

BAB II

LANDASAN TEORI

2.1 Penelitian Terdahulu

Terdapat beberapa penelitian terdahulu yang terangkum pada Tabel 2. 1 terkait dengan topik penelitian yang dijadikan acuan pada penelitian ini.

Tabel 2. 1 Penelitian terdahulu

Penelitian 1	
Nama Jurnal	Hindawi - <i>Computational Intelligence and Neuroscience</i> , Q1
Judul Jurnal	<i>Driver Fatigue Detection Based on Convolutional Neural Networks Using EM-CNN</i> , 2020 [44]
Nama Penulis	Zuopeng Zhao, Nana Zhou, Lan Zhang, Hualin Yan, Yi Xu, dan Zhongxin Zhang
Metode	- MTCNN untuk mendeteksi fitur wajah manusia - EM-CNN untuk mengidentifikasi keadaan mata dan mulut dari gambar <i>Region of Interest (RoI)</i> - Penggunaan PERCLOS untuk persentase bukaan mata seiring waktu dan POM untuk derajat bukaan mulut
Hasil dan Kesimpulan	Hasil eksperimental menunjukkan bahwa EM-CNN dapat mendeteksi status kelelahan pengemudi dengan efisien menggunakan gambar dengan tingkat akurasi mencapai 93%
Penelitian 2	
Nama Jurnal	Sensors (MDPI), Q2
Judul Jurnal	<i>Detection Of Drowsiness Among Drivers Using Novel Deep Convolutional Neural Network Model</i> , 2023, Vol. 23 (21) [45]
Nama Penulis	Fiaz Majeed, Umair Shafique, Mejdil Safran, Sultan Alfarhood, dan Imran Ashraf
Metode	- <i>Proposed CNN</i> untuk mendeteksi kantuk pengemudi - Dlib untuk mengidentifikasi <i>facial key</i> guna menghitung MAR - Penerapan Augmentasi data untuk meningkatkan variasi dataset yang kecil
Hasil dan Kesimpulan	Hasil eksperimental menunjukkan bahwa model CNN yang diusulkan mencapai rata-rata akurasi sebesar 95.71% pada data yang telah diterapkan teknik augmentasi data.
Penelitian 3	
Nama Jurnal	<i>Applied Sciences</i> (MDPI), Q2
Judul Jurnal	<i>Driver Fatigue Detection Based on Residual Channel Attention Network and Head Pose Estimation</i> , 2021, Vol. 11 (19) [32]
Nama Penulis	Mu Ye, Weiwei Zhang, Pengcheng Cao, Kangan Liu
Metode	- RetinaFace untuk menangkap wajah beserta lima fitur wajah dan mengekstrak area mata dan mulut. - RCAN untuk mendeteksi kondisi mata dan mulut

	- PnP untuk memperkirakan sudut kepala - PERCLOS dan POM menilai kondisi kelelahan berdasarkan ambang batas
Hasil dan Kesimpulan	Penggabungan RetinaFace dan metode RCAN menghasilkan akurasi 98% untuk mengklasifikasikan kondisi mata, mulut, dan pose wajah. Sistem deteksi kelelahan pengemudi ini memiliki akurasi dan ketahanan deteksi yang tinggi.
Penelitian 4	
Nama Jurnal	Heliyon, Q1
Judul Jurnal	<i>Computer Vision-Based Approach To Detect Fatigue Driving And Face Mask For Edge Computing Device</i> , 2022, Vol. 8 (10) [46]
Nama Penulis	Ashiqur Rahman, Mamun Bin Harun Hriday, Riasat Khan
Metode	- Dlib untuk mengukur tingkat kelelahan berdasarkan bukaan mata dan mulut - Custom CNN untuk mengklasifikasikan penggunaan masker wajah - Pengujian <i>real-time</i> dilakukan dengan berbagai variasi situasi nyata
Hasil dan Kesimpulan	Penelitian telah melalui serangkaian uji coba dengan berbagai variasi situasi nyata seperti sudut pandang, jarak pengambilan gambar, penggunaan kacamata, dan pencahayaan. Hasil uji coba mendapatkan akurasi sebesar 97,44% untuk mendeteksi kelelahan dan 97,90% untuk identifikasi masker wajah.
Penelitian 5	
Nama Jurnal	Hindawi - <i>Security and Communication Networks</i> , Q2
Judul Jurnal	<i>Driver Fatigue Detection Based on Facial Key Points and LSTM</i> , 2021, Vol. 2021 (8) [47]
Nama Penulis	Long Chen, Guojiang Xin, Yuling Liu, dan Junwei Huang
Metode	- MTCNN untuk mendeteksi wajah manusia - Dlib untuk menemukan <i>landmark</i> wajah guna mengekstraksi fitur kelelahan - Penggunaan LSTM untuk mendapatkan hasil akhir kelelahan berdasarkan fitur yang disusun menjadi urutan fitur temporal - Proses kelelahan diidentifikasi selama 3 detik
Hasil dan Kesimpulan	Eksperimen menghasilkan bahwa algoritma deteksi kelelahan yang diusulkan mencapai akurasi yang lebih baik dalam akurasi dibandingkan dengan metode lain. Akurasi rata-rata dari metode yang diusulkan mencapai 93% pada <i>self-built</i> dataset.
Penelitian 6	
Nama Jurnal	IET <i>Image Processing</i> , Q2
Judul Jurnal	<i>Research on fatigue detection based on visual features</i> , 2022, Vol. 16 (4) [48]
Nama Penulis	Guangzhe Zhao, Yanqing He, Hanting Yang, Yong Tao
Metode	- Single Shot Multi-Box Detector (SSD) digunakan untuk mendeteksi wajah - Dlib untuk mengekstrak area fitur wajah (mata dan mulut) - VGG16 CNN untuk mempelajari fitur kelelahan dari area mata dan mulut
Hasil dan Kesimpulan	Hasil dari penelitian menunjukkan bahwa mata dan mulut adalah bagian dari <i>visual features</i> yang berperan penting dalam mendeteksi kelelahan. Metode pendeteksian wajah yang menggabungkan mata dan mulut mencapai tingkat akurasi 91.88%.
Penelitian 7	
Nama Jurnal	<i>Journal of Information Systems Engineering and Business Intelligence</i> , S1

