoritma Naive Bayes," *Jurnal Sains Komputer dan Informatika*, vol. 6, no. 1, pp. 267-281, 2022.
- [14] Miftahuddin, A. Pratama and I. Setiawan, "Analisis Hubungan Antara Kelembaban Relatif Dengan Beberapa Variabel Iklim Dengan Pendekatan Korelasi Pearson Di Samudera Hindia," *Jurnal Siger Matematika*, vol. 02, no. 01, pp. 25 - 33, 2021.
- [15] M. K. Alim and D. B. Arianto, "Analisis Korelasi Antara Faktor Ekonomi dan Distribusi Penduduk di Jawa Timur Tahun 2020 Menggunakan Metode Korelasi Pearson," *Kohesi: Jurnal Sains dan Teknologi*, vol. 1, no. 4, pp. 20-30, 2023.
- [16] I. Nanda, PENELITIAN TINDAKAN KELAS UNTUK GURU INSPIRATIF, Indramayu: CV Adanu Abimata, 2021.
- [17] C. A. Yanti and I. J. Akhri, "PERBEDAAN UJI KORELASI PEARSON, SPEARMAN DAN KENDALL TAU DALAM MENGANALISIS KEJADIAN DIARE," *Jurnal Endurance : Kajian Ilmiah Problema Kesehatan*, vol. 6, no. 1, pp. 51-58, 2021.

Otomatisasi Pemisah Minyak dan Air Laut Menggunakan Separator pada Kapal Berbasis IoT

Hana Fitri Fikriyah¹, Anggit Yustyawan², Muhammad Farizul Hadi³, Faisal Budiman⁴,
Brahmantya Aji Pramudita⁵

^{1,2,3,4,5} Program Studi Teknik Elektro, Fakultas Teknik Elektro, Universitas Telkom Bandung

E-mail: ¹[*gripaice@student.telkomuniversity.ac.id](mailto:gripaice@student.telkomuniversity.ac.id), ²anggityustyawan@student.telkomuniversity.ac.id,
³mfarizulhadi@student.telkomuniversity.ac.id, ⁴faisalbudiman@telkomuniversity.ac.id,
⁵brahmantyaajip@telkomuniversity.ac.id

Abstrak – Perancangan alat pemisah minyak dan air laut berbasis Internet of Things (IoT) telah dilakukan untuk menanggulangi permasalahan tumpahan minyak pada ekosistem air. Sistem dirancang menggunakan separator sebagai metode pemisahan minyak dan air laut dengan memanfaatkan berat massa cairan minyak dan air laut, sistem yang dirancang memiliki kelebihan dapat mendeteksi keberadaan alat pemisah minyak dan air laut serta dapat mengukur minyak yang telah dipisahkan, kedua data tersebut nantinya akan dikirim menggunakan LoRa agar dapat dilihat di aplikasi kodular pada handphone. Hasil penelitian menunjukkan minyak yang bisa dipisahkan sebesar 91%. Harapan sistem dirancang dapat memisahkan minyak yang terdapat pada ekosistem air tanpa mencemari kembali ekosistem air tersebut.

Kata Kunci — Internet of Things, Pemisah Minyak dan Air Laut, Separator

1. PENDAHULUAN

Pencemaran ekosistem air karena tumpahan minyak merupakan permasalahan serius yang dapat merugikan ekosistem air. Berbagai faktor seperti operasi kapal tanker (penggunaan air ballast), kegiatan perbaikan dan perawatan kapal (docking), terminal bongkar muat di tengah laut, saluran pembuangan air bilga (berisi minyak dan pelumas dari proses mesin), pemusnahan kapal, dan faktor lainnya dapat menyebabkan pencemaran air laut akibat tumpahan minyak [1]. Dampak dari pencemaran minyak terhadap ekosistem air dapat mencakup kematian massal ikan di tambak atau keramba, serta jenis kerang yang memiliki kemampuan migrasi yang rendah untuk menghindari dampak tumpahan minyak [1]. Selain merugikan ekosistem, pencemaran minyak pada ekosistem air juga dapat menimbulkan kerugian ekonomi. Sebagai contoh, tragedi Exxon Valdez di Prince William Sound, Alaska, Amerika Serikat pada tahun 1989 menyebabkan tumpahan sekitar 41 juta liter minyak mentah. Dampak ekonominya mencakup biaya pemulihan yang mencapai sekitar 6,2 miliar Dolar AS (sekitar Rp 9,5 triliun dengan kurs waktu itu), termasuk ganti rugi sebesar 1,1 miliar Dolar AS (sekitar Rp 1,6 triliun) [2].

Berbagai cara telah dilakukan untuk menyelesaikan polusi minyak pada air oleh para peneliti sebelumnya. Sebagai contoh paling sederhana dengan cara membakar minyak, tetapi cara ini sangat berbahaya karena menimbulkan api yang sangat besar dan juga menghasilkan gas beracun akibat pembakaran [3]. Contoh lainnya dengan cara surfaktan atau bahan kimia yang biasanya digunakan pada sabun yang dapat menstabilkan campuran minyak dan air, namun penggunaan surfaktan sebagai pemisah minyak dan air dapat menimbulkan permasalahan lainnya pada ekosistem air akibat terpapar oleh bahan kimia [4].

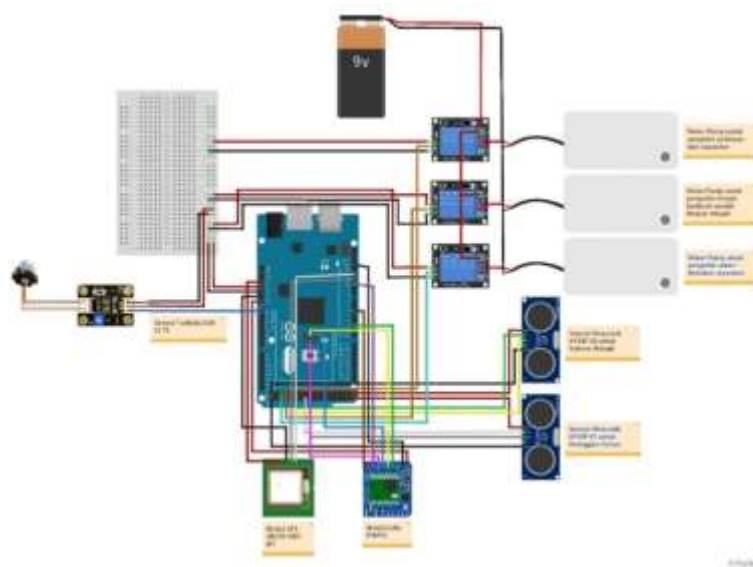
Penelitian ini berfokus pada perancangan alat “Pemisah minyak dan air laut menggunakan separator berbasis Internet of Things (IoT)”. Sistem terdiri dari separator sebagai wadah tempat pemisah minyak dan air laut dengan memanfaatkan massa jenis cairan. Mikrokontroler Arduino mega, sensor turbidity, sensor ultrasonik, LoRa dan modul GPS bertujuan sebagai pendeteksi dan *user interface*. Alat dirancang untuk memisahkan minyak dan air dengan metode separator, dimana metode ini dapat memisahkan minyak hanya dengan memanfaatkan massa jenis cairan. Minyak yang sudah terapung diatas air pada separator nantinya akan disedot oleh *water pump* untuk dipindahkan kedalam wadah khusus minyak. Pada wadah ini akan ada sensor ultrasonik yang berfungsi sebagai pengukur volume jumlah minyak. Data hasil pengukuran sensor ultrasonik nantinya akan dikirimkan melalui LoRa bersama dengan data dari modul GPS yang dapat dimonitoring di *handphone* pada aplikasi kodular, hal ini bertujuan agar dapat mempermudah pengguna [5] dalam memonitoring jumlah minyak yang sudah terpisah dan memonitoring keberadaan alat. Diharapkan alat yang dirancang dapat memisahkan minyak yang terdapat pada ekosistem air tanpa mencemari kembali ekosistem air tersebut.

2. METODE PENELITIAN

Penelitian ini memiliki fokus pada pengembangan alat pemisah minyak dan air otomatis berbasis *Internet of Things* (IoT), yang mengintegrasikan sistem elektronika dengan sistem IoT. Alat ini bertujuan untuk melakukan pemisahan minyak dari air secara otomatis. Secara umum, metodologi pembuatan alat ini dapat dijelaskan sebagai berikut:

2.1 Pengujian Komponen

Gambar 1 menunjukkan *wiring diagram* antara komponen elektronika dalam suatu rangkaian. Secara umum, rangkaian ini terdiri dari sensor, mikrokontroler, dan aktuator, dengan fungsi masing-masing dijelaskan lebih lanjut dalam Tabel 1. Komponen elektronika ini kemudian digabungkan menjadi sebuah perangkat elektronika untuk digunakan sebagai alat pemisah minyak dan air otomatis. Seluruh sensor telah melalui proses kalibrasi dan penyesuaian.



Gambar 1. *Wiring diagram*

Tabel 1. Komponen Elektronika pada Alat dan Fungsinya

Komponen Elektronika	Tipe/Model	Fungsi
Sensor Ultrasonik	HYSRF 05	Pengukur volume yang sudah dipisahkan dari air dan ketinggian separator
Sensor GPS	UBLOK NEO M7	Pendeteksi lokasi alat
Sensor Turbidity	SEN0175	Pendeteksi minyak
Modul LoRa	RFM95	Modul komunikasi untuk pengiriman data volume minyak dan keberadaan alat
Water Pump	Mini Water Pump 12V	Sebagai penyedot cairan minyak dan air ke dalam separator dan juga ke dalam wadah khusus minyak (hasil pemisahan)

Proses kalibrasi dan penyesuaian pada alat yang sudah dirangkai menjadi satu terbagi menjadi dua, yaitu:

1. Kalibrasi yang dilakukan pada sensor ultrasonik guna mencari besar persentase galat, hasil galat yang didapat kemudian dijadikan faktor koreksi pada pembacaan sensor ultrasonik yang akan digunakan

untuk membaca ketinggian cairan dan volume dari minyak yang sudah dipisahkan, adapun presentasi galat dihitung menggunakan persamaan (1) sebagai berikut:

$$\% \text{ error} = \frac{\text{jarak sesungguhnya} - \text{jarak sensor}}{\text{jarak sesungguhnya}} \times 100\% \dots\dots\dots(1)$$

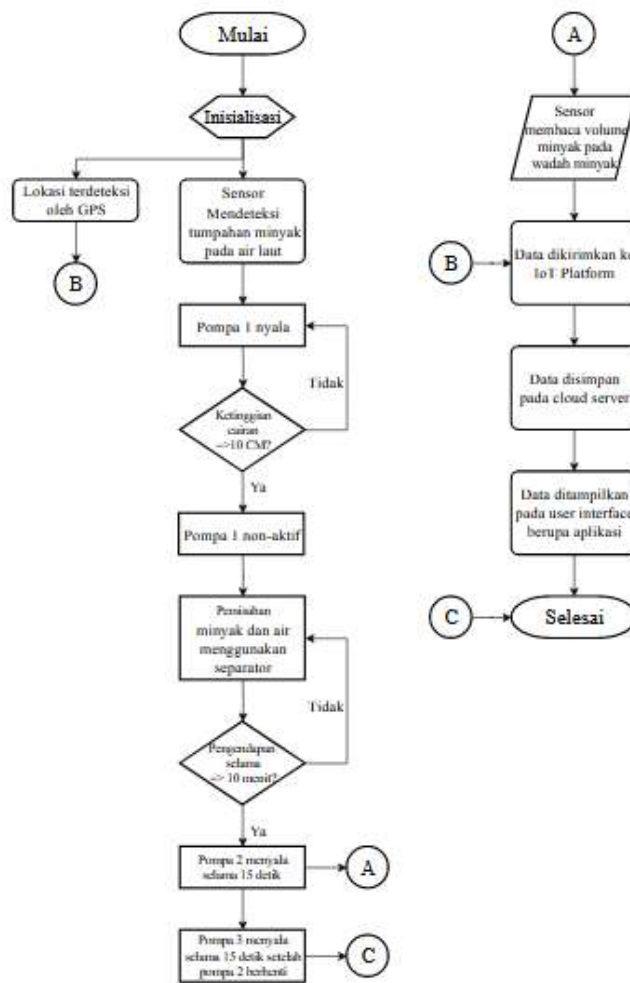
4.2. Persamaan (2) berguna untuk menghitung seberapa efisien alat pemisah minyak dan air menggunakan separator yang memanfaatkan massa jenis cairan dalam proses pemisahannya. Adapun rumus dari persamaan (2) untuk menghitung efisiensi alat pemisah minyak dan air sebagai berikut:

$$\frac{\text{minyak tersaring}}{\text{minyak awal}} \times 100\% \dots\dots\dots(2)$$

Selain proses kalibrasi dan penyesuaian terdapat juga proses pengujian yang dilakukan pada sensor *turbidity*, dimana pengujian dilakukan ketika sensor mendeteksi minyak maka akan menyalakan water pump, pendeteksian ini didasarkan oleh nilai tegangan yang dikeluarkan sensor *turbidity*, jika dia mendeteksi minyak maka tegangan yang dikeluarkan akan melebihi 4,2 V sedangkan jika mendeteksi air biasa akan mengeluarkan tegangan dibawah 4V.

2.2 Alur Sistem

Gambar 2 menampilkan representasi alur proses operasional dari perangkat pemisah minyak dan air otomatis yang telah dibuat. Terlihat bahwa pembacaan sensor yang relevan akan menggerakkan aktuatur pompa sebagai respons, memicu proses pemisahan minyak dan air secara otomatis dengan menggunakan separator. Setelah pemisahan, volume minyak yang dihasilkan diukur secara otomatis dengan bantuan sensor ultrasonik. Data otomatis dari pengukuran volume ini selanjutnya dikirim melalui platform IoT dan dapat diakses melalui aplikasi mobile.



Gambar 2. Diagram Alir Sistem

3. HASIL DAN PEMBAHASAN

Gambar 3 menunjukkan hasil rancangan perangkat pemisah minyak dan air otomatis. Terlihat dari gambar bahwa terdapat tiga wadah utama; wadah yang berisi campuran minyak dan air (*oily water*) sebagai simbol tumpahan minyak di air, wadah yang memuat separator (tempat terjadinya proses pemisahan minyak), dan wadah hasil pemisahan minyak yang telah berhasil dilakukan. Dalam aplikasinya, perangkat pemisah minyak dan air ini berpotensi efisien dan efektif untuk dalam lingkungan yang telah dibuat atau pada tahap *scaling up* untuk diterapkan dalam penanganan polusi air. Prinsip kerja pemisahan minyak dan air laut ini melibatkan penggunaan sensor *turbidity* sebagai pengindera minyak dalam air. Sensor ini akan mendeteksi keberadaan minyak, dan informasi tersebut akan diteruskan ke mikrokontroler. Mikrokontroler kemudian mengaktifkan pompa air untuk menarik minyak melalui penggunaan relay. Minyak yang ditarik oleh pompa air dialirkan ke separator untuk mengalami proses pemisahan dengan memanfaatkan perbedaan massa jenis cairan. Minyak yang mengapung di atas cairan kemudian dideteksi oleh sensor ultrasonik untuk mengukur ketinggiannya. Apabila minyak terdeteksi, pompa air akan diaktifkan kembali untuk mengalirkan minyak ke dalam wadah khusus minyak. Wadah ini dilengkapi dengan sensor ultrasonik yang berfungsi untuk mengukur volume minyak dalam tabung. Sementara itu, GPS juga berperan dalam mendeteksi lokasi alat pemisah minyak dan air. Data mengenai volume minyak dan keberadaan alat tersebut kemudian dikirimkan melalui jaringan LoRa. Data yang diterima oleh LoRa selanjutnya dikirimkan ke aplikasi Kodular melalui platform Firebase.



Gambar 3. Hasil Perancangan Alat

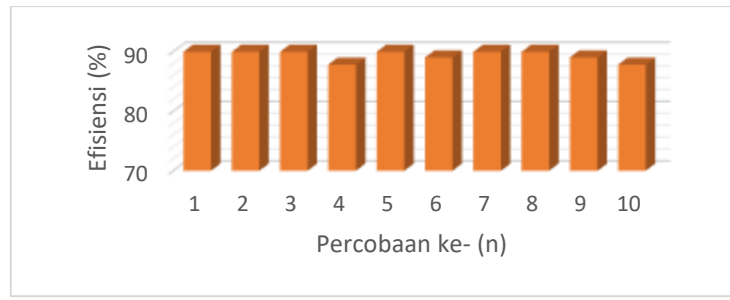


Gambar 4. Hasil Pengujian Sensor Ultrasonik

Pada penelitian ini, sensor ultrasonik dimanfaatkan untuk pengukuran volume serta pengukur ketinggian cairan didalam separator. Proses pengujian kalibrasi dilakukan untuk mengetahui besar galat yang diberikan sensor dengan membandingkan pembacaan sensor dengan pembacaan standar panjang. Pengujian dilakukan dari jarak 3-30 cm dengan interval perbedaan setiap 3 cm. Hasil pengujian yang ditampilkan pada Gambar 4 menunjukkan bahwa besar persentase galat yang didapat menggunakan persamaan (1) adalah 0,68%.

Hasil galat yang didapat kemudian dijadikan faktor koreksi pada pembacaan sensor ultrasonik yang nantinya akan mengukur ketinggian cairan dan volume minyak yang sudah dipisahkan. Faktor koreksi nantinya akan dimasukkan kedalam kodingan pada proses dilaksanakan, hal ini nantinya akan mempengaruhi angka pengukuran sensor ultrasonik yang menjadi lebih baik dari pada sebelum dimasukkan nilai faktor koreksi didalam kodingannya.

Hasil pengujian pemisahan minyak dan air otomatis ditunjukkan pada Gambar 5. Pengujian dilakukan berulang sebanyak 10 kali untuk melihat repeatability dan recyclability dengan kondisi pengujian awal adalah, *oily water* dengan rasio sebanyak 2:1 (minyak:air). Dapat dilihat bahwa hasil pemisahan minyak berhasil dilakukan dengan efisiensi rata-rata mencapai 91% berdasarkan perhitungan persamaan (2).



Gambar 5. Hasil Pengujian Pemisahan Minyak dan Air

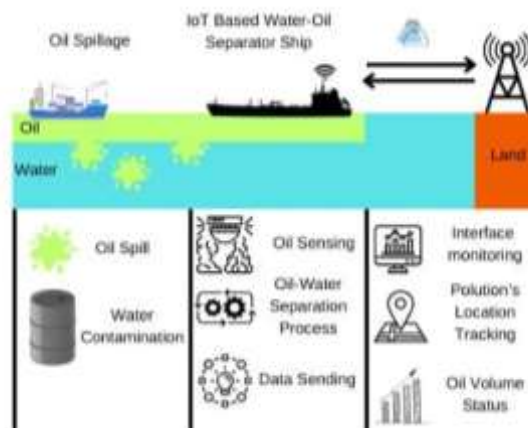
Pengujian pemisahan minyak dan air yang dilakukan pada separator dengan memanfaatkan massa jenis cairan membuat pemisahan minyak tidak 100% terpisah, hal ini juga dikarenakan pada proses pemindahan minyak ke wadah khusus minyak dilakukan oleh sensor ultrasonik sebagai sensor yang menentukan kapan *water pump* menyedot minyak untuk pindah, Dimana sensor ultrasonik tidak mengetahui mana cairan yang minyak dan mana yang air. Kedepannya pada separator digunakan sensor yang dapat membedakan minyak dan air, agar penyedotan oleh *water pump* lebih efisien.

Hasil dari pengujian pemisahan minyak dan air secara otomatis akan dikirim ke aplikasi kodular di handphone melalui modul LoRa, adapun tampilan pada aplikasi kodular seperti yang ditunjukkan pada Gambar 6. Pada aplikasi ini juga menunjukkan data lokasi berupa longitude dan latitude dari keberadaan alat.



Gambar 6. Tampilan Mobile Aplikasi

Mempertimbangkan hasil efektivitas pemisahan minyak dan air di atas, Berikut disampaikan potensi pengembangan yang dapat dilakukan dari alat yang telah dirancang. Gambar 7 menunjukkan ekosistem pemisahan minyak dan air yang berpeluang untuk diimplementasikan ketika terjadi adanya pencemaran air laut karena adanya tumpahan minyak.



Gambar 7. Ilustrasi Potensi Pengembangan Alat untuk Diimplementasikan sebagai Kapal Pemisah Minyak dan Air

3. SIMPULAN

Perancangan alat pemisah minyak dan air otomatis telah dilakukan pengkalibrasian sensor dan percobaan pemisahan minyak dan air pada separator. Dari pengkalibrasian dan percobaan tersebut didapatkan beberapa kesimpulan yang dapat diambil sebagai berikut.

1. Hasil pengujian kedua sensor ultrasonik didapat persentase rata-rata galat sebesar 0,68%.
2. Hasil pengujian pemisahan minyak dan air menunjukkan efisiensi rata-rata mencapai 91%
3. Hasil pengujian pengiriman data menunjukkan bahwa data dapat terkirim ke aplikasi.
4. Rencana pengembangan ke depan sensor untuk mendeteksi minyak diubah menjadi non kontak dan sensor yang menjadi pengatur kapan minyak dipindahkan ke wadah khusus minyak merupakan sensor yang dapat mendeteksi minyak.

4. SARAN

1. Penelitian selanjutnya terkait pemisahan minyak dan air menggunakan separator dapat menggunakan sensor non kontak pada bagian mendeteksi cairan minyak dan air, dikarenakan sensor *turbidity* merupakan sensor kontak yang memiliki kekurangan dalam mengubah data yang dideteksi dengan waktu singkat, akibatnya cairan minyak masih tertempel pada sensor menyebabkan sensor masih mendeteksi minyak dan *water pump* masih menyedot cairan kedalam separator.
2. Sensor yang digunakan sebagai pengatur kapan minyak dipindahkan ke wadah khusus minyak adalah sensor yang dapat mendeteksi minyak. Sehingga pemindahan minyak ke wadah khusus minyak dan dilakukan secara maksimal.

DAFTAR PUSTAKA

- [1] Kementerian Kelautan dan Perikanan RI, “Tumpahan Minyak (Oil Spill).” Accessed: Dec. 10, 2023. [Online]. Available: kkp.go.id
- [2] L. Alfa, “Perhitungan Biaya Kerugian Akibat Tumpahan Minyak Montara di Pesisir Nusa Tenggara Timur.”
- [3] A. P. Karlapudi *et al.*, “Role of biosurfactants in bioremediation of oil pollution-a review,” *Petroleum*, vol. 4, no. 3. KeAi Communications Co., pp. 241–249, Sep. 01, 2018. doi: 10.1016/j.petlm.2018.03.007.
- [4] Johannes Bagas Ardy Nugraha, “Sistem Pemisah Air dan Minyak Otomatis Menggunakan Nanofilter dan Sensor Ultrasonik.” Accessed: Dec. 10, 2023. [Online]. Available: <https://openlibrary.telkomuniversity.ac.id/pustaka/166093/sistem-pemisah-air-dan-minyak-otomatis-menggunakan-nanofilter-dan-sensor-ultrasonik.html>
- [5] H. T. Kristanto, A. S. Mustakim, M. Y. K. Huda, and R. Helilintar, “Perancangan Aplikasi Android Untuk Pemilihan Tempat Wisata Di Kota Kediri,” 2023, Accessed: Dec. 27, 2023. [Online]. Available: <https://proceeding.unpkediri.ac.id/index.php/stains/article/view/2846>

PERANCANGAN SISTEM INFORMASI JDIIH BERBASIS WEB DENGAN METODE *PROTOTYPE*

Ronaldo Arsad¹, Muhamad Son Muare²

^{1,2}Sistem Informasi, Fakultas Sains Dan Teknologi, Universitas Islam Negeri Radan Fatah Palembang
E-mail: ^{*1}ronaldoarsad6@gmail.com, ²muhamadsonmuari@radenfatah.ac.id

Abstrak – Perkembangan hukum dan perundang-undangan yang pesat menuntut kebutuhan akan Sistem Informasi yang efisien dan terintegrasi. Pada perumda tirta musi palembang belum terdapat JDIIH (Jaringan Dokumentasi Informasi Hukum). sehingga penelitian ini bertujuan untuk membuat Jaringan Dokumentasi dan Informasi Hukum sebagai wadah pendayagunaan dokumen hukum secara tertib, terpadu, dan berkesinambungan, serta merupakan sarana pemberian pelayanan informasi hukum secara lengkap, akurat, mudah, dan cepat. Pengembangan JDIIH menggunakan metode *prototype* secara berurut sesuai dengan tahapan metode tersebut pengembangan sistem ini menggunakan PHP, XAMPP, MySQL, dan Sublimetext3. Penelitian ini menghasilkan sistem JDIIH (Jaringan Dokumentasi dan Informasi Hukum) yang berisikan tentang nformasi, berita, dan surat pemberitahuan yang berada pada PDAM Tirta Musi Palembang berbasis website. Maka dari hal tersebut penelitian ini penting di lakukan dengan harapan dapat mempermudah dalam pencarian informasi hukum.

Kata Kunci —MySQL, Metode *Prototype*, Sistem Informasi, Website.

1. PENDAHULUAN

Perkembangan teknologi yang sangat cepat berpengaruh pada segala aspek kehidupan, terutama dalam penyampaian informasi bagi organisasi atau instansi yang memerlukan sistem yang dapat mengolah informasi dengan cepat, akurat, dan tepat [1]. Meskipun demikian, perkembangan JDIIH pada era digital di PDAM Tirta Musi Palembang masih terbatas. Oleh kerna itu peneliti mengembangkan sistem jdiih berbasis *website* dengan demikian menjadi solusi yang tepat untuk meningkatkan efesiesnsi dalam mengelola jdiih pada PDAM tirta musi palembang.

Dalam penyelenggaraan pelayanan publik, inovasi juga bisa diartikan sebagai proses memikirkan dan mengimplementasikan kebijakan publik yang original, penting, dan berdampak[2]. Dalam perancangan ini penulis membuat website JDIIH akan mempermudah dalam pencarian dan penelusuran peraturan perundang-undangan dan produk hukum lainnya. Sebagai media pemberitahuan informasi hukum di PDAM Palembang.

Belum ada survey kepuasan penggunaan aplikasi, sehingga admin tidak mengetahui kepuasan respon pengguna dalam menggunakan aplikasi, [1].

Salah satu sistem yang umum digunakan oleh instansi-instansi pemerintahan untuk mempublikasikan produk-produk hukumnya adalah sistem Jaringan Dokumen dan Informasi Hukum (JDIIH). Sistem ini menyimpan dan merekam data semua jenis produk hukum seperti peraturan perundangan-undangan, putusan pengadilan, yurisprudensi, monografi hukum, peraturan dan keputusan lainnya, dan lain sebagainya[3]. Dalam sistem jdiih terdapat fitur yang berisikan tentang informasi, berita, dan surat pemberitahuan di diharapkan dapat mempermudah dalam pencarian informasi hukum.

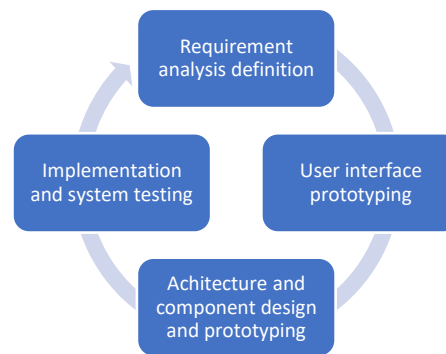
Disetiap organisasi pemerintah saat ini memiliki media pelayanan informasi yang berbasis TIK (Teknologi Informasi dan Komunikasi) sebagai upaya untuk melaksanakan tugas dan fungsinya secara prima..[4] Dengan pembuatan laporan secara terkomputerisasi akan lebih efektif dan efisien jika dibandingkan secara manual dengan menggunakan dokumen kertas karena akan lebih memakan waktu, tenaga, dan biaya.

Untuk itu peneliti bertujuan mengembangkan sebuah sistem JDIIH yang di diharapkan sistem tersebut dapat mempermudah dalam pencarian informasi, berita, surat pemberitahuan. Dengan itu peneliti menggunakan metode *prototype*. Berdasarkan masalah tersebut penulis berencana mempermudah pencarian data menggunakan *website* JDIIH.

2. METODE PENELITIAN

Metode *prototype* sangat baik digunakan untuk menyelesaikan masalah kesalah pahaman antara user dan analis yang timbul akibat user tidak mampu mendefinisikan secara jelas kebutuhannya[5]. Sebagian user kesulitan mengungkapkan keinginannya untuk mendapatkan aplikasi yang sesuai dengan kebutuhannya.

Kesulitan ini yang perlu diselesaikan oleh analis dengan memahami kebutuhan user dan menerjemahkannya ke dalam bentuk model (*prototype*).



Gambar 1. Metode *Prototype*

Beberapa langkah menggunakan metode *prototype* yang harus dilakukan untuk melakukan penelitian adalah

1. *Requirement analysis definition*

Penelitian di mulai dengan merumuskan masalah dengan melakukan studi literatur dan juga menentukan batasan masalah. Kemudian dari aktivitas sebelumnya maka akan didapatkan kebutuhan perusahaan dan juga tujuan dari penelitian dari kedua hal tersebut akan dilakukan analisis kebutuhan sistem yang sesuai dengan kebutuhan perusahaan.

2. *User interface prototyping*

Jika analisis kebutuhan sistem teridentifikasi maka dimulai pembuatan *prototype* user interface dari aplikasi untuk kemudian akan dilakukan pengujian oleh user secara langsung, jika user telah puas dengan user *interface prototype* yang telah dibuat maka akan dilakukan tahap selanjutnya.

3. *Achitecture and component design and prototyping*

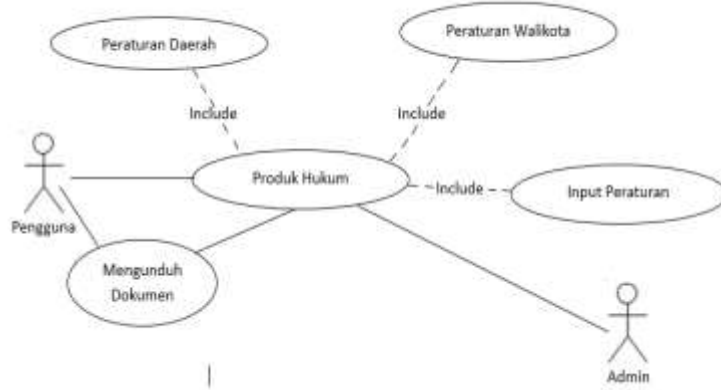
Pada tahapan ini klien akan dilakukan pengecekan terhadap *prototype* yang sudah dibangun dengan maksud memastikan sistem yang dirancang sudah sesuai dengan tujuan dan keperluan dari klien. Apabila *prototype* yang dibangun belum sesuai dengan keinginan klien maka akan dilakukan koreksi serta perbaikan dengan kembali. Tahap ini akan menyempurnakan tampilan *input* dan *output* yang belum sesuai ataupun perlu penambahan fitur baru.

4. *Implementation and system testing*

Pada tahap ini dilakukan pembuatan aplikasi berdasarkan dari model *prototype* sebelumnya yang telah di evaluasi oleh user dan sesuai dengan keinginan user. Setelah aplikasi telah dibangun maka akan dilakukan pengujian yang akan menghasilkan produk akhir dan selanjutnya akan di lakukan evaluasi kembali oleh user. Jika sesuai dengan kebutuhan user maka dilakukan implementasi ke perusahaan.

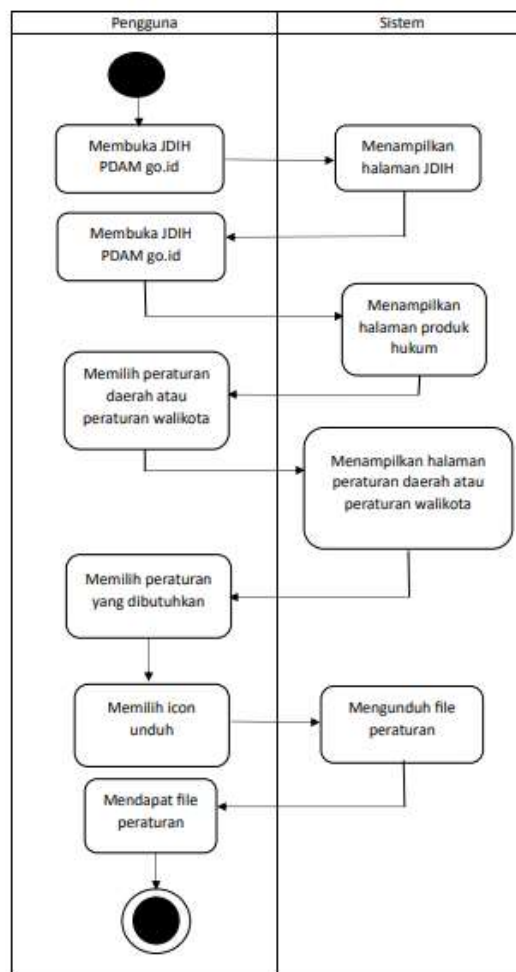
2.1 *Requirment Analysis Definition*

Pada bagian ini menjelaskan hasil dari perancangan website Jaringan Dan Informasi Hukum untuk mengatasi permasalahan publikasi peraturan pada PDAM yang selama ini tidak memiliki wadah untuk mempublikasikan peraturan dan informasi melalui internet. *Requirement Analysis* atau analisa kebutuhan adalah tahapan dimana peneliti melakukan observasi yang bertujuan menemukan suatu konflik yang ada dalam sistem inventory. pada tahap ini analisis dibagi menjadi dua yaitu analisis kebutuhan dan analisis spesifikasi. Analisis kebutuhan ialah kegiatan observasi serta wawancara narasumber terkait dan menggunakan studi literatur. isu serta data yang didapatkan berasal analisa kebutuhan akan digunakan menjadi point krusial buat selanjutnya dijadikan spesifikasi fitur apa saja yang akan dirancang pada penelitian ini. Analisis spesifikasi ialah analisa terhadap software serta hardware apa aja yg dipergunakan buat menciptakan sistem yg dikembangkan[6].



Gambar 2. Use Case Diagram

2.2 User Interface Prototyping

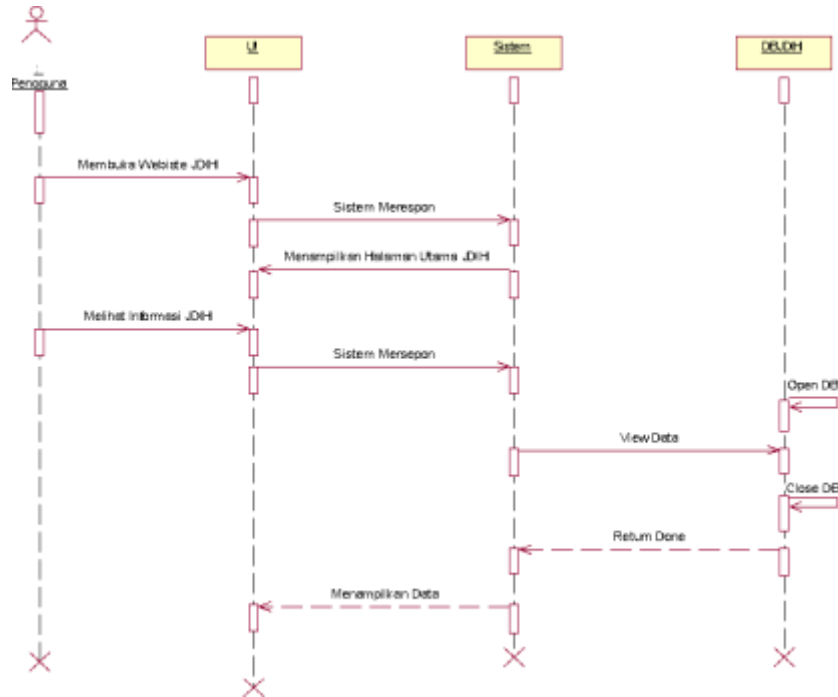


Gambar 3. Activity Diagram

Pada aplikasi Jaringan Dokumentasi dan Informasi Hukum. Pengguna hanya memiliki akses untuk melihat informasi-informasi hukum dan mengunduh dokumen peraturan yang ada didalam web Jaringan Dokumentasi dan Informasi Hukum PDAM. *Activity* Diagram merupakan rancangan aliran aktivitas atau aliran kerja dalam sebuah sistem yang akan dijalankan. *Activity* Diagram juga digunakan untuk mendefinisikan atau mengelompokan aliran tampilan dari sistem tersebut. Alur aktivitas aplikasi JDIH pada PDAM digambarkan dari pengguna masuk ke aplikasi hingga pengguna dapat mengunduh dokumen hukum.

2.3 sequence diagram

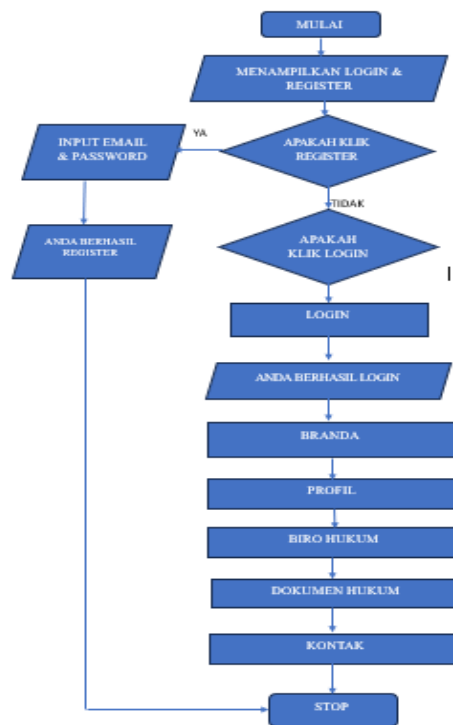
Pada *sequence* diagram Pengguna menggambarkan bagaimana interaksi dan komunikasi pada setiap objek-objek yang ada pada sistem dalam melakukan akses peraturan daerah/walikota sehingga dapat diunduh.



Gambar 4. Sequence Diagram

2.4 Flowchart

Flowchart ini bertujuan untuk menggambarkan urutan penyelesaian masalah dengan sederhana dan jelas. Tahap ini menjelaskan serta memastikan sebuah program memiliki alur sendiri serta memiliki acuan menyusun dan mengembangkan *website*. Berikut *flowchart* nya.



Gambar 5. Flowchart

2.5 Implementation & Unit Testing

Merupakan tahapan dimana dilakukan pengkodean dari desain sistem yang telah dibuat. Disini peneliti membuat *source code program* kedalam sub program seperti acara dari *create, read, update* serta *delete*. sub acara yg telah didesain selanjutnya akan diuji satu persatu apakah berjalan sesuai kegunaannya masing – masing. Bahasa pemrograman yang dipergunakan pada pengembangan sistem *inventory* ini merupakan *Java* dan *PHP*[6].

3. HASIL DAN PEMBAHASAN

Hasil dan pembahasan dari rancangan *interfase* bertujuan untuk mengembangkan rancangan tampilan *website* yang akan di buat, Berikut ini adalah beberapa rancangan :

3.1.1 Halaman Login

Pada halaman *login* merupakan tampilan awal ketika pengguna mengakses website sistem informasi jdih.



Gambar 6. Halaman Login

3.1.2 Halaman Registrasi

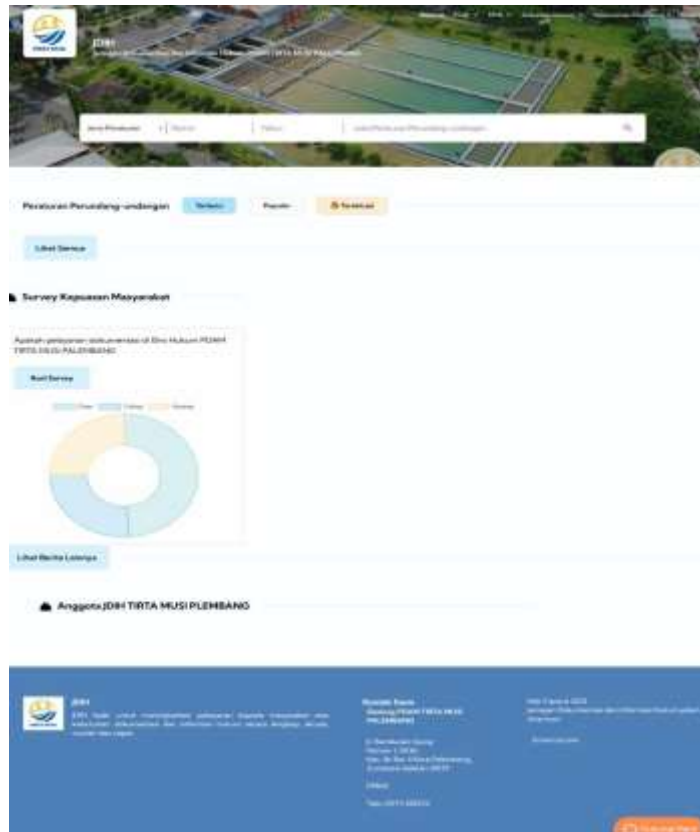
Pada halaman ini terlihat form *registrasi* apabila bagi pengguna yang belum mempunyai akun maka harus *registrasi* terlebih dahulu.



