

## Evaluasi Kinerja Algoritma Klasifikasi Deep Learning dalam Prediksi Diabetes

Tuahta Hasiholan Pinem<sup>1\*</sup>, Zico Pratama Putra<sup>2</sup>

<sup>1,2</sup>Fakultas Teknologi Informasi, Universitas Nusa Mandiri  
Email: <sup>1</sup>14220020@nusamandiri.ac.id, <sup>2</sup>zico.zpp@nusamandiri.ac.id  
Penulis Korespondensi\*

(received: 25-06-24, revised: 04-09-24, accepted: 09-12-24)

### Abstrak

Penelitian yang bertujuan untuk mengembangkan dan mengevaluasi algoritma model prediksi diabetes telah dilakukan dengan menggunakan algoritma model *K-Nearest Neighbor Classifier*, *Naive Bayes*, *Regresi Logistik*, *SVM*, dan *Neural Network*. Dataset yang digunakan didapatkan dari Kaggle yang terdiri dari 768 data pasien yang dibagi menjadi data training 60%, data validation 20%, dan data test 20%. Hasil penelitian menunjukkan bahwa akurasi tertinggi diperoleh oleh model *Regresi Logistik* dan *Neural Network*, masing-masing sebesar 73% dan 72%. Model *Regresi Logistik* unggul dalam presisi untuk kelas non-diabetes dan *recall* untuk kelas diabetes, sedangkan model *Neural Network* menunjukkan keseimbangan performa yang baik antara presisi dan *recall* untuk kedua kelas. Model *Naive Bayes* juga menunjukkan performa yang kompetitif dengan akurasi 72% dan *recall* tinggi untuk kelas diabetes, model ini dapat menjadi pilihan yang baik dalam situasi yang memprioritaskan deteksi positif. Kinerja yang lebih rendah ditunjukkan oleh model KNN dan SVM jika dibandingkan dengan model lainnya. Masalah utama yang diangkat dalam penelitian ini adalah pentingnya meningkatkan akurasi prediksi diabetes untuk mendukung deteksi dini dan pengobatan. Secara keseluruhan, model *Regresi Logistik* dan *Neural Network* diidentifikasi sebagai model yang paling potensial untuk prediksi diabetes, dengan *Regresi Logistik* menunjukkan efektivitas yang tinggi dalam mengidentifikasi kasus non-diabetes, sementara *Neural Network* memberikan keseimbangan performa yang baik di kedua kelas.

**Kata Kunci:** Diabetes, *K-Nearest Neighbors*, *Machine Learning*, *Naive Bayes*, *Neural Network*, *Regresi Logistik*, *Support Vector Machine*

### Abstract

The research focused on developing and evaluate diabetes prediction models utilizing algorithms such as *K-Nearest Neighbor Classifier*, *Naive Bayes*, *Logistic Regression*, *SVM*, and *Neural Network*. The dataset used was obtained from Kaggle, consisting of 768 patient records, divided into 60% training data, 20% validation data, and 20% test data. The results showed that the highest accuracy was achieved by the *Logistic Regression* and *Neural Network* models, at 73% and 72%, respectively. The *Logistic Regression* model excelled in precision for the non-diabetes class and recall for the diabetes class, while the *Neural Network* model demonstrated a good balance between precision and recall for both classes. The *Naive Bayes* model also showed competitive performance with 72% accuracy and high recall for the diabetes class, making it a good choice in situations prioritizing positive detection. Lower performance was observed in the *KNN* and *SVM* models compared to the others. The primary issue highlighted in this study is the importance of improving diabetes prediction accuracy to support early detection and personalized treatment. Overall, the *Logistic Regression* and *Neural Network* models were identified as the most promising for diabetes prediction, with *Logistic Regression* showing high effectiveness in identifying non-diabetes cases, while *Neural Network* provided a well-balanced performance across both classes.

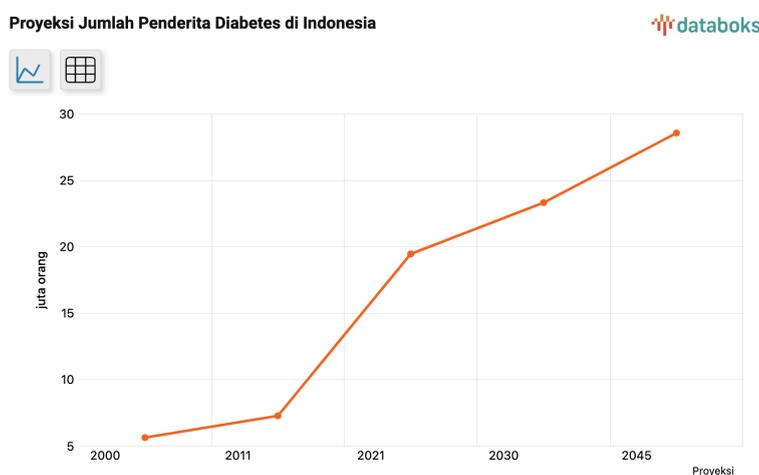
**Keywords:** Heart Disease, *K-Nearest Neighbors*, *Machine Learning*, *Naive Bayes*, *Neural Network*, *Regresi Logistik*, *Support Vector Machine*

## 1. PENDAHULUAN

Diabetes merupakan penyakit kronis yang dikenal dengan gula darah tinggi dan merupakan salah satu penyakit tidak menular (PTM) yang membahayakan tubuh manusia [1]. Prediksi diabetes merupakan bidang penelitian yang penting, dengan studi berfokus pada penggunaan berbagai model dan algoritma untuk memperkirakan kemungkinan mengembangkan prediksi penyakit. Penelitian telah menunjukkan bahwa faktor-faktor seperti "Kehamilan," "Glukosa," "BMI," "Fungsi Silsilah Diabetes," "Usia," dan "Tekanan Darah" merupakan penentu signifikan diabetes [2], Setiap tahunnya, banyak pasien yang kehilangan nyawa karena

menderita penyakit diabetes. Penyakit ini menyebabkan komplikasi mikrovaskular (*retinopati, nefropati, neuropati*) dan makrovaskular (penyakit koroner, penyakit pembuluh darah perifer, kejadian serebrovaskular) yang mempengaruhi banyak sistem organ dan memerlukan strategi pengobatan yang komprehensif. Diabetes tipe 1 memerlukan terapi insulin, sedangkan pengobatan diabetes tipe 2 berfokus pada perubahan gaya hidup, pengelolaan berat badan, obat penurun glukosa, dan pengurangan risiko kardiovaskular. Kondisi medis yang ditandai oleh kadar gula darah yang tinggi dan dapat dikategorikan menjadi beberapa tipe utama yaitu Diabetes Tipe 1, Diabetes Tipe 2, Diabetes Gestasional, Diabetes Monogenik dan Diabetes Sekunder. Masing-masing tipe diabetes ini memerlukan pendekatan penanganan yang berbeda untuk mengelola kadar gula darah dan mencegah komplikasi jangka Panjang [3]. Perawatan jangka panjang yang efektif memerlukan kolaborasi antara pasien dan profesional kesehatan dalam sistem layanan kesehatan yang kuat [4]. Berbagai upaya telah dilakukan untuk mengurangi penyebaran penyakit yang dipengaruhi oleh gaya hidup sehari-hari.

