

BAB II

TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Pada penelitian sebelumnya memiliki tujuan agar dapat membandingkan hasil riset yang sudah ada dan mendapatkan inspirasi untuk penelitian yang selanjutnya. Kajian ini turut mendukung penelitian saat ini dalam mengidentifikasi posisinya dan menunjukkan keorisinalitasnya. Penelitian sebelumnya memiliki peran signifikan dalam membimbing dan memperkaya penelitian saat ini. Dengan mempelajari penelitian terdahulu, peneliti dapat memahami konteks sejarah perkembangan pengetahuan dalam bidang terkait, mengidentifikasi celah pengetahuan yang belum terjawab, serta membangun landasan teoritis yang kuat (Muannif, Suhar, Bahrul, Fauzi, 2021). Selain itu, penelitian terdahulu membantu peneliti untuk menghindari duplikasi penelitian yang sudah ada, mengidentifikasi metode penelitian yang relevan, dan dapat meningkatkan kualitas penelitian dengan mengambil dari pengalaman peneliti sebelumnya. Oleh karena itu, penelitian sebelumnya adalah tahap penting dalam mendukung pengembangan pengetahuan dan kontribusi yang lebih baik dalam bidangnya.

2.1.1 Perubahan layanan manual menjadi Chatbot NLP di PT.XYZ (Dhafa et al., 2024)

Pada penelitian ini bertujuan untuk meningkatkan efisiensi operasional dan kepuasan pelanggan melalui pengembangan chatbot berbasis AI dan NLP di PT. XYZ agar dapat menggantikan sistem layanan pelanggan manual. Penelitian ini mencakup analisis kebutuhan, pengumpulan data, pengembangan model NLP, serta implementasi chatbot. Pergeseran ke chatbot ini mengatasi kelemahan sistem manual, seperti respons lambat, biaya tinggi, inkonsistensi jawaban, dan keterbatasan jam operasional. Manfaat yang diharapkan meliputi peningkatan efektivitas layanan, kepuasan pelanggan, dan pengurangan biaya. Hasilnya, chatbot

mampu memberikan respons yang relevan, meski belum bisa mencakup setiap skenario percakapan. Tetapi penelitian ini tidak merinci model NLP yang digunakan. Selain itu, belum ada skenario fail-safe saat chatbot gagal menjawab, dan pengujian hanya sebatas *black-box*, tanpa metrik kuantitatif objektif.

2.1.2 Implementasi Chatbot untuk Pendaftaran Mahasiswa Baru Menggunakan *Recurrent Neural Network* (Zuraiyah et al., 2020)

Penelitian berikut ini bertujuan menciptakan sistem otomatis yang dapat mempermudah calon mahasiswa mengakses informasi pendaftaran yang lebih spesifik dari situs web universitas melalui pengembangan chatbot berbasis Model *Recurrent Neural Network* (RNN) digunakan untuk dapat melakukan klasifikasi teks. Hasil pengujian menunjukkan performa chatbot yang solid, mencapai nilai akurasi 88%, *presisi* 95%, dan *recall* 92%. Angka-angka ini mengindikasikan efektivitas sistem dalam menyajikan informasi yang relevan dan akurat, sehingga mempermudah akses bagi pengguna. Implementasi chatbot ini terbatas pada platform Telegram, belum mencakup platform lain, membatasi jangkauan penggunaannya.

2.1.3 Perancangan Chatbot *Booking Order* di *Carwash* dengan NLP (Samuel & Susanto, 2024)

Penelitian berfokus pada penggunaan algoritma *Naive Bayes* untuk mengklasifikasi pertanyaan pelanggan melalui perancangan chatbot yang memanfaatkan *Natural Language Processing* (NLP) dalam sistem pemesanan *carwash*. Tujuan utama riset ini adalah mengatasi masalah antrean dan fluktuasi harga yang kerap dihadapi bisnis *carwash*, dengan menyediakan layanan informasi real-time kepada pelanggan. Hasil penelitian menunjukkan bahwa chatbot yang dikembangkan mendapat respons positif dari pengguna. Tetapi penelitian ini tidak menjelaskan metrik performa chatbot secara detail, seperti akurasi, *precision*, atau *recall*. Selain itu, arsitektur NLP tidak diperinci, sehingga tidak diketahui apakah sistem ini berbasis aturan (rule-based) atau *machine learning*.

2.1.4 Penerapan Chatbot *Deep Learning* untuk Penerimaan Mahasiswa Baru di Universitas Indo Global Mandiri

(Desrika et al., 2023) melakukan penelitian mengenai pengembangan sistem chatbot yang memanfaatkan algoritma *Deep Learning*, khususnya model *Artificial Neural Network* (ANN). Tujuan utama penelitian ini adalah untuk mempermudah calon mahasiswa dalam mengakses informasi dan meningkatkan efisiensi layanan administrasi universitas. Manfaat yang diharapkan dari implementasi chatbot ini adalah peningkatan aksesibilitas informasi bagi calon mahasiswa. Hasil penelitian menunjukkan bahwa chatbot memiliki nilai akurasi 86% dalam menjawab pertanyaan pengguna. Meskipun demikian, ditemukan beberapa kelemahan dalam mengenali teks atau pertanyaan yang tidak sesuai dengan kaidah EYD.

2.1.5 Pemanfaatan *Deep Feed-Forward Neural Network* untuk Chatbot Informasi Akademik (Ruvita et al., 2023)

Penelitian berikut ini memiliki tujuan untuk dapat menyediakan informasi akademik yang mudah diakses bagi komunitas akademik, terutama dalam hal menghadapi tantangan kurangnya layanan informasi yang terpusat dan pembatasan akibat pandemi COVID-19. Hasil yang di dapat dari penelitian ini menggunakan dataset sebanyak 2059 entri dan mencapai nilai akurasi pelatihan sebesar 94%, dengan metrik evaluasi berupa *Precision* (0.88), *Recall* (0.89), dan *F1-Score* (0.88). Chatbot ini diimplementasikan sebagai aplikasi berbasis web menggunakan *framework* Flask.

2.1.6 Aplikasi Chatbot untuk Layanan Informasi Akademik menggunakan AIML (Guntoro et al., 2020)

Tujuan utama penelitian ini adalah menyediakan akses informasi yang cepat mengenai pendaftaran mahasiswa, program akademik, dan pertanyaan relevan lainnya melalui pengembangan aplikasi chatbot yang dirancang untuk meningkatkan kualitas layanan bagi mahasiswa dan masyarakat umum. Hasil pengujian fungsionalitas menunjukkan tingkat

keberhasilan 100%, sementara pengujian penerimaan pengguna (UAT) mencatat tingkat kepuasan 95%. Pengujian menunjukkan bahwa chatbot yang dibangun dengan *Artificial Intelligence Markup Language* (AIML) efektif dalam memberikan jawaban yang tepat dan cepat kepada pengguna serta menjadikannya alat yang bermanfaat, khususnya bagi calon mahasiswa. Tetapi, perlu diketahui bahwa AIML bersifat *rule-based*, yang berarti chatbot ini cenderung kurang fleksibel untuk menangani percakapan yang lebih kompleks atau di luar aturan yang telah ditentukan. Selain itu, sistem ini tidak menggunakan pembelajaran mesin, sehingga kurang adaptif terhadap pertanyaan-pertanyaan baru yang tidak ada dalam *knowledge base* yang sudah dibangun.

2.1.7 Penerapan Chatbot *Multilayer Perceptron* untuk Layanan Fakultas Teknologi Informasi Universitas Budi Luhur (Maulidia & Sayfrullah, 2024)

Tujuan dari penelitian ini adalah mengatasi lambatnya respons dosen pembimbing terhadap pertanyaan mahasiswa yang seringkali masih disampaikan secara manual melalui email atau tatap muka dengan mengembangkan chatbot menggunakan algoritma *Multilayer Perceptron* (MLP). Hasil penelitian menunjukkan bahwa chatbot yang diimplementasikan dengan MLP mampu mencapai akurasi 1.000 dan loss 0.0117. Angka-angka ini mengindikasikan bahwa model tersebut sangat efektif dalam menyediakan layanan informasi yang dibutuhkan. Namun, akurasi model yang terlalu ideal (1.000) menimbulkan pertanyaan dan membutuhkan verifikasi lebih lanjut untuk memastikan tidak terjadi overfitting.

2.1.8 Pemanfaatan NLP untuk Chatbot di SMK YPE Nusantara Slawi (Apriliani et al., 2023)

Pada penelitian berikut inibertujuan untuk melakukan optimalisasi penyampaian informasi sekolah secara interaktif dan efisien melalui teknologi *Natural Language Processing* (NLP). Model NLP dalam penelitian

ini dibangun menggunakan arsitektur neural network multilayer dan dioptimasi dengan algoritma *Stochastic Gradient Descent* (SGD). Evaluasi performa dilakukan menggunakan metrik akurasi dan loss. Hasilnya menunjukkan bahwa chatbot yang dikembangkan mampu merespons pertanyaan pengguna secara cepat dan relevan serta tahap preprocessing untuk representasi numerik menggunakan *Bag of Word*. Pembahasan ini terlalu fokus pada klaim kualitatif tanpa didukung oleh angka atau metrik performa hasil pengujian.

2.1.9 Implementasi Chatbot Informasi Al-Qur'an dengan TF-IDF dan Neural Network (Hasibuan et al., 2024)

Penelitian berikut ini berfokus pada kombinasi metode TF-IDF untuk pembobotan kata dan *Neural Network* sebagai model klasifikasi dalam pengembangan chatbot. Tujuan utamanya adalah untuk meningkatkan pengalaman pengguna dalam mengakses informasi spesifik Al-Qur'an, seperti ayat, tafsir, dan terjemahan, tanpa perlu pencarian manual. Chatbot ini diimplementasikan pada platform Telegram sebagai antarmuka pengguna. Pengujian yang dilakukan menunjukkan bahwa chatbot memberikan tanggapan yang tepat terhadap 38 dari 46 pertanyaan, menghasilkan akurasi 82,6% dan *loss* 0.0423. Penelitian ini menyimpulkan bahwa kombinasi TF-IDF dan *Neural Network* efektif dalam menyediakan layanan informasi keagamaan yang relevan dan responsif. Meski demikian, penelitian ini belum membahas adanya pembelajaran berkelanjutan atau pembaruan *knowledge base* secara otomatis, yang akan menjadi penting untuk menjaga relevansi dan cakupan informasi seiring waktu

Tabel 2.1 Review Literatur

No	Nama Penulis	Judul	Metode	Rangkuman
1.	(Dhafa, Ifal, Bayu, Wisnu, Aries, 2024)	Perubahan layanan manual menjadi	<i>Natural Language Processing</i> (NLP)	Chatbot mampu memberikan respon yang cepat dan akurat serta

		Chatbot NLP di PT.XYZ		mengurangi biaya operasional. Namun, chatbot tidak dapat sepenuhnya mencakup seluruh bagian percakapan
2.	(Tjut, Dian, Degi, 2020)	Implementasi Chatbot menggunakan <i>Recurrent Neural Network</i> untuk Pendaftaran Mahasiswa Baru	RNN	Chatbot yang dikembangkan mampu menjawab pertanyaan dengan akurasi 88%, presisi 95%, dan <i>recall</i> 92%, menunjukkan efektivitas dalam memberikan informasi yang relevan.
3.	(Samuel & Susanto, 2024)	Perancangan Chatbot <i>Booking Order</i> di <i>Carwash</i> dengan NLP	<i>Natural Language Processing</i> (NLP) dan algoritma <i>Naive Bayes</i>	Chatbot yang dikembangkan mendapat respons positif dari pengguna, menunjukkan bahwa sistem ini dapat membantu memudahkan proses booking order

4.	(Desrika, Zulkifli, Ricky, 2023)	Penerapan Chatbot <i>Deep Learning</i> untuk Penerimaan Mahasiswa Baru di Universitas Indo Global Mandiri	<i>Deep learning</i> dan arsitektur <i>Artificial Neural Network</i> (ANN).	Hasil uji coba menunjukkan bahwa chatbot yang telah dibuat memiliki akurasi sebesar 88%, presisi 95%, dan <i>recall</i> 92% dalam menjawab pertanyaan terkait pendaftaran mahasiswa baru
5.	(Ruvita, Dede, Ahmad, 2023)	Pemanfaatan <i>Deep Feed-Forward Neural Network</i> untuk Chatbot Informasi Akademik	<i>Deep FeedForward Neural Network</i>	Chatbot mencapai akurasi pelatihan sebesar 94%, dengan <i>Precision</i> 0.88, <i>Recall</i> 0.89, dan <i>F1Score</i> 0.88.
6.	(Guntoro, Loneli, Lisawita, 2020)	Aplikasi Chatbot Kampus untuk Layanan Informasi Akademik menggunakan AIML	AIML	Chatbot ini sanggup memberikan respons atas pertanyaan pengguna berdasarkan informasi yang telah diinputkan, dan dapat diakses

				secara daring melalui peramban web
7.	(Maulidia, Sayfrullah, 2024)	Penerapan Chatbot <i>Multilayer Perceptron</i> untuk Layanan Fakultas Teknologi Informasi Universitas Budi Luhur	<i>Multilayer Perceptron</i>	Chatbot mencapai akurasi 1.000 dan <i>loss</i> 0.0117, menunjukkan efektivitas dalam memberikan respons yang tepat terhadap pertanyaan mahasiswa.
8.	(Apriliani, Handayani, Saputra, 2023)	Pemanfaatan NLP untuk Chatbot di SMK YPE Nusantara Slawi	NLP & SGD	sistem chatbot berhasil menghasilkan akurasi model sebesar 98.48% dan <i>loss</i> training sebesar 0.0554, menunjukkan performa sistem yang tinggi dan responsif terhadap pertanyaan pengguna.
9.	(Hasibuan, Armansyah, Fauzan, 2024)	Implementasi Chatbot Informasi Al-Qur'an dengan TF-IDF dan	TF-IDF dan <i>Neural Network</i>	Hasil Penelitian menunjukkan hasil akurasi respon sebesar 82,6%, yang dapat

		<i>Neural Network</i>		memberikan jawaban spesifik seperti ayat, tafsir, dan terjemahan secara interaktif kepada pengguna.
--	--	-----------------------	--	-----------------------------------------------------------------------------------------------------

Research Gap:

Berdasarkan kajian literatur yang telah dilakukan, belum ditemukan studi yang secara khusus menggabungkan metode TF-IDF dengan model *Artificial Neural Network* (ANN) dan optimasi menggunakan algoritma *Stochastic Gradient Descent* (SGD) dalam pengembangan sistem chatbot berbasis NLP, sehingga hal ini dapat menjadi celah penelitian yang dapat dikaji lebih lanjut.

2.2 Teori Terkait

Teori merupakan kumpulan ide yang saling terkait yang digunakan untuk menjelaskan fenomena tertentu. Dalam penelitian, teori berfungsi sebagai peta jalan yang dapat memandu peneliti dalam memahami, menganalisis, dan memprediksi hasil dari sebuah penelitian. Teori terkait memungkinkan peneliti untuk merumuskan pertanyaan penelitian yang jelas, memilih metode penelitian yang tepat, dan menginterpretasikan data secara objektif. Selain itu, teori juga membantu peneliti untuk menghubungkan temuan penelitiannya dengan penelitian sebelumnya, sehingga dapat bermanfaat dalam memperkaya bidang keilmuan.

2.2.1 Chatbot

Program komputer yang berfungsi menyimulasikan percakapan antarmanusia disebut chatbot (Amalia, Wibowo, 2019). Arsitektur umum chatbot terdiri dari komponen-komponen seperti *Natural Language Understanding* (NLU) untuk dapat memahami input pengguna, dialog management yang memiliki fungsi untuk mengelola aliran percakapan, dan *Natural Language Generation* (NLG) untuk menghasilkan suatu respons yang relevan. Terdapat beberapa jenis chatbot, yaitu:

- a. *Rule-based*: Chatbot yang beroperasi berdasarkan aturan-aturan yang sudah ditentukan sebelumnya.
- b. *Retrieval-based*: Chatbot yang beroperasi dengan cara mencari jawaban yang paling relevan dari sumber data yang digunakan.
- c. *Generative chatbot*: Chatbot yang mampu menghasilkan teks baru berdasarkan model bahasa yang telah dilatih dengan data yang lebih kompleks.

Setiap jenis chatbot memiliki perbedaan kelebihan dan kekurangan masing-masing. Chatbot *Rule-based* sangat mudah dalam pengembangannya tetapi kurang fleksibel dalam menghadapi suatu pertanyaan yang tidak sesuai aturan yang sudah ditentukan.

2.2.2 Natural Language Processing (NLP)

NLP (Natural Language Processing) adalah bidang ilmu komputer yang meneliti bagaimana komputer dapat berinteraksi dengan bahasa manusia. Hal ini memungkinkan mesin untuk memahami dan menghasilkan bahasa seperti yang digunakan manusia. (Furqan, Sriani, Shidqi, 2023). Dalam pengembangan chatbot, NLP memiliki peran penting dalam memahami input teks dari pengguna, mengolahnnya, dan menghasilkan respons yang sesuai.

Proses-proses dasar dalam NLP meliputi tokenisasi yaitu pemecahan teks perkata, *stemming* yaitu pengurangan kata ke dalam bentuk dasarnya, *lemmatization* yaitu pengurangan kata ke dalam bentuk lemma atau akar katanya, dan *vectorization* yang berarti representasi kata dalam bentuk vektor numerik (Rosyadi, Amrullah, Marcus, Affandi, 2020).

Namun, NLP juga memiliki tantangan seperti contoh dalam hal ambiguitas Bahasa dan kompleksitas semantik. Ambiguitas dapat muncul dalam bentuk homonim yang berarti kata yang sama tetapi memiliki makna yang berbeda atau sinonim yang berarti persamaan kata. Kompleksitas

semantik berkaitan dengan pemahaman makna suatu kata yang mendalam dalam konteks bahasan tertentu.

2.2.3 **Artificial Neural Network (ANN)**

ANN, atau *Artificial Neural Network*, adalah sebuah model komputasi yang struktur dan fungsinya mencontoh otak manusia. Model ini tersusun atas banyak neuron yang terhubung dan berkoordinasi untuk mencapai tujuan tertentu. (Kukreja et al, 2016). ANN adalah jenis model yang paling sederhana namun cukup efektif dalam menyelesaikan masalah klasifikasi. ANN terdiri dari beberapa lapisan neuron, termasuk lapisan *input*, lapisan tersembunyi, dan lapisan *output*.

2.2.4 **Stochastic Gradient Descent (SGD)**

Stochastic Gradient Descent (SGD) adalah algoritma optimasi yang digunakan untuk melatih model pembelajaran mesin, termasuk ANN. SGD memperbarui parameter model secara bertahap berdasarkan gradien dari fungsi *loss* pada setiap sampel data (Admojo, Sulistya, 2022). Algoritma ini bekerja dengan menghitung gradien (turunan pertama) dari fungsi *loss* terhadap setiap parameter model, seperti bobot dan bias. Gradien ini menunjukkan arah dan besarnya "kemiringan" dari fungsi *loss*, sehingga model dapat bergerak ke arah yang berlawanan untuk mencapai titik terendah (minimum). Dibandingkan dengan algoritma optimasi lainnya seperti *batch gradient descent* dan *mini-batch gradient descent*, dengan perbandingan sebagai berikut:

- a. *Batch Gradient Descent* memperbarui parameter setelah memproses seluruh dataset. Metode ini menjamin konvergensi yang stabil, tetapi sangat lambat dan memerlukan banyak memori saat dataset berukuran besar.
- b. *Mini-Batch Gradient Descent* menggunakan subset data (mini-batch) untuk menghitung gradien. Metode ini menawarkan keseimbangan yang baik antara kecepatan dan stabilitas, menjadikannya pilihan yang paling umum digunakan.

- c. SGD memperbarui parameter model untuk setiap sampel data secara berurutan. Hal ini menghasilkan pembaruan parameter yang sangat sering dan cepat.

SGD memiliki keunggulan, seperti konvergensi yang lebih cepat dan kemampuan untuk menangani dataset yang cukup besar.

2.2.5 *Term Frequency – Inverse Document Frequency*

TF-IDF adalah teknik yang dapat digunakan untuk representasikan sebuah teks umum dalam NLP yang menghitung seberapa penting sebuah kata tersebut (term) dalam satu dokumen relatif terhadap sebuah korpus. Hal ini dilakukan dengan cara mengalikan dua komponen utama yaitu:

1. *Term Frequency* (TF) berfungsi untuk mengukur seberapa sering sebuah kata muncul di dalam satu dokumen. Semakin sering sebuah kata tersebut muncul, semakin tinggi nilai TF-nya, menunjukkan bahwa kata tersebut relevan dengan topik di dokumen tersebut.
2. *Inverse Document Frequency* (IDF) memiliki fungsi untuk penyeimbang dengan mengukur seberapa umum atau langka sebuah kata dalam seluruh kumpulan dokumen. Kata-kata yang umum seperti "dan," "yang," atau "di," yang muncul di hampir semua dokumen, akan memiliki nilai IDF rendah, mendekati nol. Ini karena kata-kata tersebut tidak memberikan informasi unik yang dapat membedakan satu dokumen dari yang lainnya. Sebaliknya, kata-kata yang jarang muncul di koleksi dokumen, tetapi muncul dalam dokumen dengan topik yang spesifik, akan mendapatkan nilai IDF yang tinggi.

Untuk mendapatkan skor TF-IDF yaitu dilakukan dengan cara mengalikan nilai TF dan IDF, kita mendapatkan skor TF-IDF. Skor ini akan tinggi jika sebuah kata sering muncul dalam sebuah dokumen (TF tinggi) tetapi jarang muncul di dokumen lain (IDF tinggi). Hasilnya TF-IDF efektif dalam mengekstraksi kata kunci penting (Wibowo, Witanti, Susilawati, 2024). Yang

dapat digunakan untuk klasifikasi maksud (intent classification) pada chatbot. Kata-kata dengan skor TF-IDF tinggi adalah petunjuk kuat tentang topik atau tujuan dari sebuah pertanyaan.

2.2.6 Confusion Matrix

Untuk menilai kinerja model klasifikasi, digunakan *Confusion Matrix*. Matriks ini menampilkan hasil prediksi dalam format tabel, merinci empat kategori: *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN), yang memberikan gambaran tentang performa prediksi. benar dan salah untuk setiap kelas, membantu dalam mengidentifikasi kesalahan (Tolla 2025). *Confusion Matrix* berguna untuk memahami kinerja model dalam konteks masalah klasifikasi dengan dua kelas atau lebih.

Pada tahap pengujian menggunakan *confusion matrix*, langkah pertama adalah melakukan prediksi terhadap sejumlah data uji, di mana data ini berisi sampel input beserta label kelas (intent) yang benar. Setiap input diproses oleh model untuk menghasilkan prediksi *intent*, yang kemudian dibandingkan dengan label sebenarnya. Hasil prediksi dan label sebenarnya ini diolah menjadi *confusion matrix*, yaitu sebuah tabel yang menunjukkan seberapa sering model berhasil memprediksi *intent* dengan benar (benar positif/benar negatif) dan seberapa sering model salah memprediksi (salah positif/salah negatif) untuk setiap kelas.

Confusion matrix memberikan informasi detail tentang performa model dalam bentuk empat nilai utama: *true positive* (TP) untuk prediksi benar yang sesuai, *false positive* (FP) untuk prediksi salah yang seharusnya tidak ada, *false negative* (FN) untuk data yang tidak terprediksi dengan benar padahal seharusnya ada, dan *true negative* (TN) untuk prediksi salah yang terhindar. Dari hasil ini, kita dapat menghitung metrik penting, seperti akurasi, presisi, *recall*, dan *F1 score*, yang memberikan gambaran menyeluruh tentang keakuratan dan ketelitian model.

2.2.7 *Classification Report*

Classification report merupakan alat evaluasi penting dalam pengujian model klasifikasi, yang memberikan gambaran menyeluruh mengenai kinerja model dalam membedakan masing-masing kelas (intent) berdasarkan hasil prediksi terhadap data uji. Laporan ini menyajikan sejumlah metrik utama seperti *precision*, *recall*, *f1-score*, dan *support* untuk setiap kelas secara terpisah, sehingga memungkinkan evaluasi performa model yang lebih mendalam dan terperinci dibandingkan hanya menggunakan satu metrik global.

1. *Precision*

Precision adalah perbandingan antara jumlah prediksi yang benar untuk suatu kelas dibandingkan dengan seluruh prediksi yang diberikan untuk kelas tersebut (Nasution & Hayaty, 2019). Metrik ini menilai seberapa akurat suatu model ketika memutuskan bahwa sebuah input tersebut termasuk dalam suatu kelas tertentu.

Secara matematis, *precision* dirumuskan sebagai:

$$Precision = \frac{TP}{TP + FP}$$

TP (True Positive) adalah jumlah prediksi benar dalam suatu kelas tertentu, FP (False Positive) adalah jumlah prediksi yang salah, yaitu kondisi saat model mengira input termasuk ke dalam kelas tersebut padahal tidak. Nilai *precision* yang tinggi dapat menunjukkan bahwa model jarang salah dalam membuat prediksi positif, sehingga nilai ini berfungsi untuk menghindari kesalahan prediksi yang keliru (false alarm).

2. *Recall*

Recall adalah perbandingan antara jumlah prediksi yang benar pada suatu kelas dengan seluruh data yang seharusnya masuk ke dalam kelas tersebut. Metrik ini menunjukkan seberapa baik suatu model dalam menangkap semua contoh dari suatu kelas tertentu.

Rumus *recall* adalah sebagai berikut:

$$Recal = \frac{TP}{TP + FN}$$

FN (False Negative) adalah jumlah data dari suatu kelas yang tidak berhasil diprediksi oleh model. Nilai *recall* yang tinggi menunjukkan bahwa model dapat mengenali sebagian besar data yang benar

3. *F1-score*

F1-score merupakan rata-rata harmonik dari *precision* dan *recall*, dan digunakan untuk dapat memberikan hasil nilai yang menyeluruh mengenai keseimbangan antara keduanya (Putri, Taryo & Hindasyah, 2025). Nilai *F1-score* akan tinggi jika nilai *precision* dan nilai *recall* tinggi, dan akan menurun jika salah satunya nilai dari *precision* dan *recall* rendah.

Rumus *F1-score* adalah:

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

F1-score sangat berguna jika digunakan dalam konteks klasifikasi yang tidak seimbang, yaitu ketika jumlah data pada tiap kelas berbeda-beda. Metrik ini menghubungkan keperluan akan ketepatan dan kelengkapan klasifikasi secara bersamaan.