Gambar 7. Halaman Registrasi

3.1.3 Halaman Branda

Tampilan pada halaman branda untuk melihat peraturan yang ada pada pdam tirta musu Palembang bisa di telusuri oleh pengguna sesuai dengan peraturan yang di dikeluarkan, ada pilihan nomr, tahun, judul, ketika kita klik tombol pencarian maka akan muncul peraturan yang kita cari apabila telah di apload oleh admin.



Gambar 8. Halaman Beranda

3.1.4 Halaman Profil

Merupakan tampilan pada halaman profil yang berisi tentang sejarah singkat dari pdam tirta musipalembang, visi, misi, dan penghargaan yang telah di dapat oleh pdam tirta musipalembang.



Gambar 9. Halaman Profil

3.1.5 Halaman PPID Biro Hukum

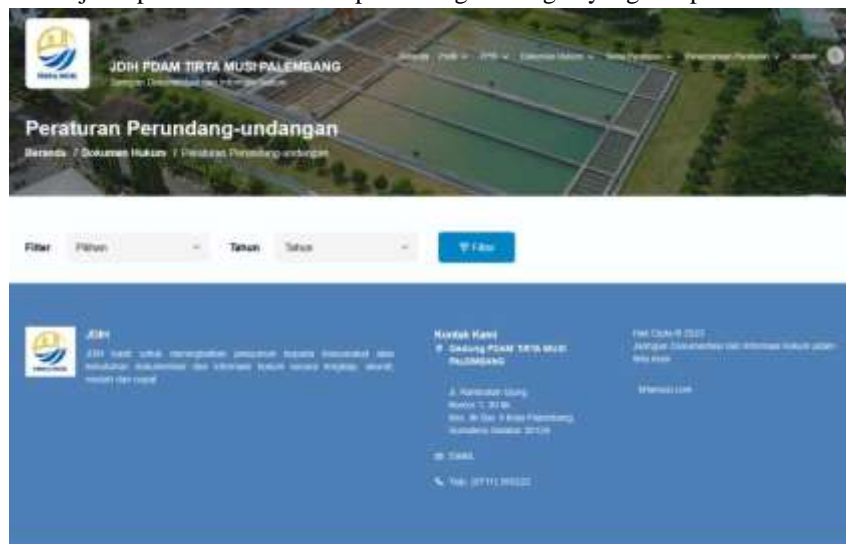
Tampilan pada halaman PPID Biro hukum, dimana pada halaman ini adalah peraturan hukum yang ada pdam tirta musi Palembang.



Gambar 10. Halaman PPID Biro Hukum

3.1.6 Halaman Dokumen Hukum

merupakan tampilan halaman dokumen hukum yang berisi tentang peraturan perundang-undangan, pengguna dapat melihat jenis peraturan dan tahun perundang-undangan yang ada pdam tirta musi Palembang.



Gambar 11. Halaman Dokumen Hukum

3.1.7 Halaman Kontak

Tampilan halaman kontak, yang mana berisi alamat kantor pdam, telpon, email, website, dan instagram



Gambar 12. Halaman Kontak

3.2 Pengujian dan penggunaa aplikasi

Setelah selesai melakukan pembuatan program yang menghasilkan rancangan muka selanjutnya dibuatkan pengujian yang dimana untuk mengetahui apakah program itu berhasil atau tidak.

Tabel 1 Pengujian aplikasi

Test Case Description	Test Case	Expected Result	Actual Result	Kesimpulan
Login sebagai pengguna	User : Pass : 123	Sistem akan berpindah ke tampilan awal branda pengguna	Pengguna akan berpindah ke tampilan awal pengguna	Berhasil
Pengguna memilih branda	Jenis peraturan, nomor, tahun, judul.	Sistem akan berpindah pada tampilan surat jdih	Sistem menampilkan surat jdih	Berhasil
Pengguna memilah ke halaman profil	Tentang jdih	Sistem akan berpindah ke tampilan profil pdam	Sistem akan menampilkan sejarah singkat pdam, visi, misi, dan penghargaan	Berhasil
Pengguna memilih halaman biro hukum	Peraturan, judul, file	Sistem akan berpindah ke tampilan biro hukum pdam	Sistem akan menampilkan halaman tentang peraturan pdam.	Berhasil

Test Case Description	Test Case	Expected Result	Actual Result	Kesimpulan
Pengguna memilih halaman dokumen hukum	Pilihan, tahun	Sistem akan berpindah kr tampilan peraturan perundang undangan	Sistem akan menampilkan peraturan perundang undangan yang di cari	Berhasil
Pengguna memilih halaman kontak	Nama, alamat email, no hp, instansi.	Sistem akan berpindah ke tampilan kontak	Sistem akan menampilkan nama, alamat email, no hp dan alamat instansi.	Berhasil

4. SIMPULAN

Berdasarkan hasil analisis dan perancangan seluruh kegiatan kerja praktek dapat di tarik kesimpulannya yaitu sebagai berikut :

1. Merancang suatu *system* informasi yang di rancang menggunakan *prototype* ini bertujuan untuk memudahkan pengguna dalam melihat peraturan perundang-undangan hukum yang ada pada PDAM Tirta Musi Palembang,
2. Membuat sistem lebih berjalan seiring perkembangan teknologi secara efektif dan efisien.
3. Merancang sistem informasi JDIH ini diharapkan bisa mempermudah aktivitas pekerjaan di PDAM dan sistem yang di rancang ini masih dapat di kembangkan lagi secara sistem agar dapat lebih baik lagi..

5. SARAN

Terdapat beberapa saran dari penulis untuk perancangan sistem informasi yang di usulkan oleh penulis antara lain:

1. Agar pengembangan lebih baik, disarankan untuk melakukan pengembangan lebih lanjut dengan menggunakan metode atau pengembangan sistem lain agar menjadi perbandingan referensi agar rancangan bias bekerja semaksimal mungkin.
2. Serta adanya pemeliharaan terhadap sistem yang telah di rancang..

DAFTAR PUSTAKA

- [1] A. A. Bouty, A. A. Kadim, B. S. Rijal, M. S. Tuloli, and R. Kaharu, "Pengembangan Aplikasi Jaringan dan Dokumentasi Informasi Hukum dengan Metode User Centered Design," *Jambura J. Informatics*, vol. 5, no. 1, pp. 68–76, 2023, doi: 10.37905/jji.v5i1.17923.
- [2] K. Umam, G. H. Wibowo, and I. W. Suardinata, "Penerapan Aplikasi JDIH di Desa Tambong Sebagai Media Publikasi Peraturan Desa ke Masyarakat," *Abdimasku J. Pengabd. Masy.*, vol. 6, no. 2, p. 402, 2023, doi: 10.33633/ja.v6i2.1040.
- [3] R. R. Putra and A. Ariesmansyah, "INOVASI SEKTOR PUBLIK DALAM PENGELOLAAN JDIH (Jaringan Dokumentasi Dan Informasi Hukum) DI BAGIAN HUKUM SEKRETARIAT DAERAH KABUPATEN SUMEDANG," *J. Ilmu Sos.*, vol. 21, no. 2, pp. 47–56, 2023, [Online]. Available: <http://jurnaldialektika.com/>
- [4] P. K. Tengah, "Abstrak".
- [5] E. W. Fridayanthie, H. Haryanto, and T. Tsabitah, "Penerapan Metode Prototype Pada Perancangan Sistem Informasi Penggajian Karyawan (Persis Gawan) Berbasis Web," *Paradig. - J. Komput. dan Inform.*, vol. 23, no. 2, pp. 151–157, 2021, doi: 10.31294/p.v23i2.10998.
- [6] H. Kristanto, A. S. Mustakim, and ..., "Perancangan Aplikasi Android Untuk Pemilihan Tempat Wisata Di Kota Kediri," ... *Teknol. \& ...*, vol. 2, pp. 27–34, 2023, [Online]. Available: <https://proceeding.unpkediri.ac.id/index.php/stains/article/view/2846>

Optimasi Model Prediksi Kesuksesan Startup Menggunakan StandartScaler Transform

Wulan Sri Lestari¹

¹Teknologi Informasi, Fakultas Informatika, Universitas Mikroskil Medan

E-mail: *¹wulan.lestari@mikroskil.ac.id

Abstrak – Pertumbuhan pesat startup di era teknologi modern menimbulkan tantangan kompleks dalam memprediksi kesuksesan suatu startup. Berdasarkan data yang diperoleh, 90 dari 100 startup gagal mencapai tingkat pendanaan yang diinginkan. Keberhasilan sebuah startup tidak hanya mencerminkan prestasi bisnis individu tetapi juga memiliki dampak yang luas pada dinamika ekonomi global. Oleh karena itu, perlu dikembangkan model prediksi yang handal sebagai dasar pengambilan keputusan yang efektif dalam pengembangan startup. Untuk mengatasi masalah ini, penelitian ini menggunakan integrasi Deep Neural Networks dengan StandartScaler Transform untuk mengoptimalkan model prediksi kesuksesan startup. StandartScaler Transform digunakan untuk menormalkan distribusi data sebelum diterapkan pada Deep Neural Networks. Hasil eksperimen pada model prediksi kesuksesan startup menunjukkan peningkatan akurasi sebesar 35,25% pada tahap training, dan peningkatan tambahan sebesar 36,95% pada tahap testing. Model prediksi kesuksesan startup yang dibangun berhasil mengatasi masalah overfitting/underfitting, sehingga menjadikannya sebagai dasar untuk pengembangan aplikasi prediksi kesuksesan startup.

Kata Kunci — Kesuksesan Startup, Deep Neural Networks, StandartScaler Transform, Optimasi, Prediksi

1. PENDAHULUAN

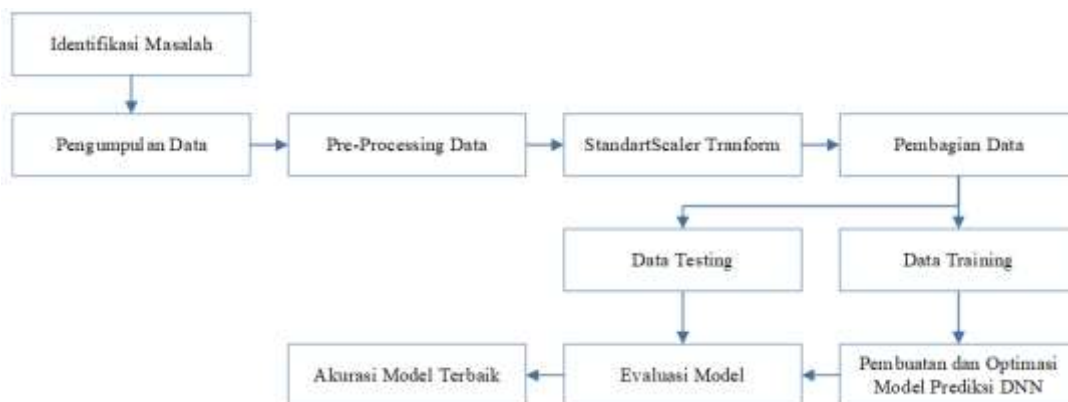
Perkembangan yang pesat dalam teknologi dan dinamika pasar global telah menciptakan lonjakan jumlah startup di berbagai sektor industri. Startup merupakan entitas bisnis yang terbentuk oleh sekelompok individu dengan tujuan mengembangkan produk atau layanan melalui pemanfaatan teknologi informasi [1]. Dalam kondisi ketidakpastian, startup memiliki tingkat risiko kegagalan yang signifikan karena produk atau layanan yang dikembangkan mungkin tidak sesuai dengan kebutuhan pengguna atau tidak mampu bersaing dengan perusahaan sejenis [2]. Keberhasilan suatu startup tidak hanya tergantung pada kebaruan ide yang diusungnya, melainkan juga pada kapasitas untuk mengelola dan menganalisis data guna membuat keputusan strategis yang efisien. Di era bisnis yang penuh dengan pertumbuhan startup yang pesat, peran dan dampaknya terhadap inovasi dan ekonomi semakin signifikan. Lonjakan jumlah startup mencerminkan kehidupan yang dinamis dalam ekosistem bisnis. Namun, seiring dengan pertumbuhan ini, hadir tantangan kompleks dan ketidakpastian yang dihadapi oleh startup. Secara keseluruhan, 90 dari 100 startup mengalami kegagalan dalam mencapai tingkat pendanaan yang diinginkan [3]. Keberhasilan suatu startup tidak hanya memiliki dampak pada kesuksesan bisnis individu tersebut, melainkan juga memberikan pengaruh luas terhadap dinamika ekonomi global. Oleh karena itu, dibutuhkan pengembangan model prediksi yang handal untuk memahami variabel-variabel yang mempengaruhi keberhasilan startup, sebagai dasar pengambilan keputusan yang baik dalam pengembangan startup.

Terdapat beberapa penelitian yang telah dilakukan sebelumnya dalam membangun model prediksi kesuksesan startup. Pada tahun 2021, Adhitya et al. melakukan penelitian prediksi kesuksesan startup dengan memanfaatkan *Decision Tree*, *K-NN*, dan *Naïve Bayes*, yang menghasilkan tingkat akurasi masing-masing sebesar 79,29%, 66,69%, dan 64,21% [4]. Dalam penelitian tersebut [4], jumlah variabel yang digunakan untuk pembuatan model prediksi sebanyak 19 dari total 49 variabel dalam *dataset*. Bangdiwala et al., membangun model prediksi startup menggunakan *Decision Trees*, *Random Forest*, *Gradient Boost*, *Logistic Regression*, dan *MLP Neural networks*. Hasil penelitian Bangdiwala et al., mendapatkan nilai akurasi sebesar 92% [5]. Ross et al, melakukan penelitian menggunakan algoritma *machine learning* untuk menghasilkan model prediksi startup dan mendapatkan nilai akurasi sekitar 80 – 89% [6]. Pada tahun 2022, Wulan & Halim menggunakan *Deep Neural Networks* untuk membangun model prediksi kesuksesan startup. Hasil penelitian menunjukkan bahwa model yang dihasilkan tidak mengalami *underfitting/overfitting* dengan tingkat akurasi sebesar 83,93% [7]. Untuk meningkatkan nilai akurasi dari penelitian [7], maka pada penelitian ini digunakan *StandartScaler Transform*.

StandartScaler Transform [8] melakukan transformasi pada *dataset* sehingga nilai rata-rata distribusinya menjadi nol dan deviasi standarnya menjadi satu. Nilai yang telah mengalami transformasi diperoleh dengan mengurangi nilai aslinya dari rata-rata dan membaginya dengan deviasi standar. Thara et, al. [9], menggunakan *StandartScaler Transform* untuk proses *feature scaling* pada model auto deteksi gejala kejang menggunakan *Deep Neural Networks*. Hasil percobaan yang dilakukan [9], diperoleh bahwa *StandartScaler Transform* menghasilkan akurasi yang lebih baik dibandingkan *RobustScaler*, *Normalizer*, dan *MinMaxScaler*.

Salah satu tahapan penting dalam mempersiapkan data sebelum diproses ke *Deep Neural Networks* adalah menerapkan *StandartScaler Transform*. Pendekatan ini berperan dalam normalisasi distribusi data, menciptakan dasar yang lebih optimal untuk *Deep Neural Networks* dalam memahami pola-pola yang kompleks. Oleh karena itu, penelitian ini berfokus pada integrasi *Deep Neural Networks* dan *StandartScaler Transform* dengan tujuan mengoptimalkan model prediksi kesuksesan *startup*. Melalui kombinasi *Deep Neural Networks* dan efektivitas *StandartScaler Transform*, diharapkan penelitian ini dapat memberikan sumbangan signifikan terhadap perkembangan metodologi prediksi dalam lingkungan *startup*, memberikan dampak positif pada pertumbuhan dan keberlanjutan ekosistem *startup*.

2. METODE PENELITIAN



Gambar 1. Tahapan Penelitian

Untuk menghasilkan model prediksi kesuksesan *startup* dengan nilai akurasi yang lebih baik, maka dilakukan tahapan penelitian seperti yang terlihat pada Gambar 1.

Berikut ini merukan penjelasan dari setiap tahapan yang dilakukan.

2.1 Identifikasi Masalah

Proses identifikasi masalah merupakan tahapan awal yang penting dilakukan untuk menganalisis apa yang menjadi masalah pada penelitian ini seperti yang telah dijabarkan pada bagian pendahuluan.

2.2 Pengumpulan Data

Pengumpulan data bertujuan untuk mengumpulkan data yang tepat sesuai dengan kebutuhan penelitian. Data yang digunakan dalam penelitian ini adalah data sekunder yang diambil dari website <https://www.kaggle.com/datasets/manishkc06/startup-success-prediction>. *Dataset Startup Success Prediction* adalah kumpulan data yang mencakup perusahaan *startup* yang beroperasi di Amerika Serikat. Dalam *dataset* ini, terdapat 923 data yang mencakup 597 *startup* yang berhasil diakuisisi dan 326 *startup* yang telah ditutup. *Dataset* ini memiliki 49 variabel, dimana 48 variabel digunakan sebagai atribut, sedangkan 1 variabel (status) berfungsi sebagai kelas.

2.3 Pre-Processing Data

Berikut merupakan proses *pre-processing* yang dilakukan.

1. Pengecekan *Missing Value*

Missing value adalah kondisi di mana suatu variabel dalam data tidak memiliki nilai atau informasi yang lengkap, yang dapat disebabkan oleh kesalahan pengumpulan data atau ketidakhadiran responden. Penting untuk memeriksa dan menangani *missing value* sebelum melanjutkan ke proses analisis data agar mendapatkan data yang berkualitas.

2. Konversi Tipe Data

Pada tahap ini dilakukan proses konversi data-data kategorial *string* dan *object* menjadi data kategorial numerik sehingga dapat lebih mudah dipahami oleh model yang akan dibangun. Proses konversi dilakukan dengan menggunakan fungsi *label encoder*.

3. Pemilihan Variabel/Atribut

Pada tahap pemilihan variabel/atribut, pada penelitian ini digunakan *correlation matrix*. *Correlation matrix* merupakan metode statistik yang digunakan untuk mengevaluasi hubungan antara dua atau lebih variabel. Dalam *correlation matrix*, nilai korelasi berkisar antara -1 hingga 1 yang artinya jika dua buah variabel memiliki nilai 1, maka korelasi sempurna positif, yang berarti bahwa dua variabel bergerak bersamaan dalam arah yang sama. Jika nilai 0, maka tidak ada korelasi yang menunjukkan tidak ada

hubungan linier antara dua variabel. Sedangkan jika nilai -1, maka korelasi sempurna negatif, yang berarti bahwa dua variabel bergerak berlawanan arah. Proses ini dilakukan untuk menghapus data atau variabel yang tidak relevan dengan model prediksi kesuksesan *startup*.

2.4 StandartScaler Transform

StandartScaler transform merupakan metode umum yang digunakan dalam *machine learning* untuk menghasilkan distribusi normal pada variabel-variabel numerik dengan nilai rata-rata 0 dan standart deviasi 1. Proses ini dilakukan untuk membantu mengoptimalkan model prediksi kesuksesan *startup* yang akan dibangun menggunakan *Deep Neural Networks* sehingga dapat menghasilkan hasil prediksi yang akurat.

2.5 Pembagian Data

Setelah proses *StandartScaler transform* selesai dilakukan, tahap selanjutnya adalah membagi dataset menjadi 70% data *training* dan 30% data *testing*. *Data training* akan digunakan saat proses pembuatan model prediksi kesuksesan *startup* menggunakan *Deep Neural Networks* dan data *testing* digunakan untuk evaluasi model yang telah dihasilkan.

2.6 Pembuatan dan Optimasi Model Prediksi Kesuksesan Startup

Proses selajutnya adalah pembuatan dan optimasi model prediksi kesuksesan *startup* menggunakan data *training* yang telah dipre-processing dan melewati proses *StandartScaler Transform*. Model *Deep Neural Networks* yang dihasilkan bertujuan untuk menghasilkan akurasi yang tinggi untuk memprediksi apakah sebuah *startup* akan sukses atau akan tutup.

2.7 Evaluasi Model

Evaluasi model dilakukan dengan menggunakan *confusion matrix*. *Confusion matrix* merupakan tabel evaluasi yang digunakan untuk melihat kinerja model klasifikasi pada *supervised learning* kedalam kategori yang benar atau salah. *Confusion matrix* yang digunakan dalam penelitian ini adalah akurasi dan presisi seperti terlihat pada tabel 1.

Tabel 1. *Confusion Matrix*

Actual Class	Assigned Class	
	Positive	Negative
Positive	True Positive	False Negative
Negative	False Positive	True Negative

$$\text{Akurasi} = (TP + TN) / (TP+TN+FP+FN) \dots\dots\dots(1)$$

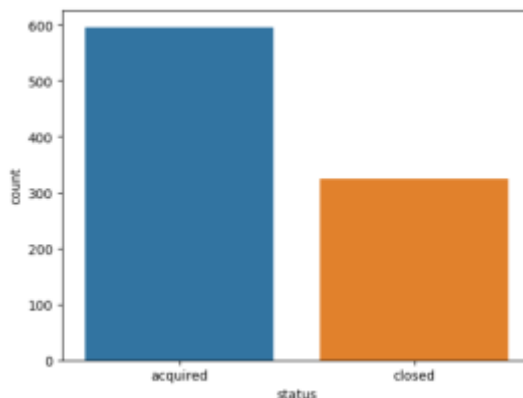
$$\text{Presisi} = (TP + TN) / (TP + FP) \dots\dots\dots(2)$$

3. HASIL DAN PEMBAHASAN

Berikut ini merupakan hasil dan pembahasan yang diperoleh dari penelitian ini.

3.1 Data Yang Digunakan

Pada penelitian ini menggunakan menggunakan *dataset* yang diperoleh dari website *Kaggle* [15] berisi 923 data yang mencakup 597 *startup* yang berhasil diakuisisi dan 326 *startup* yang telah ditutup seperti terlihat pada Gambar 2.



Gambar 2. Jumlah *Startup* Sukses dan *Closed*

3.2 Pre-Processing Data

Pada pre-processing data, proses pertama yang dilakukan adalah pengecekan *missing value*. Pada proses pengecekan *missing value*, ditemukan bahwa terdapat lima variabel atau atribut dengan *missing value*. Contohnya, variabel 'Unnamed:6' memiliki 53,41% *missing value*, 'closed_at' memiliki 63,71% *missing value*, 'age_first_milestone_year' memiliki 16,47% *missing value*, 'last_first_milestone_year' memiliki 16,47% *missing value*, dan 'state_code' memiliki 0,11% *missing value*. Setelah melakukan analisis, ditemukan bahwa variabel 'Unnamed:6', 'state_code', dan 'closed_at' tidak memiliki pengaruh signifikan terhadap kesuksesan *startup*, sehingga ketiga variabel tersebut dihapus. Selanjutnya, untuk 'age_first_milestone_year' dan 'last_first_milestone_year', nilai-nilai yang hilang diisi dengan 0 karena *startup* yang memiliki *missing value* pada kedua variabel tersebut belum memiliki *milestone*.

Setelah proses pengecekan *missing value*, selanjutnya dilakukan proses konversi tipe data untuk memastikan bahwa seluruh variabel objek/string yang digunakan dikonversi menjadi variabel numerik. Variabel 'category_code' merupakan variabel kategorial string/object yang akan diubah menjadi variabel kategorial numerik untuk memudahkan model memahami data dengan baik. Proses selanjutnya adalah dilakukan proses pemilihan variabel/atribut. Berdasarkan hasil analisis *Correlation matrix*, dipilih 36 variabel berikut yang akan digunakan dalam proses pembuatan model prediksi kesuksesan *startup*: 'latitude', 'longitude', 'labels', 'category_code', 'age_first_funding_year', 'age_last_funding_year', 'age_first_milestone_year', 'age_last_milestone_year', 'relationship', 'funding_rounds', 'funding_total_usd', 'milestones', 'is_CA', 'is_NY', 'is_MS', 'is_TX', 'is_otherstate', 'is_software', 'is_web', 'is_mobile', 'is_enterprise', 'is_advertising', 'is_gamesvideo', 'is_ecommerce', 'is_biotech', 'is_consulting', 'is_othercategory', 'has_vc', 'has_angel', 'has_roundA', 'has_roundB', 'has_roundC', 'has_roundD', 'avg_participants', 'is_top500', dan 'status'.

3.3 StandartScaler Transform

StandartScaler merupakan teknik pre-processing data yang berfungsi untuk mengubah variabel/atribut numerik pada *dataset* sehingga memiliki nilai *mean* = 0 dan standart deviasi = 1. Proses yang dilakukan pada tahap ini adalah StandartScaler menghitung *mean* dan standart deviasi dari setiap variabel/atribut dalam *dataset*. Setelah itu dilakukan transformasi data untuk menghasilkan nilai *mean* = 0 dan standart deviasi = 1. Proses tranformasi data ini penting untuk membantu menyeragamkan seluruh variabel/atribut dalam *dataset* yang dapat membantu algortima *Deep Neural Networks* untuk dapat menghasilkan model prediksi yang lebih akurat.

3.4 Pembuatan dan Optimasi Model Prediksi

Untuk membangun dan optimasi model prediksi digunakan dua skenario berikut ini:

1. Pembuatan model prediksi kesuksesan *startup* menggunakan *Deep Neural Networks*
2. Pembuatan model prediksi kesuksesan *startup* menggunakan *Deep Neural Networks* dengan StandartScaler Transform

Pembuatan model dilakukan dengan metode ekperimen untuk mendapatkan akurasi terbaik. Berdasarkan hasil ekperimen yang telah dilakukan, untuk membangun model *Deep Neural Networks* digunakan parameter berikut:

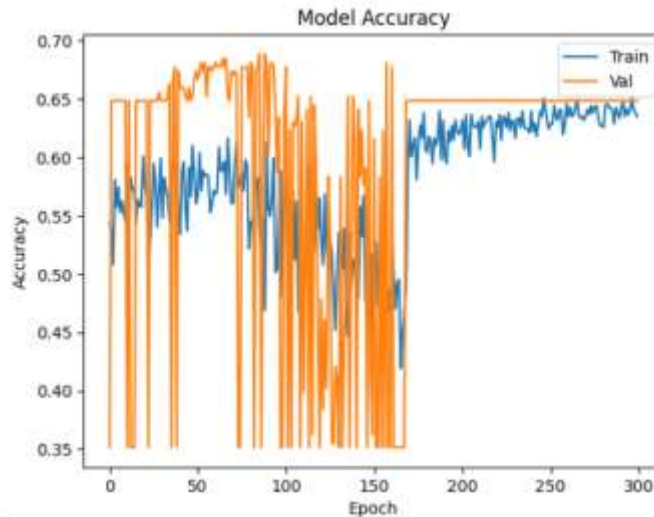
1. *Input layer* dengan jumlah sebanyak 35 sesuai dengan jumlah variabel/atribut yang digunakan.
2. 4 *hidden layer* dengan masing-masing 100 *neuron*
3. 1 *output layer* untuk menghasilkan luaran apakah *startup* akan *acquired* atau *closed*
4. *Relu* digunakan sebagai fungsi aktivasi pada *input* dan *hidden layer* sedangkan *sigmoid* digunakan sebagai fungsi aktivasi pada *output layer*
5. Adam digunakan sebagai fungsi optimasi model
6. Jumlah *epoch* = 300

Pembuatan model dilakukan dengan menggunakan 70% data *training* yang telah diproses sebelumnya. Tabel 2 merupakan hasil pembuatan model dengan 2 skenario yang telah disebutkan di atas.

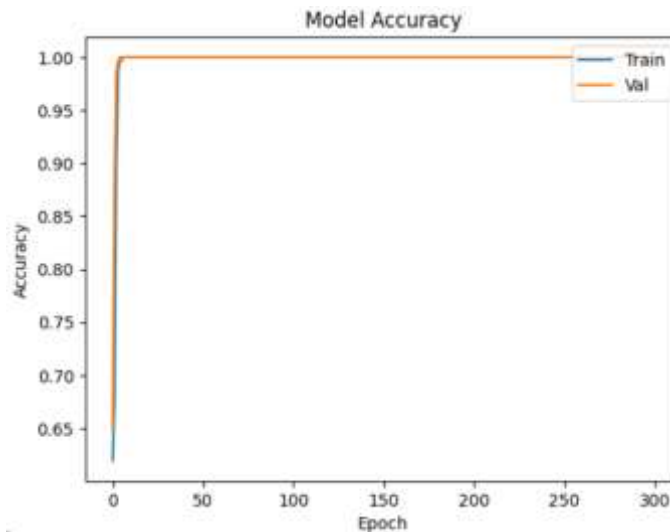
Tabel 2. Hasil Pembuatan Model Prediksi

Metode	Variabel/Atribut	Akurasi (%)	Presisi (%)
DNN	35	64.75	0
DNN + StandartScaler Transform	35	100	100

Gambar 3, merupakan hasil akurasi pembuatan model prediksi dengan *Deep Nerual Networks* dan Gambar 4, merupakan hasil akurasi pembuatan model prediksi dengan *Deep Neural Networks* dengan StandartScaler Transform.



Gambar 3. Hasil Akurasi Model Prediksi Dengan *Deep Neural Networks*



Gambar 4. Hasil Akurasi Model Prediksi Dengan *Deep Neural Networks* dan *StandartScaler Transform*

Berdasarkan Gambar 3 dapat dilihat bahwa model prediksi dengan *Deep Neural Networks* mengalami *underfitting/overfitting* sedangkan Gambar 4, menunjukkan bahwa model prediksi dengan *Deep Neural Networks* dan *StandartScaler Transform* sudah *bestfitting* yang artinya model dapat dilanjutkan ke proses evaluasi dengan data *testing*. Pada tahap *training*, terjadi peningkatan akurasi sebesar 35,25%.

3.5 Evaluasi dan Pembahasan

Model yang telah dibangun selanjutnya di evaluasi menggunakan 30% data *testing*. Tabel 3 merupakan hasil evaluasi model.

Tabel 3. Hasil Evaluasi Model Prediksi

Metode	Variabel/Atribut	Akurasi (%)	Presisi (%)
DNN	35	63.04	0
DNN + StandartScaler Transform	35	100	100

Berdasarkan hasil pengujian yang telah dilakukan pada Tabel 3, diperoleh bahwa akurasi model prediksi kesuksesan *startup* dengan *Deep Neural Networks* dan *StandartScaler Tranform* lebih baik dibandingkan dengan hanya menggunakan *Deep Neural Networks*.

4. SIMPULAN

Eksperimen dan evaluasi yang telah dilaksanakan menunjukkan bahwa optimalisasi model prediksi kesuksesan *startup* melalui *Deep Neural Networks* berhasil ditingkatkan dengan pemanfaatan *StandartScaler Transform*. Selama proses pengembangan model, terjadi peningkatan signifikan pada tingkat akurasi sekitar 35,25%, dan pada tahap pengujian, tercatat peningkatan akurasi sebesar 36,96%. Model prediksi kesuksesan

startup yang diajukan berhasil menghindari masalah *overfitting/underfitting*. Oleh karena itu, model prediksi ini dapat diandalkan dan berpotensi menjadi dasar yang kuat untuk pengembangan aplikasi prediksi kesuksesan *startup*.

5. SARAN

Adapun saran-saran yang dapat diberikan untuk penelitian selanjutnya adalah:

1. Untuk penelitian selanjutnya model ini dapat diterapkan untuk dataset startup pada kategori tertentu saja untuk menguji apakah model prediksi yang dibangun dapat menghasilkan nilai akurasi yang tinggi atau tidak.
2. Model prediksi yang dibangun juga dapat diuji dengan menggunakan dataset yang lain seperti dataset dari *angelist.co* dan lain-lain untuk memprediksi performance sebuah startup.

UCAPAN TERIMA KASIH

Peneliti mengungkapkan apresiasi kepada Universitas Mikroskil atas dukungan yang diberikan selama pelaksanaan penelitian ini.

DAFTAR PUSTAKA

- [1] Y. Birkman, “Hello, Startup: A Programmer's Guide to Building Products, Technologies, and Teams,” O'Reilly Media, Inc., 2015.
- [2] Goldenia, et al., “Implementasi Algoritma Support Vector Machine dalam Memprediksi Keberhasilan Suatu Startup Berdasarkan Status Akuisisi,” Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA), 2021.
- [3] E. Ramalakshmi, and S. R. Kamidi, “Predictions for Startups”, International Journal of Engineering & Technology, vol. 7, 2018.
- [4] A. P. Adhitya, K. Ainiyah, and K. F. H. Holle, “Analisis Perbandingan Algoritma Decision Tree, kNN, dan Naive Bayes untuk Prediksi Kesuksesan Start-up”, JISKA, vol. 6, No. 3, Pp. 178 – 188, 2021.
- [5] M. Bangdiwala, Y. Mehta, S. Agrawal and S. Ghane, “Predicting Success Rate of Startups using Machine Learning Algorithms,” 2022 2nd Asian Conference on Innovation in Technology (ASIANCON), Ravet, India, 2022, pp. 1-6, doi: 10.1109/ASIANCON55314.2022.9908921.
- [6] G. Ross, S. Das, D. Sciro, and H. Raza, “CapitalVX: A machine learning model for startup selection and exit prediction,” The Journal of Finance and Data Science, vol. 7, Pp. 94-114, 2021.
- [7] W. S. Lestari & A. Halim, “Prediksi Kesuksesan Startup Menggunakan Deep Neural Network,” Jurnal SIFO Mikroskil, vol. 23, no. 2, 2022.
- [8] scikit-learn developers. (2023, Des) Compare the effect of different scalers on data with outliers [online]. Available: https://scikit-learn.org/stable/auto_examples/preprocessing/plot_all_scaling.html#standardscaler
- [9] T. D. K. Thara, B. G. P. Sudha and F. Xiong, “Auto-detection of Epileptic Seizure Events Using Deep Neural Network with Different Feature Scaling Techniques,” Pattern Recognition Letters, 2019.

Review Komprehensif: Ekstraksi Fitur GLCM, GLRLM, dan LBP untuk Pendeteksian Korosi

Mizanul Ridho Aohana¹, Ratu Nisful Laily Hidayah², Melki Jonathan Andara³, Nadya Amara⁴, Fitri Bimantoro⁵

^{1,2,3,4,5}Teknik Informatika, Fakultas Teknik, Universitas Mataram

E-mail: ¹mizanul.aohana@gmail.com, ²ratunisful16@gmail.com, ³melkijonathan2@gmail.com,
⁴nadyamaraa15@gmail.com, ⁵bimo@unram.ac.id

Abstrak – Penelitian ini mengeksplorasi solusi untuk mengatasi masalah korosi yang dapat mengancam pertumbuhan industri di Indonesia. Dengan tingginya tingkat investasi dalam sektor-sektor kunci seperti Transportasi, Gudang, Telekomunikasi, dan Pertambangan, risiko korosi menjadi perhatian serius. Korosi dapat menyebabkan kerusakan signifikan pada infrastruktur dan peralatan, menciptakan beban finansial yang besar jika tidak ditangani dengan cepat. Dalam penelitian ini, diajukan metode deteksi dini korosi dengan memanfaatkan ekstraksi fitur GLRLM, GLCM, dan LBP. Melibatkan model pendeteksi berbasis MLP, penelitian ini menitikberatkan pada sektor-sektor investasi utama di Indonesia. Hasil eksperimen menunjukkan keberhasilan metode ini, dengan tingkat akurasi berturut-turut sebesar 74,02%, 79,61%, dan 84,36%. Metode deteksi dini ini memberikan kontribusi besar terhadap upaya pencegahan korosi sejak dini, memungkinkan industri untuk mengidentifikasi potensi kerusakan pada tahap awal. Hal ini dapat mengurangi biaya perbaikan dan penggantian aset yang terkena dampak, memberikan keuntungan finansial bagi perusahaan dan negara secara keseluruhan. Penelitian ini relevan dengan pertumbuhan industri di Indonesia dan dapat menjadi landasan bagi strategi pencegahan korosi yang lebih efektif, khususnya dalam sektor-sektor yang menjadi pilar investasi utama negara. Dengan adanya solusi ini, diharapkan dapat meningkatkan ketahanan dan keberlanjutan infrastruktur industri di Indonesia.

Kata Kunci — Corrosion Detection, GLRLM, GLCM, LBP, MLP

1. PENDAHULUAN

Industri adalah penolong pertumbuhan perekonomian yang signifikan bagi Indonesia, Badan Statistika Pusat mencatat Jumlah perusahaan industri manufaktur skala menengah dan besar pada tahun 2022 mencapai sekitar 29 ribu usaha atau perusahaan [1]. Menurut data yang dikeluarkan oleh Kementerian Investasi atau Badan Koordinasi Penanaman Modal (BKPM) mengungkapkan realisasi penyertaan modal dalam negeri (PMDN) mencapai Rp. 178,2. Kemudian, secara sektor, PMDN mendominasi di sektor Transportasi, Gudang dan Telekomunikasi Rp 24,8 triliun; Pertambangan Rp. 22,3 triliun, Perumahan, Kawasan Industri dan Perkantoran Rp. 16,6 triliun [2]. Dengan tingginya tingkat investasi dalam sektor-sektor kunci seperti transportasi, gudang, telekomunikasi, dan pertambangan, risiko korosi menjadi perhatian serius.

Korosi merupakan fenomena penurunan kondisi logam akibat reaksi elektrokimia dengan lingkungan pada udara terbuka, disebut juga dengan korosi atmosfer. Korosi atmosfer dipengaruhi akibat topografi dan iklim lingkungan seperti temperatur, kelembapan, dan kandungan bahan kimia dalam udara sehingga dapat mempengaruhi laju korosi [3]. Korosi merupakan ancaman serius bagi perindustrian yang berkaitan langsung dengan logam, struktur baja, dan perlengkapan lainnya. Kerusakan akibat korosi dapat mengakibatkan penurunan kinerja, biaya perawatan yang tinggi, serta risiko kecelakaan kerja yang berpotensi fatal.

Berdasarkan permasalahan tersebut diperlukan sebuah solusi untuk mendeteksi korosi, pemetaan area dan level karat sehingga penanganan dapat dilakukan lebih cepat dan efisien. Oleh karena itu, penelitian ini akan membahas solusi deteksi dini korosi secara komprehensif dengan menggunakan metode ekstraksi fitur GLCM, GLRLM, dan LBP. Pendekatan ini memungkinkan analisis gambar korosi secara mendalam, menghasilkan informasi yang beragam mengenai struktur dan pola korosi. Proses pendeteksian korosi melibatkan analisis gambar yang telah mengalami kerusakan akibat korosi, dengan setiap metode ekstraksi fitur memberikan kontribusi informasi unik. Metode GLCM membantu mengukur perubahan tekstur benda akibat karat dengan menganalisis tingkat keabuan yang *co-occur* pada jarak dan sudut tertentu. Peningkatan korelasi pada jarak tertentu bisa menandakan daerah yang terkena karat [4], [5] dan metode GLRLM membantu mendeteksi pola linear pada citra permukaan yang terkena karat. Karat cenderung terbentuk dalam garis atau pola yang mengikuti permukaan benda. Dengan mengukur panjang *run* tingkat keabuan dalam berbagai arah, GLRLM membantu mengidentifikasi daerah dengan karakteristik linear, yang bisa menunjukkan adanya karat [6], [7]. Alasan menggunakan GLCM dan GLRLM terkait dengan pendeteksian korosi pada pipa menggunakan teknik tersebut menghasilkan tingkat akurasi korosi rata-rata sebesar 92,81% [6]. Penelitian ini juga menggunakan MLP, berdasarkan salah satu penelitian membuktikan penggunaan MLP dengan GLCM menghasilkan akurasi sebesar 73,5% pada klasifikasi wayang [5]. Penggunaan LBP digunakan untuk mendeteksi perubahan pola biner piksel

yang diakibatkan oleh karat [8], memungkinkan identifikasi daerah dengan variasi tekstur kompleks, yang menandakan kemungkinan adanya karat.

Pada penelitian identifikasi citra bahan kulit hewan, *Local Binary Pattern* (LBP) dan *Grey Level Run Length Matrix* (GLRLM) digunakan sebagai metode ekstraksi fitur tekstur menggunakan klasifikasi *Artificial Neural Network* (ANN). Penelitian ini menggunakan gambar bahan kulit hewan yang diklasifikasi menjadi lima kategori yaitu kulit sapi, babi, domba, kambing, dan kanguru. Akurasi yang dicapai dengan penerapan metode LBP dan GLRLM untuk mengidentifikasi citra kulit hewan adalah 60% [9].

Penelitian lain mengaplikasikan GLRLM pada pendeteksian COVID dengan menggunakan GLRLM dan *Extreme Learning Machine* (ELM). Tahap ekstraksi fitur menggunakan metode GLRLM memvisualisasikan fitur SRE dan LRE. Terlihat bahwa nilai SRE pada citra normal lebih kecil dibandingkan pada citra COVID-19, sementara nilai LRE pada citra normal lebih tinggi dibandingkan pada citra COVID-19. Hal ini mengindikasikan bahwa citra COVID-19 memiliki tekstur yang lebih kasar dari pada citra normal dan hasil dari *Learning Machine* menghasilkan akurasi sebesar 92%, sensitivitas 89.62%, dan spesifisitas 94.68%. Ini menandakan bahwa model ini memiliki kinerja yang sangat baik dalam mendeteksi COVID-19 melalui citra *X-Ray* [10].