Judul Jurnal	<i>Fatigue Detection on Face Image Using FaceNet Algorithm and K-Nearest Neighbor Classifier</i> , 2021, Vol. 7 (1) [20]
Nama Penulis	Faisal Dharma Adhinata, Diovianto Putra Rakhmadani, Danur Wijayanto
Metode	FaceNet untuk feature extraction
Hasil dan Kesimpulan	Hasil dari kombinasi algoritma FaceNet dan algoritma klasifikasi yang digunakan memberikan hasil akurasi yaitu 94,68%. Terdapat kelemahan penggunaan FaceNet pada pendeteksian wajahnya, yaitu sulit untuk menangkap area wajah jika latar belakang dalam kondisi yang tidak kondusif.
Penelitian 8	
Nama Jurnal	Journal of Automobile Engineering, Q2
Judul Jurnal	<i>Driver fatigue detection based on convolutional neural network and face alignment for edge computing device</i> , 2021, Vol. 1 (1) [49]
Nama Penulis	Xiaofeng Li, Jiahao Xia, Libo Cao, Guanjun Zhang, Xiexing Feng
Metode	- CNN untuk mendeteksi dan mengklasifikasi keadaan wajah - Penggunaan EAR dan MAR untuk mengukur parameter kelelahan pengemudi - PERCLOS untuk durasi penutupan mata dan Frekuensi menguap (POM)
Hasil dan Kesimpulan	Metode yang diusulkan berhasil mendeteksi kelelahan dengan akurasi rata-rata 89.55% pada dataset YawDD, 300W, dan DriverEyes.
Penelitian 9	
Nama Jurnal	Sensors (MDPI), Q2
Judul Jurnal	<i>An Explainable Student Fatigue Monitoring Module with Joint Facial Representation</i> , 2023, Vol. 23 (7) [50]
Nama Penulis	Xiaomian Li, Jiaqin Lin, Zhiqiang Tian, Yuping Lin
Metode	- CNN untuk menghasilkan fitur spasial-temporal dan mengklasifikasikan gejala - Memadukan hasil klasifikasi dengan <i>Maximum a Posteriori</i> - NTHU-DDD dataset untuk data <i>training-validation</i> dan OSFM (Self-built)
Hasil dan Kesimpulan	Metode yang dikembangkan menghasilkan tingkat akurasi mencapai 94.47% pada data OSFM
Penelitian 10	
Nama Jurnal	<i>Journal of Real-Time Image Processing</i> , Q2
Judul Jurnal	<i>Temporal And Spatial Feature Based Approaches In Drowsiness Detection Using Deep Learning Technique</i> , 2021, Vol. 18 (1) [51]
Nama Penulis	Nageshwar Nath Pandey, Naresh Babu Muppalaneni
Metode	- CNN digunakan untuk ekstraksi fitur spasial - LSTM digunakan untuk ekstraksi fitur temporal
Hasil dan Kesimpulan	Penggunaan CNN dan LSTM untuk model berbasis fitur spasial menghasilkan akurasi yang sangat tinggi 97.5%, sedangkan model berbasis fitur temporal menunjukkan kinerja yang lebih cepat namun akurasi yang rendah 79.9%

Berdasarkan penelitian terdahulu yang tercantum pada Tabel 2. 1, terdapat beberapa penelitian yang telah melakukan klasifikasi kondisi kelelahan dengan

menggunakan algoritma *deep learning* CNN dengan berbagai arsitektur dan kombinasi teknik yang berbeda-beda. Penerapan algoritma arsitektur CNN yang disesuaikan pada berbagai dataset untuk klasifikasi kondisi kelelahan menunjukkan hasil yang cukup baik [44] [49] [50] [51]. Pengukuran tingkat kelelahan pada wajah dapat dilakukan dengan berbagai pendekatan, menggunakan seluruh wajah tanpa memecah fitur wajah atau menggunakan fitur wajah seperti mata dan mulut. Penelitian yang mengembangkan model deteksi kelelahan berdasarkan wajah telah dilakukan dan menghasilkan akurasi sebesar 89.55% [49], sedangkan penelitian lainnya yang mengembangkan model deteksi kelelahan berdasarkan fitur wajah, seperti mata dan mulut, atau salah satunya, menunjukkan hasil yang lebih baik dengan rata-rata akurasi diatas 90% [44] [32] [46]. Oleh karena itu, penelitian ini akan menerapkan gabungan fitur mata dan mulut untuk mendapatkan keakuratan klasifikasi kelelahan yang lebih baik. Penelitian yang membahas kombinasi Dlib untuk mengukur tingkat kelelahan berdasarkan persamaan EAR dan MAR juga kerap dilakukan. Kemampuan Dlib dalam mendeteksi 68 titik koordinat wajah, membuat perhitungan tingkat kelelahan lebih mudah untuk dilakukan, banyak penelitian yang memanfaatkan kemampuan ini untuk mengukur tingkat kelelahan pada mata dan mulut. Penelitian ini akan melakukan deteksi dini gejala kelelahan berdasarkan fitur wajah, yaitu mata dan mulut, untuk mengidentifikasi kelelahan pada mahasiswa, sebagaimana dilakukan pada penelitian sebelumnya [50] [51].

Berdasarkan penelitian sebelumnya yang terdapat pada Tabel 2. 1, terdapat gap yang menjadi landasan kebaruan dari penelitian ini. Penelitian ini akan menerapkan algoritma RetinaFace untuk deteksi wajah berdasarkan 5 fitur wajah, penelitian [32] telah membuktikan performa RetinaFace dalam mendeteksi wajah dalam berbagai situasi kondisi, kemampuan RetinaFace dapat mengatasi kelemahan pada penelitian [20], namun algoritma ini terbilang masih baru dan masih jarang digunakan pada penelitian lainnya. Penggunaan Dlib yang telah teruji pada beberapa penelitian sebelumnya [45] [46] [47] [48] akan diterapkan pada penelitian ini untuk memastikan label setiap gambar telah sesuai dengan tingkat kelelahannya, menggunakan persamaan EAR dan MAR. Untuk meningkatkan jumlah dan variasi data, teknik augmentasi akan diterapkan pada dataset *training*. Penelitian ini akan

menerapkan arsitektur CNN yang telah teruji keakuratannya dalam mengklasifikasikan gambar terutama wajah. Kombinasi Dlib, RetinaFace, dan Arsitektur *pre-trained* CNN serta *CNN Custom layer*, belum pernah dilakukan sebelumnya, sehingga menjadi kebaruan untuk penelitian ini.

2.1 Tinjauan Teori

2.1.1 Kelelahan

Kelelahan adalah salah satu gejala penyakit umum yang dapat menyerang semua kalangan. Kelelahan dapat terjadi pada kondisi individu yang bugar maupun individu dengan kondisi medis akut dan kronis. Kelelahan merupakan salah satu hal penting yang perlu ditangani dengan baik. Kelelahan dapat menyebabkan berbagai permasalahan seperti menurunnya efisiensi kerja, produktivitas dan kemampuan kerja [52]. Kelelahan berarti kondisi tubuh individu merasakan satu hal yang tidak sesuai dengan kapasitasnya, kondisi ini memberikan berbagai sensasi kepada tubuh untuk meredakan kondisi kelelahan ini. Seperti halnya, sensasi tubuh yang merasa perlu untuk istirahat atau sensasi ketidaksesuaian antara usaha yang dikeluarkan dengan kapasitas yang dimiliki tubuh [53]. Kelelahan adalah gejala umum yang berpotensi melemahkan kondisi tubuh seseorang, sehingga dapat mempengaruhi kualitas hidup terkait kesehatan individu [54]. Umumnya, individu menggambarkan kelelahan sebagai kelesuan, kekurangan energi, kelelahan, kantuk, kecenderungan untuk tertidur di siang hari, hingga kebiasaan tertidur Ketika dalam kondisi melakukan aktivitas normal sehari-hari [55]. Kelelahan memiliki beberapa jenis, diantaranya kelelahan fisik, kelelahan mental, dan kelelahan emosional [53].

a. Penyebab kelelahan

Penyebab utama kelelahan yang terus-menerus adalah gangguan tidur, gangguan pernapasan pada saat tidur, depresi, kurang berolahraga, dan stress psikososial. Namun, kasualitasnya tidak selalu dapat ditetapkan dengan jelas, dikarenakan kelelahan memiliki berbagai potensi penyebab

yaitu, biologis, psikologis, dan sosial. Selain itu, perlu diperhatikan bahwa interaksi yang biasa dilakukan sehari-hari juga dapat menjadi penyebab dari kelelahan, seperti pengobatan, lebih banyak mengambil waktu untuk istirahat dan tidur, dan aktivitas fisik [55]. Kelelahan juga dapat terjadi dari jam kerja yang Panjang atau kurangnya pekerjaan [56].

