

**PERANCANGAN SISTEM *BUSINESS INTELLIGENCE*
BERBASIS *CHATBOT* DENGAN *RULE-BASED QUERY*,
NATURAL LANGUAGE PROCESSING, DAN *TIME SERIES*
FORECASTING UNTUK ANALISIS DAN PREDIKSI
PERMINTAAN LAYANAN PELANGGAN**

Proposal Tugas Akhir

Oleh

**Thalita Zahra Sutejo
18222023**



**PROGRAM STUDI SISTEM DAN TEKNOLOGI INFORMASI
SEKOLAH TEKNIK ELEKTRO DAN INFORMATIKA
INSTITUT TEKNOLOGI BANDUNG
Oktober 2025**

LEMBAR PENGESAHAN

PERANCANGAN SISTEM *BUSINESS INTELLIGENCE* BERBASIS *CHATBOT* DENGAN *RULE-BASED QUERY*, *NATURAL LANGUAGE PROCESSING*, DAN *TIME SERIES* *FORECASTING* UNTUK ANALISIS DAN PREDIKSI PERMINTAAN LAYANAN PELANGGAN

Proposal Tugas Akhir

Oleh

Thalita Zahra Sutejo
18222023

Program Studi Sistem dan Teknologi Informasi
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung

Proposal Tugas Akhir ini telah disetujui dan disahkan
di Bandung, pada tanggal 28 Oktober 2025

Pembimbing

Dr. Ir. Dimitri Mahayana, M.Eng.

NIP. 196808091991021001

DAFTAR ISI

DAFTAR GAMBAR	iv
DAFTAR TABEL	v
I PENDAHULUAN	1
I.1 Latar Belakang	1
I.2 Rumusan Masalah	3
I.3 Tujuan	4
I.4 Batasan Masalah	5
I.5 Metodologi	5
I.5.1 Tahapan Investigasi dan Pengumpulan Fakta	6
I.5.2 Studi Literatur	7
I.5.3 Perancangan Sistem	7
I.5.4 Implementasi Sistem	8
I.5.5 Pengujian dan Evaluasi	8
I.5.6 Dokumentasi dan Penyusunan Laporan	9
II STUDI LITERATUR	10
II.1 Business Intelligence dan Arsitektur Sistem	10
II.1.1 Konsep Business Intelligence	10
II.1.2 Arsitektur Business Intelligence Hibrida	10
II.1.3 Semantic Layer dalam Business Intelligence	11
II.1.3.1 Komponen Inti Semantic Layer	11
II.2 Rule-Based Query Optimization	12
II.2.1 Konsep Rule-Based Query	12
II.2.2 AI-Driven Query Optimization	12
II.2.3 Automated Query Rewriting	12
II.3 Chatbot dan Conversational AI	13
II.3.1 Konsep dan Arsitektur Chatbot	13
II.3.2 Komponen Utama Chatbot	13
II.3.2.1 Natural Language Understanding (NLU)	13
II.3.3 Multi-task Learning untuk NLU	13
II.4 Natural Language Processing untuk Chatbot	14
II.4.1 Konsep NLP dalam Aplikasi Chatbot	14
II.4.2 Implementasi NLP dalam Layanan Pelanggan	14

II.4.3	Model BERT untuk Intent Classification	14
II.4.4	Pattern Matching dalam NLP	15
II.5	Time Series Forecasting	15
II.5.1	Konsep Time Series dan Peramalan	15
II.5.2	Model ARIMA untuk Time Series	15
II.5.3	Model SARIMA untuk Data Musiman	16
II.5.4	Model LSTM untuk Time Series	16
II.5.5	Perbandingan ARIMA, SARIMA, dan LSTM	17
II.5.5.1	Studi Empiris Perbandingan Model	17
II.5.5.2	Model Hibrida ARIMA-LSTM	17
II.5.5.3	Perbandingan Karakteristik Model	18
II.6	Metrik Evaluasi	18
II.6.1	Metrik untuk Klasifikasi NLP	18
II.6.1.1	Accuracy (Akurasi)	18
II.6.1.2	Precision (Presisi)	18
II.6.1.3	Recall (Daya Ingat)	19
II.6.1.4	F1-Score	19
II.6.2	Metrik untuk Time Series Forecasting	19
II.6.2.1	Mean Absolute Error (MAE)	19
II.6.2.2	Root Mean Square Error (RMSE)	19
II.6.2.3	Mean Absolute Percentage Error (MAPE)	19
II.7	Integrasi Sistem dan Penelitian Terkait	20
II.7.1	Sistem Hibrida Rule-Based dan NLP	20
II.7.2	Customer Churn Prediction	20

DAFTAR GAMBAR

DAFTAR TABEL

II.1	Perbandingan Metode Peramalan Deret Waktu	18
------	---	----

BAB I

PENDAHULUAN

I.1 Latar Belakang

Transformasi digital di sektor bisnis telah mengubah cara perusahaan menganalisis dan memanfaatkan data untuk pengambilan keputusan strategis. *Business Intelligence* (BI) kini menjadi tulang punggung operasional organisasi modern yang ingin tetap kompetitif di pasar global. Pasar global *Business Intelligence* diproyeksikan mencapai USD 36,82 miliar pada tahun 2025 dan tumbuh menjadi USD 116,25 miliar pada tahun 2033, dengan tingkat pertumbuhan tahunan gabungan (CAGR) sebesar 14,98% (Straits Research 2024). Pertumbuhan ini menunjukkan bahwa organisasi di seluruh dunia semakin menyadari nilai strategis sistem analitik berbasis data untuk meningkatkan efisiensi operasional dan daya saing.

Di Indonesia, adopsi teknologi kecerdasan buatan (*Artificial Intelligence/AI*) dan digitalisasi juga mengalami akselerasi signifikan. Pemerintah Indonesia memproyeksikan bahwa AI akan menghasilkan manfaat ekonomi hingga USD 366 miliar selama dekade mendatang (Intimedia 2024). Selain itu, menurut laporan *East Ventures—Digital Competitiveness Index 2025*, lebih dari 80% pelaku bisnis di Indonesia telah mengintegrasikan AI ke dalam operasi bisnis, meskipun hanya sekitar 13% yang menggunakan AI pada tingkat lanjutan (East.vc 2025).

Salah satu tantangan utama dalam implementasi BI adalah kompleksitas akses dan interpretasi data bagi pengguna nonteknis. Sistem BI konvensional sering memerlukan keahlian khusus dalam bahasa kueri (seperti SQL) dan pemahaman mendalam tentang struktur basis data, sehingga menjadi hambatan bagi manajer dan staf operasional yang membutuhkan *insight* cepat untuk keputusan harian. Untuk mengatasi hal tersebut, teknologi *Conversational AI* berbasis *chatbot* muncul sebagai solusi yang menjanjikan. Pasar *Conversational AI* global diperkirakan mencapai USD 49,7 miliar pada tahun 2025 dan tumbuh dengan CAGR sekitar 23,6% hingga 2032 (Qaltivate 2025).

Implementasi *chatbot* dalam konteks layanan pelanggan dan operasional internal telah menunjukkan hasil yang positif. Ringkasan industri menunjukkan bahwa *chatbot* dapat menjawab hingga 80% pertanyaan standar pelanggan, mempercepat waktu respons, dan memangkas beban kerja agen manusia (Fullview 2025). Dalam konteks BI, *chatbot* berfungsi sebagai antarmuka percakapan yang memungkinkan pengguna internal, seperti manajer penjualan, tim layanan pelanggan, dan analis

bisnis, mengakses data analitik, laporan, dan prediksi melalui kueri bahasa alami tanpa keahlian teknis. Studi (Iqbal, Siregar, dan Rismayanti 2021) menunjukkan bahwa penerapan *natural language processing* (NLP) pada *chatbot* meningkatkan efisiensi dan akurasi pelayanan pelanggan secara signifikan.

Teknologi NLP menjadi fondasi yang memungkinkan sistem memahami dan merespons kueri pengguna dalam bahasa alami. Pasar NLP global diproyeksikan tumbuh dari USD 35,43 miliar pada tahun 2024 menjadi USD 438,08 miliar pada tahun 2034, dengan CAGR sekitar 28,6% (Precedence Research 2025). Dalam aplikasi BI, NLP memungkinkan sistem menginterpretasikan *intent* pengguna, mengekstraksi entitas penting dari kueri, dan menghasilkan respons relevan berdasarkan konteks percakapan, baik melalui pendekatan *rule-based* untuk kueri terstruktur maupun melalui model pembelajaran mesin untuk kueri yang lebih kompleks.

Selain kemampuan analisis deskriptif dan diagnostik, kebutuhan akan kemampuan prediktif dalam BI juga meningkat. Analisis prediktif memungkinkan organisasi mengantisipasi permintaan layanan pelanggan, merencanakan alokasi sumber daya, dan mengidentifikasi potensi *churn* sebelum terjadi. Dalam konteks ini, permintaan layanan pelanggan didefinisikan sebagai minat, pesanan, dan kebutuhan pelanggan terhadap produk atau layanan yang ditawarkan perusahaan, yang dapat diukur melalui data pemesanan, target penjualan, tingkat *churn*, dan pendapatan pelanggan. Pasar *time series forecasting* secara global diperkirakan tumbuh dari sekitar USD 11,17 miliar (2024) menjadi USD 36,9 miliar pada 2032, dengan CAGR ~16,12% (WiseGuyReports 2024).

Metode *forecasting* yang lazim meliputi model statistik seperti ARIMA dan SARIMA, serta model *deep learning* (misalnya LSTM dan GRU) untuk menangkap pola temporal. Pada kasus prediksi *customer churn* di industri telekomunikasi, (P. Lalwani dkk. 2022) menunjukkan bahwa *random forest*, SVM, dan XGBoost dapat mencapai akurasi tinggi, sehingga memungkinkan tindakan proaktif terhadap potensi keluarnya pelanggan.

Integrasi tiga komponen utama, *rule-based query*, NLP, dan *time series forecasting*, ke dalam satu sistem BI berbasis *chatbot* menawarkan solusi komprehensif bagi kebutuhan analisis dan prediksi di lingkungan bisnis modern. Sistem demikian berpotensi memperkuat pengambilan keputusan proaktif berbasis data serta meningkatkan efisiensi operasional internal perusahaan.

I.2 Rumusan Masalah

Sebagian besar sistem *Business Intelligence* untuk analisis indikator permintaan terkait pelanggan di lingkungan internal perusahaan pada saat ini masih mengandalkan pelaporan manual, belum terintegrasi dengan antarmuka percakapan berbasis *chatbot*, dan minim pemanfaatan kecerdasan buatan untuk analisis prediktif serta interaksi berbasis bahasa alami. Kondisi ini memunculkan sejumlah permasalahan sebagai berikut.

1. **Keterbatasan akses analitik berbasis percakapan.**

Tidak tersedia sistem yang mampu memproses dan memahami permintaan data serta analisis bisnis secara otomatis dari pengguna internal melalui percakapan dalam bahasa alami. Hal ini menyebabkan pegawai mengalami kendala dalam memperoleh *insight* terkait pesanan (order), pencapaian target penjualan, *churn* pelanggan, serta pendapatan secara cepat dan intuitif untuk mendukung pengambilan keputusan harian.

2. **Ketiadaan fitur prediksi yang mudah diakses pengguna nonteknis.**

Belum terdapat fitur prediksi indikator utama bisnis terkait pelanggan, seperti peramalan jumlah pesanan, *churn* pelanggan, dan target pendapatan berbasis pembelajaran mesin atau peramalan deret waktu yang dapat diakses oleh pengguna nonteknis. Keadaan ini menyebabkan perencanaan kapasitas, evaluasi target, dan penetapan strategi bisnis perusahaan menjadi kurang responsif terhadap tren serta dinamika bisnis pada masa yang akan datang.

3. **Minimnya integrasi *rule-based query* dan *natural language processing*.**

Integrasi *rule-based query* dan pemrosesan bahasa alami (*natural language processing*) dalam sistem BI internal masih sangat terbatas, sehingga proses pencarian data, penyaringan metrik utama, serta penyusunan laporan kinerja bisnis masih harus dilakukan secara manual dan berulang. Keadaan tersebut menimbulkan keterlambatan pengambilan keputusan serta penyajian data yang kurang selaras dengan kebutuhan analisis waktu nyata oleh manajemen.

4. **Kurangnya analitik interaktif dan *insight* prediktif waktu nyata.**

Sistem BI yang ada belum menyediakan *dashboard* analitik interaktif dan *insight* prediktif secara waktu nyata, baik melalui antarmuka percakapan *chatbot* maupun visualisasi data, yang mudah diakses oleh seluruh pengguna internal. Akibatnya, pemangku kepentingan sering kehilangan kesempatan untuk mendeteksi perubahan pola permintaan, anomali pesanan, *churn* yang mendadak, atau potensi pendapatan yang belum tergarap optimal.

Bagaimana merancang dan mengembangkan sistem *Business Intelligence* berbasis *chatbot* yang terintegrasi dengan teknologi *rule-based query*, pemrosesan bahasa

alami (*natural language processing*), dan algoritma peramalan deret waktu, sehingga seluruh pengguna internal perusahaan dapat menganalisis, memantau, dan memprediksi indikator bisnis pelanggan, seperti pesanan, target, *churn*, dan pendapatan, secara waktu nyata melalui antarmuka percakapan, serta memperoleh *insight* berbasis data yang akurat, efisien, dan relevan untuk mendukung pengambilan keputusan bisnis dan perencanaan strategis perusahaan?

I.3 Tujuan

Tujuan utama tugas akhir ini adalah merancang dan mengembangkan sistem *Business Intelligence* berbasis *chatbot* yang terintegrasi dengan teknologi *rule-based query*, pemrosesan bahasa alami (*Natural Language Processing/NLP*), dan algoritma peramalan deret waktu (*time series forecasting*) untuk memfasilitasi pengguna internal perusahaan dalam menganalisis dan memprediksi indikator bisnis pelanggan secara interaktif dan waktu nyata (*real-time*).

Secara terperinci, tujuan yang hendak dicapai dalam tugas akhir ini adalah sebagai berikut.

1. Merancang arsitektur sistem *Business Intelligence* berbasis *chatbot* yang mampu memproses permintaan analisis data dan memberikan *insight* berbasis data kepada pengguna internal perusahaan secara otomatis.
2. Mengembangkan modul pemrosesan bahasa alami sehingga pengguna internal dapat mengajukan pertanyaan dan permintaan analisis bisnis menggunakan bahasa alami melalui antarmuka *chatbot*.
3. Mengimplementasikan mekanisme *rule-based query* untuk menangani permintaan data yang bersifat terstruktur dan berulang dengan respons yang cepat dan konsisten.
4. Mengintegrasikan model peramalan deret waktu (*time series forecasting*) untuk menghasilkan prediksi terhadap indikator bisnis pelanggan, seperti volume pesanan, pencapaian target penjualan, tingkat *churn* pelanggan, dan proyeksi pendapatan.
5. Membangun antarmuka percakapan dan *dashboard* visualisasi yang menampilkan hasil analisis serta prediksi secara interaktif dan mudah dipahami oleh pengguna internal.
6. Menguji dan mengevaluasi kinerja sistem dari segi akurasi prediksi, kecepatan respons, serta kemudahan penggunaan (*usability*) oleh pengguna internal perusahaan.

I.4 Batasan Masalah

Dalam memfokuskan ruang lingkup penelitian dan memastikan kelayakan pelaksanaan tugas akhir, batasan-batasan penelitian ditetapkan sebagai berikut.

1. **Pengguna sistem.** Sistem dirancang khusus untuk pengguna internal perusahaan (manajer, analis bisnis, dan staf operasional), bukan untuk pelanggan eksternal.
2. **Bahasa.** Sistem *chatbot* hanya mendukung interaksi dalam Bahasa Indonesia dan Bahasa Inggris.
3. **Data.** Data yang digunakan terbatas pada data historis pelanggan yang mencakup informasi pesanan (*order*), target penjualan, tingkat *churn*, dan pendapatan dalam kurun waktu minimal dua tahun terakhir.
4. **Jenis permintaan (*query*).** Sistem hanya menangani permintaan terkait analisis deskriptif, diagnostik, dan prediktif terhadap indikator bisnis pelanggan. Sistem tidak mencakup fungsi transaksional atau modifikasi data.
5. **Metode peramalan.** Penggunaan metode peramalan deret waktu dibatasi pada algoritma ARIMA, SARIMA, dan LSTM atau kombinasi dari ketiganya.
6. **Kapasitas *intent chatbot*.** Sistem dirancang untuk mengenali maksimal 20 hingga 30 *intent* utama yang berkaitan dengan analisis dan prediksi indikator bisnis pelanggan.
7. **Integrasi sistem.** Sistem yang dikembangkan merupakan purwarupa mandiri (*standalone*) dan belum terintegrasi secara penuh dengan sistem ERP (*Enterprise Resource Planning*) atau CRM (*Customer Relationship Management*) yang ada di perusahaan.
8. **Platform.** Purwarupa sistem diimplementasikan pada platform berbasis *web* dengan menggunakan *framework* yang umum untuk pengembangan *chatbot* dan visualisasi data.
9. **Evaluasi.** Evaluasi sistem dilakukan melalui pengujian fungsional, pengukuran akurasi prediksi, analisis waktu respons, serta survei kepuasan pengguna dengan jumlah responden terbatas (minimal 20 hingga 30 pengguna internal).
10. **Keamanan.** Implementasi fitur keamanan dan privasi data dibatasi pada tingkat dasar, seperti autentikasi pengguna dan enkripsi koneksi, tanpa mencakup standar keamanan tingkat *enterprise* seperti ISO 27001.

I.5 Metodologi

Metodologi pelaksanaan tugas akhir ini mengadopsi kerangka kerja hibrida yang menggabungkan *CRISP-DM* (*Cross-Industry Standard Process for Data Mining*)

untuk tahapan pengolahan data dan analitik dengan prinsip *Agile* untuk pengembangan perangkat lunak. Pendekatan ini dipilih agar proses penelitian berjalan secara iteratif, luwes, dan tetap sistematis dalam menyelesaikan permasalahan yang telah dirumuskan.

CRISP-DM (*Cross-Industry Standard Process for Data Mining*) adalah metodologi standar lintas industri untuk penambangan data yang menyediakan pendekatan terstruktur dalam enam fase utama: pemahaman bisnis (*business understanding*), pemahaman data (*data understanding*), persiapan data (*data preparation*), pemodelan (*modeling*), evaluasi (*evaluation*), dan penyebaran (*deployment*) (DataScience-PM, 2024). Metodologi ini bersifat iteratif, yang memungkinkan pengulangan fase jika diperlukan untuk meningkatkan kualitas hasil.

Agile adalah pendekatan pengembangan perangkat lunak yang menekankan keluwesan, kerja sama, dan pengiriman nilai secara berkelanjutan melalui prinsip-prinsip iterasi pendek (*sprint*), kolaborasi dengan pemangku kepentingan, dan adaptasi terhadap perubahan kebutuhan (Asana, 2025).

Dalam penelitian ini, **CRISP-DM digunakan** untuk mengelola aspek penambangan data dan analitik prediktif (khususnya pengembangan model peramalan deret waktu dan NLP), sementara **Agile digunakan** untuk mengelola pengembangan perangkat lunak *chatbot* dan sistem BI secara keseluruhan dengan pembagian kerja dalam *sprint* iteratif.

I.5.1 Tahapan Investigasi dan Pengumpulan Fakta

Tahapan ini bertujuan untuk mengumpulkan data dan informasi faktual sebagai landasan perumusan masalah. Kegiatan yang dilakukan meliputi:

- a. **Observasi sistem yang ada**, yaitu pengamatan terhadap sistem *Business Intelligence* yang saat ini digunakan organisasi untuk mengidentifikasi keterbatasan dan kebutuhan pengguna internal.
- b. **Wawancara dengan pemangku kepentingan**, dilakukan secara terstruktur dengan minimal lima hingga sepuluh pemangku kepentingan internal (manajer layanan pelanggan, analis data, staf operasional) untuk menggali kebutuhan fungsional dan ekspektasi terhadap sistem baru.
- c. **Survei kebutuhan pengguna**, disebarkan kepada minimal tiga puluh hingga lima puluh calon pengguna untuk mengidentifikasi kebutuhan informasi, preferensi interaksi, dan kebutuhan analitik.
- d. **Analisis data historis**, dilakukan untuk menelaah volume dan tren permintaan, pola musiman, serta kelengkapan dan kualitas data yang tersedia.

- e. **Dokumentasi temuan**, mengompilasi seluruh hasil observasi, wawancara, survei, dan analisis dalam bentuk laporan yang akan dilampirkan.

I.5.2 Studi Literatur

Tahapan ini bertujuan untuk mengumpulkan, mengelompokkan, dan menyaring literatur yang relevan dengan topik penelitian. Proses yang dilakukan meliputi:

- a. **Pencarian literatur sistematis** pada basis data akademik (IEEE Xplore, ScienceDirect, SpringerLink, ACM Digital Library, Google Scholar) menggunakan kata kunci terkait *business intelligence*, *chatbot*, *natural language processing*, dan *time series forecasting*. Kriteria inklusi adalah artikel telaah sejawat (*peer-reviewed*) tahun 2020–2025 yang relevan dengan topik.
- b. **Pengelompokan literatur** berdasarkan tema utama: konsep *Business Intelligence*, teknologi *chatbot* dan *conversational AI*, pemrosesan bahasa alami, metode peramalan deret waktu, integrasi sistem, dan studi kasus implementasi.
- c. **Penapisan dan analisis kualitas** literatur berdasarkan faktor dampak, jumlah sitasi, metodologi, dan relevansi terhadap penelitian.
- d. **Sintesis literatur** untuk membangun kerangka teoretis, mengidentifikasi praktik terbaik dan teknologi terkini, serta menyusun tinjauan pustaka pada Bab II.
- e. **Dokumentasi proses** dalam bentuk matriks literatur dan pengelolaan referensi menggunakan *reference manager*.

I.5.3 Perancangan Sistem

Tahapan ini meliputi perancangan arsitektur dan komponen sistem secara terperinci, mencakup:

- a. **Perancangan arsitektur sistem**, yaitu penyusunan arsitektur berlapis yang mengintegrasikan *chatbot*, pemrosesan bahasa alami, mesin *rule-based query*, dan peramalan deret waktu, serta penentuan teknologi dan kerangka kerja yang akan digunakan.
- b. **Perancangan modul *chatbot***, meliputi komponen pemahaman bahasa alami (*Natural Language Understanding*), manajemen dialog (*Dialogue Management*), pembangkitan bahasa alami (*Natural Language Generation*), serta pendefinisian *intent* dan entitas yang akan dikenali.
- c. **Perancangan mesin *rule-based query***, yaitu pendefinisian aturan untuk kueri terstruktur dan berulang serta pemetaan antara *intent* dengan templat kueri.
- d. **Perancangan model peramalan deret waktu**, meliputi penentuan variabel target, perancangan praolah data, pemilihan arsitektur model (ARIMA, SA-

RIMA, LSTM), dan strategi evaluasi model.

- e. **Perancangan basis data dan gudang data**, yaitu perancangan skema basis data untuk menyimpan data pelanggan, hasil prediksi, dan data pelatihan NLP, serta perancangan proses ETL (*Extract, Transform, Load*).
- f. **Perancangan antarmuka pengguna**, meliputi pembuatan *wireframe* dan *moc-up* untuk antarmuka *chatbot* dan *dashboard* visualisasi yang intuitif.

I.5.4 Implementasi Sistem

Tahapan ini meliputi pengembangan sistem berdasarkan rancangan yang telah ditetapkan, mencakup:

- a. **Pengembangan *backend***, yaitu implementasi mesin *rule-based query*, model NLP untuk klasifikasi *intent* dan pengenalan entitas, model peramalan deret waktu, serta penyediaan API *backend*.
- b. **Pengembangan *frontend***, meliputi implementasi antarmuka *chatbot* dan *dashboard* visualisasi pada platform *web* dengan tata letak responsif.
- c. **Pengembangan basis data**, yaitu implementasi skema basis data, proses ETL untuk integrasi data, dan *data pipeline* untuk pemrosesan data.
- d. **Integrasi komponen**, meliputi integrasi *chatbot* dengan mesin *rule-based query* dan NLP, integrasi model peramalan dengan sistem BI, serta implementasi mekanisme tembolok (*caching*) untuk performa.
- e. **Implementasi RAG (*Retrieval-Augmented Generation*)** secara opsional, jika diperlukan untuk meningkatkan kemampuan *chatbot* dalam mengambil informasi dari basis pengetahuan.

I.5.5 Pengujian dan Evaluasi

Tahapan ini bertujuan untuk menguji dan mengevaluasi kinerja sistem secara menyeluruh, meliputi:

- a. **Pengujian unit**, yaitu pengujian setiap komponen secara individual untuk memastikan fungsi berjalan dengan benar.
- b. **Pengujian integrasi**, untuk memverifikasi interaksi antarkomponen dan kebenaran aliran data dalam sistem.
- c. **Pengujian fungsional**, yaitu validasi pemenuhan seluruh kebutuhan fungsional melalui skenario ujung-ke-ujung (*end-to-end*).
- d. **Evaluasi model NLP**, menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* untuk mengukur kinerja klasifikasi *intent*.
- e. **Evaluasi model peramalan deret waktu**, menggunakan metrik MAE (*Mean Absolute Error*), RMSE (*Root Mean Square Error*), dan MAPE (*Mean Absolute Percentage Error*) untuk mengukur akurasi prediksi.

- f. **Evaluasi performa sistem**, meliputi pengukuran waktu respons, *throughput*, dan latensi, serta pengujian beban (*load testing*) untuk menilai skalabilitas.
- g. **Evaluasi kebergunaan (*usability*)**, melalui *User Acceptance Testing* (UAT) dengan minimal dua puluh hingga tiga puluh pengguna internal dan pengukuran *System Usability Scale* (SUS).
- h. **Analisis dan interpretasi hasil**, yaitu kompilasi hasil pengujian, perbandingan terhadap target, dan identifikasi kelebihan serta area perbaikan sistem.

I.5.6 Dokumentasi dan Penyusunan Laporan

Tahapan akhir meliputi dokumentasi dan penyusunan laporan tugas akhir, mencakup:

- a. **Dokumentasi teknis**, yaitu penyusunan dokumentasi API, dokumentasi kode, panduan pengguna (*user manual*), dan panduan penyebaran (*deployment guide*).
- b. **Penyusunan laporan tugas akhir**, meliputi penulisan laporan sesuai format yang ditetapkan (Bab I hingga Bab VI) dengan penyuntingan dan pemeriksaan kelengkapan.
- c. **Persiapan presentasi**, meliputi pembuatan *slide* presentasi, penyusunan skenario demonstrasi sistem, dan kurasi materi pendukung.
- d. **Revisi dan finalisasi**, yaitu revisi berdasarkan masukan pembimbing, finalisasi laporan dan sistem, serta persiapan sidang tugas akhir.

Seluruh proses metodologi ini akan didokumentasikan dengan baik untuk memastikan keterbukaan, kemampuan direproduksi, dan akuntabilitas penelitian.

BAB II

STUDI LITERATUR

Bab ini menyajikan kajian teoretis dan tinjauan pustaka yang relevan dengan topik penelitian. Pembahasan meliputi konsep *Business Intelligence*, teknologi chatbot dan *Conversational AI*, pemrosesan bahasa alami, serta peramalan deret waktu yang mendukung analisis dan prediksi permintaan layanan pelanggan.

II.1 Business Intelligence dan Arsitektur Sistem

II.1.1 Konsep Business Intelligence

Business Intelligence (BI) merupakan sistem teknologi yang mengumpulkan, mengorganisasi, dan menganalisis data dari berbagai sumber dalam suatu bisnis untuk membantu perusahaan mengubah data mentah menjadi wawasan yang berguna sehingga mereka dapat membuat keputusan yang lebih baik (Sorour 2020). Evolusi BI dari sistem berbasis aturan tradisional ke platform berbasis kecerdasan buatan telah meningkatkan secara signifikan analitik prediktif, otomatisasi, dan kemampuan pengambilan keputusan waktu nyata di seluruh industri (Khaddam dan Alzghoul 2025).

Menurut Sorour (2020), implementasi BI di institusi pendidikan tinggi telah menunjukkan kemampuan untuk memantau aktivitas jaminan kualitas dan mendukung pengambilan keputusan strategis. Penelitian Shah (2025) menunjukkan bahwa organisasi yang mengimplementasikan arsitektur analitik hibrida melaporkan peningkatan kemampuan penanganan data sebesar 85% dibandingkan sistem tradisional, dengan konsistensi dan standar kualitas data yang terjaga.

II.1.2 Arsitektur Business Intelligence Hibrida

Arsitektur BI hibrida mengintegrasikan sistem BI tradisional dengan kemampuan kecerdasan buatan untuk meningkatkan pengambilan keputusan, analitik prediktif, dan efisiensi operasional. Shah (2025) menyajikan pendekatan terstruktur yang memanfaatkan model pembelajaran mesin bersama pelaporan BI tradisional untuk menjembatani kesenjangan antara analisis historis dan wawasan berbasis data waktu nyata.

Komponen utama arsitektur BI hibrida meliputi (Shah 2025):

1. **Lapisan Sumber Data:** Mencakup sumber internal (sistem CRM, perangkat lunak keuangan) dan eksternal (data media sosial, laporan pasar).
2. **Gudang Data (*Data Warehouse*):** Fasilitas penyimpanan terpusat untuk data yang terorganisasi dan siap digunakan untuk pelaporan.

3. **Mesin Analitik:** Bertanggung jawab untuk pengenalan pola dan analisis tren, memberikan wawasan prediktif untuk perencanaan strategis.
4. **Lapisan Visualisasi:** Dashboard interaktif dan laporan yang meningkatkan adopsi pengguna melalui presentasi data yang intuitif.

Penelitian menunjukkan bahwa implementasi arsitektur analitik hibrida menghasilkan peningkatan 27% dalam proses pengambilan keputusan dan peningkatan 31% dalam metrik kinerja organisasi secara keseluruhan (Shah 2025).

II.1.3 Semantic Layer dalam Business Intelligence

Semantic layer adalah antarmuka berorientasi bisnis yang menjembatani kesenjangan antara model data yang kompleks dan pengguna bisnis, bertindak sebagai lapisan abstraksi yang menerjemahkan struktur data teknis ke dalam istilah dan konsep bisnis yang familiar (Databricks 2025). Lapisan semantik menyediakan pandangan bisnis terpadu terhadap data di seluruh organisasi, terlepas dari data berada atau bagaimana secara teknis terstruktur.

Menurut AtScale (2025), lapisan semantik menyederhanakan konsep dan teknis data untuk pengguna bisnis sehingga mereka tidak perlu mengubah data bisnis yang mendasarinya untuk bekerja dengan cara baru. Lapisan ini memungkinkan pengguna bisnis untuk berinteraksi dengan data menggunakan terminologi yang mereka pahami tanpa perlu memahami struktur teknis basis data yang mendasarinya.

II.1.3.1 Komponen Inti Semantic Layer

Lapisan semantik memiliki lima komponen inti yang bertindak sebagai fondasi struktural dan teknis (Labs 2024):

1. **Definisi Model Semantik:** Menciptakan representasi logis dari domain bisnis, memetakan struktur basis data teknis ke konsep bisnis. Alih-alih bekerja dengan tabel mentah seperti `usr_tbl` atau `trx_hist`, entitas seperti Customer atau Order didefinisikan yang merangkum kompleksitas yang mendasarinya.
2. **Manajemen Metadata:** Menangani informasi tentang data, seperti deskripsi bidang, garis keturunan data (*data lineage*), frekuensi pembaruan, dan metrik kualitas.
3. **Lapisan Logika Bisnis (*Business Logic Layer*):** Berisi aturan perhitungan spesifik untuk metrik bisnis.
4. **Lapisan Akses Data (*Data Access Layer*):** Menerjemahkan permintaan bisnis menjadi kueri basis data yang dioptimalkan, menerapkan filter keamanan yang diperlukan.

5. **Mekanisme *Caching***: Memeriksa apakah hasil perhitungan sudah tersedia dalam tembolok sebelum menjalankan kueri.

II.2 Rule-Based Query Optimization

II.2.1 Konsep Rule-Based Query

Sistem optimasi kueri berbasis aturan (*rule-based query optimization*) mengandalkan heuristik yang telah ditentukan sebelumnya dan aturan transformasi untuk menghasilkan rencana eksekusi kueri. Pendekatan ini menggunakan sekumpulan aturan yang dikodekan untuk menentukan bagaimana kueri harus diproses berdasarkan pola yang dikenali dalam struktur kueri (Shah 2024).

Menurut Finance dan Gardarin (1994), optimisator kueri berbasis aturan menggunakan bahasa aturan tingkat tinggi yang seragam untuk memodelkan transformasi kueri dan strategi optimasi. Pendekatan ini memungkinkan pemodelan berbagai strategi pencarian dan teknik optimasi dalam kerangka kerja yang terpadu.

II.2.2 AI-Driven Query Optimization

Evolusi dari optimasi berbasis aturan tradisional menuju teknik berbasis kecerdasan buatan menawarkan kemampuan adaptabilitas waktu nyata dan pengambilan keputusan berbasis intelijen (Shah 2024). Sistem berbasis AI menganalisis data eksekusi kueri historis untuk memprediksi konsumsi sumber daya dan menyesuaikan rencana eksekusi secara dinamis.

Shah (2024) menunjukkan bahwa optimisator berbasis AI dapat meningkatkan efisiensi pemanfaatan sumber daya dari 72% (optimasi berbasis aturan) menjadi 89%, dengan peningkatan efisiensi penulisan ulang kueri dari 15% menjadi 40%. Model pembelajaran mesin menganalisis data eksekusi kueri masa lalu untuk mengidentifikasi pola yang dapat digunakan untuk mengoptimalkan kueri masa depan.

II.2.3 Automated Query Rewriting

Penulisan ulang kueri otomatis (*automated query rewriting*) adalah teknik optimasi berbasis AI yang meningkatkan kinerja basis data dengan merestrukturisasi kueri SQL yang tidak efisien tanpa mengubah hasil yang dimaksudkan (Shah 2024). Model pembelajaran mendalam menilai rencana eksekusi kueri untuk mendeteksi ketidakefisienan seperti *join* yang berlebihan, agregasi yang tidak perlu, atau kondisi penyaringan yang tidak optimal.

Keuntungan utama penulisan ulang kueri berbasis AI adalah kemampuannya untuk mengidentifikasi operasi yang berlebihan dan menghilangkan komputasi yang tidak

perlu. Dengan menulis ulang kueri ke dalam struktur yang lebih efisien, optimisator berbasis AI mengurangi waktu eksekusi dan meminimalkan overhead komputasi (Shah 2024).

II.3 Chatbot dan Conversational AI

II.3.1 Konsep dan Arsitektur Chatbot

Chatbot adalah agen percakapan yang menggunakan *Named Entity Recognition* (NER) dan model klasifikasi *intent* yang terintegrasi ke dalam NLU untuk memahami maksud pengguna dan mengekstraksi entitas bernama (Ali dkk. 2020). Penelitian Jumriyah dkk. (2025) menunjukkan bahwa pengguna yang berinteraksi dengan chatbot berbasis bahasa alami melaporkan tingkat kepuasan dan efisiensi yang lebih tinggi dibandingkan dengan mereka yang menggunakan metode layanan pelanggan tradisional.

II.3.2 Komponen Utama Chatbot

II.3.2.1 Natural Language Understanding (NLU)

NLU adalah komponen arsitektur chatbot yang bertanggung jawab untuk memahami dan memproses masukan bahasa manusia (Ali dkk. 2020; Rizou, Giannoulakis, dan Tsapatsoulis 2023). Dua sub-tugas inti NLU dalam konteks agen percakapan adalah:

1. **Named Entity Recognition (NER):** Mengidentifikasi entitas yang telah ditentukan sebelumnya dari teks masukan pengguna, seperti nama orang, lokasi, organisasi, tanggal, atau metrik bisnis (Rizou, Giannoulakis, dan Tsapatsoulis 2023).
2. **Intent Classification:** Menentukan niat atau tujuan pengguna dari masukan mereka, seperti apakah pengguna ingin "memeriksa pesanan", "memprediksi penjualan", atau "menganalisis churn pelanggan" (Perdana dkk. 2025).

II.3.3 Multi-task Learning untuk NLU

Pendekatan *multi-task learning* menggabungkan tugas klasifikasi *intent* dan pengenalan entitas menggunakan model bersama berdasarkan representasi bersama dan dependensi tugas (Perdana dkk. 2025). Penelitian Annisa dkk. (2024) menunjukkan bahwa kombinasi *intent classification* dan NER menggunakan mekanisme *multi-task learning* dalam bentuk model *Dual Intent and Entity Transformer* (DIET) dapat mengatasi masalah representasi numerik yang berbeda yang digunakan dalam setiap model.

Hasil penelitian Perdana dkk. (2025) menunjukkan efektivitas pendekatan yang diusulkan dengan beberapa eksperimen, baik dari klasifikasi *intent* maupun pengenalan

entitas bernama. Kombinasi hyperparameter optimal terdiri dari *warm-up step* sebesar 60, probabilitas *early stopping* sebesar 10, *weight decay* sebesar 0,001, bobot kerugian NER sebesar 0,58, dan bobot kerugian klasifikasi *intent* sebesar 0,4.

II.4 Natural Language Processing untuk Chatbot

II.4.1 Konsep NLP dalam Aplikasi Chatbot

Natural Language Processing (NLP) telah merevolusi chatbot menjadi instrumen penting baik untuk perusahaan maupun individu dengan meningkatkan kapasitas chatbot untuk meniru interaksi mirip manusia melalui teknik seperti identifikasi *intent*, analisis sentimen, dan pemahaman kontekstual (Shinde dan Ambika 2025). NLP memungkinkan chatbot untuk memproses, memahami, dan menghasilkan bahasa manusia secara efisien.

II.4.2 Implementasi NLP dalam Layanan Pelanggan

Penelitian Mulyatun, Kurniawan, dan Hudaya (2021) mengimplementasikan pendekatan NLP pada aplikasi chatbot sebagai alat bantu *customer service* menggunakan metode *Fuzzy String Matching* sebagai media penalarannya. Teknologi chatbot merupakan salah satu bentuk aplikasi NLP yang mempelajari komunikasi antara manusia dengan komputer melalui bahasa alami.

Implementasi chatbot berbasis NLP menggunakan model BERT untuk meningkatkan pelayanan menunjukkan bahwa chatbot dapat memberikan respons yang relevan dan akurat, mempercepat proses konsultasi awal (Ibadurrahman 2025). Sistem dikembangkan dengan metode Agile, dengan tahapan termasuk analisis kebutuhan, desain antarmuka pengguna, pelatihan model, dan implementasi chatbot.

II.4.3 Model BERT untuk Intent Classification

BERT (*Bidirectional Encoder Representations from Transformers*) adalah model yang dilatih secara dua arah untuk memahami konteks kata dari kiri dan kanan. Proyek menggunakan BERT untuk klasifikasi *intent* dalam chatbot kontekstual yang menggunakan NLU untuk memahami maksud pengguna dan merespons dengan sesuai (Vasist dan Krishnan 2021).

Penelitian Vasist dan Krishnan (2021) menunjukkan bahwa akurasi percakapan lebih baik dengan BERT dan dapat ditingkatkan untuk dataset besar. Model BERT memiliki keuntungan utama karena dilatih pada Wikipedia dan *Book Corpus*, dengan kode model *open-source* yang memecahkan beberapa rekor untuk tugas berbasis bahasa yang sulit.

II.4.4 Pattern Matching dalam NLP

Implementasi metode NLP berbasis pola (*pattern-based*) pada chatbot menggunakan pola bahasa untuk memahami dan merespons masukan pengguna. Penelitian Rahayu, Supriyono, dan Prasatya (2024) mengevaluasi kinerja menggunakan metrik BLEU, presisi BLEU-2 mencapai nilai 0,75 atau 75%. Hasil ini menunjukkan bahwa metode berbasis pola mampu menghasilkan respons yang relevan tetapi masih memerlukan penyempurnaan untuk mencapai akurasi yang lebih tinggi.

Data percakapan dikumpulkan dan dianalisis untuk mengidentifikasi pola umum, yang kemudian dipetakan ke dalam respons yang sesuai. Pengujian dilakukan dengan mengukur tingkat kecocokan *bigram* antara respons chatbot dan respons referensi (Rahayu, Supriyono, dan Prasatya 2024).

II.5 Time Series Forecasting

II.5.1 Konsep Time Series dan Peramalan

Time series adalah serangkaian titik data yang diindeks dalam urutan waktu, dan peramalan deret waktu adalah proses menggunakan model untuk memprediksi nilai masa depan berdasarkan nilai yang diamati sebelumnya. Peramalan akurat sangat penting untuk perencanaan strategis dan pengambilan keputusan di berbagai domain bisnis.

II.5.2 Model ARIMA untuk Time Series

ARIMA (*AutoRegressive Integrated Moving Average*) adalah model statistik populer untuk peramalan deret waktu yang menggabungkan tiga komponen: AutoRegressive (AR), Integrated (I), dan Moving Average (MA). Model ARIMA dinyatakan sebagai $ARIMA(p, d, q)$:

- p = jumlah *lag* observasi dalam model (*order* dari komponen AR)
- d = jumlah kali deret waktu harus dibedakan untuk menjadi stasioner
- q = ukuran jendela *moving average* (*order* dari komponen MA)

Persamaan umum model ARIMA dapat ditulis sebagai:

$$\phi(B)(1 - B)^d X_t = \theta(B)\epsilon_t \quad (II.1)$$

X_t adalah nilai deret waktu pada waktu t , B adalah operator *backshift*, $\phi(B)$ adalah polinomial autoregresif, $\theta(B)$ adalah polinomial *moving average*, ϵ_t adalah *white noise error term*, dan d adalah *order* pembedaan.

II.5.3 Model SARIMA untuk Data Musiman

SARIMA (*Seasonal ARIMA*) adalah ekstensi dari ARIMA yang dirancang untuk menangani data deret waktu musiman dengan lebih baik. SARIMA dinyatakan sebagai $SARIMA(p, d, q)(P, D, Q, m)$, (p, d, q) adalah parameter non-musiman dan (P, D, Q) adalah parameter musiman, dengan m sebagai jumlah langkah waktu untuk satu siklus musiman penuh.

Persamaan SARIMA dapat ditulis sebagai:

$$\phi(B)\Phi(B^m)(1-B)^d(1-B^m)^D X_t = \theta(B)\Theta(B^m)\epsilon_t \quad (II.2)$$

$\Phi(B^m)$ dan $\Theta(B^m)$ adalah polinomial autoregresif dan *moving average* musiman.

II.5.4 Model LSTM untuk Time Series

LSTM (*Long Short-Term Memory*) adalah jenis jaringan saraf tiruan berulang yang dirancang khusus untuk mengatasi masalah *vanishing gradient* dalam RNN tradisional. LSTM mampu mempelajari dependensi jangka panjang dalam data sekuens, menjadikannya sangat efektif untuk peramalan deret waktu.

Arsitektur LSTM terdiri dari sel memori dan tiga gerbang:

- **Gerbang Lupa (*Forget Gate*):** Menentukan informasi apa dari sel memori yang harus dibuang.
- **Gerbang Masukan (*Input Gate*):** Menentukan informasi baru apa yang akan disimpan dalam sel memori.
- **Gerbang Keluaran (*Output Gate*):** Menentukan bagian mana dari sel memori yang akan menjadi keluaran.

Persamaan matematis LSTM:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (II.3)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (II.4)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (II.5)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (II.6)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (II.7)$$

$$h_t = o_t * \tanh(C_t) \quad (II.8)$$

f_t adalah gerbang lupa, i_t adalah gerbang masukan, o_t adalah gerbang keluaran, C_t

adalah keadaan sel, h_t adalah keadaan tersembunyi, σ adalah fungsi sigmoid, dan W serta b adalah parameter bobot dan *bias* yang dipelajari.

II.5.5 Perbandingan ARIMA, SARIMA, dan LSTM

II.5.5.1 Studi Empiris Perbandingan Model

Penelitian Siami-Namini, Tavakoli, dan Namin (2018) melakukan studi empiris yang menunjukkan bahwa algoritma berbasis *deep learning* seperti LSTM mengungguli algoritma tradisional seperti ARIMA. Rata-rata pengurangan tingkat kesalahan yang diperoleh oleh LSTM adalah antara 84-87% bila dibandingkan dengan ARIMA, menunjukkan keunggulan LSTM atas ARIMA.

Sirisha dkk. (2022) membandingkan ketiga model untuk prediksi profit dan memperoleh akurasi sebesar 93,84% (ARIMA), 94,378% (SARIMA), dan 97,01% (LSTM). Penelitian ini menemukan bahwa LSTM memberikan akurasi tertinggi untuk prediksi profit. Hasil metrik kinerja menunjukkan (Sirisha dkk. 2022):

- ARIMA: MAE = 33,05; RMSE = 39,28; $R^2 = 0,912$
- SARIMA: MAE = 31,48; RMSE = 38,06; $R^2 = 0,921$
- LSTM: MAE = 28,63; RMSE = 34,91; $R^2 = 0,937$

Sunendar, Wibirama, dan Setiawan (2025) melakukan perbandingan ARIMA, LSTM, dan GRU, dengan hasil menunjukkan bahwa LSTM mengungguli model lain dengan MAPE sebesar 10,76%, diikuti oleh ARIMA pada 11,23% dan GRU pada 11,47%. Temuan ini mengonfirmasi bahwa LSTM memiliki kemampuan generalisasi terbaik, terutama dalam menangani lompatan mendadak dan struktur temporal yang kompleks.

II.5.5.2 Model Hibrida ARIMA-LSTM

Model hibrida ARIMA-LSTM menggabungkan keunggulan komplementer dari kedua pendekatan, menyediakan kerangka kerja yang lebih kuat dan akurat untuk peramalan deret waktu (Alharbi 2025). ARIMA, dikenal karena efektivitasnya dalam mengidentifikasi pola linear dan tren, membentuk fondasi model dengan menangani elemen stasioner dan linear dari deret waktu.

Di sisi lain, LSTM dirancang khusus untuk menangani dependensi non-linear dan hubungan temporal jangka panjang dalam data. Dengan menggabungkan kedua pendekatan ini, model hibrida ARIMA-LSTM mendapatkan manfaat dari kemampuan pemodelan linear ARIMA dan pengenalan pola non-linear LSTM (Alharbi 2025). Integrasi ini memungkinkan model untuk menangkap dengan efektif baik fluktuasi jangka pendek maupun tren jangka panjang.

II.5.5.3 Perbandingan Karakteristik Model

Dubey, Kumar, dan Narang (2021) melakukan studi komparatif yang menunjukkan bahwa LSTM lebih baik dalam menangkap pola musiman dan tren jangka panjang dibandingkan dengan ARIMA dan SARIMA untuk data konsumsi energi. Tabel II.1 menunjukkan perbandingan karakteristik metode peramalan berdasarkan sintesis dari berbagai penelitian.

Tabel II.1 Perbandingan Metode Peramalan Deret Waktu

Metode	Kelebihan	Kekurangan	MAPE Tipikal
ARIMA	<ul style="list-style-type: none"> – <i>Interpretable</i> – Tidak memerlukan data besar – Cepat untuk dilatih 	<ul style="list-style-type: none"> – Sulit menangkap pola non-linear – Memerlukan data stasioner 	11–15%
SARIMA	<ul style="list-style-type: none"> – Menangani musiman – <i>Interpretable</i> – Baik untuk pola periodik 	<ul style="list-style-type: none"> – Memerlukan pengetahuan – Parameter banyak 	10–14%
LSTM	<ul style="list-style-type: none"> – Menangkap dependensi – Baik untuk pola kompleks – Tidak perlu stasioner 	<ul style="list-style-type: none"> – Memerlukan data besar – <i>Black box</i> – Lama untuk dilatih 	8–12%

II.6 Metrik Evaluasi

II.6.1 Metrik untuk Klasifikasi NLP

Dalam evaluasi model klasifikasi *intent* dalam NLP, metrik yang umum digunakan meliputi (Ali dkk. 2020; Perdana dkk. 2025):

II.6.1.1 Accuracy (Akurasi)

Proporsi prediksi yang benar dari total prediksi:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (\text{II.9})$$

II.6.1.2 Precision (Presisi)

Proporsi prediksi positif yang benar:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (\text{II.10})$$

II.6.1.3 Recall (Daya Ingat)

Proporsi kasus positif aktual yang diprediksi dengan benar:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (\text{II.11})$$

II.6.1.4 F1-Score

Rata-rata harmonik dari *precision* dan *recall*:

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (\text{II.12})$$

$TP = \text{True Positive}$, $TN = \text{True Negative}$, $FP = \text{False Positive}$, dan $FN = \text{False Negative}$.

II.6.2 Metrik untuk Time Series Forecasting

Dalam evaluasi model peramalan deret waktu, metrik yang umum digunakan meliputi (Sirisha dkk. 2022; Sunendar, Wibirama, dan Setiawan 2025):

II.6.2.1 Mean Absolute Error (MAE)

Rata-rata dari nilai absolut kesalahan:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (\text{II.13})$$

II.6.2.2 Root Mean Square Error (RMSE)

Akar kuadrat dari rata-rata kesalahan kuadrat:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (\text{II.14})$$

II.6.2.3 Mean Absolute Percentage Error (MAPE)

Rata-rata persentase kesalahan absolut:

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (\text{II.15})$$

y_i adalah nilai aktual, \hat{y}_i adalah nilai prediksi, dan n adalah jumlah observasi.

II.7 Integrasi Sistem dan Penelitian Terkait

II.7.1 Sistem Hibrida Rule-Based dan NLP

Sistem hibrida yang menggabungkan pendekatan *rule-based* dan NLP menawarkan keseimbangan antara kontrol deterministik dan fleksibilitas berbasis pembelajaran mesin. Pendekatan *rule-based* cocok untuk kueri terstruktur yang sering diulang, sementara NLP menangani kueri yang lebih kompleks dan bervariasi (Shah 2024).

Penelitian Shah (2024) menunjukkan bahwa optimisator berbasis AI dapat meningkatkan efisiensi pemanfaatan sumber daya dari 72% (optimasi berbasis aturan) menjadi 89%, dengan peningkatan efisiensi penulisan ulang kueri dari 15% menjadi 40%. Sistem berbasis AI juga meningkatkan akurasi deteksi anomali dari 60% menjadi 95% dan mengurangi biaya infrastruktur cloud sebesar 30%.

II.7.2 Customer Churn Prediction

Prediksi *customer churn* adalah aplikasi penting dari pembelajaran mesin dalam analisis bisnis pelanggan. Penelitian Praveen Lalwani dkk. (2022) menunjukkan bahwa *Random Forest*, SVM, dan XGBoost mampu mencapai akurasi tinggi dalam memprediksi *churn* pelanggan pada industri telekomunikasi, memungkinkan perusahaan bertindak lebih proaktif terhadap potensi pelanggan yang akan keluar.

Fitur-fitur yang umum digunakan untuk prediksi *churn* meliputi *tenure* (lama berlangganan), nilai transaksi rata-rata, frekuensi transaksi, riwayat keluhan atau masalah layanan, demografi pelanggan, dan pola penggunaan produk atau layanan (Praveen Lalwani dkk. 2022).

DAFTAR PUSTAKA

- Alharbi, Nada. 2025. "Prediction of Stock Market Using LSTM, ARIMA, and Hybrid ARIMA-LSTM Models: A Comparative Analysis". *Natural Sciences Publishing* 19 (1): 45–58. <https://doi.org/10.18576/amis/190105>.
- Ali, Mohd Hilmi Hasan, Masnizah Zolkepli, Nor Badrul Bin Anuar, dan Mohamad Zulhaimi Zainal. 2020. *Chatbot: A Conversational Agent Employed with Named Entity Recognition Model Using Recurrent Neural Network*. arXiv preprint. arXiv: 2007.04248 [cs.CL]. <https://arxiv.org/abs/2007.04248>.
- Annisa, Puput Triandini, dkk. 2024. "Kombinasi Intent Classification dan Named Entity Recognition dengan Pendekatan Multi-task Learning pada Dataset ATIS". *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)* 11 (5): 1055–1064. <https://doi.org/10.25126/jtiik.2024117985>.
- AtScale. 2025. "Why Semantic Layers? A Key Technical Architecture". Diakses pada October 28, 2025. <https://www.atscale.com/blog/semantic-layers-key-technical-architecture/>.
- Databricks. 2025. "The Role of Semantic Layers in Modern Data Analytics". Diakses pada October 28, 2025. <https://www.databricks.com/glossary/semantic-layer>.
- Dubey, Ashish Kumar, Abhishek Kumar, dan Sandeep Narang. 2021. "Study and Analysis of SARIMA and LSTM in Forecasting Time Series Data". *Sustainable Energy Technologies and Assessments* 47:101474. <https://doi.org/10.1016/j.seta.2021.101474>.
- East.vc. 2025. "The potential of AI to drive Indonesia's digital competitiveness growth". Diakses pada October 28, 2025. <https://east.vc/news/insights/the-potential-of-ai-to-drive-indonesias-digital-competitiveness-growth>.
- Finance, Beatrice, dan Georges Gardarin. 1994. "A Rule-Based Query Optimizer with Multiple Search Strategies". *Information Systems* 19 (5): 465–486. [https://doi.org/10.1016/0169-023X\(94\)90045-0](https://doi.org/10.1016/0169-023X(94)90045-0).
- Fullview. 2025. "80+ AI Customer Service Statistics & Trends in 2025". Diakses pada October 28, 2025. <https://www.fullview.io/blog/ai-customer-service-stats>.

- Ibadurrahman. 2025. "Implementasi Chatbot Berbasis Natural Language Processing dengan Model BERT untuk Meningkatkan Pelayanan". *Jurnal Informatika Teknologi dan Sains* 7 (3): 1521–1528. <https://doi.org/10.51401/jinteks.v7i3.8176>.
- Intimedia. 2024. "Indonesia accelerates digital transformation with AI adoption to achieve Golden Indonesia 2045". Diakses pada October 28, 2025. <https://intimedia.id/read/indonesia-accelerates-digital-transformation-with-ai-adoption-to-achieve-golden-indonesia-2045>.
- Iqbal, M., M. N. H. Siregar, dan R. Rismayanti. 2021. "Implementation of Natural Language Processing for Chatbots in Customer Service". *Journal of Computer Science and Informatics Engineering* 1 (1): 1–9. <https://doi.org/10.64803/jocsaic.v1i1.4>.
- Jumriyah dkk. 2025. "The Influence of Natural Language-Based Chatbot Usage on Customer Satisfaction and Service Efficiency". *Journal of Innovation in Learning and Technology* 2 (2): 78–89. <https://doi.org/10.61650/jiltech.v2i2.821>.
- Khaddam, Ibrahim, dan Mohammad Alzghoul. 2025. "Artificial Intelligence-Driven Business Intelligence: Systematic Literature Review". *International Journal of Energy Economics and Policy* 15 (1): 45–54. <https://doi.org/10.32479/ijeeep.19820>.
- Labs, dbt. 2024. "Understanding Semantic Layer Architecture". Diakses pada October 28, 2025. <https://www.getdbt.com/blog/semantic-layer-architecture>.
- Lalwani, P., M. K. Mishra, J. S. Chadha, dan P. Sethi. 2022. "Customer Churn Prediction System: A Machine Learning Approach". *Computing* 104 (8): 1–24. <https://doi.org/10.1007/s00607-021-00908-y>.
- Lalwani, Praveen, Manas Kumar Mishra, Jasroop Singh Chadha, dan Pratyush Sethi. 2022. "Customer Churn Prediction System: A Machine Learning Approach". *Computing* 104 (2): 271–294. <https://doi.org/10.1007/s00607-021-00908-y>.
- Mulyatun, Mulyatun, Dedi Kurniawan, dan Agus Hudaya. 2021. "Pendekatan Natural Language Processing pada Aplikasi Chatbot sebagai Alat Bantu Customer Service". *Jurnal Teknologi Sistem Informasi dan Aplikasi* 4 (1): 55–62. <https://doi.org/10.32493/jtsi.v4i1.7751>.

- Perdana, Rizki Maulana, dkk. 2025. "Multi-task Learning for Named Entity Recognition and Intent Classification in Task-Oriented Dialogue Systems". *Journal of Information Systems Engineering and Business Intelligence* 11 (1): 143–155. <https://doi.org/10.20473/jisebi.11.1.143-155>.
- Precedence Research. 2025. "Natural Language Processing Market Size to Hit USD 438.08 Bn by 2034". Diakses pada October 28, 2025. <https://www.precedenceresearch.com/natural-language-processing-market>.
- Qaltivate. 2025. "Conversational AI market summary: 2025 edition". Diakses pada October 28, 2025. <https://qaltivate.com/blog/conversational-ai-market/>.
- Rahayu, Ika Yuniar, Supriyono, dan Siva Prasatya. 2024. "Implementation of Pattern-Based Natural Language Processing for Chatbot Responses". *Jurnal Sistem Informasi dan Informatika (Systems and Information Journal)* 12 (2): 738–748. <https://doi.org/10.33480/josai.v12i2.6739>.
- Rizou, Spyridoula, Stelios Giannoulakis, dan Nicolas Tsapatsoulis. 2023. "Efficient Intent Classification and Entity Recognition for E-Commerce Queries Using Transformer-Based Models". *SN Computer Science* 4 (6): Article 728. <https://doi.org/10.1016/j.socs.2023.100728>.
- Shah, Nisarg. 2024. "AI-Driven Autonomous Database Management: Self-Tuning Query Optimization, Predictive Maintenance, and Intelligent Workload Distribution". *World Journal of Advanced Research and Reviews* 24 (1): 1519–1533. <https://doi.org/10.30574/wjarr.2024.24.1.3430>.
- . 2025. "Hybrid Analytics Architecture: Integrating Traditional BI with AI-Driven Insights". *World Journal of Advanced Engineering Technology and Sciences* 14 (02): 256–271. <https://doi.org/10.30574/wjaets.2025.14.2.0037>.
- Shinde, Amruta R., dan S. Ambika. 2025. "Applications of Natural Language Processing in Chatbot". *Journal of Information Systems and Engineering Management* 10 (2): 25936. <https://doi.org/10.55267/iadt.07.15936>.
- Siami-Namini, Sima, Neda Tavakoli, dan Akbar Siami Namin. 2018. "A Comparison of ARIMA and LSTM in Forecasting Time Series". Dalam *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 1394–1401. IEEE. <https://doi.org/10.1109/ICMLA.2018.00227>.

- Sirisha, U. M., dkk. 2022. "Profit Prediction Using ARIMA, SARIMA and LSTM Models in Time Series Forecasting: A Comparison". *International Journal of Current Research and Review (IJCRR)* 14 (24): 32–42. <https://doi.org/10.31782/IJCRR.2022.142410>.
- Sorour, Mohamed K. 2020. "The Role of Business Intelligence and Analytics in Higher Education Decision Making". Dalam *2020 IEEE Global Engineering Education Conference (EDUCON)*, 1849–1856. IEEE. <https://doi.org/10.1109/EDUCON45650.2020.9194157>.
- Straits Research. 2024. "Business Intelligence Market Size & Outlook, 2025–2033". Diakses pada October 28, 2025. <https://straitsresearch.com/report/business-intelligence-market>.
- Sunendar, Agus, Sunu Wibirama, dan Noor Akhmad Setiawan. 2025. "Comparison of ARIMA, LSTM, and GRU Models for Sales Prediction in Indonesian Banking Stock Market". *PILAR: Physics, Informatics, and Applied Research* 3 (1): 67–78. <https://doi.org/10.25273/pilar.v3i1.6412>.
- Vasist, P. Naga Sai, dan Ajay Sukumar Krishnan. 2021. "Intent Classification using BERT for Chatbot Application". Dalam *EAI International Conference on Multimedia, Signal Processing and Communication Technologies*, 1–10. <https://doi.org/10.4108/eai.7-12-2021.2314563>.
- WiseGuyReports. 2024. "Time Series Forecasting Market". Estimasi: USD 9.62 miliar (2023) → USD 36.9 miliar (2032), CAGR ~16.12%. Diakses pada October 28, 2025. <https://www.wiseguyreports.com/reports/time-series-forecasting-market>.