Pada penelitian deteksi kematangan buah melon dengan algoritma *Support Vector Machine* Berbasis Ekstraksi Fitur GLCM. Menggunakan teknik ekstraksi fitur GLCM untuk menganalisis tekstur buah melon dan menerapkan algoritma SVM untuk mengatasi klasifikasi citra digital. Hasil penelitian menunjukkan bahwa teknik ekstraksi fitur GLCM dapat membantu dalam deteksi kematangan buah melon dengan tingkat akurasi sebesar 78,67% [11].

Pada penelitian Implementasi *Gray Level Co-Occurrence Matrix* (GLCM) untuk Klasifikasi Penyakit Daun Padi GLCM digunakan untuk ekstraksi fitur pada citra keabu-abuan (*grayscale*) dari daun padi yang terinfeksi penyakit. akurasi klasifikasi menggunakan metode GLCM dan KNN diperoleh akurasi sebesar 93,3% [12].

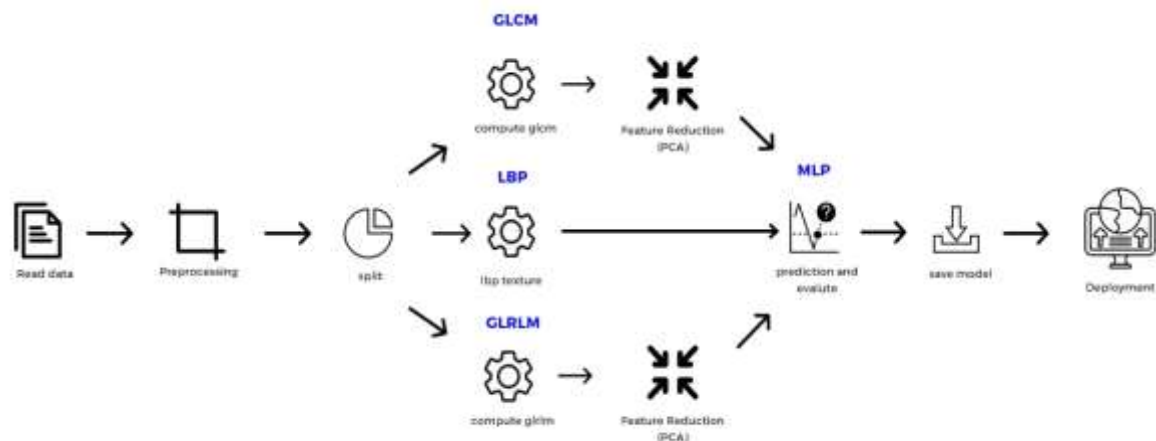
Penelitian lain yaitu Klasifikasi Batik dengan Ekstraksi Fitur Tekstur *Local Binary Pattern* dan Metode *K-Nearest Neighbor*. Metode *local binary pattern* dapat digunakan untuk mengenali pola tekstur pada batik sebagai fitur pada citra. Hasil yang didapat dari penggunaan metode LBP pada pengolahan citra digital batik adalah fitur tekstur yang dapat digunakan untuk membedakan kelas batik dengan akurasi sebesar 65% menggunakan metode *K-Nearest Neighbor* (KNN)[13].

Hasil dari analisis ini diharapkan dapat memberikan pemahaman yang lebih mendalam tentang kondisi korosi pada tingkat mikro, memungkinkan langkah-langkah penanganan yang lebih terarah dan efektif. Untuk mengatasi masalah di industri diperlukan sebuah sistem untuk mendeteksi korosi pada benda. Berdasarkan referensi yang sudah dipaparkan sebelumnya peneliti mencoba untuk melakukan komparatif analisis menggunakan metode GLCM, GLRLM dan LBP.

2. METODE PENELITIAN

2.1 Proposed Work

Dalam penelitian ini, peneliti mengusulkan tahap proses yang dimulai dengan membaca data, selanjutnya dilakukan *preprocessing*. Berdasarkan Gambar 1, *preprocessing* yang dilakukan yaitu *me-resize dataset* untuk mempercepat pemrosesan data. *Dataset* di *resize* dengan ukuran 256 x 256 yang kemudian datanya akan disimpan dalam folder yang berbeda. *Dataset* tersebut kemudian di-*split* menjadi beberapa folder untuk memproses *dataset* dengan berbagai metode. Model yang diperoleh dari pemrosesan tersebut kemudian disimpan dan dapat di *deploy* untuk kepentingan lainnya.



Gambar 1. Alur Penelitian

2.2 Pengumpulan *Dataset*

Pengumpulan data pada penelitian ini dilakukan dengan menerapkan metode *Image Retrieval*. *Image Retrieval* merupakan sistem penting untuk mengambil gambar serupa dengan melakukan pencarian dan penjelajahan di dalam basis data besar. *Image Retrieval* dapat menjadi alat yang dapat diandalkan untuk mengoptimalkan penggunaan akumulasi gambar, dan menemukan metode yang efisien untuk mengambil gambar [14].

2.3 *Dataset Preprocessing*

Preprocessing merupakan suatu teknik untuk meningkatkan kualitas gambar agar memperoleh hasil terbaik dari hasil yang diperoleh. Metode klasik yang umum digunakan meliputi deteksi tepi, peningkatan kontras, pengurangan *noise* gambar, dan normalisasi warna gambar. Sering digunakan untuk memfokuskan dan memperoleh gambar dari objek penelitian, dan metode deviasi standar sering digunakan untuk menampilkan warna gambar dengan lebih jelas [15].

Langkah pertama dalam mengelola *dataset* adalah melakukan pengecekan satu per satu terhadap entri data untuk mengidentifikasi duplikasi. Setelah itu, dilakukan pengecekan terhadap kevalidan label pada *dataset*. Data yang memiliki label yang salah akan segera dihapus atau dipindahkan. Terakhir, proses *resizing* gambar dilakukan untuk melakukan *preprocessing*, sehingga ukuran gambar disamakan untuk memudahkan proses pengolahan selanjutnya [16].

2.4 Metode Ekstraksi

Dalam penelitian ini, digunakan 3 metode ekstraksi antara lain adalah *Gray Level Run Length Matrix* (GLRLM), *Gray Level Co-Occurrence Matrix* (GLCM) dan *Local Binary Pattern* (LBP).

2.2.1 GLRLM

GLRLM merupakan metode ekstraksi tekstur yang memiliki *output* berupa atribut statistik orde tinggi dengan mengestimasi piksel piksel dengan derajat keabuan yang sama pada arah tertentu. GLRLM mampu menggambarkan tekstur yang merupakan ciri penting untuk menunjukkan informasi struktur permukaan suatu citra [17]. Keunggulan metode GLRLM mampu membedakan citra halus dan citra kasar, dimana citra yang memiliki tekstur halus akan memiliki sedikit piksel tetangga yang intensitasnya sama sedangkan yang memiliki tekstur kasar akan memiliki banyak piksel tetangga yang intensitasnya sama [18], [19].

2.2.2 GLCM

Gray Level Co-Occurrence Matrix (GLCM) merupakan suatu metode yang terbentuk berdasarkan citra *Grayscale*. *Matrix* tersebut menghitung frekuensi kemunculan suatu nilai *pixel* horizontal terhadap *pixel* vertikal yang bersebelahan maupun diagonal [7] *Gray Level Co-Occurrence Matrix* (GLCM) merupakan metode yang menguntungkan untuk melakukan analisis tekstur citra, namun parameter dalam menghitung GLCM suatu citra dapat dipilih dari jarak yang luas, yang menghasilkan komputasi dalam jumlah besar [20].

2.2.3 LBP

Metode *Local Binary Pattern* pertama kali diperkenalkan oleh Ojala et al. Menurut Ojala et al. *Local Binary Pattern* merupakan metode yang digunakan sebagai ukuran tekstur *grayscale* yang terbukti efektif dan invarian terhadap pencahayaan yang berbeda. LBP dimanfaatkan untuk deskripsi tekstur dan didukung oleh komposisi pola mikro yang dapat dijelaskan oleh sebuah operator [21].

Semua LBP adalah deskriptor tekstur yang dapat juga digunakan untuk mewakili objek, karena gambar objek dapat dilihat sebagai sebuah komposisi *micro texture pattern* yaitu suatu operator non parametrik yang menggambarkan tata ruang lokal citra. Metode LBP mengubah tekstur dari suatu citra *input* menjadi nilai biner, dan nilai tersebut mewakili bagian dari piksel piksel suatu objek yang membentuk sebuah lingkaran dan memiliki pusat sebagai acuan terhadap nilai-nilai tersebut [17].

2.5 *Principal Component Analysis*

Metode *Principal Component Analysis* (PCA) merupakan sebuah metode yang digunakan untuk menyederhanakan suatu data. Metode ini ditemukan oleh Karl Pearson pada tahun 1901 dan digunakan dalam biologi. Sejak teori ini ditemukan kembali oleh Karhunen pada tahun 1947 dan dikembangkan lebih lanjut oleh Loewe pada tahun 1963, maka teori ini disebut juga transformasi Karhunen-Loewe dalam bidang ilmu telekomunikasi. Analisis komponen utama (PCA) adalah teknik statistik yang mengubah sebagian besar variabel yang awalnya berkorelasi menjadi sekumpulan variabel baru yang lebih kecil dan saling independen. Oleh karena itu, analisis komponen utama (PCA) membantu mereduksi data dan memfasilitasi interpretasi data [22]

PCA adalah kombinasi linier dari variabel pertama. Secara geometris, kombinasi linier ini merupakan sistem koordinat baru yang diperoleh dengan memutar sistem koordinat asli sebanyak kali. Metode PCA sangat berguna ketika data Anda yang ada memiliki jumlah variabel yang banyak dan korelasi antar variabel. Perhitungan analisis komponen utama didasarkan pada perhitungan nilai *eigen* dan vektor *eigen* yang mewakili

sebaran data dari *dataset*. PCA digunakan untuk menjelaskan struktur matriks varians-kovarians dari suatu set variabel melalui kombinasi linier dari variabel-variabel tersebut. Secara umum PCA dapat berguna untuk seleksi fitur dan interpretasi variabel-variabel [23].

2.6 Algoritma *Machine Learning*

MLP adalah salah satu arsitektur ANN paling populer. Hal ini diyakini sebagai yang pertama memperkenalkan MLP menggunakan algoritma pelatihan *backpropagation*. Jaringan *multilayer* biasanya menggunakan fungsi aktivasi *sigmoid* di lapisan tersembunyi. Fungsi-fungsi ini terkadang disebut fungsi "*squashing*" karena fungsi ini memampatkan rentang masukan tak terbatas menjadi rentang keluaran terbatas. Fungsi *sigmoid* dicirikan oleh kemiringan yang mendekati 0 seiring dengan meningkatnya masukan. Hal ini menyebabkan masalah saat menggunakan penurunan paling curam untuk melatih jaringan *multilayer* dengan fungsi *sigmoid*. Hal ini disebabkan karena ukuran gradien yang sangat kecil dapat mengakibatkan perubahan bobot yang kecil dan bias yang jauh dari optimal [24].

Arsitektur MLP-ANN terdiri dari sekumpulan *node* yang saling berhubungan yang terdiri dari tiga jenis lapisan: *input*, lapisan tersembunyi, dan *output*. Berikut penjelasan masing-masing *layer*:

a. *Input Layer*

Lapisan masukan yang menerima nilai masukan dari setiap record dalam data Anda. Jumlah *node* masukan sesuai dengan jumlah variabel prediktor.

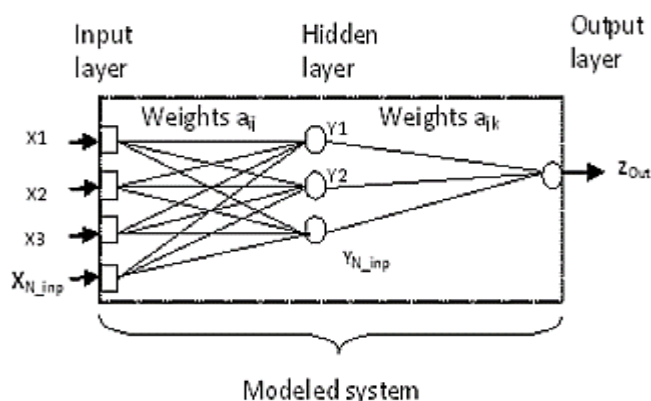
b. *Hidden Layer*

Lapisan tersembunyi mengubah nilai masukan dalam jaringan. Setiap *node* pada lapisan tersembunyi terhubung ke *node* lapisan tersembunyi sebelumnya atau dari *node* lapisan masukan, dan terhubung ke *node* lapisan tersembunyi berikutnya atau *node* lapisan keluaran. Jumlah lapisan tersembunyi bisa mencapai.

c. *Output Layer*

Garis yang terhubung ke lapisan keluaran berasal dari lapisan tersembunyi atau lapisan masukan dan mengembalikan nilai keluaran yang sesuai dengan variabel prediktor. *Output* dari lapisan *output* biasanya berupa nilai *floating point* antara 0 dan 1 [25].

Ilustrasi berbagai komponen MLP dengan lapisan tersembunyi. Lapisan masukan hanya mengirimkan nilai masukan x_i ke unit-unit yang ada pada lapisan tersembunyi dan tidak melakukan operasi apapun terhadap sinyal masukan. Lapisan tersembunyi menerima sinyal dari *node* lapisan masukan dan mengubahnya menjadi sinyal, yang dikirim ke semua *node* keluaran dan diubah menjadi keluaran. Bobot koneksi dari masukan ke *node* lapisan tersembunyi dan dari lapisan tersembunyi ke lapisan keluaran disesuaikan menggunakan algoritma penurunan paling curam. Algoritma ini digunakan untuk menyelesaikan permasalahan *nonlinier* [25].



Gambar 2. Arsitektur *Multi-Layer Perceptron*

Algoritma *machine learning* yang digunakan untuk melatih fitur korosi dalam penelitian ini adalah *Multi-Layer Perceptron*. *Multi-Layer Perceptron* atau MLP merupakan salah satu model *supervised learning* dalam *Artificial Neural Network* (ANN) yang memiliki jumlah neuron yang dapat disesuaikan dengan kebutuhan. Alasan dari pemilihan algoritma ini adalah kemampuan MLP dalam menangani tugas prediksi kompleks dan dapat belajar dari pola yang rumit dalam data. Dalam konteks ini, ekstraksi fitur menggunakan metode GLCM (*Gray Level Co-occurrence Matrix*), GLRLM (*Gray Level Run Length Matrix*), dan LBP (*Local Binary Pattern*) memberikan representasi yang kaya dan informatif tentang struktur tekstur gambar korosi.

MLP memiliki keunggulan dalam mengekstraksi dan memodelkan pola yang kompleks dari data bertipe gambar, sehingga dapat secara efektif memanfaatkan informasi yang dihasilkan oleh metode ekstraksi fitur seperti GLRLM, GLCM, dan LBP. Dengan kemampuan ini, MLP dapat memahami hubungan antara fitur-fitur

ekstraksi dan tingkat korosi yang terkait, sehingga memungkinkan untuk melakukan prediksi korosi dengan tingkat akurasi yang tinggi.

Dengan menggunakan MLP untuk melatih model prediksi korosi, penelitian ini dapat mengoptimalkan proses pembelajaran mesin untuk menemukan pola-pola yang kompleks dan tersembunyi dalam data ekstraksi fitur. Hasilnya diharapkan dapat memberikan prediksi yang akurat terkait tingkat korosi berdasarkan fitur-fitur yang diekstraksi menggunakan GLCM, GLRLM, dan LBP.

2.7 Metrik Evaluasi

Untuk mengukur keakuratan dari model yang diperoleh, digunakan *matrix* evaluasi yaitu *confusion matrix*, *Confusion Matrix* adalah sebuah tabel yang digunakan untuk mengevaluasi performa dari suatu model klasifikasi pada *machine learning*. *Confusion matrix* memuat empat nilai evaluasi utama, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN).

Tabel 1. *Confusion Matrix*

	Prediction	
	<i>Positive</i>	<i>Negative</i>
<i>Positive</i>	TP	TN
<i>Negative</i>	FP	FN

Dengan menggunakan nilai-nilai evaluasi dari Tabel 1, dapat dihitung matriks evaluasi berupa akurasi (*accuracy*), presisi (*precision*), dan *recall*. Berikut merupakan persamaan yang digunakan untuk menghitung nilai evaluasi.

Accuracy dihitung dengan membagi jumlah *True Positive* (TP) dan *True Negative* (TN) oleh total nilai keseluruhan. dinyatakan dalam persamaan 1.

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \dots\dots\dots(1)$$

Recall akan mengukur tingkat positif aktual dari hasil prediksi, untuk menghitung nilai tersebut, dinyatakan dalam persamaan 2.

$$Recall = \frac{TP}{TP+FN} \dots\dots\dots(2)$$

Precision memberikan nilai tingkat positif yang sebenarnya dari hasil prediksi, persamaannya dinyatakan dalam persamaan 3.

$$Precision = \frac{TP}{TP+FP} \dots\dots\dots(3)$$

ROC atau *Receiver Operating Characteristic* merupakan sebuah grafik yang mengilustrasikan kemampuan diagnostik dari sistem pengklasifikasi biner karena variasi ambang diskriminasi. dinyatakan dalam persamaan 4.

$$ROC = \frac{P(x|positive)}{P(x|negative)} \dots\dots\dots(4)$$

Dalam penelitian ini, penulis akan menggunakan akurasi, presisi, *recall* dan ROC sebagai metrik evaluasi untuk semua algoritma yang akan diuji [26]

2.8 Area Under Curve (AUC)

Area Under Curve (AUC) adalah ukuran dari luas area di bawah suatu kurva, dan nilainya selalu berada di antara 0 dan 1. AUC dihitung dengan menggabungkan luas trapesium yang terbentuk oleh titik-titik yang mewakili sensitivitas dan spesifisitas. Metrik ini mengevaluasi kinerja diskriminatif dengan memperkirakan probabilitas model dalam mengklasifikasikan sampel yang dipilih secara acak dari populasi positif atau negatif.

AUC berfungsi sebagai indikator akurasi model prediksi, dan perhitungannya melibatkan penilaian seberapa baik model dapat membedakan antara kelas. Nilai AUC berkisar antara 0 dan 1, di mana nilai yang lebih tinggi menunjukkan klasifikasi yang lebih kuat. AUC pada dasarnya merupakan ukuran kemampuan model untuk menyusun peringkat instansi positif lebih tinggi daripada instansi negatif [27].

3. HASIL DAN PEMBAHASAN

3.1 Hasil dan Diskusi

Berdasarkan metodologi yang telah dijelaskan sebelumnya, *dataset* yang digunakan dalam penelitian ini diperoleh melalui proses pengambilan data dari *Google Image* menggunakan kata kunci "karat". *Dataset* ini terbagi menjadi dua kelas, yakni kelas korosi dan kelas non-korosi, dengan jumlah gambar pada masing-masing label secara berturut-turut sebanyak 990 data untuk kelas korosi dan 829 data untuk kelas non-korosi. Proses pengumpulan *dataset* melalui metode *scraping* menghasilkan gambar-gambar karat yang berasal dari berbagai objek dengan bentuk yang beragam. Contoh *dataset* dapat dilihat pada Gambar 3 dan Gambar 4



Gambar 3 *Dataset* Non-Korosi



Gambar 4 *Dataset* Korosi

Gambar 3 dan 4 merupakan gambar *dataset* yang digunakan dalam mekanisme percobaan yang akan dilakukan. Tingginya variasi dalam *dataset* ini diharapkan memberikan keuntungan dalam proses pelatihan model, karena model akan terlatih dengan berbagai bentuk gambar atau objek. Keberagaman ini diharapkan dapat membantu model untuk memahami karakteristik karat dengan lebih baik, termasuk aspek-aspek seperti warna, tekstur, dan bentuk dari gambar-gambar tersebut.

3.2 Analisis Komparatif

Dalam mekanisme *training* model pada penelitian ini, akan digunakan algoritma *Multi-Layer Perceptron*. Tabel 2. menunjukkan arsitektur yang digunakan pada masing-masing fitur. Standarisasi arsitektur ini dilakukan untuk mendapatkan hasil analisis yang lebih komprehensif.

Tabel 2. Arsitektur Model MLP

Fitur	<i>Learning Rate</i>	<i>Activation</i>	<i>Epoch</i>	<i>Hidden Layer</i>	<i>Solver</i>
GLRLM	<i>constant</i>	relu	200	200	adam
GLCM	<i>constant</i>	relu	200	300,	adam
LBP	<i>constant</i>	relu	200	100,50	adam

Learning rate constant pada Tabel 2 menunjukkan laju pembelajaran dari model akan tetap sama sepanjang pelatihan. Fungsi aktivasi pada lapisan *hidden layer* adalah relu atau *rectified linear*, fungsi ini yang akan mengkonversi setiap nilai *input* negatif menjadi nol dan membiarkan nilai positif tidak berubah. *Hidden layer* yang digunakan dalam setiap mekanisme *training* sudah disesuaikan dengan hasil performa yang paling optimal, sehingga digunakan jumlah *hidden layer* tersebut sebagai acuan.

Berdasarkan arsitektur yang sudah dibangun tersebut, maka dilakukan berbagai percobaan untuk mendapatkan hasil akurasi tertinggi dari ketiga metode GRLRM, GLCM, dan LBP. Proses ini melibatkan kombinasi berbagai fitur, pengulangan, evaluasi ulang, dan penyesuaian pada fitur-fitur yang digunakan, Metode GLCM menggunakan 24 fitur awal yang kemudian mengalami proses reduksi dengan PCA, menghasilkan 15 komponen. Sementara GLRLM menggunakan 44 fitur awal yang juga mengalami proses reduksi dengan PCA dan menghasilkan 15 komponen. Sedangkan untuk Metode LBP (*Local Binary Patterns*) tidak memiliki fitur terpisah seperti layaknya GLRLM dan GLCM. LBP sendiri sebenarnya adalah sebuah metode untuk mengekstraksi fitur dari citra kemudian merepresentasikannya ke dalam sebuah vektor. Teknik ini bekerja dengan cara membandingkan intensitas piksel dengan tetangganya secara lokal dan menghasilkan kode biner yang merepresentasikan pola tekstur di area tersebut. Dari percobaan yang sudah dilakukan, hasil percobaan dapat dilihat dalam Tabel 3.

Tabel 3. Fitur Hasil PCA

Percobaan	Daftar Fitur	Acc	P	R	AUC
GLRLM	SRE0,SRE45, SRE90,SRE135, LRE0, LRE45,LRE90, LRE135, GLN0,GLN45, GLN90, GLN135, RLN0,RLN45, RLN90, RLN135, RP0, RP45, RP90, RP135, LGLRE0, LGLRE45, LGLRE90, LGLRE135, HGL0, HGL45, HGL90, HGL135, SRLGLE0, SRLGLE45, SRLGLE90, SRLGLE135, SRHGLE0, SRHGLE45, SRHGLE90, SRHGLE135, LRLGLE0, LRLGLE45, LRLGLE90, LRLGLE135, LRHGLE0, LRHGLE45, LRHGLE90, LRHGLE135	74.02%	72.81%	82.29%	83.54%
GLCM	Contrast0, Contrast45 Contrast90, Contrast135, Dissimilarity0, Dissimilarity45, Dissimilarity90, Dissimilarity135, Homogeneity0, Homogeneity45, Homogeneity90, Homogeneity135, Energy0, Energy45 Energy90, Energy135, Correlation0, Correlation45, Correlation90, Correlation135, ASM0, ASM45, ASM90, ASM135	77.09%	78.35%	79.17%	85.76%
GLRLM with PCA	SRLGLE90, SRLGLE0, SRLGLE45, SRLGLE135, SRE45, SRE135, RLN45, LRE45, RLN135, LRE135, GLN45, RP45, HGL135, SRHGLE135, LRE0	74.02%	73.24%	81.25%	82.50%
GLCM with PCA	Correlation135, Contrast135, Homogeneity135, Dissimilarity135, Energy135, Correlation90, Contrast90, Contrast0, ASM135, Energy0, ASM0, Energy90, Homogeneity0, Dissimilarity90, Correlation0	79.61%	79.60%	83.33%	86.96%
LBP	-	84.36%	83.01%	89.06%	91.22%

Dari Tabel 3. hasil terbaik diperoleh menggunakan LBP dengan akurasi sebesar 84,36% dan AUC sebesar 91.22%. Disusul oleh GLCM with PCA dengan akurasi sebesar 79.61% dan GLCM dengan akurasi sebesar 77.09%. Hasil ini menunjukkan bahwa konsep LBP dalam mengenali pola hubungan biner lokal antara piksel dan tetangganya mampu menggambarkan perubahan piksel akibat korosi dengan lebih baik. Hal ini disebabkan karena LBP dirancang secara mikro untuk mengenali tekstur yang ada pada setiap piksel citra.

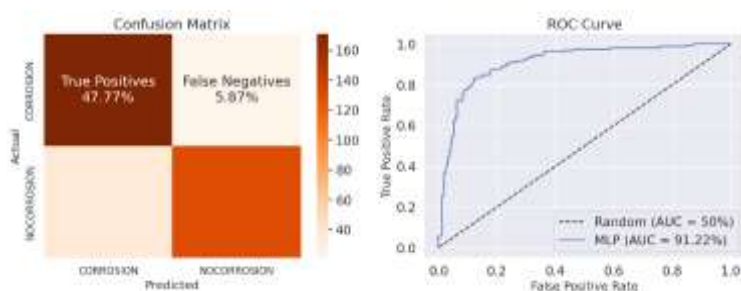
Sedangkan, jika melihat perbandingan antara algoritma GLCM dan GLRLM, performa terbaik diberikan oleh GLCM. Perbedaan ini kemungkinan besar disebabkan oleh kemampuan GLCM untuk menangkap hubungan spasial kompleks antara intensitas piksel yang mungkin lebih relevan dalam membedakan area korosi daripada informasi yang diukur oleh GLRLM tentang panjang dan jumlah *run* piksel yang mengalami korosi.

Selain itu, hal yang tidak kalah menarik ditunjukkan oleh penggunaan PCA dalam percobaan algoritma GLCM dan GLRLM. Dari analisis data yang terdapat dalam tabel 3, terlihat bahwa penerapan *Principal Component Analysis* (PCA) mampu meningkatkan tingkat akurasi GLCM dan GLRLM. Meskipun demikian, dapat dicatat bahwa metode *Local Binary Pattern* (LBP) tetap menjadi pilihan unggul dengan tingkat akurasi tertinggi untuk pendeteksian karat dibandingkan dengan penggunaan PCA pada metode *Gray Level Co-occurrence Matrix* (GLCM) atau *Gray Level Run Length Matrix* (GLRLM).

Berikut adalah grafik hasil *training* dengan menggunakan metode *Local Binary Pattern*, detail grafik dapat dilihat pada Gambar 5 dan Gambar 6.



Gambar 4. Grafik Loss



Gambar 5. Confusion Matrix dan ROC Curve

Gambar 4. menunjukkan bahwa loss yang didapatkan dalam pelatihan ini mengalami penurunan yang sangat halus dari 0.7 ke nilai loss 0.15, akurasi *training* yang diperoleh sebesar 94% dan 84% untuk test *accuracy* menggunakan LBP. Sedangkan Gambar 5. menunjukkan bahwa hasil *confusion matrix* dari validasi yang sudah dilakukan, kemudian terdapat kurva ROC yang menggambarkan hubungan antara *True Positive Rate* (*Sensitivity*) dan *False Positive Rate* ($1 - \textit{Specificity}$). Tren yang ditunjukkan oleh kurva ROC ini sangat baik, hal ini direpresentasikan dari hasil AUC yang mendapatkan skor sebesar 91.22%.

4. SIMPULAN

Berdasarkan hasil penelitian, dicapai tingkat akurasi tertinggi sebesar 94% untuk *training* dan 84% untuk test menggunakan metode *Local Binary Pattern* (LBP), diikuti oleh *Gray Level Co-occurrence Matrix* (GLCM) dengan akurasi *training* sebesar 83% dan test sebesar 79%. Sementara itu, metode *Gray Level Run Length Matrix* (GLRLM) menunjukkan akurasi terendah, dengan *training accuracy* 79% dan test *accuracy* 74%. Perbedaan hasil ini dapat diatribusikan pada karakteristik masing-masing metode ekstraksi fitur. Dalam konteks deteksi karat dengan variasi objek yang beragam dalam gambar, LBP terbukti menjadi metode ekstraksi fitur yang paling efektif untuk mengenali korosi. Kesimpulan ini memberikan wawasan penting dalam pemilihan metode ekstraksi fitur yang sesuai untuk aplikasi deteksi korosi, dengan implikasi pada peningkatan kinerja model dalam memahami pola tekstur dan ciri-ciri kritis dalam citra karat.

5. SARAN

Untuk meningkatkan kualitas dan efektivitas sistem deteksi korosi, disarankan untuk melibatkan eksplorasi lebih lanjut dengan mengimplementasikan algoritma *training* alternatif seperti *Learning Vector Quantization* (LVQ), *Support Vector Machine* (SVM) ataupun dengan algoritma *deep learning*. Selain itu, dapat dilakukan proses *hypertuning* parameter pada model dan melakukan segmentasi dalam proses *preprocessing* agar dapat memberikan peningkatan signifikan dalam akurasi dan efisiensi. Dalam konteks *dataset*, dianjurkan untuk melakukan pembenahan pada kualitas data dengan melakukan standarisasi *dataset* dan menangani potensi anomali serta menambah jumlah data untuk mencakup lebih banyak variasi citra korosi. Penggabungan saran-saran ini diharapkan dapat memperkuat kinerja sistem deteksi korosi dan meningkatkan kemampuannya untuk mengenali dengan akurat pola-pola tekstur yang berkaitan dengan korosi.

DAFTAR PUSTAKA

- [1] Badan Pusat Statistik, Direktori Industri Manufaktur Indonesia 2022. Jakarta: Badan Pusat Statistik, 2022. Accessed: Dec. 26, 2023. [Online]. Available: <https://www.bps.go.id/id/publication/2022/09/30/cbc730b4a2e4ebc36749998c/direktori-industri-manufaktur-indonesia-2022.html>
- [2] CNBC Indonesia, "BKPM: Investasi Dalam Negeri Capai Rp178 T di Kuartal III," 20 October 2023.
- [3] A. M. S. C. A. S. A. R. N. I. T. S. F. S. H. Affandi, "Analisa Korosi Atmosferik Baja Karbon Rendah di Kecamatan Medan Belawan," vol. 14, Nov. 2020.
- [4] "Wayang Image Classification Using MLP Method and GLCM Feature Extraction," *Journal of Computer Science, Information Technology and Telecommunication Engineering*, Sep. 2020, doi: 10.30596/jcositte.v1i2.5131.
- [5] N. D. Hoang, "Image Processing-Based Pitting Corrosion Detection Using Metaheuristic Optimized Multilevel Image Thresholding and Machine-Learning Approaches," *Math Probl Eng*, vol. 2020, 2020, doi: 10.1155/2020/6765274.
- [6] J. P. Sari, A. Erlansari, and E. P. Purwandari, "Identifikasi Citra Digital Kura-Kura Sumatera Dengan Perbandingan Ekstraksi Fitur Glcm dan Glrlm Berbasis Web," 2021. [Online]. Available: www.ejournal.unib.ac.id/index.php/pseudocode

-
- [7] Marwan Noor Fauzy, “Perbandingan Tekstur Citra Dalam Klasifikasi Motif Batik Menggunakan Metode Gray Level Co-Occurrence Matrix (GLCM) dan Gray Level Run Length Matrix (GLRLM),” Program PascaSarjana Universitas Amikom Yogyakarta, Yogyakarta, 2020.
- [8] S. H. Khaleefah, S. A. Mostafa, A. Mustapha, and M. F. Nasrudin, “Review of local binary pattern operators in image feature extraction,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 19, no. 1, pp. 23–31, 2020, doi: 10.11591/ijeecs.v19.i1.pp23-31.
- [9] M. Hafizd, M. M. Santoni, and A. Muliawati, *Level Run Length Matrix Untuk Identifikasi Citra Bahan Kulit Hewan*.
- [10] Y. S. HARIYANI, S. HADIYOSO, and T. S. SIADARI, “Deteksi Penyakit Covid-19 Berdasarkan Citra X-Ray Menggunakan Deep Residual Network,” *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*, vol. 8, no. 2, p. 443, May 2020, doi: 10.26760/elkomika.v8i2.443.
- [11] R. A. Saputra, D. Puspitasari, and T. Baidawi, “Deteksi Kematangan Buah Melon dengan Algoritma Support Vector Machine Berbasis Ekstraksi Fitur GLCM.” [Online]. Available: <http://ejournal.bsi.ac.id/ejurnal/index.php/infortech200>
- [12] A. A. Huda, B. Setiaji, and F. R. Hidayat, “Implementasi Gray Level Co- Occurrence Matrix (GLCM) Untuk Klasifikasi Penyakit Daun,” 2022. [Online]. Available: www.ejournal.unib.ac.id/index.php/pseudocode
- [13] M. T. Kanugroho, M. A. Rahman, and R. C. Wihandika, “Klasifikasi Batik dengan Ekstraksi Fitur Tekstur Local Binary Pattern dan Metode K-Nearest Neighbor,” 2022. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [14] B. I. Khaleel, “Color Image Retrieval Based On Fuzzy Neural Networks and Swarm Intelligence,” *IJUM Engineering Journal*, vol. 23, no. 1, pp. 116–128, 2022, doi: 10.31436/IJUM.EJ.V23I1.1802.
- [15] Z. A. A. R. Z. G. D. Ucta Pradema Sanjaya, “Optimasi Convolutional Neural Network dengan Standard Deviasi untuk Klasifikasi Pneumonia pada Citra X-rays Paru,” *Fakultas Sains dan Teknologi Universitas Nahdlatul Ulama Sunan Giri*, vol. 7, no. 3, Jul. 2023.
- [16] A. Brijith, “Data Preprocessing for Machine Learning.” [Online]. Available: <https://www.researchgate.net/publication/375003512>
- [16] O. Pribadi, “Bulletin Of Computer Science Research Aplikasi Pembeda Daging Sapi dan Babi dengan Metode Color Moment dan Local Binary Pattern Histogram,” *Media Online*, vol. 3, no. 5, pp. 336–342, 2023, doi: 10.47065/bulletincsr.v3i5.260.
- [17] Idamatul Bashiroh, “Deteksi Osteoporosis Melalui Analisis Tekstur Citra Tulang Manusia Pada Wanita Pasca Menopause dengan Menggunakan Metode Ekstraksi Fitur Gray Level Run Length Matrix (GLRLM) dan Klasifikasi Adaptive Neuro Fuzzy Inference System (Anfis),” *Universitas Islam Negeri Maulana Malik Ibrahim Malang, Malang*, 2023.
- [18] A. Puteri Bilqis Salsabila et al., “Identifikasi Citra Jenis Bunga menggunakan Algoritma KNN dengan Ekstraksi Warna HSV dan Tekstur GLCM,” *Technomedia Journal (TMJ)*, doi: 10.33050/tmj.v6i1.
- [19] L. B. Coelho, D. Zhang, Y. Van Ingelgem, D. Steckelmacher, A. Nowé, and H. Terry, “Reviewing machine learning of corrosion prediction in a data-oriented perspective,” *npj Materials Degradation*, vol. 6, no. 1. Nature Publishing Group, Dec. 01, 2022. doi: 10.1038/s41529-022-00218-4.
- [20] D. Retnoningrum, A. W. Widodo, and M. A. Rahman, “Ekstraksi Ciri Pada Telapak Tangan Dengan Metode Local Binary Pattern (LBP),” 2019. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [21] A. Z. W. S. D. S. M. S. Noya Van Delsen, “Penerapan Metode Analisis Komponen Utama untuk Mereduksi Faktor-faktor Inflasi di Kota Ambon,” vol. 11, no. 2, Dec. 2017.
- [22] M. Z. Nasution, “Penerapan Principal Component Analysis (PCA) dalam Penentuan Faktor Dominan yang Mempengaruhi Prestasi Belajar Siswa (Studi Kasus: SMK Raksana 2 Medan),” *Jurnal Teknologi Informasi*, vol. 3, no. 1, 2019.
- [23] P. Hosseinzadeh Talaee, “Multilayer perceptron with different training algorithms for streamflow forecasting,” *Neural Comput Appl*, vol. 24, no. 3–4, pp. 695–703, Mar. 2014, doi: 10.1007/s00521-012-1287-5.
- [24] A. Setiadi Jurusan Manajemen Informatika Akademi Manajemen Informatika dan Komputer Bina Sarana Informatika Karawang Jl Banten No, “PENERAPAN ALGORITMA MULTILAYER PERCEPTRON UNTUK DETEKSI DINI PENYAKIT DIABETES.” [Online]. Available: <http://archive.ics.uci.edu/ml/>.
- [25] M. Hasnain, M. F. Pasha, I. Ghani, M. Imran, M. Y. Alzahrani, and R. Budiarto, “Evaluating Trust Prediction and Confusion Matrix Measures for Web Services Ranking,” *IEEE Access*, vol. 8, pp. 90847–90861, 2020, doi: 10.1109/ACCESS.2020.2994222.
- [26] D. A. Setiawan, H. Fitriyah, and W. Kurniawan, “Sistem Klasifikasi Jenis Karat Menggunakan Metode Decision Tree Berbasis Raspberry Pi,” 2019. [Online]. Available: <http://j-ptiik.ub.ac.id>
-

Deteksi Parasit Malaria Menggunakan Metode *Gray Level Co-Occurance Matrix* (GLCM)

Ahmad Arsyad Surgi Mukti¹, Ghina Briliana Fatin Octariana², Krisna Dian Sukmana³, Fitri Bimantoro⁴

^{1,2,4}Teknik Informatika, Fakultas Teknik, Universitas Mataram

³Teknik Informatika, Fakultas Teknik, Universitas Kuningan

E-mail: *¹ahmadarsyad@mhs.unram.ac.id, ²f1d021093@student.unram.ac.id,

³krisnadian12@gmail.com, ⁴bimo@unram.ac.aid

Abstrak – Malaria adalah salah satu penyakit menular paling mematikan di dunia. Deteksi dini malaria sangat penting untuk mencegah penyebaran penyakit ini. Tradisionalnya, deteksi malaria dilakukan dengan pemeriksaan mikroskopis darah. Namun, metode ini membutuhkan waktu dan usaha yang relatif lama. Dalam penelitian ini, deteksi parasit malaria dilakukan menggunakan metode Gray Level Co-occurrence Matrix (GLCM). Metode GLCM digunakan untuk mengekstraksi fitur tekstur dari gambar sel darah merah yang terinfeksi malaria. Fitur tekstur ini kemudian digunakan untuk melatih model klasifikasi. Hasil penelitian menunjukkan bahwa metode GLCM dapat digunakan untuk mendeteksi parasit malaria dengan akurasi 89%. Studi ini menyarankan bahwa metode GLCM memiliki potensi sebagai metode deteksi malaria yang lebih cepat dan akurat.

Kata Kunci — Deteksi, malaria, sel darah merah, tekstur

1. PENDAHULUAN

Kesehatan menjadi aspek krusial dalam kehidupan manusia yang memiliki nilai tinggi. Namun, terdapat perilaku dari pihak-pihak terkait yang mengabaikan aspek kesehatan mereka dengan tidak mengadopsi gaya hidup sehat. Hal ini dapat mengakibatkan penyakit yang mereka derita terlambat didiagnosis, sehingga mencapai tingkat kronis yang sulit untuk disembuhkan. Padahal, umumnya setiap penyakit menunjukkan gejala awal yang dapat diidentifikasi oleh pihak terkait. Sebagai contoh, dalam kasus gigitan nyamuk, perilaku yang kurang peduli terhadap gaya hidup sehat, seperti pembuangan sampah sembarangan di saluran air rumah dan kurangnya kebersihan saluran tersebut, dapat mengakibatkan kurang lancarnya drainase. Faktor ini memungkinkan nyamuk untuk membentuk koloni dan berkembang biak dengan cepat. Gejala awal yang umumnya muncul akibat gigitan nyamuk, seperti demam berdarah, malaria, atau chikungunya, seringkali serupa dan sulit untuk dibedakan, termasuk pusing, nyeri, demam tinggi secara tiba-tiba, penurunan nafsu makan, kelemahan, serta kelelahan [1].

Malaria adalah penyakit menular yang disebabkan oleh parasit Plasmodium. Penyakit ini menyebar lewat gigitan nyamuk yang terinfeksi parasit. Infeksi malaria dapat terjadi hanya dengan satu gigitan nyamuk saja. Penyakit ini tidak menular secara langsung dari satu individu ke individu lainnya [2]. Malaria disebabkan oleh protozoa dengan jenis plasmodium. Terdapat 4 jenis plasmodium yang menyebabkan malaria pada manusia yaitu: *Plasmodium Falciparum*, *Plasmodium Vivax*, *Plasmodium Malariae*, dan *Plasmodium Ovale*. *Plasmodium* dengan jenis *Falciparum* dapat sangat berbahaya karena paling banyak menyebabkan kematian [3].