Secara ilmiah, kelelahan terjadi dikarenakan dari deformasi mikrostruktur lokal akibat pembebanan *cyclic compound*, kerapatan dislokasi yang dihasilkan secara signifikan lebih tinggi daripada pembebanan *monotonic* [57]. Selain itu, kelelahan dapat terjadi karena ketidakseimbangan antara energi yang dibutuhkan untuk kontraksi (ATP) dan suplai oksigen oleh aliran darah. suplai oksigen dihambat oleh peningkatan tekanan intramuskular, menghambat aliran darah ke otot yang berkontraksi. Dalam keadaan ini, mekanisme anaerobik untuk pembentukan asam laktat terbentuk, untuk itu otot perlu istirahat untuk memecah asam laktat. Pada penelitian lain, ditemukan juga bahwa kelelahan terjadi dikarenakan peningkatan kadar ion kalium [58].

b. Jenis-jenis kelelahan

- 3) Kelelahan Fisik, ketidakmampuan fisik untuk berfungsi secara optimal. Kelelahan fisik berawal dari kelelahan otot dari aktivitas fisik yang terjadi secara bertahap dan tergantung pada tingkat kebugaran seseorang, serta faktor penyebab lainnya seperti kurangnya waktu tidur dan kondisi kesehatan secara keseluruhan. Kelelahan fisik disebabkan oleh aktivitas fisik, seperti berdiri berjam-jam, berlutut dengan laptop dalam waktu yang cukup lama, mengubah atau mengangkat beban yang berlebihan, dan seperti gejala ketidaknyamanan umum serta kesulitan melakukan aktivitas yang membutuhkan kekuatan fisik. Kelelahan ini selalu dikaitkan dengan pekerjaan fisik dan menyebabkan penurunan kemampuan untuk melakukan aktivitas fisik [59].
- 4) Kelelahan Mental, suatu proses kumulatif yang mengarah pada berkurangnya tingkat usaha, efisiensi, dan kesadaran dalam

aktivitas mental. Hal ini dapat mengakibatkan penurunan produktivitas bahkan kecelakaan. Melakukan tugas yang melelahkan secara kognitif menguras otak dan sumber daya kognitif yang tidak memadai untuk diisi ulang dengan cukup cepat untuk melakukan tugas berikutnya [60]. Kelelahan mental membuat seseorang berada pada suatu kondisi ketika ia kehilangan kemampuan untuk melakukan tugas-tugas yang membutuhkan kerja mental. Kondisi ini ditandai dengan penurunan konsentrasi, berkurangnya toleransi risiko dan kinerja, serta berkurangnya kesadaran. Kelelahan mental dapat diakibatkan oleh pekerjaan yang sangat kurang atau terlalu banyak tugas yang menuntut, dan tugas yang membosankan serta tidak menarik [61].

- 5) Kelelahan Emosional, keadaan emosi konstan yang mengarah pada stress psikologis (jiwa) karena tekanan pekerjaan. Emosi seseorang dapat terkuras dalam menghadapi situasi, masalah, kebutuhan pribadi dan profesional yang terjadi secara berkelanjutan atau terus-menerus. Kelelahan ini berbeda dari gejala biasanya, seperti perasaan cemas ketika memulai pekerjaan. Ketika seseorang mengalami kelelahan emosional, biasanya mereka akan menjadi mudah frustrasi atau marah pada diri sendiri. Kelelahan ini biasanya terjadi apabila seseorang yang merasa kekurangan sumber daya untuk melakukan pekerjaan diperlukan dan merasa lelah serta terkuras secara emosional [62].

c. Gejala kelelahan

Perasaan lelah, mengantuk, dan bosan muncul dikarenakan gejala kelelahan. Gejala kelelahan terbagi menjadi tiga antara lain, adanya gangguan kinerja, hilangnya motivasi, dan kelelahan fisik [63] [64].

- 1) Gangguan kinerja, Gejala ini biasanya ditandai dengan adanya rasa berat di kepala, seluruh bagian tubuh terasa lelah, kaki seperti menarik beban, terus-menerus menguap, kebingungan, lesu, mata

berat, mengantuk, dan ketidakmampuan untuk menyeimbangkan tubuh ketika berdiri.

- 2) Hilangnya motivasi, ditandai dengan kesulitan berpikir, terasa lelah saat ingin berbicara, merasa cemas, tidak mampu berkonsentrasi atau sulit fokus, mudah lupa, kurang percaya diri, tidak mampu mengendalikan sikap, dan kurangnya ketekunan dalam bekerja.
- 3) Kelelahan fisik, penurunan kondisi fisik yang bervariasi pada setiap individu, seperti adanya sakit kepala, bahu kaku, sakit pada bagian punggung, sesak nafas, haus, pening, dan tubuh terasa tidak fit.

2.1.2 Wajah

Wajah merupakan salah satu bagian sistem biometrik, wajah memiliki ciri dan bentuk yang unik, sehingga keunikan ini digunakan sebagai pengenalan dan identitas setiap individu [65]. Wajah setiap orang memiliki bentuk yang berbeda, bentuk wajah dapat dengan mudah untuk dikenali dengan mengenali pola wajah [66]. Wajah adalah bagian depan kepala manusia yang memiliki beberapa fitur mulai dari dahi sampai dagu, meliputi rambut, dahi, alis, pelipis, mata, hidung, mulut, pipi, gigi, kulit, dan dagu. Wajah dapat merepresentasikan ekspresi, penampilan, dan identitas seseorang [67].

a. Ekspresi wajah

Ekspresi wajah merupakan salah satu bentuk komunikasi non-verbal yang menggambarkan perasaan batin seseorang dalam kehidupan sehari-hari [68]. Manifestasi ekspresi wajah ini terjadi secara bersamaan di berbagai bagian wajah, termasuk alis, mata, hidung, mulut, dan dagu, untuk itu setiap fitur berkontribusi pada gambaran keseluruhan ekspresi yang dirasakan [69]. Secara khusus, kondisi mata dan mulut yang terbuka sangat penting dalam merepresentasikan kondisi wajah seseorang hingga mempengaruhi ekspresi wajah secara keseluruhan [70]. Mata maupun

mulut memainkan peran penting dalam menyampaikan emosi dan perasaan seseorang [71].