Data dari International Diabetes Federation (IDF) menunjukkan jumlah penderita diabetes di dunia pada tahun 2021 mencapai 537 juta. Angka ini diprediksi akan terus meningkat mencapai 643 juta di tahun 2030 dan 783 juta pada tahun 2045. Menurut IDF, Indonesia menduduki peringkat kelima negara dengan jumlah diabetes terbanyak dengan 19,5 juta penderita di tahun 2021 dan diprediksi akan menjadi 28,6 juta pada 2045, seperti ditunjukkan pada Gambar 1. Persoalan ini menjadi perhatian dari Kementerian Kesehatan, mengingat diabetes melitus merupakan ibu dari segala penyakit.



Gambar 1. Proyeksi jumlah diabetes di Indonesia [5]

Gambar 1 menunjukkan bahwa jumlah penderita diabetes di Indonesia sangat tinggi dan merupakan beban yang sangat besar yang dapat ditangani sendiri oleh dokter spesialis/subspesialis atau bahkan seluruh tenaga kesehatan yang ada. Mengingat DM berdampak pada kualitas sumber daya manusia dan sangat meningkatkan biaya kesehatan, semua pihak, baik masyarakat maupun pemerintah, harus ikut serta dalam pemberantasan DM, khususnya upayanya pencegahan. Menurut American Diabetes Association (ADA) 2010, diabetes adalah sekelompok penyakit metabolik yang ditandai dengan hiperglikemia akibat kelainan sekresi insulin, kerja insulin, atau keduanya. Berbagai upaya telah dilakukan untuk mengurangi penyebaran penyakit yang dipengaruhi oleh gaya hidup sehari-hari.

Upaya pencegahan dan diagnosis dini penyakit DM perlu ditingkatkan, sehingga penting untuk memahami kriteria diagnostik *diabetes mellitus* (DM) berdasarkan kadar glukosa darah, seperti yang ditunjukkan dalam Tabel 1.

Tabel 1. Kadar glikosa darah sewaktu dan puasa sebagai diagnosis DM (mg/dL)

Jenis Pengambilan Sampel		Bukan DM	Belum pasti DM	DM
Kadar Glukosa darah sewaktu (mg/dL)	Plasma vena	<100	100 - 199	≥ 200
	Darah kapiler	<90	90 - 199	≥ 200
Kadar glukosa darah puasa (mg/dL)	Plasma vena	<100	100 - 125	≥ 126
	Darah kapiler	<90	90 - 99	≥ 100

Tabel 1 menunjukkan kriteria diagnosis diabetes berdasarkan kadar glukosa darah dalam berbagai kondisi dan jenis pengambilan sampel. Untuk glukosa darah sewaktu, plasma vena dianggap tidak diabetes jika kadarnya kurang dari 100 mg/dL, belum pasti diabetes jika kadarnya antara 100-199 mg/dL, dan diabetes jika kadarnya 200 mg/dL atau lebih. Pengukuran darah kapiler dianggap tidak diabetes jika kurang dari 90 mg/dL, belum pasti diabetes jika antara 90-199 mg/dL, dan diabetes jika 200 mg/dL atau lebih. Untuk glukosa darah puasa, plasma vena dianggap tidak diabetes jika kadarnya kurang dari 100 mg/dL, belum pasti diabetes jika antara 100-125 mg/dL, dan diabetes jika 126 mg/dL atau lebih. Pengukuran darah kapiler dianggap tidak diabetes jika kurang dari 90 mg/dL, belum pasti diabetes jika antara 90-99 mg/dL, dan diabetes jika 100 mg/dL atau lebih [6].

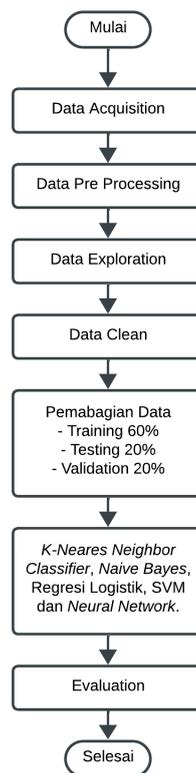
Di sektor kesehatan, teknologi mulai memainkan peran penting melalui kecerdasan buatan (AI). Meskipun tidak ada yang bisa menggantikan profesi petugas kesehatan, AI diciptakan untuk mendorong perubahan dan meningkatkan hasil yang baik [7]. Integrasi algoritma pembelajaran mendalam dengan data medis tidak hanya akan meningkatkan keakuratan diagnosis diabetes, tetapi juga membuka jalan bagi deteksi dini dan strategi pengobatan yang dipersonalisasi. Hal ini pada akhirnya memberikan manfaat bagi pasien dengan meminimalkan risiko yang terkait dengan penyakit tersebut. Penelitian ini bertujuan untuk mengembangkan dan mengevaluasi model prediksi dengan memanfaatkan data tersebut untuk membantu menguji data pasien diabetes positif dan negatif menggunakan metode *K-Nearest Neighbor Classifier*, *Naive Bayes*, Regresi Logistik, SVM, dan *Neural Network*.