Berdasarkan data dari Kementerian Kesehatan Republik Indonesia tahun 2021 mengatakan bahwa pada tahun 2010 terdapat 465.700 kasus positif malaria, sedangkan pada tahun 2020 kasus positif malaria menurun menjadi 235.700. Pencegahan penyakit malaria dapat dilakukan dengan cara menggunakan obat nyamuk, pakaian yang proper, dan memasang jaring nyamuk [4].

Diagnosis definitif malaria tetap melibatkan deteksi parasit malaria dalam darah. Tabel 1 memberikan "daftar keinginan" yang menentukan apa yang mungkin dianggap sebagai fitur ideal dari suatu metode diagnostik yang umumnya berlaku. Idealnya, tes untuk malaria dapat mendeteksi sebanyak 1 parasit darah, dapat diterapkan untuk semua spesies malaria dan dapat mengidentifikasi spesies yang ada, tidak memberikan hasil negatif palsu dan sangat sedikit positif palsu, memberikan hasil definitif dalam waktu 20 menit atau kurang, memberikan ukuran kuantitatif dari tingkat infeksi saat ini, dan tidak dipengaruhi oleh infeksi malaria sebelumnya pada subjek. Selanjutnya, karena sebagian besar kasus malaria terjadi di daerah yang kurang berkembang, metode ini harus dapat dipindahkan, mudah dilakukan di daerah tanpa infrastruktur, dan sangat murah [5].

Setelah terinfeksi, malaria akan berkembang pesat, dengan risiko serius untuk berkembang menjadi parah dan menyerang otak dengan gejala neurologis untuk infeksi *Plasmodium Falciparum*. Oleh karena itu, diagnosis malaria yang tepat waktu sangatlah penting. Meskipun malaria dapat didiagnosis dengan berbagai cara, tes diagnostik malaria yang ada saat ini masih perlu ditingkatkan, termasuk mengurangi biaya, meningkatkan spesifisitas, dan meningkatkan kemudahan penggunaan. Mendeteksi keberadaan parasit adalah kunci untuk mendiagnosis malaria. Selain itu, penting untuk mengidentifikasi spesies parasit dan adanya potensi infeksi campuran, serta mengamati tahapan perkembangan parasit *Plasmodium Falciparum* dalam kaitannya dengan tingkat keparahan penyakit. Menghitung parasit untuk menentukan tingkat parasitemia tidak hanya penting untuk mengidentifikasi infeksi dan mengukur tingkat keparahannya, namun juga memungkinkan pemantauan pasien dengan mengukur kemanjuran obat dan potensi resistensi obat. Klasifikasi dan analisis citra merupakan salah satu tahap yang paling penting pada pendeteksian plasmodium penyebab penyakit malaria dalam sel darah merah manusia [5].

Ekstraksi ciri berfungsi sebagai pendeteksi ciri dari suatu citra. Ciri yang dapat digunakan untuk membedakan objek satu dengan objek lainnya, diantaranya adalah ciri bentuk, ukuran, ciri geometri, ciri tekstur, dan warna. Pada penelitian ini menggunakan ekstraksi ciri tekstur Gray Level CoOccurrence Matrix (GLCM). Masing-masing objek diekstrak cirinya berdasarkan parameter-parameter tertentu dan dikelompokkan pada kelas tertentu. Nilai dari parameter tersebut kemudian dijadikan sebagai data masukan dalam proses identifikasi klasifikasi [6].

Proses klasifikasi citra berdasarkan analisis tekstur pada umumnya memerlukan tahap ekstraksi karakteristik yang terdiri dari tiga jenis metode, yaitu metode statistik, metode spektral, dan metode struktural. GLCM merupakan metode yang banyak digunakan dalam penelitian analisis tekstur pada gambar yang diperkenalkan oleh Haralick pada tahun 1973. Konsep sederhana dari GLCM adalah dapat menghitung variasi piksel dengan intensitas yang disebut i dan kesamaan piksel atau j , pada jarak atau d dan orientasi dari sudut. Pada dasarnya GLCM digunakan pada 4 orientasi sudut tertentu, yaitu sudut 0° , 45° , 90° , dan 135° . Metode GLCM (*Gray Level Co-occurrence Matrix*) termasuk dalam metode statistik dimana dalam perhitungan statistiknya menggunakan sebaran derajat keabu-abuan (*histogram*) dengan cara mengukur derajat kontras, granularitas, dan kekasaran suatu bidang hubungan efisiensi. antar piksel pada gambar. Paradigma statistik ini memiliki kegunaan yang tidak terbatas, sehingga cocok untuk subtype tekstur alami yang tidak terstruktur dan aturan yang ditetapkan (struktur mikro) [7].

Metode statistik terdiri dari ekstraksi sifat orde pertama dan ekstraksi sifat orde kedua. Ekstraksi sifat orde pertama dilakukan melalui histogram citra sedangkan ekstraksi sifat statistik orde kedua dilakukan dengan matriks kookurensi, yaitu matriks perantara yang merepresentasikan hubungan resistansi antar piksel pada citra dalam berbagai arah. orientasi dan jarak spasial. Penelitian ini mengimplementasikan metode *Gray Level Co-occurrence Matrix* (GLCM) untuk tahap klasifikasi dan ekstraksi fitur pada sistem berdasarkan ciri tekstur. Berdasarkan uraian di atas, maka pada penelitian ini diusulkan suatu sistem yang dapat mendeteksi parasit malaria pada sel dara manusia dengan menerapkan metode *Gray Level Co-occurrence Matrix* (GLCM) untuk mengklasifikasi plasmodium penyebab penyakit malaria dalam sel darah merah manusia [8].

2. METODE PENELITIAN

2.1 Sumber Data

Dataset yang digunakan berjudul *malaria cells images* dan dipublikasikan di *Kaggle*. *Dataset* terdiri dari dua kelas, yaitu kelas *parasitized* dan *uninfected*. *Dataset* berupa citra sel darah yang masing-masing kelas berjumlah 13.780 citra.

2.2 Gray Level Co-Occurance Matrix (GLCM)

Metode *Gray Level Co-Occurance Matrix* (GLCM) merupakan salah satu teknik ekstraksi tekstur pada gambar. Tujuan dari ekstraksi tekstur ini adalah untuk meraih informasi kunci dari suatu gambar sebelum digunakan dalam proses berikutnya. GLCM menggunakan sejumlah fitur pendekatan statistik, seperti energi, entropi, kontras, dan lain sebagainya. Tahapan awal dalam perhitungan fitur GLCM melibatkan konversi gambar RGB menjadi gambar skala keabuan. Langkah berikutnya adalah pembuatan matriks *co-occurrence* dengan menentukan relasi spasial antara piksel referensi dan piksel tetangga berdasarkan sudut θ dan jarak d . Selanjutnya, dibentuk matriks simetris dengan menambahkan matriks *co-occurrence* dengan transposenya. Matriks simetris kemudian dinormalisasi dengan menghitung probabilitas setiap elemen dalam matriks. Tahapan terakhir melibatkan perhitungan fitur GLCM, dimana setiap fitur dihitung dengan menggunakan satu piksel jarak dalam empat arah, yaitu 0° , 45° , 90° , dan 135° , untuk mendeteksi *co-occurrence*.

2.3 Alur Penelitian



Gambar 1. Alur Penelitian

2.4 Studi Literatur

Pada penelitian ini, dapat dilihat tahap awal yaitu melakukan studi literatur yang bertujuan untuk mempelajari metode yang akan digunakan pada penelitian ini serta perancangan sistem pengumpulan dataset.

2.5 Pengumpulan *Dataset*

Pada percobaan kali ini digunakan dataset foto plasmodium sel darah, dataset didapatkan dari platform kaggle.



Gambar 2. Sel Darah Terkena Parasit



Gambar 3. Sel Darah Tidak Terkena Parasit

2.6 Pembangunan model/Proses Testing

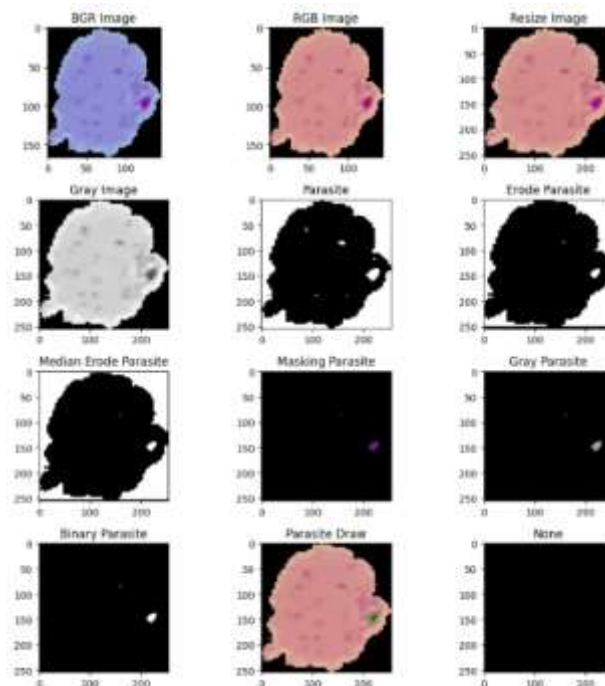
Setelah dilakukan pengumpulan data maka akan selanjutnya data akan diolah dengan melakukan *preprocessing* data. Selanjutnya data latih akan melakukan proses pelatihan model dari *dataset* yang sudah diproses sebelumnya.

2.7 Proses Testing

Pada tahapan *testing* ini akan dilakukan pada data yang telah dipisahkan menjadi data uji, testing dilakukan untuk pendeteksian serta pengenalan terhadap parasit yang ada pada citra plasmodium darah sehingga nantinya akan didapatkan hasil akhir berupa nilai *accuracy* dari proses *training* tersebut dan hasil akan di analisis dan di evaluasi.

3. HASIL DAN PEMBAHASAN

Pembangunan model pada penelitian ini menggunakan *Gray Level Co-Occurrence Matrix* (GLCM), sebelum itu, peneliti melakukan *preprocessing* data dengan beberapa metode yaitu: *bgr2rgb*, *rgb2gray*, *resize*, *threshold*, *erosi*, *median*, *masking*, dan *draw contour*. Dari *preprocessing* data ini didapatkan hasil gambar sel malaria sebagai berikut:



Gambar 4. Data Gambar Hasil *Preprocessing*

Pada gambar diatas merupakan hasil dari preprocessing pada gambar sel yang terdeteksi memiliki parasit malaria, gambar hasil ekstraksi fitur yang menggunakan metode GLCM pada citra parasit di atas menunjukkan pola tekstur yang khas dari parasit tersebut. Pola tekstur tersebut dapat dilihat dari nilai fitur GLCM yang dihitung, yaitu kontras, homogenitas, energi, dan entropi.

Setelah dilakukannya preprocessing data, selanjutnya data diproses dengan menggunakan glcm dengan 4 derajat yaitu: 0° , 45° , 95° , 135° . Berikut merupakan hasil dari proses GLCM nya.

Name	Label	Contrast0	Dissimilarity0
C100P61T	parasitized	11.849.828.624.363.100	19.878.647.522.001.300
C101P62T	parasitized	9.670.081.828.006.760	17.981.164.119.191.300
C101P62T	parasitized	14.103.038.443.723.900	18.783.078.585.765.200
C101P62T	parasitized	13.637.593.021.460.500	22.944.727.497.298.500
C101P62T	parasitized	9.220.957.233.287	18.983.171.221.244.700

Gambar 5. Hasil Fitur GLCM

Kontras menunjukkan perbedaan intensitas antara piksel yang berdekatan. Nilai kontras yang tinggi menunjukkan bahwa terdapat perbedaan intensitas yang signifikan antara piksel yang berdekatan. Pada gambar hasil ekstraksi fitur, nilai kontras parasit relatif tinggi, terutama pada arah 0° dan 90°. Hal ini menunjukkan bahwa parasit memiliki perbedaan intensitas yang signifikan antara piksel yang berdekatan, baik secara horizontal maupun vertikal.

Homogenitas menunjukkan tingkat kesamaan intensitas antara piksel yang berdekatan. Nilai homogenitas yang tinggi menunjukkan bahwa intensitas piksel yang berdekatan relatif sama. Pada gambar hasil ekstraksi fitur, nilai homogenitas parasit relatif rendah, terutama pada arah 0° dan 90°. Hal ini menunjukkan bahwa parasit memiliki tingkat kesamaan intensitas yang rendah antara piksel yang berdekatan, baik secara horizontal maupun vertikal.

Energi menunjukkan keseluruhan intensitas piksel yang berdekatan. Nilai energi yang tinggi menunjukkan bahwa terdapat intensitas piksel yang berdekatan yang cukup tinggi. Pada gambar hasil ekstraksi fitur, nilai energi parasit relatif tinggi, terutama pada arah 0° dan 90°. Hal ini menunjukkan bahwa parasit memiliki intensitas piksel yang berdekatan yang cukup tinggi, baik secara horizontal maupun vertikal.

Entropi menunjukkan tingkat tidak beraturannya intensitas piksel yang berdekatan. Nilai entropi yang tinggi menunjukkan bahwa intensitas piksel yang berdekatan tidak teratur. Pada gambar hasil ekstraksi fitur, nilai entropi parasit relatif tinggi, terutama pada arah 0° dan 90°. Hal ini menunjukkan bahwa intensitas piksel yang berdekatan pada parasit tidak teratur, baik secara horizontal maupun vertikal.

Secara keseluruhan, pola tekstur parasit yang ditunjukkan oleh nilai fitur GLCM tersebut dapat digunakan untuk membedakan parasit dari objek lain. Pola tekstur tersebut dapat digunakan untuk pengembangan aplikasi pengenalan parasit, seperti aplikasi deteksi parasit pada tanaman atau hewan. Selain data gambar didapatkan juga data angka sebagai berikut:

	precision	recall	f1-score	support
parasitized	0.83	0.97	0.89	30
uninfected	0.97	0.82	0.89	34
accuracy			0.89	64
macro avg	0.90	0.90	0.89	64
weighted avg	0.90	0.89	0.89	64

Gambar 6. Hasil Testing Proses GLCM

Data dalam gambar menunjukkan presisi, akurasi, dan f1-score dukungan dari model klasifikasi. Presisi adalah ukuran seberapa akurat model tersebut memprediksi kelas yang benar. Di sini, presisi untuk kelas *parasitized* adalah 0,83, yang berarti bahwa 83% dari waktu, model memprediksi kelas *parasitized* dengan benar. Presisi untuk kelas *uninfected* adalah 0,97, yang berarti bahwa 97% dari waktu, model memprediksi kelas *uninfected* dengan benar.

Recall adalah ukuran seberapa banyak kelas yang benar yang diprediksi oleh model. Di sini, *recall* untuk kelas *parasitized* adalah 0,97, yang berarti bahwa model memprediksi 97% dari *parasitized* yang sebenarnya. *Recall* untuk kelas *uninfected* adalah 0,82, yang berarti bahwa model memprediksi 82% dari *uninfected* yang sebenarnya.

F1-score adalah ukuran gabungan dari presisi dan recall. Di sini, *f1-score* untuk kelas *parasitized* adalah 0,89, yang berarti bahwa model memprediksi kelas *parasitized* dengan akurasi dan sensitivitas yang baik. *F1-score* untuk kelas *uninfected* adalah 0,89, yang berarti bahwa model memprediksi kelas *uninfected* dengan akurasi dan sensitivitas yang baik.

Support adalah jumlah data yang digunakan untuk menghitung *presisi*, *recall*, dan *f1-score*. Di sini, *support* untuk kelas *parasitized* adalah 30, yang berarti bahwa ada 30 data *parasitized* yang digunakan untuk menghitung *presisi*, *recall*, dan *f1-score*. *Support* untuk kelas *uninfected* adalah 34, yang berarti bahwa ada 34 data *uninfected* yang digunakan untuk menghitung *presisi*, *recall*, dan *f1-score*.

4. SIMPULAN

Berdasarkan hasil penelitian dan pengujian sistem yang dilakukan, dapat disimpulkan bahwa hasil deteksi malaria menggunakan metode *Gray Level Co-Occurance Matrix* (GLCM) dengan beberapa preprocessing yang dilakukan mendapatkan akurasi sebesar 89%. Dengan akurasi yang cukup baik, penelitian ini dapat membantu dalam pendeteksian malaria sebagai tindakan untuk mengurangi akibat fatal yang disebabkan oleh penyakit malaria.

5. SARAN

Pada penelitian berikutnya, peneliti dapat menggunakan metode yang berbeda untuk melakukan analisis dan menjadikan ini sebagai rujukan terkait dengan perbedaannya.

DAFTAR PUSTAKA

- [1] Y.B. Utomo, G.W. Harsanto, “Penerapan Metode Certainty Factor Dan Naïve Bayes Untuk Mendiagnosa Penyakit Akibat Gigitan Nyamuk” *Generation Journal*, Vol.4 No.2, 2020
- [2] P. Gascoyne, J. Satayavivad, M. Ruchirawat, “Microfluidic approaches to malaria detection”, *Acta Tropica* 89 (2004) 357–369.
- [3] N. Ain Banyal, Surianti, A. Rachmat Dayat, “Klasifikasi Citra Plasmodium Penyebab Penyakit Malaria Dalam Sel Darah Merah Manusia Dengan Menggunakan Metode Multi Class Support Vector Machine (Svm)”, *ILKOM Jurnal Ilmiah* Volume 8 Nomor 2 (Agustus 2016).
- [4] J. Bana Abraham, “Plasmodium Detection Using Simple CNN and Clustered GLCM Features”, *Electrical and Information Engineering*.
- [5] A. Negi, K. Kumar, and P. Chauhan, “Deep Learning-Based Image Classifier for Malaria Cell Detection”, *Machine Learning for Healthcare Applications*, 2021.
- [6] Q. Shandy, S.S. Panna, Y. Malago, “Penerapan Metode Grey Level Co-Occurrence Matriks (GLCM) dan K-Nearest Neighbor (K-NN) Untuk Mendeteksi Tingkat Kematangan Buah Belimbing Bintang”, *Jurnal Nasional cosPhi*, Vol. 3 No. 1, 2019
- [7] A. Prasetyo, R. Rosnelly, Wanayumini, “Identification of Malaria Parasite Patterns With Gray Level Co-Occurance Matrix Algorithm (GLCM)”, *Jurnal Rekayasa Sistem dan Teknologi Informasi*, Vol. 6 No. 3 (2022) 359 – 369, 2022.
- [8] Qurina Firdaus., et al, “Lung Cancer Detection Based On Ct-Scan Images with Detection Features Using Gray Level Conoccurrence Matrix (GlcM) and Support Vector Machine (Svm) Methods” *Indonesian Journal of Electrical Engineering and Computer Science* Vol. 16, 2020.

Klasifikasi Jeruk Segar dan Busuk Melalui GLCM dan HSV dengan Menggunakan Metode ANN

Alisyia Kornelia Ulandari¹, Ghina Kamilah Ramdhani², Wahyuningsih³, M. Naufal Arwansyuri⁴, Fitri Bimantoro⁵

^{1,2,3,4,5}Teknik Informatika, Fakultas Teknik, Universitas Mataram

E-mail: ¹alisyiaulandari@gmail.com, ²ghinagr6@gmail.com, ³wahyuningsi2707@gmail.com,
⁴arwansyurinaufal@gmail.com, ⁵bimo@unram.ac.id

Abstrak – Evolusi pesat teknologi komputer berdampak besar pada berbagai sektor, termasuk pertanian, di mana inovasi digital memiliki peran yang sangat penting. Dalam penelitian ini, kami membahas klasifikasi kualitas jeruk menggunakan teknik computer vision. Dengan memanfaatkan dataset yang terdiri dari 1.466 gambar jeruk segar dan 1.595 gambar jeruk busuk, penelitian ini menggunakan metode ekstraksi fitur GLCM dan HSV untuk mengekstrak informasi tekstur dan warna. Fokusnya adalah menyederhanakan proses klasifikasi berdasarkan warna dan tekstur kulit untuk meningkatkan akurasi. Penelitian ini berpotensi besar untuk meningkatkan penilaian terhadap kualitas jeruk, memberikan keuntungan yang signifikan bagi penjual, pembeli, dan komunitas peneliti di sektor pertanian. Hasil penelitian ini menunjukkan kinerja model yang memuaskan, dengan akurasi mencapai 86,88% dengan penerapan metode ANN.

Kata Kunci — ANN, GLCM, HSV, Jeruk

1. PENDAHULUAN

Kemajuan teknologi komputer, termasuk interaksi antara manusia dan komputer, telah mengalami perkembangan pesat dalam beberapa tahun terakhir. Bahkan dalam sektor pertanian, teknologi komputer telah banyak digunakan, terutama dalam aplikasi seperti deteksi penyakit, pengklasifikasian mutu, penentuan berat, dan identifikasi jenis buah-buahan dan sayuran. Buah-buahan merupakan salah satu kelompok produk pertanian yang memiliki peran penting di Indonesia. Kebanyakan orang menggemari buah-buahan karena kandungan nutrisinya yang bermanfaat bagi kesehatan tubuh [1]. Buah-buahan juga berperan penting dalam perkembangan ekonomi nasional, diantaranya, jeruk adalah buah kedua yang paling banyak dihasilkan setelah pisang [2].

Data dari Kementerian Pertanian RI tahun 2017 mencatat bahwa produksi jeruk mencapai 20.142.06 kilogram dari jumlah tanaman hasil sebanyak 24.135.162 pohon, dengan produktivitas sekitar 83,46 kilogram per pohon. Jeruk juga dikenal memiliki kandungan nutrisi yang sangat baik, terutama vitamin C yang diperlukan oleh manusia untuk menjaga kesehatan tubuh [2]. Setiap buah memiliki ciri-ciri yang digunakan dalam menilai kualitas dan mutunya, seperti ukuran, bentuk, warna kulit, serta potensi kerusakan atau cacat pada buah tersebut [3].

Dalam proses pengklasifikasi mutu buah jeruk, diperlukan pengolahan citra digital buah jeruk. Data citra buah jeruk ini digunakan dalam proses pelatihan sistem untuk kemudian dapat mengklasifikasi kualitas buah. Oleh karena itu, penting untuk menggunakan metode ekstraksi ciri yang efektif untuk mendapatkan informasi tekstur yang berkualitas. Hasil dari proses ekstraksi ciri ini kemudian dimanfaatkan dalam upaya mengklasifikasikan mutu buah jeruk. Dalam konteks ini, data yang digunakan mencakup 1.466 citra buah jeruk segar dan 1.595 citra buah jeruk busuk untuk analisis dan pengolahan [2].

Dalam upaya mengurangi ketidak validitas data, klasifikasi mutu buah jeruk didasarkan pada warna kulitnya, menjadi metode yang digunakan untuk menentukan kualitas buah jeruk. Pendekatan yang saat ini digunakan dalam identifikasi buah jeruk masih bersifat manual atau tradisional, seperti menilai buah berdasarkan warna kulit, tekstur kulit, dan pori-porinya. Namun, pendekatan manual ini memiliki kelemahan karena sifatnya yang subjektif dan tidak konsisten, sehingga tingkat akurasi rendah. Oleh karena itu, diperlukan suatu metode yang lebih efisien untuk mengelompokkan buah jeruk berdasarkan tingkat kematangan, terutama dengan mempertimbangkan warna kulitnya. Dengan berkembangnya teknologi komputer, proses ini dapat disederhanakan. Dalam penelitian ini, diterapkan ilmu komputer vision yang dapat membantu individu yang mungkin memiliki pengetahuan terbatas dalam menganalisis citra buah jeruk untuk menilai kualitas buah tersebut [5].

Pada penelitian klasifikasi buah jeruk segar dan busuk menggunakan ekstraksi fitur GLCM dan HSV, Kombinasi antara ekstraksi fitur GLCM dan fitur warna HSV dapat diterapkan untuk meningkatkan akurasi klasifikasi. GLCM adalah metode yang digunakan untuk mengekstraksi fitur tekstur dari citra, yang berfokus pada hubungan antara intensitas piksel dalam citra. Dalam konteks klasifikasi buah jeruk segar dan busuk, fitur tekstur dapat membantu dalam mengenali perbedaan dalam struktur dan tekstur kulit buah. Sementara itu, model warna HSV digunakan untuk mengekstraksi informasi warna dalam citra. Kombinasi informasi tekstur dan

warna dapat memberikan pemahaman yang lebih lengkap tentang karakteristik buah jeruk. Dengan menggabungkan kedua jenis fitur ini akan memungkinkan untuk menangkap lebih banyak informasi yang relevan dalam citra buah jeruk. Dan juga Seperti yang ditemukan dalam penelitian sebelumnya, penggunaan Entropi dalam ekstraksi fitur dapat meningkatkan akurasi klasifikasi [3].

Oleh karena itu, penerapan Entropi dalam ekstraksi fitur GLCM dapat membantu dalam mengidentifikasi ketidakteraturan tekstur pada citra buah jeruk. Oleh karena itu, solusi yang dapat ditawarkan pada penelitian ini berjudul Klasifikasi Buah Jeruk Segar dan Busuk Menggunakan Ekstraksi Ciri GLCM dan HSV dengan menggunakan metode ANN [10]. Dengan adanya penelitian ini diharapkan dapat mengklasifikasi buah segar dan busuk serta menjadi dasar penelitian yang akan datang ataupun menjadi dasar pengetahuan tentang klasifikasi buah segar dan busuk yang dapat membantu penjual, pembeli dan peneliti.

2. METODE PENELITIAN

2.1 Studi Literatur

Penelitian dimulai dengan studi literatur untuk memahami metode yang akan digunakan. Penulis memilih model berdasarkan perbandingan performa dari studi literatur, mempertimbangkan aspek seperti dataset dan sumberdaya yang diperlukan.

Pada ekstraksi ciri HSV dan RGB dalam Pendeteksian Kematangan Buah Jeruk Dengan Fitur Citra Kulit Buah terdapat 30 dataset sebagai sampel yang dibagi menjadi 3 diantaranya 10 sample jeruk terlalu matang, 10 sample jeruk matang, dan 10 sample jeruk mentah. Pada penelitian ini proses klasifikasi citra kulit jeruk menggunakan aplikasi GUI MATLAB dengan menghasilkan tiga klasifikasi nilai rentang Hue. Warna hijau, yang diwakili oleh rentang nilai Hue antara (30.224 - 68.68), menunjukkan jeruk mentah. Sedangkan, warna kuning kemerahan, dengan nilai Hue antara (11.914 - 29.688), menandakan jeruk yang sudah matang. Terakhir, warna merah tua gelap, dengan nilai Hue antara (0.627-8.991), menunjukkan jeruk yang sudah terlalu matang [7].

Retno melakukan pengujian ciri menggunakan GLCM dengan memanfaatkan metode ANN pada 103 citra jeruk, yang terdiri 4 jenis penyakit yaitu Scab, Canker, Black Spot (bintik hitam) dan Greening. Penelitian ini menunjukkan bahwa penambahan Entropi dapat meningkatkan tingkat akurasi sebanyak 6,73%. Dimana, akurasi tanpa Entropi adalah 71,09%, sedangkan dengan Entropi mencapai 77,82%. Hal ini menunjukkan bahwa Entropi memiliki pengaruh yang besar dalam meningkatkan akurasi dengan mengidentifikasi ketidakteraturan tekstur pada setiap citra [8].

Pada penelitian kematangan buah jeruk menggunakan metode naive bayes dengan melakukan ekstraksi fitur GLCM. Penelitian ini menggunakan dataset citra jeruk sebanyak 130 citra. Penelitian ini berhasil mengklasifikasikan kematangan buah jeruk keprok dengan sukses menggunakan ekstraksi fitur GLCM dan metode Naïve Bayes. Sistem ini efektif dan efisien. Saran untuk penelitian selanjutnya adalah menambahkan fitur seperti LBP dan dataset yang lebih besar [9].

Penelitian berikutnya Klasifikasi Citra Buah berbasis fitur warna HSV dengan memanfaatkan metode SVM untuk melakukan klasifikasi buah. Datasetnya berisi buah apel, jeruk, dan semangka berjumlah 97 citra yang berasal dari vicos.si dan dari google image. Didapatkan fitur model warna HSV, dengan cara konversi RGB ke HSV kemudian dibentuk Histogram HSV dengan 16 bin hue, 4 bin saturation, dan 4 bin value. Klasifikasi buah dengan menggunakan fitur HSV dan similarity dengan SVM sebagai metodenya diketahui hasil klasifikasi yang cukup baik, yaitu sebesar 100% [6].

Dari studi literatur yang telah dilakukan, klasifikasi buah jeruk melibatkan beberapa metode dan pendekatan untuk meningkatkan akurasi. Penelitian yang menggunakan fitur warna HSV dalam klasifikasi buah jeruk segar dan busuk mencapai akurasi yang cukup tinggi. Demikian pula, penggunaan GLCM dalam ekstraksi ciri pada beberapa penelitian klasifikasi buah segar dan busuk juga memberikan hasil akurasi yang tinggi.

Dari beberapa studi literatur yang telah dilakukan, penelitian ini bertujuan melakukan klasifikasi buah jeruk segar dan busuk dengan menggunakan metode ANN. Pendekatan ini melibatkan ekstraksi fitur menggunakan HSV dan GLCM, diharapkan dapat menjadi pendekatan yang komprehensif dalam mengklasifikasikan buah jeruk segar dan busuk. Pendekatan ini diharapkan mampu memanfaatkan informasi warna dan tekstur secara optimal.

2.2 Data

Pada penelitian ini menggunakan dataset buah jeruk yang diambil dari *source kaggle* yang terdiri dari dua kelas yaitu buah jeruk segar dan buah jeruk busuk, dimana masing masing kelas terdiri dari 1466 citra buah segar dan 1595 citra buah busuk. citra dataset buah jeruk segar dan busuk ditampilkan pada Gambar 1. dan Gambar 2.



Gambar 1. Citra Buah Jeruk Segar

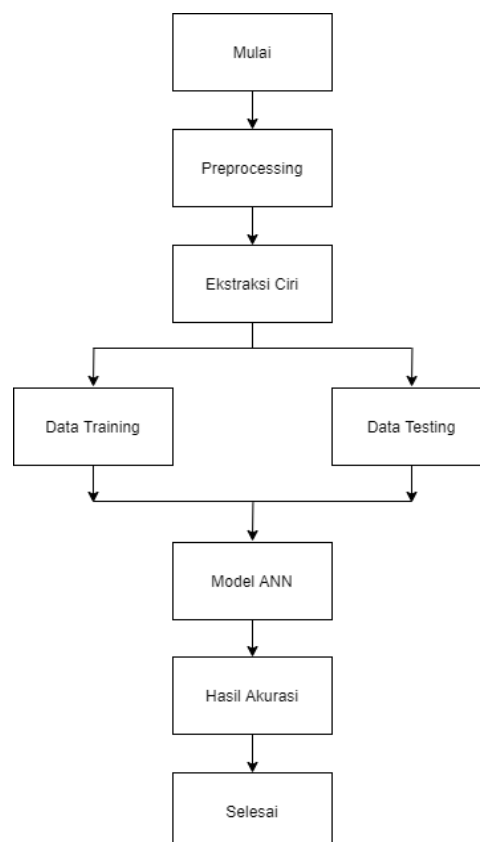


Gambar 2. Citra Buah Jeruk Busuk

Pada dataset dilakukan tahap preprocessing dimana dilakukan *rename* terhadap nama citra, kemudian dilakukan *resize* terhadap citra menjadi ukuran 128x128 piksel.

2.3 Perancangan Sistem

Perancangan sistem yang kami ajukan sesuai dengan diagram pada gambar 3. Dari Gambar 3. terlihat bahwa langkah awalnya adalah melakukan tahap *preprocessing* data pada citra dimana dilakukan *resize* pada citra dengan ukuran 128x128 piksel. Selanjutnya, citra tersebut akan dilakukan ekstraksi ciri menggunakan HSV dan GLCM, kemudian hasilnya akan disimpan dalam format CSV yang akan digunakan untuk melakukan klasifikasi buah jeruk segar dan busuk menggunakan metode ANN.



Gambar 3. Diagram Alur Kerja Sistem

2.4 Skenario Pengujian

Arsitektur pengujian melibatkan tiga jenis arsitektur yang berbeda. Setiap arsitektur menggunakan layer dense dengan jumlah neuron yang berbeda-beda. Perbedaan jumlah neuron ini dapat mempengaruhi kapasitas dan kompleksitas model. Rincian dari ketiga arsitektur tersebut dapat dilihat pada tabel di bawah ini:

Tabel 1. Arsitektur Pengujian

Arsitektur	Danse
Arsitektur 1	Danse 128, Danse 2
Arsitektur 2	Danse 64, Danse 2
Arsitektur 3	Danse 32, Danse 2

Skenario pengujian melibatkan 12 percobaan yang terbagi dalam tiga arsitektur berbeda. Setiap arsitektur menjalani empat skenario yang berbeda, dengan variasi rasio data latih (70:30 dan 80:20), serta penerapan dropout sebanyak 20% dan 50%. Penggunaan dropout pada tingkat 20% dan 50% diharapkan dapat membantu mengurangi risiko overfitting pada model yang digunakan. Dropout 20% dapat meningkatkan kapasitas model, memungkinkannya memahami pola yang lebih kompleks, sedangkan dropout 50% dapat membuat model lebih tahan terhadap variasi dan noise. Rincian dari ke-12 percobaan tersebut dapat ditemukan pada tabel di bawah ini:

Tabel 2. Skenario Percobaan

Skenario	Skenario Percobaan	Arsitektur
Skenario 1	Split train:test = 70:30	Arsitektur 1
	Dropout: 20%	
Skenario 2	Split train:test = 70:30	Arsitektur 1
	Dropout: 50%	
Skenario 3	Split train:test = 80:20	Arsitektur 1
	Dropout: 20%	
Skenario 4	Split train:test = 80:20	Arsitektur 1
	Dropout: 50%	
Skenario 5	Split train:test = 70:30	Arsitektur 2
	Dropout: 20%	
Skenario 6	Split train:test = 70:30	Arsitektur 2
	Dropout: 50%	
Skenario 7	Split train:test = 80:20	Arsitektur 2
	Dropout: 20%	
Skenario 8	Split train:test = 80:20	Arsitektur 2
	Dropout: 50%	
Skenario 9	Split train:test = 70:30	Arsitektur 3
	Dropout: 20%	
Skenario 10	Split train:test = 70:30	Arsitektur 3
	Dropout: 50%	
Skenario 11	Split train:test = 80:20	Arsitektur 3
	Dropout: 20%	
Skenario 12	Split train:test = 80:20	Arsitektur 3
	Dropout: 50%	

3. HASIL DAN PEMBAHASAN

Percobaan dilakukan dengan menggunakan beberapa arsitektur sesuai dengan skenario yang telah dibuat. Hasil percobaan dapat dilihat pada tabel di bawah ini:

Tabel 1. Arsitektur Pengujian

Arsitektur	Train	Test
Skenario 1	85,15%	85,09%
Skenario 2	85,34%	85,31%
Skenario 3	81,51%	81,17%
Skenario 4	83,28%	83,67%

Skenario 5	82,16%	83,13%
Skenario 6	72,92%	70,94%
Skenario 7	84,92%	83,19%
Skenario 8	82,18%	80,42%
Skenario 9	86,88%	86,13%
Skenario 10	83,70%	84,01%
Skenario 11	79,04%	77,97%
Skenario 12	71,28%	71,77%

Pemilihan rasio split data menjadi 80% untuk data train dan 20% untuk data testing diterapkan karena pertimbangan dataset yang digunakan tidak terlalu bervariasi dan kompleksitas masalah tidak terlalu tinggi.

Pembagian data 70% untuk data train dan 30% untuk data testing merupakan hal umum yang terbukti efektif dalam pembangunan model seperti dalam kasus klasifikasi buah jeruk segar dan busuk ini. pembagian data ini disesuaikan dengan ukuran dataset dan kebutuhan dalam proses pengerjaan dan juga membantu mengurangi *overfitting*.

Dengan pemberian 70% data untuk train, model memiliki kesempatan yang lebih baik untuk memahami pola dan fitur dalam data yang dapat meningkatkan kemampuannya untuk membuat prediksi yang akurat. split data 30% untuk testing memberikan ukuran yang cukup besar untuk mengevaluasi kinerja model dengan baik. Jumlah ini dianggap cukup untuk memberikan gambaran yang valid tentang seberapa baik model dapat menggeneralisasi pada data.

Dari 12 percobaan pada tabel di atas dilakukan 50 *epoch* percobaan dan dapat disimpulkan bahwa akurasi tertinggi didapat pada skenario 9 dengan menggunakan Arsitektur 3 yaitu Dense 32 dengan *split* data *train* dan *test* sebesar 70:30 dan menggunakan *dropout* 20%. Pada skenario 9 menghasilkan akurasi yang cukup baik yaitu sebesar 86,88% pada train-nya dan 86,13 pada test-nya. Dapat dilihat juga penggunaan dropout 50% cenderung merugikan performa model terutama pada skenario 5, 10, dan 12. Hal ini menunjukkan bahwa dropout yang terlalu tinggi dapat mengurangi kapasitas model untuk belajar.

4. SIMPULAN

Berdasarkan percobaan yang telah dilakukan, performa terbaik didapatkan pada percobaan ke-9 setelah dilakukan 12 kali percobaan. Percobaan ini melibatkan penggunaan skenario pembagian data train dan test 70:30 dan menerapkan arsitektur model ke-3. Skenario pembagian data 70:30 memberikan hasil yang optimal. Dengan menggunakan 70% data untuk train dan 30% untuk testing, model mampu menghasilkan akurasi yang baik pada dataset. Pemilihan arsitektur model ke-3 memberikan kontribusi positif terhadap peningkatan performa. Ini menunjukkan bahwa desain arsitektur model secara spesifik berpengaruh pada hasil akhir. Didapatkan hasil akurasi pada data train sebesar 86,88% dan hasil akurasi pada data testing sebesar 86,13%. Ini mengindikasikan bahwa model tidak hanya dapat mempelajari dengan baik dari data train, tetapi juga mampu menggeneralisasi data dengan baik. Dapat disimpulkan bahwa model yang dikembangkan telah mengalami optimasi dengan baik dan mampu memberikan kinerja yang memuaskan dalam penelitian klasifikasi buah jeruk segar dan busuk.

5. SARAN

Saran untuk penelitian selanjutnya perlu dicoba lebih beragam lagi ekstraksi ciri yang digunakan untuk mendapatkan kombinasi fitur yang lebih baik. Penelitian ini juga dapat dikembangkan dengan metode lain kedepannya seperti Convolutional Neural Network (CNN).

DAFTAR PUSTAKA

- [1] S. Napitu, R. Paramita Panjaitan, P. A. Nulhakim, and M. Khalik Lubis, "Klasifikasi Buah Jeruk Segar dan Busuk Berdasarkan RGB dan HSV Menggunakan Metode KNN," *J. SAINTEKOM*, vol. 13, no. 2, pp. 214–221, 2023, doi: 10.33020/saintekom.v13i2.420.
- [2] R. Widodo, A. W. Widodo, and A. Supriyanto, "Pemanfaatan Ciri Gray Level Co-Occurrence Matrix (GLCM) Citra Buah Jeruk Keprok (*Citrus reticulata* Blanco) untuk Klasifikasi Mutu," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 11, pp. 5769–5776, 2018, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/3420>
- [3] A. Hadinegoro and D. A. Rizaldilhi, "Pengaruh HSV Pada Pengolahan Citra Untuk Kematangan Buah Cabai," *Build. Informatics, Technol. Sci.*, vol. 3, no. 3, pp. 155–163, 2021, doi: 10.47065/bits.v3i3.1020.

-
- [4] K. Komal, “GLCM Algorithm and SVM Classification Method for Orange Fruit Quality Assessment,” *Int. J. Eng. Res. Technol.*, vol. 8, no. 09, pp. 697–703, 2019, [Online]. Available: <https://www.semanticscholar.org/paper/7e8c0e774b93e518201cd05055ed97fdf61a23c1>
- [5] K. P. Siwilopo and H. Marcos, “Membandingkan Klasifikasi Pada Buah Jeruk Menggunakan Metode Convolutional Neural Network Dan K-Nearest Neighbor,” *Komputa J. Ilm. Komput. dan Inform.*, vol. 12, no. 1, pp. 57–64, 2023, doi: 10.34010/komputa.v12i1.9068.
- [6] Meiriyama, “Klasifikasi Citra Buah berbasis fitur warna HSV dengan klasifikatorSVM,” *J. Komput. Terap.*, vol. 4, no. 1, pp. 50–61, 2018, [Online]. Available: <http://jurnal.pcr.ac.id>
- [7] R. Rahmadewi, G. L. Sari, and H. Firmansyah, “Pendeteksian Kematangan Buah Jeruk Dengan Fitur Citra Kulit Buah Menggunakan Transformasi Ruang Warna HSV,” *JTEV (Jurnal Tek. Elektro dan Vokasional)*, vol. 5, no. 1.1, p. 166, 2019, doi: 10.24036/jtev.v5i1.1.107560.
- [8] R. Wahyusari, S. Nuralimah, P. Studi Informatika, and S. Tinggi Teknologi Ronggolawe, “Classification of Citrus Disease Using Feature Extraction Co-Occurrence Matrix (GLCM) and Neural Network (NN) Algorithms,” *Pros. SENDIKO (Seminar Nas. Has. Penelit. Pengabdi. Masy. Bid. Ilmu Komputer)*, vol. 2, pp. 90–96, 2023.
- [9] A. R. K. Haba and K. C. Pelangi, “Sistem Cerdas Dalam Klasifikasi Kematangan Buah Jeruk Berdasarkan Fitur Ekstraksi Glcm Dengan Metode Naïve Bayes,” *J. Teknol. dan Manaj. Inform.*, vol. 5, no. 2, 2019, doi: 10.26905/jtmi.v5i2.3935.
- [10] A. W. Putri, “Implementasi Artificial Neural Network (ANN) Backpropagation Untuk Klasifikasi Jenis Penyakit Pada Daun Tanaman Tomat,” *MATHunesa J. Ilm. Mat.*, vol. 9, no. 2, pp. 344–350, 2021, doi: 10.26740/mathunesa.v9n2.p344-350.