2.1.3 Pengolahan citra

Teknologi pengolahan citra (image Processing) semakin berkembang, kini manusia mampu menciptakan sistem yang dapat mengenali citra digital. Citra digital adalah citra yang memproses secara digital atau oleh komputer. Pengolahan citra adalah pemrosesan citra yang menggunakan komputer untuk menghasilkan citra dengan kualitas yang lebih baik. Citra kaya akan informasi, tetapi sering kali citra memiliki penurunan kualitas (quality loss), seperti berlubang, noise, buram atau kurang tajam, dan lain sebagainya, sehingga menyebabkan citra sulit untuk ditafsirkan karena berkurangnya informasi dari citra tersebut [72]. Dalam beberapa tahun terakhir, pengenalan citra digital telah menerima banyak perhatian. Berbagai bidang telah menerapkan pengenalan digital untuk bidang dan objek yang berbeda, seperti di bidang pertanian menentukan kualitas cabai yang baik, pengenalan anggrek, pengenalan plat nomor, hingga pengenalan wajah [73].

a. Deteksi objek

Deteksi objek memegang peranan penting dalam visi komputer serta menjadi aplikasi penting di berbagai bidang, mulai dari penelitian ilmiah hingga implementasi industri praktis. Bidang ini meliputi deteksi wajah, identifikasi teks, pengenalan pejalan kaki untuk sistem keselamatan kendaraan, analisis merek dengan pengenalan logo, deteksi konten video yang kompleks, serta analisis citra medis untuk mendiagnosis penyakit [74]. Deteksi objek tidak hanya klasifikasi kategori objek, tetapi juga melakukan prediksi lokasi pada setiap objek dalam suatu gambar dengan menggunakan kotak pembatas (*bounding box*) [75]. Deteksi objek tidak hanya bertujuan untuk menemukan dan mengenali objek dari kelas target yang ditentukan di dalam suatu gambar, tetapi juga melibatkan lokalisasi dan klasifikasi objek dari satu atau beberapa kelas target [76].

2.2 Framework, Algoritma, dan Evaluasi

2.2.1 Framework

a. CRISP-DM

Cross-Industry Standard Process for Data Mining telah menjadi standar dalam proses data mining, berfungsi sebagai strategi umum yang sistematis dalam pemecahan masalah pada sektor bisnis maupun penelitian [77]. Manfaat dari penerapan CRISP-DM adalah pengurangan biaya dan waktu dalam proyek *data mining*, meminimalisir kebutuhan pengetahuan, serta memfasilitasi proses seperti pelatihan yang lebih cepat, transfer pengetahuan yang lebih efisien, penyusunan dokumentasi yang lebih terstruktur, serta mengaplikasikan praktik terbaik dalam proyek-proyek *data mining* [78]. CRISP-DM memiliki beberapa karakteristik yang menjadikannya sangat berguna dalam penggalian dan analisis bukti. *Framework* ini menguraikan menyediakan model proses yang bisa diterapkan secara universal, dengan memberikan fleksibilitas untuk disesuaikan dengan kebutuhan spesifik proyek [79]. CRISP-DM terdiri dari enam fase yang berjalan secara iteratif sebagai berikut [80],

1) *Business understanding*

Fase pemahaman bisnis memfokuskan pada identifikasi tujuan proyek. Tahap ini melibatkan diskusi dengan pemangku kepentingan untuk memastikan kejelasan terkait dengan hal yang ingin dicapai, serta merencanakan pendekatan proyek yang sesuai dengan kebutuhan bisnis, sekaligus merancang rencana pelaksanaan yang efisien.

2) *Data understanding*

Fase pemahaman data adalah tahapan pengumpulan data relevan, memastikan kualitas data memenuhi standar yang dibutuhkan untuk analisis, serta melibatkan eksplorasi mendetail untuk mengidentifikasi karakteristik data serta potensi masalah pada data, seperti nilai yang hilang, outlier, atau kesalahan dalam data.

3) *Data preparation*

Fase persiapan data melibatkan transformasi data mentah menjadi dataset untuk pemodelan. Tahapan ini mencakup pembersihan data untuk memperbaiki data yang bermasalah, normalisasi, serta seleksi fitur untuk menentukan variabel paling relevan.

4) *Modeling*

Fase pemodelan adalah tahapan untuk membangun model dengan dataset yang sudah dipersiapkan. Proses ini termasuk pemilihan algoritma yang cocok dan penyesuaian parameter untuk mencapai kinerja optimal. Fase ini berakhir dengan pemilihan model terbaik berdasarkan evaluasi kinerja model, lalu diuji dalam tahap evaluasi lebih lanjut.

5) *Evaluation*

Fase evaluasi bertujuan untuk menguji kinerja model untuk menentukan seberapa baik model dalam memenuhi tujuan bisnis. Evaluasi ini melibatkan analisis terhadap hasil yang dihasilkan oleh model untuk memastikan bahwa temuan yang diperoleh relevan.

6) *Deployment*

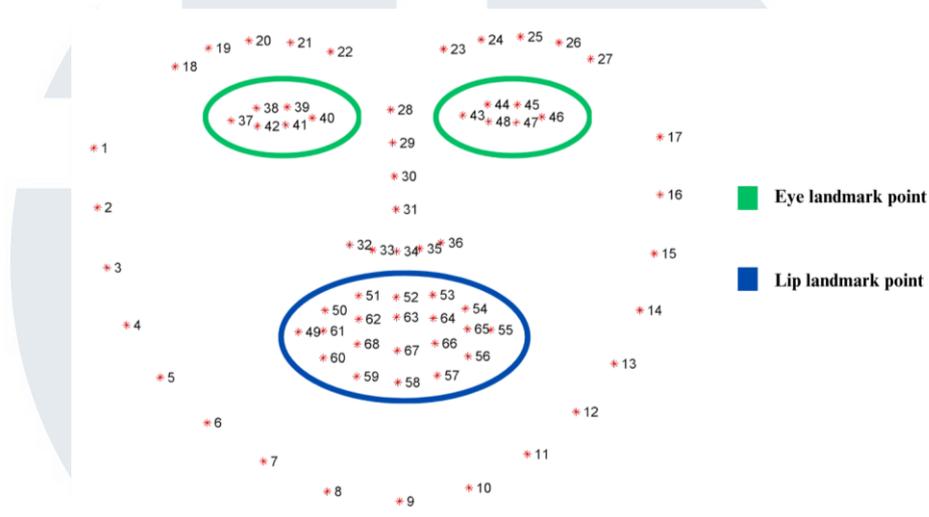
Fase *deployment* melibatkan integrasi model yang telah dikembangkan untuk diimplementasikan di kehidupan sehari-hari. Tahapan ini mencakup pengendalian teknis untuk efisiensi penggunaan model, serta penyusunan dokumentasi detail yang menjelaskan penggunaan dan *maintenance* model.