Penelitian sebelumnya telah banyak dilakukan untuk mengoptimalkan prediksi diabetes dengan membandingkan berbagai algoritma klasifikasi, seperti *K-Neighbors Classifier*, *Naive Bayes*, Regresi Logistik, dan *Support Vector Machine* (SVM) [8]. Studi-studi ini umumnya menggunakan dataset seperti *Pima Indians Diabetes Database* untuk mengevaluasi kinerja model dalam hal akurasi, presisi, recall, dan F1-score. Misalnya, penelitian oleh Patel dkk. menemukan bahwa SVM menunjukkan kinerja terbaik dalam hal akurasi dan F1-score dibandingkan dengan algoritma lain [9]. Di sisi lain, *Naive Bayes* sering kali lebih cepat dalam proses pelatihan dan prediksi, meskipun dengan akurasi yang sedikit lebih rendah. *K-Nearest Neighbor Classifier* dikenal karena kesederhanaannya dan performanya yang kompetitif pada dataset kecil, sementara Regresi Logistik sering digunakan karena interpretabilitasnya yang tinggi. Keseluruhan hasil dari berbagai penelitian menunjukkan bahwa tidak ada satu algoritma yang unggul secara universal, dan pemilihan model terbaik sering kali bergantung pada karakteristik spesifik dataset dan tujuan analisis.

Studi sebelumnya yang membandingkan berbagai algoritma klasifikasi seperti *K-Neighbors Classifier*, *Naive Bayes*, Regresi Logistik, dan SVM, penelitian ini menambahkan *Neural Network* ke dalam perbandingan algoritma, yang belum banyak dilakukan secara komprehensif dalam studi sebelumnya. Penelitian ini juga mengeksplorasi kombinasi parameter *Neural Network*, termasuk jumlah *node*, *learning rate*, dan *batch size*, untuk menganalisis pengaruhnya terhadap kinerja model, yang mungkin belum dieksplorasi secara mendalam dalam konteks prediksi diabetes. Penelitian ini menekankan pentingnya memahami ketidakpastian dalam estimasi kinerja saat mengevaluasi pengklasifikasi pada kumpulan data empiris, aspek yang mungkin kurang diperhatikan dalam studi-studi sebelumnya.

## 2. METODOLOGI PENELITIAN

Pada penelitian ini terdapat beberapa langkah, antara lain: *data acquisition*, *data exploration*, *modelling*, dan *evaluation*. Dataset yang digunakan adalah dari Kaggle yaitu terdiri dari 768 data pasien dengan beberapa variabel atau atribut prediktor medis, dataset yang digunakan untuk klasifikasi diabetes terdiri dari dua kelas, yaitu diabetes (label 1) dan non-diabetes (label 0). Berdasarkan hasil analisis, terdapat ketidakseimbangan jumlah data antara kedua kelas tersebut, di mana kelas non-diabetes memiliki 500 sampel, sementara kelas diabetes hanya memiliki 268 sampel. Ketidakseimbangan ini menunjukkan bahwa proporsi data pada kelas non-diabetes hampir dua kali lipat lebih banyak dibandingkan dengan kelas diabetes. Adapun Langkah-langkah penelitian seperti ditunjukkan pada Gambar 2.



Gambar 2. Langkah – Langkah Penelitian

**2.1. Data acquisition**

Data yang digunakan pada penelitian ini berupa dataset yang berasal dari kaggle dengan format .csv [10]. Karakteristik atribut atau variabel pada dataset ditunjukkan pada Tabel 2 dan Tabel 3. Tahapan penelitian ini merupakan prosedur dalam pencapaian melalui empat tahapan yaitu pengumpulan data, perancangan system, pengujian akurasi dan prediksi dengan metode *K-Neares Neighbor Classifier*, *Naive Bayes*, Regresi Logistik, SVM dan *Neural Network*.

Tabel 2. Parameter prediksi

Parameter	Satuan
<i>(Pregnancies/Kehamilan)</i>	Total Kehamilan
<i>Glucose/Glukosa</i>	Mg/dl
<i>Blood Pressure/Tekanan Darah</i>	Mmhg
<i>Skin Thickness/Ketebalan Kulit</i>	Mm
Insulin	Mu U/ml
BMI/Indeks Masa Tubuh	Kg/m2

Tabel 3. Data

No.	Pregnancies	Glucose	Blood Pressure	Skin Thickness	Insulin	BMI	Diabetes Pedigree Function	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.228	33	1

Tabel 2 menunjukkan parameter-parameter medis yang sering digunakan dalam screening atau pemeriksaan untuk mendeteksi diabetes mellitus yang mengukur faktor risiko penyakit diabetes. Tabel 3 memberikan informasi mengenai beberapa pasien yang didiagnosis dengan diabetes, serta beberapa fitur yang berkaitan dengan kondisi kesehatan mereka. Tabel ini mencakup jumlah kehamilan (*Pregnancies*), kadar glukosa (*Glucose*), tekanan darah (*Blood Pressure*), ketebalan kulit (*Skin Thickness*), kadar insulin (*Insulin*), indeks massa tubuh (BMI), fungsi keturunan diabetes (*Diabetes Pedigree Function*), usia (*Age*), dan hasil diagnosis (*Outcome*). Misalnya, baris pertama menunjukkan seorang pasien dengan 6 kali kehamilan, kadar glukosa 148 mg/dL, tekanan darah 72 mm Hg, ketebalan kulit 35 mm, kadar insulin 0, BMI 33.6, fungsi keturunan diabetes 0.627, dan berusia 50 tahun dengan hasil diagnosis positif diabetes (*Outcome 1*).

## 2.2. Data Exploration

Setelah proses akuisisi data, langkah berikutnya adalah melakukan eksplorasi data secara cermat. Pada tahap eksplorasi ini, dataset yang diunduh dari situs Kaggle diperlakukan dengan memprioritaskan penghapusan data duplikat dan penanganan nilai yang hilang (*missing values*), yang krusial untuk memastikan integritas data sebelum lanjut ke analisis lebih lanjut. Selain itu, untuk mempertahankan distribusi kelas yang seimbang dalam dataset, digunakan teknik stratified sampling saat membagi dataset menjadi set pelatihan, validasi, dan pengujian dengan rasio 60:20:20. Teknik ini memastikan bahwa proporsi setiap kelas dijaga dengan baik di setiap subset, menghindari potensi bias dalam pembagian dataset yang dapat mempengaruhi kualitas dan hasil akhir dari model pembelajaran mesin yang akan dikembangkan.

## 2.2. Data Clean (Pembersihan Data)

Data yang telah dieksplorasi kemudian dibersihkan dari hal-hal yang dapat mengganggu analisis, seperti data yang tidak relevan, duplikasi, atau outlier yang tidak diinginkan. Tahap ini memastikan data yang digunakan dalam analisis benar-benar berkualitas.

## 2.3. Pembagian data

Data yang telah bersih kemudian dibagi menjadi tiga bagian, yaitu data pelatihan (60%), data pengujian (20%), dan data validasi (20%), yang digunakan untuk melatih, menguji, dan memvalidasi model agar terhindar dari overfitting.

## 2.3. Modelling

Pemodelan merupakan salah satu langkah dalam membuat model dari sistem yang akan diklasifikasikan. Algoritma *K-Neares NeighborClassifier*, *Naive Bayes*, Regresi Logistik, SVM dan *Neural Network* digunakan dalam penelitian ini.

### a. *K-Nearest Neighbor*

*K-Nearest Neighbor* (KNN) merupakan algoritma yang banyak digunakan di berbagai bidang seperti analisis data, prediksi penyakit dan prediksi grafik. Ini melibatkan penentuan titik *K* dalam kumpulan data yang paling dekat dengan titik data tertentu berdasarkan ukuran kesamaan. Metode ini dapat ditingkatkan dengan teknik seperti teknik hashing sensitif lokal *multi-layer* untuk meningkatkan efisiensi dan akurasi pencarian [8]. Formula *K-Nearest Neighbor* ditunjukkan pada persamaan (1).

$$dis(x_1, x_2) = \sqrt{\sum_{i=0}^n (x_{1i} - x_{2i})^2} \quad (1)$$

### b. *Naive Bayes*

*Naive Bayes* merupakan model klasifikasi dokumen yang populer karena kesederhanaan dan efisiensinya [11]. Meskipun secara tradisional digunakan untuk pengelompokan dan klasifikasi, kemajuan terkini telah menunjukkan potensinya untuk pembelajaran probabilistik umum dan inferensi, memberikan akurasi dan waktu pembelajaran yang sebanding dengan jaringan *Bayesian* yang tidak bergantung pada konteks, namun kecepatan inferensi jauh lebih cepat [12]. Formula *Naive Bayes* ditunjukkan pada persamaan (2).

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) \cdot P(A)}{P(B)} \quad (2)$$

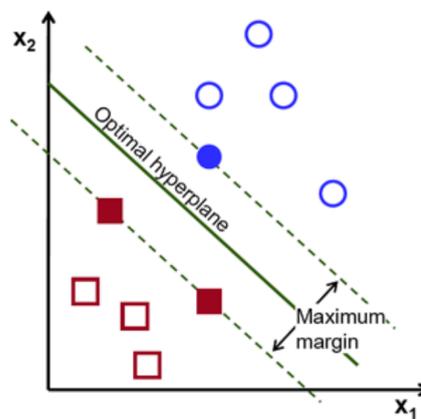
**c. Regresi Logistik**

Regresi logistik merupakan metode statistik ampuh yang banyak digunakan dalam penelitian medis untuk menganalisis pengaruh variabel independen terhadap hasil biner [13]. Hal ini memungkinkan untuk mengukur pengaruh unik setiap variabel dengan mengidentifikasi kombinasi linier terkuat dari faktor-faktor yang terkait dengan hasil yang diamati. Metode ini serbaguna dan mempertimbangkan variabel kontinu dan kategorikal [14]. Formula *Regresi Logistik* ditunjukkan pada persamaan (3).

$$f(x) = \frac{1}{1+e^{-x}} \quad (3)$$

**d. Support Vector Machine (SVM)**

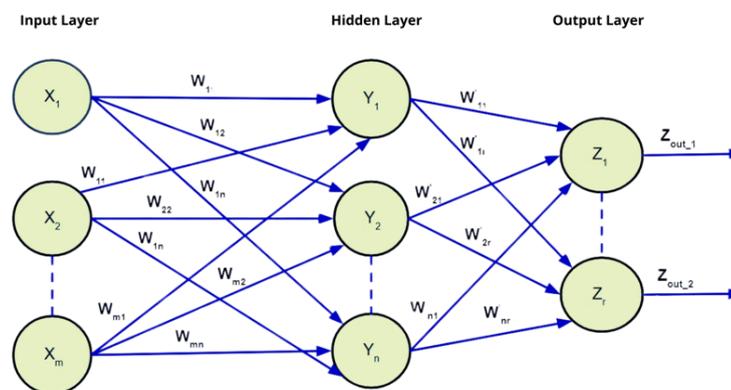
*Support Vector Machines* (SVM) merupakan algoritma yang kuat dan akurat dalam penambangan data, dengan aplikasi dalam klasifikasi dan regresi [15]. SVM menemukan *hyperplane* yang memisahkan titik data yang diberi label secara optimal ke dalam kelas, mendukung vektor pada margin, dan dapat digunakan untuk pemerataan nonlinier dalam sistem komunikasi yang dalam dari beberapa kasus mengungguli jaringan saraf [16] seperti yang ditunjukkan pada Gambar 3.



Gambar 3. *Hyperplane Support Vector Machine* (SVM) [16]

**e. Neural Network**

*Neural Network* (NN) atau Jaringan saraf merupakan metode pembelajaran mesin dasar yang terinspirasi oleh sistem saraf biologis, terdiri dari neuron yang saling berhubungan yang memproses data untuk memecahkan berbagai masalah [17]. Sementara NN tradisional menggunakan fungsi aktivasi sederhana seperti fungsi sigmoid, kemajuan terkini mengusulkan penggunaan regresi proses *Gaussian aditif* untuk menghasilkan fungsi aktivasi individu yang optimal untuk setiap neuron, meningkatkan efisiensi komputasi dan ekspresi [18]. Proses *Learning Jaringan Neural* ditunjukkan pada Gambar 4.



Gambar 4. Proses *Learning Jaringan Neural*

## 2.4. Evaluation (Evaluasi)

Setelah model dilatih dan diuji, performanya dievaluasi menggunakan metrik seperti akurasi, presisi, recall, F1-score, dan lainnya. Tahap ini bertujuan untuk menentukan seberapa baik model dalam memprediksi atau mengklasifikasikan data. Evaluasi dilakukan dengan memilih salah satu metrik akurasi, presisi, recall atau f1-score berdasarkan penghitungan true positif (TP), true negative (TN), false positive (FP), dan false negative (FN) dalam matriks confusion.

### a. Confusion Matrix

*Confusion Matrix* adalah metode untuk mengevaluasi model pembelajaran mesin dengan membandingkan kelas yang diprediksi dengan kelas yang sebenarnya [19]. Namun, *Confusion Matrix* tradisional tidak sepenuhnya mendukung struktur data yang kompleks, seperti label hierarki dan multi-label, sehingga mendorong pengembangan pendekatan baru seperti matriks konfusi hierarki dan matriks noise shift [20]. Inovasi ini bertujuan untuk meningkatkan evaluasi model dengan mempertimbangkan kekhususan masalah klasifikasi hierarki dan menggabungkan pengaruh karakteristik instans pada kinerja anotator. Selain itu, metode ini menekankan pentingnya memahami ketidakpastian dalam estimasi kinerja saat mengevaluasi pengklasifikasi pada kumpulan data empiris, menunjukkan bagaimana ketidakpastian ini dapat menyembunyikan perbedaan kinerja pengklasifikasi [21].

*Confusion Matrix* digunakan untuk menghitung berbagai metrik kinerja dalam evaluasi model klasifikasi. Metrik-metrik tersebut meliputi Akurasi (*Accuracy*), Presisi (*Precision*), Recall (*Sensitivitas*), dan *F1-Score*. Akurasi mengukur seberapa tepat model dalam mengklasifikasikan data secara keseluruhan, dinyatakan sebagai rasio prediksi benar terhadap total data. Presisi mengukur proporsi data yang diprediksi positif yang sebenarnya positif dari semua data yang diprediksi positif. *Recall* mengukur proporsi data yang diprediksi positif yang sebenarnya positif dari semua data yang sebenarnya positif. *F1-Score* adalah rata-rata harmonik dari presisi dan recall, memberikan keseimbangan antara kedua metrik tersebut. Perhitungan Akurasi (*Accuracy*), Presisi (*Precision*), Recall (*Sensitivitas*), dan *F1-Score* secara berturut-turut ditunjukkan pada persamaan (4) sampai persamaan (7).

### b. Accuracy (Akurasi)

Akurasi mengukur seberapa sering model membuat prediksi yang benar.

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

### c. Precision (Presisi)

Presisi menunjukkan proporsi prediksi positif yang benar.

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

Presisi tinggi berarti bahwa ketika model memprediksi positif, kemungkinan besar prediksi tersebut benar.

### d. Recall (Sensitivitas atau TPR - True Positive Rate)

Recall mengukur seberapa baik model dalam mendeteksi semua data positif.

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

Recall tinggi menunjukkan bahwa model berhasil mengidentifikasi sebagian besar data positif.

### e. F1-Score

*F1-Score* adalah rata-rata harmonis dari presisi dan recall, memberikan keseimbangan antara keduanya.

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

*F1-Score* digunakan ketika kita ingin mempertimbangkan keseimbangan antara presisi dan recall. Metrik-metrik ini membantu dalam mengevaluasi dan membandingkan kinerja berbagai model klasifikasi.

Di mana TP adalah *True Positive*, TN adalah *True Negative*, FP adalah *False Positive* dan FN adalah *False Negative*.

## 3. HASIL DAN PEMBAHASAN

Hasil tinjauan dan analisis karakteristik variabel pada Tabel 2 sehingga mendapatkan Tabel 3 menunjukkan lima data teratas untuk kumpulan data dengan delapan variabel independen dan satu variabel dependen. 8

variabel independen tersebut adalah *pregnancies*, *glucose*, *blood pressure*, *skin thickness*, *insulin*, BMI (*body mass index*), *diabetes pedigree function*, dan *age*. Sedangkan 1 variabel dependen adalah *outcome*. Proses pembangunan model dalam penelitian ini melibatkan beberapa tahapan, mulai dari eksplorasi data hingga evaluasi model. Untuk *K-Nearest Neighbor* (KNN), data yang diunduh dari Kaggle diolah melalui eksplorasi, pembagian set data secara *stratified*, dan pemilihan parameter "K" optimal, kemudian dievaluasi dengan metrik akurasi, presisi, *recall*, dan *F1-score*. *Naive Bayes* mengikuti proses yang serupa, dengan penekanan pada asumsi independensi kondisional dari fitur. Regresi Logistik menggunakan fungsi probabilitas logistik untuk klasifikasi biner, sementara *Support Vector Machine* (SVM) memilih kernel yang sesuai untuk memisahkan kelas dengan *hyperplane optimal*. Neural Network dibangun dengan konfigurasi lapisan dan neuron yang tepat. Model-model ini dievaluasi dengan metrik yang sama, dan hasilnya menunjukkan *Regresi Logistik* dan *Neural Network* memiliki akurasi tertinggi dalam memprediksi diabetes. Setiap variabel memiliki rentang nilai yang berbeda-beda. Model dievaluasi menggunakan akurasi, presisi, *recall*, dan *F1-Score* untuk memberikan gambaran yang komprehensif tentang kinerja model. Dalam penelitian ini, menggunakan kernel linear untuk algoritma *Support Vector Machine* (SVM). Kernel linear merupakan pilihan yang sederhana dan efisien, yang memungkinkan SVM untuk menemukan hyperplane terbaik untuk memisahkan kelas-kelas dalam dataset. Kernel ini cocok untuk memisahkan data dengan hyperplane linear dan biasanya menjadi pilihan pertama sebelum mengeksplorasi kernel yang lebih kompleks.

Kernel linear pada SVM memetakan data ke dalam ruang fitur yang sama dengan dimensi aslinya, sehingga memungkinkan untuk menemukan solusi yang optimal tanpa menambah kompleksitas komputasional yang berlebihan. Ini sangat berguna terutama ketika kita berhadapan dengan dataset yang tidak terlalu besar dan kompleks, seperti yang digunakan dalam penelitian ini.

Dari hasil eksperimen dan analisis, terlihat bahwa model SVM dengan kernel linear memberikan akurasi yang kompetitif, walaupun tidak sebaik Regresi Logistik dan Neural Network dalam konteks dataset tertentu yang digunakan dalam penelitian ini. Hal ini menunjukkan bahwa kernel linear dapat menjadi pilihan yang tepat untuk beberapa kasus, meskipun tidak selalu memberikan hasil terbaik dibandingkan dengan kernel lain. Modeling dilakukan pada data *Train Set* (Pelatihan), *Validation Set* (Validasi), dan *Test Set* (Pengujian) dengan persentasi masing-masing data adalah 60% *train set*, 20% *Validation set*, dan 20% *Test set* dari keseluruhan data berjumlah 768, jumlah data training sebanyak 460, *Validation set* yang digunakan untuk *modelling* sebanyak 153 data dan *test set* yang digunakan 155 data. model ini menggunakan semua variabel independen karena hampir semua variabel memiliki korelasi yang mendekati.

### 3.1. Pemodelan *Machine Learnings*

Setelah tahap membangun permodelan menggunakan metode *K-Neares Neighbor Classifier*, *Naive Bayes*, Regresi Logistik, SVM dan *Neural Network* di lakukan pengukuran performa hasil *confusion matrix* yaitu berupa nilai rata-rata akurasi, presisi, *recall*, dan *F1-score* ditunjukkan pada Tabel 4.

Tabel 4. Evaluasi

Metrik	KNN	Naive Bayes	Regresi Logistik	SVM
Precision (0)	0.78	0.81	0.83	0.79
Recall (0)	0.79	0.74	0.73	0.74
f1-score (0)	0.79	0.77	0.78	0.76
Precision (1)	0.60	0.59	0.59	0.57
Recall (1)	0.59	0.69	0.72	0.63
f1-score (1)	0.60	0.73	0.65	0.60
Accuracy	0.72	0.72	0.73	0.70

Hasil dalam tabel metrik evaluasi performa model (KNN, *Naive Bayes*, *Regresi Logistik*, dan SVM) menunjukkan kinerja masing-masing model dalam mengklasifikasikan dua kelas yang diamati yaitu kelas 0 dan kelas 1. Kelas 0 menunjukkan hasil prediksi negatif (tidak memiliki diabetes) dan kelas 1 menunjukkan hasil prediksi positif (mempunyai diabetes) dari dataset.

Model *Regresi Logistik* memiliki nilai presisi (*precision*) pada kelas 0 (tidak memiliki diabetes) tertinggi yaitu 0.83, yang menunjukkan bahwa dari semua prediksi yang model buat sebagai tidak memiliki diabetes, sebanyak 83% di antaranya benar-benar tidak memiliki diabetes. Sedangkan, model KNN memiliki *recall* (sensitivitas) yang lebih tinggi untuk kelas 0 yaitu 0.79, artinya model tersebut lebih baik dalam mengidentifikasi sebagian besar data yang sebenarnya tidak memiliki diabetes. *F1-score* untuk kelas 0

menunjukkan harmonisasi antara presisi dan *recall*, dengan nilai tertinggi dimiliki oleh model *Regresi Logistik* yaitu 0.78.

Model Naive Bayes memiliki nilai *recall* pada kelas 1 (mempunyai diabetes) tertinggi yaitu 0.69, yang menunjukkan kemampuannya dalam mengidentifikasi sebagian besar data yang sebenarnya memiliki diabetes. tetapi, model KNN memiliki presisi yang lebih tinggi untuk kelas 1 yaitu 0.60, menunjukkan bahwa dari semua prediksi yang model buat sebagai memiliki diabetes, sebanyak 60% di antaranya benar-benar memiliki diabetes. *F1-score* untuk kelas 1 menunjukkan konsistensi antara presisi dan *recall*, dengan model *Naive Bayes* memiliki nilai tertinggi yaitu 0.63.

Secara keseluruhan, kinerja model dapat dibandingkan berdasarkan metrik akurasi (*Accuracy*), di mana model *Regresi Logistik* memiliki akurasi yang relatif tinggi yaitu 0.73, menunjukkan kemampuan model ini dalam mengklasifikasikan dengan benar kedua kelas 0 dan kelas 1.4

### 3.2. Hyperparameter Tuning of The Neural Network Results

Penyetelan hiperparameter dalam jaringan saraf merupakan proses penting yang secara signifikan mempengaruhi kinerja model dan pemanfaatan sumber daya, optimalisasi hiperparameter sangat penting untuk mencapai perilaku generalisasi yang baik dalam jaringan saraf, menegakkan bias induktif yang tepat, regularisasi, dan peningkatan kinerja secara keseluruhan, terutama dengan data terbatas pengukuran performa hasil *hyperparameter* dari *Neural Network* seperti yang ditunjukkan pada Tabel 5.

Tabel 5. Penyetelan konfigurasi *Hyperparameter* pada data Diabetes

Nodes	Dropout	Learning Rate	Batch Size	Loss	Accuracy
16	0	0.01	32	0.44	0.79
			64	0.43	0.78
			128	0.42	0.79
		0.005	32	0.44	0.77
			64	0.44	0.77
			128	0.44	0.80
		0.001	32	0.44	0.80
			64	0.44	0.79
			128	0.45	0.79
32	0	0.01	32	0.47	0.75
			64	0.45	0.78
			128	0.44	0.77
		0.005	32	0.46	0.78
			64	0.43	0.79
			128	0.42	0.81
		0.001	32	0.45	0.78
			64	0.45	0.79
			128	0.44	0.77
64	0	0.01	32	0.47	0.77
			64	0.45	0.79
			128	0.45	0.77
		0.005	32	0.46	0.77
			64	0.46	0.77
			128	0.44	0.78
		0.001	32	0.48	0.76
			64	0.45	0.80
			128	0.45	0.78

Tabel 5 menunjukkan hasil penelitian pada model *machine learning* dengan berbagai kombinasi parameter. Tabel ini menguraikan hasil dari pengujian model dengan jumlah node yang berbeda (16, 32, dan 64), tanpa

menggunakan dropout, serta dengan variasi pada *learning rate* (0.01, 0.005, dan 0.001) dan *batch size* (32, 64, dan 128). Hasil dari eksperimen ini diukur dalam dua metrik utama: *Loss* dan *Accuracy*.

Jumlah node yang diuji adalah 16, 32, dan 64. Berdasarkan hasil pada table 4, kita dapat mengamati bahwa peningkatan jumlah node tidak selalu meningkatkan akurasi model secara signifikan. Misalnya, pada node 16, model mencapai akurasi tertinggi sebesar 0.80 dengan *learning rate* 0.001 dan *batch size* 32. Pada node 32, akurasi tertinggi juga mencapai 0.81 dengan *learning rate* 0.005 dan *batch size* 128. Sementara itu, pada node 64, akurasi tertinggi adalah 0.80 dengan *learning rate* 0.001 dan *batch size* 64. Hal ini menunjukkan bahwa jumlah node yang lebih banyak tidak selalu berkorelasi langsung dengan peningkatan performa model, dan ada kombinasi parameter lain yang mempengaruhi hasil.

*Learning rate* yang diuji adalah 0.01, 0.005, dan 0.001. Dari hasil eksperimen, *learning rate* yang lebih kecil (0.001 dan 0.005) cenderung memberikan hasil akurasi yang lebih baik dibandingkan dengan *learning rate* yang lebih besar (0.01). Contohnya, pada node 16, *learning rate* 0.001 memberikan akurasi 0.80 untuk *batch size* 32, sedangkan *learning rate* 0.01 hanya memberikan akurasi 0.79 untuk *batch size* yang sama. Ini menunjukkan bahwa *learning rate* yang lebih kecil membantu model untuk belajar dengan lebih baik dan mengurangi *loss* dengan lebih efektif.

*Batch size* yang diuji adalah 32, 64, dan 128. Hasilnya menunjukkan bahwa *batch size* yang lebih besar (128) cenderung memberikan akurasi yang lebih baik dalam beberapa kasus. Misalnya, pada node 32 dengan *learning rate* 0.005, *batch size* 128 memberikan akurasi tertinggi sebesar 0.81. Namun, ada juga kasus di mana *batch size* yang lebih kecil (32) memberikan hasil yang baik, seperti pada node 16 dengan *learning rate* 0.001 yang mencapai akurasi 0.80. Ini menunjukkan bahwa ada *trade-off* antara *batch size* dan efisiensi pembelajaran model, di mana *batch size* yang lebih besar membantu dalam stabilitas pembelajaran, tetapi *batch size* yang lebih kecil juga bisa efektif tergantung pada kombinasi parameter lainnya.

Peneliti dapat menyimpulkan dari Tabel 5 bahwa performa model dipengaruhi oleh kombinasi dari jumlah *node*, *learning rate*, dan *batch size*. Hasil eksperimen model *machine learning* menunjukkan variasi performa berdasarkan jumlah *node* (16, 32, 64), *learning rate* (0.01, 0.005, 0.001), dan *batch size* (32, 64, 128) tanpa dropout. Secara umum, *learning rate* yang lebih kecil (0.001 dan 0.005) cenderung memberikan akurasi yang lebih baik, dengan akurasi tertinggi mencapai 0.81 pada *node* 32, *learning rate* 0.005, dan *batch size* 128. Peningkatan jumlah *node* tidak selalu meningkatkan akurasi secara signifikan, dan *batch size* yang lebih besar (128) seringkali memberikan hasil lebih baik.

Hasil penelitian ini memiliki kontribusi secara praktis dan teoritis dalam bidang prediksi diabetes dengan pendekatan *machine learning*. Secara praktis, penelitian ini merekomendasikan penggunaan dan evaluasi berbagai algoritma klasifikasi, yaitu *K-Nearest Neighbor*, *Naïve Bayes*, *Regresi Logistik*, SVM, dan *Neural Network*, untuk meningkatkan akurasi diagnosis diabetes serta mendukung deteksi dini dan strategi pengobatan yang dipersonalisasi. Secara teoritis, penelitian ini memperluas pemahaman tentang perbandingan kinerja antar algoritma dalam konteks prediksi diabetes, menunjukkan bahwa tidak ada satu algoritma yang secara universal unggul. Pemilihan model terbaik bergantung pada karakteristik dataset dan tujuan analisis.

#### 4. KESIMPULAN

Dalam penelitian ini, peneliti mengevaluasi kinerja beberapa model *machine learning* untuk memprediksi diabetes berdasarkan dataset yang telah dibagi menjadi *training set*, *validation set*, dan *test set*. Dataset terdiri dari total 768 data, dengan 460 data digunakan untuk pelatihan, 153 data untuk validasi, dan 155 data untuk pengujian. Lima model yang diuji meliputi *K-Nearest Neighbor* (KNN), *Naive Bayes*, *Regresi Logistik*, *Support Vector Machine* (SVM), dan *Neural Network*. Kinerja model dievaluasi berdasarkan metrik *precision*, *recall*, *F1-score*, dan akurasi. Hasil penelitian menunjukkan bahwa *Regresi Logistik* dan *Neural Network* memiliki akurasi tertinggi yaitu sebesar 80%. Hal ini menjadikan bahwa *Regresi Logistik* dan *Neural Network* adalah pilihan terbaik dalam hal akurasi keseluruhan. *Regresi Logistik* unggul dalam *precision* untuk kelas 0 (non-diabetes) dengan nilai 0.83 dan juga memiliki *recall* yang baik untuk kelas 1 (diabetes) sebesar 0.72, yang menunjukkan bahwa model ini mampu mengidentifikasi kasus diabetes dengan baik. *Neural Network* menunjukkan performa yang seimbang dengan *precision* dan *recall* yang cukup baik untuk kedua kelas, menghasilkan *F1-score* yang konsisten dan akurasi keseluruhan yang tinggi. *Naive Bayes* juga menunjukkan performa yang kompetitif dengan akurasi sebesar 72%. Model ini memiliki *recall* yang tinggi untuk kelas 1 sebesar 0.69, yang penting untuk mengurangi *false negatives* (FN) dalam deteksi diabetes. KNN dan SVM menunjukkan performa yang lebih rendah dibandingkan dengan model lainnya. KNN memiliki akurasi sebesar 72% tetapi menunjukkan *precision* dan *recall* yang lebih rendah untuk kelas 1. SVM memiliki akurasi terendah sebesar 70% dan performa yang kurang baik untuk kelas 1. Secara keseluruhan, *Regresi Logistik* dan *Neural Network* merupakan model yang paling menjanjikan untuk prediksi diabetes dalam dataset ini. *Regresi Logistik* sangat baik dalam mengidentifikasi non-diabetes (kelas 0) dengan *precision* yang tinggi, sementara *Neural*

*Network* menunjukkan keseimbangan yang baik antara precision dan recall untuk kedua kelas. *Naive Bayes*, dengan *recall* yang tinggi untuk kelas 1, dapat menjadi pilihan yang baik dalam situasi di mana deteksi positif (diabetes) sangat kritikal. Pemilihan model akhir harus disesuaikan dengan kebutuhan spesifik aplikasi, mempertimbangkan *trade-offs* antara *precision*, *recall*, dan akurasi keseluruhan.

#### DAFTAR PUSTAKA

- [1] A. Oktaviana, D. P. Wijaya, A. Pramuntadi and D. Heksaputra, "Prediksi Penyakit Diabetes Melitus Tipe 2 Menggunakan Algoritma K-Nearest Neighbor (K-NN)," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, pp. 812-818, 3 Juli 2024.
- [2] A. Massaro, N. Magaletti, G. Cosoli, V. O. Giardinelli and A. Leogrande, "Methods and Metrics for the Prediction of Diabetes," *Preprints*, no. doi:10.20944/preprints202206.0202.v1, pp. 1-22, 14 Juni 2022.
- [3] N. A. ElSayed, G. Aleppo, V. R. Aroda, R. R. Bannuru, F. M. Brown, D. Bruemmer, B. S. Collins, J. L. Gaglia, M. E. Hilliard, D. Isaacs, E. L. Johnson, S. Kahan, K. Khunti, S. K. Lyons and L., "Classification and Diagnosis of Diabetes: Standards of Care in Diabetes—2023," *Diabetes Care*, vol. 46, pp. 19-40, Januari 2023.
- [4] R. Wagh, J. M. Wagh, S. Upadhyay, C. Prajapati and U. Parit, "COMPLICATIONS IN DIABETES : A REVIEW," *Indian Journal of Applied Research*, vol. 12, no. 10, pp. 19-31, Oktober 2022.
- [5] R. Pahlevi, "Jumlah Penderita Diabetes di Indonesia Diproyeksikan Capai 28,57 Juta pada 2045," 24 November 2021. [Online]. Available: <https://databoks.katadata.co.id/datapublish/2021/11/24/jumlah-penderita-diabetes-di-indonesia-diproyeksikan-capai-2857-juta-pada-2045>. [Accessed 16 Juni 2024].
- [6] J. Jasmani and T. Rihiantoro, "Edukasi dan Kadar Glukosa Darah pada Pasien Diabetes," *Jurnal Keperawatan*, vol. XII (1), no. ISSN 1907 - 0357, pp. 140-148, April 2016.
- [7] A. Panesar, *Machine Learning and AI for Healthcare - Big Data for Improved Health Outcomes*, Berkeley, California: Apress, 2021.
- [8] A. Asmarani, M. I. Permana, A. Putri, M. R. Wijaya, E. Rasywir, D. Meisak and Y. Pratama, "Implementasi Algoritma K-Nearst Neighbor Untuk Memprediksi Penyakit Diabetes," *Jurnal Informatika Dan Rekayasa Komputer (JAKAKOM)*, vol. 2 (2), pp. 231-239, September 2022.
- [9] H. O. P. A. S. Patel, "Diabetes Prediction Using Machine Learning," in *roceedings of Second International Conference on Computing, Communications, and Cyber-Security*, Singapura, 2021.
- [10] MrSimple, "Diabetes Prediction [Data set]," Kaggle, 2024. [Online]. Available: <https://www.kaggle.com/dsv/7655840>. [Accessed Juni 2024].
- [11] S. Vishwakarma and S. Ganguly, "Optimal partition of feature using Bayesian classifier," *arXiv*, vol. 1, pp. 1-7, 27 April 2023.
- [12] A. Atthohiroh, R. Ayu and S. Maharani, "Penerapan Metode Naive Bayes dalam Memprediksi Penyakit Jantung," *Jurnal Teknisi : Jurnal Teknologi Komputer dan Sistem Informasi*, vol. 3 (1), pp. 8-13, Februari 2023.
- [13] J. C. Stoltzfus, "Logistic Regression: A Brief Primer," *Academic Emergency Medicine*, vol. 18, no. 10; Doi: <https://doi.org/10.1111/j.1553-2712.2011.01185.x>, pp. 1099-1104, 13 Oktober 2011.
- [14] V. Bewick, L. Cheek and J. Ball, "Review Statistics review 14: Logistic regression," *Critical Care*, vol. 9, pp. 112-118, 13 Januari 2005.
- [15] J. M. Davis, "Support Vector Machines for Classification and Quantitative Analysis," *Microscopy and Microanalysis*, vol. 20, no. DOI: <https://doi.org/10.1017/S1431927614005005>, pp. 656 - 657, Agustus 2014.
- [16] L. Nguyen, "Tutorial on Support Vector Machine," *Applied and Computational Mathematics*, vol. 6, pp. 1-15, 17 Juni 2017.
- [17] S. Manzhos and M. Ihara, "Neural Network with Optimal Neuron Activation Functions Based on Additive Gaussian Process Regression," *The Journal of Physical Chemistry A*, vol. 127 (37), pp. 7823-7835, 12 September 2023.
- [18] M. Thorat, S. Pandit and S. Balote, "Artificial Neural Network: A brief study," *Asian Journal of Convergence in Technology*, vol. VIII (3), no. III, pp. 12-16, 2022.
- [19] K. Riehl, M. Neunteufel and M. Hemberg, "Hierarchical confusion matrix for classification performance evaluation," *arXiv*, vol. 1, pp. 1-8, 15 Juni 2023.
- [20] J. Li, H. Sun and J. Li, "Beyond confusion matrix: learning from multiple annotators with awareness of instance features," *Machine Learning*, vol. 112, pp. 1053-1075, 7 Juli 2021.

- [21] J. Görtler, F. Hohman, D. Moritz, K. Wongsuphasawat, D. Ren, R. Nair, M. Kirchner and K. Patel, "Neo: Generalizing Confusion Matrix Visualization to Hierarchical and Multi-Output Labels," in *CHI '22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 2022.
- [22] World Health Organization, "Cardiovascular diseases," World Health Organization, 22 Maret 2023. [Online]. Available: [https://www-who-int.translate.google/?\\_x\\_tr\\_sl=en&\\_x\\_tr\\_tl=id&\\_x\\_tr\\_hl=id&\\_x\\_tr\\_pto=sc&\\_x\\_tr\\_hist=tr ue#tab=tab\\_3](https://www-who-int.translate.google/?_x_tr_sl=en&_x_tr_tl=id&_x_tr_hl=id&_x_tr_pto=sc&_x_tr_hist=tr ue#tab=tab_3). [Accessed 11 Juni 2024].
- [23] L. Yahaya, N. D. Oye and J. E. Garba, "A Comprehensive Review on Heart Disease Prediction Using Data Mining and Machine Learning Techniques," *American Journal of Artificial Intelligence*, vol. 4 (1), no. doi: 10.11648/j.ajai.20200401.12, pp. 20-29, 23 April 2020.
- [24] Y. Amelia, "Perbandingan Metode Machine Learning Untuk Mendeteksi Penyakit Jantung," *Idealis: Indonesia Journal Information System*, vol. 6 (2), no. ISSN 2684-7280, pp. 220-225, Juli 2023.
- [25] M. A. A.-R. Asif, M. M. Nishat, F. Faisal, R. R. Dip, M. H. Udoy, M. F. Shikder and R. Ahsan, "Performance Evaluation and Comparative Analysis of Different Machine Learning Algorithms in Predicting Cardiovascular Disease," *Engineering Letters*, vol. 29, no. 2, pp. 1-11, Juni 2021.
- [26] D. Lowd and P. Domingos, "Naive Bayes Models for Probability Estimation," in *ICML '05: Proceedings of the 22nd international conference on Machine learning*, 2005.