Klasifikasi Tumor Otak Menggunakan CNN Dengan Arsitektur Resnet50

Mohammad Liyananta S.¹, Muhammad Shata' Hibrizi², Nurun Latifah³, Rosalina⁴, Fitri Bimantoro⁵

^{1,2,3,4,5}Teknik Informatika, Fakultas Teknik, Universitas Mataram

E-mail: ¹nantasid14@gmail.com, ²shatahibrizi@gmail.com, ³latifahnurun80@gmail.com,
⁴rosalinaaa1101@gmail.com, ⁵bimo@unram.ac.id

Abstrak – Penelitian ini mengusulkan penggunaan Convolutional Neural Network (CNN) dengan model ResNet-50 untuk mengklasifikasikan jenis tumor otak berdasarkan gambar MRI. Dataset terdiri dari empat kelas: Glioma, Hipofisis, Meningioma, dan Normal. Metode penelitian melibatkan pengumpulan data, preprocessing, desain arsitektur CNN, pelatihan model, dan evaluasi. Hasil pengujian menunjukkan bahwa model mampu mengklasifikasikan jenis tumor otak dengan akurasi yang memuaskan. Penerapan ResNet-50 meningkatkan kinerja dengan mengatasi masalah hilangnya gradien. Berdasarkan penelitian tersebut, klasifikasi tumor otak menggunakan CNN dengan arsitektur Resnet50 dapat mendukung deteksi dini tumor otak untuk meningkatkan akurasi diagnostik. Pada penelitian ini akurasi terbaik diperoleh sebesar 96% pada percobaan epoch ke-11.

Kata Kunci — brain tumor, CNN, classification, MRI images, ResNet-50

1. PENDAHULUAN

Perkembangan teknologi saat ini membawa pengaruh pada dunia medis karena berperan aktif untuk memudahkan dokter dalam melakukan pendeteksian suatu penyakit. Salah satu penyakit yang dapat dideteksi awal adalah penyakit tumor otak. Penyakit Tumor otak merupakan sebuah penyakit yang memperlihatkan adanya ketidaknormalan dari pertumbuhan sel otak secara tidak wajar dan tidak terkendali di dalam ataupun di area sekitar otak. Apabila penyakit tumor otak tidak ditangani dengan cepat di awal maka dapat berdampak pada terganggunya fungsi otak secara signifikan dan dapat mengancam kehidupan penderitanya. Penyakit tumor otak terbagi menjadi beberapa jenis, yaitu Glioma, Pituitary, dan Meningioma. Apabila dilihat dari citra pada ketiga jenis tumor otak tersebut, akan menunjukkan citra yang hampir mirip. Namun, telah ditemukan perbedaan di antara ketiga jenis citra tumor otak tersebut.

Tumor otak dibagi menjadi dua yaitu, tumor otak primer dan sekunder. Tumor otak primer merupakan perubahan sel yang tidak normal dan tidak terkontrol yang berasal dari sel otak itu sendiri. Sedangkan, tumor otak sekunder merupakan tumor yang menyebar ke otak dari kanker tubuh bagian lain. Kasus tumor otak di dunia semakin meningkat setiap tahunnya. Di Indonesia, terhitung ada 300 pasien setiap tahunnya yang terdiagnosis tumor otak. Bukan hanya orang dewasa, tetapi tumor otak juga menyerang anak-anak dengan usia yang tergolong muda. Banyak orang mengabaikan gejala yang disebabkan oleh tumor otak [1].

Terdapat upaya yang biasa dokter gunakan untuk melakukan pengklasifikasian jenis tumor otak, yaitu dengan melakukan pengamatan langsung (biopsi). Namun proses untuk melakukan pengamatan langsung ini memerlukan waktu yang terbilang lama karena membutuhkan uji laboratorium selama 10-15 hari. Selain itu pengamatan langsung oleh dokter dapat memberikan resiko terjadinya kesalahan. Oleh karena itu diperlukan solusi tambahan berupa penggunaan metode Convolutional Neural Network (CNN) pada deep learning yang dapat membantu dokter untuk melakukan klasifikasi dan diagnosa terhadap jenis tumor otak yang dialami oleh pasien dengan hasil kesalahan yang minim [2].

Convolutional Neural Network (CNN) adalah salah satu metode populer dari deep learning untuk pengenalan pola citra. Dalam melakukan proses ekstraksi fitur yang kompleks pada klasifikasi citra pada skala besar dengan otomatis dan efisien, metode CNN sangat baik untuk digunakan, karena metode deep learning dapat membedakan dengan jelas suatu citra yang mempunyai karakteristik yang serupa dan sulit dikenali apabila menggunakan metode machine learning tradisional. Selain itu, deep learning juga dapat secara objektif melakukan ekstraksi fitur dengan sendirinya dan bisa mengerjakan pemrosesan data gambar dalam dua dimensi secara langsung.

Beberapa penelitian terkait sudah dilakukan untuk mengklasifikasikan tumor otak. Penelitian dengan judul “Klasifikasi Tumor Otak Menggunakan *EfficientNet-B-3*” dilakukan dengan metode CNN. Metode tersebut digunakan untuk mengklasifikasikan jenis citra tumor otak. Setelah dilakukan pengujian, diperoleh hasil akurasi sebesar 99.7% dengan nilai F1-Score terbesar mencapai 99.6% [1].

Penelitian lainnya yang mengklasifikasikan tumor otak. Penelitian ini menggunakan metode CNN dengan arsitektur MobileNetV2, dan menggunakan 3167 citra MRI tumor otak yang terdiri dari 4 kelas yakni Glioma Tumor, Meningioma Tumor, No Tumor, dan Pituitary Tumor. Model dilatih menggunakan adam optimizer, dengan 32 batch dan 30 epoch, dengan mengevaluasi performa model menggunakan metrik akurasi, presisi, dan recall. Pada penelitian ini didapatkan nilai akurasi sebesar 88.64%, precision 90% dan recall 89% [3].

Adapun penelitian lainnya yaitu klasifikasi penyakit kulit dengan menggunakan metode Convolutional Neural Network. Penelitian tersebut melakukan pengembangan sistem untuk identifikasi penyakit kulit sehingga mampu melakukan klasifikasi terhadap kondisi cacar air, scabies, campak, dan jerawat dengan menerapkan metode CNN. Hasil pengujian diperoleh yaitu dengan nilai akurasi mencapai 96.53% [4].

CNN memiliki beragam arsitektur, pada penelitian yang berjudul “Convolutional Neural Network pada Klasifikasi Sidik Jari menggunakan ResNet-50”, CNN yang menggunakan arsitektur ResNet-50 dapat mengklasifikasikan citra sidik jari dengan akurasi 95% [5]. Kemudian arsitektur ResNet-50 juga digunakan pada penelitian dengan judul “Penerapan Metode Residual Network (ResNet) dalam Klasifikasi Penyakit pada Daun Gandum”, arsitektur ini dapat mengklasifikasikan penyakit daun gandum dengan akurasi 98% [6].

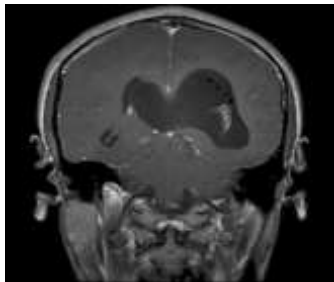
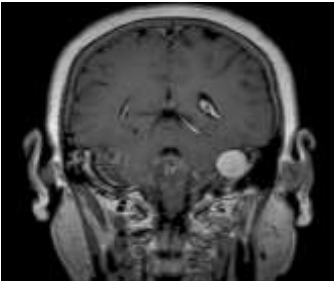
Berdasarkan latar belakang yang telah diuraikan diatas, maka dilakukan penelitian untuk klasifikasi jenis penyakit tumor otak menggunakan metode Convolutional Neural Network (CNN) menggunakan arsitektur ResNet-50 untuk melakukan klasifikasi jenis tumor otak ke dalam 4 kelas yaitu Glioma Tumor, Meningioma Tumor, No Tumor dan Pituitary Tumor. Dengan tujuan membantu dan memudahkan dokter dalam melakukan diagnosa penyakit tumor otak pada pasien sehingga diharapkan dapat menekan jumlah kasus penyakit tumor otak yang meningkat.

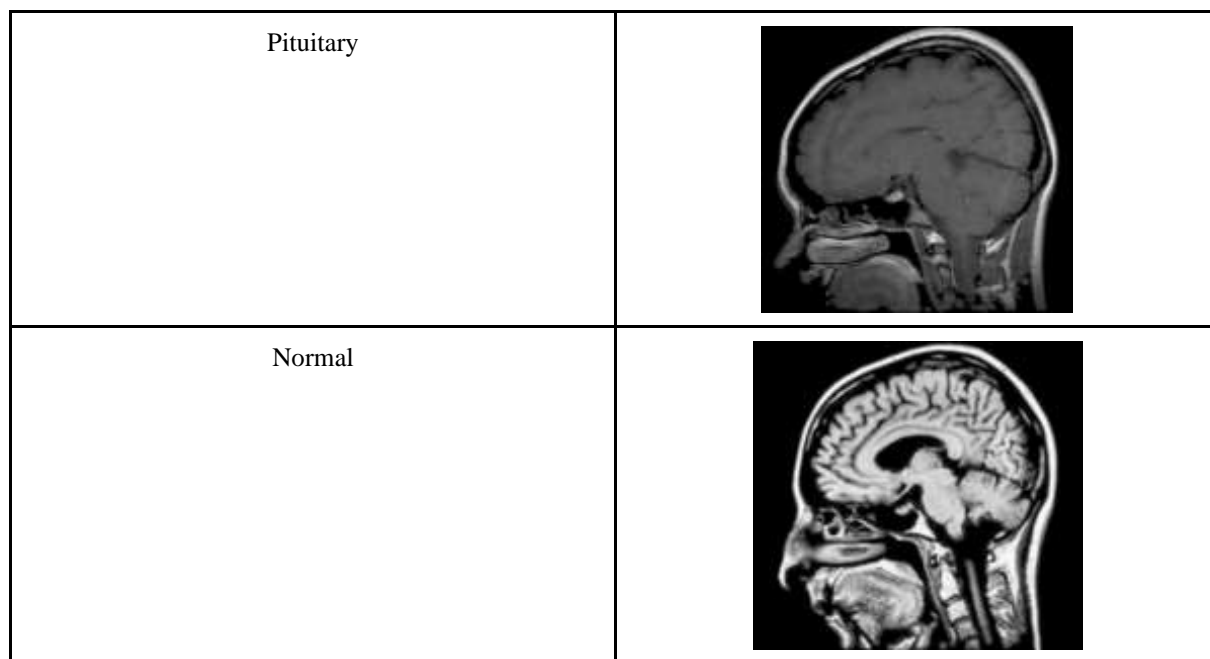
2. METODE PENELITIAN

2.1 Sumber Data

Data yang digunakan adalah dataset citra MRI tumor otak yang diambil dari website kaggle, pada tautan berikut: <https://www.kaggle.com/datasets/thomasdubail/brain-tumors-256x256>. Dataset tersebut terdiri dari 4 kelas, yaitu glioma_tumor, meningioma_tumor, pituitary_tumor, dan normal. Persebaran data citra pada masing-masing kelas, yaitu 901 citra pada kelas glioma_tumor, 913 citra pada kelas meningioma_tumor, 844 citra pada pituitary_tumor, dan 438 citra pada kelas normal.

Tabel 1. Dataset Citra Tumor Otak

Kelas	Citra
Glioma	
Meningioma	



2.2 Tumor Otak

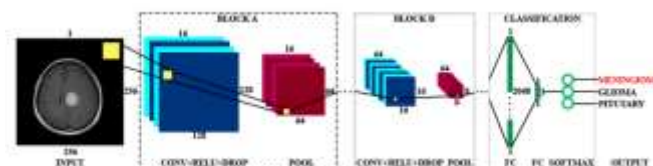
Tumor adalah hasil pertumbuhan tidak normal dari sel-sel yang merupakan komponen dasar dalam pembentukan jaringan dan organ dalam tubuh. Dalam kasus tumor otak, sel-sel yang tidak biasa berkembang dan membentuk benjolan di sekitar otak, yang bisa mengganggu fungsi normal dari otak itu sendiri.

Beberapa jenis tumor otak antara lain Glioma, Pituitary dan Meningioma. 3 jenis tumor otak ini jika dilihat dari citranya ketiganya hampir mirip. Tetapi para ahli radiologi dan juga dokter spesialis bedah berhasil menemukan bahwa ada perbedaan antara citra Glioma, Pituitary dan Meningioma.

Glioma adalah jenis tumor otak orang dewasa yang paling umum, terhitung 78% dari tumor otak ganas. Tumor tersebut muncul dari sel pendukung otak, yang disebut glia. Sel-sel ini dibagi lagi menjadi astrosit, sel endependymal, dan sel oligodendrogial (atau oligo). Meningiomas adalah tumor intrakranial jinak yang paling umum, terdiri dari 10 hingga 15% dari semua neoplasma otak, meskipun sebagian kecil merupakan tumor ganas. Tumor ini berasal dari meninges, yaitu struktur mirip membran yang mengelilingi otak dan sumsum tulang belakang. Pituitary adalah tumor intrakranial yang paling umum setelah glioma, meningioma, dan schwannoma. Sebagian besar pituitary adenoma merupakan tumor jinak dan tumbuh cukup lambat. Bahkan tumor ganas pituitary jarang menyebar ke bagian tubuh yang lain. Adenoma sejauh ini merupakan penyakit paling umum yang menyerang jaringan pituitary. Tumor tersebut biasanya menyerang orang-orang berusia 30-an atau 40-an bahkan orang dewasa. Sebagian besar tumor ini dapat diobati sampai hilang [7].

2.3 Convolutional Neural Network (CNN)

CNN merupakan salah satu tipe jaringan saraf yang sering dipakai dalam mengolah gambar. Fungsinya adalah mengenali serta mendeteksi objek dalam sebuah gambar. Meskipun pada dasarnya mirip dengan jaringan saraf biasa, CNN terdiri dari neuron yang memiliki bobot, nilai bias, dan fungsi aktivasi. Secara umum, CNN terdiri dari lapisan konvolusi, pooling, dan fully connected layer. Adapun arsitektur yang dimiliki oleh Convolutional Neural Network sebagai berikut [8].



Gambar 1. Arsitektur CNN[12]

2.4 Resnet50

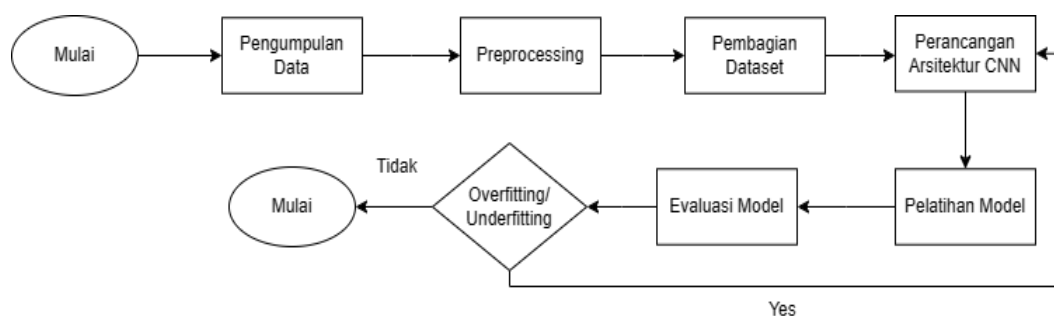
ResNet50 merupakan salah satu arsitektur dari Convolutional Neural Network (CNN) yang memperkenalkan sebuah konsep baru, yaitu shortcut connections. Munculnya konsep shortcut connections pada arsitektur ResNet-50 memiliki keterkaitan dengan vanishing gradient problem yang terjadi ketika usaha

memperdalam struktur suatu network dilakukan. Bagaimanapun, memperdalam suatu network dengan tujuan meningkatkan performanya tidak bisa dilakukan hanya dengan cara menumpuk layer. Semakin dalam suatu network, hal tersebut dapat memunculkan vanishing gradient problem yang bisa membuat gradient menjadi sangat kecil, berakibat pada menurunnya performa atau akurasi [9].

Karena itu ResNet memperkenalkan konsep shortcut connections dan dalam konsep ini fitur yang merupakan input dari layer sebelumnya juga dijadikan sebagai input terhadap output dari layer tersebut. Cara ini dilakukan sebagai solusi untuk meminimalisir hilangnya fitur-fitur penting pada saat proses konvolusi. Secara keseluruhan ResNet-50 terdiri dari 5 stage proses konvolusi yang kemudian dilanjutkan average pooling dan diakhiri dengan fully connected layer sebagai layer prediksi.

2.5 Alur Penelitian

Pengembangan penelitian ini dilakukan dengan program yang mengaplikasikan bahasa pemrograman python. Program tersebut dirancang dengan mengimplementasikan algoritma Convolutional Neural Network. Metodologi pada penelitian ini dilakukan dengan beberapa tahapan, yaitu pengumpulan data, preprocessing, pembagian dataset, perancangan arsitektur CNN, pelatihan model, dan evaluasi model. Tahapan penelitian dapat dilihat pada diagram alir pada Gambar



Gambar 2. Alur Penelitian

2.6.1 Pengumpulan Data

Tahap awal dari penelitian adalah melakukan pengumpulan data. Data yang digunakan adalah dataset citra MRI tumor otak yang diambil dari website kaggle, pada tautan berikut: <https://www.kaggle.com/datasets/thomasdubail/brain-tumors-256x256>. Dataset tersebut terdiri dari 4 kelas, yaitu glioma_tumor, meningioma_tumor, pituitary_tumor, dan normal. Persebaran data citra pada masing-masing kelas, yaitu 901 citra pada kelas glioma_tumor, 913 citra pada kelas meningioma_tumor, 844 citra pada pituitary_tumor, dan 438 citra pada kelas normal.

2.6.2 Preprocessing

Tahap awal dari penelitian adalah melakukan pengumpulan data. Proses preprocessing sebelumnya sudah dilakukan dengan menghapus data redundan dan dilakukan resize 256x256.

2.6.3 Pembagian Dataset

Dataset dibagi menjadi dua bagian yaitu data training dan data testing. Sebanyak 80% data digunakan untuk melatih model, dan 20% persennya digunakan untuk pengujian.

2.6.4 Perancangan Arsitektur CNN

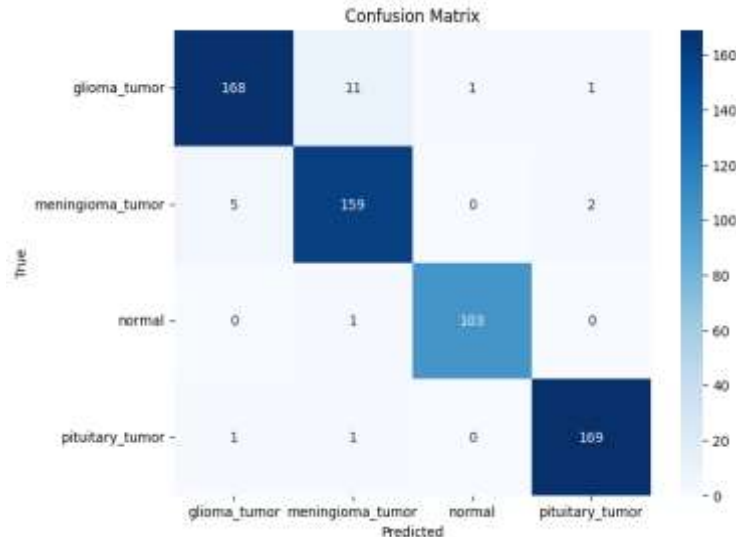
Penelitian ini mengadopsi model Arsitektur ResNet, di mana ResNet (Residual Network) merupakan salah satu arsitektur Convolutional Neural Network (CNN) yang telah terbukti efektif dalam menangani permasalahan vanishing gradient dan memungkinkan pelatihan jaringan yang lebih dalam. ResNet menonjolkan konsep residual blocks yang memungkinkan aliran informasi yang lebih lancar selama proses pelatihan [9].

Convolution layer pada ResNet juga memiliki karakteristik yang berbeda. ResNet menggunakan ketebalan filter yang sesuai dengan input image dan menerapkan residual connections untuk mempercepat pelatihan. Arsitektur ResNet memanfaatkan struktur blok residual untuk mencapai performa yang unggul dalam menghadapi kompleksitas tugas pengenalan gambar.

Diagram Arsitektur ResNet dapat diilustrasikan dengan blok-blok residual yang saling terhubung, menciptakan jaringan yang mendalam namun tetap dapat dilatih secara efisien. Pendekatan ini memungkinkan ResNet untuk mengatasi tantangan pelatihan jaringan yang mendalam tanpa mengalami penurunan performa.

3. HASIL DAN PEMBAHASAN

Hasil dari pengujian yang dilakukan sebanyak 11 epoch, yang di mana awalnya pengujian di set 50 epoch dengan mengaktifkan penghentian dini yang mana pelatihan akan dihentikan lebih awal apabila tidak ada peningkatan yang terjadi pada akurasi validasi setelah 3 epoch. Pada model yang dijalankan akan mengambil nilai akurasi terbaik dan didapatkan akurasi sebesar 96.30% pada epoch ke-11.



Gambar 3. Confusion Matrix

Hasil confusion matrix di atas diinterpretasikan sebagai berikut, pada diagonal utama yang mencerminkan jumlah prediksi yang benar pada masing-masing kelas, kelas “glioma_tumor” diprediksi benar sebanyak 168 kali, kelas “meningioma_tumor” diprediksi benar sebanyak 159 kali, kelas “normal” diprediksi benar sebanyak 103, dan kelas “pituitary_tumor” diprediksi benar sebanyak 169 kali.

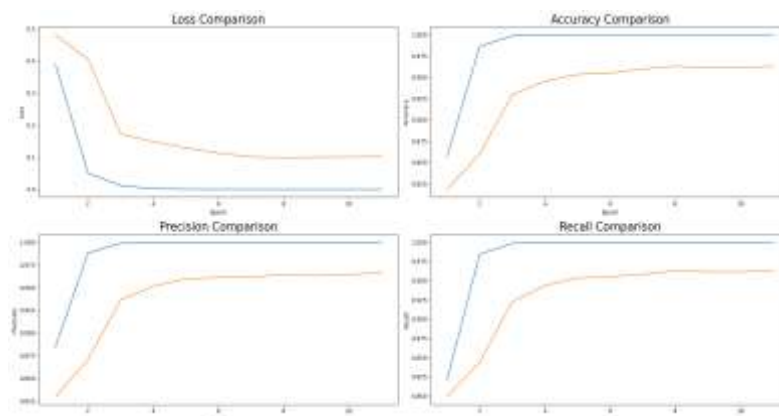
Namun, masih terdapat beberapa kesalahan yang dapat dilihat di luar dari diagonal utama pada confusion matriks tersebut, seperti pada kelas “glioma_tumor” terdapat 11 sampel data yang diprediksi sebagai “meningioma_tumor” dan sebagai kelas “normal” sebanyak 1 kali, pada kelas “meningioma_tumor” terdapat 5 sampel yang diprediksi sebagai “glioma_tumor” dan sebagai kelas “pituitary_tumor” sebanyak 2 kali, pada kelas “normal” terdapat 1 sampel data yang diprediksi sebagai kelas “meningioma_tumor”, dan pada kelas “pituitary_tumor” terdapat 1 sampel yang diprediksi sebagai “glioma_tumor” dan sebagai “meningioma_tumor” sebanyak 1 kali.

Sehingga berdasarkan confusion matrix tersebut, apabila divisualisasikan menggunakan *classification report* maka hasil pelatihan diperoleh sebagai berikut:

Tabel 2. Classification Report

	Precision	Recall	F1-Score	Support
glioma_tumor	0.97	0.93	0.95	181
meningioma_tumor	0.92	0.96	0.94	166
normal	0.99	0.99	0.99	104
pituitary_tumor	0.98	0.99	0.99	171
Accuracy: 0.96				

Berdasarkan pelatihan yang telah dilakukan menunjukkan hasil bahwa model memiliki performa yang baik dengan presisi, recall, dan F1-score yang tinggi untuk setiap kelas. Accuracy yang diperoleh adalah sebesar 96%.



Gambar 4. Grafik perbandingan *loss*, *accuracy*, *precision*, dan *recall*.

Grafik perbandingan *loss*, *accuracy*, *precision*, dan *recall* dari pelatihan dan validasi dari multiclass classification menggunakan CNN Resnet-50. Visualisasi grafik tersebut digunakan untuk mengetahui performa dari model machine learning selama pelatihan dilakukan. Grafik tersebut dapat menunjukkan bagaimana proses yang dilalui oleh model seiring berjalannya waktu dan dapat memberikan informasi indikasi apakah model yang dibuat mengalami masalah *overfitting* atau *underfitting*.

4. SIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, yakni menggunakan metode CNN dengan arsitektur Resnet50 dengan dataset tumor otak yang terdiri dari 4 kelas yakni glioma_tumor, meningioma_tumor, pituitary_tumor, dan normal, dengan split data 80 : 20 menghasilkan nilai akurasi terbaik sebesar 96% pada epoch ke-11.

5. SARAN

Pengembangan model dapat ditingkatkan dengan menerapkan metode Transfer Learning menggunakan arsitektur seperti InceptionV3, InceptionV4, atau GoogleNet. Selain itu, peningkatan model juga dapat dicapai melalui penambahan dataset atau pengenalan kategori penyakit otak yang lebih luas.

DAFTAR PUSTAKA

- [1] R. Andre, B. Wahyu, and R. Purbaningtyas, "Klasifikasi Tumor Otak Menggunakan Convolutional Neural Network Dengan Arsitektur Efficientnet-B3," *J. IT*, vol. 11, no. 3, pp. 55–59, 2021, [Online]. Available: <https://jurnal.umj.ac.id/index.php/just-it/index>
- [2] I. B. L. M. Suta, R. S. Hartati, and Y. Divayana, "Diagnosa Tumor Otak Berdasarkan Citra MRI (Magnetic Resonance Imaging)," *Maj. Ilm. Teknol. Elektro*, vol. 18, no. 2, 2019, doi: 10.24843/mite.2019.v18i02.p01.
- [3] M. N. Winnarto, M. Mailasari, and A. Purnamawati, "Klasifikasi Jenis Tumor Otak Menggunakan Arsitektur Mobilenet V2," *J. SIMETRIS*, vol. 13, no. 2, pp. 1–12, 2022.
- [4] M. A. Hanin, R. Patmasari, and R. Y. Nur, "Sistem Klasifikasi Penyakit Kulit Menggunakan Convolutional Neural Network (CNN)," *e-Proceeding Eng.*, vol. 8, no. 1, pp. 273–281, 2021.
- [5] N. D. Miranda, L. Novamizanti, and S. Rizal, "Convolutional Neural Network Pada Klasifikasi Sidik Jari Menggunakan Resnet-50," *J. Teknik. Informatika.*, vol. 1, no. 2, pp. 61–68, 2020, doi: 10.20884/1.jutif.2020.1.2.18.
- [6] A. Ridhovan and A. Suharso, "Penerapan Metode Residual Network (Resnet) Dalam Klasifikasi Penyakit Pada Daun Gandum," *JUPI (Jurnal Ilmiah. Penelitian. dan Pembelajaran Informatika.*, vol. 7, no. 1, pp. 58–65, 2022, doi: 10.29100/jupi.v7i1.2410.
- [7] A. Nada Nafisa, E. Nia Devina Br Purba, F. Aulia Alfarisi Harahap, N. Adawiyah Putri, I. Komputer, and F. Matematika dan Ilmu Pengetahuan Alam, "Implementasi Algoritma Convolutional Neural Network Arsitektur Model Mobilenetv2 Dalam Klasifikasi Penyakit Tumor Otak Glioma, Pituitary Dan Meningioma," *J. Teknol. Informasi, Komputer, danApl.(JTika)*, vol.5,no.1,pp.53–61,2023.
- [8] M. M. Badža and M. C. Barjaktarović, "Classification of brain tumors from mri images using a convolutional neural network," *Applied Sciences (Switzerland)*, vol. 10, no. 6, Mar. 2020, doi: 10.3390/app10061999.
- [9] Faiz Nashrullah, Suryo Adhi Wibowo, and Gelar Budiman, "The Investigation of Epoch Parameters in ResNet-50 Architecture for Pornographic Classification," *J. Comput. Electron. Telecommun.*, vol. 1, no. 1, pp. 1–8, 2020, doi: 10.52435/complete.v1i1.51.

Pendeteksian Kecurangan Ujian Melalui CCTV Menggunakan Algoritma YOLOv5

Fitri Bimantoro¹, I Gede Pasek Suta Wijaya², Mizanul Ridho Aohana³

^{1,2,3}Teknik Informatika, Fakultas Teknik, Universitas Mataram

E-mail: ¹[*¹bimo@unram.ac.id](mailto:bimo@unram.ac.id), ²[g²psutawijaya@unram.ac.id](mailto:gpsutawijaya@unram.ac.id), ³[m³izanul.aohana@gmail.com](mailto:mizanul.aohana@gmail.com)

Abstrak – Penggunaan teknologi di sektor pendidikan, khususnya ketika ujian, masih menghadapi tantangan berupa tingkat kecurangan yang tinggi. Salah satu penerapan teknologi ini adalah pendeteksian kecurangan saat ujian menggunakan CCTV. YOLO merupakan salah satu metode yang cukup handal untuk melakukan deteksi objek, dan YOLOv5 adalah salah satu varian YOLO yang mampu memberikan performa baik pada perangkat yang minim. Penelitian ini mengusulkan penerapan algoritma YOLOv5 untuk mendeteksi kecurangan melalui CCTV. Penelitian ini fokus pada efisiensi dan performa, dengan membandingkan tiga varian YOLOv5, yaitu YOLOv5l, YOLOv5m, dan YOLOv5s. Dataset yang digunakan merupakan rekaman video CCTV yang berada pada ruang kelas, dimana dataset ini terdiri dari 5 kelas (1 kelas normal, dan 4 kelas tindakan kecurangan). Pengujian dilakukan dengan membandingkan performa dari ketiga varian YOLOv5. Berdasarkan hasil pengujian, beban komputasi YOLOv5s saat pelatihan adalah 9,1 ms, 7x jauh lebih kecil dari pada YOLOv5l dan 3x lebih kecil dari YOLOv5m. Selain itu performa YOLOv5s lebih baik dibandingkan dengan YOLOv5l dan YOLOv5m, dengan akurasi, AP, AR dan mAP:50 sebesar 0,43, 0,492, 0,431, dan 0,549 secara berurutan. Hasil tersebut menegaskan bahwa YOLOv5s yang terbaik baik secara beban komputasi maupun performa. Meskipun demikian, perlu dilakukan perbaikan kualitas dan kuantitas dataset dan juga metode untuk meningkatkan performa dari pendeteksi kecurangan melalui CCTV ini.

Kata Kunci — deteksi, kecurangan, ujian, CCTV, YOLOv5

1. PENDAHULUAN

Penggunaan teknologi di era digitalisasi ini memang sangat masif terjadi, hampir seluruh sektor kehidupan sudah menerapkan teknologi dalam proses optimasi yang dilakukan. Optimalisasi anggaran, efisiensi kerja, dan pengurangan beban kerja menjadi salah satu parameter kunci dalam penerapan teknologi. Salah satu sektor penggunaan teknologi adalah sektor Pendidikan. Hingga saat ini, penggunaan teknologi telah terbukti dapat mengoptimalkan proses belajar mengajar [1] dan terbukti dapat meningkatkan kualitas penyerapan materi pada siswa [2]. Tidak hanya dari sisi siswa, staff pengajar juga dapat menggunakan teknologi untuk meningkatkan pengalaman belajar yang diberikan [3].

Dalam dunia Pendidikan, ujian merupakan salah satu aspek penting dalam institusi pendidikan untuk mengukur seberapa jauh kemampuan dan pengetahuan siswa terkait materi yang sudah diberikan [4]. Ujian tulis [5] merupakan salah satu metode yang paling banyak digunakan sampai saat ini [6]. Walaupun beberapa kampus atau sekolah sudah melakukan ujian berbasis komputer [7]. Kendati demikian, ujian tulis dan digital tetap memiliki persentase kecurangan yang tinggi dalam pelaksanaannya. Hal ini bisa disebabkan oleh berbagai faktor, beberapa diantaranya seperti rendahnya rasa efikasi diri [8], [9] dan ketakutan siswa menghadapi kegagalan dalam mengerjakan ujian [10]. Faktor tersebut merupakan faktor internal yang didorong oleh kepribadian masing-masing siswa, namun terdapat pula faktor eksternal yang dapat mengakibatkan hal tersebut yakni rendahnya tingkat konsistensi proktor dalam mengawasi jalannya ujian.

Pemasangan Closed Circuit Television (CCTV) di ruang kelas merupakan salah satu upaya untuk mempermudah melakukan pengawasan ujian. Namun, dalam beberapa kasus, sering kali manusia atau proktor yang mengawasi ujian sering mengalami rasa jenuh dan bosan dalam melihat aktifitas yang monoton. Oleh karena itu, diperlukan sebuah algoritma visi komputer untuk mengoptimasi proses tersebut melalui CCTV [11]. Salah satu algoritma yang populer dalam dunia visi komputer saat ini adalah You Only Look Once (YOLO) [12], [13]. Pada implementasinya, YOLO sering kali digunakan untuk melakukan pendeteksian dan pengenalan objek, beberapa diantaranya seperti deteksi masker dan jarak sosial [14], [15], deteksi kecelakaan lalu lintas [16], deteksi ketertarikan pakaian [17], dan menghitung jumlah kendaraan yang lewat di jalan raya [18]. Berdasarkan informasi dari beberapa referensi di atas, maka algoritma ini juga sangat memungkinkan untuk diterapkan dalam mendeteksi aktivitas kecurangan siswa saat mengikuti ujian di dalam ruangan kelas [11].

Melihat pada penelitian yang dilakukan sebelumnya, Arinaldi A et al. [19] melakukan penelitian tentang kecurangan ujian berdasarkan gestur tubuh menggunakan Algoritma 3D CNN dan XGBoost. Dataset yang digunakan terdiri dari 6 class, yakni *Exchange Paper*, *Use Cheat Seet*, *Talking Friend*, *Giving Code*, *Looking Friend*, dan *No Cheat*. Penelitian ini berhasil memperoleh akurasi terbaik sebesar 81,1% menggunakan algoritma XGBoost disusul dengan 3D CNN dengan akurasi sebesar 70%.

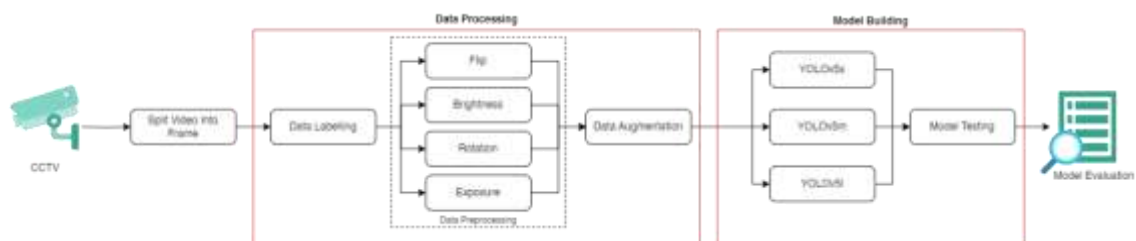
Kemudian, mengacu pada penelitian terbaru lainnya, Aqeel et al. [20] membuat model deteksi kecurangan elektronik yang memungkinkan pengawas dapat melakukan Deteksi dan Pelacakan Siswa/Orang, Mendeteksi aktivitas yang mencurigakan, Membuat peringatan, dan Menandai kehadiran. Deteksi siswa dari video dilakukan dengan menggunakan YOLOv7 dan pelacak DeepSort digunakan untuk melacak orang yang terdeteksi oleh algoritma YOLOv7. Akurasi yang didapatkan pada penelitian ini adalah 81.67%.

Studi komparasi terkait YOLOv5 dan YOLOv7 dilakukan dengan menggunakan dataset yang berbeda yakni untuk pendeteksian penyakit padi [21] dan dataset *Remote Weapon Station* [22] menghasilkan kesimpulan yang serupa. Berdasarkan pengujian yang dilakukan dengan dataset penyakit padi YOLOv5 dan YOLOv7 memiliki performa yang hampir sama. Selain itu, YOLOv5 memberikan kinerja yang lebih baik pada lingkungan dengan sumber daya yang minim[21]. Pada penelitian yang menggunakan pengenalan objek senjata sebagai objektifnya, YOLOv5 juga mampu memberikan performa yang lebih baik dibandingkan dengan YOLOv7. Pada penelitian ini dicatatkan bahwa akurasi YOLOv5 4% lebih baik daripada YOLOv7[22].

Berdasarkan beberapa referensi yang diuraikan diatas, penelitian ini akan fokus menggunakan algoritma YOLOv5. Pemilihan algoritma ini didasarkan pada tingkat eksistensinya dalam permasalahan visi komputer saat ini, khususnya pada permasalahan object detection dan object classification. Salah satu hal yang menarik dalam pemilihan algoritma ini adalah ukuran model YOLOv5 yang relatif kecil dan memiliki performa yang bagus dalam melakukan prediksi citra.

2. METODE PENELITIAN

Penelitian ini mengungkap YOLOv5 sebagai solusi dalam mengoptimasi proses pengawasan ujian menggunakan CCTV. Detail dari metode penelitian yang diusulkan dapat dilihat pada gambar di bawah ini.



Gambar 1. Penelitian yang Diusulkan

Berdasarkan Gambar 1. penelitian ini akan menggunakan beberapa metode, diantaranya Pengumpulan Dataset, Data Pre-Processing, Object Detection dan Object Tracking menggunakan YOLOv5, dan evaluasi menggunakan metrik yang sudah ditentukan. Penelitian ini akan difokuskan untuk mengevaluasi kinerja yang diberikan oleh model YOLOv5 untuk mendeteksi kegiatan kecurangan di dalam kelas.

2.1 Pengumpulan Dataset

Penelitian ini menggunakan dataset yang dikumpulkan sendiri oleh peneliti melalui Closed Circuit Television (CCTV) dengan resolusi 1920x1080, 25 fps. Pengambilan dataset diambil di ruangan A3-01 yang berlokasi di Gedung A, Program Studi Teknik Informatika, Fakultas Teknik, Universitas Mataram. Ruangan ini memiliki kapasitas 35 orang dan memiliki luas sekitar 40 m³.



Gambar 2. Pengambilan Dataset

Gambar 2. merupakan gambar tangkapan layar dari video dataset yang diambil pada ruangan A3-01. Jika mengacu pada gambar tersebut, dataset yang diperoleh cukup jelas untuk dilakukan proses pelatihan model. Setelah video dikumpulkan, akan dilakukan proses perubahan video ke dalam bentuk frame untuk memudahkan proses training model.

2.2 Data Pre-Processing

Preprocessing merupakan suatu teknik untuk meningkatkan kualitas gambar agar memperoleh hasil terbaik dari hasil yang diperoleh. Metode klasik yang umum digunakan meliputi deteksi tepi, peningkatan kontras, pengurangan noise gambar, dan normalisasi warna gambar. Sering digunakan untuk memfokuskan dan memperoleh gambar dari objek penelitian, dan metode deviasi standar sering digunakan untuk menampilkan warna gambar dengan lebih jelas [23].

Langkah awal yang dilakukan dalam proses Pre-Processing adalah perubahan format dataset, yang semula berformat “.mp4” menjadi potongan-potongan frame dalam format “.jpg”. Proses ini dilakukan menggunakan bahasa python dengan interval pengambilan frame setiap 7 detik.

Setelah melakukan transformasi format, penulis melakukan anotasi pada dataset gambar, anotasi ini bertujuan untuk mengambil objek manusia yang ada dalam gambar dan memisahkannya ke dalam kelas-kelas yang sudah ditentukan. Dalam implementasinya, penulis menggunakan bantuan *tools Roboflow* untuk membantu mempermudah proses anotasi dataset.



Gambar 3. Anotasi Dataset

Gambar 3. merupakan salah satu dataset yang telah menjalani proses anotasi. Proses anotasi ini akan mengelompokkan data ke dalam lima kelas, yaitu "normal," "cheating," "exchange paper," "giving code," dan "looking friends." Data-data ini dianotasi secara menyeluruh, mulai dari kepala hingga tangan peserta ujian, sehingga secara jelas memperlihatkan gestur tubuh yang diperagakan oleh peserta ujian.

Setelah proses anotasi selesai, proses selanjutnya adalah proses pre-processing dan augmentasi. Proses ini melibatkan 3 operasi manipulasi citra seperti flip horizontal, rotasi (antara -5° hingga 5°), peningkatan kecerahan (antara -25% hingga 25%), dan peningkatan exposure (antara -10% hingga 10%). Selain itu, dilakukan pula augmentasi pada dataset sebanyak tiga kali, sehingga jumlah total data yang telah dianotasi mencapai 10.252. Jika dijabarkan lebih lanjut, kelas "normal" memiliki 3.975 gambar, "Looking Friend" memiliki 3.897 gambar, "Cheat Sheet" memiliki 1.890 gambar, "Giving Code" memiliki 372 gambar, dan kelas terakhir, "Exchange Paper," memiliki 136 gambar.

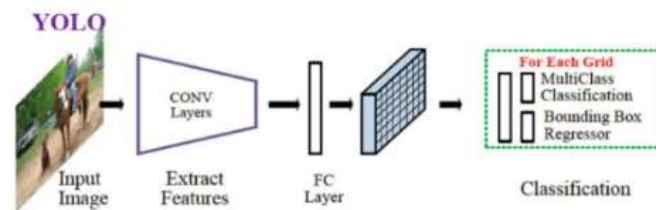
2.3 Algoritma YOLOv5

Pengenalan objek secara real time adalah elemen penting dalam sistem visi komputer [20]. Sedangkan pelacakan objek adalah salah aspek fundamental dalam konsep visi komputer yang mencoba mendeteksi dan melacak setiap objek dalam rangkaian gambar [24]. Pengenalan objek melibatkan pengklasifikasian dan

penentuan lokasi objek yang potensial dalam satu gambar secara bersamaan [25]. Terdapat banyak sekali algoritma yang dapat digunakan dalam object detection seperti Faster R-CNN, SSD, RetinaNet, and YOLO [20].

YOLOv5 adalah salah satu variasi model YOLO yang dirilis oleh ultralytics dan saat ini YOLOv5 merupakan model paling populer yang digunakan dalam dunia deteksi objek dan pelacakan objek. Saat ini pengembangan algoritma YOLO sudah sampai pada YOLOv8, tetapi hal yang menarik adalah YOLOv5 masih banyak digunakan oleh berbagai kalangan untuk berbagai kebutuhan. Arsitektur model yang efisien dan kecil dari segi ukuran menjadikan YOLOv5 masih menjadi primadona untuk melakukan berbagai pekerjaan. Oleh karena itu, peneliti tertarik untuk menggunakan algoritma YOLOv5 dalam kasus pendeteksian kecurangan ujian melalui CCTV.

Secara umum, arsitektur YOLO dapat digambarkan sebagai pendekatan deteksi objek yang memproses gambar secara keseluruhan dan menghasilkan prediksi objek dalam satu feedforward pass melalui neural network. Pendekatan ini memungkinkan deteksi objek yang cepat dan efisien. Dengan menggunakan berbagai varian YOLO, peneliti akan melakukan evaluasi kinerja dan keakuratan setiap model dalam mendeteksi perilaku mencurigakan selama ujian. Berikut arsitektur YOLO [26].

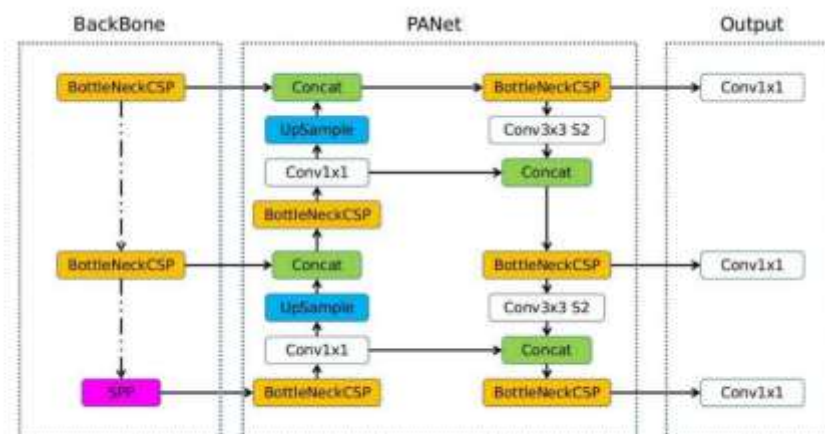


Gambar 4. Arsitektur Umum YOLO

Gambar 1. Memperlihatkan bahwa algoritma YOLO dimulai dengan memasukkan gambar sebagai langkah pertama. Masukan ini kemudian akan melewati serangkaian tahap ekstraksi fitur di dalam Convolutional Neural Network (CNN). Tahap ini bertujuan untuk mengidentifikasi fitur-fitur unik yang ada dalam gambar, seperti tepi objek, tekstur, warna, dan pola lainnya. Hasil dari tahap ekstraksi fitur ini adalah representasi data berdimensi tinggi yang menggambarkan gambar dengan cara yang lebih abstrak.

Setelah fitur-fitur tersebut berhasil diekstraksi, langkah selanjutnya adalah mengirimkannya ke lapisan Fully Connected Layer (FC) dalam Neural Network. Lapisan FC ini kemudian berfungsi untuk menggabungkan dan memproses fitur-fitur yang telah diambil sebelumnya. Dalam proses ini, pengolahan akan melibatkan perhitungan yang lebih kompleks dan pembentukan representasi data yang lebih abstrak. Output dari lapisan FC inilah yang akan menentukan hasil akhir dalam proses klasifikasi pada algoritma YOLO. Hasil output ini melibatkan informasi terkait lokasi objek (koordinat bounding box), kelas objek yang terdeteksi, dan tingkat kepercayaan (confidence) terhadap deteksi tersebut. Dengan kata lain, YOLO menggunakan fitur-fitur yang telah diproses melalui lapisan FC untuk menghasilkan informasi komprehensif tentang objek-objek dalam gambar, mencakup jenis objek, lokasinya, dan seberapa yakinnya algoritma terhadap deteksinya.

Jika dianalisis lebih jauh, penggunaan spesifik terhadap arsitektur YOLOv5 dapat divisualisasikan ke dalam bentuk Gambar 5.



Gambar 5. Arsitektur YOLOv5

Perbedaan utama dari YOLOv5 jika dibandingkan dengan model sejenis dan model sebelumnya terletak pada penggunaan CrossStage Partial Network (CSPNet) sebagai tulang punggung model dan Path Aggregation Network (PANet) sebagai leher agregasi fitur. Arsitektur model dapat dilihat pada Gambar 1. Perbaikan-perbaikan ini menghasilkan ekstraksi fitur yang lebih optimal dan peningkatan yang signifikan dalam nilai mAP score [27].

Dengan pendekatan ini, YOLOv5 dapat secara efisien dan akurat mendeteksi objek-objek dalam gambar dengan menggabungkan ekstraksi fitur yang canggih dengan pengolahan informasi dalam lapisan fully connected untuk memberikan hasil klasifikasi yang kuat.

2.4 Metrik Evaluasi

Untuk mengevaluasi kinerja model dalam penelitian ini, digunakan beberapa metrik alternatif seperti Confusion Matrix untuk mengukur kinerja klasifikasi model, Average Precision (AP), dan mean Average Precision (mAP) untuk mengukur seberapa jauh model mampu melakukan deteksi objek secara presisi. Ketiga evaluasi ini dilakukan secara berurutan untuk mendapatkan hasil evaluasi performa yang optimal.

Confusion Matrix merupakan salah satu skema evaluasi yang paling sering digunakan dalam machine learning, khususnya supervised learning [28]. Confusion matrix terdiri dari 4 nilai evaluasi, yakni True Positive (TP), True Negatif (TN), False Positif (FP) dan False Negatif (FN) [29].

Tabel 1. *Confusion Matrix Theory*

	Prediction		
	Positive	Negative	
	Actual	TP	TN
	Negative	FP	FN

Tabel 1. Menunjukkan komposisi penyusunan yang ada pada confusion matrix, berdasarkan tabel tersebut dapat dihitung matrix evaluasi berupa accuracy, precision dan recall. Persamaan (1), (2), dan (3) merupakan persamaan yang digunakan untuk menghitung nilai evaluasi [30].

$$Acc = \frac{TP+TN}{TP+FN+FP+TN} \dots\dots\dots(1)$$

$$P = \frac{TP}{TP+FP} \dots\dots\dots(2)$$

$$R = \frac{TP}{TP+FN} \dots\dots\dots(3)$$

Accuracy (Acc) merupakan salah satu metode pengujian berdasarkan tingkat kedekatan keseluruhan antara nilai prediksi dengan nilai actual. Precision (P) didefinisikan sebagai ketepatan prediksi yang memiliki label benar dengan hasil positif. Sedangkan Recall (R) merepresentasikan bagaimana kuantitas dari kualitas pada seluruh proses prediksinya [31].

Dalam algoritma YOLO, salah satu indikator utama untuk menilai performa model adalah dengan menghitung nilai Mean Average Precision (mAP) [32]. Metrics mAP adalah metrik yang umum digunakan dalam tugas deteksi objek untuk mengukur sejauh mana model mampu mendeteksi dan mengklasifikasikan objek dengan akurasi tinggi [33].

Secara konsep, mAP akan menghitung presisi rata-rata (Average Precision) untuk setiap kelas objek [34] Presisi mengukur sejauh mana deteksi yang dilakukan oleh model benar-benar merupakan objek yang dimaksud, dan ini dihitung dengan menghitung jumlah true positive (TP) dibagi dengan jumlah total prediksi positif yang dibuat oleh model [35].

$$AP = \sum_{i=1}^N P(i)\Delta R(i) \dots\dots\dots(4)$$

$$mAP = \frac{1}{N} \sum_{i=1}^N AP(img_i) \dots\dots\dots(5)$$

Dimana AP mendefinisikan average precision, sedangkan N mendefinisikan jumlah class dan img_i melambangkan image pada tiap iterasi. Nilai AP dan mAP memiliki rentang 0-1, sedangkan nilai ideal pada kedua tolak ukur tersebut adalah 1.

3. HASIL DAN PEMBAHASAN

Pengujian yang dilakukan pada penelitian ini dengan membandingkan beban komputasi, performa dan menganalisis hasil prediksi dari 3 jenis YOLOv5 yaitu YOLOv5l, YOLOv5m, dan YOLOv5s. Berdasarkan Tabel 1, beban komputasi dari ketiga jenis YOLOv5 dihitung berdasarkan jumlah parameter, GFLOPS dan waktu eksekusi ketika preproses, pelatihan dan pengujian. Tabel 3 menunjukkan bahwa YOLOv5s memiliki jumlah parameter yang jauh lebih kecil dari YOLOv5 yang lainnya, hampir 7 kali lebih kecil dari YOLOv5l dan

3x lebih kecil dari YOLOv5m. Hal ini tentu mempengaruhi waktu pelatihannya, dapat dilihat waktu pelatihan meningkat sesuai dengan jumlah parameter yang dimiliki masing-masing model. Namun hal tersebut tidak mempengaruhi secara signifikan waktu preproses dan pengujian. Parameter lainnya yang menyatakan beban komputasi adalah GFLOPs, berdasarkan Tabel 3 YOLOv5l memiliki GFLOPs tertinggi, sedangkan YOLOv5m dan YOLOv5s memberikan GFLOPs yang serupa. Secara umum, Tabel 1 menunjukkan bahwa YOLOv5s memiliki beban komputasi yang paling ringan dibandingkan 2 varian YOLOv5 yang lainnya.

Tabel 3. Komparasi Komputasi YOLOv5

Algoritma	Parameter	GFLOPs	Waktu rata-rata		
			preproses	pelatihan	pengujian
YOLOv5l	46.13 M	107.7	0.4 ms	40.6 ms	1.6 ms
YOLOv5m	20.87 M	47.9	0.4 ms	22.0 ms	1.6 ms
YOLOv5s	7.02 M	45.153	0.4 ms	9.1 ms	1.2 ms

Tabel 4 menunjukkan performa untuk ketiga YOLOv5 berdasarkan nilai akurasi, rata-rata presisi (AP), rata-rata recall (AR), dan mAP:50. Nilai akurasi, AP dan AR yang dihasilkan oleh YOLOv5s lebih baik daripada kedua YOLOv5 lainnya. Sedangkan YOLOv5l dan YOLOv5m memiliki performa yang serupa pada metrik evaluasi ini. Namun hal ini berbeda ketika melihat nilai mAP:50, walaupun tidak signifikan, YOLOv5l dan YOLOv5s memberikan nilai mAP yang serupa sedangkan YOLOv5m memiliki nilai yang sedikit lebih rendah. Dapat disimpulkan berdasarkan Tabel 4, secara performa YOLOv5s memberikan performa yang lebih baik dibandingkan YOLOv5l dan YOLOv5m.

Tabel 4. Metrik Evaluasi YOLOv5

Algoritma	Akurasi	AP	AR	mAP:50
YOLOv5l	0,39	0,401	0,391	0,555
YOLOv5m	0,39	0,406	0,385	0,497
YOLOv5s	0,43	0,492	0,431	0,549

Tabel 5 memberikan informasi tentang hasil prediksi 5 kelas dari YOLOv5 yang terdiri dari *Cheat Sheet* (CS), *Exchange Paper* (EP), *Giving Code* (GC), *Looking Friend* (LF) dan Normal (N). Kelas LF merupakan kelas yang konsisten memberikan nilai ketepatan yang tinggi dibandingkan kelas yang lainnya pada setiap metode, sedangkan kelas GC merupakan kelas dengan nilai ketepatan prediksi terendah, bahkan hanya YOLOv5s yang mampu memberikan nilai prediksi diatas 0. YOLOv5l baik dalam menentukan kelas normal dengan ketepatan 76%, sedangkan YOLOv5m sangat baik dalam melakukan prediksi LF. Namun secara umum, YOLOv5s mampu menghasilkan prediksi bukan 0 pada setiap kelas, dan juga unggul dalam melakukan prediksi pada kelas EP dan GC dimana kedua kelas ini memiliki jumlah dataset yang paling sedikit. Walaupun pada kelas yang lainnya tidak unggul, namun perbedaan jumlah prediksi yang diberikan tidaklah signifikan.

Tabel 5. Confusion Matrix YOLOv5s

		YOLOv5-L						YOLOv5-M						YOLOv5-S					
		CS	EP	GC	LF	N	BG	CS	EP	GC	LF	N	BG	CS	EP	GC	LF	N	BG
Actual	CS	0,71	0,06	0,03	0,04	0,05	0,3	0,69	0,08	0,01	0,05	0,07	0,28	0,69	0,06	0,03	0,04	0,08	0,25
	EP	0	0,08	0	0	0	0	0	0,08	0	0	0	0	0	0,25	0	0	0	0,03
	GC	0	0	0	0	0	0	0	0	0	0	0	0	0	0,03	0,11	0	0,01	0,02
	LF	0,1	0,58	0,62	0,79	0,15	0,42	0,11	0,64	0,65	0,82	0,17	0,5	0,11	0,56	0,57	0,78	0,14	0,39
	N	0,14	0,03	0,17	0,11	0,76	0,27	0,12	0	0,13	0,08	0,72	0,22	0,14	0	0,17	0,11	0,75	0,31
	BG	0,05	0,25	0,18	0,05	0,04	0	0,08	0,19	0,21	0,05	0,04	0	0,06	0,11	0,13	0,05	0,03	0

Berdasarkan analisis yang telah dilakukan, YOLOv5s mampu memberikan performa yang lebih optimal daripada kedua varian YOLOv5 yang lain. Hal ini terlihat bagaimana arsitektur YOLOv5s yang kecil mampu membuat beban komputasinya jadi jauh lebih ringan. Walaupun dengan arsitektur yang kecil, YOLOv5s

memberikan performa yang baik, selain itu YOLOv5s mampu mendeteksi dengan baik kelas EP dan GC yang tidak mampu dideteksi oleh YOLOv5l dan YOLOv5m.

4. SIMPULAN

YOLOv5 merupakan varian YOLO yang baik digunakan untuk pendeteksian objek. Dalam menentukan kecurangan melalui CCTV, YOLOv5 mampu memberikan performa yang cukup baik walaupun masih jauh dari kata ideal. Hal ini terlihat dari akurasi, AP, AR, dan mAP terbaik sebesar 0,43, 0,492, 0,431, dan 0,555. Salah satu yang menjadi penyebabnya adalah adanya ketidakseimbangan jumlah data pada setiap kelasnya. Selanjutnya, YOLOv5s, yang merupakan varian YOLOv5 dengan arsitektur paling kecil mampu memberikan yang terbaik dari kedua model lainnya. Dengan keunggulan jumlah parameter paling sedikit, YOLOv5s memberikan performa yang paling baik walaupun tidak signifikan. Selain itu, YOLOv5s mampu memprediksi dua kelas yang tidak dapat diprediksi oleh YOLOv5l dan YOLOv5m.

5. SARAN

Berdasarkan penelitian yang telah dilakukan, rencana penelitian kedepannya yakni dengan memperbaiki kualitas dataset dengan upaya memperbanyak jumlah dataset terutama pada kelas-kelas dengan jumlah data paling sedikit. Selain itu, perlu dilakukan pengujian dengan menggunakan metode-metode deep learning yang lainnya disertai *tuning* pada hyperparameternya. Diharapkan dengan melakukan hal tersebut akan membuat performa dari pendeteksi kecurangan melalui CCTV akan meningkat.

DAFTAR PUSTAKA

- [1] H. Cai and I. King, "Education Technology for Online Learning in Times of Crisis," in *Proceedings of 2020 IEEE International Conference on Teaching, Assessment, and Learning for Engineering, TALE 2020*, Institute of Electrical and Electronics Engineers Inc., Dec. 2020, pp. 758–763. doi: 10.1109/TALE48869.2020.9368387.
- [2] L. Al-Labadi and S. Sant, "Enhance Learning Experience Using Technology In Class," *J Technol Sci Educ*, vol. 11, no. 1, pp. 44–52, Sep. 2021, doi: 10.3926/jotse.1050.
- [3] M. Alahmari, "Exploring the Influential Factors Affecting Staff Willingness to Adopt Augmented Reality," *International Journal of Information and Education Technology*, vol. 13, no. 7, pp. 1078–1084, Jul. 2023, doi: 10.18178/ijiet.2023.13.7.1907.
- [4] S. R. Sletten, "Rethinking Assessment: Replacing Traditional Exams with Paper Reviews," *J Microbiol Biol Educ*, vol. 22, no. 2, Sep. 2021, doi: 10.1128/jmbe.00109-21.
- [5] É. Cambron-Goulet, J. P. Dumas, É. Bergeron, L. Bergeron, and C. St-Onge, "Guidelines for Creating Written Clinical Reasoning Exams: Insight from a Delphi Study," *Health Professions Education*, vol. 5, no. 3, pp. 237–247, Sep. 2019, doi: 10.1016/j.hpe.2018.09.001.
- [6] R. Ellis, J. Cleland, D. SG. Scrimgeour, A. J. Lee, J. Hines, and P. A. Brennan, "Establishing the predictive validity of the intercollegiate membership of the Royal Colleges of surgeons written examination: MRCS Part A," *The Surgeon*, Aug. 2023, doi: 10.1016/j.surge.2023.07.004.
- [7] A. Manurung, A. Fildzah, and J. Rajagukguk, "Computer Based Test (CBT) System for Student Academic Examination," European Alliance for Innovation n.o., Oct. 2019. doi: 10.4108/eai.18-10-2018.2287186.
- [8] A. Fadillah, "Hubungan Antara Efikasi Diri Dengan Perilaku Menyontek Pada Mahasiswa," *Psikoborneo: Jurnal Ilmiah Psikologi*, vol. 7, no. 4, pp. 657–664, 2019.
- [9] D. Septiana, "HUBUNGAN EFIKASI DIRI DAN STRES AKADEMIK DENGAN PERILAKU MENYONTEK SISWA," UNIVERSITAS ISLAM NEGERI RADEN INTAN LAMPUNG, Lampung, 2022. Accessed: Nov. 18, 2023. [Online]. Available: <http://repository.radenintan.ac.id/17318/>
- [10] C. M. Ifeagwazi, J. C. Chukwuorji, S. O. Egbodo, and M. B. Nwoke, "Peer pressure, fear of failure and examination cheating behavior in the university: Does gender make the difference?," *Cognition, Brain, Behavior. An Interdisciplinary Journal*, vol. 23, no. 1, pp. 43–62, 2019, doi: 10.24193/cbb.2019.23.03.
- [11] J. Nishchal, S. Reddy, and P. N. Navya, "Automated Cheating Detection in Exams using Posture and Emotion Analysis," in *2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, IEEE, Jul. 2020, pp. 1–6. doi: 10.1109/CONECCT50063.2020.9198691.
- [12] C. Wu, Q. Luo, and D. Shou, "Image Recognition Technology and Development of YOLO Algorithm," in *Advances in Transdisciplinary Engineering*, IOS Press BV, Dec. 2022, pp. 32–38. doi: 10.3233/ATDE221146.

- [13] T. Diwan, G. Anirudh, and J. V. Tembhurne, "Object detection using YOLO: challenges, architectural successors, datasets and applications," *Multimed Tools Appl*, vol. 82, no. 6, pp. 9243–9275, Mar. 2023, doi: 10.1007/s11042-022-13644-y.
- [14] V. Senthilkumar, G. Kanagaraj, T. Primya, J. Joycema, M. B. Joan, and J. A. Vicram, "Application of AI And Computer Vision To Face Mask And Social Distance Detection in CCTV Video Streams," in *2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA)*, IEEE, Oct. 2021, pp. 1–5. doi: 10.1109/ICAECA52838.2021.9675746.
- [15] B. Hardiansyah and A. Primasetya, "Sistem Deteksi Penggunaan masker (Face Mask Detection) Menggunakan Algoritma Deep Learning YOLOv4," *STAINS (SEMINAR NASIONAL TEKNOLOGI & SAINS)*, vol. 2, no. 1, pp. 313–318, Jan. 2023, Accessed: Dec. 25, 2023. [Online]. Available: <https://proceeding.unpkediri.ac.id/index.php/stains/article/view/2820>
- [16] A. Srinivasan, A. Srikanth, H. Indrajit, and V. Narasimhan, "A Novel Approach for Road Accident Detection using DETR Algorithm," in *2020 International Conference on Intelligent Data Science Technologies and Applications, IDSTA 2020*, Institute of Electrical and Electronics Engineers Inc., Oct. 2020, pp. 75–80. doi: 10.1109/IDSTA50958.2020.9263703.
- [17] E. P. Ijjina, A. S. Joshi, and G. Kanahasabai, "Detection of Customer Interested Garments in Surveillance Video using Computer Vision," in *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Kharagpur, India: IEEE, Jul. 2020, pp. 1–5. doi: 10.1109/ICCCNT49239.2020.9225571.
- [18] N. Chhadikar, P. Bhamare, K. Patil, and S. Kumari, "Image processing based Tracking and Counting Vehicles," in *2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore, India: IEEE, Jun. 2019, pp. 335–339. doi: 10.1109/ICECA.2019.8822070.
- [19] A. Arinaldi and M. I. Fanany, "Cheating Video Description Based on Sequences of Gestures," in *2017 5th International Conference on Information and Communication Technology, ICoICT 2017*, Institute of Electrical and Electronics Engineers Inc., Oct. 2017. doi: 10.1109/ICoICT.2017.8074679.
- [20] A. Zaffar, M. Jawad, M. Shabbir, P. O. Factory, and W. Cantt, "A Novel CNN-RNN Model for E-Cheating Detection Based on Video Surveillance," *UW JOURNAL OF COMPUTER SCIENCE*, vol. 5, no. 1, Jun. 2023, Accessed: Sep. 11, 2023. [Online]. Available: <https://uwjcs.org.pk/index.php/ojs/article/view/64>
- [21] M. T. Mulyani, "Analisis Perbandingan Algoritma Yolov5 Dan Yolov7 Dalam Pendeteksian Penyakit Padi Pada Raspberry Pi," Jakarta, Jul. 2023. Accessed: Dec. 25, 2023. [Online]. Available: https://repository.pnj.ac.id/id/eprint/12477/1/PEMBUKAAN%20sd%20BAB%20I_Meutia%20Tri%20Mulyani.pdf
- [22] O. E. Olorunshola, M. E. Irhebhude, and A. E. Ewwiekpaefe, "A Comparative Study of YOLOv5 and YOLOv7 Object Detection Algorithms 1*," *Journal of Computing and Social Informatics*, vol. 2, no. 1, p. 1, 2023.
- [23] U. P. Sanjaya, Z. Alawi, A. R. Zayn, and G. Dirgantara, "Optimasi Convolutional Neural Network dengan Standard Deviasi untuk Klasifikasi Pneumonia pada Citra X-rays Paru," *Generation Journal*, vol. 7, no. 3, pp. 40–47, Oct. 2023, doi: <https://doi.org/10.29407/gj.v7i3.20183>.
- [24] Z. Soleimanitaleb, M. A. Keyvanrad, and A. Jafari, "Object Tracking Methods:A Review," in *2019 9th International Conference on Computer and Knowledge Engineering (ICCKE)*, IEEE, Oct. 2019, pp. 282–288. doi: 10.1109/ICCKE48569.2019.8964761.
- [25] Z. Q. Zhao, P. Zheng, S. T. Xu, and X. Wu, "Object Detection with Deep Learning: A Review," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 11. Institute of Electrical and Electronics Engineers Inc., pp. 3212–3232, Nov. 01, 2019. doi: 10.1109/TNNLS.2018.2876865.
- [26] O. E. Olorunshola, M. E. Irhebhude, and A. E. Ewwiekpaefe, "A Comparative Study of YOLOv5 and YOLOv7 Object Detection Algorithms," *Journal of Computing and Social Informatics*, vol. 2, no. 1, pp. 1–12, Feb. 2023, doi: 10.33736/jcsi.5070.2023.
- [27] D. Dlužnevskij, P. Stefanovč, and S. Ramanauskaite, "Investigation of YOLOv5 efficiency in iPhone supported systems," *Baltic Journal of Modern Computing*, vol. 9, no. 3, pp. 333–344, 2021, doi: 10.22364/bjmc.2021.9.3.07.
- [28] M. Hasnain, M. F. Pasha, I. Ghani, M. Imran, M. Y. Alzahrani, and R. Budiarto, "Evaluating Trust Prediction and Confusion Matrix Measures for Web Services Ranking," *IEEE Access*, vol. 8, pp. 90847–90861, May 2020, doi: 10.1109/ACCESS.2020.2994222.
- [29] R. Ilahi, F. Bimantoro, R. Dwiyanaputra, and R. Farinda, "Classification of Pringgasela Typical Songket Using Multi Texton Co-occurrence Descriptor and K-Nearest Neighbor," in *Proceedings of the First Mandalika International Multi-Conference on Science and Engineering 2022, MIMSE 2022 (Informatics and Computer Science)*, Dordrecht: Atlantis Press International BV, 2022, pp. 352–366. doi: 10.2991/978-94-6463-084-8_30.

-
- [30] G. Tripathi and R. Kumar, “Early Prediction of Diabetes Mellitus Using Machine Learning,” in *2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, IEEE, Jun. 2020, pp. 1009–1014. doi: 10.1109/ICRITO48877.2020.9197832.
- [31] L. Farokhah, “IMPLEMENTASI K-NEAREST NEIGHBOR UNTUK KLASIFIKASI BUNGA DENGAN EKSTRAKSI FITUR WARNA RGB,” *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, vol. 7, no. 6, pp. 1129–1135, 2019, doi: 10.25126/jtiik.202072608.
- [32] J. Lee and K. il Hwang, “YOLO with adaptive frame control for real-time object detection applications,” *Multimed Tools Appl*, vol. 81, no. 25, pp. 36375–36396, Oct. 2022, doi: 10.1007/s11042-021-11480-0.
- [33] M. Durve, F. Bonaccorso, A. Montessori, M. Lauricella, A. Tiribocchi, and S. Succi, “Tracking droplets in soft granular flows with deep learning techniques,” *The European Physical Journal Plus*, vol. 136, no. 8, p. 864, Aug. 2021, doi: 10.1140/epjp/s13360-021-01849-3.
- [34] D. Gros *et al.*, “Anaphora Resolution,” *International Journal of Information Retrieval Research*, vol. 8, no. 3, pp. 33–45, Jul. 2018, doi: 10.4018/IJIRR.2018070103.
- [35] M. Zhu, “Recall, Precision and Average Precision,” Canada, Aug. 2004. Accessed: Sep. 14, 2023. [Online]. Available: https://www.researchgate.net/publication/228874142_Recall_precision_and_average_precision

Analisis Sentimen di Twitter: Mengungkap Persepsi dan Emosi Publik Seputar Konflik Palestina-Israel

Rizki Azhar¹, Muhammad Farid Wijayanto²

^{1,2}Teknik Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Nusantara PGRI Kediri

E-mail: *¹rizkiazhar024@gmail.com, ²faridwijayaaa@gmail.com

Abstrak – Konflik yang terjadi antara Palestina dan Israel masih menjadi konflik yang belum terselesaikan hingga saat ini. Konflik dimulai dengan berdirinya negara Israel yang diakui PBB pada tahun 1948 di tanah yang sebelumnya dikuasai Inggris, dan konflik dimulai dengan serangan negara-negara Arab terhadap Israel dan memprovokasi perang. Mayoritas penduduk Palestina beragama Islam dan mendapat dukungan dari negara-negara Timur Tengah dan mayoritas penduduknya beragama Islam. Terlebih lagi, Israel terus berjuang dalam konflik ini dengan dukungan negara-negara Barat. Opini publik merupakan sumber informasi penting untuk membentuk opini mengenai konflik ini. Analisis sentimen terhadap data media sosial seperti tweet adalah metode yang berguna untuk mengubah sejumlah besar data menjadi informasi yang berguna. Analisis sentimen menggunakan pemrosesan bahasa alami (NLP) membantu memahami emosi dan mengkategorikannya sebagai positif atau negatif, memberikan informasi penting untuk evaluasi merek, evaluasi, dan penyaringan opini yang mendukung perusahaan dan masyarakat. Indonesia adalah negara yang menjunjung tinggi perdamaian sesuai pembukaan UUD 1945, tetapi masyarakat terus memperdebatkan konflik antara Palestina dan Israel.

Kata Kunci — Analisis Sentimen, Israel, Konflik, NLP, Palestine

1. PENDAHULUAN

Palestina dan Israel merupakan dua negara yang masih memiliki konflik perang yang belum terselesaikan hingga saat ini. Negara Israel berdiri pada tahun 1948 ketika perserikatan bangsa-bangsa (PBB) telah menyetujui terkait pendiriannya di tanah Palestina yang awalnya di bawah naungan Inggris. Setelah pendirian negara Israel negara-negara Arab yang terdiri dari Lebanon, Suriah, Yordania, Mesir, dan Irak langsung menyerang negara Israel yang berdiri ketika disahkan oleh Perserikatan Bangsa-Bangsa (PBB). Palestina yang mayoritas penganut agama Islam mendapat dukungan dari negara-negara Timur Tengah dan negara yang mayoritas penduduknya beragama Islam. Disisi lain Israel yang masih dalam naungan Inggris mendapat dukungan dari negara-negara barat, dari sini banyak polemik yang terjadi dalam konflik setelah berlangsungnya peperangan sampai saat ini [1]. Peperangan yang tak kunjung selesai saat ini menimbulkan banyak opini yang beragam dari masyarakat. Dengan opini tersebut, masyarakat juga memiliki rasa keberpihakan terhadap negara konflik. Di Indonesia perdebatan ini masih saja terjadi, pada hakikatnya Indonesia adalah negara yang menjunjung tinggi asas perdamaian seperti yang tertuang pada Pembukaan UUD 1945 alinea pertama yang berbunyi penjajahan diatas dunia harus dihapuskan.

Pandangan dan tanggapan dari masyarakat adalah sumber daya berharga. Melalui tweet yang dipublikasikan oleh masyarakat, para peneliti memiliki kesempatan untuk mendapatkan gambaran menyeluruh tentang pandangan pengguna melalui analisis mendalam terhadap setiap tweet yang dibagikan. Di tengah tanggapan masyarakat yang beraneka ragam, terutama adanya sikap kebencian terhadap salah satu pihak, hal ini dapat memperdalam jurang perpecahan dalam masyarakat. Kondisi ini disebabkan oleh isu-isu sensitif yang sering kali terkait dengan konflik atau peperangan yang sedang terjadi. Namun, karena volume data yang besar, diperlukan suatu metode untuk mengubah data tersebut menjadi informasi yang berarti. Salah satu metode yang digunakan untuk hal ini adalah analisis sentimen. Analisis sentimen adalah proses menganalisis teks digital untuk menentukan apakah pesan teks bersifat positif, negatif, atau netral [2]. Saat ini, analisis sentimen menjadi topik hangat yang banyak diminati oleh peneliti dengan tujuan menyediakan informasi dari dataset yang tidak terstruktur. Penelitian ini merupakan bagian dari penelitian yang disebut *finer grained sentiment analysis* yang menganalisis teks komentar [3].

Penelitian yang dilakukan Rian Tineges, dkk pada tahun 2020 membahas tentang objek layanan indihome pada twitter. Penelitian tersebut bertujuan untuk mendapatkan model klasifikasi sentimen menggunakan SVM, dan untuk mengetahui seberapa besar akurasi yang dihasilkan oleh metode SVM yang diterapkan pada analisis sentimen, serta untuk mengetahui seberapa puas pengguna layanan Indihome berdasarkan Twitter. Setelah dilakukan pengujian dengan metode SVM hasilnya adalah *accuracy* 87%, *precision* 86%, *recall* 95%, *error rate* 13%, dan *f1-score* 90% [4].

Penelitian lain tentang analisis sentimen dilakukan oleh Aldiansyah Putra, dkk pada tahun 2021 dengan sebuah objek kebijakan PPKM pada media sosial twitter. Penelitian tersebut bertujuan untuk mengetahui

pendapat Masyarakat terhadap kebijakan pemerintah tentang PPKM, Dari hasil pemodelan menggunakan 3000 data diketahui akurasi adalah 64%, tidak terlalu tinggi jika dibandingkan dengan analisis sentimen yang telah dilakukan di penelitian-penelitian sebelumnya. Namun, algoritma SVM masih mampu mengenali kata-kata penting untuk sentimen negatif seperti ‘tolak PPKM’, ‘pelan-pelan’ dan ‘rumah makan’. Metode SVM dapat memprediksi 2000 tweet menjadi 57%-60% berlabel positif dan sisanya negatif, dimana hasilnya tidak jauh berbeda dengan data training [5].

Penelitian analisis sentimen terhadap aplikasi Gojek pada Twitter oleh Nur Fitriyah, dkk tahun 2020 bertujuan menggali respon pengguna terhadap layanan. Hasil pengujian menunjukkan kernel RBF mencapai akurasi tertinggi dalam pelabelan manual (79.19%, κ 16.52%) dan *sentiment scoring* (79.19%, κ 21%). Performa model klasifikasi dapat diukur dari nilai *overall accuracy* dan *kappa accuracy* yang semakin tinggi, menandakan kinerja yang lebih baik [6].

Berdasarkan latar belakang dan penelitian sebelumnya, penelitian ini merinci rumusan masalah utama, yaitu bagaimana mengembangkan model klasifikasi sentimen menggunakan *Support Vector Machine* (SVM). Dan fokus penelitian ini terletak pada data yang terkait dengan kata kunci "Palestina" dan "Israel" dalam bahasa Indonesia di platform Twitter, dengan penerapan metode klasifikasi SVM. Adapun tujuan utama penelitian ini adalah untuk melibatkan pemodelan klasifikasi sentimen menggunakan SVM, evaluasi akurasi metode SVM, serta pemahaman opini masyarakat terkait konflik Palestina dan Israel berdasarkan analisis Twitter.

2. METODE PENELITIAN

2.1 Diagram Alur Proses Penelitian

Terdapat beberapa tahapan dalam penelitian kali ini diantaranya tahap pengumpulan data, *text preprocessing*, *vektorisasi*, pembagian data, pemodelan metode klasifikasi, dan terakhir tahap evaluasi, sebagaimana diilustrasikan pada Gambar 1.



Gambar 1. Proses Alur Diagram

Penelitian berupa kumpulan fakta atau data yang didapat dari suatu objek penelitian. Dalam hal ini objek penelitian yang dimaksud adalah data tweet yang mengandung kata “hancurkan gaza”, “Kemenangan IDF”, “Selamatkan warga Palestina”, “Palestina Israel Gaza Hamas IDF ” yang kemudian datanya akan dijadikan sebagai sumber data. Dalam melakukan penelitian ini diperlukan alur proses penelitian, agar dalam proses penelitian berjalan sesuai dengan rencana.

2.2 Pengumpulan Data

Tahap pengumpulan data dengan melakukan crawling melalui API Twitter dengan menggunakan modul tweepy pada Bahasa pemrograman Python. Data yang di crawl adalah data berdasar kumpulan tweet dengan kata kunci “hancurkan gaza”, ”Kemenangan IDF”, “Selamatkan warga Palestina”, “Palestina Israel Gaza Hamas IDF”. Setelah itu dataset tweet yang masih mentah disimpan dalam bentuk CSV (*Comma Separated Values*) yang nantinya akan diproses lebih lanjut.

2.3 Text Preprocessing

Proses pengumpulan data dikarenakan dataset tersebut termasuk dalam *unstructured* data (data tidak terstruktur). Sebelum dilakukan analisis lebih lanjut, terlebih dahulu dataset melalui proses *text preprocessing* untuk menghilangkan serta mengatasi *noisy* data agar hasil perhitungan optimal[4]. Adapun alur proses text preprocessing antara lain, *case folding*, *filtering*, *tokenizing*, dan *stopword removal*. Setelah data bersih maka dilakukan proses pelabelan secara manual. Berikut adalah alur diagram dari *text preprocessing*.



Gambar 2. Diagram Alur *Text Preprocessing*

1. *Cleansing*

Proses *cleansing* dalam sistem berguna untuk membersihkan data tweet seperti angka, tanda baca, link, *hashtag*, *mention*, *emoticon* dan menghasilkan kata yang akan diolah pada proses selanjutnya. Hasil proses *cleansing* adalah menghilangkan kata dari karakter yang tidak ada artinya dan akan diganti oleh karakter spasi [7].

2. *Case Folding*

Case folding merupakan proses yang dilakukan untuk mengubah setiap kata yang ada di dalam dataset menjadi huruf kecil menggunakan fungsi *lowercase* [8].

3. *Filtering*

Filtering merupakan tahapan yang bertujuan untuk menghilangkan kata umum yang biasa sering muncul dalam jumlah banyak dan tidak memiliki makna menggunakan algoritma *stoplist* (membuang kata yang kurang penting) atau *wordlist* (menyimpan kata penting).

4. *Tokenization*

Karena Twitter memungkinkan pengguna untuk menulis teks pendek saja, tweet sering kali hadir dengan tata bahasa dan singkatan khusus, sehingga pengguna dapat menyampaikan pesan dengan kata-kata sesedikit mungkin. *Tokenizer* tradisional dan *POS tagger* mungkin tidak memadai untuk pra-pemrosesan tweet dan ada kebutuhan untuk alternatif yang dapat mengenali token, *hashtag*, *emotikon*, dan URL tweet [9].

5. *Stopword Removal*

Proses *stopword removal* ini digunakan untuk menghilangkan kata yang tidak berpengaruh dalam proses sentimen. Kata dari data tweet dibandingkan dengan kata yang terdapat dalam *database stopwords*, hasil dari proses ini adalah menghilangkan kata yang terdeteksi sama pada kata *stopword* yang ada pada *database*.

2.4 Pembobotan Kata TF-IDF

Pada pembobotan kata dilakukan proses pemberian nilai atau bobot terhadap setiap kata pada tweet yang telah melewati tahap *preprocessing*. Metode yang digunakan dalam pemberian bobot pada *term* yaitu TF-IDF. Pembobotan ini bertujuan untuk memberikan nilai kepada sebuah *term* dimana nilai dari *term* tersebut akan dijadikan sebagai *input* pada proses klasifikasi.

$$TF - IDF(d, t) = TD(d, t) * IDF(t) \dots\dots\dots(1)$$

Dimana :

$$TF(d, t) = \frac{\text{jumlah kata } t \text{ pada dokumen } d}{\text{total kata pada dokumen } d} \dots\dots\dots(2)$$

$$IDF(t) = \frac{j\text{total dokumen}}{\text{jumlah dokumen yang mengandung kata } t} \dots\dots\dots(3)$$

Dan:

$t = \text{ kata}$

$d = \text{ dokumen}$

2.5 Klasifikasi SVM

Setelah dilakukan proses pembobotan TF IDF, selanjutnya masuk ke tahap analisis data menggunakan algoritma yang telah ditentukan yaitu *support vector machine*. Pada tahap ini data dibagi menjadi dua bagian, yaitu *testing* dan *training*. Dalam tahap ini juga akan dilakukan proses klasifikasi untuk menentukan label positif dan negatif dari *term* yang sudah melewati proses pembobotan TF IDF untuk mendapatkan nilai akurasi dari algoritma *Support Vector Machine*.