2.2.2 Algoritma

a. **DLib**

Dlib adalah *library* C++ yang berfokus pada penggunaan lintas platform dalam pengembangan dengan *image processing*, *machine learning*, dan manipulasi struktur data. Dlib sering digunakan karena keakuratannya dalam deteksi wajah dan kemampuannya secara konsisten

melokalisasi fitur wajah, serta prediksi bentuk wajah dengan menggunakan model deteksi *landmark* wajah 68 titik koordinat [5]. Integrasi Dlib memungkinkan sistem untuk melakukan deteksi serta pengenalan wajah [81].



Gambar 2. 1 68 *Facial Landmark*
Sumber: [82] [83]

Gambar 2. 1 merepresentasikan 68 titik koordinat wajah dengan nomor yang merupakan output deteksi dari Dlib. Titik-titik koordinat ini, terutama pada bagian mata dan mulut sangat penting dalam penelitian ini karena digunakan sebagai perhitungan *Eye Aspect Ratio* (EAR) dan *Mouth Aspect Ratio* (MAR) [5] [83],

1) EAR

Eye Aspect Ratio (EAR) adalah metode yang digunakan untuk mengukur tingkat bukaan mata seseorang dengan memanfaatkan titik-titik koordinat di sekitar mata, guna membandingkan jarak vertikal dan horizontal mata. Secara khusus, EAR dihitung dengan menggunakan enam titik koordinat yang terletak di sekitar mata, titik P1 dan P4 digunakan untuk menentukan jarak horizontal, sedangkan titik P2, P3, P5, dan P6 digunakan untuk mendapatkan jarak vertikal. EAR dihitung dengan mengambil

rasio jarak vertikal rata-rata terhadap jarak horizontal. Perhitungan tersebut ditunjukkan pada Rumus 2. 1.

$$EAR = \frac{||P1-P5||+||P2-P4||}{2||P0-P3||} \quad (2. 1)$$

2) MAR

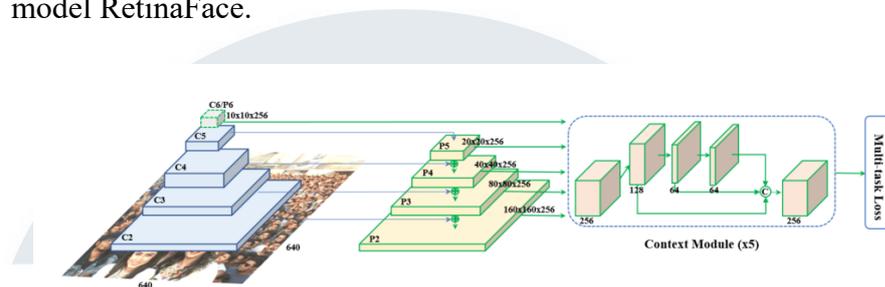
Mouth Aspect Ratio (MAR) merupakan metode pengukuran yang serupa dengan *Eye Aspect Ratio* (EAR), tetapi digunakan untuk menilai bukaan mulut berdasarkan titik-titik koordinat pada area mulut. MAR dihitung dengan menggunakan titik koordinat yang terletak di sekitar mulut, titik P1 dan P7 digunakan untuk menentukan lebar mulut, yaitu jarak horizontal antara sudut mulut kiri dan kanan. Jarak vertikal diukur antara bibir atas dan bibir bawah dengan titik P0 dan P4, P2 dan P6, serta P3 dan P5, seperti yang ditunjukkan pada Rumus 2. 2.

$$MAR = \frac{||P0-P4||+||P2-P6||+||P3-P5||}{3||P1-P7||} \quad (2. 2)$$

b. RetinaFace

RetinaFace adalah model pengenalan wajah selangkah lebih baru dan canggih yang ditawarkan oleh InsightFace pada tahun 2019 yang didasarkan pada struktur RetinaNet [84]. Pada awal rilis, implementasi didasarkan pada mxnet. Kemudian, setelah diterbitkan kembali oleh Stanislas Bertrand, implementasi ulang didasarkan pada tensorflow [85]. RetinaFace melakukan pelokalan wajah berbasis pixel pada rasio wajah yang berbeda menggunakan *extra-supervised and self-supervised multi-task learning* untuk memprediksi skor wajah, lima fitur wajah, serta memprediksi posisi 3D dan korespondensi wajah untuk masing-masing pixel. RetinaFace dapat dengan mudah menangani wajah dalam berbagai pose, namun sulit untuk kondisi yang terlalu kompleks. RetinaFace dapat berjalan secara real-time pada satu inti CPU untuk gambar beresolusi VGA [33]. RetinaFace dirancang berdasarkan piramida fitur dengan

context module yang berdiri sendiri. Gambar 2. 2 menunjukkan struktur model RetinaFace.



Gambar 2. 2 Struktur model RetinaFace
Sumber: [86]

RetinaFace memanfaatkan pendekatan *multi-task learning* dalam fungsi *loss*-nya untuk meningkatkan kinerja deteksi wajah, pengenalan landmark, dan perhitungan estimasi pose secara bersamaan. Fungsi *loss* ini merupakan gabungan dari tiga komponen: *loss* untuk deteksi wajah (L_{det}), *loss* untuk landmark ($L_{landmarks}$), dan *loss* untuk pose (L_{pose}), yang dirumuskan dengan persamaan pada Rumus 2. 3:

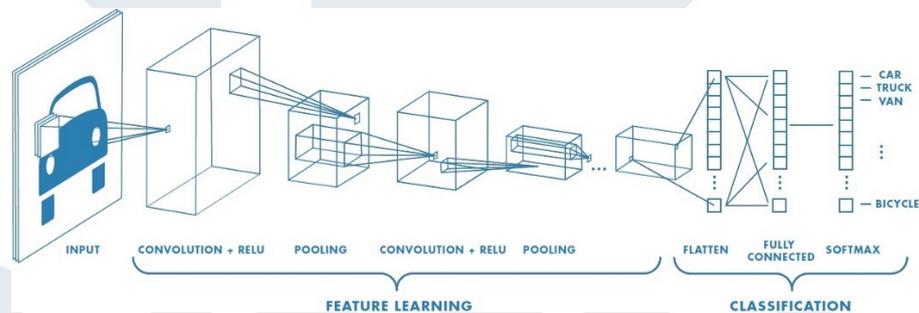
$$L_{total} = L_{det} + \lambda_1 L_{landmark} + \lambda_2 L_{pose} \quad (2.3)$$

λ_1 dan λ_2 menentukan bobot relatif dari *loss landmark* dan pose dalam kalkulasi *loss* total. Komposisi RetinaFace terdiri dari beberapa lapisan utama, termasuk:

- 1) *Backbone Layer*: Umumnya memanfaatkan ResNet atau MobileNet yang sudah dilatih untuk mengambil fitur dasar.
- 2) *Feature Pyramid Network* - FPN: Memperbaiki kemampuan pendeteksian lintas skala objek.
- 3) *Context Module*: Meningkatkan detail fitur wajah dengan konteks yang lebih luas, sehingga akurasi meningkat.
- 4) *Multi-task Loss*: Digunakan untuk melakukan deteksi wajah, identifikasi landmark, dan perhitungan pose secara bersamaan

c. CNN

CNN adalah algoritma deep learning yang berfokus pada gambar dan teks, algoritma ini menjadi salah satu algoritma deep learning yang sangat populer untuk masalah pengenalan gambar. CNN adalah tiruan dari jenis jaringan saraf yang digunakan dalam pengenalan dan pemrosesan gambar [87]. CNN meniru cara sel saraf berkomunikasi dengan neuron yang saling berhubungan, dan CNN memiliki konstruksi yang sama. Keunikan CNN dari *Neural Network* lainnya operasi konvolusi yang menerapkan filter ke setiap bagian dari input sebelumnya untuk mengekstrak pola. CNN terdiri dari lapisan input, setidaknya satu lapisan convolutional, dan lapisan output [88]. Gambar 2. 3 adalah ilustrasi dari arsitektur CNN, namun tidak semua penelitian menerapkan keseluruhan arsitekturnya, beberapa memodifikasinya agar dapat menghasilkan akurasi yang lebih baik [89].



Gambar 2. 3 Arsitektur CNN
Sumber: Matlab

Arsitektur yang terdapat pada CNN terbagi menjadi 2 bagian besar, yaitu *Feature Learning* dan *Classification*.

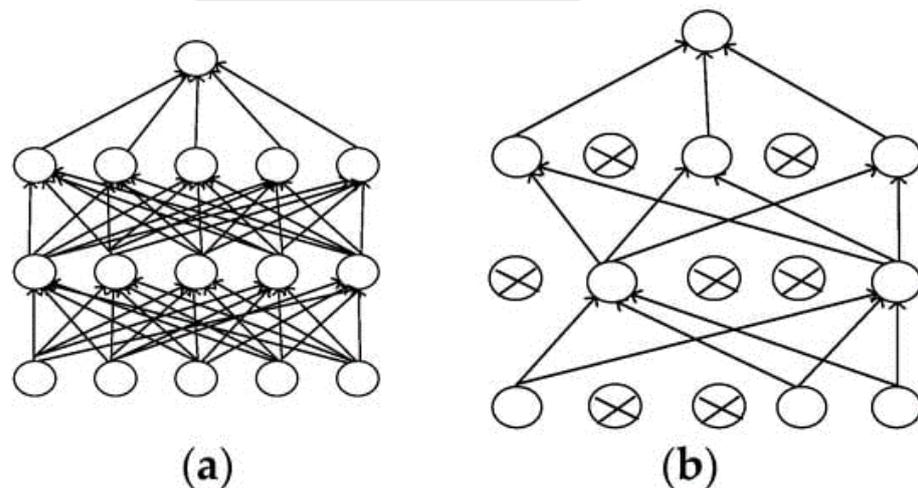
1) *Feature Learning*

Proses yang melakukan *encoding* dari sebuah input gambar menjadi *feature* dalam bentuk angka yang merepresentasikan gambar tersebut (*Feature Extraction*). Pada bagian ini terdapat dua layer *Convolutional Layer* dan *Pooling Layer*.

2) Classification

Bagian yang mengumpulkan semua hasil dari dari *feature map* dengan melakukan *flatten* atau *reshape* menjadi sebuah vektor agar dapat digunakan sebagai input *fully-connected*. *Fully Connected layer* akan mengolah data sehingga dapat diklasifikasikan.

Selanjutnya, terdapat teknik regularisasi *neural network* yaitu *Dropout Regularization*, beberapa neuron dipilih secara *random* dan tidak digunakan selama *training* dilakukan. *Dropout* adalah proses yang akan mencegah terjadinya *overfitting* dan juga mempercepat proses dengan cara menghilangkan *neuron* yang dipilih secara acak. Gambar 2. 4 adalah ilustrasi perbedaan ketika teknik *Dropout* belum dilakukan dan sudah dilakukan. Gambar 2. 4 (a) merupakan *neural network* tanpa teknik *dropout*, sedangkan Gambar 2. 4 (b) merupakan *neural network* dengan teknik *dropout* [90].

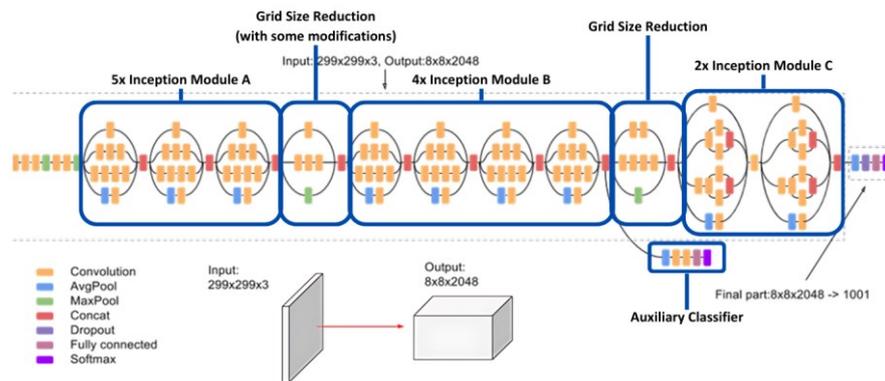


Gambar 2. 4 Perbedaan *neural network* ketika dilakukan teknik *Dropout*
Sumber: [90]

d. InceptionV3

Model InceptionV3 adalah arsitektur pengenalan citra yang dikembangkan oleh Google dan telah dilatih pada lebih dari 1000 kelas dan lebih dari 1,4 juta gambar. Proses klasifikasi dalam InceptionV3

melibatkan *fully-connected* dan *softmax layers*. Dalam proses pelatihannya, model ini mengekstrak fitur penting dari gambar input dan kemudian menerapkan klasifikasi berdasarkan fitur tersebut [91]. Arsitektur InceptionV3 dirancang untuk mengoptimalkan penggunaan data komputasi dengan memperkenalkan konsep *Convolution* terfaktor yang mengurangi jumlah parameter dalam jaringan, sehingga meningkatkan efisiensi komputasi. Konvolusi besar digantikan dengan konvolusi yang lebih kecil untuk menjaga efisiensi jaringan, misalnya konvolusi asimetris seperti pergantian konvolusi 3x3 dengan kombinasi konvolusi 1x3 diikuti dengan konvolusi 3x1, sehingga jumlah parameter berkurang 33%. Selain itu, pengurangan ukuran grid dengan *pooling* juga mendukung pengurangan dimensi data [92].



