

## BAB II TINJAUAN PUSTAKA

### 2.1 Studi Literatur

Pada sub bab ini menjelaskan secara rinci mengenai penelitian-penelitian terdahulu yang menjadi acuan bagi penelitian yang sedang dilaksanakan. Hasil temuan dari beberapa studi sebelumnya disajikan dalam berikut.

Tabel 2. 1. Penelitian Terdahulu

No	Judul	Penulis	Kontribusi	Hasil
1	The Efficacy of Machine-Learning-Supported Smart System for Heart Disease Prediction	Nurul Absar dkk [12]	Penelitian ini mengadopsi empat model machine learning (Random Forest, Decision Tree, AdaBoost, dan K-Nearest Neighbor) untuk mendeteksi penyakit jantung. Dikembangkan algoritma umum untuk menganalisis kekuatan faktor-faktor relevan yang memengaruhi prediksi penyakit jantung. Penulis juga membuat sistem cerdas berbasis cloud hosting Streamlit yang dapat menyederhanakan pekerjaan dokter dalam diagnosis.	Empat model machine learning mampu memprediksi penyakit jantung dengan akurasi tinggi: RF 99,03%, DT 96,10%, AB 100%, KNN 100% (CHSLB) dan 97,82% (Cleveland). Pre-processing data meningkatkan akurasi. Penelitian ini memperkuat dasar pengembangan sistem dengan penambahan data, k-fold cross-validation, dan evaluasi potensi overfitting.
2	Diabetes Classification Using Machine Learning Techniques	Methaporn Phongying dan Sasiprapa Hirrote [13]	Menggabungkan hyperparameter tuning dan interaction terms antara BMI dan riwayat keluarga diabetes dengan variabel lain. Pendekatan ini meningkatkan performa model secara	Semua metode klasifikasi mengalami peningkatan performa saat interaction terms digunakan. Random Forest mencatat akurasi 97,5%, precision 97,4%,

			signifikan dan dapat diadaptasi untuk penyakit kronis lain, membuka peluang pengembangan sistem screening otomatis berbasis data medis.	recall 96,6%, F1-score 97%. Variabel interaksi berbasis faktor risiko utama terbukti meningkatkan akurasi prediksi diabetes.
3	Prostate Cancer Diagnosis via Visual Representation of Tabular Data and Deep Transfer Learning	Moumen El-Melegy dkk [14]	Mengembangkan metode diagnosis kanker prostat non-invasif berbasis data klinis sederhana dan kuesioner gejala pasien. Memperkenalkan Tab2Visual, transformasi data tabular menjadi gambar untuk memanfaatkan transfer learning pada CNN. Menggabungkan model terbaik ke dalam stacking classifier serta menyediakan aplikasi web publik berbasis Streamlit.	Model TabPFN, LightGBM, dan XGBoost (F1-score 0.849, AUC 0.87) ditingkatkan dengan Tab2Visual dan EfficientNet V2 (F1-score 0.857, AUC 0.899). Stacking classifier mencapai F1-score 0.907 dan AUC 0.911. PSA dan gejala IEBladder menjadi prediktor utama.
4	TabLLM: Few-shot Classification of Tabular Data with Large Language Models	Stefan Hegselman dkk [15]	Memperkenalkan TabLLM, kerangka kerja yang memanfaatkan LLM untuk klasifikasi data tabular dalam skenario zero-shot dan few-shot. Dikembangkan sembilan teknik serialization untuk mengubah tabel menjadi bahasa alami. Evaluasi dilakukan pada 10 dataset publik dan 3 tugas prediksi risiko kesehatan nyata.	TabLLM unggul pada zero-shot dan very-few-shot (hingga 256 contoh pelatihan), dengan selisih AUC rata-rata >5 poin dibanding metode lain. Text Template menjadi metode serialization terbaik. Performa tetap unggul pada dataset kesehatan dengan data terbatas, namun

				metode klasik dapat menyaingi saat data besar karena batas panjang input LLM.
--	--	--	--	---

## 2.2 Gagal Jantung

Gagal jantung merupakan suatu sindrom klinis yang ditandai dengan ketidakmampuan jantung memompa darah secara optimal guna memenuhi kebutuhan metabolic jaringan, atau hanya mampu melakukannya dengan peningkatan tekanan pengisian. Kondisi ini dapat dipicu oleh gangguan struktural maupun fungsional pada jantung, seperti kardiomiopati, penyakit jantung koroner, hipertensi kronis, atau kelainan katup. Manifestasi klinis utama mencakup sesak napas, kelelahan, dan edema perifer. Berdasarkan fraksi ejeksinya, gagal jantung diklasifikasikan menjadi gagal jantung dengan fraksi ejeksi menurun (HF<sub>r</sub>EF) dan gagal jantung dengan fraksi ejeksi terpelihara (HF<sub>p</sub>EF), yang masing-masing memiliki perbedaan dalam mekanisme patofisiologi serta pendekatan terapinya [12].

## 2.3 Diabetes

Diabetes Adalah kelompok gangguan metabolic yang ditandai oleh hiperglikemia kronis akibat kelainan pada sekresi insulin, aksi insulin, atau kombinasi keduanya. Keadaan ini mengganggu metabolisme karbohidrat, lemak, protein, serta berpotensi menimbulkan komplikasi jangka panjang pada organ vital, termasuk ginjal, mata, saraf, dan sistem pembuluh darah. Faktor risiko yang berperan meliputi predisposisi genetic, pola makan yang kurang sehat, obesitas, serta gaya hidup sedentary. Berdasarkan klasifikasinya, diabetes dibedakan menjadi tipe 1, tipe 2, dan diabetes gestasional, dengan tipe 2 sebagai bentuk yang paling dominan secara global, seiring meningkatnya prevalensi obesitas dan rendahnya aktivitas fisik [13].

## 2.4 Kanker

Kanker merupakan sekelompok penyakit yang ditandai oleh pertumbuhan serta pembelahan sel yang tidak terkendali, disertai kemampuan sel untuk menginvasi jaringan di sekitarnya atau bermetastasis ke organ lain melalui sistem sirkulasi darah dan limfatik. Kondisi ini umumnya terjadi akibat akumulasi mutasi genetic pada gen yang mengatur pertumbuhan sel, seperti proto-onkogen dan gen penekan tumor. Kanker dapat berkembang pada berbagai organ dengan perbedaan etiologi, prognosis, serta respons terhadap terapi. Faktor risiko yang berkontribusi meliputi paparan karsinogen, predisposisi genetic, infeksi virus tertentu dan gaya hidup tidak sehat, termasuk kebiasaan merokok serta konsumsi makanan tinggi lemak [14].

## 2.5 Data Tabular

Data tabular merupakan bentuk data yang tersusun dalam format tabel, terdiri atas baris (rows) dan kolom (columns), dimana setiap baris merepresentasikan suatu entitas atau catatan, sedangkan setiap kolom menggambarkan atribut atau variabel dari entitas tersebut. Format ini banyak digunakan karena mudah dibaca, diolah, dan dianalisis, baik secara manual maupun dengan bantuan perangkat lunak analisis data. Isi data tabular dapat berupa nilai numerik, teks, maupun kategori, dan sering dimanfaatkan di berbagai bidang seperti kesehatan, bisnis, serta penelitian ilmiah. Dalam ranah *machine learning*, data tabular menjadi salah satu format utama untuk pelatihan model, dimana setiap variabel berperan sebagai fitur yang membantu model dalam mengenali pola dan menghasilkan prediksi [16].

## 2.6 Machine Learning

*Machine Learning* (ML) adalah salah satu cabang dari kecerdasan buatan yang berfokus pada pengembangan algoritma dan model yang memungkinkan sistem untuk belajar dari data serta meningkatkan kinerjanya tanpa memerlukan pemrograman secara eksplisit. Pendekatan ini memanfaatkan metode statistik, optimasi, dan analisis pola untuk mengidentifikasi hubungan antar variabel, serta melakukan prediksi atau klasifikasi. Pada penelitian ini, *machine learning* (ML)

diterapkan untuk menganalisis data kesehatan dengan tujuan mengenali pola klinis dan memprediksi risiko penyakit secara lebih tepat, sehingga dapat mendukung proses pengambilan keputusan medis berbasis data [14].

### **2.6.1 Logistic Regression**

Logistic Regression adalah sebuah algoritma yang digunakan untuk menganalisis data yang bersifat kategoris, biasanya datanya bersifat biner seperti 0 atau 1, ya atau tidak, benar atau salah [20]. Dalam machine learning, Logistic Regression merupakan model statistik berbasis probabilitas umum yang digunakan untuk memecahkan masalah klasifikasi dengan menentukan hubungan antara variabel bebas dan variabel terikat [19]. Logistic Regression banyak digunakan di berbagai bidang termasuk kedokteran, ilmu sosial, dan machine learning untuk tugas seperti memprediksi hasil penyakit, perilaku pelanggan, dan masalah klasifikasi biner [17].

### **2.6.2 Support Vector Machine (SVM)**

SVM atau Support Vector Machine merupakan salah satu sistem prediksi serta metode pembelajaran yang berada di dalam ruang hipotesis berdimensi tinggi untuk beroperasi dengan menggunakan fungsi-fungsi linier pada ruang fitur (feature space). Teknik ini merupakan salah satu di antara metode yang relatif masih baru dibandingkan dengan teknik-teknik lainnya. Untuk membedakan kedua kelompok tersebut, akan sangat membantu jika kita menemukan hyperlane terbaik. Dengan menggunakan penghitungan margin hyperlane, kami menentukan garis hyperlane yang paling efektif untuk membagi data menjadi dua kelompok [18].

### **2.6.3 Naive Bayes**

Naive Bayes adalah metode klasifikasi yang mendasarkan keputusan pada teorema Bayes, dengan anggapan bahwa semua variabel bersifat independen. Meskipun anggapan ini sering tidak sepenuhnya benar, algoritma ini tetap dapat memberikan hasil yang baik, cepat, dan efisien. Naive Bayes banyak digunakan pada data yang besar dan bervariasi, termasuk untuk analisis teks atau data medis [17] [18].

#### 2.6.4 XGBoost

XGBoost adalah versi canggih dari gradient boosting yang dirancang untuk menghasilkan prediksi seakurat mungkin dengan waktu komputasi yang efisien. Algoritma ini membangun model secara bertahap dari pohon-pohon keputusan kecil, lalu menggabungkannya menjadi model yang kuat. XGBoost juga dilengkapi fitur untuk mencegah overfitting, sehingga sering dipilih untuk kompetisi dan proyek machine learning skala besar [19].

#### 2.7 Serialisasi

Serialisasi data tabular adalah proses mengubah data yang tersimpan dalam bentuk tabular menjadi representasi teks yang menyerupai bahasa alami. Proses ini dilakukan dengan cara menggabungkan setiap atribut beserta nilainya ke dalam kalimat deskriptif sehingga data yang semula bersifat numerik maupun kategorikal dapat dipahami dalam bentuk narasi. Sebagai contoh, data tabular yang memuat informasi seperti usia, kebiasaan tidur, atau kondisi kesehatan seseorang dapat dituliskan ulang menjadi kalimat, misalnya “Pasien berusia 45 tahun, tidur 6 jam per hari, dan memiliki tekanan darah tinggi.” Dengan adanya serialisasi, data tabular yang awalnya kaku dan terstruktur dapat diproses sebagai teks, sehingga lebih mudah diinterpretasikan dan digunakan dalam berbagai tugas analisis berbasis teks. [15].

#### 2.8 Teks Reprsntasi

Representasi teks merupakan proses mengubah informasi berbasis bahasa alami menjadi bentuk numerik yang terstruktur sehingga dapat diproses oleh *mechine learning* model maupun *Large Language model*. Transformasi ini tidak hanya mengkonversi kata menjadi angka, tetapi juga mempertahankan makna semantik, konteks, dan hubungan antar kata melalui teknik seperti *embedding* berbasis *transformer*, yang mampu menangkap nuansa makna secara dinamis. Dengan pendekatan ini, data teks dapat diintegrasikan secara efektif dengan data tabular, memungkinkan analisis multimodal yang lebih akurat, terutama dalam domain medis di mana deskripsi klinis dan data numerik saling melengkapi dalam mendukung pengambilan keputusan [20].

### **2.8.1. Term Frequency–Inverse Document Frequency (TF IDF)**

Metode Term Frequency–Inverse Document Frequency (TF-IDF) digunakan untuk memberi bobot pada kata dalam dokumen berdasarkan tingkat kepentingannya dibandingkan seluruh kumpulan dokumen (corpus). Konsep ini menggabungkan dua komponen: TF (Term Frequency), yang menghitung frekuensi kemunculan kata dalam sebuah dokumen, dan IDF (Inverse Document Frequency), yang menurunkan bobot kata-kata yang sering muncul di banyak dokumen. Pendekatan ini membantu menonjolkan kata yang relevan dan mengurangi pengaruh kata umum yang tidak informatif [20].

### **2.8.2. Bag of Word**

Bag of Words (BoW) merupakan teknik representasi teks yang mengubah dokumen menjadi sekumpulan kata unik beserta jumlah kemunculannya, tanpa mempertimbangkan urutan atau struktur kalimat. Dalam metode ini, semua kata dari kumpulan dokumen dikumpulkan menjadi vocabulary, lalu setiap dokumen digambarkan sebagai vektor frekuensi kata. Pendekatan ini sering dimanfaatkan untuk tugas seperti klasifikasi teks dan analisis sentimen karena kesederhanaannya. Meski begitu, BoW cenderung mengabaikan konteks dan makna urutan kata, serta dapat menghasilkan dimensi yang besar ketika kosakatanya luas [20].

### **2.8.3. Word2Vec**

Word2Vec merupakan metode representasi kata ke dalam bentuk vektor numerik dengan tujuan menangkap hubungan semantik dan sintaktik antar kata. Pendekatan ini menggunakan jaringan saraf sederhana untuk mempelajari pola kemunculan kata dalam suatu korpus teks melalui dua arsitektur utama, yaitu Continuous Bag of Words (CBOW) yang memprediksi kata target berdasarkan konteks, dan Skip-Gram yang memprediksi konteks berdasarkan kata target. Hasil representasi yang dihasilkan memungkinkan kata dengan makna serupa memiliki posisi yang berdekatan dalam ruang vektor, sehingga lebih efektif dibandingkan metode berbasis frekuensi dalam menangkap makna kata [20].

#### 2.8.4. N-Grams

N-Grams merupakan model representasi teks yang memecah kalimat atau dokumen menjadi urutan kata atau karakter sebanyak  $n$  elemen secara berurutan. Pendekatan ini digunakan untuk menangkap konteks lokal dalam teks, di mana  $n$  menunjukkan panjang unit yang diambil, misalnya unigram ( $n=1$ ), bigram ( $n=2$ ), atau trigram ( $n=3$ ). Dengan mempertahankan urutan kata, N-Grams mampu merepresentasikan hubungan antar kata yang berdekatan, sehingga lebih kaya informasi dibandingkan metode yang mengabaikan urutan seperti Bag of Words. Namun, peningkatan nilai  $n$  dapat menyebabkan dimensi data menjadi sangat besar dan memerlukan sumber daya komputasi yang lebih tinggi [20].

### 2.9 Large Language Model

Large Language Model (LLM) merupakan model kecerdasan buatan berskala besar yang dirancang untuk memproses dan menghasilkan bahasa alami dengan tingkat kefasihan serta koherensi tinggi. Model ini dibangun dengan arsitektur *deep learning* berbasis transformer, yang memungkinkan analisis hubungan antar kata dalam konteks panjang dan kompleks. LLM dilatih menggunakan korpus teks berukuran masif yang mencakup berbagai domain, sehingga memiliki kemampuan memahami struktur bahasa, makna kata, hingga konteks semantik secara mendalam. Berkat keunggulan tersebut, LLM dapat diterapkan pada berbagai tugas pemrosesan bahasa alami seperti penerjemahan, peringkasan teks, klasifikasi, ekstraksi informasi, hingga sistem tanya jawab. Dalam penerapannya, terdapat sejumlah teknik yang digunakan untuk meningkatkan performa model, antara lain *few-shot learning* yang memungkinkan model belajar dari sejumlah kecil contoh, *zero-shot learning* yang memungkinkan adaptasi tanpa contoh pelatihan, *prompt engineering* untuk merancang instruksi input yang lebih efektif, serta *fine-tuning* guna menyesuaikan model pada domain atau tugas tertentu. Pemilihan teknik tersebut sangat dipengaruhi oleh ketersediaan data, tujuan penelitian, serta kompleksitas permasalahan yang dihadapi, sehingga LLM tidak hanya menjadi alat serbaguna dalam pemrosesan bahasa, tetapi juga

membuka peluang luas bagi pengembangan sistem cerdas di berbagai bidang [15] [21].

### **2.9.1 Prompt Engineering**

Prompt Engineering adalah proses merancang dan mengoptimalkan instruksi (prompt) yang diberikan kepada LLM agar model dapat menghasilkan keluaran yang relevan dan akurat. Teknik ini memerlukan pemahaman mendalam mengenai perilaku model, sehingga perumusan prompt dapat membantu memandu LLM dalam menyelesaikan tugas dengan benar. Dalam konteks klasifikasi data tabular medis, prompt yang efektif dapat meningkatkan interpretasi data dan memperbaiki akurasi hasil [15] [21].

### **2.9.2 Few-Shot**

Few-shot learning merujuk pada pendekatan pembelajaran mesin yang memungkinkan model mempelajari pola atau menyelesaikan tugas hanya dengan sejumlah kecil contoh data pelatihan. Dalam konteks Large Language Model (LLM), metode ini memanfaatkan kemampuan generalisasi model yang telah dilatih pada korpus teks berskala besar. Pemberian beberapa contoh (shots) yang relevan membantu LLM memahami instruksi dan menerapkan pengetahuan yang telah diperoleh sebelumnya untuk menghasilkan prediksi atau keluaran yang akurat. Pendekatan ini memiliki keunggulan dari segi efisiensi kebutuhan data, sehingga sesuai untuk domain dengan keterbatasan data berlabel seperti penelitian medis, bahasa daerah, maupun kasus dengan batasan privasi tinggi [15].

### **2.9.3 Zero-Shot**

Zero-shot learning mengacu pada pendekatan pembelajaran mesin yang memungkinkan model menyelesaikan tugas tanpa menerima contoh data pelatihan spesifik untuk tugas tersebut. Pada Large Language Model (LLM), teknik ini memanfaatkan pengetahuan luas yang diperoleh dari pelatihan pada korpus teks berskala besar. Model dapat langsung menafsirkan instruksi atau pertanyaan baru dengan mengandalkan pemahaman bahasa alami dan penalaran kontekstual. Pendekatan ini memberikan keuntungan signifikan pada situasi di mana data

berlabel tidak tersedia, misalnya untuk bahasa yang jarang digunakan, istilah teknis khusus, atau domain yang memiliki keterbatasan akses data [15].

#### **2.9.4 Fine Tuning**

Fine-tuning merupakan proses penyesuaian parameter model pembelajaran mesin yang telah dilatih sebelumnya (pre-trained model) agar selaras dengan kebutuhan atau domain tertentu. Pada Large Language Model (LLM), tahap ini dilakukan dengan memberikan data tambahan yang relevan sehingga model dapat mengadaptasi pengetahuan umum menjadi kemampuan yang lebih spesifik terhadap tugas yang diinginkan. Pendekatan ini sering digunakan untuk meningkatkan kinerja model dalam konteks khusus, seperti analisis data medis, klasifikasi dokumen hukum, atau pemrosesan bahasa dalam bidang industri tertentu [15] [21].

#### **2.9.5 Large Language Model Meta AI (LLaMA)**

LLaMA (Large Language Model Meta AI) merupakan salah satu pengembangan Large Language Model yang dirilis oleh Meta dengan tujuan menyediakan model berkapasitas tinggi namun efisien dalam penggunaan sumber daya komputasi. Desain arsitekturnya mengacu pada prinsip transformer-based architecture yang telah dioptimalkan agar mampu memberikan kinerja setara, bahkan pada beberapa skenario melebihi, model sekelasnya dengan jumlah parameter lebih besar. Keunggulan LLaMA terletak pada skalabilitas dan fleksibilitasnya, sehingga dapat diadaptasi untuk berbagai skema fine-tuning, baik few-shot, zero-shot, maupun instruksi khusus yang kompleks. Ketersediaannya secara terbuka bagi peneliti memperluas peluang eksplorasi dalam pengembangan sistem kecerdasan buatan, termasuk penerapan pada analisis data medis tabular yang memerlukan interpretabilitas dan presisi tinggi [21].

#### **2.9.6 Gemini**

Gemini adalah Large Language Model sumber terbuka yang dikembangkan oleh Google DeepMind dengan arsitektur transformer yang efisien dan fleksibel. Model ini tersedia dalam berbagai ukuran parameter untuk menyesuaikan kebutuhan komputasi, serta mendukung few-shot dan zero-shot learning sehingga

mampu memberikan prediksi yang akurat meskipun data pelatihan terbatas. Keunggulan tersebut membuat Gemini cocok digunakan dalam penelitian maupun implementasi industri, termasuk integrasi data teks dan tabular untuk analisis prediktif [21].

### 2.9.7 Qwen

Qwen adalah Large Language Model sumber terbuka yang dikembangkan oleh Alibaba Cloud dengan kemampuan pemrosesan bahasa alami multi-bahasa, termasuk Bahasa Indonesia dan bahasa-bahasa internasional lainnya. Menggunakan arsitektur transformer yang dioptimalkan untuk efisiensi dan skalabilitas, Qwen mendukung berbagai ukuran parameter serta mampu melakukan few-shot dan zero-shot learning untuk menghasilkan respons yang relevan dan kontekstual. Fleksibilitas ini menjadikannya efektif untuk berbagai skenario, mulai dari pemrosesan teks murni hingga integrasi dengan data tabular dalam sistem analisis prediktif [22].

### 2.10 Evaluasi Model

Tahap evaluasi merupakan aspek yang sangat penting untuk memastikan bahwa model yang dibangun tidak hanya akurat, tetapi juga adil serta bebas dari bias. Proses evaluasi ini berperan dalam memahami dampak model terhadap berbagai kelompok, sehingga prediksi yang dihasilkan bersifat inklusif dan mampu mendukung pengambilan keputusan yang lebih adil di dunia nyata. Dalam penelitian ini, evaluasi dilakukan dengan menggunakan classification report yang mencakup metrik accuracy, precision, recall, f1-score, serta AUC. Penggunaan classification report memberikan gambaran yang komprehensif mengenai tingkat ketepatan dan kelengkapan model dalam mengklasifikasikan data.

#### a) Accuracy

*Accuracy* digunakan untuk mengukur proporsi jumlah prediksi yang benar (baik positif maupun negatif) terhadap keseluruhan data yang diuji. Ini adalah metrik yang paling umum digunakan, karena mudah dipahami dan memberikan gambaran umum kinerja model. Berikut rumus untuk menghitung *accuracy*:

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

**b) Precision**

*Precision* digunakan untuk mengukur proporsi dari prediksi positif yang benar-benar merupakan kelas positif. Ini menunjukkan keandalan model dalam memberikan label positif. Berikut rumus untuk menghitung *precision* :

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

**c) Recall**

*Recall* mengukur seberapa baik model mendeteksi seluruh kasus positif yang benar-benar ada dalam data. Berikut rumus untuk menghitung *recall* :

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

**d) F1-Score**

*F1-Score* merupakan rata-rata harmonik dari *precision* dan *recall*, yang berguna ketika diperlukan keseimbangan antara keduanya, terutama saat data tidak seimbang. Berikut rumus untuk menghitung *F1-Score* :

$$\text{F1-Score} = 2 \times \frac{\text{precision} \times \text{recall}}{(\text{precision} + \text{recall})}$$