2.6 Evaluasi

Evaluasi dalam penelitian ini bertujuan untuk mengetahui kinerja atau performa dari model yang diusulkan. Metode yang digunakan untuk evaluasi pada penelitian ini adalah *confusion matrix*. *Confusion matrix* merupakan sebuah tabel yang memberikan informasi perbandingan hasil klasifikasi yang dilakukan oleh sistem (prediksi) dengan hasil klasifikasi yang sebenarnya. Dalam *confusion matrix*, terdapat 4 istilah nilai hasil klasifikasi yang nantinya akan digunakan untuk menghitung nilai akurasi, presisi, recall, dan f1-score, yaitu *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)*.

$$Accuracy (A) = \frac{(TP+FN)}{(TP+FP+FN+TN)} \dots\dots\dots(4)$$

$$Precision (P) = \frac{(TP)}{(TP+FP)} \dots\dots\dots(5)$$

$$Recall (R) = \frac{(TP)}{(TP+FN)} \dots\dots\dots(6)$$

3. HASIL DAN PEMBAHASAN

3.1 Preprocessing

1. Cleansing

Proses *cleansing* dalam sistem berguna untuk membersihkan data tweet seperti angka, tanda baca, *link*, *hashtag*, *mention*, *emoticon* dan menghasilkan kata yang akan diolah pada proses selanjutnya. Hasil proses *cleansing* adalah menghilangkan kata dari karakter yang tidak ada artinya dan akan diganti oleh karakter spasi. Hasil dari proses *cleansing* disajikan pada Tabel 1 berikut.

Tabel 1. Tabel Hasil *Cleansing* Data Dan *Case Folding*

<i>Full Text</i>	<i>Cleansing dan Case Folding</i>
Gaza (2005) IDF memaksa semua warga Israel keluar dari Gaza dan memberikan Gaza 100% ke Palestina. Hari ini (7 Okt 2023) 75% warga Gaza setuju dan merayakan pembantaian Hamas ke	gaza idf memaksa warga israel gaza gaza palestina okt warga gaza setuju merayakan pembantaian hama warga sipil israel membumi ratakan wilayah daerah kosong

<i>Full Text</i>	<i>Cleansing dan Case Folding</i>
<p>warga sipil Israel Waktunya membumi ratakan wilayah ini dan dipastikan menjadi daerah kosong https://t.co/ApuGZgzlTZ</p> <p>@netanyahu Anda mengakui bahwa Israel teroris dan Anda seorang komando yang meminta idf menyerang Gaza, alih alih ingin menyerang Hamas tapi tak kena. Faktanya Anda justru membunuh warga sipil Palestina. Ingat! Anda tidak hidup selamanya bapak tua #StopsilencingPalestine #FreePalestine</p> <p>Penyanyi Zionis Israel Rinat Bar tampil menghibur tentara IDF dan menyanyikan "Ya Yahya Sinwar, semoga kamu mati besok, bersama pejuang Palestina dan seluruh Rakyat Gaza" Mereka gembira atas pembantaian Gaza. Nb: Yahya Sinwar adalah pemimpin Hamas di Gaza @ireallyhateyou https://t.co/aA3lfkvk8N</p>	<p>mengakui israel teroris komando idf menyerang gaza alih alih menyerang hama kena faktanya membunuh warga sipil palestina hidup tua</p> <p>penyanyi zionis israel rinat bar tampil menghibur tentara idf menyanyikan ya yahya sinwar semoga mati besok pejuang palestina rakyat gaza gembira pembantaian gaza nb yahya sinwar pemimpin hama gaza</p>

2. Case Folding

Case folding dilakukan untuk mengubah seluruh ukuran huruf pada kata menjadi suatu bentuk ukuran huruf yang sama. Karena tidak semua tweet konsisten dalam penggunaan ukuran huruf. Fungsi *case folding* ini akan mengubah semua huruf menjadi *lowercase* atau huruf kecil. Hasil dari proses *case folding* disajikan pada Tabel 1 diatas.

3. Tokenization

Proses *tokenizing* dilakukan untuk memisahkan deretan kata di dalam kalimat, paragraf atau halaman menjadi token atau potongan kata tunggal. Pada saat bersamaan, *tokenizing* juga membuang karakter selain huruf seperti tanda baca. Hasil dari proses *tokenization* disajikan dalam Tabel 3 berikut.

Tabel 3. Tabel Hasil Tokenization Data

<i>Full Text</i>	<i>Tokenization</i>
gaza idf memaksa warga israel gaza gaza palestina	“gaza”, “idf”, “memaksa”, “warga”, “israel” “gaza”, “gaza”, “palestina”

4. Stopword Removal

Proses *stopword removal* ini digunakan untuk menghilangkan kata yang tidak berpengaruh dalam proses sentimen. Kata dari data tweet dibandingkan dengan kata yang terdapat dalam *database stopwords*, hasil dari proses ini adalah menghilangkan kata yang terdeteksi sama pada kata *stopword* yang ada pada *database*. Hasil dari proses *stopword removal* dapat dilihat pada Tabel 4 berikut.

Tabel 4. Tabel Hasil Stopword Removal Data

<i>Full Text</i>	<i>Stopword Removal</i>
Gaza (2005) IDF memaksa semua warga Israel keluar dari Gaza dan memberikan Gaza 100% ke Palestina. Hari ini (7 Okt 2023) 75% warga Gaza setuju dan merayakan pembantaian Hamas ke warga sipil Israel Waktunya membumi ratakan wilayah ini dan dipastikan menjadi daerah kosong https://t.co/ApuGZgzlTZ	gaza idf memaksa warga israel gaza palestina okt setuju merayakan pembantaian hama sipil membumi ratakan wilayah daerah kosong

5. Stemming

Proses terakhir dalam *preprocessing* adalah *stemming*, fungsi dari *stemming* untuk mengubah kata yang berlebihan menjadi kata dasar agar memudahkan dalam proses pembobotan nantinya. Fungsi ini menggunakan algoritma Nazief dan Adriani yaitu library Sastrawi. Hasil dari proses *stemming* dapat dilihat pada Tabel 5 berikut.

Tabel 5. Tabel Hasil *Stemming Data*

Full Text	Stemming
gaza idf memaksa warga israel gaza gaza palestina	gaza idf paksa warga israel gaza palestina
mengakui israel teroris komando idf menyerang	aku israel terror komando idf serang gaza

3.2 Pembobotan TF – IDF

Pembobotan ini bertujuan untuk memberikan nilai kepada sebuah *term* dimana nilai dari *term* tersebut akan dijadikan sebagai *input* pada proses klasifikasi. Hasil dari proses pembobotan menggunakan TF-IDF didapatkan nilai sentiment pada setiap kalimat, dan disajikan pada Tabel 9 berikut.

Tabel 6. Tabel hasil pembobotan data

Full Text	Sentiment
gaza idf memaksa warga israel gaza gaza palestina okt warga gaza setuju merayakan pembantaian hama warga sipil israel membumi ratakan wilayah daerah kosong	Negatif
mengakui israel teroris komando idf menyerang gaza alih alih menyerang hama kena faktanya membunuh warga sipil palestina hidup tua	Negatif

3.3 Klasifikasi dan Pengujian

Dalam proses klasifikasi *Support Vector Machine* dibutuhkan bobot setiap kata yang ada pada data *training*. Nilai bobot setiap *term* ini didapat dari proses pembobotan TF-IDF yang akan digunakan dalam pembuatan model dari algoritma *Support Vector Machine*. Proses dari klasifikasi *Support Vector Machine* disajikan melalui *source code* pada Gambar 4 berikut.

```
from sklearn import svm
from sklearn.model_selection import cross_val_score

clf = svm.SVC(kernel="linear")

cross_val_score(clf, X_train, y_train, cv=10)
```

Gambar 3. Baris kode metode SVM

Hasil dari proses klasifikasi *Support Vector Machine* berupa *validation score*. dan disajikan pada Tabel 7 berikut.

Tabel 7. Tabel validasi score

0.91549296	0.92957746	0.90140845	0.83098592	0.90140845
0.85915493	0.88732394	0.90140845	0.88732394	0.90140845

3.4 Evaluasi

Pengujian kinerja dari *Support Vector Machine* dilakukan dengan metode *confusion matrix* berdasarkan data *testing* yang akan diproses oleh model klasifikasi *Support Vector Machine* yang telah dibuat sebelumnya. Metode *confusion matrix* menghasilkan empat nilai yaitu *True Positive*, *True Negative*, *False Positive* dan *False Negative*. Selanjutnya dilakukan perhitungan mulai dari *accuracy*, *precision* dan *recall*. Berikut adalah hasil dari

pengujian dengan data training 90% dan data testing 10%. Hasil dari evaluasi algoritma *Support Vector Machine* disajikan pada Tabel 11 berikut.

Tabel 10. Tabel hasil Evaluasi

Evaluasi	Hasil
F1 Score	0.8695652173913044
Accuracy	0.8860759493670886
Precision	0.8953758718677345
Recall	0.8860759493670886

Tabel 10 menunjukkan bahwa hasil pengujian untuk metode klasifikasi Support Vector Machine mendapatkan F1 Score 89%, Accuracy 88%, Precision 89%, dan Recall 88%.

4. SIMPULAN

Berdasarkan hasil dan pembahasan dapat ditarik kesimpulan mengenai analisis sentimen terhadap konflik Palestina dan Israel di Twitter menggunakan metode Support Vector Machine sebagai berikut:

- a. Penerapan metode Support Vector Machine untuk analisis sentimen terhadap konflik Palestina dan Israel, Peneliti memperoleh hasil evaluasi dengan menggunakan Confusion Matrix. Akurasi sebesar 88%, ketepatan antara hasil prediksi dengan data sebenarnya (precision) sebesar 89%, tingkat keberhasilan sistem dalam memprediksi sebuah data (recall) sebesar 88%, serta tingkat kesalahan semua data yang diprediksi (error rate) sebesar 10%. Nilai perbandingan rata-rata precision dan recall (f1-score) adalah sebesar 90%.
- b. Berdasarkan hasil dan pembahasan menunjukkan bahwa metode SVM dapat diterapkan untuk analisis sentimen data Twitter berbahasa Indonesia.

5. SARAN

Penelitian ini jalankan dengan metode SVM dan masih memerlukan peningkatan. Oleh karena itu, untuk penelitian mendatang, pertimbangkan untuk mempertimbangkan penggunaan metode lain yang dapat memberikan hasil yang lebih optimal.

DAFTAR PUSTAKA

- [1] Muhammad Jamaluddin and Erik Ilham Habibillah, "Pengaruh Kebijakan Perserikatan Bangsa-Bangsa (Pbb) Dalam Konflik Palestina-Israel," *J. Tapis Teropong Aspir. Polit. Islam*, vol. 19, no. July, pp. 1–23, 2020.
- [2] Hendrastuty, N., Isnain, A. R., & Rahmadhani, A. Y. (2021). Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine. *Jurnal Informatika*.
- [3] A. Surahman, "PENGEMBANGAN MARKET SEGMENTASI UNTUK MENCAPAI KEUNGGULAN BERSAING PADA E-MARKETPLACE," *J. Komput. dan Inform.*, vol. 15, no. 1, pp. 118–126, 2020.
- [4] R. Tingegs, A. Triayudi, and I. D. Sholihati, "Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM)," *J. Media Inform. Budidarma*, vol. 4, no. 3, p. 650, 2020, doi: 10.30865/mib.v4i3.2181.
- [5] A. Putra, D. Haeirudin, H. Khairunnisa, and R. Latifah, "Analisis Sentimen Masyarakat Terhadap Kebijakan PPKM Pada Media Sosial Twitter Menggunakan Algoritma SVM," *Semin. Nas. Sains dan Teknol. 2021*, no. November, pp. 1–6, 2021.
- [6] N. Fitriyah, B. Warsito, and D. A. I. Maruddani, "Analisis Sentimen Gojek Pada Media Sosial Twitter Dengan Klasifikasi Support Vector Machine (Svm)," *J. Gaussian*, vol. 9, no. 3, pp. 376–390, 2020, doi: 10.14710/j.gauss.v9i3.28932.
- [7] F. D. Ananda and Y. Pristyanto, "Analisis Sentimen Pengguna Twitter Terhadap Layanan Internet Provider Menggunakan Algoritma Support Vector Machine," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 20, no. 2, pp. 407–416, 2021, doi: 10.30812/matrik.v20i2.1130.
- [8] M. I. Fikri, T. S. Sabrila, Y. Azhar, and U. M. Malang, "Comparison of the Naïve Bayes Method and Support Vector Machine on Twitter Sentiment Analysis," *SMATIKA J. STIKI Inform. J.*, vol. 10, no. 2, pp. 71–76, 2020.
- [9] M. A. Reinfeld, "a C Ase S Tudy of the Grp 5000," *Technology*, vol. 5, no. 3123, pp. 548–554, 2011.

KLASIFIKASI DAN PENGENALAN POLA PENYAKIT CABAI DENGAN METODE CNN (*Convolution Neural Network*)

Bella Nurbuana Tri Cahya Ningrum¹, Erlina Nasrinatun Ni'mah², Miranda Putri Arifin³,
Made Ayu Dusea Widya Dara⁴

^{1,2,3,4}Teknik Informatika, Fakultas Teknik, Universitas Nusantara PGRI Kediri

E-mail: *¹bellanurbuana@gmail.com, ²erlinann15@gmail.com, ³mirandaputriarifin@gmail.com,
⁴madedara@gmail.com

Abstrak – Penelitian ini mengusulkan sistem klasifikasi menggunakan *Convolutional Neural Network* (CNN) untuk identifikasi penyakit pada buah cabai. Dalam pengujian dengan 162 citra, algoritma mencapai akurasi 100%. Namun, saat diuji dengan 236 citra, termasuk data penyakit lalat yang belum dikenali sistem, akurasi menurun drastis menjadi 45%. Penyakit seperti bercak abu, virus kuning cabai, dan layu bakteri telah berhasil diidentifikasi dengan presisi tinggi. Meskipun demikian, tantangan muncul ketika data penyakit baru diuji, menyebabkan penurunan signifikan dalam performa sistem. Hasil ini menggarisbawahi pentingnya pengembangan model yang responsif terhadap penyakit baru serta kebutuhan akan dataset yang lebih luas dan representatif. Sistem ini menawarkan potensi besar dalam mendukung petani cabai dalam mendeteksi penyakit, namun membutuhkan peningkatan agar dapat mengenali variasi penyakit yang lebih luas dan tetap handal dalam situasi lingkungan.

Kata Kunci — Cabai, CNN, Klasifikasi, Performa

1. PENDAHULUAN

Deteksi penyakit tanaman merupakan hal terpenting dalam memelihara dan merawat tanaman. Penyakit yang tidak bisa terdeteksi dan dibiarkan begitu saja akan mengakibatkan kerusakan pada tanaman. Kerusakan tersebut akan berdampak bagi ekonomi dan petani. Beberapa penelitian telah dilakukan untuk mendapat pengetahuan tentang berbagai penyakit tanaman, sehingga perawatan pada tanaman menjadi lebih mudah dengan mengetahui ciri-ciri dari setiap penyakit, dan cara pengobatan tanaman, hingga cara pencegahan penyakit [1].

Perkembangan teknologi saat ini telah mendorong cara untuk mendeteksi penyakit menggunakan suatu sistem yang dapat dijalankan melalui *smartphone* atau komputer. Sistem deteksi ini sangat direkomendasikan karena hasil akurasi yang dihasilkan cukup akurat. Banyak penelitian yang menggunakan metode *Deep Learning* sebagai pondasi untuk membuat sistem. Banyak penelitian yang dilakukandengan memanfaatkan sistem klasifikasi untuk mendiagnosa suatu penyakit dengan menggunakan objek gambar sebagai masukannya dan sistem ini dibangun untuk menggantikan peran seorang pakar. Dengan menggunakan sistem klasifikasi ini dapat mempermudah dalam mendiagnosa penyakit pada tanaman cabai.

Adapun beberapa penelitian yang dilakukan oleh peneliti sebelumnya, yaitu penelitian yang dilakukan oleh Uly Nuhanatika yang berjudul “Penentuan Tingkat Kematangan Cabe Rawit (*Capsicum frutescens* L.) Berdasarkan *Gray Level Co-Occurrence Matrix*” bertujuan untuk klasifikasi tingkat kematangan cabai rawit menggunakan ekstraksi fitur warna dan tekstur. Ekstraksi fitur warna diambil dari nilai mean saturasi, sedangkan ekstraksi fitur tekstur menggunakan nilai fitur *Gray level Co-Occurrence Matrix* (GLCM). Menghasilkan pelatihan sistem yang mampu mengklasifikasi tingkat kematangan cabai rawit dengan akurasi sebesar 81,4% dan akurasi proses pengujian cabai rawit sebesar 74,2% [2]. Selanjutnya penelitian yang dilakukan oleh Joelyan Vicky, Frisca Ayu dan Bagas Julianto yang berjudul “Implementasi Pendeteksi Penyakit pada Daun Alpukat Menggunakan Metode CNN” dimana proses implementasi yang dilakukan dengan model dan testing data pada data daun alpukat yang terkena penyakit mempunyai hasil yang baik dengan menggunakan metode CNN mampu mendeteksi penyakit pada daun alpukat dengan akurasi sebesar 80% [3]. Selanjutnya penelitian yang dilakukan oleh Alya Zalvadila, Purnawansyah, Lukman Syafie, Herdianti Darwis yang berjudul “Klasifikasi Penyakit Tanaman Bawang Merah Menggunakan Metode SVM dan CNN” penelitian ini menggunakan CNN dan SVM dengan kernel RBF, linear, dan polynominal yang memberikan nilai accuracy, precision, recall, dan F1 score yang masing-masing bernilai 100%. SVM diawali dengan proses ekstraksi fitur menggunakan GLCM, apabila SVM menggunakan kernel sigmoid akan menghasilkan performa yang jauh lebih buruk [4].

Petani sering menghadapi tantangan dalam mengidentifikasi secara akurat penyakit yang menyerang tanaman cabai. Keterbatasan pengetahuan dan pengalaman petani dapat menyebabkan kesulitan dalam

membedakan gejala penyakit yang serupa. Oleh karena itu, kebutuhan akan sistem klasifikasi penyakit tanaman cabai menjadi penting. Dengan adanya sistem ini, petani dapat dengan cepat dan akurat mengidentifikasi jenis penyakit yang menyerang tanaman cabai mereka. Hal ini tidak hanya memungkinkan tindakan pencegahan yang lebih dini dan tepat, tetapi juga membantu dalam optimalisasi penggunaan sumber daya, mengurangi kerugian hasil panen, dan secara keseluruhan meningkatkan produktivitas pertanian dengan cara yang lebih efisien dan berkelanjutan.

Berdasarkan permasalahan yang telah disebutkan diatas maka peneliti membuat penelitian tentang klasifikasi dan pengolahan citra penyakit pada tanaman cabai menggunakan metode CNN (*Convolution Neural Network*) dengan proses deteksi warna yang diawali dengan konversi ruang warna citra RGB (*Red, Green, Blue*) menjadi HSV (*Hue, Saturation, Value*). Pada proses klasifikasi warna dilakukan berdasarkan pengelompokan nilai Hue, selanjutnya proses thresholding yang menghasilkan citra biner. Setelah itu dilakukan proses ekstrasi ciri morfologi dari citra biner berdasarkan parameter eccentricity dan metric. Hasil penelitian diharapkan mampu membedakan jenis penyakit pada tanaman cabai.

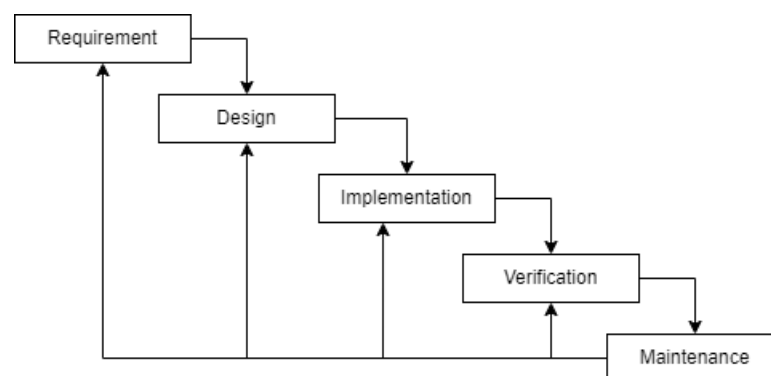
2. METODE PENELITIAN

2.1 Pengumpulan Data

Metode yang digunakan untuk pengumpulan data pada penelitian ini adalah studi pustaka, yaitu dalam pencarian teori dan pengumpulan data, peneliti menyimpulkan informasi dari kepustakaan yang berhubungan dengan penelitian tersebut. Sumber yang digunakan oleh penulis yaitu dari buku, jurnal penelitian yang sebelumnya yang berkaitan dengan topik penelitian tersebut, serta sumber lainnya yang di dapatkan melalui internet.

2.2 Metode Waterfall

Dalam perancangan sistem klasifikasi ini menggunakan metode *waterfall*. Metode ini terdiri dari 5 proses dalam perancangan sistem yaitu *requirement*, *design*, *implementation*, *verification*, dan *maintenance*, Seperti yang terlihat pada gambar 1.



Gambar 1. Metode Waterfall

2.3 Pengolahan Citra

a. Segmentasi Warna RGB

Segmentasi Warna RGB adalah proses pemisahan objek pada citra berwarna menggunakan model RGB (*Red, Green, Blue*). Pada setiap piksel dalam model warna RGB memiliki 3 Saluran warna: merah, hijau dan biru. Segmentasi warna RGB bertujuan untuk mengidentifikasi dan memisahkan objek dalam citra berdasarkan perbedaan warna dalam saluran warna RGB.

b. Ruang warna HSV

Ruang warna HSV merupakan model warna yang digunakan untuk mempresentasikan warna dalam citra. Dalam ruang warna HSV ini memisahkan warna menjadi 3 komponen utama, yaitu Hue (nuansa), Saturation (saturasi), dan Value (nilai).

a. Hue mempresentasikan jenis warna yang ada dalam citra dengan menunjukkan warna yang sebenarnya, yaitu merah, hijau, biru, kuning, cyan, magenta. Hue memiliki rentang angka dari 0 hingga 360 derajat.

b. Saturasi mengukur sejauh mana warna itu jenuh atau tidak. Nilai saturasi yang tinggi menunjukkan warna yang jenuh dan kuat, sementara nilai saturasi rendah menghasilkan warna yang pudar atau abu-abu.

- c. Value menggambarkan kecerahan dari warna yang nilai nya berkisar dai 0 sampai 100%, 0 adalah hitam dan 100% adalah putih

Ruang warna HSV diperoleh dari ruang warna RGB dengan persamaan 1, 2, 3, 4 dan 5:

$$r = \frac{R}{R+G+B}, g = \frac{G}{R+G+B}, b = \frac{B}{R+G+B} \dots\dots\dots(1)$$

$$V = \max (r, g, b) \dots\dots\dots(2)$$

$$S = \left\{ \begin{array}{l} 0, \text{ jika } V = 0 \\ 1 - \frac{\min(r,g,b)}{v}, v > 0 \end{array} \right\} \dots\dots\dots(3)$$

$$H = \left\{ \begin{array}{l} 0, \text{ jika } S = 0 \\ \frac{60*(g-b)}{S*V}, \text{ jika } V = r \\ 60 * \left[2 + \frac{b-r}{S*V} \right], \text{ jika } V = g \\ 60 * \left[4 + \frac{r-g}{S*V} \right], \text{ jika } V = b \end{array} \right\} \dots\dots\dots(4)$$

$$H = H + 360 \text{ jika } H < 0 \dots\dots\dots(5)$$

2.4 Algoritma CNN

Convolution Neural Network merupakan model jaringan saraf tiruan yang digunakan untuk mengelola citra dan pengenalan pola. CNN digunakan untuk ekstraksi fitur-fitur visual penting dari citra, hal ini sangat berguna dalam berbagai sistem aplikasi pengenalan objek, klasifikasi citra, dan segmentasi citra [1]. Algoritma CNN dipergunakan untuk Deep Learning dengan mengambil input dan memberikan bobot dan bias yang dapat dipelajari pada fitur dalam dibandingkan dengan klasifikasi algoritma lainnya [5]. Filter metode primitive direkayasa secara manual, dengan pelatihan yang cukup, dengan Langkah – Langkah sebagai berikut :

- (1) Inisialisasi parameter model, langkah yang setara dengan menyuntikkan noise ke dalam model
- (2) for $i = 1, 2, \dots n$ (n adalah nomor epoch)

- a. Performa dengan *forward propagation* :

$\forall i$, hitung nilai prediksi dari a melalui jaringan saraf tiruan : $\hat{y} i \theta$

Evaluasi fungsi dapat dilihat pada persamaan 2 :

$$J(\theta) = \frac{1}{m} \sum_{i=1}^m \delta(\hat{y}_i \theta, y_i) \dots\dots\dots(6)$$

Dimana m adalah ukuran dari data training dan θ adalah model parameter dan (*) fungsi perkalian.

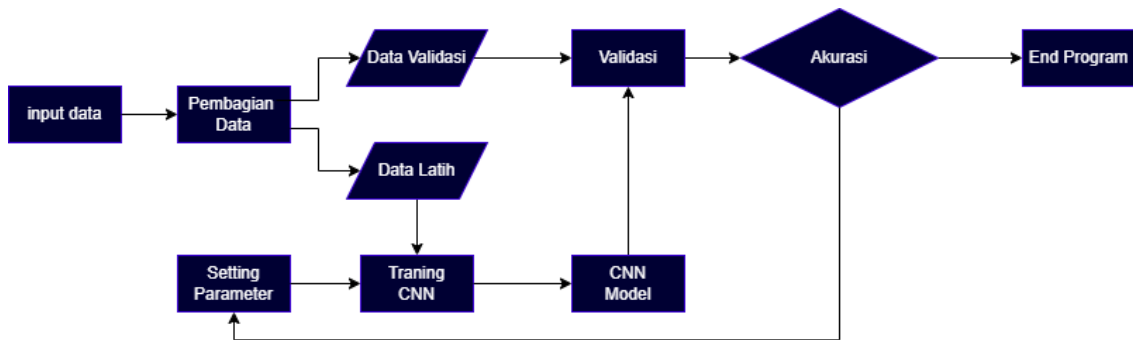
- b. Performa dengan *back propagation* :

Terapkan metode penurunan untuk memperbarui parameter dengan persamaan 7 :

$$\theta =: G () \dots\dots\dots(7)$$

2.5 Flowchart

Pada gambar 2 akan ditampilkan diagram alur proses pengolahan data, pembersihan data, pelatihan dan pengujian, serta evaluasi untuk mencapai akurasi terbaik dalam mengidentifikasi citra penyakit pada cabai mentah dan kering.. [6]



Gambar 2. Flowchart

3. HASIL DAN PEMBAHASAN

3.1 Implementasi Sistem

Dalam penelitian ini, digunakan data citra digital cabai jenis Asmoro yang diambil melalui kamera Hp dengan resolusi sebesar 104 Megapixel. Citra digital tersebut dibagi menjadi dua kategori, yaitu kering dan mentah, dengan total 162 citra dalam sampel, terdiri dari 105 citra kering dan 57 citra mentah. Data ini kemudian terbagi menjadi dua bagian, yakni data training dan data testing, dengan perbandingan 80% untuk data training dan 20% untuk data testing. Pada proses klasifikasi citra digital, digunakan perangkat lunak Jupiter dan Visual Studio Code. Visual Studio Code memanfaatkan beberapa package, antara lain TensorFlow Keras, Cv2, dan Numpy, untuk mendukung proses analisis. Tahapan analisis dalam penelitian ini melibatkan langkah-langkah sebagai berikut:

1. *Import Library:*

Import library yang diperlukan seperti *train_test_split* untuk membagi data menjadi set pelatihan dan pengujian, *LabelEncoder* untuk mengubah label menjadi bentuk biner, dan beberapa komponen dari *TensorFlow* untuk membangun model *neural network*.

```
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from sklearn.metrics import classification_report
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import InputLayer, Flatten, Dense, Conv2D, MaxPool2D, Dropout
from tensorflow.keras.optimizers import Adam
import numpy as np
import cv2
import glob
import matplotlib.pyplot as plt
```

Gambar 3. *Import Library*

2. *Load Dataset:*

Definisikan path dataset dan daftar label terkait. Selanjutnya, lakukan iterasi untuk membaca setiap gambar dalam dataset. Selama iterasi, ubah ukuran setiap gambar menjadi (32, 32), dan tambahkan gambar yang telah diubah ukurannya beserta labelnya ke dalam array data dan labels secara bersamaan.



Gambar 4. Dataset Cabai Kering



Gambar 5. Dataset Cabai Mentah

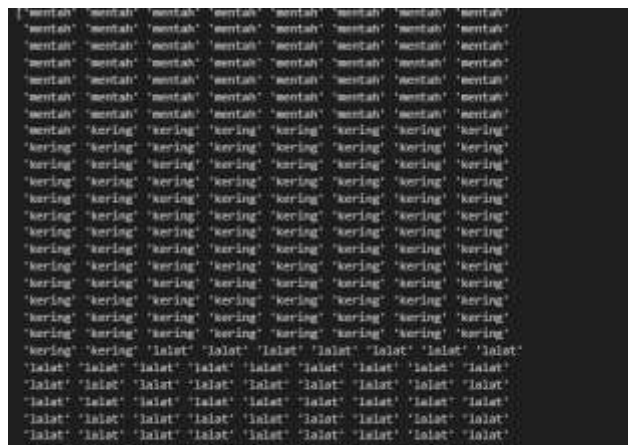
3. *Preprocessing Data:*
Konversi data menjadi *array NumPy* dan normalisasi nilai pikselnya ke rentang [0, 1]. Menggunakan *LabelEncoder* untuk mengubah label menjadi *format biner*.
4. *Pembagian Data:*
Memisahkan data menjadi set pelatihan (x_{train} , y_{train}) dan set pengujian (x_{test} , y_{test}) menggunakan *train_test_split*.
5. *Pembangunan Model:*
Membangun model *sequential* menggunakan *TensorFlow*. Model terdiri dari beberapa layer *Convolutional Neural Network (CNN)* untuk mengekstrak fitur dan beberapa layer *Fully Connected* untuk klasifikasi.
6. *Compile Model:*
Mengompilasi model dengan menggunakan fungsi kerugian (*binary_crossentropy*) dan pengoptimal (*Adam*). *Metric* yang diukur adalah akurasi.
7. *Pelatihan Model:*
Melakukan pelatihan model dengan data pelatihan (x_{train} , y_{train}) selama beberapa *epoch*. Hasil pelatihan disimpan dalam variabel *H*.
8. *Evaluasi Model:*
Menggunakan data pengujian (x_{test} , y_{test}) untuk mengevaluasi model. Mencetak laporan klasifikasi dan menggambar grafik akurasi pada set pelatihan dan pengujian.
9. *Prediksi dengan Data Baru:*
Menggunakan model yang telah dilatih untuk melakukan prediksi pada gambar baru (mentah (16).jpg). Menampilkan hasil prediksi dengan label yang sesuai pada gambar tersebut.
10. *Tampilkan Hasil:*
Menampilkan gambar hasil prediksi dengan menambahkan teks label di atasnya.



Gambar 6. Hasil Gambar Terpilih

3.2 Analisa Sistem

Hasil Analisa sistem menunjukkan kemampuan pembacaan banyak data baru yang diinputkan melalui preprocessing untuk klasifikasi penyakit pada buah cabai [7]. Pengujian ini dilakukan dengan testing dataset yang berisi 74 data citra yang belum dikenali serta data yang sudah dikenali oleh sistem.



Gambar 7. Pengujian Data Baru

Data baru yang berisi penyakit lalat telah dilakukan pengenalan ke dalam sistem, kemudian di hitung akurasi dengan dataset yang sudah dikenali oleh sistem sehingga mendapatkan akurasi yang berubah sedikit lebih turun dari akurasi sebelumnya, yaitu sebagai berikut :

```
# menghitung nilai akurasi model terhadap data test
predictions = model.predict(x_test, batch_size=32)
target = (predictions > 0.5).astype(np.int64) # or np.int32 or other appropriate types
print(classification_report(y_test, target, target_names=label_list))

0.6s
1/2 [=====] - 0s 20ms/step
precision    recall  f1-score   support

 mentah      1.00     0.70     0.82         28
  kering     0.50     1.00     0.67         17
   lalat     0.00     0.00     0.00          11

 accuracy          0.58
 macro avg          0.57
weighted avg          0.58
```

Gambar 8. Akurasi data baru ditambahkan

```
# menghitung nilai akurasi model terhadap data test
predictions = model.predict(x_test, batch_size=17)
target = (predictions > 0.5).astype(np.int64) # or np.int32 or other appropriate type
print(classification_report(y_test, target, target_names=label_list))
```

```
2/2 [-----] - 0s 0kB/step
```

	precision	recall	f1-score	support
mentah	1.00	1.00	1.00	28
kering	1.00	1.00	1.00	13
accuracy			1.00	31
macro avg	1.00	1.00	1.00	31
weighted avg	1.00	1.00	1.00	31

Gambar 9. Akurasi data baru yang sudah dikenali

Berdasarkan perhitungan pada analisa sistem diatas, dapat diambil kesimpulan bahwa akurasi algoritma Convolutional Neural Network dalam klasifikasi penyakit buah cabai ini dengan menggunakan 162 citra yang sudah dikenali oleh sistem memiliki akurasi 1,0 atau 100%, namun jika ditambahkan data baru sebanyak 74 data, sehingga memiliki total jumlah data keseluruhan adalah 236 data, maka akurasi yang terjadi menurun menjadi 0,4 atau 45%.

4. SIMPULAN

Berdasarkan hasil implementasi yang telah dilakukan, kesimpulan yang didapatkan dari hasil penelitian yang telah dilakukan adalah sebagai berikut:

1. Sistem sudah berhasil mengimplementasikan metode CNN untuk mendeteksi penyakit pada cabai.
2. Dengan adanya sistem ini dapat membantu mengetahui sejauh mana performa CNN dalam mengklasifikasikan data buah cabai dalam bentuk citra
3. Model CNN yang telah dibuat berhasil mendapatkan hasil akurasi yang dapat dilihat perbandingannya untuk menjadi pertimbangan penelitian kedepannya.
4. Data yang digunakan masih belum terlalu mempengaruhi perubahan yang lebih spesifik dan kompleks

5. SARAN

Berdasarkan hasil pembuatan sistem klasifikasi dan pengenalan pola untuk mendeteksi penyakit pada cabai menggunakan metode CNN, dalam sistem masih terdapat kesalahan maupun kekurangan. Oleh karena itu ada beberapa saran yang dapat diberikan yaitu bisa mengimplementasikan model tersebut dalam sistem berbasis android ataupun web agar mudah di akses siapa saja.

Hasil uji yang sudah diterapkan dapat membantu kemudahan mengetahui seberapa akurat dan seberapa baik performa dari CNN untuk klasifikasi ini. Kedepannya peneliti dapat membuat perhitungan hasil akurasi yang lebih unggul lagi dari penelitian sebelumnya.

DAFTAR PUSTAKA

[1] A. T. R. Dzaky, “Deteksi Penyakit Tanaman Cabai Menggunakan Metode Convolutional Neural Network,” *e-Proceeding Eng.*, vol. 8, no. 2, pp. 3040–3055, 2021.

[2] Z. E. Fitri, U. Nuhanatika, A. Madjid, and A. M. N. Imron, “Penentuan Tingkat Kematangan Cabe Rawit (*Capsicum frutescens* L.) Berdasarkan Gray Level Co-Occurrence Matrix,” *J. Teknol. Inf. dan Terap.*, vol. 7, no. 1, pp. 1–5, 2020, doi: 10.25047/jtit.v7i1.121.

[3] J. V. P. Putra, F. Ayu, and B. Julianto, “Implementasi Pendeteksi Penyakit pada Daun Alpukat Menggunakan Metode CNN,” *Stain. (Seminar Nas. Teknol. Sains)*, vol. 2, no. 1, pp. 155–162, 2023.

[4] A. Zalvadila, L. Syafie, and H. Darwis, “Klasifikasi Penyakit Tanaman Bawang Merah Menggunakan Metode SVM dan CNN,” *J. Inform. J. Pengemb. IT*, vol. 8, no. 3, pp. 255–260, 2023.

- [5] P. Kalkura, P. R. B, S. K. N, Surya, and M. Ramyashree, “Pest control management system using organic pesticides,” *Glob. Transitions Proc.*, vol. 2, no. 2, pp. 175–180, 2021, doi: 10.1016/j.gltp.2021.08.058.
- [6] D. Irfansyah, M. Mustikasari, and A. Suroso, “Arsitektur Convolutional Neural Network (CNN) Alexnet Untuk Klasifikasi Hama Pada Citra Daun Tanaman Kopi,” *J. Inform. J. Pengemb. IT*, vol. 6, no. 2, pp. 87–92, 2021.
- [7] J. Kecerdasan, T. Informasi, D. H. Firdaus, B. Imran, L. D. Bakti, and E. Suryadi, “Klasifikasi Penyakit Katarak Pada Mata Menggunakan Metode Convolutional Neural Network (Cnn) Berbasis Web Web-Based Classification of Cataract in the Eyes Using Convolutional Neural Network (Cnn) Method,” *J. Kecerdasan Buatan dan Teknol. Inf.*, vol. 1, no. 3, pp. 18–26, 2022.

Implementasi Convolutional Neural Network (CNN) dalam Pengenalan Pola Penulisan Tangan

Salis Nilam Amartama¹, Alvi Nurul Hidayah², Putri Kartika Sari³, Risky Aswi Ramadhani⁴

^{1,2,3,4}Teknik Informatika, Fakultas Teknik, Universitas Nusantara PGRI Kediri

E-mail: ¹salisnilam@gmail.com, ²alvinurulhidayah04@gmail.com, ³kartikasputri840@gmail.com,
⁴riskyaswiramadhani@gmail.com

Abstrak – Era modern ini semakin mendorong manusia untuk mengembangkan suatu inovasi yang lebih baru berteknologi Artificial Intelligence (AI). Pengenalan pola penulisan tangan merupakan bidang penelitian yang terus berkembang dengan pesat, terutama dalam konteks pengembangan sistem otomatisasi deteksi tulisan tangan. Salah satu pendekatan efektif untuk mencapai tujuan adalah menggunakan Convolutional Neural Network (CNN), yang merupakan sebuah jenis arsitektur jaringan saraf yang terbukti efektif dalam memproses data citra. Metodologi penelitian mencakup pengumpulan dataset besar yang mencakup berbagai gaya tulisan tangan, pemrosesan citra, dan pelatihan model CNN. Pada model yang telah dibuat mengalami suatu hal yang disebut *overfitting* atau “terlalu belajar”, dengan kata lain model terlalu mempelajari data training dalam jumlah besar sehingga melakukan pengenalan data-data yang lebih detail dan kurang berguna. Dengan adanya *overfitting* ini, model yang dibuat tetap bisa mengenali pola tulisan tangan namun masih perlu beberapa pengembangan lagi untuk mendapatkan hasil yang lebih akurat. Dengan demikian, penelitian ini memberikan kontribusi positif terhadap pengembangan teknologi pengenalan pola penulisan tangan, memperluas pemahaman tentang implementasi CNN dalam konteks ini, dan membuka peluang untuk pengembangan lebih lanjut di masa depan.

Kata Kunci — convolutional neural network, overfitting, tulisan tangan

1. PENDAHULUAN

Artificial Intelligence (AI) merupakan topik yang sedang diperbincangkan dan semakin berkembang dalam akhir-akhir ini. Hal ini disebabkan karena AI yang merujuk pada kemampuan mesin atau komputer untuk meniru kecerdasan manusia, termasuk dengan kemampuan untuk belajar, merencanakan, menyelesaikan masalah, dan beradaptasi dengan lingkungan yang berubah. Berbagai teori mengenai pendekatan AI berkembang pesat. Salah satu pendekatan AI yaitu *machine learning* yang terdiri dari *supervised learning* dan *unsupervised learning*. Saat ini, AI seperti pembelajaran mesin, jaringan saraf, pembelajaran mendalam, robotika, keamanan informasi, data besar, komputasi awan, Internet, dan forensik semuanya merupakan hotspot teknologi informasi dan komunikasi (TIK) dan topik menarik. Pembelajaran mesin (*machine learning*) adalah pendekatan kecerdasan buatan yang dapat meniru atau menggantikan perilaku manusia untuk memecahkan masalah, sedangkan pembelajaran mendalam (*deep learning*) adalah teknik pembelajaran mesin yang meniru saraf manusia, bagian mendasar dari otak [1]. Metode neural network merupakan salah satu pendekatan supervised learning yang cukup berkembang pesat. Model ini terdiri dari sejumlah unit pemrosesan sederhana yang disebut neuron atau node, yang terhubung secara hierarkis dalam beberapa lapisan untuk memproses informasi.

Penelitian ini menggunakan metode *Convolution Neural Network (CNN)*. CNN adalah metode yang paling banyak digunakan dan telah terbukti efisien untuk pemrosesan data citra dalam pengenalan pola. Pada konteks ini penelitian bertujuan untuk menelaah penerapan metode CNN dalam pengenalan pola tulisan tangan. Fokus utama dari penelitian ini adalah untuk memahami bagaimana CNN dapat dipergunakan dalam mengenali dan menganalisis huruf-huruf yang ada pada penulisan tangan, dengan tujuan meningkatkan akurasi, dan kecepatan dalam pengenalan.

Penelitian sebelumnya yang dilakukan oleh [2] dengan judul identifikasi kepribadian dari tulis tangan menggunakan *Euclidean Distance* yang bertujuan untuk mengidentifikasi tulisan tangan seseorang menggunakan metode *euclidean distance* serta merta menguji kinerja *euclidean distance* itu sendiri dalam mengidentifikasi tulisan tangan seseorang. Pada penelitian ini menggunakan beberapa tahap processing yang dilakukan untuk perubahan gambar dari RGB ke *grayscale* dan melakukan deteksi tepi. Dengan menggunakan data uji 15 data yang tidak dikenali 2 dan yang dikenali sebanyak 13, dengan 4 class metrik dan akurasi yang didapatkan sebesar 86.7%.

Penelitian yang dilakukan oleh [3] dengan judul analisa performa penulisan tangan angka berdasarkan jumlah iterasi menggunakan metode *Convolution Neural Network* yang bertujuan untuk membuat suatu sistem analisa akurasi performa pengenalan tulisan tangan angka menggunakan metode CNN. Pada penelitian ini Performa meningkat signifikan pada iterasi antara 0 – 20, sedangkan pada iterasi 100-1000 tidak. Hasil menunjukkan bahwa semakin banyak besar jumlah iterasi yang dilakukan semakin baik performa yang dihasilkan. Untuk itu pada penelitian ini bertujuan untuk mengimplementasi huruf-huruf yang ada pada penulisan tangan seseorang dengan menerapkan metode CNN.

2. METODE PENELITIAN

2.1 Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) adalah metode klasifikasi yang memberikan label saat melakukan pembelajaran dan termasuk dalam supervised learning [4]. Secara umum, CNN terdiri dari dua lapisan dalam model pengenalan pola, yaitu lapisan ekstraksi fitur dan lapisan klasifikasi [5]. Lapisan ekstraksi fitur terdiri dari lapisan konvolusi, sedangkan lapisan klasifikasi terdiri dari lapisan multi-layer perceptron.

CNN dikembangkan sebagai pengembangan dari Multilayer Perceptron (MLP) yang dirancang khusus untuk mengolah data dua dimensi dalam bentuk citra [6]. CNN telah terbukti sangat efektif dalam pengenalan pola pada citra dan telah digunakan dalam berbagai aplikasi seperti pengenalan wajah, deteksi objek, dan pengenalan tulisan tangan. Dalam CNN, lapisan konvolusi bertanggung jawab untuk mengekstraksi fitur-fitur penting dari citra dengan menerapkan operasi konvolusi pada piksel-piksel citra. Operasi konvolusi ini melibatkan filter atau kernel yang digeser melintasi citra untuk menghasilkan peta fitur [7]. Peta fitur ini kemudian digunakan sebagai input untuk lapisan klasifikasi, yang dapat berupa lapisan fully connected atau lapisan multi-layer perceptron untuk melakukan klasifikasi berdasarkan fitur-fitur yang diekstraksi. Dalam proses pelatihan CNN, bobot dan parameter model diubah secara iteratif menggunakan algoritma backpropagation dan optimisasi gradien untuk meminimalkan kesalahan prediksi [7]. Dengan demikian, CNN dapat belajar secara otomatis untuk mengenali pola-pola yang kompleks dan abstrak dalam citra. CNN telah menjadi salah satu metode yang paling sukses dalam bidang pengolahan citra dan pengenalan pola. Keunggulan utamanya adalah kemampuan untuk secara efektif mengekstraksi fitur-fitur penting dari citra dan melakukan klasifikasi dengan akurasi yang tinggi.

Dalam pengenalan pola tulisan tangan, Convolutional Neural Network (CNN) adalah metode klasifikasi yang digunakan. CNN telah digunakan dalam berbagai penelitian untuk mengenali pola tulisan tangan, termasuk dalam pengenalan karakter huruf dan angka [8]. CNN menggunakan lapisan konvolusi untuk mengekstraksi fitur-fitur penting dari citra tulisan tangan. Lapisan konvolusi ini menerapkan operasi konvolusi pada citra untuk menghasilkan peta fitur yang mencerminkan pola-pola penting dalam tulisan tangan. Peta fitur ini kemudian digunakan sebagai input untuk lapisan klasifikasi [9], seperti lapisan multi-layer perceptron, yang melakukan klasifikasi berdasarkan fitur-fitur yang diekstraksi.

2.2 Sumber Data

Data yang digunakan dalam penelitian ini, dibagi menjadi dua data yaitu data training dan validasi. Data training digunakan untuk mempelajari pengenalan tulisan-tulisan tangan yang umum ditulis oleh manusia. Data validasi atau data testing digunakan untuk menguji pengenalan pola pada data training. Kedua data ini berasal dari tulisan tangan penulis ditambah dengan dataset dari Internet. Data ini dibagi lagi menjadi 26 bagian mulai dari kategori huruf A sampai huruf Z.

2.3 Pre-processing Data

Tahapan pertama yaitu melakukan proses training data pada data citra training dan data citra validasi, kemudian data citra diubah ke ukuran 32x32pixel. Dari data training dan validasi dipisahkan menjadi data features dan data label. Data label diubah menjadi bentuk representasi biner. Data features di normalisasi dengan membagi tiap pixel dengan 255.0. dan diperoleh hasil dimensi dari masing-masing data features. Kemudian mengimplementasikan model Neural Network menggunakan library keras, dengan menggunakan model sequential, yang berarti layer ditambahkan satu per satu secara berurutan. Lapisan konvolusi menggunakan teknik dropout untuk mencegah overfitting. Fungsi aktivasi yang digunakan adalah ReLU. Untuk mengetahui informasi rinci setiap layer dalam model menggunakan fungsi summary(), yang akan menampilkan jumlah parameter yang digunakan serta ukuran dan bentuk dari keluaran setiap layer. Total params merupakan parameter yang dipelajari model, sedangkan trainable params jumlah parameter yang akan diperbarui selama pelatihan. Data tersebut sudah siap untuk dilatih dengan menggunakan algoritma optimisasi adam.

2.4 Proses CNN

Untuk melatih data yang sudah didapat, terdapat serangkaian proses pemrosesan data. Pertama dimulai dengan membaca gambar menggunakan OpenCV, kemudian dikonversi ke skala keabuan yang kemudian di lakukan tresholding dan dilasi. Mencari kontur eksternal untuk mengidentifikasi kontur-kontur pada gambar, kontur diurutkan dari kiri ke kanan. Dialkukan iterasi pada tiap kontur, termasuk penggambaran kotak batas, normalisasi dan resizing sebelum diprediksi menggunakan model CNN yang telah di latih. Pada model CNN yang digunakan, akan dilakukan normalisasi dan preprocessing sebagai data input, karakter yang didapat diprediksi untuk mendapatkan label karakter dari output model. Menggunakan fungsi “get_word” untuk mengonfersi daftar karakter menjadi suatu kata yang utuh. Seluruh proses tersebut divisualisasikan dengan menampilkan gambar asli yang telah diberi kotak batas pada karakter-karakter yang terdeteksi. Penelitian menggunakan CNN dalam pengenalan pola tulisan tangan telah menunjukkan hasil yang baik dalam hal akurasi. CNN dapat belajar secara otomatis untuk mengenali pola-pola yang kompleks dan abstrak dalam tulisan tangan.

2.5 Iterasi Epoch

Data dilatih dengan fungsi fit pada data training dan validasi, dengan jumlah epochs=50. Setelah pelatihan data selesai, informasi loss dan akurasi pada setiap epoch akan tersimpan dalam objek history, yang dapat digunakan untuk menganalisis performa model selama pelatihan. Menampilkan hasil perbandingan akurasi train dan validasi. Dengan menggunakan 4000 data training dan 1000 data validasi, dan proses iterasi berjalan sebanyak 50 iterasi, setiap kali iterasi akurasi pada data validasi mengalami kenaikan dan penurunan. Saat proses iterasi berakhir mendapatkan akurasi sebesar 94,44% untuk data training dan 86,46% untuk data validasi.

```
4375/4375 [=====] - 105s 23ms/step - loss: 0.5488 - accuracy: 0.8225 - val_loss: 0.6664 - val_accuracy: 0.8443
Epoch 2/50
4375/4375 [=====] - 108s 25ms/step - loss: 0.3005 - accuracy: 0.8939 - val_loss: 0.6442 - val_accuracy: 0.8634
Epoch 3/50
4375/4375 [=====] - 130s 31ms/step - loss: 0.2620 - accuracy: 0.9055 - val_loss: 0.6311 - val_accuracy: 0.8497
Epoch 4/50
4375/4375 [=====] - 113s 26ms/step - loss: 0.2390 - accuracy: 0.9128 - val_loss: 0.5321 - val_accuracy: 0.8738
Epoch 5/50
4375/4375 [=====] - 145s 33ms/step - loss: 0.2247 - accuracy: 0.9167 - val_loss: 0.5900 - val_accuracy: 0.8700
Epoch 6/50
4375/4375 [=====] - 134s 31ms/step - loss: 0.2153 - accuracy: 0.9190 - val_loss: 0.5893 - val_accuracy: 0.8722
```

Gambar 1. Proses Iterasi

```
Epoch 41/50
4375/4375 [=====] - 171s 39ms/step - loss: 0.1439 - accuracy: 0.9430 - val_loss: 0.6708 - val_accuracy: 0.8837
Epoch 46/50
4375/4375 [=====] - 176s 40ms/step - loss: 0.1438 - accuracy: 0.9434 - val_loss: 0.6313 - val_accuracy: 0.8793
Epoch 47/50
4375/4375 [=====] - 176s 40ms/step - loss: 0.1429 - accuracy: 0.9437 - val_loss: 0.6530 - val_accuracy: 0.8784
Epoch 48/50
4375/4375 [=====] - 173s 40ms/step - loss: 0.1442 - accuracy: 0.9434 - val_loss: 0.6852 - val_accuracy: 0.8793
Epoch 49/50
4375/4375 [=====] - 167s 38ms/step - loss: 0.1425 - accuracy: 0.9434 - val_loss: 1.0174 - val_accuracy: 0.8839
Epoch 50/50
4375/4375 [=====] - 149s 36ms/step - loss: 0.1413 - accuracy: 0.9444 - val_loss: 1.0176 - val_accuracy: 0.8646
```

Gambar 2. Hasil Iterasi ke-50

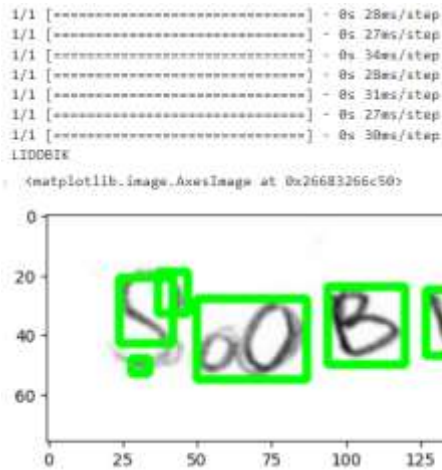
2.6 Recognition dan Processing

Pada proses recognition dan processing, membuat fungsi untuk mengurutkan kontur-kontur dalam gambar menggunakan suatu metode pengurutan berdasarkan sumbu x atau y. Selanjutnya membuat fungsi untuk memproses dalam mengenali karakter-karakter pada gambar. Dalam fungsi ini dapat mengenali, mengurutkan, dan memprediksi menggunakan model yang telah di latih. Daftar darakter yang dikenali akan ditandai dengan kotak batas, dan karakter yang dikenali akan dikembalikan dalam bentuk string sehingga bisa di baca dengan mudah.

3. HASIL DAN PEMBAHASAN

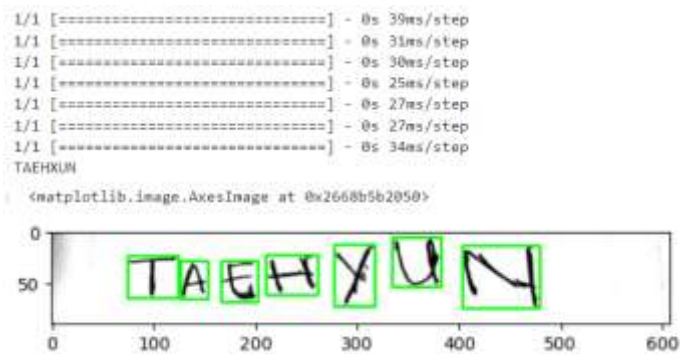
Pada saat proses iterasi data validasi mengalami penurunan akurasi dari pada data training, hal tersebut bisa dikatakan sebagai overfitting. Overfitting sendiri bisa dikatakan sebagai kemampuan model untuk menggeneralisasi ke data baru. Dalam penelitian ini, model yang dibuat mengalami overfitting, dengan kata

Pada percobaan kedua, model tidak dapat mengenali tanda garis horizontal (-), dan diidentifikasi menjadi huruf I.



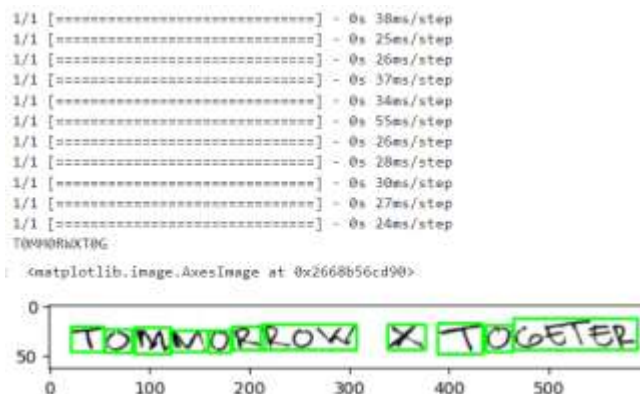
Gambar 6. Uji Coba 2

Pada percobaan ketiga, huruf S diidentifikasi menjadi beberapa bagian hal tersebut terjadi karena penekanan tulisan yang berbeda.



Gambar 7. Uji Coba 3

Pada percobaan keempat, sudah hampir sempurna bisa mengenali huruf dengan baik, namun karena terjadi overfitting berdampak yang signifikan pada pengenalan huruf.



Gambar 8. Uji Coba 4

Pada percobaan kelima, dengan menggunakan beberapa kata dalam satu kertas. Dapat mengenali huruf dengan baik. Namun karena tidak ada celah antar huruf pada akhir kata, huruf tersebut diidentifikasi menjadi satu kata.

4. SIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan mengenai implementasi metode CNN pada pengenalan pola tulisan tangan, diperoleh kesimpulan sebagai berikut :

1. Pola pengenalan tulisan tangan dapat dikenali dengan cukup baik, dengan melakukan beberapa percobaan dengan penulisan tangan secara manual.
2. Model mengalami overfitting, yang mengakibatkan penurunan akurasi pada data validasi.
3. Terdapat simbol yang teridentifikasi menjadi huruf lain.

5. SARAN

Berdasarkan penjelasan materi diatas, adapun saran untuk penelitian kedepan untuk dapat memperbaiki kekurangan dalam penelitian ini yaitu :

1. Me-regularize atau merubah model untuk mengurangi overfitting. Namun bisa juga dengan mengganti dataset dengan data lain.
2. Dapat mengenali simbol dan mengenali pola tulisan tangan yang belum jelas (blur).

DAFTAR PUSTAKA

- [1] A. Ahmad, 2017. "Mengenal artificial intelligence, machine learning, neural network, dan deep learning," *Yayasan Cahaya Islam, Jurnal Teknologi Indonesia*, vol. 3, 1-5.
- [2] A. I. Maulana, 2023. "Identifikasi Kepribadian Dari Tulisan tangan Menggunakan Eucliden Distance," *Prosiding Seminar Nasional Teknologi dan Sains*, vol. 2, 177-182.
- [3] S. Prihatiningsih, N. S. M, F. Andriani and N. Nugraha, 2019. "Analisa Performa Pengenalana Tulisan Tangan Angka Berdasarkan Jumlah Iterasi Menggunakan Metode CNN," *Jurnal Ilmiah Teknologi dan Rekayasa*, vol. 24, 58-66.
- [4] M. Arsal, B. A. Wardijono and D. Anggraini, 2020. "Face Recognition Untuk Akses Pegawai Bank Menggunakan Deep Learning Dengan Metode CNN," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 6, 55-63.
- [5] N. Kasim and G. S. Nugraha, 2021. "Pengenalan Pola Tulisan Tangan Aksara Arab Menggunakan Metode CNN," *Jurnal Teknologi Informasi, Komputer, dan Aplikasinya*, vol. 3.
- [6] A. Arkadia, S. A. Damayanti and D. S. Prasvita, 2021. "Klasifikasi Buah Mangga Badami Untuk Menentukan Tingkat Kematangan dengan Metode CNN," *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)*, vol. 2, 158-165.
- [7] I. L. Rahmatullah, "repository.uinjkt.ac.id," 20 Juni 2022. [Online]. Available: <https://repository.uinjkt.ac.id/dspace/bitstream/123456789/62399/1/IMAM%20LUTFI%20RAHMATULLAH-FST.pdf>. [Accessed 05 Desember 2023].
- [8] F. Ilham and N. Rochmawati, 2020. "Transliterasi AKsara Jawa Tulisan Tangan ke Tulisan Latin Menggunakan CNN," *Journal of Informatics and Computer Science*, vol. 1, 200-208.
- [9] S. C. Pradana, "dspace.uui.ac.id," 21 Maret 2018. [Online]. Available: <https://dspace.uui.ac.id/handle/123456789/6474?show=full>. [Accessed 5 Desember 2023].

Pengembangan Media *Game* Bertema *Serious Game* Sebagai Pembelajaran Berbasis *Ice Breaking*

Candra Bagus Pratama¹

¹Teknik Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Nusantara PGRI Kediri

E-mail: *¹chandbags848@gmail.com

Abstrak – *Game* akhir- akhir ini telah menjadi alat yang cukup efektif dalam konteks pendidikan. *Serious Game* adalah salah satu pendekatan yang inovatif meningkatkan kognitif siswa selain *game* bertema puzzle. Penelitian ini bertujuan untuk menghasilkan media pembelajaran berupa *serious game*. Jenis penelitian yang digunakan adalah penelitian dan pengembangan (R&D). Memiliki langkah – langkah berdasarkan Bord dan Gall namun berhenti pada tahapan ke 3 dari 10 tahapan. Meliputi *Research and information collecting, Planning dan Develop preliminary form of product*. *Serious Game* ini dikembangkan dengan *software unity* dengan tahap perencanaan melalui tahap persiapan, tahap analisis, dan tahap desain memerlukan waktu 3 bulan. Produk media yang dikembangkan dapat dikolaborasi dengan proses pembelajaran sekolah untuk memajukan kognitif siswa. Penelitian yang dilakukan masih diperlukan pengembangan dan penerapan sehingga masih diperlukan penelitian lanjutan untuk dapat menyempurnakan produk yang ada.

Kata Kunci — Kognitif, Pembelajaran *Ice Breaking*, Pendidikan, *Serious Game*

1. PENDAHULUAN

Dalam kehidupan sehari – hari, sangat banyak pekerjaan yang menggunakan komputer. Antara lain bidang komunikasi, transportasi, industri, kesenian, pertanian bahkan sampai dalam bidang pendidikan. Dalam dunia pendidikan, komputer sangat banyak dimanfaatkan menjadi media sebagai sarana pembelajaran. Komputer merupakan salah satu alat bantu dalam mengajar, Media pembelajaran perangkat keras dan perangkat lunak yang digunakan pada perangkat komputer. Perangkat lunak dipakai untuk media pembelajaran sebagai sumber belajar. Salah satu contoh dari program pembelajaran yaitu melalui *game*. Perkembangan teknologi yang semakin maju di bidang *game*, sehingga *game* dapat dijadikan sebagai media pembelajaran.

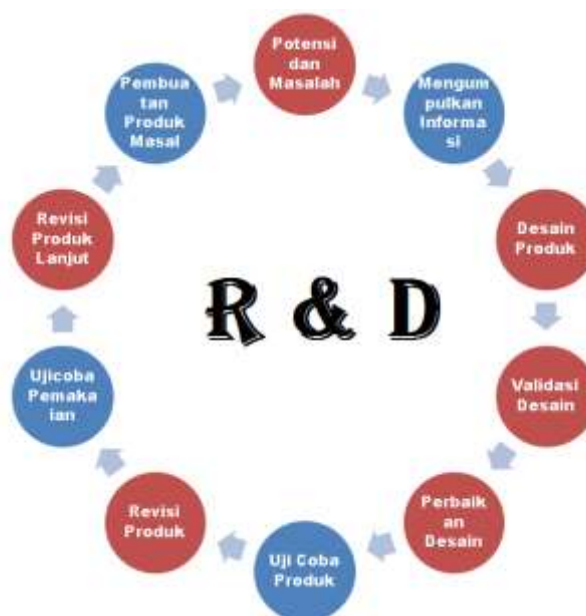
Kemampuan kognitif merupakan salah satu aspek perkembangan manusia yang sangat fundamental. Kemampuan kognitif sebagai aspek perkembangan manusia yang berkaitan dengan pengetahuan, segala proses psikologis yang berkaitan bagaimana individu mempelajari dan memikirkan lingkungannya. Kreativitas sebagai kemampuan untuk menciptakan sesuatu yang baru, sebagai kemampuan untuk memberi gagasan baru yang dapat diterapkan dalam pemecahan masalah, atau sebagai kemampuan untuk melihat hubungan baru antara unsur yang sudah ada sebelumnya [1]. Segala individu memainkan peran aktif dalam menyusun pengetahuan dari kegiatan yang dilakukan setiap hari, sehingga individu dapat bereksplorasi dalam membangun dan mengembangkan kemampuan kognitif [2].

Serious games merupakan *game* yang memiliki maksud atau tujuan yang lebih dari sekedar sarana hiburan. *Game* jenis ini biasanya digunakan sebagai alat pelatihan dan terkadang juga digunakan sebagai sarana promosi sebuah produk. Permainan edukasi pada *serious game* bisa digunakan menjadi contoh media pendidikan yang berfungsi sebagai alat pembelajaran. Namun kondisi tersebut masih belum didukung dengan ketersediaan media pembelajaran yang digunakan di sekolah [3]. Pada *serious game* selain unsur hiburan juga membawa pesan pembelajaran dan pendidikan. Walaupun unsur pendidikan lebih menonjol, unsur edukasi yang terdapat *game* jauh dari suatu pelajaran formal sehingga *game* menjadi tidak terlalu kaku dan lebih menarik. Setelah memainkan *game* ini, diharapkan pemain akan paham topik yang ada di dalam *game* yang biasanya disertai dengan solusi dan model pemecahannya. Salah satu konten yang ada pada *serious game* ini, pemain akan mendapatkan pengalaman seolah-olah terlibat secara langsung.

Ice breaking dalam pembelajaran merupakan cara yang digunakan untuk mencairkan suasana yang kurang kondusif selama proses pembelajaran berlangsung, kegiatan ini dapat dikombinasikan dengan materi, kondisi siswa dan lingkungan belajarnya [4]. Kegiatan *ice breaking* dalam pembelajaran dapat membangkitkan motivasi belajar dan mendapatkan hasil belajar yang lebih baik [5].

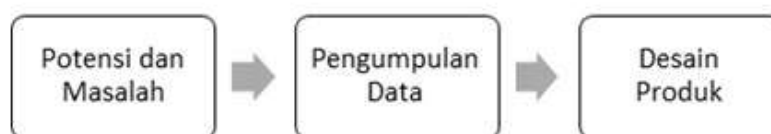
2. METODE PENELITIAN

Penelitian dilakukan menggunakan jenis penelitian dan pengembangan (*Research and Development*). Borg dan Gall (2003:771) menjelaskan, secara detail tentang penelitian pengembangan. Yaitu merupakan sebuah proses yang digunakan untuk mengembangkan dan memvalidasi produk pendidikan yang mencakup tidak hanya objek material, buku teks, film instruksional, tetapi juga prosedur dan proses, seperti metode pengajaran atau metode pengorganisasian instruksi.



Gambar 1. Skema Tahapan Penelitian Borg and Gall

Langkah - langkah dalam proses penelitian dan pengembangan yang digunakan berdasarkan Borg dan Gall, namun berhenti pada tahapan yang ke-3 dari 10 total tahapan. Meliputi *Research and information collecting*, *Planning*, dan *Develop preliminary form of product*. Instrumen pengumpulan data yang digunakan adalah non tes. Teknik pengumpulan, yaitu dengan menggunakan studi literasi.



Gambar 2. Tiga Tahapan yang diambil penulis

2.1 Penelitian dan pengumpulan Informasi (*Research and information collecting*)

Tahapan awal melalui kegiatan survei literatur dan analisis kebutuhan studi literatur bertujuan untuk meningkatkan pemahaman secara teoritis tentang produk yang akan dihasilkan, sehingga terdapat jaminan produk dikembangkan dengan teori yang kuat dan dapat dipertanggung jawabkan.

2.2 Perencanaan (*Planning*)

Berisi tabel perencanaan yang dimana menjelaskan secara rinci tentang kegiatan apa saja yang dilakukan dan berapa lama proses penelitian dilakukan. Pencatatan dimulai dari tahap awal sampai dari tahap ketiga atau akhir dari metode penelitian yang diambil penulis.

2.3 Pengembangan bentuk awal produk (*Develop preliminary form of product*)

Meliputi desain produk yang dikembangkan penulis yaitu sebuah game bertema *serious game*. Membuat rancangan awal seperti alur cerita, mekanik game, *flow*, sampai dari *layout Game*. Setelah itu baru membuat desain atau *game* melalui *unity* menambahkan beberapa tombol, animasi, aset dalam game, dan kebutuhan isi *game* yang lainnya.

3. HASIL DAN PEMBAHASAN

3.1 Penelitian dan pengumpulan informasi (*Research and information collecting*)

Tahapan awal melalui kegiatan survei literatur dan analisis kebutuhan studi literatur bertujuan untuk meningkatkan pemahaman secara teoritis tentang produk yang akan dihasilkan, sehingga terdapat jaminan produk dikembangkan dengan teori yang kuat dan dapat dipertanggung jawabkan. Adapun kegiatan yang dilakukan adalah melakukan analisis melalui *website* tentang tema *game* yang ramai seperti misalnya horor atau petualangan dikalangan siswa sebagaimana dirasa sesuai dengan permasalahan penelitian.

Untuk mendukung referensi yang relevan, survei literasi juga dilakukan dengan mereview beberapa informasi melalui jurnal ilmiah. Analisa kebutuhan berdasarkan survei literatur yang telah dilakukan, maka keseluruhan data yang diperoleh diolah dalam analisa kebutuhan yang terdiri dari analisis pengguna dan analisis pembelajaran. Berdasarkan analisis yang ada adapun hasil akhir yang diambil sebagai analisis kebutuhan adalah pembuatan game bertema *serious game* sebagai pelatihan kognitif yang meliputi pemecahan masalah, meningkatkan kreatifitas, dan daya ingat siswa. Adapun pertimbangan pemilihan materi tersebut adalah karakteristik materi pembelajaran yang dianggap unik dan jarang diajarkan kepada siswa. Untuk membuat media pembelajaran *serious game* ini, *software* yang digunakan adalah *unity*.

3.2 Perencanaan (*Planning*)

Tabel perencanaan yang dimana menjelaskan secara rinci tentang kegiatan apa saja yang dilakukan dan berapa lama proses penelitian dilakukan. Pencatatan dimulai dari tahap awal sampai dari tahap ketiga atau akhir metode penelitian yang diambil penulis. Tahap dimulai dari tahap persiapan yang meliputi isian apa saja beserta dengan waktunya lalu dilanjutkan dengan tahap analisis dan desain produk beserta isian dan waktu proses pembuatannya akan dijelaskan pada tabel berikut :

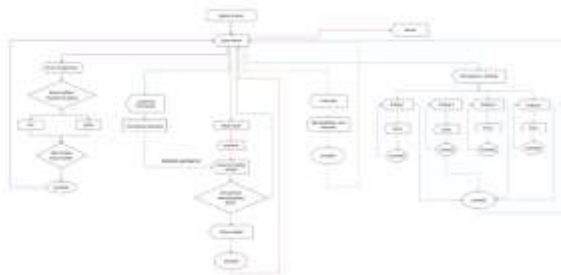
Tabel 1. Perencanaan Penelitian dan Pengembangan

Tahapan	Kegiatan	Waktu
Tahap Persiapan	Tahap ini meliputi pengajuan judul, pembuatan perancangan penelitian	Oktober 2023
Tahap Analisis	Tahap ini meliputi analisis pengguna, dan analisis media pembelajaran	November 2023
Tahap Desain	Tahap ini meliputi pembuatan produk serta pengecekan <i>bug</i> atau bisa disebut <i>trial and error</i>	Desember 2023

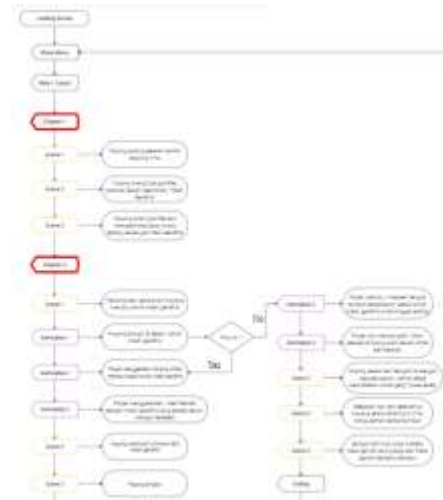
Sumber: perencanaan penelitian

3.3 Pengembangan bentuk awal produk (*Develop preliminary form of product*)

Rancangan awal seperti alur cerita, mekanik *game*, *flow*, sampai dari *layout Game*. Setelah itu baru membuat desain atau *game* melalui *unity* menambahkan beberapa tombol, animasi serta beberapa aset dalam *game*.



Gambar 3. Rancangan *Flow* Tampilan Awal Pada *Game*



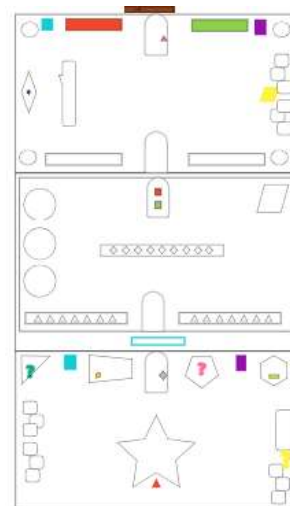
Gambar 4. Rancangan cerita atau *Flow Story Game*

Pada gambar 3 dan 4 merupakan *flow* dari *game* yang dibuat, untuk gambar 3 menjelaskan *Gameflow screen* atau tampilan awal menu ketika memasuki *game* yaitu meliputi tombol Mulai, Lanjutkan, Pengaturan, Pencapaian, Cara bermain, dan Keluar. Sedangkan untuk gambar 4 merupakan sebagian dari *Flowstory game* yang memiliki cerita dengan 2 rute atau jalan cerita dengan *ending* yang berbeda dan setiap apa yang dilakukan karakter dalam *game* menentukan *ending* dari cerita di dalam *game*.

Game Mechanic :

- Player dapat menggerakkan karakter dengan menekan tombol A-S-W-D / ← → ↑ ↓
- Player bisa mengambil ruangan
- Tespat Kiblat tidak bisa ditebak/ditahak, untuk melewatiya baru serouta atau mengindai
- Player bisa mengambil kayu bakar dekat perapian dengan klik tombol 'E'
- Player bisa mengambil kasek yang tersembunyi di atas dengan klik tombol 'E'
- Player bisa mengambil bensin yang tersembunyi di atas dengan klik tombol 'E'
- Player bisa mengambil lilin yang tergeletak di tempat ritual dengan klik tombol 'E'
- Player bisa interaksi dengan foto, klik tombol 'E' untuk melihat
- Player bisa interaksi dengan kertas sebekan, klik tombol 'E' untuk membaca
- Player bisa interaksi dengan lukisan, klik tombol 'E' untuk melihat dan membaca
- Player bisa melihat isi tas dengan klik tombol 'T' atau mengasahat kuncir ke gambar tas
- Player harus menyusun secarik kertas yang tertinggal pada rak buku lalu menentabkan kuderma untuk membuka kunci padlock, untuk membaca bus menggerakkan klik tombol 'E'
- Peta dapat dibuka dengan cara menyelesaikan Puzzle untuk gendak Padlock

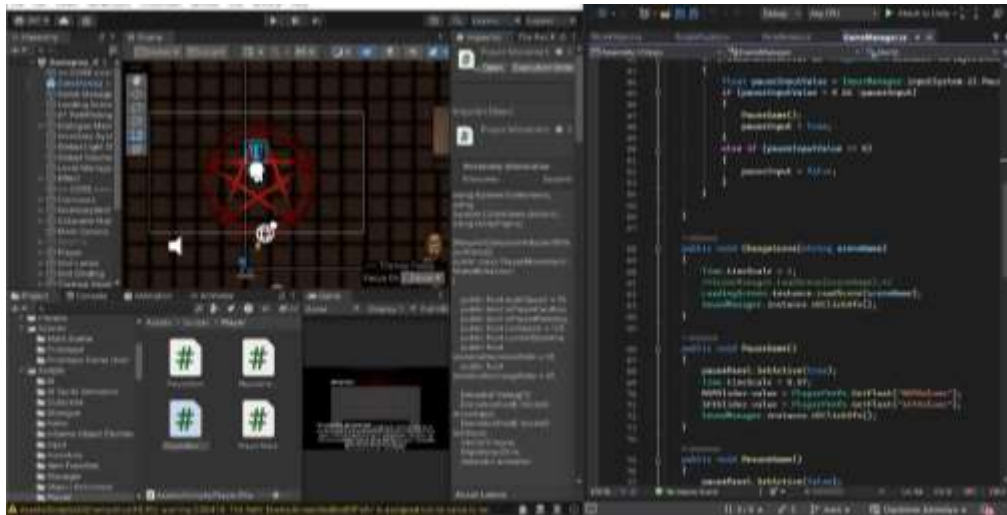
Gambar 5. Rancangan Mekanik pada *Game*



Gambar 6. Rancangan *Layout* pada *Game*

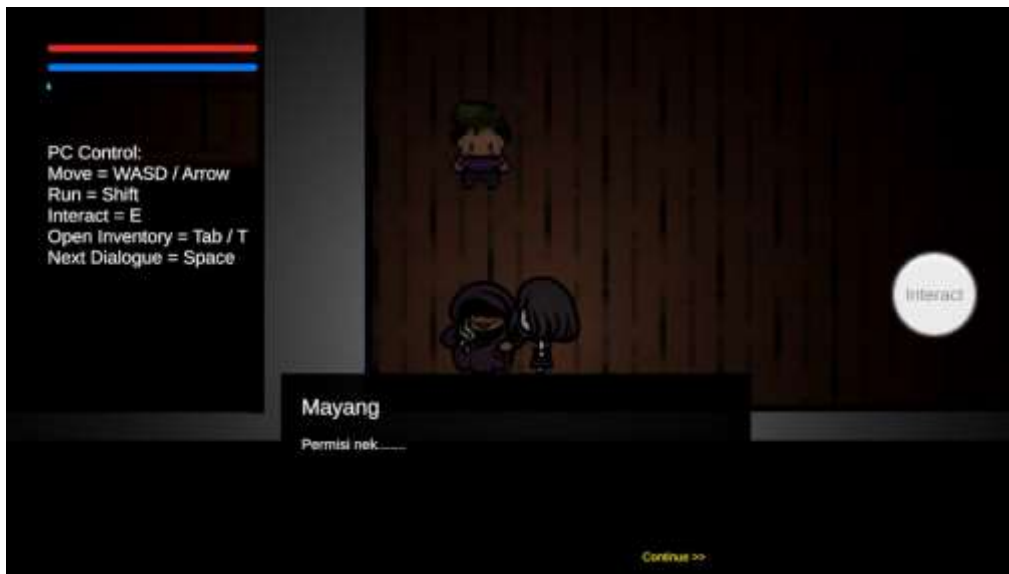
Pada gambar 5 merupakan bagian mekanik dari salah satu ruangan pada *game*, dengan penjelasan apa saja yang bisa dilakukan karakter dalam ruangan beserta bagaimana cara pemain keluar dari ruangan tersebut. Seperti cara menggerakkan karakter, apa saja benda yang bisa diinteraksi sampai menyelesaikan *puzzle* untuk keluar dari ruangan tersebut.

Sedangkan pada gambar 6 merupakan desain rancangan *layout* dari beberapa ruangan. Hal ini dilakukan sebagai panduan dari tempat dari *game* yang akan dibuat, seperti penempatan barang interaksi sampai barang yang dapat diambil karakter. Bukan hanya barang saja, tetapi penempatan rintangan juga akan dimasukkan pada desain *layout* ini. Meliputi *clue* untuk *puzzle* dan barang – barang yang akan dikombinasikan hingga menjadi satu barang seperti korek, kayu bakar dan bensin jika disatukan akan menjadi obor.



Gambar 7. Perancangan Game Pada Unity

Pada gambar 7 merupakan proses dari rancangan game yang akan dibuat pada unity. Proses diatas ditunjukkan beberapa coding pada bagian ruangan satu, meliputi background music apa yang akan dimainkan, sound effect apa yang dimainkan ketika berinteraksi dengan salah satu barang, bagaimana cara pemain menggerakkan karakter, tombol apa saja yang digunakan untuk berinteraksi sampai pengecekan bug atau error yang menyebabkan karakter tidak nyaman untuk memainkan game ini.



Gambar 8. Tampilan Game Ketika Memasuki Permainan

Pada gambar 8 yaitu penyelesaian game serta pengecekan bug dan error, dengan masuk ke dalam game lalu memainkan game dari awal sampai selesai. Setelah itu jika menemukan bug, error atau dirasa ada beberapa hal yang kurang sesuai dengan desain awal, maka kembali pada unity untuk melakukan perbaikan. Hal ini dilakukan berulang kali sehingga menjadi game yang sesuai dengan desain awal dan tentunya dengan sedikit berimprovisasi agar lebih baik untuk keseluruhan game ini.

4. SIMPULAN

Pembelajaran terus dikembangkan dengan berbagai cara yang berbeda, meliputi metode yang digunakan dalam pembelajaran, model seperti yang digunakan dalam pembelajaran, dan media yang digunakan dalam pembelajaran. Perancangan *Game* ini adalah wujud kepedulian akan pentingnya suatu media pembelajaran dalam menunjang kognitif siswa, sebuah pembelajaran dengan memanfaatkan sebuah *game* bertemakan *serious game* yang memiliki konten *puzzle* untuk siswa dapat memahami pentingnya kognitif pada siswa. Banyak yang harus dipelajari untuk menghasilkan *game* yaitu.

1. Mengetahui konten apa yang akan digunakan
2. Aplikasi apa yang akan dikembangkan menjadi *game*

5. SARAN

Penelitian yang saya lakukan masih memerlukan banyak perbaikan baik itu apa yang ada didalam maupun diluar *game*. Mungkin selanjutnya membuat konsep perancangan yang lebih matang dan kemudian memproduksi konsep tersebut menjadi sebuah *game*.

DAFTAR PUSTAKA

- [1] Munandar, U. (2009). Pengembangan Kreativitas Anak Berbakat. Jakarta: Rineka Cipta.
- [2] Desmita. 2010. Psikologi Perkembangan. Bandung: PT Remaja Rosdakarya.
- [3] Pratama, L. D., Lestari, W., & Bahauddin, A. 2019. Game Edukasi: Apakah membuat belajar lebih menarik?. *At-Ta'lim: Jurnal Pendidikan*, 5(1), 39-50.
- [4] Febriandari, E. I. 2018. Pengaruh kreativitas guru dalam menerapkan ice breaking dan motivasi belajar terhadap hasil belajar siswa sekolah dasar. *Briliant: Jurnal Riset dan Konseptual*, 3(4), 485-494
- [5] Susannah, R., & Alarifin, D. H. 2014. Penerapan permainan penyegar (ice breaking) dalam pembelajaran fisika untuk meningkatkan motivasi dan hasil belajar. *Jurnal Pendidikan Fisika*, 2(1).
- [6] Fanani, Achmad. 2010. "Ice Breaking Dalam Proses Belajar Mengajar". *Jurnal*. Vol.1 No.11, VI, hlm. 4-69.
- [7] Hamzah B. Uno. 2019. Model Pembelajaran (Menciptakan Proses Belajar Mengajar yang Kreatif, dan Efektif), Jakarta: Bumi Aksara. Hlm. 2
- [8] Said, M. 2010. 80+ Ice breaker Games; Kumpulan Permainan Pengunggah Semangat. Yogyakarta: Andi Publisher.