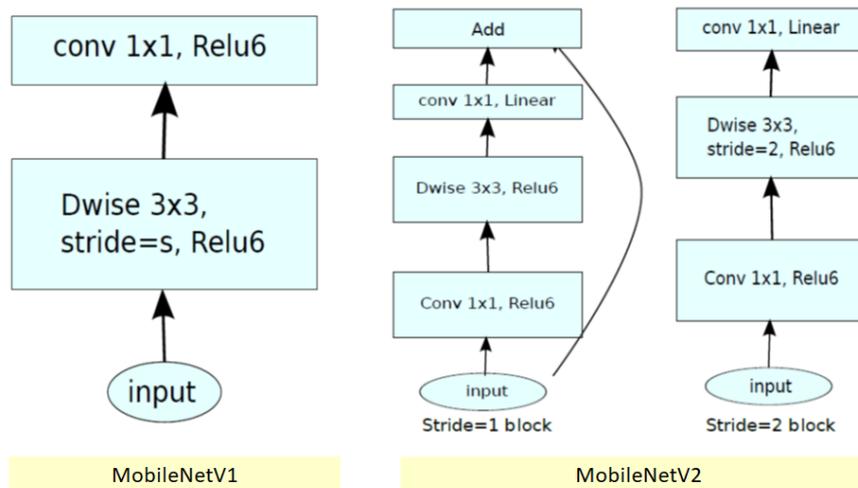
Gambar 2. 5 Arsitektur InceptionV3
Sumber: [93]

Gambar 2. 5 Adalah representasi dari komposisi berbagai modul dalam arsitektur InceptionV3 sehingga membentuk jaringan. Komposisi tersebut terdiri dari Modul A - *5x Inception* yang menjadi lima *block* pertama dari model yang masing-masing memiliki beberapa operasi *convolution* dan *pooling* yang dijalankan secara paralel, lalu hasilnya akan digabungkan. Selain itu, terdapat Modul B - *4x Inception* yang memiliki struktur serupa namun dengan beberapa modifikasi pada filter *convolution* dan *pooling*. Selanjutnya, *Grid Size Reduction* diterapkan di antara modul A dan B, serta setelah modul B, pengurangan ukuran grid

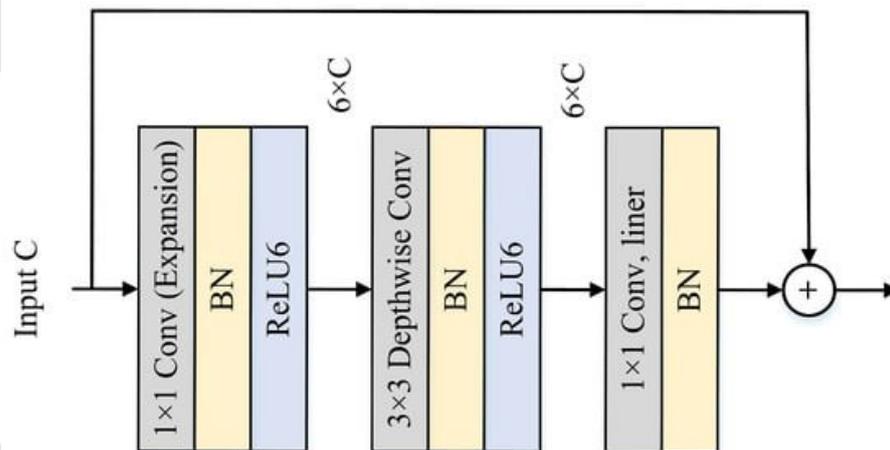
biasanya melibatkan operasi *convolution* dan *pooling* dengan *stride* yang lebih besar untuk mengurangi dimensi spasial dari data yang masuk. Modul terakhir adalah $2x$ *Inception*, disini setiap modul dirancang untuk penanganan fitur pada level yang lebih tinggi dengan ukuran grid yang sudah dikurangi. Selain itu, terdapat *classifier* tambahan yang biasanya digunakan dalam proses pelatihan untuk memberikan pembelajaran tambahan dan mencegah terjadinya masalah *vanishing gradients* di lapisan yang lebih dalam di akhir arsitektur ini, ditambahkan lapisan *fully connected* dan lapisan *softmax* yang digunakan untuk klasifikasi akhir [93].

e. MobileNetV2

MobileNetV2 merupakan evolusi dari versi sebelumnya yang mengadopsi teknik *Depthwise Separable Convolutions* (DSC) yang meningkatkan efisiensi dan efektivitas, sehingga MobileNetV2 dapat meminimalkan ukuran model dan mempercepat komputasi tanpa mengorbankan akurasi. Arsitektur ini dikembangkan dengan tujuan mengatasi degradasi informasi yang biasa terjadi pada lapisan non-linier dengan memanfaatkan *Shortcut Connection* atau *Inverted Residuals*, yaitu sebuah inovasi struktur ketika urutan operasi *Convolution* ditukar. Selain itu, MobileNetV2 mengadopsi *linier Bottleneck*, yaitu menggunakan lapisan linier tanpa fungsi aktivasi non-linier seperti ReLU di akhir *Residual Block*. Pendekatan ini mempertahankan informasi penting saat transisi dari ruang dimensi tinggi ke ruang dimensi rendah yang rentan dirusak oleh ReLU [94]. Oleh karena itu, MobileNetV2 memperkenalkan *Bottleneck Residual Block* dengan urutan 1×1 (*Expansion*) $\rightarrow 3 \times 3$ (*Depthwise Separable Convolutions*) $\rightarrow 1 \times 1$ (*Compression*) dengan transformasi linier yang berbeda dari *Residual Block* standar yang melakukan *Compression* sebelum *Expansion*. *Block* tersebut akan menggantikan aktivasi ReLU untuk menghindari *Information Loss* seperti pada Gambar 2. 6.



Gambar 2. 6 Perbedaan MobileNetV1 dan MobileNetV2
Sumber: [95]



Gambar 2. 7 *Residual Block* pada arsitektur MobileNetV2
Sumber: [92]

Gambar 2. 7 menggambarkan komponen dari *Residual Block* terbalik yang ada dalam arsitektur MobileNetV2. Komposisi awal dari lapisan MobileNetV2 adalah *Convolution* 1x1 untuk ekspansi *channel*, diikuti dengan *Batch Normalization* dan aktivasi ReLU6. Setelah itu, terdapat lapisan *Depthwise Convolution* 3x3 mengisolasi dan memproses setiap *channel* secara terpisah untuk efisiensi komputasi, diikuti dengan *Batch Normalization* dan ReLU6 [92].

f. *Transfer Learning*

Transfer Learning merupakan salah satu cara untuk meningkatkan efisiensi dan efektivitas dalam pembuatan model, khususnya *deep learning*. *Transfer Learning* memanfaatkan pengetahuan yang telah diperoleh dari pemecahan masalah pada satu sumber untuk meningkatkan proses pembelajaran pada sumber yang berbeda tapi terkait (target) [96] [97]. *Transfer Learning* memungkinkan model yang telah dilatih untuk mengenali objek dalam foto dengan menyesuaikan agar dapat mengenali objek dalam video, meminimalisir kebutuhan untuk pelatihan dari awal. *Transfer Learning* sering kali digunakan untuk menyesuaikan model-model *deep learning* yang sudah ada dan terlatih pada dataset besar untuk pelatihan [98]. *Transfer Learning* menjadi jalan pintas dalam pengembangan model karena dapat mengembangkan model dengan model yang sudah memiliki arsitektur dasar dan bobot yang telah melalui pelatihan awal, sehingga pengembangan model dapat berfokus pada penyesuaian model agar sesuai dengan kebutuhan spesifik penelitian. *Transfer Learning* menjadi strategi untuk menghemat waktu dan sumber daya, serta meningkatkan kemungkinan model baru untuk bekerja dengan baik dalam berbagai kondisi dataset [99].

