

**PERANCANGAN SISTEM *BUSINESS INTELLIGENCE*
BERBASIS *RULE-BASED CHATBOT* DENGAN *TIME SERIES*
FORECASTING DAN TEKNOLOGI BAHASA ALAMI
UNTUK ANALISIS DAN PREDIKSI KINERJA BISNIS**

Proposal Tugas Akhir

Oleh

**Thalita Zahra Sutejo
18222023**



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

LEMBAR PENGESAHAN

PERANCANGAN SISTEM *BUSINESS INTELLIGENCE* BERBASIS *RULE-BASED CHATBOT* DENGAN *TIME SERIES* *FORECASTING* DAN TEKNOLOGI BAHASA ALAMI UNTUK ANALISIS DAN PREDIKSI KINERJA BISNIS

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 3 Desember 2025

Pembimbing

Dr. Ir. Dimitri Mahayana, M.Eng.

NIP. 196808091991021001

DAFTAR ISI

DAFTAR GAMBAR	v
DAFTAR TABEL	vi
DAFTAR KODE	vii
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	6
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	11
II.1 <i>Business Intelligence</i> dan Arsitektur Sistem	11
II.1.1 Konsep <i>Business Intelligence</i>	11
II.1.2 Arsitektur <i>Business Intelligence</i> Hibrida	12
II.1.3 <i>Semantic Layer</i> dalam <i>Business Intelligence</i>	13
II.1.3.1 Komponen Inti <i>Semantic Layer</i>	13
II.2 <i>Rule-Based Query</i> dan Implementasinya pada <i>Chatbot</i>	14
II.2.1 Konsep <i>Rule-Based Query</i> dalam Sistem Basis Data	14
II.2.2 <i>Chatbot</i> Berbasis Aturan dan Pencocokan Pola	14
II.2.3 Integrasi <i>Rule-Based Query</i> dalam Arsitektur <i>Chatbot BI</i>	15
II.3 <i>Pattern-Based Intent Classification</i> dengan <i>Confidence Level</i>	16
II.3.1 Konsep <i>Intent Classification</i>	16
II.3.2 Klasifikasi <i>Intent</i> Berbasis <i>Transformer</i>	17
II.3.3 <i>Pattern Matching</i> untuk <i>Intent Classification</i>	18
II.3.4 <i>Confidence Score</i> dalam <i>Intent Classification</i>	19
II.3.4.1 <i>Confidence Threshold</i>	19
II.3.5 Evaluasi Performa <i>Intent Classification</i>	20

II.3.5.1	Metrik Evaluasi Klasifikasi <i>Intent</i> Multi-Kelas	20
II.3.5.2	<i>Confusion Matrix</i> dan <i>Confidence Score Analysis</i>	23
II.3.5.3	Target Performa <i>Intent Classification</i>	24
II.4	<i>Template-Based Natural Language Generation</i>	24
II.4.1	Konsep <i>Natural Language Generation</i>	25
II.4.2	<i>Template-Based Natural Language Generation</i>	25
II.4.3	<i>Template Rewriting</i> dengan <i>Pre-trained Language Models</i>	25
II.4.4	Keamanan dan Privasi Data dalam NLG Internal	26
II.4.4.1	<i>Differential Privacy</i> untuk NLG	26
II.4.4.2	Strategi Keamanan Data dalam NLG	27
II.4.4.3	Mitigasi Risiko dalam NLG Internal	27
II.4.5	Evaluasi Kualitas <i>Natural Language Generation</i>	27
II.4.5.1	Metrik Otomatis untuk Evaluasi NLG	28
II.4.5.2	Evaluasi Subjektif Berbasis Pengguna	29
II.4.5.3	Kriteria Performa NLG dalam Literatur	30
II.5	<i>Time Series Forecasting</i>	31
II.5.1	Konsep <i>Time Series</i> dan Peramalan	32
II.5.2	Model Peramalan Deret Waktu Berbasis <i>Deep Learning</i>	32
II.5.2.1	LSTM (<i>Long Short-Term Memory</i>) untuk Peramalan Deret Waktu	32
II.5.2.2	GRU (<i>Gated Recurrent Unit</i>) sebagai Alternatif LSTM	33
II.5.2.3	<i>Temporal Fusion Transformer</i> (TFT) untuk Data Multivariat	34
II.5.2.4	Model Hibrida dan <i>Ensemble</i>	35
II.5.2.5	Perbandingan Karakteristik Model Peramalan <i>Deep Learning</i>	35
II.5.2.6	Pemilihan Model dalam Konteks Penelitian	40
II.5.3	Evaluasi Performa <i>Time Series Forecasting</i>	40
II.5.3.1	Metrik Evaluasi <i>Forecasting</i>	40
II.5.3.2	<i>Skill Score</i> terhadap Garis Dasar	42
II.5.3.3	Pelaporan per Horizon dan per Entitas Prioritas	42
II.5.3.4	<i>Residual Analysis</i>	43
II.5.3.5	<i>Time Series Cross-Validation</i>	43
II.5.3.6	Target Performa <i>Forecasting</i>	44
II.6	Integrasi Sistem dan Penelitian Terkait	45
II.6.1	Arsitektur <i>Chatbot</i> Hibrida	45
II.6.1.1	Komponen Arsitektur <i>Chatbot</i> Modern	45
II.6.1.2	<i>Chatbot</i> Hibrida dengan AI dan <i>Human Oversight</i>	46
II.6.1.3	Pasar <i>Chatbot</i> AI Global	46
II.6.2	Integrasi <i>Pattern-Based Intent</i> dan <i>Template-Based NLG</i>	46

II.6.2.1	Alur Kerja Terintegrasi	47
II.6.2.2	Evaluasi Sistem Terintegrasi	47
II.6.3	<i>Customer Churn Prediction</i> dengan <i>Machine Learning</i>	48
II.6.3.1	Model <i>Machine Learning</i> untuk <i>Churn Prediction</i>	48
II.6.3.2	Evaluasi Model <i>Churn Prediction</i>	48
II.6.3.3	Fitur-fitur untuk <i>Churn Prediction</i>	49
II.6.3.4	<i>Real-Time Churn Prediction</i>	49
II.6.4	Integrasi <i>Churn Prediction</i> dengan <i>Chatbot BI</i>	50
III	ANALISIS MASALAH	51
III.1	Analisis Kondisi Saat Ini	51
III.2	Analisis Kebutuhan	54
III.2.1	Identifikasi Masalah Pengguna	54
III.2.2	Kebutuhan Fungsional	56
III.2.3	Kebutuhan Nonfungsional	60
III.3	Analisis Pemilihan Solusi	64
III.3.1	Alternatif Solusi	65
III.3.2	Analisis Penentuan Solusi	70
III.3.2.1	Solusi 1: <i>Chatbot</i> Berbasis Aturan Sederhana	70
III.3.2.2	Solusi 2: <i>Chatbot BI</i> dengan Klasifikasi Maksud Berbasis Pola, Pembangkitan Bahasa Alami Berbasis Templat, dan Peramalan Deret Waktu	72
III.3.2.3	Solusi 3: Dasbor BI Interaktif Tradisional dengan Peningkatan . . .	76
III.3.2.4	Matriks Perbandingan <i>Balanced Scorecard</i>	82
III.3.2.5	Kesimpulan Pemilihan Solusi	84

DAFTAR GAMBAR

I.1	Metodologi <i>CRISP-DM</i>	6
II.1	Bahasa Kueri untuk Penemuan Aturan dalam Basis Data	14
II.2	<i>Pattern-Based Intent Classification</i>	16
II.3	<i>Natural Language Generation</i>	24
II.4	<i>Time Series Forecasting</i>	31

DAFTAR TABEL

II.1	Perbandingan Karakteristik Model Peramalan Berbasis <i>Deep Learning</i>	36
III.1	Kebutuhan Fungsional Sistem <i>Chatbot</i> BI	57
III.2	Kebutuhan Nonfungsional Sistem <i>Chatbot</i> BI	60
III.3	Perbandingan Alternatif Solusi Teknologi	65
III.4	Analisis <i>Balanced Scorecard</i> untuk Aturan Sederhana	71
III.5	Analisis <i>Balanced Scorecard</i> untuk BI Komprehensif	73
III.6	Analisis <i>Balanced Scorecard</i> untuk BI Tradisional	77
III.7	Analisis <i>Balanced Scorecard</i> untuk LLM	79
III.8	Perbandingan <i>Balanced Scorecard</i> Alternatif Solusi	83

DAFTAR KODE

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 Desember 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 Ventures July 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, Ahmad, dan Khan 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% (WiseGuy Reports 2024).

Metode *forecasting* yang lazim meliputi model statistik dan model *deep learning* untuk menangkap pola temporal. Pada kasus prediksi *customer churn* di industri telekomunikasi, baik Random Forest maupun LSTM efektif dalam prediksi *churn*, tetapi kegunaannya bergantung pada sifat data dan batasan bisnis. LSTM lebih baik bila interaksi pelanggan historis dan tren kepuasan menjadi kunci dalam pengambilan keputusan (Mena dkk. 2023).

Integrasi tiga komponen utama, *Rule-Based Query*, *Natural Language Generation*, 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 generation*

Integrasi *rule-based query* dan pemrosesan bahasa alami (*natural language generation*) 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 oleh manajemen.

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

alami dengan (*natural language generation*), dan algoritma peramalan deret waktu sehingga seluruh pengguna internal perusahaan dapat menganalisis, memantau, dan memprediksi indikator bisnis pelanggan, seperti pesanan, target, *churn*, dan pendapatan 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 Generation/NLG*), dan algoritma peramalan deret waktu (*time series forecasting*) untuk memfasilitasi pengguna internal perusahaan dalam menganalisis dan memprediksi indikator bisnis pelanggan.

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

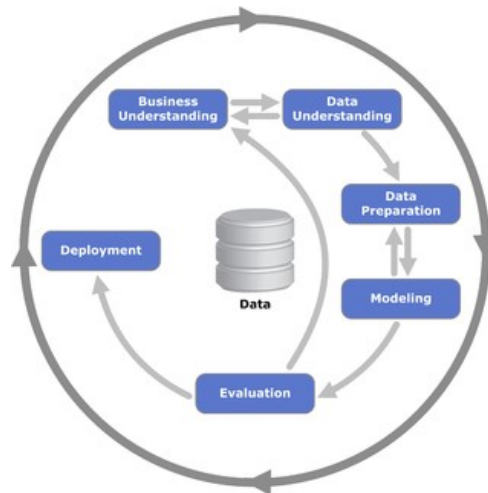
1. Merancang arsitektur sistem *Business Intelligence* berbasis *chatbot* yang mampu memproses permintaan analisis data dan memberikan *insight* berbasis data kepada pengguna internal perusahaan secara otomatis.
2. Mengimplementasikan modul pemrosesan bahasa alami sehingga pengguna internal dapat mengajukan pertanyaan dan permintaan analisis bisnis menggunakan bahasa alami melalui antarmuka *chatbot*.
3. Mengimplementasikan modul 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. Mengintegrasikan model yang menghasilkan bahasa tertulis atau lisan alami dari data terstruktur.
6. Merancang antarmuka percakapan yang dapat menampilkan percakapan.
7. 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. Sistem dirancang khusus untuk pengguna internal perusahaan (manajer, analis bisnis, dan staf operasional), bukan untuk pelanggan eksternal.
2. Sistem *chatbot* hanya mendukung interaksi dalam Bahasa Indonesia dan Bahasa Inggris.
3. Sistem menggunakan data yang terbatas pada data historis pelanggan yang mencakup informasi pesanan (*order*), target penjualan, tingkat *churn*, dan pendapatan dalam kurun waktu minimal dua tahun terakhir.
4. Sistem hanya menangani permintaan terkait analisis deskriptif, diagnostik, dan prediktif terhadap indikator bisnis pelanggan. Sistem tidak mencakup fungsi transaksional atau modifikasi data.
5. Sistem menggunakan metode peramalan deret waktu berbasis deep learning yang dibatasi pada algoritma LSTM, GRU, Temporal Fusion Transformer (TFT), atau kombinasi dari algoritma-algoritma tersebut.
6. Sistem dirancang untuk mengenali maksimal 20 hingga 30 *intent* utama yang berkaitan dengan analisis dan prediksi indikator bisnis pelanggan.
7. 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. Sistem diimplementasikan pada platform berbasis *web* dengan menggunakan *framework* yang umum untuk pengembangan *chatbot*.
9. Sistem dievaluasi melalui pengujian fungsional, pengukuran akurasi prediksi, analisis waktu respons, serta survei kepuasan pengguna dengan jumlah responden terbatas (minimal 20 hingga 30 pengguna internal).
10. Sistem mengimplementasikan 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



Gambar I.1 Metodologi *CRISP-DM*

Metodologi pelaksanaan tugas akhir ini mengadopsi kerangka kerja *CRISP-DM* (*Cross-Industry Standard Process for Data Mining*) untuk tahapan pengolahan data dan analitik. 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. Dalam penelitian ini, **CRISP-DM digunakan** untuk mengelola aspek penambangan data dan analitik prediktif (khususnya pengembangan model peramalan deret waktu dan NLP).

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 generation*, 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. **Pengimplementasian *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, 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 *mockup* untuk antarmuka *chatbot*.

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*.
- 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.

I.5.5 Pengujian dan Evaluasi

Tahapan ini bertujuan menilai kinerja sistem secara menyeluruh—mencakup aspek fungsionalitas, akurasi, performa, dan kebergunaan.

- a. **Pengujian unit**, yaitu setiap komponen diuji secara terpisah agar fungsinya terverifikasi dengan benar dan stabil.

- b. **Pengujian integrasi**, yaitu Interaksi antarmodul diverifikasi melalui skenario ujung-ke-ujung, dari masukan percakapan hingga keluaran berupa kueri yang dieksekusi, prediksi, narasi, dan visualisasi.
- c. **Pengujian fungsional**, yaitu Pemenuhan seluruh kebutuhan fungsional divalidasi menggunakan himpunan skenario representatif yang mencerminkan pertanyaan bisnis prioritas.
- d. **Evaluasi model NLP (klasifikasi intent multikelas, satu label per ujaran)**, yaitu hasil dilaporkan sebagai *top-1* dan *top-k* accuracy; *precision*, *recall*, dan *F1-score* per kelas yang kemudian dirata-ratakan dengan pendekatan *macro*, *micro*, dan *weighted*. Disajikan pula *confusion matrix* berukuran $K \times K$, evaluasi ambang kepercayaan untuk mekanisme *fallback*/UNKNOWN, serta ringkasan kalibrasi probabilitas.
- e. **Evaluasi model peramalan deret waktu**, yaitu protokol yang digunakan ialah *rolling-origin backtesting* pada beberapa horizon prediksi dengan *seasonal naive* sebagai garis dasar. Laporan mencakup MAE, RMSE, sMAPE, dan MASE, dilengkapi *skill score* terhadap garis dasar. Hasil disajikan per horizon dan per entitas prioritas (median dan rentang antar kuartil).
- f. **Evaluasi performa sistem**, yaitu waktu respons, *throughput*, dan latensi diukur pada beban realistis; *load testing* dilakukan untuk menilai skalabilitas.
- g. **Evaluasi kebergunaan (usability)**, yaitu *User Acceptance Testing* melibatkan sedikitnya 20–30 pengguna internal. Hasil dirangkum menggunakan skor *System Usability Scale* (SUS) dan umpan balik kualitatif terkait kemudahan, kejelasan, dan alur kerja.

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 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*, klasifikasi *intent* berbasis pola, pembangkitan bahasa alami berbasis templat, serta peramalan deret waktu yang mendukung analisis dan prediksi permintaan layanan pelanggan. Kajian literatur ini disusun secara sistematis untuk memberikan landasan teoretis yang kuat bagi pengembangan sistem yang diusulkan, dengan fokus pada integrasi *pattern-based intent classification*, *template-based natural language generation*, dan *time series forecasting* dalam satu platform terintegrasi.

II.1 *Business Intelligence* dan Arsitektur Sistem

Bagian ini membahas konsep dasar *Business Intelligence*, evolusinya, serta arsitektur sistem yang mendukung implementasi BI modern. Pembahasan mencakup komponen-komponen utama arsitektur BI dan peran *semantic layer* dalam menjembatani kesenjangan antara data teknis dan kebutuhan bisnis. Pemahaman yang komprehensif tentang arsitektur BI menjadi fondasi penting dalam merancang sistem yang mampu mengintegrasikan berbagai komponen teknologi secara efektif.

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).

Implementasi BI di institusi pendidikan tinggi telah menunjukkan kemampuan untuk memantau aktivitas jaminan kualitas dan mendukung pengambilan keputusan strategis (Sorour 2020). Penelitian Shah (2025b) 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. Pasar BI global terus menunjukkan pertumbuhan yang signifikan, dengan proyeksi mencapai USD 53,8 miliar pada tahun 2033, meningkat dari USD 25,9 miliar pada tahun 2024, dengan CAGR sebesar 8,48% (Straits Research 2024). Pertumbuhan ini menunjukkan relevansi dan kebutuhan akan solusi BI yang inovatif dalam menghadapi kompleksitas data bisnis modern.

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. Pendekatan ini menyajikan struktur yang memanfaatkan model pembelajaran mesin bersama pelaporan BI tradisional untuk menjembatani kesenjangan antara analisis historis dan wawasan berbasis data waktu nyata (Shah 2025b).

Komponen utama arsitektur BI hibrida mencakup beberapa lapisan yang saling terintegrasi (Shah 2025b). Lapisan sumber data mencakup sumber internal seperti sistem CRM dan perangkat lunak keuangan, serta sumber eksternal seperti data media sosial dan laporan pasar. Gudang data berfungsi sebagai fasilitas penyimpanan terpusat untuk data yang terorganisasi dan siap digunakan untuk pelaporan. Mesin analitik bertanggung jawab untuk pengenalan pola dan analisis tren, memberikan wawasan prediktif untuk perencanaan strategis. Lapisan visualisasi menyediakan dashboard interaktif dan laporan yang meningkatkan adopsi pengguna melalui presentasi data yang intuitif.

Implementasi arsitektur analitik hibrida menghasilkan peningkatan 27% dalam proses pengambilan keputusan dan peningkatan 31% dalam metrik kinerja organisasi secara keseluruhan (Shah 2025b). Integrasi komponen-komponen ini membentuk fondasi yang kuat untuk sistem BI modern yang responsif dan dapat diandalkan. Keberhasilan implementasi arsitektur hibrida ini membuka jalan bagi pengembangan sistem yang lebih kompleks dan terintegrasi dengan teknologi percakapan.

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 lokasi data atau bagaimana secara teknis terstruktur.

Lapisan semantik menyederhanakan konsep dan teknis data untuk pengguna bisnis sehingga mereka tidak perlu mengubah data bisnis yang mendasarinya untuk bekerja dengan cara baru (AtScale 2025). Lapisan ini memungkinkan pengguna bisnis untuk berinteraksi dengan data menggunakan terminologi yang mereka pahami tanpa perlu memahami struktur teknis basis data yang mendasarinya. Kemampuan ini sangat penting dalam konteks sistem *chatbot* BI, dengan memfasilitasi interpretasi kueri bahasa alami menjadi operasi basis data yang sesuai.

II.1.3.1 *Komponen Inti Semantic Layer*

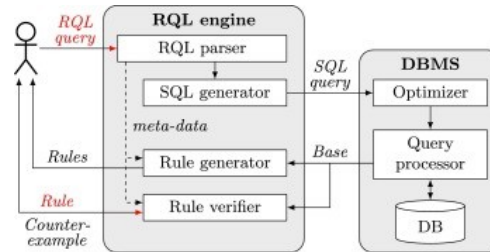
Lapisan semantik memiliki lima komponen inti yang bertindak sebagai fondasi struktural dan teknis (Labs 2024). Komponen pertama adalah definisi model semantik yang menciptakan representasi logis dari domain bisnis, memetakan struktur basis data teknis ke konsep bisnis. Alih-alih bekerja dengan tabel mentah seperti `usr_tbl` atau `trx_hist`, entitas seperti *Customer* atau *Order* didefinisikan untuk merangkum kompleksitas yang mendasarinya.

Komponen kedua adalah manajemen metadata yang menangani informasi tentang data, seperti deskripsi bidang, garis keturunan data (*data lineage*), frekuensi pembaruan, dan metrik kualitas. Lapisan logika bisnis sebagai komponen ketiga berisi aturan perhitungan spesifik untuk metrik bisnis. Komponen keempat adalah lapisan akses data yang menerjemahkan permintaan bisnis menjadi kueri basis data yang dioptimalkan, menerapkan filter keamanan yang diperlukan. Komponen terakhir adalah mekanisme *caching* yang memeriksa apakah hasil perhitungan sudah tersedia dalam tembolok sebelum menjalankan kueri.

Implementasi *semantic layer* yang efektif memungkinkan organisasi untuk mencapai konsistensi dalam pelaporan dan analisis, sekaligus mengurangi kompleksitas teknis yang dihadapi oleh pengguna bisnis. Pemahaman tentang *semantic layer* ini menjadi dasar penting dalam merancang sistem klasifikasi *intent* yang akan dibahas

pada bagian selanjutnya.

II.2 Rule-Based Query dan Implementasinya pada Chatbot



Gambar II.1 Bahasa Kueri untuk Penemuan Aturan dalam Basis Data

II.2.1 Konsep Rule-Based Query dalam Sistem Basis Data

Secara historis, optimisasi kueri pada sistem manajemen basis data relasional banyak mengandalkan pendekatan berbasis aturan (*rule-based*) dan berbasis biaya (*cost-based*) untuk memilih rencana eksekusi kueri yang efisien (Yang dkk. 2024; Shah 2025a). Pendekatan *rule-based* menggunakan seperangkat aturan heuristik yang ditetapkan pakar, seperti urutan penerapan seleksi dan proyeksi, penggunaan indeks, atau restrukturisasi *join*, untuk menyederhanakan dan mengoptimalkan kueri tanpa melakukan estimasi biaya numerik yang rinci (Yang dkk. 2024). Aturan-aturan tersebut, misalnya, dapat mengharuskan operator seleksi ditempatkan sedekat mungkin dengan sumber data atau menganjurkan penggunaan indeks ketika kondisi penyaringan sesuai, sehingga mengurangi jumlah baris yang harus diproses pada tahap-tahap berikutnya.

Meskipun perkembangan terkini banyak mengarah pada optimisasi kueri berbasis kecerdasan buatan, sejumlah kajian menegaskan bahwa teknik berbasis aturan masih relevan dan sering digunakan sebagai fondasi, terutama untuk pola kueri yang berulang dan terstruktur dengan baik (Shah 2025a; Lubis dkk. 2025). Dalam konteks sistem *Business Intelligence*, aturan-aturan kueri dapat dirancang untuk memetakan permintaan bisnis yang sering muncul (misalnya, “total penjualan per bulan per segmen pelanggan”) ke pola kueri yang terstandarisasi terhadap gudang data, sehingga mempercepat proses pengambilan data dan menjaga konsistensi hasil analisis.

II.2.2 Chatbot Berbasis Aturan dan Pencocokan Pola

Pada ranah *chatbot*, pendekatan berbasis aturan (*rule-based chatbot*) merupakan salah satu arsitektur paling awal dan masih banyak digunakan untuk skenario tugas-

terarah (*task-oriented*) dengan ruang lingkup percakapan yang terbatas (Singh dkk. 2025; Ouali dkk. 2024). *Chatbot* berbasis aturan umumnya bekerja dengan mencocokkan masukan pengguna terhadap pola-pola teks yang telah didefinisikan sebelumnya menggunakan teknik pencocokan pola (*pattern matching*), pencarian kata kunci, atau ekspresi reguler. Ketika suatu pola terpenuhi, sistem akan mengeksekusi aturan yang terkait, baik berupa pemilihan respons templat maupun pemanggilan prosedur tertentu di *backend*.

Kajian sistematis mengenai *chatbot* berbahasa Arab menunjukkan bahwa pendekatan *pre-scripted rules* memetakan ujaran pengguna ke aturan atau pola yang telah dipersiapkan, kemudian memilih respons dari himpunan jawaban yang sudah tersedia (Ouali dkk. 2024). Pendekatan ini juga digunakan pada berbagai *chatbot* klasik seperti ELIZA dan ALICE, serta pada bahasa skrip khusus seperti AIML (*Artificial Intelligence Markup Language*) yang memungkinkan pengembang mendefinisikan pola masukan dan templat keluaran secara eksplisit (Singh dkk. 2025). Keunggulan utama pendekatan berbasis aturan terletak pada kontrol penuh terhadap perilaku sistem, kemudahan untuk menjamin bahwa respons tidak keluar dari batas pengetahuan yang diizinkan, serta kesesuaian untuk domain yang sangat terstruktur.

Penelitian terbaru di Indonesia juga mengimplementasikan pendekatan pencocokan pola berbasis aturan untuk mengembangkan *chatbot* layanan informasi internal. Misalnya, pengembangan aplikasi Sido *chatbot* menggunakan pola aturan dan pencocokan pola untuk menjawab pertanyaan umum pengguna, disertai pengujian keberhasilan respons melalui uji fungsional dan kuesioner pengguna (Ananta dkk. 2025). Studi lain mengimplementasikan kerangka kerja Rasa untuk membangun sistem *FAQ bot* yang menggabungkan pemetaan intent dengan aturan dialog dan integrasi ke platform pesan seperti Telegram, menunjukkan bahwa arsitektur hibrida berbasis aturan dan pembelajaran mesin dapat memberikan fleksibilitas sekaligus menjaga struktur dialog yang terkontrol (Talenggoran dkk. 2025).

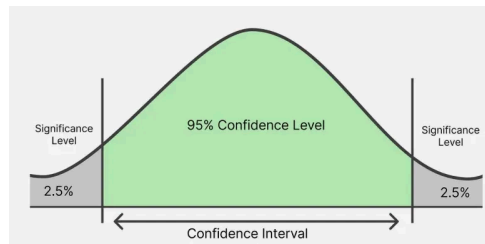
II.2.3 Integrasi *Rule-Based Query* dalam Arsitektur Chatbot BI

Dalam konteks *Business Intelligence*, beberapa kajian mengusulkan penggunaan *chatbot* sebagai antarmuka percakapan untuk eksplorasi data dan kolaborasi analitik, di mana komponen percakapan (*conversational agent*) bekerja berdampingan dengan agen eksplorasi data dan agen rekomendasi (Cherednichenko dkk. 2023). Pendekatan ini sering kali memanfaatkan arsitektur berbasis aturan untuk menerjemahkan permintaan dalam bahasa alami menjadi perintah atau kueri terstruktur terhadap sistem BI dan gudang data. Aturan-aturan tersebut dapat mencakup pe-

metaan intent tertentu ke templat kueri, pengisian parameter kueri dari entitas yang diekstraksi, serta penentuan agregasi dan dimensi pelaporan berdasarkan frasa waktu atau hierarki bisnis yang disebutkan pengguna.

Chatbot berbasis aturan memerlukan basis pengetahuan yang menyimpan aturan-aturan yang dirancang secara manual untuk menjamin kekokohan dan konsistensi perilaku sistem (Singh dkk. 2025). Dalam kerangka *chatbot* BI, basis pengetahuan ini tidak hanya berisi pasangan pola–jawaban, tetapi juga aturan pemetaan ke *rule-based query engine* yang mengakses gudang data. Studi-studi tersebut mendukung pendekatan yang digunakan dalam penelitian ini, yaitu menggabungkan klasifikasi intent berbasis pola dengan *rule-based query* yang terdokumentasi dengan baik, sehingga sistem dapat menjawab pertanyaan bisnis terstruktur secara andal dan transparan, sambil tetap membuka ruang pengembangan lebih lanjut untuk integrasi teknik pembelajaran mesin yang lebih canggih jika cakupan intent dan kompleksitas kebutuhan analitik meningkat.

II.3 *Pattern-Based Intent Classification dengan Confidence Level*



Gambar II.2 *Pattern-Based Intent Classification*

Bagian ini membahas pendekatan klasifikasi *intent* berbasis pola untuk sistem *chatbot*, dengan fokus pada mekanisme pencocokan pola, penggunaan *confidence score*, dan evaluasi performa. Klasifikasi *intent* merupakan komponen kritis dalam memahami maksud pengguna dari kueri bahasa alami dan menentukan respons yang sesuai berdasarkan data internal perusahaan.

II.3.1 Konsep *Intent Classification*

Klasifikasi *intent* adalah proses menentukan niat atau tujuan pengguna dari masukan mereka, yang merupakan tugas fundamental dalam sistem *chatbot* percakapan (Ouaddi, Benabbou, dan Sael 2025). Inti dari *chatbot* berbasis AI adalah komponen NLU yang mengklasifikasikan *intent* pengguna untuk menghasilkan respons yang

sesuai. Tugas klasifikasi ini sangat penting karena menentukan alur percakapan dan tindakan yang akan diambil sistem.

Pasar *Conversational AI* global diproyeksikan mencapai USD 49,9 miliar pada tahun 2030, meningkat dari USD 10,7 miliar pada tahun 2023, dengan CAGR sebesar 25,2% (Qaltivate 2025). Pertumbuhan eksponensial ini didorong oleh meningkatnya adopsi teknologi AI dalam layanan pelanggan dan operasional bisnis. *Chatbot* dapat menjawab hampir 80% dari semua pertanyaan standar, yang secara signifikan mengurangi beban kerja agen manusia dan mempercepat waktu respons (Fullview 2025). Statistik ini menunjukkan potensi besar klasifikasi *intent* dalam meningkatkan efisiensi operasional dan kepuasan pengguna.

II.3.2 Klasifikasi *Intent* Berbasis *Transformer*

Klasifikasi *intent* berbasis *Transformer* merupakan pendekatan modern yang memanfaatkan arsitektur jaringan saraf dengan mekanisme perhatian diri (*self-attention*) untuk memahami konteks ujaran pengguna secara dua arah dan menyeluruh (Merugudkk. 2024). Model-model seperti BERT (*Bidirectional Encoder Representations from Transformers*), RoBERTa, IndoBERT, dan berbagai *sentence transformer* telah menunjukkan kinerja yang unggul pada tugas klasifikasi teks, termasuk klasifikasi *intent* dalam sistem percakapan (Ozerova 2022; Gehweiler 2024). Pendekatan ini merepresentasikan setiap kalimat sebagai vektor berdimensi tinggi yang memuat informasi semantik kaya, kemudian menambahkan lapisan klasifikasi di bagian akhir untuk memetakan representasi tersebut ke label *intent* yang telah didefinisikan.

Berbagai penelitian melaporkan bahwa model berbasis *Transformer* mampu melampaui pendekatan tradisional seperti LSTM, CNN, maupun metode berbasis fitur manual dalam hal akurasi, skor F1, dan kemampuan menangkap nuansa konteks (Gehweiler 2024). Studi yang membandingkan BERT, RoBERTa, dan IndoBERT untuk klasifikasi *intent chatbot* berbahasa Indonesia menunjukkan bahwa IndoBERT dapat mencapai akurasi hingga sekitar 0,94 berkat proses prapelatihan yang selaras dengan struktur bahasa Indonesia dan korpus yang digunakan. Penelitian lain yang memanfaatkan MiniLM dan RoBERTa untuk modul deteksi *intent* dalam asisten percakapan melaporkan peningkatan akurasi 10–15 persen dibandingkan solusi sebelumnya, sekaligus menekankan pentingnya memperhatikan kecepatan inferensi dan kebutuhan perangkat keras saat model *Transformer* diintegrasikan ke dalam sistem produksi (Ozerova 2022).

Perkembangan model bahasa besar berbasis *Transformer* juga mendorong kajian

terbaru yang mengevaluasi penggunaan *sentence transformer*, model BERT terkontrasif, dan bahkan model bahasa besar generatif untuk tugas deteksi *intent* pada agen percakapan tugas-orientasi (Merugu dkk. 2024; Gehweiler dkk. 2025). Hasil kajian tersebut menyoroti adanya kompromi yang jelas antara kualitas prediksi, kemampuan menangani *intent* di luar cakupan (*out-of-scope*), latensi sistem, serta konsumsi sumber daya komputasi, sehingga pemilihan arsitektur perlu mempertimbangkan kebutuhan dan batasan operasional sistem yang dibangun (Merugu dkk. 2024).

Dalam konteks penelitian ini, klasifikasi *intent* berbasis *Transformer* diposisikan sebagai pendekatan yang relevan dan dikaji secara teoretis, tetapi tidak diimplementasikan secara langsung pada prototipe sistem. Pertimbangan utama adalah ruang lingkup pertanyaan yang relatif terbatas pada kebutuhan internal, pola kueri yang cenderung berulang, serta prioritas untuk menjaga efisiensi pemrosesan dan menghemat penggunaan sumber daya komputasi berbiaya tinggi seperti unit pemroses grafis (GPU) (Ozerova 2022; Gehweiler 2024). Oleh karena itu, penelitian ini memilih pendekatan klasifikasi *intent* berbasis pola dengan mekanisme nilai keyakinan (*pattern-based intent classification with confidence scoring*) yang lebih ringan secara komputasi, sambil menjadikan pendekatan berbasis *Transformer* sebagai acuan pengembangan lanjutan ketika cakupan *intent* dan beban sistem meningkat di masa mendatang.

II.3.3 *Pattern Matching untuk Intent Classification*

Pendekatan berbasis pola (*pattern matching*) dalam klasifikasi *intent* menggunakan aturan dan templat yang telah ditentukan sebelumnya untuk mencocokkan masukan pengguna dengan kategori *intent* yang sesuai. Sistem ini menyimpan pola-pola pertanyaan atau *utterances* untuk setiap *intent*, kemudian membandingkan *input* pengguna dengan pola-pola yang telah didefinisikan untuk menemukan kecocokan terbaik.

Pendekatan *pattern matching* efektif untuk menangani kueri terstruktur dan berulang dalam domain spesifik. Implementasi dapat dilakukan menggunakan ekspresi reguler, pencocokan *n-gram*, atau algoritma kesamaan string seperti *Fuzzy String Matching* (Mulyatun, Kurniawan, dan Hudaya 2021). Keuntungan utama pendekatan ini adalah kecepatan eksekusi yang tinggi, prediktabilitas hasil, dan kontrol penuh atas logika pencocokan, menjadikannya pilihan yang tepat untuk sistem yang memerlukan respons cepat dan deterministic based pada data internal perusahaan.

II.3.4 *Confidence Score* dalam *Intent Classification*

Confidence score atau skor keyakinan adalah nilai numerik antara 0 dan 1 yang menunjukkan tingkat kepercayaan sistem terhadap prediksi klasifikasi *intent* (Khosla dkk. 2022). Skor ini dihitung berdasarkan tingkat kecocokan antara *input* pengguna dengan pola yang telah didefinisikan. Semakin tinggi kecocokan, semakin tinggi nilai *confidence score*, dan semakin besar kemungkinan respons yang diberikan adalah benar.

Sistem menghitung *confidence score* dengan membandingkan *input* pengguna terhadap semua pola *utterances* atau frasa pelatihan yang tersedia. Sistem kemudian memilih *intent* dengan skor tertinggi sebagai prediksi. Implementasi *confidence score* memungkinkan sistem untuk mengidentifikasi kapan prediksi mungkin tidak akurat dan memerlukan mekanisme *fallback* (Kuligowska dan Kowalczyk 2024).

II.3.4.1 *Confidence Threshold*

Confidence threshold adalah nilai ambang batas minimum yang harus dicapai oleh *confidence score* agar sistem memicu respons *intent* tertentu. Nilai threshold adalah angka numerik antara 0 dan 1, dengan nilai default yang umum digunakan adalah 0,45 atau 45% (Teneo Developers 2022). Sistem yang lebih konservatif dapat menggunakan threshold lebih tinggi seperti 0,70 atau 70% untuk mengurangi risiko respons yang salah.

Jika *confidence score* terbaik jatuh di bawah threshold, sistem akan memicu interaksi *fallback*, biasanya berupa pesan yang meminta klarifikasi atau memberitahu pengguna bahwa sistem tidak memahami pertanyaan. Pengaturan threshold yang tepat sangat penting untuk menyeimbangkan antara tingkat otomasi dan akurasi respons (Cyara 2024).

Analisis threshold membantu menentukan pengaturan yang optimal berdasarkan *Key Performance Indicators* (KPI) bisnis. Organisasi dapat menyesuaikan threshold berdasarkan tiga pendekatan utama. Pendekatan pertama adalah menyesuaikan threshold untuk mencapai persentase target respons yang benar. Pendekatan kedua adalah membatasi persentase maksimal respons yang salah. Pendekatan ketiga adalah mengoptimalkan trade-off antara tingkat otomasi dan risiko kesalahan berdasarkan toleransi bisnis (Cyara 2024; Genesys 2023).

II.3.5 Evaluasi Performa *Intent Classification*

Evaluasi performa klasifikasi *intent* dilakukan dengan mengukur beberapa metrik kunci yang mencerminkan akurasi dan keandalan sistem dalam memprediksi niat pengguna. Evaluasi yang komprehensif memastikan bahwa sistem dapat beroperasi sesuai dengan standar yang ditetapkan dan memberikan pengalaman pengguna yang optimal.

II.3.5.1 Metrik Evaluasi Klasifikasi *Intent* Multi-Kelas

Model klasifikasi *intent* dalam penelitian ini dievaluasi menggunakan metrik standar yang disesuaikan untuk masalah klasifikasi multi-kelas (38 kelas *intent* yang saling eksklusif), bukan multi-label. Pendekatan evaluasi harus mengakomodasi bahwa setiap kueri pengguna hanya dapat diklasifikasikan ke dalam satu kelas yang benar, sehingga metrik klasifikasi biner harus diperluas menggunakan strategi *one-vs-all* dan agregasi yang tepat (Grandini, Bagli, dan Visani 2020; Evidently AI 2025).

Perbedaan utama antara klasifikasi multi-kelas dan biner adalah pada interpretasi elemen matriks kebingungan (*confusion matrix*) dan agregasi metrik per-kelas. Dalam 38 kelas, matriks kebingungan berukuran 38×38 , dengan setiap baris merepresentasikan kelas hasil prediksi dan setiap kolom merepresentasikan kelas sebenarnya. Prediksi yang benar terletak pada diagonal utama, sedangkan prediksi salah tersebar pada sel off-diagonal. Pola sebaran kesalahan pada matriks kebingungan multi-kelas memberikan wawasan berharga tentang persamaan atau kebingungan antar kelas, misalnya prediksi kelas “Penjualan-Bulanan” yang sering disalahklasifikasikan sebagai “Penjualan-Triwulanan” menunjukkan kesamaan semantik yang perlu ditangani dengan aturan atau templat yang lebih jelas (Evidently AI 2025).

Dalam menghitung metrik per kelas- i , masalah klasifikasi multi-kelas diubah menjadi masalah biner secara virtual melalui pendekatan *one-vs-all* (Grandini, Bagli, dan Visani 2020):

- a. **TP _{i}** (*True Positive*): Jumlah sampel kelas- i yang diprediksi benar sebagai kelas- i (elemen diagonal matriks kebingungan pada baris dan kolom i)
- b. **FP _{i}** (*False Positive*): Jumlah sampel kelas selain i yang diprediksi sebagai kelas- i (jumlah seluruh baris i pada matriks kebingungan kecuali elemen diagonal)
- c. **FN _{i}** (*False Negative*): Jumlah sampel kelas- i yang diprediksi salah sebagai kelas lain (jumlah seluruh kolom i pada matriks kebingungan kecuali elemen diagonal)

Dalam konteks Agregasi Metrik (*Macro-Average* dengan *Micro-Average*) dan adanya 38 kelas dalam sistem ini, diperlukan strategi agregasi untuk memperoleh metrik keseluruhan performa sistem. Dua pendekatan utama yang umum digunakan dalam klasifikasi multi-kelas adalah *macro-averaging* dan *micro-averaging* (Evidently AI 2025)

Pendekatan ***Micro-Averaging*** ini menghitung metrik (*Precision*, *Recall*, *F1-Score*) untuk setiap kelas secara individual, kemudian mengambil rata-rata aritmatik dari 38 nilai metrik tersebut. Strategi ini memberikan bobot yang sama untuk setiap kelas, sehingga mengutamakan performa pada kelas minoritas dan mencegah kelas mayoritas mendominasi hasil evaluasi (Grandini, Bagli, dan Visani 2020).

$$\text{Presisi}_{\text{makro}} = \frac{1}{38} \sum_{i=1}^{38} \text{Presisi}_i \quad (\text{II.1})$$

$$\text{Daya Ingat}_{\text{makro}} = \frac{1}{38} \sum_{i=1}^{38} \text{Daya Ingat}_i \quad (\text{II.2})$$

$$\text{Skor-F1}_{\text{makro}} = \frac{1}{38} \sum_{i=1}^{38} \text{Skor-F1}_i \quad (\text{II.3})$$

Pendekatan ***Micro-Averaging*** menjumlahkan seluruh nilai TP, FP, dan FN dari semua kelas terlebih dahulu, kemudian menghitung metrik secara global berdasarkan jumlahan tersebut. Strategi ini memberikan bobot yang sama untuk setiap sampel, sehingga metrik cenderung didominasi oleh kelas mayoritas dan mengabaikan performa pada kelas minoritas. Dalam klasifikasi multi-kelas, mikro-rata presisi = mikro-rata daya ingat = akurasi (Evidently AI 2025).

$$\text{Presisi}_{\text{mikro}} = \frac{\sum_{i=1}^{38} \text{TP}_i}{\sum_{i=1}^{38} (\text{TP}_i + \text{FP}_i)} \quad (\text{II.4})$$

$$\text{Daya Ingat}_{\text{mikro}} = \frac{\sum_{i=1}^{38} \text{TP}_i}{\sum_{i=1}^{38} (\text{TP}_i + \text{FN}_i)} \quad (\text{II.5})$$

Sebagai pendekatan komplementer, rata-rata tertimbang menghitung metrik per-kelas kemudian mengambil rata-rata yang diboboti oleh jumlah sampel di setiap kelas atau ***Weighted-Averaging***. Ini mengakomodasi keseimbangan antara kepedu-

lian terhadap kelas minoritas dan representasi data yang ada (Grandini, Bagli, dan Visani 2020).

Sistem menggunakan mekanisme pemberian skor keyakinan (*confidence scoring*) dalam rentang 0 hingga 1 untuk setiap prediksi klasifikasi intent. Tingkat keyakinan tertinggi menunjukkan kelas yang diprediksi oleh sistem. Dalam menentukan apakah prediksi dapat diterima atau perlu klarifikasi, ditetapkan *threshold* pada nilai 0,70 (Santosa, Nusantara, dan Imron 2025):

- a. Jika tingkat keyakinan $\geq 0,70$: Prediksi dianggap “positif” dan diterima untuk kelas yang diprediksi; sistem memberikan respons berdasarkan *intent* yang teridentifikasi
- b. Jika tingkat keyakinan $< 0,70$: Prediksi dianggap “tidak cukup yakin”; sistem meminta klarifikasi kepada pengguna atau menawarkan alternatif *intent* yang mungkin

Dengan demikian, untuk setiap kelas-*i*, perhitungan TP, FP, dan FN disesuaikan sebagai berikut:

- a. **TP_i**: Prediksi kelas-*i* dengan tingkat keyakinan $\geq 0,70$ yang ternyata benar
- b. **FP_i**: Prediksi kelas-*i* dengan tingkat keyakinan $\geq 0,70$ yang ternyata salah
- c. **FN_i**: Sampel kelas-*i* yang sebenarnya tidak terprediksi dengan tingkat keyakinan $\geq 0,70$ (termasuk prediksi kelas lain atau tingkat keyakinan rendah yang ditolak sistem)

Visualisasi matriks kebingungan untuk 38 kelas memberikan informasi penting tentang pola kesalahan klasifikasi (Evidently AI 2025), yaitu diagonal utama menampilkan jumlah prediksi benar per kelas, sementara intensitas warna pada sel off-diagonal menunjukkan kelas-kelas mana yang sering dikacaukan satu sama lain. Sebagai contoh, jika ditemukan frekuensi tinggi kesalahan pada sel (Penjualan-Bulanan, Penjualan-Triwulanan), hal ini menunjukkan bahwa dua kelas tersebut memiliki kesamaan semantik yang tinggi dan memerlukan aturan diferensiasi yang lebih spesifik atau contoh pelatihan yang lebih banyak.

Setiap kelas *intent* memiliki kepentingan strategis yang berbeda dalam konteks bisnis sehingga penelitian ini menggunakan kombinasi metrik berikut (Grandini, Bagli, dan Visani 2020):

- a. **Skor-F1 Makro** sebagai metrik utama: Mengutamakan performa yang merata di semua 38 kelas, memastikan bahwa kelas-kelas minoritas tidak terabaikan
- b. **Skor-F1 Tertimbang** sebagai metrik sekunder: Mempertimbangkan distribu-

si frekuensi kelas dalam data percobaan, mencerminkan performa yang lebih realistis terhadap data produksi

- c. **Skor-F1 Mikro** sebagai metrik verifikasi: Harus sama dengan akurasi keseluruhan; digunakan untuk memvalidasi perhitungan
- d. **Matriks Kebingungan**: Untuk analisis kualitatif pola kesalahan dan identifikasi peningkatan yang diperlukan pada aturan atau templat tertentu

Berdasarkan studi industri dan best-practices untuk sistem klasifikasi *intent* dalam *chatbot* bisnis, target performa yang ditetapkan untuk penelitian ini adalah (Santosa, Nusantara, dan Imron 2025; Ozerova 2022):

- a. **Skor-F1 Makro** $\geq 0,85$: Menjamin performa yang baik dan seimbang di seluruh kelas, termasuk kelas-kelas dengan sampel pelatihan yang lebih sedikit
- b. **Skor-F1 Tertimbang** $\geq 0,90$: Menjamin performa yang baik pada kelas-kelas mayoritas yang mewakili mayoritas pertanyaan pengguna dalam praktik
- c. **Skor-F1 Mikro** (Akurasi Keseluruhan) $\geq 0,88$: Target akurasi minimal untuk sistem dapat diterima dalam produksi
- d. **Matriks Kebingungan**: Diharapkan menunjukkan lebih dari 80% nilai prediksi berada pada diagonal utama, mengindikasikan bahwa mayoritas prediksi benar

Pencapaian target-target ini akan diverifikasi melalui pengujian pada data validasi yang terpisah dan independen dari data pelatihan model, memastikan bahwa hasil evaluasi tidak bias dan representatif terhadap performa sistem pada data baru di lingkungan produksi.

II.3.5.2 *Confusion Matrix dan Confidence Score Analysis*

Confusion matrix menyediakan visualisasi detail tentang performa klasifikasi untuk setiap kelas *intent*. Analisis *confusion matrix* memungkinkan identifikasi *intent* yang sering salah diklasifikasikan dan pola kesalahan sistematis yang dapat diperbaiki melalui penyesuaian pola atau threshold.

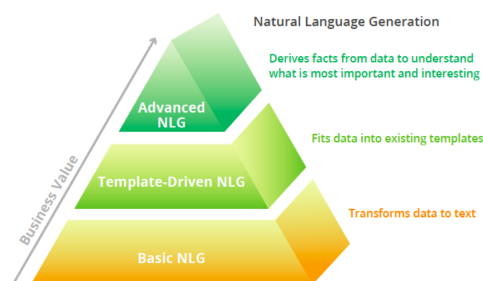
Analisis distribusi *confidence score* penting untuk memahami bagaimana sistem membedakan antara *intent* yang benar (in-domain) dan pertanyaan di luar domain (out-of-domain atau OOD). Sistem yang baik harus menghasilkan *confidence score* tinggi untuk *intent* in-domain dan score rendah untuk OOD (Khosla dkk. 2022; Yutao Zhang dkk. 2022). Masalah *overconfidence*, dengan sistem menghasilkan score tinggi bahkan untuk sampel OOD yang abnormal, perlu dimitigasi melalui kalibrasi model yang tepat.

II.3.5.3 Target Performa *Intent Classification*

Berdasarkan literatur dan praktik industri terkini, target performa yang ditetapkan untuk sistem klasifikasi *intent* berbasis pola mencakup beberapa metrik kunci (Spot Intelligence 2023; Nabiilah, Adiwijaya, dan Maharani 2025). *Accuracy* harus mencapai minimal 85%, yang berarti minimal 85% dari semua prediksi *intent* harus benar. *F1-Score* ditargetkan minimal 0,80 untuk memastikan keseimbangan antara presisi dan *recall*. *Precision* ditetapkan minimal 0,82, yang berarti minimal 82% dari *intent* yang diprediksi positif benar-benar positif. *Recall* ditargetkan minimal 0,78, yang berarti minimal 78% dari *intent* positif aktual berhasil terdeteksi.

Metrik tambahan terkait *confidence score* meliputi *Average Confidence Score* untuk prediksi benar minimal 0,75, dan *Coverage Rate* atau persentase kueri yang mencapai threshold minimal 80%. *Response Time* harus di bawah 100 milidetik untuk 95% kueri agar memberikan pengalaman pengguna yang responsif. Pencapaian target-target ini memastikan bahwa sistem klasifikasi *intent* dapat memberikan pengalaman pengguna yang akurat dan responsif, dengan tingkat kesalahan yang minimal dan kemampuan untuk menangani variasi pertanyaan pengguna yang luas.

II.4 *Template-Based Natural Language Generation*



Gambar II.3 *Natural Language Generation*

Bagian ini membahas pendekatan pembangkitan bahasa alami berbasis templat untuk menghasilkan variasi respons dari data internal perusahaan. Pembahasan mencakup konsep NLG berbasis templat, keamanan data dalam implementasi NLG internal, dan evaluasi kualitas teks yang dihasilkan. Pendekatan ini memastikan bahwa respons yang dihasilkan bervariasi dan natural sambil tetap menjaga keamanan dan privasi data perusahaan.

II.4.1 Konsep *Natural Language Generation*

Natural Language Generation (NLG) adalah proses menghasilkan teks bahasa alami dari representasi data terstruktur. Dalam konteks *chatbot* Business Intelligence, NLG digunakan untuk mengubah hasil kueri basis data menjadi respons bahasa natural yang mudah dipahami pengguna (Kale dan Rastogi 2020). NLG memungkinkan sistem untuk menyajikan informasi numerik dan data terstruktur dalam format naratif yang lebih mudah dicerna.

Pasar NLP global diproyeksikan tumbuh dari USD 35,43 miliar pada tahun 2024 menjadi USD 438,08 miliar pada tahun 2034, dengan CAGR sebesar 28,6% (Precedence Research 2025). Pertumbuhan eksponensial ini didorong oleh meningkatnya adopsi teknologi AI, *cloud computing*, dan kebutuhan akan analisis data tidak terstruktur di berbagai sektor industri. Pertumbuhan pasar ini menunjukkan relevansi dan urgensi pengembangan sistem NLG untuk mendukung kebutuhan bisnis modern.

II.4.2 *Template-Based Natural Language Generation*

Sistem NLG berbasis templat memetakan *input* non-linguistik secara langsung ke struktur linguistik yang berisi "celah" atau *placeholders* yang diisi selama *output*. Pendekatan ini menggunakan templat yang telah ditentukan sebelumnya dengan slot yang dapat diisi dengan data aktual dari hasil kueri basis data.

Contoh implementasi templat sederhana mencakup struktur seperti "Penjualan pada [bulan] mencapai [nilai], meningkat [persentase]% dari bulan sebelumnya." Sistem kemudian mengisi placeholder dengan data aktual: "Penjualan pada Oktober 2025 mencapai Rp 5,2 miliar, meningkat 15% dari bulan sebelumnya." (Kale dan Rastogi 2020).

Keuntungan utama pendekatan berbasis templat adalah kontrol penuh atas output, kecepatan generasi yang tinggi, tidak memerlukan *dataset* pelatihan besar, dan kemudahan maintenance dan modifikasi. Sistem dapat menghasilkan variasi output dengan mendefinisikan multiple templat untuk skenario yang sama, menghindari respons yang monoton (Kapoor dan Shetty 2023).

II.4.3 *Template Rewriting dengan Pre-trained Language Models*

Pendekatan lanjutan dalam NLG berbasis templat melibatkan penggunaan model bahasa pre-trained untuk menulis ulang templat sederhana menjadi teks yang lebih

koheren dan natural. Metode ini menggabungkan keuntungan templat dengan fleksibilitas model bahasa neural (Kale dan Rastogi 2020; Rebuffel dkk. 2020).

Arsitektur sistem mencakup tiga tahapan utama. Tahap pertama adalah modul kebijakan menghasilkan sekumpulan tindakan berdasarkan konteks. Tahap kedua adalah templat sederhana mengkonversi setiap tindakan menjadi *utterance* bahasa alami. Tahap ketiga adalah model *encoder-decoder* seperti T5 menulis ulang gabungan *utterances* menjadi respons percakapan yang koheren (Kale dan Rastogi 2020).

Pendekatan ini memungkinkan sistem untuk menghasilkan respons yang secara semantik benar dari templat, kemudian model bahasa memperbaiki koherensi dan kealamian teks. Metode *template rewriting* telah menunjukkan peningkatan sample efficiency yang signifikan dibandingkan metode end-to-end, dengan jumlah templat yang tumbuh linear terhadap jumlah slot dibandingkan pertumbuhan kombinatorial (Rebuffel dkk. 2020).

II.4.4 Keamanan dan Privasi Data dalam NLG Internal

Implementasi NLG dalam lingkungan perusahaan memerlukan pertimbangan keamanan yang ketat untuk melindungi data sensitif dan mencegah kebocoran informasi. Sistem NLG internal harus dirancang dengan prinsip *privacy-by-design* untuk memastikan data perusahaan tetap aman.

II.4.4.1 Differential Privacy untuk NLG

Differential Privacy (DP) adalah kerangka kerja yang memberikan jaminan privasi secara matematis, memastikan bahwa kontribusi individu dalam *dataset* dibatasi (Feyisetan dkk. 2022; Mireshghallah dkk. 2022). Dalam konteks NLG, DP dapat diterapkan melalui dua pendekatan utama. Pendekatan pertama adalah pelatihan model dengan DP-SGD yang menambahkan noise pada gradien selama pelatihan. Pendekatan kedua adalah perturbasi representasi teks yang menambahkan noise pada *embedding* sebelum generasi.

Parameter privasi ϵ mengontrol tingkat privasi yang diberikan, dengan nilai ϵ yang lebih kecil memberikan privasi yang lebih kuat tetapi dapat mengurangi utility model. Implementasi DP dalam NLG memerlukan trade-off yang cermat antara privasi dan kualitas output (Luo dan Liu 2024).

II.4.4.2 Strategi Keamanan Data dalam NLG

Implementasi NLG internal yang aman memerlukan beberapa strategi keamanan berlapis (IT Convergence 2024). Strategi pertama adalah enkripsi data dengan mengenkripsi data saat transit menggunakan TLS/SSL dan saat istirahat untuk melindungi data sensitif. Strategi kedua adalah kontrol akses yang ketat dengan menerapkan autentikasi dan otorisasi yang robust, menggunakan prinsip *least privilege*. Strategi ketiga adalah data minimization dengan hanya menggunakan data minimum yang diperlukan untuk generasi respons. Strategi keempat adalah *template-based approach* yang membatasi generasi hanya pada templat yang telah diverifikasi, mencegah generasi teks bebas yang dapat membocorkan informasi sensitif.

Strategi kelima adalah audit dan monitoring dengan mencatat semua aktivitas NLG untuk deteksi anomali dan audit keamanan. Strategi terakhir adalah *secure deployment environment* dengan mengamankan infrastruktur server, firewall, dan menjaga software tetap up-to-date (IT Convergence 2024; Amazon Web Services 2024).

II.4.4.3 Mitigasi Risiko dalam NLG Internal

Risiko utama dalam implementasi NLG internal mencakup kebocoran data sensitif melalui output yang dihasilkan, *prompt injection* yang dapat memanipulasi model untuk menghasilkan output tidak diinginkan, dan *data tampering* yang dapat menyebabkan respons AI yang salah. Mitigasi risiko ini memerlukan pendekatan multi-layer (Fujitsu - PostgreSQL Fastware 2024).

Teknik mitigasi mencakup pemfilteran output untuk memeriksa dan menyensor informasi sensitif sebelum ditampilkan ke pengguna. *Input validation* digunakan untuk memvalidasi semua *input* dan menolak pola yang mencurigakan. *Template whitelisting* membatasi generasi hanya pada templat yang telah disetujui. Implementasi *role-based access control* (RBAC) memastikan pengguna hanya dapat mengakses data sesuai peran mereka. Regular security assessment dilakukan untuk menguji kerentanan sistem secara berkala (European Data Protection Board 2025; Velotix 2025).

II.4.5 Evaluasi Kualitas *Natural Language Generation*

Evaluasi kualitas keluaran *Natural Language Generation* (NLG) merupakan aspek penting untuk memastikan bahwa teks yang dihasilkan tidak hanya benar secara faktual, tetapi juga mudah dipahami, wajar, dan bermanfaat bagi pengguna. Dalam konteks penelitian ini, fungsi utama modul NLG adalah menyajikan hasil analisis

dan prediksi dalam bentuk jawaban yang lebih ramah pengguna, sehingga penilaian kualitas dari sudut pandang pengguna menjadi sangat sentral. Meskipun demikian, sejumlah penelitian terdahulu menunjukkan bahwa kombinasi antara metrik evaluasi otomatis dan penilaian manusia (subjektif) memberikan gambaran yang lebih komprehensif mengenai kualitas keluaran NLG (Celikyilmaz, Clark, dan Gao 2020; Sai, Mohankumar, dan Khapra 2020; Lee dkk. 2021).

II.4.5.1 Metrik Otomatis untuk Evaluasi NLG

Berbagai survei mengenai evaluasi NLG mengelompokkan metrik otomatis ke dalam beberapa kategori utama, seperti metrik berbasis kecocokan permukaan (*string-matching*), metrik berbasis pemetaan vektor (*embedding-based*), dan metrik berbasis model pembelajaran mesin (Celikyilmaz, Clark, dan Gao 2020; Sai, Mohankumar, dan Khapra 2020; Gao dkk. 2025).

Kategori pertama adalah metrik berbasis referensi yang membandingkan keluaran sistem dengan satu atau beberapa teks rujukan yang ditulis manusia. Contoh yang paling banyak digunakan adalah BLEU yang mengukur kecocokan *n-gram* antara keluaran dan teks rujukan, dengan rumus umum (Sai, Mohankumar, dan Khapra 2020):

$$\text{BLEU} = BP \times \exp \left(\sum_{n=1}^N w_n \log p_n \right)$$

dengan p_n menyatakan presisi *n-gram* dan BP adalah faktor penalti panjang (*brevity penalty*). Metrik lain seperti METEOR memasukkan faktor sinonim dan *stemming*, sedangkan ROUGE berfokus pada daya ingat (*recall*) dan banyak digunakan untuk evaluasi peringkasan teks (Celikyilmaz, Clark, dan Gao 2020; Sai, Mohankumar, dan Khapra 2020).

Kategori kedua adalah metrik berbasis penyajian vektor (*embedding-based*), yang menggunakan representasi neural untuk mengukur kesamaan semantik, bukan hanya kesamaan permukaan. Contohnya adalah BERTScore yang memanfaatkan *embedding* BERT untuk menghitung kesamaan kontekstual antara token keluaran dan token rujukan. Metrik ini umumnya menunjukkan korelasi yang lebih baik dengan penilaian manusia dibandingkan metrik *string-matching* murni, terutama ketika terdapat banyak cara wajar untuk mengekspresikan makna yang sama (Celikyilmaz, Clark, dan Gao 2020; Gao dkk. 2025).

Meskipun berguna untuk perbandingan model secara cepat dan konsisten, berbagai studi meta-evaluasi menegaskan bahwa metrik otomatis ini hanya menangkap seba-

gian aspek kualitas dan sering kali berkorelasi sedang atau bahkan rendah dengan penilaian manusia, terutama pada tugas dialog dan percakapan yang bersifat terbuka (Celikyilmaz, Clark, dan Gao 2020; Sai, Mohankumar, dan Khapra 2020; Xiao dkk. 2023). Oleh karena itu, metrik otomatis umumnya direkomendasikan sebagai pelengkap, bukan pengganti, evaluasi subjektif oleh manusia.

II.4.5.2 Evaluasi Subjektif Berbasis Pengguna

Sejumlah survei dan studi empiris menegaskan bahwa evaluasi manusia tetap dianggap sebagai standar emas (*gold standard*) dalam penilaian kualitas keluaran NLG, khususnya untuk aspek yang bersifat kualitatif seperti kewajaran bahasa, keterpahaman, dan kegunaan bagi pengguna akhir (Lee dkk. 2021; Celikyilmaz, Clark, dan Gao 2020; Morrison dkk. 2021). Evaluasi subjektif ini umumnya dilakukan dengan meminta penilai manusia memberikan skor pada skala Likert (biasanya 1–5 atau 1–7) terhadap beberapa dimensi kualitas teks.

Praktik evaluasi manusia yang umum pada sistem NLG dan sejumlah dimensi yang sering digunakan, antara lain (Lee dkk. 2021):

- a. **Kewajaran/Naturalness**: Sejauh mana teks terdengar wajar seperti ditulis penutur asli dan tidak tampak kaku atau *robotik*.
- b. **Kelancaran/Fluency**: Sejauh mana teks bebas dari kesalahan tata bahasa dan struktur kalimat yang janggal.
- c. **Kecukupan/Adequacy atau Informativeness**: Sejauh mana teks menyampaikan informasi yang tepat dan memadai terkait masukan.
- d. **Koherensi/Coherence**: Sejauh mana kalimat-kalimat dalam teks saling tersambung secara logis dan membentuk alur yang mudah diikuti.

Kerangka evaluasi seperti ConSiDERS dan panduan evaluasi subjektif yang dapat direproduksi menekankan pentingnya perancangan instruksi penilaian yang jelas, pemilihan skala yang konsisten, dan penggunaan beberapa penilai untuk meningkatkan reliabilitas (Elangovan dkk. 2023; Morrison dkk. 2021). Dalam surveinya juga digarisbawahi bahwa untuk banyak tugas NLG, korelasi antara metrik otomatis dan penilaian manusia tidak cukup tinggi sehingga penilaian manusia tetap diperlukan, terutama ketika tujuan utama sistem adalah meningkatkan mutu pengalaman pengguna, bukan mengoptimalkan skor metrik otomatis tertentu (Sai, Mohankumar, dan Khapra 2020).

Dalam konteks *chatbot* berbasis AI, sejumlah penelitian di bidang kualitas layanan dan pengalaman pengguna mengaitkan kualitas percakapan alami dengan dimensi

seperti *friendliness*, *usefulness*, kejelasan jawaban, serta persepsi kemudahan penggunaan. Studi tentang kualitas layanan AI-*chatbot* menunjukkan bahwa kemampuan percakapan yang wajar, relevansi informasi, dan kecepatan respons berkontribusi signifikan terhadap kepuasan, kepercayaan, dan niat pengguna untuk terus menggunakan layanan tersebut (Syarifudin dkk. 2024; Nama dan Penulis 2024). Hal ini sejalan dengan fokus penelitian ini, di mana fungsi utama modul NLG adalah membuat jawaban sistem lebih ramah pengguna dan mudah dipahami, bukan menghasilkan variasi bahasa yang kreatif tanpa batas.

Dengan demikian, evaluasi subjektif dalam konteks *chatbot* BI dapat difokuskan pada beberapa dimensi utama berikut:

- a. **Kewajaran Bahasa:** Apakah jawaban terdengar alami dan tidak kaku.
- b. **Kejelasan dan Keterpahaman:** Apakah isi jawaban mudah dipahami oleh pengguna nonteknis.
- c. **Kegunaan/Usefulness:** Apakah jawaban benar-benar membantu pengguna menjawab pertanyaan bisnis yang diajukan.
- d. **Kenyamanan Interaksi:** Sejauh mana gaya bahasa dan struktur jawaban mendukung pengalaman percakapan yang menyenangkan.

Dimensi-dimensi tersebut dapat diukur menggunakan skala Likert oleh sejumlah penilai internal pada tahap pengujian sistem, mengikuti praktik yang direkomendasikan dalam literatur evaluasi NLG dan kualitas *chatbot* (Lee dkk. 2021; Syarifudin dkk. 2024).

II.4.5.3 Kriteria Performa NLG dalam Literatur

Survei-survei terkini mengenai evaluasi NLG menekankan bahwa tidak ada satu nilai ambang (*threshold*) numerik yang bersifat universal untuk menyatakan bahwa suatu sistem “cukup baik”, karena nilai metrik sangat bergantung pada tugas, data, dan desain sistem yang dievaluasi (Celikyilmaz, Clark, dan Gao 2020; Sai, Mohankumar, dan Khapra 2020; Gao dkk. 2025). Oleh karena itu, karya-karya ilmiah umumnya:

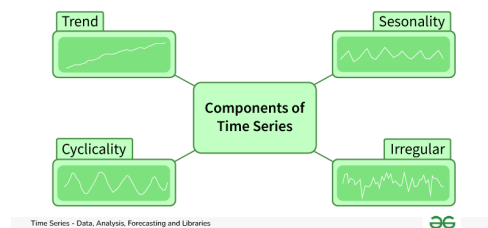
- a. Membandingkan metrik otomatis (seperti BLEU, ROUGE, BERTScore) relatif terhadap model dasar atau pendekatan lain pada tugas yang sama, dan
- b. Menggunakan penilaian manusia sebagai landasan utama, dengan rata-rata skor pada skala 1–5 atau 1–7 untuk dimensi seperti kewajaran, kelancaran, dan kecukupan, di mana skor rata-rata yang mendekati ujung atas skala lazim dipandang sebagai indikasi kualitas tinggi (Lee dkk. 2021; Celikyilmaz, Clark, dan Gao 2020).

Dalam studi-studi dialog dan percakapan, sistem yang dianggap berkualitas baik biasanya menunjukkan:

- a. Peningkatan signifikan pada metrik otomatis dibandingkan model dasar, *namun* penulis tetap menekankan bahwa korelasi metrik otomatis dengan penilaian manusia terbatas dan konteks spesifik harus diperhatikan (Sai, Mohankumar, dan Khapra 2020; Xiao dkk. 2023),
- b. Rata-rata skor penilaian manusia yang tinggi untuk dimensi kewajaran, kelancaran, dan relevansi, serta
- c. Indikator kepuasan dan persepsi kualitas layanan yang positif pada studi pengguna, seperti peningkatan kepercayaan, kenyamanan interaksi, dan niat untuk terus menggunakan *chatbot* (Syarifudin dkk. 2024; Nama dan Penulis 2024).

Berdasarkan informasi tersebut, literatur menyarankan bahwa pemilihan metrik evaluasi NLG hendaknya disesuaikan dengan tujuan sistem. Dalam sistem *Business Intelligence* berbasis *chatbot* yang memanfaatkan NLG terutama untuk menjadikan jawaban lebih ramah pengguna dan mudah dipahami, metrik subjektif yang berfokus pada persepsi pengguna (kewajaran, keterpahaman, kegunaan) dapat dipandang sebagai indikator utama, sedangkan metrik otomatis seperti BLEU atau BERTScore dapat digunakan secara opsional untuk memantau konsistensi bentuk bahasa sepanjang pengembangan (Celikyilmaz, Clark, dan Gao 2020; Lee dkk. 2021).

II.5 Time Series Forecasting



Gambar II.4 Time Series Forecasting

Bagian ini membahas konsep peramalan deret waktu, model-model statistik dan pembelajaran mendalam yang digunakan, serta evaluasi performa yang komprehensif. Pembahasan mencakup ARIMA, SARIMA, LSTM, serta perbandingan karakteristik dan performa masing-masing metode. Pemahaman yang mendalam tentang berbagai pendekatan peramalan ini menjadi fondasi penting dalam memilih dan mengimplementasikan model yang sesuai dengan karakteristik data dan kebutuhan aplikasi.

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.

Pasar *time series forecasting* global mencapai USD 9,62 miliar pada tahun 2023 dan diperkirakan tumbuh menjadi USD 36,9 miliar pada tahun 2032 dengan CAGR sebesar 16,12% (WiseGuy Reports 2024). Pertumbuhan ini menunjukkan meningkatnya adopsi teknik peramalan deret waktu di berbagai industri untuk mendukung pengambilan keputusan berbasis data, termasuk peramalan permintaan layanan pelanggan, prediksi penjualan, dan analisis tren bisnis. Proyeksi pertumbuhan pasar ini mengindikasikan relevansi dan urgensi pengembangan sistem peramalan yang akurat dan dapat diandalkan.

II.5.2 Model Peramalan Deret Waktu Berbasis *Deep Learning*

Sistem peramalan dalam penelitian ini memanfaatkan pendekatan berbasis pembelajaran mendalam (*deep learning*) yang dirancang khusus untuk menangani data deret waktu multivariat (*multivariate time series*) dengan pola temporal yang kompleks. Sistem menggunakan tiga arsitektur utama, LSTM, GRU, dan Temporal Fusion Transformer (TFT), serta kombinasi hibrida dari algoritma-algoritma tersebut untuk memaksimalkan akurasi prediksi terhadap indikator kinerja bisnis seperti volume pesanan, tingkat *churn* pelanggan, dan proyeksi pendapatan (Yunita dkk. 2025; Yun Zhang dkk. 2024).

II.5.2.1 LSTM (*Long Short-Term Memory*) untuk Peramalan Deret Waktu

LSTM adalah jenis jaringan saraf berulang (*Recurrent Neural Network*, RNN) yang dirancang untuk mengatasi masalah gradien menghilang (*vanishing gradient*) pada RNN tradisional dan mampu mempelajari dependensi jangka panjang dalam data sekuens (Yunita dkk. 2025). Arsitektur LSTM terdiri dari sel memori dan tiga gerbang utama: gerbang lupa (*forget gate*) yang menentukan informasi apa dari sel memori yang harus dibuang, gerbang masukan (*input gate*) yang menentukan informasi baru apa yang akan disimpan, dan gerbang keluaran (*output gate*) yang menentukan bagian mana dari sel memori yang akan menjadi keluaran.

Persamaan matematis LSTM dapat ditulis sebagai berikut (Yunita dkk. 2025):

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (\text{II.6})$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (\text{II.7})$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (\text{II.8})$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (\text{II.9})$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (\text{II.10})$$

$$h_t = o_t * \tanh(C_t) \quad (\text{II.11})$$

Dalam persamaan ini, f_t adalah gerbang lupa, i_t adalah gerbang masukan, o_t adalah gerbang keluaran, C_t adalah keadaan sel (*cell state*), h_t adalah keadaan tersembunyi (*hidden state*), σ adalah fungsi sigmoid, serta W dan b adalah matriks bobot dan vektor bias yang dipelajari selama pelatihan.

LSTM unggul dalam menangani data deret waktu yang kompleks dengan dependensi jangka panjang dan pola non-linear. Kemampuan model untuk mengingat informasi relevan dari masa lalu yang jauh menjadikannya sangat efektif untuk aplikasi seperti peramalan permintaan atau tingkat *churn* pelanggan yang dipengaruhi oleh berbagai faktor temporal yang kompleks (Yunita dkk. 2025).

II.5.2.2 GRU (*Gated Recurrent Unit*) sebagai Alternatif LSTM

GRU adalah varian yang lebih sederhana dari LSTM, menawarkan performa yang sebanding dengan menggunakan arsitektur yang lebih ringkas dan parameter yang lebih sedikit (Yunita dkk. 2025; Yun Zhang dkk. 2024). GRU memiliki dua gerbang utama: gerbang reset (*reset gate*) dan gerbang pembaruan (*update gate*), menggabungkan fungsi gerbang lupa dan gerbang masukan LSTM menjadi satu gerbang pembaruan.

Persamaan matematis GRU dapat ditulis sebagai (Yunita dkk. 2025):

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \quad (\text{II.12})$$

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z) \quad (\text{II.13})$$

$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t] + b) \quad (\text{II.14})$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (\text{II.15})$$

Dalam persamaan ini, r_t adalah gerbang reset, z_t adalah gerbang pembaruan, dan \tilde{h}_t

adalah calon keadaan tersembunyi. Studi empiris menunjukkan bahwa GRU sering memberikan hasil yang setara atau bahkan lebih baik daripada LSTM pada berbagai tugas peramalan deret waktu, dengan keuntungan pelatihan yang lebih cepat dan kebutuhan komputasi yang lebih rendah (Yunita dkk. 2025; Jeaab dkk. 2024).

II.5.2.3 *Temporal Fusion Transformer (TFT) untuk Data Multivariat*

Temporal Fusion Transformer (TFT) adalah arsitektur berbasis perhatian (*attention-based*) yang dirancang khusus untuk peramalan deret waktu multi-horizon pada data multivariat yang kompleks (Lim dkk. 2021; Yun Zhang dkk. 2024). TFT menggabungkan kekuatan mekanisme perhatian diri (*self-attention*) Transformer dengan lapisan berulang untuk pemrosesan lokal, memungkinkan model menangkap ketergantungan temporal pada berbagai skala waktu.

Arsitektur TFT terdiri dari beberapa komponen utama (Lim dkk. 2021; Hartanto dkk. 2024):

- a. **Variabel *Input Tipe-Campuran***: TFT menangani masukan eksogen tergantung waktu (*time-dependent exogenous features*) yang terdiri dari masukan yang tidak diketahui sebelumnya dan masukan yang diketahui, serta kovariat statis yang memberikan konteks tentang entitas yang diukur.
- b. **Jaringan Pemilihan Variabel (*Variable Selection Network*)**: Komponen ini secara adaptif memilih fitur yang paling relevan dari antara semua masukan, meningkatkan interpretabilitas dan mengurangi kebisingan dari fitur yang tidak relevan (Lim dkk. 2021).
- c. **Lapisan Gerbang Residu (*Gated Residual Networks*)**: Lapisan ini menerapkan mekanisme gerbang untuk menekan komponen yang tidak perlu, meningkatkan kemampuan model untuk fokus pada pola penting (Lim dkk. 2021).
- d. **Mekanisme *Multi-Head Attention***: Mekanisme perhatian diri memungkinkan model menangkap hubungan jangka panjang dan pola kompleks dalam data, memberikan ketepatan prediksi yang tinggi pada berbagai horizon waktu (Hartanto dkk. 2024).

Berbeda dengan RNN yang memproses data secara sekuensial, TFT dapat memproses seluruh urutan masukan secara paralel, menghasilkan pelatihan yang lebih cepat. Penelitian menunjukkan bahwa TFT unggul dalam menangani skenario peramalan multivariat yang kompleks dengan banyak fitur yang saling berinteraksi, seperti kasus prediksi kinerja bisnis yang melibatkan banyak indikator yang saling terkait (Hartanto dkk. 2024; Cheng dkk. 2025).

II.5.2.4 Model Hibrida dan *Ensemble*

Penelitian terkini menunjukkan bahwa kombinasi hibrida dari arsitektur yang berbeda sering memberikan hasil yang lebih baik daripada model tunggal, karena setiap arsitektur menangkap aspek berbeda dari pola temporal dalam data (Yunita dkk. 2025; Yun Zhang dkk. 2024).

Model hibrida **GRU-LSTM** menggabungkan keunggulan komplementer dari kedua arsitektur berulang. Studi pada prediksi suhu harian menunjukkan bahwa model hibrida GRU-LSTM mencapai RMSE sebesar 0,07499 dan MAE sebesar 0,0578 pada skenario masukan 15 hari, dengan R^2 mendekati 1 (0,9937), menunjukkan performa yang sangat baik (Diqi dkk. 2023). Keuntungan pendekatan hibrida terletak pada fleksibilitas untuk memanfaatkan kapabilitas pemrosesan lokal GRU yang efisien sekaligus pembelajaran dependensi jangka panjang yang kuat dari LSTM.

Pendekatan ensemble yang menggabungkan keluaran dari multiple model (misalnya LSTM, GRU, dan TFT) juga terbukti meningkatkan robustness dan ketepatan prediksi, terutama dalam menangani variabilitas dan ketidakpastian dalam data bisnis real-world (Yun Zhang dkk. 2024; Yunita dkk. 2025).

II.5.2.5 Perbandingan Karakteristik Model Peramalan *Deep Learning*

Berbagai penelitian empiris telah membandingkan kinerja LSTM, GRU, dan TFT pada berbagai skenario peramalan deret waktu. Tabel II.1 menyajikan perbandingan karakteristik utama dari ketiga arsitektur berdasarkan sintesis dari penelitian-penelitian terdahulu.

Tabel II.1 Perbandingan Karakteristik Model Peramalan Berbasis *Deep Learning*

Metode	Kelebihan	Kekurangan	Karakteristik
LSTM	<ol style="list-style-type: none"> 1. Menangkap dependensi jangka panjang hingga ratusan langkah waktu 2. Mengatasi masalah <i>vanishing gradient</i> melalui gerbang-gerbang adaptif 3. Fleksibel untuk menangani pola temporal kompleks dan <i>non-linear</i> 4. Mampu mempelajari urutan masukan dengan panjang variabel 	<ol style="list-style-type: none"> 1. Memerlukan volume data historis besar (minimal ribuan sampel) 2. Waktu pelatihan panjang, terutama pada <i>dataset</i> besar atau GPU terbatas 3. Sulit untuk interpretasi internal (bersifat <i>black-box</i>) 4. Proses <i>inference</i> sekuensial, tidak dapat diparalelisasi secara efisien 5. Sensitif terhadap inisialisasi <i>hyperparameter</i> 	<ol style="list-style-type: none"> 1. MAPE tipikal: 8–12% 2. Cocok untuk: data besar, pola kompleks 3. Kompleksitas model: tinggi (jutaan parameter) 4. Parameter: ratusan ribu

Bersambung ke halaman berikutnya

Tabel II.1 – Perbandingan Model Peramalan *Deep Learning* (lanjutan)

Metode	Kelebihan	Kekurangan	Karakteristik
GRU	<ol style="list-style-type: none"> 1. Arsitektur lebih sederhana dibanding LSTM (hanya dua gerbang) 2. Waktu pelatihan 15–20% lebih cepat dari LSTM pada data sama 3. Jumlah parameter 20–30% lebih sedikit, mengurangi risiko <i>overfitting</i> 4. Performa sebanding atau lebih baik dari LSTM pada banyak tugas 5. Kebutuhan memori lebih rendah 	<ol style="list-style-type: none"> 1. Tetap memerlukan data historis dalam jumlah besar 2. Juga sulit untuk interpretasi internal (<i>black-box</i>) 3. Literatur dan <i>best-practices</i> lebih terbatas dibanding LSTM 4. Performa terkadang sedikit lebih rendah pada skenario dengan ketergantungan ultra-jangka-panjang 5. <i>Inference</i> masih sekuensial 	<ol style="list-style-type: none"> 1. MAPE tipikal: 10–13% 2. Cocok untuk: data besar, komputasi terbatas 3. Kompleksitas model: sedang–tinggi 4. Parameter: 20–30% lebih sedikit dari LSTM

Bersambung ke halaman berikutnya

Tabel II.1 – Perbandingan Model Peramalan *Deep Learning* (lanjutan)

Metode	Kelebihan	Kekurangan	Karakteristik
TFT	<ol style="list-style-type: none"> 1. Dirancang khusus untuk data multivariat dengan banyak fitur terkait 2. Interpretabilitas lebih baik: mekanisme <i>attention</i> menunjukkan fitur/waktu penting 3. Paralelisasi efisien: seluruh urutan diproses sekaligus, bukan sekuensial 4. <i>Multi-horizon forecasting</i>: dapat memprediksi beberapa langkah ke depan sekaligus 5. Seleksi variabel otomatis mengurangi pengaruh fitur yang tidak relevan 	<ol style="list-style-type: none"> 1. Kompleksitas arsitektur tinggi (6–7 <i>submodule</i> berbeda) 2. Memerlukan penyetelan (<i>tuning</i>) <i>hyperparameter</i> lanjut 3. Data historis dalam jumlah sangat besar optimal untuk performa terbaik 4. Biaya komputasi (<i>computational cost</i>) tinggi (GPU umumnya diperlukan) 5. Waktu <i>training</i> lebih lama dari LS-TM/GRU karena kompleksitas 6. Memerlukan keahlian teknis lebih tinggi untuk implementasi 	<ol style="list-style-type: none"> 1. MAPE tipikal: 7–11% 2. Cocok untuk: data multivariat, fitur banyak, <i>budget</i> komputasi tinggi 3. Kompleksitas model: sangat tinggi 4. Parameter: jutaan (tergantung konfigurasi) 5. Paralelisasi: baik

Bersambung ke halaman berikutnya

Tabel II.1 – Perbandingan Model Peramalan *Deep Learning* (lanjutan)

Metode	Kelebihan	Kekurangan	Karakteristik
Hibrida (GRU-LSTM)	<ol style="list-style-type: none"> 1. Menggabungkan kekuatan dua arsitektur: efisiensi GRU dan pembelajaran jangka panjang <i>LSTM</i> 2. Hasil prediksi lebih <i>robust</i> terhadap variasi dalam data (<i>ensemble effect</i>) 3. Efisiensi komputasi baik: kompromi antara akurasi dan kecepatan 4. Fleksibilitas desain: dapat dikonfigurasi sesuai karakteristik <i>dataset</i> 5. Mengurangi risiko model bias terhadap pola tertentu 	<ol style="list-style-type: none"> 1. Desain dan implementasi lebih kompleks 2. Jumlah <i>hyperparameter</i> lebih banyak sehingga <i>tuning</i> lebih sulit 3. Risiko <i>overfitting</i> lebih tinggi jika tidak dikendalikan 4. Waktu <i>training</i> lebih lama dibanding model tunggal 5. Interpretasi model lebih sulit karena kombinasi dua arsitektur 6. Memerlukan validasi silang (<i>cross-validation</i>) yang cermat 	<ol style="list-style-type: none"> 1. MAPE tipikal: 7–10% 2. Cocok untuk: data multivariat, pola campuran, GPU menengah 3. Kompleksitas model: tinggi–sangat tinggi 4. Parameter: ratusan ribu hingga jutaan 5. <i>Trade-off</i>: akurasi vs efisiensi

Studi empiris menunjukkan bahwa pada data yang mengandung pola musiman kuat dan dependensi temporal kompleks, seperti data penjualan atau *churn* pelanggan bisnis, model LSTM dan GRU mengungguli metode statistik tradisional seperti ARIMA dengan pengurangan *error* sekitar 12–18% (Sunendar dan Rianto 2025). Penelitian pada prediksi stok Indonesia menunjukkan GRU mengungguli LSTM dengan peningkatan R^2 sebesar 10,7% dan pengurangan MAPE sebesar 18,5%, menunjukkan bahwa pada *dataset* tertentu, GRU dapat memberikan hasil lebih efisien (Sabila

dkk. 2025).

Model TFT menunjukkan performa tertinggi pada data multivariat, dengan MAPE yang mencapai 0,22% (0,0022) pada prediksi harga saham Indonesia ketika dikombinasikan dengan rekayasa fitur yang tepat (Hartanto dkk. 2024). Keunggulan TFT dalam interpretabilitas juga membuatnya menarik untuk aplikasi bisnis di mana transparansi model menjadi penting bagi *stakeholder*.

II.5.2.6 Pemilihan Model dalam Konteks Penelitian

Dalam penelitian ini, pemilihan algoritma peramalan disesuaikan dengan karakteristik data dan kebutuhan sistem. Sistem mengimplementasikan LSTM sebagai model dasar untuk menangkap pola temporal kompleks pada indikator kinerja bisnis seperti volume pesanan dan proyeksi pendapatan. GRU dapat digunakan sebagai alternatif untuk meningkatkan efisiensi pelatihan dan inferensi ketika sumber daya komputasi terbatas. TFT akan dipertimbangkan untuk skenario yang melibatkan banyak fitur eksogen yang saling terkait atau ketika interpretabilitas model menjadi prioritas utama (Yun Zhang dkk. 2024; Yunita dkk. 2025).

Pendekatan ensemble atau hibrida dapat diterapkan untuk meningkatkan robustness prediksi, terutama pada horizon prediksi yang panjang atau dalam kondisi data yang sangat berisik dan penuh ketidakpastian. Evaluasi empiris pada data historis internal akan dilakukan untuk menentukan model atau kombinasi model mana yang memberikan akurasi terbaik dan paling sesuai dengan kebutuhan operasional sistem BI (Yunita dkk. 2025).

II.5.3 Evaluasi Performa *Time Series Forecasting*

Evaluasi performa model peramalan deret waktu merupakan langkah kritis untuk memastikan bahwa prediksi yang dihasilkan memiliki akurasi yang memadai untuk mendukung pengambilan keputusan bisnis. Evaluasi yang komprehensif mencakup berbagai metrik dan teknik analisis yang saling melengkapi untuk memberikan gambaran menyeluruh tentang kualitas model.

II.5.3.1 Metrik Evaluasi *Forecasting*

Model peramalan deret waktu multivariat dievaluasi menggunakan empat metrik utama yang komprehensif: *Mean Absolute Error* (MAE), *Root Mean Square Error* (RMSE), *Symmetric Mean Absolute Percentage Error* (sMAPE), dan *Mean Absolute Scaled Error* (MASE) (Petroopoulos dkk. 2023). Setiap metrik memberikan

perspektif berbeda terhadap akurasi prediksi, sehingga kombinasinya menghasilkan penilaian yang lebih robust dan komprehensif.

Mean Absolute Error (MAE) mengukur rata-rata nilai absolut kesalahan prediksi:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (\text{II.16})$$

MAE mudah diinterpretasikan karena berada dalam satuan yang sama dengan data asli, dan *robust* terhadap *outliers*. Namun, MAE tidak memberikan penalti tambahan pada kesalahan yang sangat besar (Petropoulos dkk. 2023).

Root Mean Square Error (RMSE) mengukur akar kuadrat dari rata-rata kesalahan kuadrat:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (\text{II.17})$$

RMSE memberikan penalti yang lebih besar untuk kesalahan yang lebih besar, menjadikannya lebih sensitif terhadap *outliers* dibandingkan MAE. Metrik ini berguna ketika kesalahan besar sangat tidak diinginkan dalam konteks bisnis (Petropoulos dkk. 2023).

Symmetric Mean Absolute Percentage Error (sMAPE) mengukur persentase kesalahan secara simetris:

$$\text{sMAPE} = \frac{100\%}{n} \sum_{i=1}^n \frac{2|y_i - \hat{y}_i|}{|y_i| + |\hat{y}_i|} \quad (\text{II.18})$$

sMAPE lebih stabil daripada MAPE (*Mean Absolute Percentage Error*) ketika nilai aktual mendekati nol, karena pembagi mencakup baik nilai aktual maupun nilai prediksi. Hal ini membuat sMAPE lebih cocok untuk data multivariat dengan skala dan nilai yang bervariasi (Petropoulos dkk. 2023).

Mean Absolute Scaled Error (MASE) menormalkan kesalahan terhadap model referensi sederhana:

$$\text{MASE} = \frac{\text{MAE}}{\text{MAE}_{\text{baseline}}} \quad (\text{II.19})$$

MASE memudahkan perbandingan antar-deret waktu yang berbeda skala dan antar-model yang berbeda, karena kesalahan dinormalkan relatif terhadap garis dasar (*ba-*

seline). MASE bernilai kurang dari 1 menunjukkan bahwa model mengungguli garis dasar, sedangkan nilai lebih dari 1 menunjukkan performa yang lebih buruk (Petropoulos dkk. 2023).

Dalam formula-formula di atas, y_i adalah nilai aktual, \hat{y}_i adalah nilai prediksi, dan n adalah jumlah observasi.

II.5.3.2 *Skill Score* terhadap Garis Dasar

Selain metrik kesalahan mentah, sistem juga menghitung *skill score* yang membandingkan performa model terhadap garis dasar (*baseline*) sederhana, seperti model *naive* (prediksi nilai sebelumnya) atau *random walk* (Petropoulos dkk. 2023). *Skill score* didefinisikan sebagai:

$$\text{Skill Score} = 1 - \frac{\text{Metrik}_{\text{model}}}{\text{Metrik}_{\text{baseline}}} \quad (\text{II.20})$$

Skill score positif menunjukkan bahwa model yang diusulkan mengungguli garis dasar, sedangkan nilai negatif mengindikasikan performa yang lebih buruk dari model referensi. Pendekatan ini memberikan konteks relatif tentang keberhasilan model dan memudahkan pengambilan keputusan apakah model kompleks (seperti LSTM atau TFT) justru diperlukan atau model sederhana sudah cukup (Petropoulos dkk. 2023; Spiliotis, Makridakis, dan Petropoulos 2020).

II.5.3.3 Pelaporan per Horizon dan per Entitas Prioritas

Dalam memberikan gambaran yang kaya dan terstruktur tentang kinerja model pada berbagai skenario, hasil evaluasi dilaporkan tidak hanya sebagai nilai rata-rata global, tetapi per *forecast horizon* (cakrawala peramalan) dan per entitas prioritas (*priority entity*) (Spiliotis, Makridakis, dan Petropoulos 2020).

Dalam konteks *forecast horizon*, model dievaluasi secara terpisah untuk setiap cakrawala waktu. Hal ini penting karena akurasi model umumnya menurun seiring bertambahnya jarak prediksi, dan pola degradasi akurasi ini berguna untuk memahami batas kepercayaan pada setiap horizon (Petropoulos dkk. 2023).

Dalam konteks *priority entity*, model dievaluasi secara terpisah untuk setiap entitas yang diprioritaskan. Hal ini memungkinkan identifikasi entitas mana yang memiliki akurasi tinggi dan mana yang memerlukan penyempurnaan model (Spiliotis, Makridakis, dan Petropoulos 2020).

Dalam setiap kombinasi horizon–entitas, metrik kesalahan dilaporkan menggunakan:

- a. **Nilai Median (*Median*)**: Nilai tengah yang *robust* terhadap *outliers*, memberikan ukuran tendensi sentral yang lebih andal daripada rata-rata.
- b. **Rentang Antar Kuartil (*Interquartile Range, IQR*)**: Selisih antara kuartil ketiga dan kuartil pertama, menunjukkan penyebaran 50% tengah dari distribusi kesalahan. IQR memberikan gambaran tentang variabilitas performa model tanpa dipengaruhi nilai ekstrem.

Pendekatan ini memungkinkan analisis distribusi performa model yang lebih robust, bukan hanya bergantung pada nilai rata-rata, dan dapat mengidentifikasi situasi atau entitas di mana model mungkin mengalami tantangan (Petropoulos dkk. 2023; Spiliotis, Makridakis, dan Petropoulos 2020).

II.5.3.4 *Residual Analysis*

Analisis residual merupakan komponen penting dalam evaluasi model peramalan untuk memastikan bahwa asumsi model terpenuhi dan tidak ada pola sistematis yang tersisa dalam kesalahan prediksi. Analisis ini mencakup beberapa teknik yang saling melengkapi:

- a. **Plot Residual terhadap Waktu**: Mendeteksi pola temporal yang mungkin tersisa dalam residual. Residual yang baik harus terdistribusi secara acak di sekitar nol tanpa tren atau pola musiman yang jelas, mengindikasikan bahwa model telah menangkap sebagian besar struktur dinamis dalam data.
- b. **Uji Normalitas (*Shapiro-Wilk Test*)**: Menguji apakah residual mengikuti distribusi normal. Normalitas residual mengindikasikan bahwa model telah mengekstrak informasi sistem secara efektif dan kesalahan yang tersisa bersifat acak murni.
- c. **Uji Autokorelasi (*Ljung-Box Test*)**: Mendeteksi autokorelasi dalam residual dengan uji statistik. Autokorelasi yang signifikan mengindikasikan bahwa masih ada informasi temporal yang belum ditangkap oleh model, dan model mungkin perlu diperbaiki atau divariasikan.

Ketiga teknik ini memberikan pemeriksaan komprehensif terhadap kualitas model dan asumsi yang mendasarinya (Petropoulos dkk. 2023).

II.5.3.5 *Time Series Cross-Validation*

Validasi silang khusus untuk deret waktu (*time series cross-validation*) menggunakan pendekatan *walk-forward validation* atau *expanding window* untuk menghinda-

ri kebocoran informasi dari masa depan ke masa lalu (Spiliotis, Makridakis, dan Petropoulos 2020). Pendekatan ini membagi data secara berurutan: pada setiap iterasi, model dilatih pada subset data historis yang terus bertambah, kemudian diminta memprediksi periode mendatang. Hal ini mensimulasikan kondisi operasional (*deployment*) aktual di mana model harus memprediksi masa depan berdasarkan data masa lalu yang diketahui.

Validasi silang deret waktu memberikan estimasi performa yang lebih realistis dibandingkan validasi silang standar (misalnya *k-fold*), yang dapat menghasilkan *overfitting* karena kebocoran informasi temporal. Dengan pendekatan *walk-forward*, estimasi akurasi model pada data uji mencerminkan performa yang dapat diharapkan dalam praktik (Petropoulos dkk. 2023).

II.5.3.6 Target Performa *Forecasting*

Berdasarkan literatur dan praktik industri terkini, target performa yang ditetapkan untuk model peramalan deret waktu multivariat dalam konteks *Business Intelligence* mencakup beberapa kriteria utama (Petropoulos dkk. 2023; Spiliotis, Makridakis, dan Petropoulos 2020):

- a. **sMAPE**: Harus mencapai maksimal 15%, yang dianggap dapat diterima untuk sebagian besar aplikasi bisnis dengan prediksi deret waktu multivariat.
- b. **RMSE**: Harus seminimal mungkin dan dibandingkan dengan standar deviasi (*standard deviation*) data aktual. RMSE yang kurang dari setengah standar deviasi data aktual umumnya dianggap performa yang baik.
- c. **MAE**: Harus seminimal mungkin dan diinterpretasikan dalam konteks skala data. MAE yang secara signifikan lebih kecil dari nilai rata-rata (*mean*) data aktual menunjukkan performa yang baik.
- d. **MASE**: Harus bernilai kurang dari 1 untuk menunjukkan bahwa model mengungguli garis dasar sederhana. Nilai MASE antara 0,5 dan 1 mengindikasikan performa yang sangat baik relatif terhadap garis dasar.
- e. **Residual**: Harus lulus uji *Ljung-Box* dengan *p-value* lebih besar dari 0,05, mengindikasikan tidak ada autokorelasi signifikan yang tersisa dalam residual.
- f. **Skill Score**: Harus positif untuk semua horizon dan entitas prioritas, menunjukkan konsistensi keunggulan model terhadap garis dasar.

Pencapaian target-target ini, diukur melalui statistik median dan IQR per horizon dan per entitas prioritas, memastikan bahwa model peramalan dapat memberikan prediksi yang akurat, andal, dan konsisten untuk mendukung perencanaan strategis dan

pengambilan keputusan operasional dalam sistem *Business Intelligence* (Petrooulos dkk. 2023).

II.6 Integrasi Sistem dan Penelitian Terkait

Bagian ini membahas integrasi berbagai komponen teknologi dalam sistem *chatbot* Business Intelligence, penelitian terkait arsitektur *chatbot* hibrida, serta aplikasi pembelajaran mesin untuk prediksi *customer churn*. Pembahasan mencakup strategi integrasi, arsitektur sistem, dan studi kasus implementasi yang menunjukkan efektivitas pendekatan terintegrasi. Pemahaman tentang integrasi sistem menjadi kunci dalam mengembangkan solusi yang komprehensif dan koheren.

II.6.1 Arsitektur Chatbot Hibrida

Arsitektur *chatbot* hibrida menggabungkan pendekatan berbasis aturan (*rule-based*) dengan kemampuan kecerdasan buatan untuk menciptakan sistem yang seimbang antara kontrol deterministik dan fleksibilitas adaptif. Pendekatan hibrida ini memungkinkan sistem untuk menangani berbagai jenis kueri dengan efisiensi optimal sambil mempertahankan akurasi dan keamanan data.

II.6.1.1 Komponen Arsitektur Chatbot Modern

Arsitektur *chatbot* modern terdiri dari lima komponen utama yang bekerja secara terintegrasi (Huang 2021; XenonStack 2023). Komponen pertama adalah antarmuka pengguna (*User Interface*) yang memungkinkan pengguna berkomunikasi dengan *chatbot* melalui platform perpesanan atau antarmuka berbasis web. Komponen ini dapat mencakup konversi *speech-to-text* untuk *input* suara dan *text-to-speech* untuk output suara.

Komponen kedua adalah modul pemahaman bahasa alami (*Natural Language Understanding*) yang bertanggung jawab untuk menginterpretasikan masukan pengguna dan mengekstraksi *intent* serta entitas. Komponen ketiga adalah manajemen dialog (*Dialogue Management*) yang mengelola alur percakapan, memelihara konteks, dan menentukan tindakan yang sesuai berdasarkan *intent* yang teridentifikasi. Komponen keempat adalah *backend* yang terhubung dengan basis data, sistem eksternal, dan API untuk mengambil informasi yang diperlukan. Komponen terakhir adalah pembangkitan respons (*Response Generation*) yang menghasilkan output bahasa alami berdasarkan data yang diperoleh dan konteks percakapan (Huang 2021).

II.6.1.2 *Chatbot* Hibrida dengan AI dan *Human Oversight*

Pendekatan *chatbot* hibrida yang mengintegrasikan model NLP ringan dengan pengawasan manusia telah menunjukkan efektivitas tinggi dalam konteks pendidikan tinggi. Kerangka kerja hibrida ini menggunakan pohon keputusan, pemetaan *intent*, dan alur percakapan terstruktur untuk mengotomatisasi tugas rutin sambil mempertahankan dukungan kontekstual dan empatik melalui protokol eskalasi dinamis (Bamurange 2025).

Implementasi kerangka kerja hibrida pada Universitas Kigali menunjukkan akurasi 85% dalam menyelesaikan pertanyaan akademik dan mengurangi beban kerja staf sebesar 30% dalam simulasi (Bamurange 2025). Inovasi kunci mencakup fungsi *offline* untuk pengaturan dengan keterbatasan sumber daya, pemeriksaan kesesuaian budaya untuk menginterpretasikan kueri tidak langsung, dan strategi mitigasi bias yang selaras dengan pedoman etika. Temuan ini menekankan kelayakan platform *no-code* untuk deployment yang scalable, menekankan keseimbangan antara otomasi dan intervensi manusia.

II.6.1.3 Pasar *Chatbot* AI Global

Pasar *chatbot* berbasis AI global akan melebihi USD 3 miliar pada tahun 2025, dengan pasar *chatbot* hibrida suara dan teks mencapai USD 534 juta secara global pada tahun yang sama (Business Wire 2020). Penggunaan *Conversational AI* akan merevolusi manajemen hubungan pelanggan, dengan lebih dari 50% kueri pelanggan dapat dikelola melalui *chatbot* berbasis AI. Pertumbuhan eksponensial ini menunjukkan bahwa *chatbot* berbasis AI akan menjadi norma pada tahun 2025 ketika pusat kontak meningkatkan otomasi.

Transformasi dari antarmuka operasional tradisional ke antarmuka percakapan akan mengubah secara dramatis ekspektasi tentang bagaimana manusia berkomunikasi, mengonsumsi konten, menggunakan aplikasi, dan terlibat dalam perdagangan (Business Wire 2020). Transformasi ini akan berdampak pada hampir setiap aspek operasi pemasaran dan penjualan untuk setiap vertikal industri, termasuk sektor Business Intelligence.

II.6.2 Integrasi *Pattern-Based Intent* dan *Template-Based NLG*

Integrasi antara klasifikasi *intent* berbasis pola dan pembangkitan bahasa alami berbasis templat menciptakan alur kerja yang efisien dan aman untuk sistem *chatbot* Business Intelligence. Integrasi ini memastikan bahwa kueri pengguna diproses de-

ngan akurat sambil respons yang dihasilkan tetap bervariasi dan natural.

II.6.2.1 Alur Kerja Terintegrasi

Alur kerja sistem terintegrasi dimulai dengan penerimaan kueri pengguna melalui antarmuka *chatbot*. Modul klasifikasi *intent* berbasis pola melakukan pencocokan kueri dengan pola yang telah didefinisikan dan menghitung *confidence score*. Jika skor mencapai threshold yang ditentukan, sistem mengidentifikasi *intent* dan mengekstraksi parameter yang diperlukan untuk eksekusi kueri basis data.

Setelah *intent* teridentifikasi, sistem mengeksekusi kueri terhadap *data warehouse* untuk mengambil data internal yang relevan. Data yang diperoleh kemudian diteruskan ke modul NLG berbasis templat. Modul NLG memilih templat yang sesuai berdasarkan jenis *intent* dan konteks, mengisi *placeholder* dengan data aktual, dan menghasilkan respons bahasa alami. Dalam meningkatkan kealamian, sistem dapat menggunakan model *template rewriting* seperti T5 untuk memperbaiki koherensi dan variasi respons.

Pendekatan terintegrasi ini memastikan bahwa sistem dapat merespons dengan cepat untuk kueri terstruktur sambil menjaga keamanan data internal melalui pembatasan pada templat yang telah diverifikasi. Kontrol penuh atas output NLG mencegah kebocoran informasi sensitif dan memastikan bahwa respons tetap konsisten dengan kebijakan perusahaan.

II.6.2.2 Evaluasi Sistem Terintegrasi

Evaluasi sistem terintegrasi memerlukan pengukuran performa pada setiap komponen serta performa sistem secara keseluruhan. Metrik evaluasi mencakup akurasi klasifikasi *intent*, kualitas respons NLG, waktu respons end-to-end, dan kepuasan pengguna. Sistem yang baik harus mencapai akurasi klasifikasi *intent* minimal 85%, *factual accuracy* NLG 100%, dan waktu respons total di bawah 500 milidetik untuk 95% kueri.

Pengujian integrasi juga mencakup skenario pengujian end-to-end yang mensimulasikan percakapan nyata, pengujian beban untuk memastikan sistem dapat menangani volume kueri tinggi, dan pengujian keamanan untuk memverifikasi bahwa tidak ada kebocoran data sensitif. Pendekatan evaluasi holistik ini memastikan bahwa sistem tidak hanya berfungsi dengan baik pada tingkat komponen individual, tetapi juga memberikan pengalaman pengguna yang koheren dan memuaskan.

II.6.3 *Customer Churn Prediction* dengan *Machine Learning*

Prediksi *customer churn* adalah aplikasi pembelajaran mesin yang kritis dalam analisis bisnis pelanggan, memungkinkan organisasi untuk mengidentifikasi pelanggan berisiko dan mengambil tindakan retensi proaktif. Integrasi prediksi *churn* ke dalam sistem *chatbot* BI memberikan aksesibilitas yang mudah terhadap *insight* prediktif bagi pengguna internal.

II.6.3.1 *Model Machine Learning untuk Churn Prediction*

Beberapa model pembelajaran mesin telah terbukti efektif untuk prediksi *customer churn* (Pecan AI 2024; Kumar dan Zafar 2024). Model *Logistic Regression* dikenal karena kesederhanaannya dan kemudahan interpretasi, cocok untuk kasus dengan hubungan linear antara fitur dan *churn*. Model *Random Forest* memberikan akurasi superior dalam skenario kompleks dengan kemampuan menangani hubungan non-linear dan interaksi antar variabel. Model *Gradient Boosting Machines* (GBM) dan XGBoost sangat efektif dalam menangkap hubungan kompleks dan non-linear dengan membangun pohon secara sekuensial untuk memperbaiki kesalahan.

Studi terbaru menunjukkan bahwa XGBoost adalah algoritma paling efisien untuk prediksi *churn* pelanggan (Kumar dan Zafar 2024). Model hibrida berbasis jaringan saraf yang mengkombinasikan *Multi-Head Self-Attention*, BiLSTM, dan CNN telah diusulkan untuk meningkatkan ekstraksi fitur kompleks dan dependensi temporal (Li dkk. 2024). Model CCP-Net ini menggunakan algoritma sampling ADA-SYN untuk menyeimbangkan ukuran sampel pelanggan yang *churn* dan tidak *churn*, mengatasi dampak negatif ketidakseimbangan sampel pada performa model.

II.6.3.2 *Evaluasi Model Churn Prediction*

Evaluasi model prediksi *churn* menggunakan metrik standar pembelajaran mesin dengan pertimbangan khusus untuk ketidakseimbangan kelas. Metrik evaluasi mencakup *Accuracy*, *Precision*, *Recall*, *F1-Score*, dan *ROC-AUC*. Studi Sharma, Gupta, dan Singh (2025) menunjukkan bahwa *Random Forest* mencapai akurasi 90,30% dengan semua atribut dan 90,90% dengan *dataset* atribut yang dikurangi, mengindikasikan bahwa *dataset* dengan atribut yang dikurangi dapat berguna untuk tugas prediksi *churn*.

Studi lain menunjukkan bahwa *Random Forest* mencapai akurasi 90,30%, presisi 89,5%, *recall* 91,2%, dan *ROC-AUC* 0,94 (Rahmat dan Setiawan 2024). Model dengan *hyperparameter tuning* menggunakan teknik seperti SMOTE, ENN, dan K-

fold cross-validation menghasilkan peningkatan performa yang signifikan. Penggunaan *dataset* dengan atribut yang dikurangi tidak hanya meningkatkan akurasi tetapi juga mengurangi kompleksitas komputasi, menjadikannya lebih praktis untuk deployment produksi.

II.6.3.3 Fitur-fitur untuk *Churn Prediction*

Fitur-fitur yang umum digunakan untuk prediksi *churn* mencakup karakteristik demografis dan behavioral pelanggan (Pecan AI 2024; Sharma, Gupta, dan Singh 2025). Karakteristik demografis meliputi usia, jenis kelamin, lokasi geografis, dan tingkat pendapatan. Karakteristik behavioral mencakup *tenure* atau lama berlangganan, frekuensi transaksi, nilai transaksi rata-rata, variasi produk yang dibeli, dan pola penggunaan layanan.

Fitur tambahan yang relevan adalah riwayat keluhan atau masalah layanan, tingkat keterlibatan dengan program loyalitas, responsivitas terhadap kampanye pemasaran, dan perubahan pola penggunaan dari waktu ke waktu. *Feature engineering* yang efektif, termasuk pembuatan fitur agregat temporal dan fitur interaksi, dapat meningkatkan performa model secara signifikan.

II.6.3.4 *Real-Time Churn Prediction*

Pendekatan tradisional prediksi *churn* menggunakan pelatihan berbasis *batch* dengan *dataset* tetap yang dikumpulkan pada interval periodik. Pendekatan ini gagal menangkap sifat dinamis industri dengan preferensi pelanggan yang berubah cepat (Al-Amri, Murugesan, dkk. 2025). Model RCE (*Real-time Continual learning-based Enhancement*) telah diusulkan untuk mengatasi keterbatasan ini dengan memanfaatkan pendekatan pengembangan *event-driven* untuk prediksi *churn* waktu nyata.

RCE menggunakan mekanisme *continual learning* berbasis *replay*, memungkinkan beradaptasi dengan perilaku pelanggan baru sambil mengurangi efek *catastrophic forgetting*. Berbeda dengan model berbasis *batch* tradisional, RCE memproses aliran data yang berkelanjutan, memungkinkan bisnis untuk bereaksi dengan cepat terhadap perubahan pasar (Al-Amri, Murugesan, dkk. 2025). Pendekatan ini memastikan bahwa pembuat keputusan menerima *insight* terkini ketika perilaku pelanggan berkembang, meningkatkan efektivitas strategi retensi.

II.6.4 Integrasi *Churn Prediction* dengan *Chatbot BI*

Integrasi model prediksi *churn* ke dalam sistem *chatbot* Business Intelligence memungkinkan pengguna internal untuk dengan mudah mengakses analisis risiko *churn* melalui antarmuka percakapan yang intuitif. Pengguna dapat mengajukan kueri seperti "Tampilkan pelanggan berisiko tinggi *churn* bulan ini" atau "Analisis faktor utama yang mempengaruhi *churn* di segmen premium", dan sistem akan mengeksekusi model prediksi serta menyajikan hasil dalam format bahasa alami yang mudah dipahami.

Integrasi ini memfasilitasi pengambilan keputusan proaktif untuk retensi pelanggan, seperti penawaran promosi khusus atau peningkatan layanan untuk pelanggan berisiko tinggi. Sistem dapat menghasilkan rekomendasi tindakan berdasarkan faktor-faktor yang berkontribusi terhadap risiko *churn* individu pelanggan. *Dashboard* interaktif dapat menampilkan distribusi risiko *churn* berdasarkan segmen pelanggan, tren *churn* dari waktu ke waktu, dan efektivitas kampanye retensi yang telah dilaksanakan.

Aksesibilitas informasi prediktif melalui antarmuka percakapan menurunkan *barrier* untuk adopsi analitik prediktif di kalangan pengguna non-teknis. Pengguna bisnis tanpa latar belakang *data science* dapat memanfaatkan model pembelajaran mesin yang canggih untuk mendukung keputusan mereka tanpa perlu memahami detail teknis implementasi model. Demokratisasi akses terhadap *insight* prediktif ini meningkatkan *data literacy* organisasi dan mempercepat transformasi menuju organisasi *data-driven*.

BAB III

ANALISIS MASALAH

Bab ini menyajikan analisis mendalam terhadap kondisi sistem *Business Intelligence* saat ini, identifikasi masalah pengguna, kebutuhan fungsional dan nonfungsional sistem, serta analisis pemilihan solusi teknologi yang paling tepat. Pembahasan mencakup evaluasi sistem *Business Intelligence* yang ada, identifikasi tantangan dalam pengambilan keputusan berbasis data, dan justifikasi pemilihan pendekatan *chatbot* berbasis aturan kueri, pemrosesan bahasa alami, dan peramalan deret waktu sebagai solusi yang paling sesuai dengan kebutuhan organisasi modern.

III.1 Analisis Kondisi Saat Ini

Analisis permasalahan yang dikaji dalam penelitian ini menggunakan pendekatan CRISP-DM (*Cross-Industry Standard Process for Data Mining*). Metodologi ini terdiri dari enam tahapan utama, yaitu: pemahaman bisnis (*business understanding*), pemahaman data (*data understanding*), penyiapan data (*data preparation*), pemodelan (*modeling*), evaluasi (*evaluation*), dan penerapan (*deployment*). Masing-masing tahapan dijelaskan lebih lanjut dalam bab ini beserta kegiatan-kegiatan yang dilakukan pada setiap fase, untuk menunjukkan bagaimana sistem dikembangkan secara terstruktur guna menjawab kebutuhan analisis dan prediksi kinerja bisnis secara otomatis dan responsif.

Permasalahan besar yang diangkat dalam penelitian ini adalah ketidakefisienan dan kurangnya responsivitas dalam proses analisis data bisnis untuk pengambilan keputusan strategis. Banyak organisasi, terutama perusahaan skala menengah dan besar, masih mengandalkan sistem *business intelligence* berbasis dasbor visual statis dan pelaporan manual yang memerlukan intervensi tim teknologi informasi atau analis data untuk setiap permintaan analisis berdasarkan kebutuhan khusus. Proses ini tidak hanya memakan waktu, tetapi juga berpotensi menimbulkan keterlambatan

dalam pengambilan keputusan yang berdampak pada daya saing perusahaan dan kemampuan merespons dinamika bisnis secara proaktif.

Kekurangan dari sistem *business intelligence* konvensional ini sebagian besar bersumber dari beberapa faktor kritis. Pertama, keterbatasan akses analitik berbasis percakapan yang menyebabkan pengguna internal tidak dapat memperoleh wawasan terkait indikator bisnis pelanggan seperti pesanan, pencapaian target penjualan, tingkat kepergian pelanggan, serta pendapatan secara cepat dan intuitif. Kedua, ketiadaan fitur prediksi yang mudah diakses oleh pengguna nonteknis, sehingga perencanaan kapasitas, evaluasi target, dan penetapan strategi bisnis menjadi kurang responsif terhadap tren dan dinamika masa depan. Ketiga, minimnya integrasi antara aturan kueri berbasis dan pemrosesan bahasa alami dalam sistem BI internal, yang menyebabkan proses pencarian data, penyaringan metrik utama, serta penyusunan laporan kinerja bisnis masih harus dilakukan secara manual dan berulang. Keempat, kurangnya analitik interaktif dan wawasan prediktif waktu nyata melalui antarmuka percakapan yang mudah diakses oleh seluruh pengguna internal.

Kondisi sistem saat ini dapat dijabarkan lebih rinci sebagai berikut:

1. **Responsivitas Terbatas**

Dasbor statis hanya menampilkan metrik yang telah didefinisikan sebelumnya dan tidak dapat menjawab pertanyaan mendadak atau spesifik dari pengguna bisnis. Pengguna harus menunggu tim teknologi informasi untuk membuat kueri khusus atau laporan baru, yang dapat memakan waktu beberapa jam hingga beberapa hari.

2. **Ketergantungan pada Tim Teknis**

Setiap permintaan kueri khusus atau analisis berdasarkan kebutuhan khusus harus diproses oleh tim teknologi informasi atau analis data, menciptakan kemacetan dan memperlambat waktu respons. Hal ini menghambat kemampuan organisasi untuk mengambil keputusan secara cepat dan tepat waktu.

3. **Kurangnya Kemampuan Prediktif**

Prediksi tren kinerja bisnis seperti volume penjualan, tingkat kepergian pelanggan, atau proyeksi pendapatan masih dilakukan secara manual menggunakan lembar sebar atau alat sederhana lainnya, tanpa memanfaatkan model peramalan yang canggih berbasis pembelajaran mesin atau peramalan deret waktu.

4. **Akses yang Tidak Fleksibel**

Sistem *business intelligence* umumnya hanya dapat diakses melalui komputer pribadi dengan antarmuka khusus yang memerlukan pengetahuan teknis,

tidak mendukung akses perangkat genggam atau antarmuka percakapan yang intuitif bagi pengguna nonteknis.

5. Integrasi Data Terbatas

Data berasal dari berbagai sumber seperti sistem manajemen hubungan pelanggan, sistem perencanaan sumber daya perusahaan, dan basis data transaksi, namun belum terintegrasi sepenuhnya dalam satu mesin kueri terpadu dan mudah diakses melalui antarmuka bahasa alami.

6. Ketiadaan Antarmuka Percakapan

Sistem *business intelligence* saat ini tidak menyediakan antarmuka percakapan yang memungkinkan pengguna berinteraksi dengan data menggunakan bahasa wajar, sehingga pengguna harus memahami struktur data dan terminologi teknis untuk dapat mengakses informasi yang dibutuhkan.

Permasalahan ini dapat diatasi melalui implementasi sistem *business intelligence* berbasis *chatbot* yang mengintegrasikan teknologi aturan kueri berbasis, pemrosesan bahasa alami, dan algoritma peramalan deret waktu. Pendekatan ini memungkinkan pengguna internal untuk mengajukan pertanyaan dalam bahasa wajar, mendapatkan respons yang cepat dan akurat, serta memperoleh prediksi tren bisnis secara otomatis tanpa memerlukan keahlian teknis khusus.

Pada penelitian ini, penulis memilih untuk mengimplementasikan klasifikasi maksud berbasis pencocokan pola dengan mekanisme nilai keyakinan untuk mengidentifikasi maksud pengguna dari kueri bahasa wajar, pembangkitan bahasa alami berbasis templat untuk menghasilkan respons naratif yang beragam namun tetap aman dan terkontrol, serta model peramalan deret waktu (ARIMA, SARIMA, atau LSTM) untuk menghasilkan prediksi terhadap indikator bisnis pelanggan seperti volume pesanan, pencapaian target penjualan, tingkat kepergian pelanggan, dan proyeksi pendapatan.

Oleh karena itu, berdasarkan rumusan masalah yang telah dijelaskan pada Bab I, penelitian ini bertujuan untuk merancang dan mengimplementasikan sebuah sistem *business intelligence* berbasis *chatbot* yang mampu memfasilitasi pengguna internal dalam menganalisis dan memprediksi indikator bisnis pelanggan secara interaktif dan waktu nyata. Sistem ini diharapkan dapat mengurangi beban kerja manual, meningkatkan responsivitas pengambilan keputusan, serta memberikan solusi terintegrasi yang siap diterapkan dalam praktik bisnis riil untuk meningkatkan daya saing organisasi di era digital.

III.2 Analisis Kebutuhan

Pada tahap awal pemahaman masalah dan solusi, analisis kebutuhan menjadi langkah dasar untuk memahami fitur dan karakteristik yang harus dimiliki oleh sistem agar dapat berfungsi sesuai tujuan. Dalam konteks tugas akhir ini, sistem dirancang untuk mengotomatisasi proses analisis dan prediksi kinerja bisnis berbasis data pelanggan menggunakan teknologi kecerdasan buatan, khususnya pemrosesan bahasa alami, klasifikasi maksud berbasis pola, dan peramalan deret waktu.

Analisis kebutuhan dibagi menjadi tiga kategori utama, yaitu identifikasi masalah pengguna, kebutuhan fungsional, dan kebutuhan nonfungsional. Identifikasi masalah pengguna berfokus pada pemahaman mendalam terhadap tantangan yang dihadapi oleh pengguna internal dalam mengakses dan menganalisis data bisnis. Kebutuhan fungsional berfokus pada fitur utama yang harus dimiliki sistem agar dapat menjalankan fungsinya secara efektif, seperti kemampuan memproses pertanyaan bahasa wajar, mengklasifikasikan maksud, mengeksekusi kueri ke gudang data, menghasilkan respons naratif, serta melakukan peramalan otomatis. Sementara itu, kebutuhan nonfungsional mencakup aspek teknis dan kualitas sistem, seperti keamanan data, performa, skalabilitas, ketersediaan, dan kemudahan penggunaan.

Pemahaman yang mendalam terhadap ketiga aspek kebutuhan ini menjadi fondasi penting dalam proses desain, implementasi, dan evaluasi sistem. Berikut ini adalah analisis kebutuhan yang telah disusun berdasarkan tujuan sistem dan karakteristik pengguna.

III.2.1 Identifikasi Masalah Pengguna

Pengguna dari sistem yang dikembangkan dalam penelitian ini adalah pengguna internal perusahaan, yang meliputi manajer, supervisor, analis bisnis, dan staf operasional yang terlibat dalam pengelolaan dan analisis data kinerja bisnis terkait pelanggan. Pengguna ini berasal dari berbagai departemen seperti penjualan, layanan pelanggan, pemasaran, dan manajemen strategis yang membutuhkan akses cepat terhadap wawasan bisnis untuk mendukung pengambilan keputusan harian maupun strategis.

Saat ini, mayoritas pengguna internal masih mengandalkan sistem *business intelligence* konvensional yang berbasis dasbor visual statis dan pelaporan manual. Pendekatan ini memiliki beberapa keterbatasan signifikan yang menghambat efektivitas pengambilan keputusan:

1. **Keterlambatan Respons terhadap Pertanyaan Bisnis**

Pengguna sering mengalami keterlambatan dalam memperoleh jawaban atas pertanyaan bisnis yang mendesak. Ketika membutuhkan wawasan spesifik yang tidak tersedia di dasbor yang ada, seperti perbandingan performa penjualan antar segmen pelanggan atau analisis tren kepergian pada periode tertentu, pengguna harus menunggu tim teknologi informasi atau analis data untuk membuat kueri khusus. Waktu tunggu ini dapat berkisar antara beberapa jam hingga beberapa hari, tergantung pada kompleksitas pertanyaan dan beban kerja tim teknis. Keterlambatan ini mengakibatkan pengambilan keputusan menjadi terhambat dan peluang bisnis yang peka waktu dapat terlewatkan.

2. **Kesulitan Mengakses Analisis Berdasarkan Kebutuhan Khusus**

Banyak pertanyaan bisnis yang bersifat berdasarkan kebutuhan khusus atau unik tidak dapat dijawab oleh dasbor statis yang telah dikonfigurasi sebelumnya. Pengguna harus menjelaskan kebutuhan analisis mereka secara detail kepada tim teknologi informasi melalui surat elektronik, sistem tiket, atau komunikasi langsung, yang kemudian memerlukan waktu bagi tim teknis untuk memahami persyaratan, membuat kueri yang sesuai, dan menghasilkan laporan. Proses ini tidak efisien, memakan waktu, dan rentan terhadap kesalahpahaman antara pengguna bisnis dan tim teknis yang dapat menyebabkan hasil analisis tidak sesuai dengan kebutuhan.

3. **Ketiadaan Kemampuan Prediksi Otomatis**

Analisis prediktif untuk tren kinerja bisnis seperti proyeksi volume penjualan, prediksi tingkat kepergian pelanggan, atau estimasi pencapaian target pendapatan masih dilakukan secara manual menggunakan lembar sebar atau alat analisis sederhana. Pengguna harus mengunduh data historis, melakukan perhitungan manual, atau meminta tim ilmu data untuk membuat prediksi khusus. Proses ini tidak dapat diperluas, memakan waktu, dan prediksi yang dihasilkan tidak selalu terkini mengikuti perubahan data terbaru. Ketiadaan prediksi otomatis membuat organisasi kurang siap dalam mengantisipasi tren atau risiko bisnis yang dapat berdampak pada strategi dan alokasi sumber daya.

4. **Akses yang Tidak Ramah Pengguna Nonteknis**

Dasbor *business intelligence* yang tersedia sering kali dirancang dengan asumsi bahwa pengguna memiliki pengetahuan teknis tentang struktur data, dimensi, metrik, dan terminologi basis data. Pengguna nonteknis, seperti manajer atau staf operasional yang tidak memiliki latar belakang teknis, sering merasa kesulitan dan kewalahan dengan berbagai tapis, daftar tarik, dan opsi yang ter-

sedia. Antarmuka tidak intuitif, tidak mendukung kueri berbasis bahasa wajar, dan memerlukan kurva pembelajaran yang tinggi. Hal ini menyebabkan banyak pengguna potensial tidak memanfaatkan sistem *business intelligence* secara optimal atau bahkan menghindari penggunaannya sama sekali.

5. Fragmentasi Akses Wawasan

Wawasan bisnis yang dibutuhkan untuk pengambilan keputusan tersebar di berbagai sistem, dasbor, laporan, dan alat yang berbeda. Pengguna harus membuka berbagai aplikasi atau modul untuk mendapatkan gambaran lengkap tentang kinerja bisnis, kemudian mengintegrasikan data secara manual di lembar sebar atau membuat dokumen analisis mereka sendiri. Fragmentasi ini meningkatkan risiko ketidakkonsistenan data, kesalahan manusia dalam proses konsolidasi, dan redundansi pekerjaan karena pengguna yang berbeda mungkin melakukan analisis yang sama secara terpisah tanpa berbagi hasil.

Tantangan-tantangan ini menghadirkan kebutuhan mendesak untuk mengembangkan sistem *business intelligence* yang lebih responsif, intuitif, dan terintegrasi, dengan kemampuan *chatbot* untuk antarmuka percakapan yang ramah pengguna, klasifikasi maksud yang cerdas untuk pengalihan kueri yang efisien, pembangkitan respons naratif yang beragam dan mudah dipahami, serta peramalan otomatis untuk prediksi kinerja bisnis. Sistem yang diusulkan diharapkan dapat mengatasi keterbatasan tersebut dan memberikan nilai tambah signifikan bagi organisasi dalam meningkatkan kecepatan, akurasi, dan kualitas pengambilan keputusan berbasis data.

III.2.2 Kebutuhan Fungsional

Kebutuhan fungsional menggambarkan fitur-fitur utama yang harus dimiliki oleh sistem *business intelligence* berbasis *chatbot* agar dapat memenuhi tujuan penelitian dan mengatasi masalah pengguna yang telah diidentifikasi. Setiap kebutuhan fungsional dirancang untuk mendukung proses analisis dan prediksi kinerja bisnis secara otomatis, responsif, dan mudah diakses oleh pengguna internal dengan berbagai tingkat keahlian teknis.

Berikut merupakan daftar kebutuhan fungsional yang telah disusun berdasarkan analisis mendalam terhadap proses bisnis, alur interaksi pengguna, dan arsitektur sistem yang diusulkan:

Tabel III.1 Kebutuhan Fungsional Sistem *Chatbot* BI

ID	Nama Kebutuhan	Deskripsi
FR-1	Penerimaan Pertanyaan Bahasa Alami	<p>Tujuan: Memungkinkan pengguna mengajukan pertanyaan atau permintaan analisis dalam bahasa Indonesia atau Inggris secara wajar melalui antarmuka <i>chatbot</i>.</p> <p>Masukan: Teks pertanyaan dari pengguna dalam bentuk bahasa wajar (contoh: "Berapa total penjualan bulan ini?" atau "Prediksi tingkat kepergian pelanggan 3 bulan ke depan").</p> <p>Keluaran: Pertanyaan tersimpan dan diproses oleh sistem untuk tahap berikutnya.</p>
FR-2	Klasifikasi Maksud dan Pemberian Nilai Keyakinan	<p>Tujuan: Mengidentifikasi maksud utama pengguna dari pertanyaan menggunakan pencocokan pola dan menghitung nilai keyakinan untuk menentukan tingkat kepercayaan sistem.</p> <p>Masukan: Teks pertanyaan hasil pemrosesan awal.</p> <p>Keluaran: Maksud teridentifikasi beserta nilai keyakinan (0-1) dan parameter terkait (entitas, dimensi waktu, metrik).</p>
FR-3	Pengalihan Kueri Berdasarkan Tingkat Keyakinan	<p>Tujuan: Mengarahkan kueri ke modul pemrosesan yang sesuai berdasarkan nilai keyakinan klasifikasi maksud.</p> <p>Masukan: Maksud dan nilai keyakinan.</p> <p>Keluaran: Jika nilai keyakinan $\geq 0,70$: kueri diteruskan ke mesin kueri berbasis aturan. Jika nilai keyakinan $< 0,50$: kueri diteruskan ke modul pemrosesan bahasa alami lanjutan atau permintaan klarifikasi ke pengguna.</p>

Bersambung ke halaman berikutnya

Tabel III.1 – Kebutuhan Fungsional Sistem *Chatbot* BI (lanjutan)

ID	Nama Kebutuhan	Deskripsi
FR-4	Eksekusi Kueri ke Gudang Data	<p>Tujuan: Mengambil data relevan dari gudang data atau mart data internal berdasarkan maksud dan parameter yang telah diidentifikasi.</p> <p>Masukan: Maksud, entitas, dan parameter kueri.</p> <p>Keluaran: Data hasil kueri dalam format terstruktur (JSON, tabel).</p>
FR-5	Penciptaan Respons Naratif Berbasis Templat	<p>Tujuan: Menghasilkan respons dalam bentuk teks naratif yang mudah dipahami menggunakan pembangkitan bahasa alami berbasis templat.</p> <p>Masukan: Data hasil kueri dan konteks percakapan.</p> <p>Keluaran: Teks respons naratif yang beragam dan wajar (contoh: "Penjualan pada bulan Oktober mencapai Rp 5,2 miliar, meningkat 15% dari bulan sebelumnya").</p>
FR-6	Dukungan Penggalian Data dan Analisis Perbandingan	<p>Tujuan: Memungkinkan pengguna melakukan eksplorasi data lebih dalam dari tingkat ringkasan ke tingkat detail, serta membandingkan metrik antar periode, segmen, atau dimensi.</p> <p>Masukan: Permintaan penggalian data atau perbandingan dari pengguna.</p> <p>Keluaran: Data detail atau hasil perbandingan dengan statistik perubahan (delta, persentase, tren).</p>
FR-7	Peramalan Deret Waktu	<p>Tujuan: Menghasilkan prediksi otomatis terhadap indikator kinerja bisnis menggunakan model ARIMA, SARIMA, atau LSTM.</p>

Bersambung ke halaman berikutnya

Tabel III.1 – Kebutuhan Fungsional Sistem *Chatbot* BI (lanjutan)

ID	Nama Kebutuhan	Deskripsi
		<p>Masukan: Data historis dan cakrawala prediksi yang diminta (harian, mingguan, bulanan).</p> <p>Keluaran: Nilai prediksi, selang kepercayaan, tren, dan visualisasi grafik.</p>
FR-8	Manajemen Riwayat Percakapan dan Sesi	<p>Tujuan: Menyimpan riwayat percakapan pengguna untuk memudahkan pelacakan dan konteks percakapan yang berkelanjutan.</p> <p>Masukan: Interaksi pengguna dengan sistem.</p> <p>Keluaran: Riwayat percakapan tersimpan dan dapat diakses kembali oleh pengguna.</p>
FR-9	Autentikasi dan Otorisasi Pengguna	<p>Tujuan: Memastikan hanya pengguna yang berwenang dapat mengakses sistem dan data sesuai dengan peran mereka.</p> <p>Masukan: Kredensial pengguna (nama pengguna, sandi).</p> <p>Keluaran: Akses sistem berdasarkan kendali akses berbasis peran.</p>
FR-10	Mekanisme Pengunduran dan Penanganan Kesalahan	<p>Tujuan: Menangani kueri yang tidak dapat dipahami atau diproses dengan memberikan respons alternatif atau permintaan klarifikasi.</p> <p>Masukan: Kueri dengan nilai keyakinan rendah atau kesalahan pemrosesan.</p> <p>Keluaran: Pesan klarifikasi, saran pertanyaan alternatif, atau eskalasi ke dukungan manusia.</p>

Kebutuhan-kebutuhan fungsional di atas dirancang untuk menciptakan sistem yang komprehensif, responsif, dan mudah digunakan, yang mampu mengotomatisasi sebagian besar tugas analisis dan prediksi bisnis, mengurangi ketergantungan pada tim teknis, serta memberdayakan pengguna internal untuk mengakses wawasan yang

mereka butuhkan dengan cepat dan efisien.

III.2.3 Kebutuhan Nonfungsional

Kebutuhan nonfungsional menggambarkan karakteristik kualitas sistem yang harus dipenuhi untuk memastikan sistem dapat beroperasi dengan baik, aman, andal, dan memberikan pengalaman pengguna yang optimal. Kebutuhan nonfungsional mencakup aspek-aspek seperti performa, keamanan, ketersediaan, skalabilitas, kemudahan penggunaan, dan pemeliharaan sistem.

Berikut merupakan daftar kebutuhan nonfungsional yang telah disusun untuk sistem *Business Intelligence* berbasis *chatbot*:

Tabel III.2 Kebutuhan Nonfungsional Sistem *Chatbot* BI

ID	Nama Kebutuhan	Deskripsi
NF-1	Keamanan Data dan Privasi	Sistem harus memastikan keamanan data internal perusahaan dengan menerapkan enkripsi data saat dalam perjalanan (TLS/SSL) dan saat penyimpanan. Sistem harus mengimplementasikan kendali akses berbasis peran yang ketat untuk memastikan pengguna hanya dapat mengakses data sesuai dengan peran dan wewenang mereka. Sistem harus mencegah kebocoran data sensitif, suntikan perintah berbahaya, atau akses tidak sah. Pencatatan dan jejak audit untuk semua aktivitas kueri dan akses data harus diimplementasikan untuk kepatuhan dan analisis forensik.

Bersambung ke halaman berikutnya

Tabel III.2 – Kebutuhan Nonfungsional Sistem *Chatbot* BI (lanjutan)

ID	Nama Kebutuhan	Deskripsi
NF-2	Ketersediaan Sistem	Sistem harus memiliki tingkat ketersediaan minimal 99% agar dapat diakses kapan saja oleh pengguna internal. Sistem harus dilengkapi dengan mekanisme redundansi dan pengalihan otomatis untuk memastikan kontinuitas layanan. Jika terjadi kesalahan pada saat eksekusi kueri atau penciptaan respons, sistem harus memiliki mekanisme pengunduran yang elegan dan penanganan kesalahan yang memadai serta memberikan pesan kesalahan yang informatif kepada pengguna.
NF-3	Performa dan Waktu Respons	Sistem harus mampu memberikan waktu respons maksimal 2 detik untuk 95% dari kueri tipikal. Untuk kueri kompleks yang memerlukan peramalan atau analisis data besar, waktu respons maksimal adalah 5 detik. Sistem harus dapat memproses minimal 100 kueri per menit pada kondisi beban normal dan hingga 1000 kueri per jam pada kondisi beban puncak tanpa penurunan performa yang signifikan.
NF-4	Skalabilitas	Sistem harus mampu menangani peningkatan jumlah pengguna dan volume data secara horizontal dengan menambahkan server atau simpul tambahan. Infrastruktur harus dirancang dengan arsitektur layanan mikro atau penerapan terkontainer untuk memudahkan penskalaan. Basis data dan mesin kueri harus dioptimalkan untuk menangani volume data yang besar tanpa penurunan performa.

Bersambung ke halaman berikutnya

Tabel III.2 – Kebutuhan Nonfungsional Sistem *Chatbot* BI (lanjutan)

ID	Nama Kebutuhan	Deskripsi
NF-5	Kemudahan Penggunaan	Antarmuka <i>chatbot</i> harus intuitif dan mudah digunakan oleh pengguna nonteknis tanpa memerlukan pelatihan ekstensif. Sistem harus menyediakan panduan kontekstual dan saran pertanyaan untuk membantu pengguna merumuskan kueri. Riwayat percakapan dan manajemen sesi harus didukung untuk memudahkan pengguna melacak interaksi sebelumnya. Mekanisme umpan balik harus tersedia untuk perbaikan berkelanjutan berdasarkan masukan pengguna.
NF-6	Kemudahan Pemeliharaan	Sistem harus mudah dipelihara dengan dokumentasi kode, antarmuka program aplikasi, dan arsitektur yang lengkap dan jelas. Templat, aturan berbasis pola, dan pemetaan maksud harus mudah diperbarui oleh tim tanpa memerlukan penerapan ulang sistem atau waktu henti. Sistem harus dirancang secara modular sehingga mudah untuk menambahkan maksud baru, aturan baru, atau model peramalan baru di masa depan.

Bersambung ke halaman berikutnya

Tabel III.2 – Kebutuhan Nonfungsional Sistem *Chatbot* BI (lanjutan)

ID	Nama Kebutuhan	Deskripsi
NF-7	Kompatibilitas dan Aksesibilitas	Sistem harus dapat diakses dari berbagai platform (<i>web</i> , perangkat genggam, komputer pribadi) dan perangkat (ponsel pintar, tablet, komputer pribadi/laptop) dengan faktor bentuk yang berbeda. Antarmuka harus responsif dan mendukung berbagai ukuran layar. Sistem harus kompatibel dengan peramban umum seperti Chrome, Firefox, Edge, dan Safari, serta mendukung berbagai sistem operasi (Windows, macOS, Linux, iOS, Android).
NF-8	Keandalan	Sistem harus memiliki tingkat keandalan yang tinggi dengan kemampuan menangani kesalahan dan kondisi tidak normal tanpa berhenti. Sistem harus memiliki mekanisme pengulangan otomatis untuk proses yang gagal dan pencatatan komprehensif untuk analisis penyebab akar. Hasil prediksi dan analisis harus konsisten dan dapat direproduksi dengan data masukan yang sama.
NF-9	Integritas Data	Sistem harus memastikan bahwa data yang diambil, diproses, dan disajikan kepada pengguna adalah akurat dan tidak mengalami kerusakan. Hasil ekstraksi dari gudang data tidak boleh rusak, hilang, atau berubah saat disimpan, diproses, atau diekspor. Sistem harus mengimplementasikan validasi data dan penjumlahan cek untuk memastikan integritas data di seluruh saluran pemrosesan.

Bersambung ke halaman berikutnya

Tabel III.2 – Kebutuhan Nonfungsional Sistem *Chatbot* BI (lanjutan)

ID	Nama Kebutuhan	Deskripsi
NF-10	Ekstensibilitas	Sistem harus dirancang dengan arsitektur yang fleksibel dan modular untuk memungkinkan penambahan fitur baru di masa depan, seperti integrasi dengan sistem eksternal (ERP, CRM), penambahan model pembelajaran mesin baru, atau perluasan ke domain bisnis lainnya. Antarmuka program aplikasi yang terdokumentasi dengan baik harus disediakan untuk memfasilitasi integrasi dengan sistem lain.

Kebutuhan nonfungsional ini dirancang untuk memastikan bahwa sistem tidak hanya dapat menjalankan fungsi-fungsi yang diharapkan, tetapi juga memiliki kualitas teknis yang tinggi, aman, andal, dan memberikan pengalaman pengguna yang memuaskan. Pencapaian kebutuhan nonfungsional ini akan dievaluasi melalui pengujian performa, pengujian keamanan, pengujian beban, dan survei kepuasan pengguna pada tahap implementasi dan evaluasi sistem.

III.3 Analisis Pemilihan Solusi

Analisis permasalahan yang dikaji dalam penelitian ini menggunakan pendekatan kesesuaian masalah-solusi, yaitu dengan mencocokkan kebutuhan nyata yang terjadi di lapangan dengan solusi teknologi yang paling tepat guna. Pendekatan ini mempertimbangkan berbagai alternatif solusi yang tersedia, mengevaluasi kelebihan dan kekurangan masing-masing, serta memilih solusi yang paling optimal berdasarkan kriteria yang telah ditetapkan menggunakan metode *Balanced Scorecard*.

Pemilihan solusi yang tepat sangat penting untuk memastikan bahwa sistem yang dikembangkan dapat mengatasi masalah pengguna secara efektif, memberikan nilai tambah yang signifikan bagi organisasi, serta berkelanjutan untuk pengembangan jangka panjang. Analisis ini akan membantu memvalidasi bahwa solusi yang dipilih adalah yang paling sesuai dengan kondisi, kebutuhan, dan tujuan strategis organisasi.

III.3.1 Alternatif Solusi

Dalam rangka meningkatkan efisiensi analisis dan prediksi kinerja bisnis serta mengatasi keterbatasan sistem *business intelligence* konvensional, terdapat beberapa alternatif solusi teknologi yang dapat dipertimbangkan. Setiap alternatif memiliki karakteristik, kelebihan, dan kekurangan yang berbeda, yang perlu dievaluasi secara komprehensif untuk menentukan solusi yang paling optimal.

Berikut adalah tiga alternatif solusi utama yang telah diidentifikasi dan dianalisis berdasarkan studi literatur, praktik industri, dan kebutuhan spesifik organisasi:

Tabel III.3 Perbandingan Alternatif Solusi Teknologi

No	Nama Solusi	Deskripsi	Kelebihan	Kekurangan
1	<i>Chatbot</i> Berbasis Aturan Sederhana	Menggunakan <i>chatbot</i> dengan pencocokan pola yang sangat kaku, hanya merespons pola kalimat yang telah ditentukan sebelumnya. Sistem hanya dapat menangani kueri terstruktur dengan format tetap, tidak mampu memahami variasi bahasa wajar yang kompleks. Tidak ada model peramalan, hanya pengambilan data dari basis data.	1. Implementasi cepat dan hemat biaya 2. Mudah dipelihara untuk aturan sederhana 3. Tidak memerlukan infrastruktur kompleks 4. Kontrol penuh atas respons sistem	1. Tidak dapat diperluas untuk kebutuhan BI modern 2. Pengguna harus menyesuaikan dengan pola tetap 3. Tidak mampu menjawab kueri kompleks 4. Tidak ada kemampuan prediktif 5. Pengalaman pengguna terbatas

Bersambung ke halaman berikutnya

Tabel III.3 – Perbandingan Alternatif Solusi Teknologi (lanjutan)

No	Nama Solusi	Deskripsi	Kelebihan	Kekurangan
2	<i>Chatbot</i> BI dengan Klasifikasi Maksud Berbasis Pola, Pembangkitan Bahasa Alami Berbasis Templat, dan Peramalan Deret Waktu	Mengimplementasikan sistem BI <i>chatbot</i> yang terintegrasi dengan tiga komponen: klasifikasi maksud berbasis pola dengan pemberian nilai keyakinan, pembangkitan bahasa alami berbasis templat untuk respons yang beragam namun aman, dan model peramalan deret waktu (ARIMA, SARIMA, LSTM) untuk prediksi otomatis.	<ol style="list-style-type: none"> 1. Mampu menangani variasi bahasa wajar kompleks 2. Respons beragam dan wajar 3. Dukungan peramalan otomatis 4. Adaptif terhadap kebutuhan dinamis 5. Pengalaman pengguna unggul 6. Keamanan terjaga dengan templat 7. Dapat diperluas secara fleksibel 	<ol style="list-style-type: none"> 1. Kompleksitas implementasi lebih tinggi 2. Memerlukan usaha awal untuk desain 3. Pemeliharaan butuh keahlian berbagai bidang 4. Biaya infrastruktur lebih tinggi

Bersambung ke halaman berikutnya

Tabel III.3 – Perbandingan Alternatif Solusi Teknologi (lanjutan)

No	Nama Solusi	Deskripsi	Kelebihan	Kekurangan
3	Dasbor BI Interaktif Tradisional dengan Peningkatan	Tetap menggunakan dasbor BI visual yang sudah ada, dengan peningkatan untuk interaktivitas yang lebih baik. Pengguna dapat melakukan penggalian data, penyaringan, dan agregasi melalui kontrol antarmuka pengguna yang lebih intuitif. Menambahkan fitur analitik layanan mandiri dan aplikasi dasbor perangkat genggam.	<ol style="list-style-type: none"> 1. Teknologi matang dan terbukti 2. Dukungan komunitas dan alat kaya 3. Tata kelola data dan keamanan mapan 4. Akrab bagi organisasi dengan alat BI tradisional 	<ol style="list-style-type: none"> 1. Tidak mendukung kueri bahasa wajar 2. Tidak ada antarmuka percakapan 3. Dukungan terbatas untuk peramalan 4. Tetap memerlukan keahlian teknis 5. Tidak dapat diperluas untuk bisnis peka waktu

Bersambung ke halaman berikutnya

Tabel III.3 – Perbandingan Alternatif Solusi Teknologi (lanjutan)

No	Nama Solusi	Deskripsi	Kelebihan	Kekurangan
4	Chatbot BI Berbasis Model Bahasa Besar (Large Language Model)	Mengimplementasikan chatbot BI yang memanfaatkan arsitektur Transformer penuh dengan model bahasa besar pra-latih seperti GPT, BERT, atau varian lainnya. Sistem menggunakan pembelajaran mendalam untuk pemahaman bahasa alami yang sangat canggih, mampu menangani kueri kompleks dengan konteks multiturn, dan menghasilkan respons naratif yang sangat wajar. Model dapat disetel-halus (fine-tuned) dengan data domain spesifik untuk meningkatkan akurasi.	<ol style="list-style-type: none"> 1. Pemahaman bahasa alami paling canggih 2. Kemampuan konteks percakapan mendalam 3. Respons sangat wajar dan beragam 4. Dapat menangani kueri sangat kompleks 5. Adaptif terhadap variasi bahasa ekstrem 6. Pembelajaran berkelanjutan dari interaksi 	<ol style="list-style-type: none"> 1. Biaya komputasi sangat tinggi 2. Memerlukan infrastruktur GPU berkinerja tinggi 3. Risiko halusinasi dan respons tidak akurat 4. Sulit mengontrol keluaran secara deterministik 5. Potensi kebocoran data sensitif 6. Kompleksitas implementasi sangat tinggi 7. Ketergantungan pada penyedia model eksternal 8. Latensi respons lebih tinggi

Berikut adalah empat alternatif solusi utama yang telah diidentifikasi dan dianalisis berdasarkan studi literatur, praktik industri, dan kebutuhan fungsional dan non fungsional:

1. **Alternatif 1: *Chatbot* Berbasis Aturan Sederhana**

Solusi ini mengimplementasikan *chatbot* dengan pendekatan berbasis aturan yang sederhana, menggunakan pencocokan pola berbasis ekspresi reguler atau pencocokan langsung. Sistem hanya dapat merespons pertanyaan dengan format yang telah didefinisikan sebelumnya dan tidak memiliki kemampuan untuk memahami variasi bahasa atau konteks yang kompleks. Fitur yang tersedia terbatas pada pengambilan data sederhana dari basis data tanpa kemampuan analisis lanjutan atau prediksi. Antarmuka juga terbatas dan tidak dioptimasi untuk pengalaman pengguna yang baik pada perangkat genggam.

2. **Alternatif 2: *Chatbot* BI dengan Klasifikasi Maksud Berbasis Pola, Pembangunan Bahasa Alami Berbasis Templat, dan Peramalan Deret Waktu**

Solusi ini merupakan pendekatan komprehensif yang mengintegrasikan tiga komponen teknologi utama. Pertama, klasifikasi maksud berbasis pola dengan mekanisme pemberian nilai keyakinan untuk mengidentifikasi maksud pengguna dari kueri bahasa wajar dengan akurasi tinggi. Kedua, pembangunan bahasa alami berbasis templat untuk menghasilkan respons naratif yang variatif, wajar, dan mudah dipahami sambil menjaga keamanan data internal melalui pembatasan pada templat yang telah diverifikasi. Ketiga, model peramalan deret waktu (ARIMA, SARIMA, atau LSTM) untuk menghasilkan prediksi otomatis terhadap indikator kinerja bisnis seperti volume penjualan, tingkat kepergian pelanggan, dan proyeksi pendapatan. Sistem ini dirancang dengan arsitektur modular yang dapat diperluas dan diperbesar, mendukung akses berbagai platform (web, perangkat genggam, komputer pribadi), dan menerapkan kontrol keamanan yang ketat.

3. **Alternatif 3: Dasbor BI Interaktif Tradisional dengan Peningkatan**

Solusi ini mempertahankan pendekatan dasbor BI visual yang sudah ada dengan menambahkan peningkatan pada aspek interaktivitas dan kemudahan penggunaan. Fitur penggalan data lebih dalam, penyaringan lanjutan, dan pembuatan laporan khusus ditingkatkan dengan antarmuka yang lebih intuitif. Sistem ini juga dilengkapi dengan aplikasi dasbor perangkat genggam untuk akses data dari perangkat bergerak. Namun, solusi ini tidak menyediakan antarmuka percakapan berbasis bahasa wajar dan kemampuan prediktif yang otomatis.

4. **Alternatif 4: *Chatbot* BI Berbasis Model Bahasa Besar (*Large Language Model*)**

Berbeda dengan ketiga alternatif sebelumnya yang menggunakan pendekatan terstruktur atau semi-terstruktur, alternatif keempat ini memanfaatkan para-

digma pembelajaran mesin yang sepenuhnya berbasis data dengan arsitektur jaringan saraf tiruan yang sangat mendalam. Solusi ini mengimplementasikan *chatbot* BI yang memanfaatkan arsitektur *Transformer* penuh dengan model bahasa besar pra-latih seperti GPT, BERT, T5, atau varian lainnya. Sistem menggunakan pembelajaran mendalam untuk pemahaman bahasa alami yang sangat canggih, mampu menangani kueri kompleks dengan konteks percakapan multiturn, dan menghasilkan respons naratif yang sangat wajar menyerupai komunikasi manusia. Model dapat disetel-halus (*fine-tuned*) dengan data domain spesifik organisasi untuk meningkatkan akurasi dan relevansi respons terhadap konteks bisnis. Sistem ini memanfaatkan kemampuan pemodelan bahasa yang telah dilatih pada korpus teks masif untuk memahami nuansa, konteks implisit, dan melakukan inferensi kompleks dari pertanyaan pengguna. Namun, solusi ini memerlukan infrastruktur komputasi dengan GPU berkinerja tinggi untuk inferensi cepat, menimbulkan tantangan dalam hal kontrol deterministik terhadap keluaran, dan menghadapi risiko halusinasi di mana model dapat menghasilkan respons yang terdengar meyakinkan tetapi faktanya tidak akurat atau tidak berdasar pada data nyata. Implementasi juga memerlukan pertimbangan ketat terhadap keamanan data untuk mencegah kebocoran informasi sensitif melalui model, serta strategi mitigasi risiko untuk memastikan keandalan respons dalam konteks pengambilan keputusan bisnis kritis.

III.3.2 Analisis Penentuan Solusi

Dalam menentukan solusi terbaik untuk meningkatkan sistem *Business Intelligence* dan mengatasi masalah yang telah diidentifikasi, dilakukan analisis menggunakan metode *Balanced Scorecard* (BSC). Metode ini menilai masing-masing solusi dari empat perspektif utama yang saling melengkapi: *Financial*, *Customer*, *Internal Business Process*, dan *Learning and Growth*. Pendekatan holistik ini memastikan bahwa keputusan pemilihan solusi tidak hanya mempertimbangkan aspek biaya, tetapi juga nilai strategis jangka panjang, kepuasan pengguna, efisiensi proses bisnis, dan kapasitas inovasi organisasi.

Berikut adalah analisis mendalam untuk setiap solusi berdasarkan keempat perspektif *Balanced Scorecard*:

III.3.2.1 Solusi 1: *Chatbot* Berbasis Aturan Sederhana

Tabel III.4 Analisis *Balanced Scorecard* untuk Aturan Sederhana

Perspektif	Penilaian
<i>Financial</i>	Solusi ini memerlukan investasi awal yang rendah hingga sedang karena tidak membutuhkan infrastruktur komputasi yang kompleks atau model pembelajaran mesin yang canggih. Biaya operasional juga relatif rendah karena sistem dapat dijalankan pada pelanggan sederhana. Namun, pengembalian investasi jangka panjang rendah karena sistem tidak dapat diperluas dan akan memerlukan perancangan ulang besar ketika kebutuhan bisnis berkembang. Penghematan dari otomatisasi terbatas karena sistem hanya dapat menangani kueri yang sangat terstruktur dan sederhana, sehingga sebagian besar pekerjaan analisis masih memerlukan intervensi manual. Biaya peluang tinggi karena organisasi kehilangan peluang untuk mengoptimalkan pengambilan keputusan berbasis data secara cepat dan responsif.
<i>Customer</i> (Pengguna)	Pengalaman pengguna yang ditawarkan cukup terbatas. Pengguna harus menyesuaikan diri dengan format kueri yang kaku dan tidak fleksibel, yang tidak sesuai dengan cara berpikir dan berkomunikasi wajar mereka. Sistem tidak dapat memahami variasi pertanyaan atau konteks yang kompleks, sehingga sering kali pengguna harus mencoba beberapa kali dengan perumusan yang berbeda untuk mendapatkan hasil yang diinginkan. Hal ini menyebabkan frustrasi dan menurunkan tingkat kepuasan pengguna. Tingkat adopsi sistem cenderung rendah karena pengguna merasa sistem tidak memberikan nilai tambah yang signifikan dibandingkan dengan metode yang sudah ada.

Bersambung ke halaman berikutnya

Tabel III.4 – Analisis *Balanced Scorecard* untuk Aturan Sederhana (lanjutan)

Perspektif	Penilaian
<i>Internal Business Process</i>	Dari perspektif proses bisnis internal, solusi ini hanya menawarkan otomasi minimal untuk kueri yang sangat terstruktur dan berulang. Proses analisis manual masih tetap diperlukan untuk sebagian besar pertanyaan bisnis yang bersifat khusus atau kompleks, sehingga hambatan dalam alur analisis data tidak teratasi secara efektif. Waktu penyelesaian untuk mendapatkan wawasan bisnis tidak berkurang secara signifikan. Sistem juga tidak mendukung analisis lanjutan seperti penggalan data lebih dalam, perbandingan multi-dimensi, atau deteksi anomali, sehingga kemampuan organisasi untuk melakukan analisis mendalam tetap terbatas.
<i>Learning and Growth</i>	Solusi ini memberikan kesempatan yang sangat terbatas untuk pembelajaran dan pertumbuhan organisasi. Teknologi yang digunakan relatif sederhana dan tidak mendorong tim untuk mengembangkan keahlian baru dalam bidang kecerdasan buatan, pembelajaran mesin, atau analitik lanjutan. Organisasi tidak memperoleh keunggulan kompetitif teknologi dan tidak membangun kapabilitas yang dapat menjadi pembeda di pasar. Sistem juga tidak menyediakan landasan untuk inovasi masa depan atau integrasi dengan teknologi yang lebih canggih.

III.3.2.2 Solusi 2: *Chatbot* BI dengan Klasifikasi Maksud Berbasis Pola, Pembangunan Bahasa Alami Berbasis Templat, dan Peramalan Deret Waktu

Tabel III.5 Analisis *Balanced Scorecard* untuk BI Komprehensif

Perspektif	Penilaian
<i>Financial</i>	<p>Solusi ini memerlukan investasi awal yang lebih tinggi, mencakup biaya untuk pengembangan sistem yang kompleks, pengadaan infrastruktur komputasi yang memadai (termasuk kemungkinan GPU untuk model peramalan), dan pelatihan model kecerdasan buatan. Namun, pengembalian investasi jangka panjang sangat tinggi karena sistem memberikan nilai tambah yang signifikan melalui beberapa jalur: (1) Pengurangan dramatis dalam waktu analisis dari jam atau hari menjadi detik, yang memungkinkan pengambilan keputusan yang lebih cepat dan responsif; (2) Penghematan biaya tenaga kerja karena analis dan tim TI tidak perlu menghabiskan waktu untuk kueri rutin dan dapat fokus pada tugas-tugas yang lebih strategis; (3) Peningkatan kualitas keputusan bisnis berdasarkan wawasan prediktif yang akurat, yang dapat menghasilkan peningkatan pendapatan atau penghematan biaya yang signifikan; (4) Skalabilitas yang baik, dengan biaya tambahan untuk menambahkan pengguna atau kueri baru relatif rendah setelah sistem terbangun. Sistem ini juga memberikan keunggulan kompetitif yang dapat berdampak pada posisi pasar dan profitabilitas jangka panjang organisasi.</p>

Bersambung ke halaman berikutnya

Tabel III.5 – Analisis *Balanced Scorecard* untuk BI Komprehensif (lanjutan)

Perspektif	Penilaian
<i>Customer</i> (Pengguna)	Solusi ini menawarkan pengalaman pengguna yang sangat unggul. Antarmuka percakapan yang wajar dan akrab memungkinkan pengguna untuk mengajukan pertanyaan dengan cara yang sama seperti mereka berbicara dengan rekan manusia, tanpa perlu mempelajari tata bahasa khusus atau terminologi teknis. Sistem dapat memahami berbagai variasi pertanyaan dan konteks, sehingga mengurangi frustrasi pengguna. Respons yang dihasilkan bersifat naratif, beragam, dan mudah dipahami, bukan hanya tabel angka yang kering. Kemampuan prediktif otomatis memberikan nilai tambah yang signifikan bagi pengguna, memungkinkan mereka untuk mengantisipasi tren dan membuat keputusan proaktif. Aksesibilitas berbagai platform (web, perangkat genggam, komputer pribadi) memastikan pengguna dapat mengakses wawasan setiap membutuhkannya. Tingkat kepuasan dan adopsi pengguna diperkirakan sangat tinggi karena sistem memberikan nilai nyata dan mudah digunakan.

Bersambung ke halaman berikutnya

Tabel III.5 – Analisis *Balanced Scorecard* untuk BI Komprehensif (lanjutan)

Perspektif	Penilaian
<i>Internal Business Process</i>	<p>Dari perspektif proses bisnis internal, solusi ini menawarkan transformasi yang fundamental. Otomasi maksimal untuk hampir semua jenis kueri analitik, dari yang sederhana hingga yang kompleks, menghilangkan hambatan yang selama ini menghalangi alur analisis data. Proses bisnis dapat menjadi lebih lincah dan responsif terhadap perubahan kondisi pasar atau internal. Waktu penyelesaian untuk mendapatkan wawasan berkurang drastis, memungkinkan organisasi untuk mengidentifikasi masalah, peluang, atau tren lebih cepat dan mengambil tindakan yang tepat. Kemampuan peramalan otomatis mendukung perencanaan yang lebih akurat untuk alokasi sumber daya, penetapan target, dan strategi bisnis. Sistem juga memfasilitasi budaya berbasis data dengan keputusan dibuat berdasarkan wawasan objektif bukan hanya intuisi.</p>

Bersambung ke halaman berikutnya

Tabel III.5 – Analisis *Balanced Scorecard* untuk BI Komprehensif (lanjutan)

Perspektif	Penilaian
<i>Learning and Growth</i>	Solusi ini memberikan kesempatan tertinggi untuk pembelajaran dan pertumbuhan organisasi. Implementasi teknologi kecerdasan buatan seperti pemrosesan bahasa alami, klasifikasi maksud berbasis pola, pembangkitan bahasa alami berbasis templat, dan peramalan deret waktu memungkinkan tim untuk mengembangkan keahlian baru yang sangat relevan di era digital. Organisasi membangun kapabilitas internal dalam bidang kecerdasan buatan dan pembelajaran mesin yang dapat menjadi fondasi untuk inovasi-inovasi masa depan, seperti ekspansi ke domain lain (<i>chatbot</i> layanan pelanggan, sistem rekomendasi, deteksi penipuan). Sistem menciptakan keunggulan kompetitif teknologi yang membedakan organisasi dari kompetitor. Platform yang dibangun bersifat dapat diperluas dan dapat menjadi basis untuk pengembangan solusi analitik yang lebih canggih di masa depan. Budaya inovasi dan adopsi teknologi terdorong, memposisikan organisasi sebagai pemimpin dalam transformasi digital di industrinya.

III.3.2.3 Solusi 3: Dasbor BI Interaktif Tradisional dengan Peningkatan

Tabel III.6 Analisis *Balanced Scorecard* untuk BI Tradisional

Perspektif	Penilaian
<i>Financial</i>	Solusi ini memerlukan investasi yang sedang, terutama untuk meningkatkan antarmuka pengguna dan menambahkan fitur interaktivitas pada dasbor yang sudah ada. Biaya lisensi untuk alat BI komersial dapat cukup signifikan, terutama jika organisasi menggunakan platform seperti Tableau, Power BI, atau Qlik. Pengembalian investasi berada di tingkat menengah karena peningkatan efisiensi terbatas pada perbaikan antarmuka pengguna dan kemudahan akses, tanpa transformasi fundamental dalam cara pengguna berinteraksi dengan data. Penghematan biaya tenaga kerja terbatas karena pengguna masih memerlukan keahlian teknis untuk memanfaatkan fitur analitik layanan mandiri secara efektif. Tidak ada kemampuan prediktif otomatis yang dapat memberikan nilai tambah signifikan untuk perencanaan strategis.
<i>Customer</i> (Pengguna)	Pengalaman pengguna mengalami peningkatan dibandingkan dengan sistem lama, terutama dalam hal visualisasi data dan kemudahan navigasi. Dasbor yang lebih interaktif dengan fitur penggalian data lebih dalam dan penyaringan lanjutan memberikan fleksibilitas lebih besar bagi pengguna untuk mengeksplorasi data. Aplikasi perangkat genggam memungkinkan akses data dari perangkat bergerak, meningkatkan kemudahan akses. Namun, solusi ini masih tidak mendukung kueri bahasa wajar atau antarmuka percakapan, sehingga pengguna nonteknis tetap menghadapi kurva pembelajaran yang cukup tinggi. Sistem tidak dapat menjawab pertanyaan khusus yang kompleks tanpa konfigurasi tambahan dari tim TI. Kepuasan pengguna berada di tingkat menengah karena peningkatan yang diberikan bersifat bertahap, bukan transformatif.

Bersambung ke halaman berikutnya

Tabel III.6 – Analisis *Balanced Scorecard* untuk BI Tradisional (lanjutan)

Perspektif	Penilaian
<i>Internal Business Process</i>	Dari perspektif proses bisnis, solusi ini memberikan peningkatan yang sedang hingga tinggi dalam efisiensi operasional. Pengguna yang memiliki keahlian teknis dapat melakukan analisis sendiri tanpa selalu bergantung pada tim TI, mengurangi beban kerja analis. Fitur analitik layanan mandiri memungkinkan eksplorasi data yang lebih fleksibel untuk pengguna yang terlatih. Namun, untuk pengguna nonteknis atau pertanyaan yang sangat spesifik dan kompleks, sistem masih memerlukan intervensi manual dari tim analitik. Tidak ada dukungan untuk analisis prediktif otomatis, sehingga organisasi tidak dapat memanfaatkan kemampuan peramalan untuk perencanaan strategis yang lebih baik. Proses tetap tidak dapat diperluas untuk kebutuhan bisnis yang sangat dinamis dan peka waktu.
<i>Learning and Growth</i>	Solusi ini memberikan kesempatan sedang untuk pembelajaran dan pertumbuhan. Organisasi dapat meningkatkan literasi data tim melalui pelatihan penggunaan alat BI dan fitur analitik layanan mandiri. Penguasaan alat BI populer seperti Tableau atau Power BI dapat menjadi keahlian yang berharga bagi karyawan. Namun, solusi ini tidak mendorong inovasi teknologi yang signifikan atau pengembangan kapabilitas dalam bidang kecerdasan buatan dan pembelajaran mesin. Organisasi tidak membangun pembeda teknologi yang kuat dan tetap bergantung pada vendor eksternal untuk sebagian besar fitur dan inovasi. Platform ini juga terbatas dalam hal kemampuan perluasan untuk pengembangan solusi khusus atau integrasi dengan teknologi yang lebih canggih di masa depan.

Solusi 4: Chatbot BI Berbasis Model Bahasa Besar (Large Language Model)

Tabel III.7 Analisis *Balanced Scorecard* untuk LLM

Perspektif	Penilaian
<i>Financial</i>	<p>Solusi ini memerlukan investasi awal yang sangat tinggi, mencakup biaya lisensi atau akses API untuk model bahasa besar komersial (seperti GPT-4 atau Claude), pengadaan infrastruktur komputasi dengan GPU berkinerja tinggi untuk inferensi cepat, dan biaya pengembangan untuk integrasi sistem yang kompleks. Biaya operasional juga sangat signifikan karena setiap kueri memerlukan sumber daya komputasi yang besar, terutama untuk model dengan miliaran parameter. Untuk model berbasis API, biaya per token dapat terakumulasi dengan cepat pada skala organisasi besar. Pengembalian investasi berada pada tingkat sedang hingga tinggi dalam jangka panjang, bergantung pada seberapa baik organisasi dapat mengelola risiko halusinasi dan memastikan akurasi respons. Nilai tambah utama berasal dari kemampuan pemahaman bahasa yang sangat canggih dan fleksibilitas dalam menangani kueri kompleks yang tidak terstruktur. Namun, ketidakpastian dalam kontrol keluaran dan risiko kesalahan dapat membatasi pengembalian investasi jika tidak dikelola dengan baik. Biaya peluang juga perlu dipertimbangkan jika organisasi mengalami insiden kebocoran data atau respons yang tidak akurat yang berdampak pada keputusan bisnis.</p>

Bersambung ke halaman berikutnya

Tabel III.7 – Analisis *Balanced Scorecard* untuk LLM (lanjutan)

Perspektif	Penilaian
<i>Customer</i> (Pengguna)	<p>Solusi ini menawarkan pengalaman pengguna yang sangat unggul dalam hal kewajaran interaksi dan fleksibilitas komunikasi. Pengguna dapat mengajukan pertanyaan dengan cara yang sangat wajar, termasuk kueri yang kompleks, ambigu, atau memerlukan pemahaman konteks mendalam yang sulit ditangani sistem konvensional. Model bahasa besar dapat memahami nuansa, konteks percakapan multiturn, dan bahkan inferensi implisit dari pertanyaan pengguna. Respons yang dihasilkan sangat naratif, koheren, dan menyerupai komunikasi manusia, yang meningkatkan kepuasan pengguna secara signifikan. Namun, terdapat risiko bahwa sistem dapat menghasilkan respons yang terdengar meyakinkan tetapi faktanya tidak akurat (halusinasi), yang dapat menyesatkan pengguna dan menurunkan kepercayaan terhadap sistem. Pengguna nonteknis mungkin tidak dapat membedakan respons yang akurat dari yang dihasilkan secara spekulatif. Untuk mempertahankan kepuasan tinggi, sistem memerlukan mekanisme validasi dan verifikasi yang ketat, serta transparansi mengenai tingkat keyakinan respons.</p>

Bersambung ke halaman berikutnya

Tabel III.7 – Analisis *Balanced Scorecard* untuk LLM (lanjutan)

Perspektif	Penilaian
<i>Internal Business Process</i>	<p>Dari perspektif proses bisnis internal, solusi ini memberikan fleksibilitas tertinggi dalam menangani berbagai jenis kueri analitik, termasuk yang sangat kompleks, tidak terstruktur, atau belum pernah ditemui sebelumnya. Kemampuan model untuk memahami konteks yang dalam dan melakukan inferensi dapat mengurangi kebutuhan untuk konfigurasi manual atau pemrograman khusus untuk kasus penggunaan baru. Sistem dapat beradaptasi dengan perubahan kebutuhan bisnis dengan lebih organik. Namun, ketidakdeterministikan keluaran model bahasa besar menimbulkan tantangan dalam proses bisnis yang memerlukan konsistensi dan akurasi tinggi. Risiko halusinasi dapat menyebabkan keputusan bisnis yang salah jika tidak ada mekanisme validasi yang memadai. Latensi respons yang lebih tinggi dibandingkan solusi berbasis pola juga dapat menjadi hambatan untuk proses bisnis yang sangat sensitif terhadap waktu. Integrasi dengan sistem keamanan dan tata kelola data memerlukan perhatian khusus untuk memastikan bahwa model tidak secara tidak sengaja mengekspos informasi sensitif atau melanggar kebijakan privasi.</p>

Bersambung ke halaman berikutnya

Tabel III.7 – Analisis *Balanced Scorecard* untuk LLM (lanjutan)

Perspektif	Penilaian
<i>Learning and Growth</i>	Solusi ini memberikan kesempatan sangat tinggi untuk pembelajaran dan pertumbuhan organisasi dalam bidang teknologi mutakhir. Implementasi model bahasa besar memungkinkan tim untuk mengembangkan keahlian dalam pembelajaran mendalam, arsitektur Transformer, penyetelan-halus model, dan teknik prompt engineering yang merupakan kompetensi sangat relevan di era kecerdasan buatan generatif. Organisasi dapat memposisikan diri sebagai pelopor dalam adopsi teknologi kecerdasan buatan paling canggih. Namun, ketergantungan pada model dan infrastruktur dari penyedia eksternal (seperti OpenAI atau Anthropic) dapat membatasi kontrol organisasi terhadap teknologi inti dan menimbulkan risiko vendor lock-in. Jika menggunakan model sumber terbuka, organisasi memerlukan investasi signifikan dalam keahlian spesialisasi untuk pelatihan, penyetelan, dan pemeliharaan model. Platform ini membuka peluang untuk inovasi masa depan seperti asisten virtual multidomain, sistem rekomendasi kontekstual, atau analitik prediktif yang sangat canggih, tetapi memerlukan komitmen jangka panjang dalam pengembangan kapabilitas dan manajemen risiko.

III.3.2.4 Matriks Perbandingan *Balanced Scorecard*

Dalam memberikan gambaran yang lebih komprehensif, berikut adalah matriks perbandingan ketiga solusi berdasarkan keempat perspektif *Balanced Scorecard*, beserta karakteristik pengguna yang paling sesuai untuk masing-masing solusi:

Tabel III.8 Perbandingan *Balanced Scorecard* Alternatif Solusi

Solusi	<i>Financial</i>	<i>Customer</i>	<i>Process</i>	<i>L&G</i>	Digunakan oleh
<i>Chatbot</i> Berbasis Aturan Sederhana	Rendah- Sedang	Rendah- Sedang	Rendah	Rendah	Organisasi kecil dengan kebu- tuhan analitik sangat terbatas dan kueri yang sangat terstruktur
<i>Chatbot</i> BI dengan Klasifikasi Maksud, Pem- bangkitan Bahasa Alami, dan Peramal- an Deret Waktu	Sedang- Tinggi (ROI Tinggi)	Sangat Tinggi	Sangat Tinggi	Sangat Tinggi	Organisasi me- nengah hingga besar dengan volume data tinggi, kebutuhan analitik kom- pleks, dan fokus pada transformasi digital berbasis AI
Dasbor BI Interaktif Tradisional dengan Pe- ningkatan	Sedang	Sedang	Sedang- Tinggi	Sedang	Organisasi yang sudah memiliki investasi besar dalam alat BI tra- disional dan ingin peningkatan bertahap

Bersambung ke halaman berikutnya

Tabel III.8 – Perbandingan *Balanced Scorecard* Alternatif Solusi (lanjutan)

Solusi	<i>Financial</i>	<i>Customer</i>	<i>Process</i>	<i>L&G</i>	Digunakan oleh
Chatbot BI Berbasis Model Bahasa Besar (LLM)	Sedang-Tinggi (Biaya Tinggi)	Sangat Tinggi (dengan Risiko)	Tinggi (dengan Tantangan)	Sangat Tinggi	Organisasi besar dengan anggaran signifikan, infrastruktur AI matang, kebutuhan pemahaman bahasa sangat kompleks, dan kapabilitas manajemen risiko AI yang kuat

III.3.2.5 Kesimpulan Pemilihan Solusi

Berdasarkan analisis komprehensif menggunakan metode *Balanced Scorecard* pada keempat perspektif (*Financial*, *Customer*, *Internal Business Process*, dan *Learning and Growth*), **Solusi 2: Chatbot BI dengan Klasifikasi Maksud Berbasis Pola, Pembangkitan Bahasa Alami Berbasis Templat, dan Peramalan Deret Waktu** adalah solusi yang paling optimal dan direkomendasikan untuk diimplementasikan.

Perlu dicatat bahwa Solusi 4 (Chatbot BI Berbasis Model Bahasa Besar) menawarkan kemampuan pemahaman bahasa alami yang paling canggih dan pengalaman pengguna yang sangat unggul. Namun, solusi ini memiliki beberapa kelemahan kritis yang membuatnya kurang optimal untuk konteks penelitian ini: (1) Biaya implementasi dan operasional yang sangat tinggi, yang dapat tidak proporsional dengan kebutuhan organisasi menengah; (2) Risiko halusinasi dan respons tidak akurat yang dapat berdampak serius pada keputusan bisnis berbasis data; (3) Kesulitan dalam mengontrol keluaran secara deterministik, yang penting untuk sistem BI yang memerlukan konsistensi dan akurabilitas; (4) Potensi kebocoran data sensitif melalui model yang tidak sepenuhnya dikontrol organisasi; dan (5) Ketergantungan pada penyedia model eksternal yang menimbulkan risiko keberlanjutan jangka panjang. Meskipun teknologi ini sangat menjanjikan untuk masa depan, untuk implementasi praktis dalam konteks organisasi yang memerlukan keseimbangan antara inovasi dan stabilitas, Solusi 2 tetap menjadi pilihan yang paling optimal.

Berikut alasan utama pemilihan Solusi 2 adalah sebagai berikut:

1. Keselarasan dengan Kebutuhan

Solusi 2 secara langsung mengatasi semua masalah pengguna yang telah diidentifikasi, termasuk keterlambatan respons, kesulitan akses analisis khusus, ketiadaan kemampuan prediksi otomatis, dan akses yang tidak ramah pengguna nonteknis. Solusi ini memenuhi seluruh kebutuhan fungsional dan non-fungsional yang telah disusun secara komprehensif.

2. Skalabilitas dan Adaptabilitas

Sistem yang diusulkan dirancang dengan arsitektur modular dan dapat diperluas yang dapat diadaptasi terhadap perubahan kebutuhan bisnis di masa depan tanpa memerlukan perancangan ulang besar. Kemampuan untuk menambahkan maksud baru, aturan baru, atau model peramalan baru secara mudah memberikan fleksibilitas jangka panjang.

3. Tahan Masa Depan dan Platform untuk Inovasi

Platform yang dibangun membuka pintu untuk inovasi-inovasi masa depan seperti ekspansi ke domain lain (*chatbot* layanan pelanggan, sistem rekomendasi, deteksi penipuan), integrasi dengan sistem eksternal (ERP, CRM), atau pengembangan kemampuan analitik lanjutan lainnya. Investasi dalam platform ini tidak hanya memberikan nilai untuk kebutuhan saat ini, tetapi juga menjadi fondasi untuk transformasi digital organisasi secara keseluruhan.

4. Pengembalian Investasi dan Justifikasi Finansial

Meskipun investasi awal lebih tinggi dibandingkan alternatif lain, pengembalian investasi jangka panjang sangat menarik karena penghematan biaya tenaga kerja yang signifikan (analisis dapat fokus pada tugas strategis), pengambilan keputusan yang lebih cepat dan akurat (yang dapat berdampak langsung pada pendapatan atau penghematan biaya), dan peningkatan daya saing organisasi. Berdasarkan studi dari berbagai industri, implementasi sistem analitik prediktif berbasis kecerdasan buatan dapat meningkatkan efisiensi operasional hingga 30% dan mempercepat pengambilan keputusan hingga 85% (Shah, 2025).

5. Keunggulan Kompetitif dan Diferensiasi

Implementasi *chatbot* BI canggih dengan kemampuan kecerdasan buatan dan pembelajaran mesin memberikan keunggulan kompetitif bagi organisasi, meningkatkan kelincahan dan kecepatan pengambilan keputusan dibandingkan kompetitor yang masih menggunakan sistem konvensional. Organisasi yang mampu menganalisis data dan merespons tren pasar lebih cepat memiliki posisi yang lebih kuat dalam persaingan bisnis.

6. Adopsi Pengguna dan Kepuasan

Antarmuka percakapan berbasis bahasa wajar memiliki tingkat adopsi yang jauh lebih tinggi dibandingkan dasbor tradisional karena lebih intuitif dan tidak memerlukan pelatihan ekstensif. Statistik industri menunjukkan bahwa *chatbot* dapat menangani hingga 80% pertanyaan standar secara otomatis (Fullview, 2025), yang secara signifikan mengurangi beban kerja tim dan meningkatkan kepuasan pengguna.

7. Keamanan dan Kontrol Data

Pendekatan pembangunan bahasa alami berbasis templat memberikan kontrol penuh atas respons yang dihasilkan sistem, memastikan bahwa tidak ada kebocoran data sensitif atau informasi yang tidak seharusnya diungkapkan. Hal ini sangat penting untuk sistem yang beroperasi dengan data internal perusahaan yang bersifat rahasia.

8. Pembelajaran dan Pertumbuhan Organisasi

Implementasi solusi ini memungkinkan organisasi untuk membangun kapabilitas internal dalam bidang kecerdasan buatan, pembelajaran mesin, dan analitik lanjutan, yang merupakan keahlian kritis di era digital. Tim akan mengembangkan pengetahuan dan pengalaman yang berharga yang dapat diterapkan pada proyek-proyek inovasi lainnya di masa depan.

Dengan pemilihan Solusi 2, organisasi akan dapat mencapai transformasi digital yang signifikan dalam domain *Business Intelligence* dan analitik, meningkatkan efisiensi operasional, kecepatan pengambilan keputusan, dan memposisikan diri sebagai pemimpin dalam pemanfaatan teknologi kecerdasan buatan untuk keunggulan kompetitif di pasar yang semakin berbasis data.

Solusi yang dipilih juga sejalan dengan tren global dengan pasar kecerdasan buatan percakapan diproyeksikan mencapai USD 49,7 miliar pada tahun 2025 (Qaltivate, 2025), pasar pemrosesan bahasa alami tumbuh dengan CAGR 28,6% mencapai USD 438,08 miliar pada tahun 2034 (Precedence Research, 2025), dan pasar peramalan deret waktu diperkirakan mencapai USD 36,9 miliar pada tahun 2032 (WiseGuy Reports, 2024). Pertumbuhan pasar yang eksponensial ini menunjukkan bahwa investasi dalam teknologi-teknologi ini bukan hanya relevan untuk saat ini, tetapi juga strategis untuk jangka panjang.

DAFTAR PUSTAKA

- Amazon Web Services. 2024. "Securing Generative AI: Data, Compliance, and Privacy Considerations". Diakses pada November 4, 2025. <https://aws.amazon.com/blogs/security/securing-generative-ai-data-compliance-and-privacy-considerations/>.
- Al-Amri, Rashid, R. K. Murugesan, dkk. 2025. "Enhancing Customer Churn Analysis by Using Real-time Continual Learning". *International Journal of Advanced Computer Science and Applications* 16 (5): 1–12.
- Ananta, Y. E., dkk. 2025. "Pengembangan Aplikasi Sido Chatbot sebagai Layanan Informasi Menggunakan Pendekatan Pencocokan Pola Berbasis Aturan". *Jurnal Sistem Cerdas dan Aplikasi X (Y)*: 1–12.
- AtScale. 2025. "Why Semantic Layers? A Key Technical Architecture". Diakses pada November 4, 2025. <https://www.atscale.com/blog/semantic-layers-key-technical-architecture/>.
- Bamurange, Daniel. 2025. "Designing a Hybrid AI Chatbot Framework for Student Support: Integrating NLP and Human Oversight in African Universities". *Edinburgh Journals - Journal of Information Technology* 5 (4): 41–52. <https://doi.org/10.70619/vol5iss4pp41-52>.
- Business Wire. 2020. "The World's AI-Based Chatbot Market, 2020-2025: Conversational AI Will Improve Telemarketing Efficiency". Diakses pada November 4, 2025. <https://www.businesswire.com/news/home/20200325005452/en/>.
- Celikyilmaz, Asli, Elizabeth Clark, dan Jianfeng Gao. 2020. "Evaluation of Text Generation: A Survey". *arXiv preprint arXiv:2006.14799*.
- Cheng, Yu, dkk. 2025. "Multivariate Time Series Forecasting Through Automated Feature Extraction and Transformer-Based Modeling". *Journal of Computer Science and Software Applications* 5 (5): 1–25.

- Cherednichenko, O., dkk. 2023. “Reference Model for Collaborative Business Intelligence Virtual Assistants”. Dalam *CEUR Workshop Proceedings*, 3403:77–88.
- Cyara. 2024. “What’s the Optimum Confidence Threshold for My Chatbot”. Diakses pada November 4, 2025. <https://cyara.com/blog/optimum-chatbot-confidence-threshold/>.
- Databricks. 2025. “The Role of Semantic Layers in Modern Data Analytics”. Diakses pada November 4, 2025. <https://www.databricks.com/glossary/semantic-layer>.
- Diqi, M., dkk. 2023. “Multi-Step Vector Output Prediction of Time Series Using EMA LSTM and GRU-LSTM Hybrid Model”. *JOIN X (Y)*: 1–20.
- East Ventures. July 2025. *AI-first: Decoding Southeast Asia Trends*. White Paper. Diakses: 4 November 2025. Jakarta, Indonesia: East Ventures. <https://east.vc/press-release-3/east-ventures-launched-a-white-paper-focusing-on-ai>.
- Elangovan, Ashwin, dkk. 2023. “ConSiDERS: The Human Evaluation Framework”. Dalam *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*.
- European Data Protection Board. 2025. “AI Privacy Risks & Mitigations: Large Language Models (LLMs)”. Diakses pada November 4, 2025. <https://www.edpb.europa.eu/>.
- Evidently AI. 2025. “Accuracy, Precision, and Recall in Multi-Class Classification”. Diakses pada December 2, 2025. <https://www.evidentlyai.com/classification-metrics/multi-class-metrics>.
- Feyisetan, Oluwaseyi, Borja Balle, Thomas Drake, dan Tom Diethe. 2022. “Privacy- and Utility-Preserving Textual Analysis via Calibrated Multivariate Perturbations”. *Proceedings of the 13th ACM International Conference on Web Search and Data Mining*, 178–186. <https://doi.org/10.1145/3336191.3371856>.
- Fujitsu - PostgreSQL Fastware. 2024. “Generative AI and Data: Protecting Organizational Data”. Diakses pada November 4, 2025. <https://www.postgresql.fastware.com/blog/generative-ai-and-data>.

- Fullview. 2025. “80+ AI Customer Service Statistics & Trends in 2025”. Diakses pada November 4, 2025. <https://www.fullview.io/blog/ai-customer-service-stats>.
- Gao, Mingqi, Xinyu Hu, Jie Yin, Xiao Ruan, dan Xiaojun Wan. 2025. “LLM-based NLG Evaluation: Current Status and Challenges”. ArXiv:2402.01383, *Computational Linguistics*.
- Gehweiler, Christian. 2024. “Classification of Intent in Moderating Online Discussions”. *Expert Systems with Applications* 239:122234.
- Gehweiler, Christian, dkk. 2025. “Intent Detection for Task-Oriented Conversational Agents: A Survey”. *Expert Systems* 42 (3).
- Genesys. 2023. “Set Bot Confidence Thresholds”. Diakses pada November 4, 2025. <https://www.genesys.com/blog/post/set-bot-confidence-thresholds>.
- Grandini, Margherita, Enrico Bagli, dan Giorgio Visani. 2020. “Metrics for Multi-Class Classification: An Overview”. *arXiv preprint arXiv:2008.05756*, arXiv: 2008.05756 [cs.LG].
- Hartanto, Sigit, dkk. 2024. “Temporal Fusion Transformers for Enhanced Multivariate Time Series Stock Market Prediction”. *The Science and Information Organization* X (Y).
- Huang, Xufei. 2021. “Chatbot: Design, Architecture, and Applications”. PhD thesis, University of Pennsylvania. <https://www.cis.upenn.edu/wp-content/uploads/2021/10/Xufei-Huang-thesis.pdf>.
- Intimedia. Desember 2024. “Indonesia’s Digital Transformation in 2024: How AI is Turning Dreams into Reality”. Diakses: 4 November 2025, *Intimedia News* (). <https://intimedia.id/read/indonesias-digital-transformation-in-2024-how-ai-is-turning-dreams-into-reality>.
- Iqbal, Muhammad, Rashid Ahmad, dan Saif Ur Rehman Khan. 2021. “Combined Support Vector Machine and Pattern Matching for Arabic Question Classification”. *International Journal of Pattern Recognition and Artificial Intelligence* 35 (8): 287–299. https://doi.org/10.1007/978-3-319-99007-1_27.
- IT Convergence. 2024. “Generative AI Data Security: Key Considerations”. Diakses pada November 4, 2025. <https://www.itconvergence.com/blog/data-security-considerations-for-generative-ai/>.

- Jeaab, K., dkk. 2024. “A Comparison of LSTM, GRU, and XGBoost for Forecasting Morocco’s Yield Curve”. *Eastern-European Journal of Enterprise Technologies X (Y)*.
- Kale, Mihir, dan Abhinav Rastogi. 2020. “Template Guided Text Generation for Task-Oriented Dialogue”. Dalam *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 6505–6520. Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.emnlp-main.527>.
- Kapoor, Ankit, dan Sujala Shetty. 2023. “Implementing Natural Language Generation Through Industry-Specific Chatbots”. *Data Science and Intelligent Computing Techniques*, 55–67. <https://doi.org/10.56155/978-81-955020-2-8-6>.
- 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>.
- Khosla, Sopan, Chenwei Xu, Shikhar Toshniwal, Dinesh Manocha, Markus Dreyer, Lambert Mathias, Jing Gao, dan Asli Celikyilmaz. 2022. “Evaluating the Practical Utility of Confidence-score based Techniques for Unsupervised Open-World Classification”. Dalam *Proceedings of the Third Workshop on Insights from Negative Results in NLP*, 20–28. <https://doi.org/10.18653/v1/2022.insights-1.3>.
- Kuligowska, Karolina, dan Zdzislaw Kowalczyk. 2024. “Enhancing Chatbot Intent Classification Using Active Learning Pipeline”. *Journal of Applied Economic Sciences* 19 (2): 290–300.
- Kumar, Arnav, dan Ebad Zafar. 2024. “Predict Customer Churn with Python and Machine Learning”. *SSRN Electronic Journal*, <https://doi.org/10.2139/ssrn.5085192>.
- Labs, dbt. 2024. “Understanding Semantic Layer Architecture”. Diakses pada November 4, 2025. <https://www.getdbt.com/blog/semantic-layer-architecture>.
- Lee, Chris van der, dkk. 2021. “Human Evaluation of Automatically Generated Text: Current Trends and Best Practices”. *Natural Language Engineering* 27 (4): 1–27.

- Li, Xiaoming, Yutao Zhang, Yang Liu, dan Wei Wang. 2024. “Customer Churn Prediction Model Based on Hybrid Neural Networks”. *Scientific Reports* 14:79603. <https://doi.org/10.1038/s41598-024-79603-9>.
- Lim, Bryan, Sercan O. Arik, Nicolas Loeff, dan Tomas Pfister. 2021. “Temporal Fusion Transformers for Interpretable Multi-Horizon Time Series Forecasting”. *International Journal of Forecasting* 37 (4): 1748–1764. <https://doi.org/10.1016/j.ijforecast.2021.03.012>.
- Lubis, J. H., dkk. 2025. “Performance Optimization of ERD Designs Using Cost-Based and Rule-Based Query Optimization”. *JUTEI X* (Y): 1–10.
- Luo, Yuanshuai, dan Guiran Liu. 2024. “Data Security Strategies in Natural Language Processing Applications”. *arXiv preprint arXiv:2410.08553*, arXiv: 2410.08553.
- Mena, C. G., A. De Caigny, K. Coussement, K. W. De Bock, dan S. Lessmann. 2023. “Exploiting time-varying RFM measures for customer churn prediction with deep neural networks”. Diakses: Tanggal Akses Anda, *Annals of Operations Research* 339 (1-2): 765–787. URL_ATAU_DOI_DISINI.
- Merugu, Srujana, dkk. 2024. “Intent Detection in the Age of LLMs”. Dalam *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track*, 1–12.
- Mireshghallah, Fatemehsadat, Kartik Goyal, Shrimai Upadhyay, dan Taylor Berg-Kirkpatrick. 2022. “Differentially Private Methods in Natural Language Processing”. *Proceedings of the Fourth Workshop on Privacy in Natural Language Processing*, 1–11. <https://doi.org/10.18653/v1/2022.privatenlp-1.1>.
- Morrison, Mark, dkk. 2021. “Reproducible Subjective Evaluation”. Dalam *Proceedings of the Neural Information Processing Systems Workshop on ML Evaluation Standards*.
- 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>.

- Nabiilah, Sarah, Adiwijaya, dan Widi Maharani. 2025. "Comparative Analysis of SVM and IndoBERT for Intent Classification in E-LEMBUR Chatbot". *Journal of Software and Computer Engineering* 6 (3): 258–267. <https://doi.org/10.61628/jsce.v6i3.2058>.
- Nama, Penulis, dan Lainnya Penulis. 2024. "AI Chatbot Quality in Customer Service: Extending Service Quality Models". *Learning Gate Journal* X (Y).
- Ouaddi, Chaimae, Fatima Benabbou, dan Nawal Sael. 2025. "A Comparative Analysis and Performance Evaluation of Deep Learning Techniques for Intent Detection in Conversational AI Chatbots". *Journal of Big Data* 12 (1): 1–25. <https://doi.org/10.1186/s40537-024-00996-6>.
- Ouali, S., dkk. 2024. "Arabic Chatbots Challenges and Solutions: A Systematic Review". *International Journal of Computer Science and Mobile Computing* X (Y): 1–20.
- Ozerova, Daria. 2022. "Intent Detection Module for a Conversational Assistant". Master's thesis, Czech Technical University in Prague.
- Pecan AI. 2024. "Best Machine Learning Models for Predicting Customer Churn". Diakses pada November 4, 2025. <https://www.pecan.ai/blog/best-ml-models-for-predicting-customer-churn/>.
- Petropoulos, Fotios, dkk. 2023. "Forecast Evaluation for Data Scientists: Common Pitfalls and Best Practices". *International Journal of Forecasting* 39 (4): 1500–1518.
- Precedence Research. 2025. "Natural Language Processing Market Size to Hit USD 438.08 Bn by 2034". Diakses pada November 4, 2025. <https://www.precedenceresearch.com/natural-language-processing-market>.
- Qaltivate. 2025. "Conversational AI Market Summary: 2025 Edition". Diakses pada November 4, 2025. <https://qaltivate.com/blog/conversational-ai-market/>.
- Rahmat, Budi, dan Agus Setiawan. 2024. "Prediction Model Using Machine Learning: Analysis of Determinants of Customer Churn at PT XYZ". *Jurnal Info Sains: Informatika dan Sains* 14 (04): 745–758.
- Rebuffel, Clément, Laure Soulier, Geoffrey Scoutheeten, dan Patrick Gallinari. 2020. "A Hierarchical Model for Data-to-Text Generation". *arXiv preprint arXiv:2004.15006*, arXiv: 2004.15006.

- Sabila, A., dkk. 2025. “Deep Learning, GRU (Gated Recurrent Unit), LSTM (Long Short-Term Memory) Stock Price Prediction”. *Akademika X* (Y).
- Sai, Ananya B., Akash Kumar Mohankumar, dan Mitesh M. Khapra. 2020. “A Survey of Evaluation Metrics Used for NLG Systems”. *arXiv preprint arXiv:2008.12009*.
- Santosa, Rahmad, Adetiya Bagus Nusantara, dan Syaiful Imron. 2025. “Comparative Analysis of SVM and IndoBERT for Intent Classification in Indonesian Overtime Chatbots”. *JSCE* 6 (3): 258–270. <https://doi.org/10.61628/jsce.v6i3.2058>.
- Shah, Nisarg. 2025a. “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.
- . 2025b. “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>.
- Sharma, Rajesh, Amit Gupta, dan Pradeep Singh. 2025. “Customer Churn Prediction: A Machine Learning Approach with Data Balancing for Telecom Industry”. *International Journal of Computing* 24 (1): 1–12. <https://doi.org/10.47839/ijc.24.1.3873>.
- Singh, S. U., dkk. 2025. “A Survey on Chatbots and Large Language Models: Testing and Design Principles”. *Journal of Decision Systems* X (Y): 1–25.
- 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>.
- Spiliotis, Evangelos, Spyros Makridakis, dan Fotios Petropoulos. 2020. “Comparison of Statistical and Machine Learning Methods for Time Series Forecasting”. *International Journal of Forecasting* 36 (1): 1–15.
- Spot Intelligence. 2023. “Building a Dataset for Intent Classification in NLP: Evaluation and Model Performance”. Diakses pada November 4, 2025. <https://spotintelligence.com/2023/11/03/intent-classification-nlp/>.

- Straits Research. 2024. "Business Intelligence Market Size & Outlook, 2025–2033". Diakses pada November 4, 2025. <https://straitsresearch.com/report/business-intelligence-market>.
- Sunendar, Nendi, dan Yan Rianto. 2025. "Comparison of ARIMA, LSTM, and GRU Models for Forecasting Sales of HIT Aerosol Products". *Jurnal Pilar Nusa Mandiri* X (Y).
- Syarifudin, Muhammad, dkk. 2024. "Modeling AI-Chatbot Service Quality and Purchase Intention". *JDMHI* X (Y).
- Talenggoran, R., dkk. 2025. "Implementation of Rasa Framework to Build FAQ Bot System on Messaging Platform". *JUTEI* X (Y): 1–10.
- Teneo Developers. 2022. "Intent Classification: Confidence Threshold". Diakses pada November 4, 2025. <https://developers.teneo.ai/documentation/7.5.0/reference/conceptual-overviews/intent-classification>.
- Velotix. 2025. "Data Security in Generative AI Environments: 5 Steps". Diakses pada November 4, 2025. <https://www.velotix.ai/resources/blog/data-security-in-generative-ai-environments/>.
- WiseGuy Reports. 2024. "Time Series Forecasting Market". Diakses pada November 4, 2025. <https://www.wiseguyreports.com/reports/time-series-forecasting-market>.
- XenonStack. 2023. "Chatbots for Business: Best Practices and Top Applications". Diakses pada November 4, 2025. <https://www.xenonstack.com/insights/chatbots-applications>.
- Xiao, Ziang, dkk. 2023. "A Framework for Analyzing NLG Evaluation Metrics Using Measurement Theory". *Transactions of the Association for Computational Linguistics*.
- Yang, X., dkk. 2024. "Machine Learning for Database Management and Query Optimization". *International Journal of Computer Science and Management* X (Y): 1–12.
- Yunita, Anis, dkk. 2025. "Performance Analysis of Neural Network Architectures for Time Series Forecasting: A Comparative Study of RNN, LSTM, GRU, and Hybrid Models". *Data Science and Engineering* X (Y): 1–25.

Zhang, Yun, dkk. 2024. “A Comprehensive Survey of Time Series Forecasting: ARIMA, Deep Learning, Transformers, and Beyond”. *arXiv preprint*, eprint: 2411.05793.

Zhang, Yutao, Yang Liu, Meng Zhang, dan Xiaoming Li. 2022. “Disentangling Confidence Score Distribution for Out-of-Domain Intent Detection”. *arXiv preprint arXiv:2210.08830*, arXiv: 2210.08830.