2.2.3 Evaluasi

a. Confusion Matrix

Confusion Matrix adalah metode yang digunakan untuk melakukan pengukuran suatu model klasifikasi dengan cara mencari beberapa nilai, diantaranya nilai *precision*, *recall*, dan *accuracy* dari sebuah model klasifikasi agar dapat melakukan evaluasi benar atau salah-nya model tersebut. *Confusion matrix* menghasilkan matrix yang berisikan hasil prediksi dari nilai-nilai yang dicari dengan pembagian *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) sebagaimana ditunjukkan pada Tabel 2. 2 [100].

Tabel 2. 2 *Confusion Matrix* [90]

	Predicted		
		Yes	No
Actual	Yes	TP	FN
	No	FP	TN

Dari sebuah *confusion matrix*, akan dapat dihitung nilai akurasi untuk mengukur seberapa sering model memberikan prediksi yang benar, presisi untuk melihat seberapa banyak prediksi positif yang benar dari seluruh prediksi positif, *recall* untuk mengukur seberapa baik model dapat menemukan semua sampel positif yang sebenarnya, dan *f1-score* untuk melihat gambaran kinerja yang seimbang berdasarkan presisi dan *recall* dari model klasifikasi dengan menggunakan rumus *Accuracy* yang ditunjukkan pada Rumus 2. 4, Rumus 2. 5 *Precision*, Rumus 2. 6 *Recall* dan Rumus 2. 7 *F1-score* [101].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2. 4)$$

$$Precision = \frac{TP}{TP+FP} \quad (2. 5)$$

$$Recall = \frac{TP}{TP+FN} \quad (2. 6)$$

$$F1 - Score = 2 \times \left(\frac{Precision \times Recall}{Precision + Recall} \right) \quad (2. 7)$$

Keterangan Rumus:

TP : *True Positive* merupakan jumlah kasus yang diprediksi positif dan sebenarnya positif

TN : *True Negative* merupakan jumlah kasus yang diprediksi negatif dan sebenarnya negatif

FP : *False Negative* merupakan jumlah kasus yang diprediksi positif, tetapi sebenarnya negatif

FN : *False Negative* merupakan jumlah kasus yang diprediksi negatif, tetapi sebenarnya positif

2.3 Tools

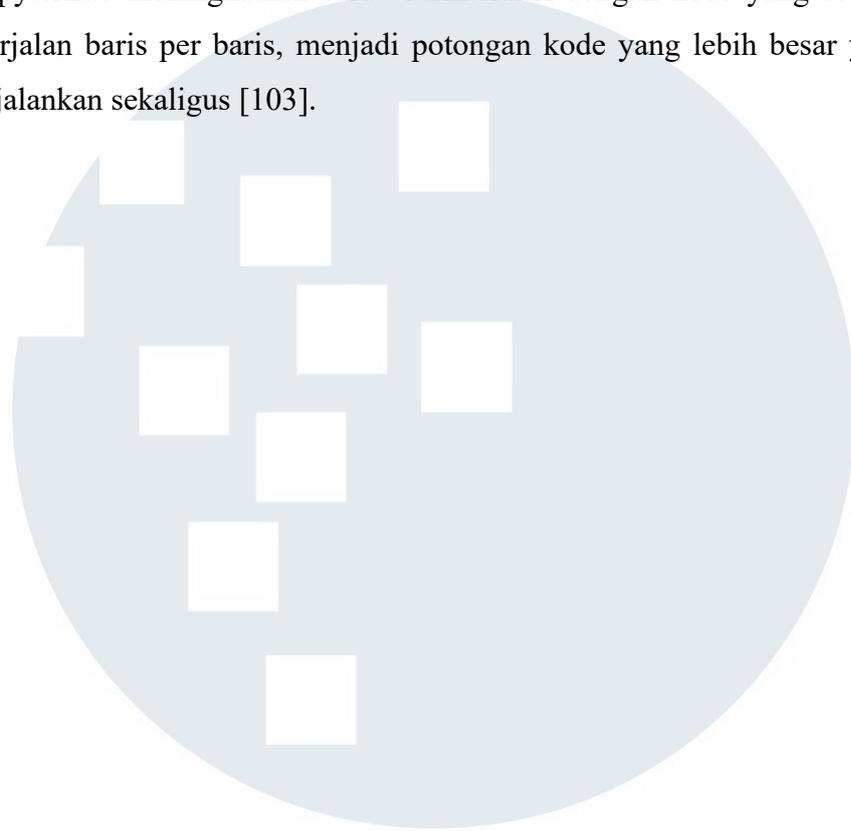
2.3.1 Python

Python adalah bahasa yang digunakan untuk melakukan tugas-tugas kompleks, bahasa ini adalah salah satu bahasa yang dapat diandalkan. Python merupakan bahasa pemrograman interpretative mudah dipelajari, python berfokus pada keterbacaan kode, dan bersifat multi-platform. Python memiliki tujuan untuk menjadi bahasa yang menggabungkan keterampilan, fungsionalitas, dengan *syntax* kode yang sangat jelas, dan *library* standar yang besar serta komprehensif. Python biasanya digunakan untuk berbagai tujuan pengembangan *software* dan dapat dijalankan di berbagai platform sistem operasi [102]. Python biasa digunakan untuk pengenalan gambar, pengenalan wajah, dan deteksi gerakan karena memiliki *library* dan fitur yang sangat cerdas. Logo bahasa pemrograman python dapat dilihat pada.

2.3.2 JupyterLab

Jupyter merupakan platform pengembangan yang dinamakan berdasarkan ketiga bahasa pemrograman, yaitu Julia, Python, dan R. Perkembangan teknologi yang ada, membuat platform ini berkembang untuk mendukung puluhan bahasa lain, serta memungkinkan pengembangan dan penelitian dalam dikerjakan secara bersamaan dalam *notebook* yang sama [103]. Jupyterlab merupakan evolusi dari Jupyter Notebook yang meningkatkan kemampuan kerja sama dan multifungsi dengan berkembangnya antarmuka web *modern* sehingga memungkinkan pengguna tidak hanya menjalankan *notebook* tetapi juga mengedit teks, mengakses terminal jarak jauh, hingga menjalankan aplikasi GUI [104]. JupyterLab memiliki

kemampuan untuk pemecahan dan eksekusi *code block* yang lebih besar. JupyterLab meningkatkan cara berinteraksi dengan kode yang sebelumnya berjalan baris per baris, menjadi potongan kode yang lebih besar yang bisa dijalankan sekaligus [103].



UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA