

BAB II

TINJAUAN PUSTAKA

2.1 Penelitian Yang Relevan

Untuk mengetahui perbandingan penelitian yang dilakukan dengan penelitian yang sudah ada, maka perlu dilakukan studi sejenis. Berikut merupakan beberapa penelitian terdahulu.

Penelitian yang dilakukan oleh (Asri et al., 2025) dengan judul “*Sentiment analysis based on Indonesian language lexicon and IndoBERT on user reviews PLN mobile application*” membahas analisis sentimen pada ulasan pengguna aplikasi PLN Mobile menggunakan kombinasi leksikon bahasa Indonesia dan model IndoBERT. Penelitian ini menggunakan 1.000 sampel ulasan dari Google Play Store, melakukan *pre-processing*, pelabelan otomatis berbasis leksikon INSET, lalu klasifikasi dengan IndoBERT pada pembagian data train, validation, dan test 80:10:10. Hasil penelitian menunjukkan bahwa model mencapai akurasi sebesar 81%, sehingga membuktikan bahwa IndoBERT cukup efektif untuk analisis ulasan PLN Mobile. Namun, penelitian ini masih berfokus pada klasifikasi sentimen umum positif, negatif, dan netral, sehingga belum mampu mengidentifikasi jenis keluhan secara lebih spesifik melalui pelabelan otomatis berbasis kategori domain.

Penelitian yang dilakukan oleh (Zhong & Goodfellow, 2024) dengan judul “*Domain-specific language models pre-trained on construction management systems corpora*” Penelitian ini membahas pengembangan model bahasa domain-spesifik yang dilatih pada corpus teks dari sistem manajemen konstruksi. Studi ini membangun beberapa corpus domain konstruksi, melakukan pre-training menggunakan model BERT dan RoBERTa, kemudian melanjutkan dengan fine-tuning pada tugas klasifikasi teks dan Named Entity Recognition (NER). Hasil penelitian menunjukkan bahwa penggunaan pre-training berbasis corpus domain-spesifik mampu meningkatkan *F1-score* klasifikasi teks sebesar 5,9% dan tugas NER sebesar 8,5% dibanding model yang hanya menggunakan corpus umum. Temuan ini menegaskan pentingnya pembangunan corpus domain-spesifik untuk meningkatkan akurasi model bahasa pada domain yang memiliki istilah dan konteks

khusus, meskipun penelitian ini masih terbatas pada domain konstruksi dan belum diterapkan pada domain keluhan layanan digital seperti PLN Mobile.

Penelitian yang dilakukan oleh (Devlin et al., 2019) dengan judul “*BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*” memperkenalkan model BERT sebagai model bahasa berbasis Transformer dengan representasi dua arah yang mampu memahami konteks kata dari sisi kiri dan kanan secara bersamaan. Penelitian ini menunjukkan bahwa BERT mencapai hasil *state-of-the-art* pada berbagai tugas NLP, termasuk GLUE, MultiNLI, dan SQuAD, serta dapat di-*fine-tune* untuk beragam tugas tanpa perubahan arsitektur yang besar. Hasil penelitian ini menjadi dasar penting bagi banyak studi lanjutan berbasis BERT, termasuk model bahasa untuk bahasa Indonesia seperti IndoBERT. Akan tetapi, penelitian ini masih bersifat **umum** dan tidak dirancang secara khusus untuk bahasa Indonesia maupun domain layanan publik seperti keluhan pelanggan PLN Mobile.

Penelitian yang dilakukan oleh (Suzuki et al., 2023) dengan judul “*Constructing and analyzing domain-specific language model for financial text mining*” Penelitian ini mengeksplorasi pembuatan dan analisis model bahasa domain-spesifik untuk ekstraksi dan klasifikasi informasi pada teks finansial. Penelitian ini membangun corpus domain-finansial, melakukan pre-training model bahasa berbasis Transformer, dan menguji performa model pada tugas ekstraksi informasi dan klasifikasi teks. Hasil penelitian menunjukkan bahwa model domain-spesifik mampu meningkatkan akurasi ekstraksi informasi serta klasifikasi teks dibanding model berbasis corpus umum. Temuan ini menegaskan efektivitas penerapan corpus domain-spesifik untuk meningkatkan performa NLP pada teks dengan terminologi dan struktur khusus, dalam hal ini domain finansial, yang relevan sebagai referensi metodologi untuk pengembangan corpus domain-spesifik pada keluhan pelanggan PLN Mobile.

Penelitian yang dilakukan oleh (Koto et al., 2020) dengan judul “*IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP*” membahas pengembangan benchmark NLP bahasa Indonesia beserta model pra-latih IndoBERT. Penelitian ini bertujuan menjawab keterbatasan sumber daya NLP bahasa Indonesia dengan menyediakan dataset evaluasi dan model pra-latih yang sesuai konteks

bahasa Indonesia. Hasil penelitian menunjukkan bahwa IndoBERT mencapai performa *state-of-the-art* pada sebagian besar tugas di benchmark IndoLEM. Temuan ini memperkuat bahwa IndoBERT layak digunakan untuk pemrosesan teks berbahasa Indonesia. Namun, penelitian ini masih berfokus pada benchmark umum NLP bahasa Indonesia, sehingga belum secara khusus mengeksplorasi pengembangan corpus domain-spesifik untuk tugas pelabelan otomatis pada domain keluhan pelanggan.

Penelitian yang dilakukan oleh (Jazuli et al., 2024) dengan judul “*Optimizing Aspect-Based Sentiment Analysis Using BERT for Comprehensive Analysis of Indonesian Student Feedback*” membahas penerapan IndoBERT untuk *aspect-based sentiment analysis* pada umpan balik mahasiswa Indonesia. Penelitian ini menggunakan 10.000 ulasan mahasiswa, melalui tahap pelabelan, *pre-processing*, pelatihan model, dan evaluasi performa. Hasil penelitian menunjukkan akurasi ekstraksi aspek sebesar 0,973 dengan *F1-score* 0,952, serta akurasi klasifikasi sentimen 0,979 dengan *F1-score* 0,974. Penelitian ini menunjukkan bahwa model BERT yang disesuaikan untuk bahasa Indonesia sangat efektif dalam menangkap aspek-aspek spesifik dari teks ulasan. Namun, penelitian ini berada pada domain umpan balik pendidikan, bukan keluhan layanan publik, dan lebih menekankan ABSA daripada pengembangan corpus domain-spesifik untuk pelabelan otomatis kategori keluhan.

Penelitian yang dilakukan oleh (Taj et al., 2025) dengan judul “*Aspect-based sentiment analysis for software requirements elicitation using fine-tuned Bidirectional Encoder Representations from Transformers and Explainable Artificial Intelligence*” membahas pemanfaatan BERT yang di-*fine-tune* untuk *aspect-based sentiment analysis* dalam elisitasi kebutuhan perangkat lunak. Penelitian ini menunjukkan bahwa pendekatan berbasis BERT dapat membantu mengidentifikasi aspek-aspek penting dari umpan balik pengguna yang relevan untuk kebutuhan pengembangan perangkat lunak, serta diperkaya dengan pendekatan *Explainable Artificial Intelligence*. Penelitian ini relevan karena memperlihatkan bahwa teks ulasan dapat diolah lebih mendalam untuk menemukan aspek atau kategori informasi tertentu. Akan tetapi, penelitian ini lebih berorientasi pada rekayasa kebutuhan perangkat lunak, bukan pada pelabelan otomatis keluhan pelanggan

PLN Mobile, serta tidak menitikberatkan pada pembangunan corpus khusus bahasa Indonesia.

Penelitian yang dilakukan oleh (Wu et al., 2024) dengan judul “*Domain-specific language models pre-trained on construction management systems corpora*” membahas pengembangan model bahasa domain-spesifik melalui *pre-training* pada corpus bidang konstruksi. Penelitian ini membangun empat corpus domain konstruksi, melakukan *transfer learning* untuk *pre-training* BERT dan RoBERTa, kemudian *fine-tuning* pada tugas klasifikasi teks dan NER. Hasil penelitian menunjukkan bahwa *domain-specific pre-training* meningkatkan F1-score sebesar 5,9% pada klasifikasi teks dan 8,5% pada NER dibanding model yang hanya menggunakan corpus umum. Penelitian ini sangat relevan karena membuktikan bahwa corpus domain-spesifik mampu meningkatkan performa model bahasa pada tugas NLP. Meskipun demikian, domain yang digunakan masih terbatas pada construction management systems, sehingga belum merepresentasikan domain keluhan pelanggan layanan digital seperti PLN Mobile.

Penelitian yang dilakukan oleh (Dwi Purnomo & Sutopo, 2024) dengan judul “*COMPARISON OF PRE-TRAINED BERT-BASED TRANSFORMER MODELS FOR REGIONAL LANGUAGE TEXT SENTIMENT ANALYSIS IN INDONESIA*” Penelitian ini membahas perbandingan berbagai model pra-latih berbasis BERT untuk analisis sentimen teks berbahasa daerah di Indonesia. Studi ini menggunakan dataset ulasan dan komentar dalam bahasa daerah, kemudian melakukan pre-processing, fine-tuning, dan evaluasi performa model menggunakan metrik akurasi dan F1-score. Hasil penelitian menunjukkan bahwa model BERT yang disesuaikan dengan data lokal mampu memberikan performa lebih baik dibanding model multilingual umum, khususnya dalam menangkap nuansa bahasa regional dan konteks lokal. Temuan ini menegaskan bahwa adaptasi model bahasa terhadap kosakata dan struktur linguistik domain tertentu penting untuk meningkatkan akurasi analisis sentimen teks berbahasa Indonesia.

Penelitian yang dilakukan oleh (Koto et al., 2021) dengan judul “*INDOBERTWEET: A Pretrained Language Model for Indonesian Twitter with Effective Domain-Specific Vocabulary Initialization*” Penelitian ini memperkenalkan model pra-latih INDOBERTWEET yang dioptimalkan untuk teks Twitter berbahasa Indonesia dengan

kosakata domain-spesifik. Penelitian membangun corpus domain-Twitter, melakukan pre-training model bahasa, dan mengevaluasi performa pada tugas klasifikasi teks dan analisis sentimen. Hasil penelitian menunjukkan bahwa inisialisasi kosakata domain-spesifik meningkatkan kemampuan model dalam memahami istilah informal, singkatan, dan konteks khas media sosial Indonesia, sehingga model mampu memberikan akurasi dan konsistensi lebih tinggi dibanding model IndoBERT umum. Temuan ini relevan untuk pengembangan corpus domain-spesifik pada keluhan pelanggan PLN Mobile, karena menunjukkan pentingnya penyesuaian kosakata terhadap karakteristik domain data.

Tabel 2. 1 State of The Art Penelitian

1.	Sentiment analysis based on Indonesian language lexicon and IndoBERT on user reviews PLN mobile application (Asri et al., 2025)							
Pre-Processing					Named Entity Recognition	Bahasa Dataset	Modeling	Evaluasi
Case Folding	Filtering	Normalization	Spelling Corrector	Tokenization				
Ya	Ya	Ya	-	Ya	Ya	Indonesia	IndoBERT	81%
Kesimpulan: Hasil penelitian menunjukkan bahwa IndoBERT mampu melakukan klasifikasi sentimen positif, negatif, dan netral dengan performa yang lebih baik dibandingkan metode pembandingan. Model IndoBERT memperoleh nilai akurasi sebesar 81%, sedangkan metode Vader Lexicon dan Naïve Bayes memperoleh akurasi sebesar 70%, sehingga menunjukkan bahwa IndoBERT efektif dalam melakukan analisis sentimen pada ulasan aplikasi PLN Mobile.								

2.	Domain-specific language models pre-trained on construction management systems corpora (Zhong & Goodfellow, 2024)							
Pre-Processing					Named Entity Recognition	Bahasa Dataset	Modeling	Evaluasi
Case Folding	Filtering	Normalization	Spelling Corrector	Tokenization				
Ya	Ya	-	-	Ya	Ya	Inggris	BERT/RoBERTa	8,5%
Kesimpulan: Hasil penelitian menunjukkan bahwa penggunaan corpus domain-spesifik mampu meningkatkan performa model dibandingkan penggunaan corpus umum. Evaluasi menunjukkan peningkatan F1-score sebesar 5,9% pada tugas text classification dan peningkatan F1-score sebesar 8,5% pada tugas Named Entity Recognition (NER), sehingga membuktikan bahwa pre-training menggunakan corpus domain dapat meningkatkan performa model NLP.								
3.	Constructing and analyzing domain-specific language model for financial text mining (Suzuki et al., 2023)							
Pre-Processing					Named Entity Recognition	Bahasa Dataset	Modeling	Evaluasi
Case Folding	Filtering	Normalization	Spelling Corrector	Tokenization				
Ya	-	Ya	-	Ya	Ya	Jepang	Transformer	90%
Kesimpulan: Hasil penelitian menunjukkan bahwa penggunaan corpus finansial dan penyesuaian vocabulary domain mampu meningkatkan kemampuan model dalam memahami istilah khusus pada berbagai tugas NLP finansial. Evaluasi penelitian menunjukkan bahwa model domain-spesifik memberikan peningkatan performa pada beberapa downstream financial tasks dibandingkan model bahasa umum.								

4.	IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP (Koto et al., 2020)							
Pre-Processing					Named Entity Recognition	Bahasa Dataset	Modeling	Evaluasi
Case Folding	Filtering	Normalization	Spelling Corrector	Tokenization				
-	-	Ya	-	Ya	Ya	Indonesia	IndoBERT	90,1%
Kesimpulan: Hasil evaluasi menunjukkan bahwa IndoBERT mampu mencapai performa terbaik (<i>state-of-the-art</i>) pada sebagian besar tugas benchmark IndoLEM dan memberikan peningkatan dibandingkan model sebelumnya seperti multilingual BERT (mBERT). Pada tugas Named Entity Recognition, IndoBERT memperoleh nilai F1-score sebesar 90,1% pada dataset NER UI dan 74,9% pada dataset NER UGM. Penelitian ini membuktikan bahwa model bahasa yang dilatih menggunakan corpus Bahasa Indonesia mampu meningkatkan kemampuan pemahaman teks dalam berbagai tugas NLP.								
5.	Optimizing Aspect-Based Sentiment Analysis Using BERT for Comprehensive Analysis of Indonesian Student Feedback (Jazuli et al., 2024)							
PreProcessing					Named Entity Recognition	Bahasa Dataset	Modeling	Evaluasi
Case Folding	Filtering	Normalization	Spelling Corrector	Tokenization				
Ya	Ya	Ya	-	Ya	Ya	Indonesia	BERT	97,3 %
Kesimpulan: Hasil penelitian menunjukkan bahwa model memperoleh performa tinggi dengan nilai accuracy ekstraksi aspek sebesar 97,3% dan F1-score sebesar 95,2%, sedangkan klasifikasi sentimen memperoleh accuracy sebesar 97,9% dan F1-score sebesar 97,4%. Penelitian ini menunjukkan bahwa penggunaan model bahasa berbasis transformer seperti IndoBERT mampu memahami konteks Bahasa Indonesia dan menghasilkan analisis sentimen berbasis aspek dengan tingkat akurasi yang tinggi.								

6.	Dynamic building defect categorization through enhanced unsupervised text classification with domain-specific corpus embedding methods (Jeon et al., 2024)							
Pre-Processing					Named Entity Recognition	Bahasa Dataset	Modeling	Evaluasi
Case Folding	Filtering	Normalization	Spelling Corrector	Tokenization				
Ya	Ya	Ya	-	Ya	Ya	Korea	sim-Classification-ftPLM	89,12 %
Kesimpulan: Hasil penelitian menunjukkan bahwa metode yang diusulkan mampu meningkatkan performa klasifikasi teks tanpa membutuhkan proses pelabelan data dalam jumlah besar, dengan nilai F1-score sebesar 89,12% untuk wean-Clustering dan 84,66% untuk sim-Classification-ftPLM, serta mengungguli metode zero-shot dan few-shot sebagai pembanding.								
7.	INDOBERTWEET: A Pretrained Language Model for Indonesian Twitter with Effective Domain-Specific Vocabulary Initialization (Koto et al., 2021)							
Pre-Processing					Named Entity Recognition	Bahasa Dataset	Modeling	Evaluasi
Case Folding	Filtering	Normalization	Spelling Corrector	Tokenization				
Ya	Ya	Ya	-	Ya	Ya	Indonesia	IndoBERTweet	86,5 %
Kesimpulan: Hasil eksperimen pada tujuh tugas NLP Twitter, termasuk sentiment analysis, emotion classification, hate speech detection, dan Named Entity Recognition menunjukkan bahwa INDOBERTWEET mampu meningkatkan performa dibandingkan model BERT umum. Metode terbaik menggunakan domain-specific vocabulary adaptation dengan domain-adaptive pre-training memperoleh performa rata-rata sebesar 86,5% pada tujuh tugas evaluasi.								

8.	Leveraging encoder-only large language models for mobile app review feature extraction (Motger et al., 2025)							
Pre-Processing					Named Entity Recognition	Bahasa Dataset	Modeling	Evaluasi
Case Folding	Filtering	Normalization	Spelling Corrector	Tokenization				
-	Ya	-	-	Ya	Ya	Inggris	Encoder-only LLM	64,6 %
Kesimpulan: Hasil evaluasi menunjukkan bahwa model XLNetlarge memperoleh performa terbaik pada skenario in-domain dengan F1-score sebesar 0,646, sedangkan pada evaluasi feature-level menghasilkan F1-score sebesar 0,600, lebih tinggi dibandingkan metode SAFE yang hanya memperoleh F1-score 0,199. Hasil penelitian menunjukkan bahwa penggunaan model berbasis Transformer dengan corpus domain-spesifik mampu meningkatkan kemampuan ekstraksi fitur dari teks ulasan aplikasi mobile dibandingkan metode tradisional.								
9.	Domain-Specific Fine-Tuning of IndoBERT for Aspect-Based Sentiment Analysis in Indonesian Travel User-Generated Content (Perwira et al., 2025)							
Pre-Processing					Named Entity Recognition	Bahasa Dataset	Modeling	Evaluasi
Case Folding	Filtering	Normalization	Spelling Corrector	Tokenization				
Ya	Ya	Ya	-	Ya	-	Indonesia	IndoBERT	81 %
Kesimpulan: Hasil eksperimen menunjukkan bahwa konfigurasi terbaik diperoleh dengan membekukan enam layer bawah IndoBERT dan melakukan fine-tuning enam layer atas, menghasilkan validation loss sebesar 0,324. Model mampu melakukan deteksi aspek dengan precision 85%-89% serta klasifikasi sentimen dengan akurasi sebesar 84%. Penelitian menunjukkan bahwa fine-tuning IndoBERT pada corpus domain tertentu mampu meningkatkan kemampuan model dalam memahami konteks dan sentimen pada ulasan pengguna berbahasa Indonesia.								

10.	Advances in Pre-trained Language Models for Domain-Specific Text Classification (Zhyar & Kertész, 2025)							
Pre-Processing					Named Entity Recognition	Bahasa Dataset	Modeling	Evaluasi
Case Folding	Filtering	Normalization	Spelling Corrector	Tokenization				
Ya	Ya	-	-	Ya	Ya	Inggris	BERT	87 %
Kesimpulan: Model domain-specific seperti BioBERT dan SciBERT menghasilkan performa lebih baik dibandingkan model BERT umum dengan akurasi mencapai 87%. Hasil tersebut menunjukkan bahwa proses pre-training menggunakan korpus yang sesuai dengan domain tertentu dapat membantu model memahami karakteristik bahasa dan istilah khusus pada suatu bidang.								

Berdasarkan tabel 2. 1 State of The Art Penelitian, dapat disimpulkan bahwa penelitian terkait analisis teks berbasis Natural Language Processing (NLP) telah mengalami perkembangan dari penggunaan model bahasa umum menuju pendekatan berbasis domain-specific corpus dan Transformer model seperti BERT serta IndoBERT. Penelitian terdahulu menunjukkan bahwa penggunaan model berbasis Transformer mampu meningkatkan performa klasifikasi teks dan Named Entity Recognition (NER), terutama ketika dilatih menggunakan data yang sesuai dengan karakteristik domain tertentu. Studi Asri et al. (2025) membuktikan bahwa IndoBERT efektif untuk menganalisis ulasan pengguna PLN Mobile dengan akurasi 81%, namun masih terbatas pada klasifikasi sentimen umum dan belum mampu mengidentifikasi kategori keluhan secara spesifik. Sementara itu, penelitian Zhong dan Goodfellow (2024) serta Suzuki et al. (2023) menunjukkan bahwa pembangunan corpus domain-spesifik dapat meningkatkan kemampuan model dalam memahami istilah khusus pada suatu bidang. Oleh karena itu, penelitian ini mengembangkan pendekatan yang berfokus pada pembangunan corpus domain-spesifik PLN Mobile melalui data ulasan pengguna, kemudian memanfaatkannya untuk meningkatkan pemahaman model IndoBERT dalam melakukan ekstraksi informasi dan analisis pengaduan pengguna secara lebih spesifik.

Penelitian ini menawarkan kontribusi berupa pembangunan corpus domain-spesifik pengaduan dan pelayanan PLN Mobile sebagai sumber data NLP berbahasa Indonesia. Corpus yang dibangun tidak hanya digunakan sebagai dataset pelatihan, tetapi juga menjadi dasar pengembangan model pelabelan otomatis menggunakan IndoBERT untuk mengidentifikasi kategori keluhan pelanggan secara lebih spesifik dibandingkan klasifikasi sentimen umum.

2.2 Landasan Teori

Sebagai landasan melakukan penelitian yang akan dilakukan maka dibahas teori yang berkaitan. Berikut ini saya uraikan landasan teori dari karya ini.

2.2.1 Sistem Informasi

Sistem informasi dalam penelitian ini dipahami bukan hanya sebagai perangkat lunak atau teknologi pengolah data, tetapi sebagai artefak sosio-teknis yang menghubungkan komponen manusia, proses organisasi, data, dan teknologi untuk menghasilkan informasi yang dapat digunakan dalam aktivitas organisasi. Perspektif ini sejalan dengan (Wissuchek & Zschech, 2024), yang menempatkan artefak sistem informasi sebagai sistem yang tersusun atas subsistem sosial dan subsistem teknis. Subsistem sosial mencakup manusia beserta pengetahuan, keterampilan, nilai, struktur, dan pola kerja, sedangkan subsistem teknis mencakup infrastruktur, perangkat keras, perangkat lunak, metode, dan mekanisme pengolahan data. Kedua subsistem tersebut saling berinteraksi untuk mengubah input menjadi output yang mendukung tujuan tertentu.

Pendekatan sosio-teknis menegaskan bahwa manfaat sistem informasi tidak hanya ditentukan oleh kecanggihan teknologi. (Rajaguru et al., 2023) menunjukkan bahwa integrasi sistem informasi antarpihak dapat meningkatkan kualitas informasi hasil peramalan dan selanjutnya meningkatkan kemampuan organisasi dalam merespons perubahan pasar. Penelitian tersebut juga menunjukkan bahwa manfaat integrasi akan lebih kuat ketika terdapat socio-technical congruence, yaitu kesesuaian antara unsur sosial organisasi dan unsur teknis yang digunakan. Dengan demikian, integrasi data dan teknologi perlu disertai kesesuaian proses kerja, kemampuan pengguna, struktur

organisasi, dan pemanfaatan informasi agar keluaran sistem benar-benar memberi nilai bagi organisasi.

Kualitas informasi menjadi penghubung penting antara integrasi sistem dan pemanfaatan informasi. Dalam temuan (Rajaguru et al., 2023), integrasi sistem informasi mendukung tersedianya informasi yang lebih tepat waktu, akurat, andal, dan terperinci. Secara konseptual, prinsip tersebut relevan dengan penelitian ini karena ulasan pengguna PLN Mobile berasal dari sumber dan bentuk teks yang beragam. Data tersebut perlu dikumpulkan, diproses, dianalisis, dan disajikan secara terstruktur sebelum dapat digunakan untuk memahami pola pengaduan. Oleh karena itu, corpus domain-spesifik, hasil Named Entity Recognition menggunakan IndoBERT, analisis pendukung, dan visualisasi dashboard dapat dipandang sebagai rangkaian teknis yang mengubah data mentah menjadi informasi yang lebih terorganisasi dan mudah diinterpretasikan.

Dari perspektif manajemen, (Killedar et al., 2025) menjelaskan bahwa integrasi sistem informasi pada proses pengambilan keputusan strategis dapat memperkuat keputusan berbasis data, meningkatkan kecepatan dan ketepatan keputusan, serta mendukung adaptabilitas organisasi. Namun, keberhasilan integrasi tidak hanya bergantung pada ketersediaan sistem. Kepercayaan manajerial terhadap keluaran sistem, literasi digital, pengalaman pengguna, budaya organisasi, dukungan kepemimpinan, serta kemampuan organisasi mengatasi resistensi dan hambatan teknologi ikut menentukan pemanfaatan sistem informasi. Hal ini memperkuat pandangan bahwa dashboard tidak cukup hanya menghasilkan visualisasi. Informasi yang disajikan harus dapat dipahami, dipercaya, dan sesuai dengan kebutuhan pengguna.

(Killedar et al., 2025) merepresentasikan efektivitas pengambilan keputusan strategis sebagai fungsi dari integrasi sistem informasi, kepercayaan manajerial, dan budaya organisasi. Hubungan tersebut dapat dituliskan secara konseptual sebagai berikut:

$$SDM = f(ISI, MT, OC)$$

Keterangan: SDM merupakan Strategic Decision-Making Effectiveness, ISI merupakan Information Systems Integration, MT merupakan Managerial Trust, dan OC merupakan Organizational Culture. Persamaan tersebut digunakan sebagai representasi konseptual, bukan sebagai persamaan matematis yang dihitung dalam penelitian ini. Intinya, kualitas

keputusan tidak berdiri sendiri sebagai hasil teknologi, tetapi dipengaruhi oleh integrasi teknologi dan faktor manusia-organisasi (Killedar et al., 2025).

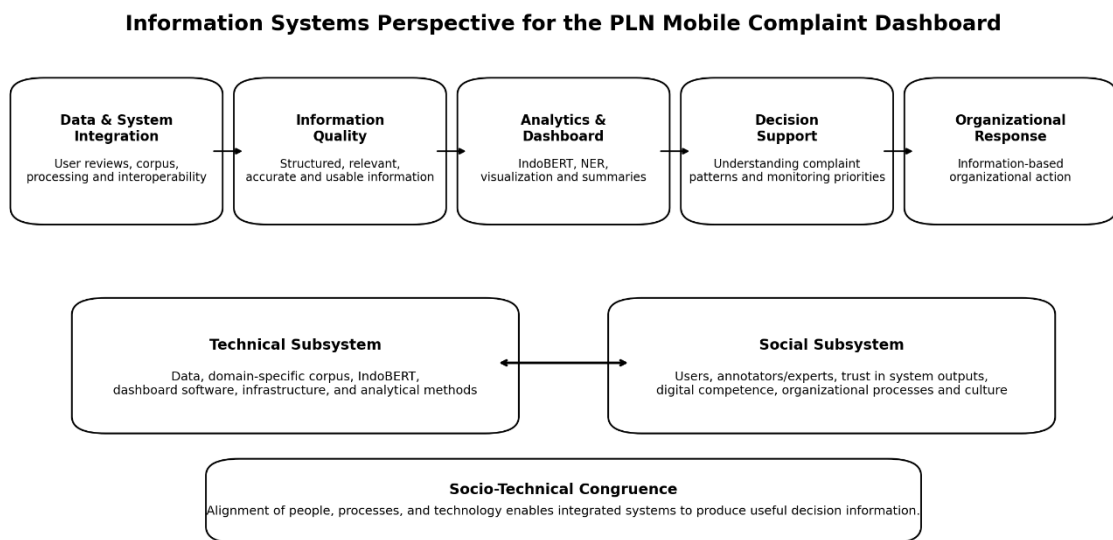
Perspektif sistem informasi juga dapat diperluas melalui sistem analitik. (Wissuchek & Zschech, 2024) meninjau 262 penelitian tentang Prescriptive Analytics Systems (PAS) dan mengidentifikasi 23 komponen penyusun serta empat archetype sistem, yaitu advisory, executive, adaptive, dan self-governing. Kajian tersebut menekankan pembagian peran dan sinergi antara manusia sebagai pengambil keputusan dengan sistem analitik. Penelitian ini tidak memosisikan Dashboard Analisis Pengaduan PLN Mobile sebagai PAS karena dashboard tidak menghasilkan rekomendasi preskriptif atau mengeksekusi keputusan secara otomatis. Namun, perspektif tersebut memperkuat posisi dashboard sebagai artefak sistem informasi yang mendukung manusia dalam memahami informasi dan melakukan analisis sebelum menentukan tindakan.

Tabel 2. 2 Sintesis Teori Sistem Informasi

Sumber	Fokus Teori/Temuan	Konsep Utama
(Rajaguru et al., 2023)	Integrasi sistem informasi memengaruhi kualitas informasi dan respons organisasi. Kesesuaian sosio-teknis memperkuat manfaat integrasi.	Information system integration, information quality, market responsiveness, socio-technical congruence
(Killedar et al., 2025)	Sistem informasi mendukung keputusan berbasis data, tetapi pemanfaatannya dipengaruhi kepercayaan manajerial, literasi digital, budaya, dan kepemimpinan.	Strategic decision-making, managerial trust, organizational culture, digital transformation
(Wissuchek & Zschech, 2024)	Sistem analitik dipandang sebagai artefak sistem informasi yang menggabungkan subsistem	IS artifact, socio-technical systems, analytics, human-system synergy

sosial dan teknis serta pembagian
peran manusia dan sistem.

Berdasarkan ketiga penelitian tersebut, sistem informasi dapat dirumuskan sebagai suatu kesatuan sosio-teknis yang mengintegrasikan data dan teknologi dengan manusia serta proses organisasi untuk menghasilkan informasi berkualitas yang mendukung pemahaman kondisi dan pengambilan keputusan. Nilai sistem muncul ketika integrasi teknis menghasilkan informasi yang relevan, sedangkan pengguna memiliki kemampuan dan kepercayaan untuk memanfaatkan informasi tersebut. Kerangka ini sesuai dengan tujuan penelitian yang tidak berhenti pada pemodelan NLP, tetapi menerapkan hasil analisis ke dalam Dashboard Analisis Pengaduan PLN Mobile.



Gambar 2. 1 Perspektif Sistem Informasi untuk Dasbor Pengaduan PLN Mobile

Tabel 2. 3 Pemetaan konsep sistem informasi terhadap penelitian Dashboard Analisis Pengaduan PLN Mobile

Konsep Sistem Informasi	Implementasi dalam Penelitian
Input dan integrasi data	Data ulasan pengguna PLN Mobile dari Google Play Store dan Platform X

	dikumpulkan dan disatukan sebagai sumber analisis.
Subsistem teknis	Preprocessing, corpus domain-spesifik, Entity Ruler, IndoBERT, NER, penyimpanan data, dan dashboard.
Subsistem sosial	Peneliti, anotator/pakar, pengguna dashboard, serta pihak yang menginterpretasikan hasil analisis.
Kualitas informasi	Hasil pengolahan disajikan dalam bentuk entitas, kategori pengaduan, tren, distribusi, sentimen pendukung, dan hasil validasi.
Interaksi manusia-sistem	Pengguna membaca, memfilter, membandingkan, dan menginterpretasikan informasi melalui fitur dashboard.
Dukungan keputusan	Dashboard menyediakan informasi terstruktur untuk membantu pemantauan pola pengaduan dan menentukan fokus evaluasi layanan.
Kesesuaian sosio-teknis	Fungsi, visualisasi, dan keluaran sistem diselaraskan dengan kebutuhan pengguna agar teknologi dapat dimanfaatkan secara efektif.

Dalam penelitian ini, dashboard berfungsi sebagai lapisan penyajian informasi yang menghubungkan hasil pengolahan teknis dengan kebutuhan pengguna. Data pengaduan yang semula tidak terstruktur diproses melalui tahapan preprocessing, pembentukan corpus domain-spesifik, pelabelan, dan pemodelan IndoBERT. Hasil tersebut kemudian divisualisasikan dalam dashboard agar pengguna dapat melihat ringkasan dataset, hasil preprocessing, karakteristik corpus, hasil NER, sentimen

pendukung, validasi pakar, serta pola pengaduan. Dengan demikian, kontribusi sistem informasi penelitian ini terletak pada integrasi antara data, analitik, model bahasa, visualisasi, dan interaksi pengguna menjadi satu media informasi yang mendukung pemantauan dan analisis layanan PLN Mobile.

2.2.2 Natural Language Processing (NLP)

Natural Language Processing (NLP) adalah cabang ilmu yang mempelajari bagaimana komputer memproses, memahami, dan menghasilkan bahasa manusia. Dalam kerangka yang lebih luas, NLP berada di persimpangan linguistik, ilmu komputer, dan kecerdasan buatan. Jurafsky dan Martin menjelaskan bahwa NLP mencakup berbagai tugas seperti klasifikasi teks, sequence labeling, information extraction, machine translation, dan question answering. Dalam penelitian ini, NLP menjadi fondasi karena data yang diolah berupa teks ulasan/keluhan pelanggan PLN Mobile yang bersifat tidak terstruktur dan perlu diubah menjadi representasi yang dapat diproses model (Daniel Jurafsky et al., 2026).

Secara konseptual, NLP bertujuan mengubah data teks mentah menjadi informasi yang bermakna. Pada penelitian ini, makna tersebut bukan hanya polaritas sentimen, tetapi juga jenis keluhan yang terkandung dalam ulasan pengguna. Karena itu, penelitian ini tidak berhenti pada text classification biasa, tetapi bergerak ke pelabelan otomatis berbasis entitas/kategori yang lebih kaya informasi (Malik & Bilal, 2024).

2.2.2.1 Text Classification

Text classification merupakan salah satu tugas utama dalam bidang Natural Language Processing (NLP) yang bertujuan untuk mengategorikan data berbentuk teks ke dalam kelas atau kategori tertentu berdasarkan karakteristik linguistik dan pola informasi yang terdapat dalam dokumen. Proses klasifikasi teks dilakukan dengan membangun suatu model yang mampu mempelajari hubungan antara representasi teks sebagai masukan (*input*) dan label kategori sebagai keluaran (*output*). Dengan demikian, sistem dapat melakukan prediksi kategori terhadap data teks baru secara otomatis (Z. Li et al., 2022).

Automatic text classification merupakan proses pengelompokan dokumen teks secara otomatis ke dalam satu atau beberapa kelas yang telah ditentukan sebelumnya berdasarkan topik atau karakteristik tertentu. Dalam perkembangan NLP modern, klasifikasi teks menjadi komponen penting karena mampu membantu proses pengorganisasian, ekstraksi informasi, dan analisis terhadap jumlah data tekstual yang besar (Silva Barbon & Akabane, 2022).

Secara umum, text classification termasuk dalam kategori tugas supervised learning, yaitu pendekatan pembelajaran mesin yang membutuhkan data berlabel sebagai dasar proses pelatihan model. Dataset teks terlebih dahulu diberikan label sesuai kategori yang telah ditentukan, kemudian model mempelajari pola hubungan antara teks dan label tersebut untuk melakukan klasifikasi terhadap data yang belum pernah dilihat sebelumnya. Kualitas hasil klasifikasi sangat dipengaruhi oleh jumlah, keseimbangan, dan kualitas data pelatihan yang digunakan. Silva Barbon dan Akabane (2022) menunjukkan bahwa karakteristik dataset, seperti distribusi jumlah sampel antar kelas, dapat memengaruhi performa model klasifikasi teks. Dataset yang tidak seimbang (*imbalanced dataset*) dapat menyebabkan model mengalami kesulitan dalam mengenali kelas tertentu karena jumlah representasi data antar kategori tidak proporsional.

Dalam penelitian ini, text classification diterapkan untuk mengategorikan ulasan pengguna aplikasi PLN Mobile berdasarkan kategori permasalahan atau layanan tertentu. Data ulasan yang telah melalui proses anotasi corpus digunakan sebagai dataset berlabel untuk melatih model klasifikasi. Melalui pendekatan ini, model dapat mempelajari pola bahasa pengguna dan melakukan prediksi kategori secara otomatis terhadap ulasan baru. Oleh karena itu, text classification menjadi tahapan penting dalam pengembangan corpus domain-spesifik karena memungkinkan corpus yang telah dianotasi dimanfaatkan untuk membangun sistem klasifikasi berbasis NLP.

Dalam penelitian ini, text classification bukan merupakan tugas utama pada pemodelan entitas. Tugas utama penelitian menggunakan Named Entity Recognition yang melakukan pelabelan pada tingkat token. Konsep text classification digunakan sebagai landasan pada analisis sentimen pendukung, karena satu label sentimen diberikan

pada tingkat ulasan. Perbedaan tersebut penting agar evaluasi klasifikasi sentimen tidak disamakan dengan evaluasi NER.

2.2.2.2 *Text Preprocessing*

Sebelum pemodelan dilakukan, data teks umumnya perlu melalui tahap preprocessing atau praproses. Dalam NLP, praproses bertujuan membersihkan dan menormalkan teks agar lebih siap digunakan sebagai input model. Tahapan ini dapat mencakup penghapusan duplikasi, pembersihan karakter yang tidak relevan, normalisasi huruf, penanganan typo, pemisahan token, serta seleksi data yang benar-benar merepresentasikan keluhan pelanggan (Garg & Sharma, 2022).

Pada tahap praproses memiliki peran yang sangat penting karena data ulasan aplikasi biasanya mengandung bahasa informal, singkatan, campuran istilah teknis, dan struktur kalimat yang tidak baku. Jika tahap ini tidak dilakukan dengan baik, maka kualitas corpus akan menurun dan model dapat belajar pola yang salah. Karena itu, pengembangan corpus domain-spesifik tidak dapat dipisahkan dari proses praproses dan kurasi data (Dąbrowski et al., 2022).

2.2.2.3 *Text Mining*

Text Mining adalah proses mengekstrak informasi dan pola penting dari data teks yang tidak terstruktur untuk mendukung analisis lebih lanjut. Dalam penelitian ini, *Text Mining* digunakan sebagai tahap eksplorasi data sebelum pembangunan corpus domain-spesifik dan proses labeling. Tujuan *Text Mining* adalah untuk menemukan istilah dan kosakata khas domain yang sering muncul pada keluhan pelanggan, menganalisis frekuensi dan distribusi kategori keluhan, serta membantu preprocessing teks, seperti identifikasi stopwords, bahasa informal, singkatan, atau kesalahan ketik, sehingga corpus domain-spesifik menjadi bersih dan representatif (Kaur & Sharma, 2023). Dalam penelitian ini, konsep text mining digunakan untuk mendukung eksplorasi karakteristik corpus melalui analisis frekuensi kata, distribusi istilah domain, ukuran vocabulary, serta visualisasi corpus. Hasil eksplorasi digunakan untuk memahami karakteristik bahasa pengguna PLN Mobile dan mengevaluasi representativitas corpus yang dibangun (Zhou et al., 2024).

2.2.2.4 Labelling

Labeling dalam konteks *Natural Language Processing* (NLP) adalah proses memberikan tag atau kategori pada unit teks agar komputer dapat mengenali makna, entitas, atau sentimen secara otomatis. Pada penelitian ini, labeling digunakan untuk menandai jenis keluhan pelanggan berdasarkan kategori domain, seperti transaksi, layanan, verifikasi, fitur aplikasi, dan keluhan teknis. Labeling dapat dilakukan secara manual untuk membangun corpus domain-spesifik yang akurat, yang kemudian digunakan untuk melatih model berbasis IndoBERT (Wan et al., 2024).

Dengan adanya labeling yang tepat, model dapat mempelajari pola bahasa domain, termasuk istilah teknis dan konteks kalimat, sehingga pelabelan otomatis menjadi lebih konsisten, efisien, dan mampu mendukung identifikasi entitas masalah pada ulasan negative dengan akurat. Misalnya, kalimat “Pembayaran token gagal meski saldo sudah terpotong” akan diberi label kategori transaksi dan sentimen negatif, sedangkan “Petugas lambat merespons gangguan listrik” akan diberi label layanan dan sentimen negatif. Labeling menjadi komponen krusial dalam membangun sistem pelabelan otomatis, karena kualitas data berlabel langsung mempengaruhi performa model IndoBERT pada tahap fine-tuning dan evaluasi menggunakan metrik seperti *accuracy*, *precision*, *recall*, dan *F1-score* (Ma et al., 2021).

2.2.2.5 Spelling Corrector

Spelling corrector merupakan salah satu tahap dalam preprocessing pada bidang Natural Language Processing (NLP) yang bertujuan untuk memperbaiki kesalahan penulisan pada teks sebelum dilakukan proses analisis lebih lanjut (Lefrandt et al., 2025). Kesalahan ejaan sering ditemukan pada data teks yang berasal dari pengguna karena adanya penggunaan bahasa informal, singkatan, kesalahan ketik (typo), penghilangan karakter, maupun penggunaan istilah tidak baku. Dalam analisis teks berbasis NLP, kesalahan ejaan dapat menyebabkan perbedaan representasi antara kata yang memiliki makna sama. Sebagai contoh, kata "aplikasi error", "aplikasi eror", dan "aplikasi erorr" dapat dianggap sebagai tiga token berbeda oleh model apabila tidak dilakukan

normalisasi. Kondisi tersebut dapat menurunkan kualitas ekstraksi fitur dan performa model (Okgetheng et al., 2022).

Spelling corrector bekerja dengan melakukan identifikasi terhadap kata yang tidak sesuai dengan kamus atau pola bahasa tertentu, kemudian mencari kandidat kata yang paling mendekati berdasarkan tingkat kemiripan karakter maupun konteks kalimat. Salah satu pendekatan yang umum digunakan adalah edit distance. Menurut (Patil et al., 2021), edit distance merupakan metode pengukuran jarak antara dua kata berdasarkan jumlah operasi perubahan minimum yang diperlukan untuk mengubah suatu kata menjadi kata lain. Operasi tersebut meliputi:

- 1) Insertion, yaitu penambahan karakter.
- 2) Deletion, yaitu penghapusan karakter.
- 3) Substitution, yaitu penggantian karakter.

Sebagai contoh: "apliksi" menjadi "aplikasi". Perubahan tersebut membutuhkan satu operasi penambahan huruf "a", sehingga memiliki jarak Levenshtein sebesar 1.

Dalam penelitian berbasis teks pengguna seperti ulasan aplikasi, spelling corrector memiliki peran penting karena data yang diperoleh berasal dari bahasa alami pengguna yang memiliki variasi penulisan. Proses koreksi ejaan dapat meningkatkan konsistensi token sehingga model NLP lebih mudah mengenali pola dan hubungan antar kata. Pada penelitian ini, spelling corrector digunakan sebagai salah satu tahap preprocessing untuk meningkatkan kualitas data ulasan PLN Mobile sebelum dilakukan proses pembentukan corpus, pelabelan entitas, dan pemodelan menggunakan IndoBERT.

2.2.3 Data Scraping

Data scraping merupakan teknik pengumpulan data secara otomatis dari sumber digital menggunakan perangkat lunak atau program komputer. Teknik ini digunakan untuk mengambil informasi yang tersedia pada halaman web, platform digital, maupun layanan berbasis aplikasi sehingga data dapat diproses lebih lanjut untuk kebutuhan

analisis. Berbeda dengan pengumpulan data secara manual, data scraping memungkinkan proses akuisisi data dalam jumlah besar dengan waktu yang lebih efisien. Teknik ini banyak digunakan dalam penelitian berbasis Natural Language Processing (NLP), terutama pada analisis teks dari media sosial, ulasan pengguna, forum daring, dan platform digital lainnya (Dhendra & Gayuh Utomo, 2025).

Menurut (Boegershausen et al., 2022), web scraping merupakan metode ekstraksi informasi dari sumber web secara otomatis dengan memanfaatkan struktur dokumen digital untuk memperoleh data yang sesuai dengan kebutuhan analisis. Metode ini banyak diterapkan dalam penelitian berbasis data karena mampu menghasilkan dataset berskala besar dari sumber yang terus berkembang.

Dalam penelitian ini, data scraping digunakan untuk memperoleh data teks berupa ulasan pengguna aplikasi PLN Mobile yang berasal dari Google Play Store dan Platform X. Data yang diperoleh kemudian digunakan sebagai dataset utama untuk proses pembangunan corpus domain-spesifik dan analisis menggunakan model IndoBERT.

2.2.4 Aplikasi PLN Mobile

PLN Mobile adalah aplikasi layanan pelanggan resmi PT PLN yang menjadi kanal utama interaksi digital antara pelanggan dan penyedia layanan ketenagalistrikan. Melalui aplikasi ini pelanggan dapat menyampaikan pengaduan gangguan, membeli token listrik, membayar tagihan, mengajukan penambahan daya, serta mengakses layanan turunan seperti ICONNET dan ListriQu. Karena fungsinya mencakup pelaporan gangguan sekaligus transaksi, teks yang dihasilkan pengguna memuat dua jenis persoalan yang berbeda, yaitu persoalan layanan di lapangan dan persoalan perangkat lunak.

Untuk keperluan penanganan internal, PLN menggunakan skema klasifikasi berjenjang atas umpan balik pengguna yang terdiri atas tujuh kelas utama dan 64 subclass. Ketujuh kelas utama tersebut adalah Service O2O yang memuat persoalan tindak lanjut layanan di lapangan, Feature Error yang memuat kegagalan fungsi aplikasi, Umum yang memuat persoalan di luar keduanya, Feature Design yang memuat masukan dan usulan atas rancangan fitur, Registration yang memuat persoalan pendaftaran dan autentikasi,

serta ICONNET dan ListriQu yang memuat persoalan pada dua layanan turunan. Skema inilah yang diadopsi sebagai kerangka anotasi pada penelitian ini, sehingga korpus yang dihasilkan selaras dengan kategori yang benar-benar digunakan dalam praktik penanganan pengaduan, bukan kategori yang disusun sendiri oleh peneliti.

Struktur taksonomi tersebut memiliki dua ciri yang berimplikasi metodologis. Pertama, taksonomi ini sesungguhnya bersumbu ganda, yaitu sumbu lini layanan dan sumbu jenis persoalan. Hal itu terlihat dari kelas ICONNET dan ListriQu yang subclass-nya justru mengulang nama kategori pada tingkat utama, seperti Registration ICONNET dan Service O2O ListriQu. Kedua, sejumlah topik yang sama muncul pada lebih dari satu kelas dan hanya dibedakan oleh maksud penutur. Persoalan pembayaran, pengaduan, Swacam, SPKLU, dan penambahan daya masing-masing tercatat pada kelas Feature Error sekaligus Feature Design, dengan pembeda berupa apakah pengguna melaporkan kegagalan atau mengusulkan perbaikan. Kedua ciri ini menjelaskan mengapa pemipihan taksonomi menjadi satu himpunan 64 label tunggal menghasilkan kategori yang sulit dibedakan, baik oleh anotator manusia maupun oleh model.

Klasifikasi tersebut disusun untuk mengakomodasi berbagai bentuk keluhan pelanggan, baik yang bersifat teknis maupun nonteknis, sehingga proses identifikasi dan penanganan pengaduan dapat dilakukan secara lebih sistematis, konsisten, terarah, dan terstandar pada seluruh unit pelayanan. Secara umum, struktur klasifikasi pengaduan tersebut dapat dijabarkan sebagai berikut:

1. *Klasifikasi Service o2o*

Tabel 2. 4 Klasifikasi *Service o2o*

Kode	Sub-Class
1A	Respon lambat terhadap laporan Gangguan
1B	Respon lambat terhadap laporan Keluhan

1C	PB/PD/PS/ Migrasi belum ada realisasi
1D	Complain terkait layanan pra/pascabayar
1E	Gangguan selesai/batal tanpa ditindaklanjuti lapangan
1F	Ketidaksesuaian layanan petugas Unit
1G	Permasalahan selesai/ batal tanpa ditindaklanjuti lapangan

2. *Klasifikasi Feature Error*

Tabel 2. 5 Klasifikasi *Feature Error*

Kode	Sub-Class
2A	Gangguan server
2B	Status payment gagal tapi saldo sudah terpotong
2C	Fungsi fitur gagal
2D	Pengaduan tidak bisa di submit/tidak bisa digunakan
2E	Fitur SPKLU
2F	PB/PD
2G	Gagal input IDPEL
2H	Swacam tidak dapat digunakan
2I	Fitur Home Charging
2J	Kesalahan koordinat lokasi
2K	Permintaan verifikasi ID tidak dapat diajukan
2L	Fitur SPLU

2M	Fitur SPBKLU
----	--------------

3. *Klasifikasi Feature Umum*

Tabel 2. 6 Klasifikasi *Feature Umum*

Kode	Sub-Class
3A	Lain-lain
3B	Tentang promo/stimulus
3C	UU PDP
3D	Marketplace
3E	Reward
3F	Keluhan Probis
3G	Website

4. *Klasifikasi Feature Design*

Tabel 2. 7 Klasifikasi *Feature Design*

Kode	Sub-Class
4A	Fitur Payment
4B	Fitur PB/PD
4C	Fitur notifikasi pemadaman
4D	Fitur Pengaduan
4E	Saran masukan User interface (UI)
4F	Fitur history pemakaian
4G	Fitur Live Chat
4H	History pembelian token/pembayaran tagihan
4I	Fitur Swacam
4J	Fitur ADD ID Pel
4K	Saran fitur baru

4L	Fitur GPS
4M	Fitur SPLU
4N	Fitur SPBKLU
4O	Tentang promo/stimulus
4P	Fitur Iconnet
4Q	Fitur SPKLU

5. *Klasifikasi Registration*

Tabel 2. 8 Klasifikasi *Registration*

Kode	Sub-Class
5A	Sulit login/tidak bisa login
5B	Sulit mendaftar/tidak bisa mendaftar
5C	Masalah OTP
5D	Masalah verifikasi
5E	No. Hp sudah terdaftar dengan nama orang lain
5F	Tidak bisa ubah data diri

6. *Klasifikasi ICONNET*

Tabel 2. 9 Klasifikasi *ICONNET*

Kode	Sub-Class
6A	Registration ICONNET
6B	Service O2O ICONNET

6C	Feature Payment ICONNET
6D	Feature Error ICONNET
6E	Others (Umum) ICONNET
6F	Feature Design ICONNET

7. *Klasifikasi ListriQu*

Tabel 2. 10 Klasifikasi *ListriQu*

Kode	Sub-Class
7A	Service O2O LISTRI-Qu
7B	Feature Error LISTRI-Qu
7C	Others (Umum) LISTRI-Qu
7D	Feature Payment LISTRI-Qu
7E	Registration LISTRI-Qu
7F	Service O2O LISTRI-Qu

2.2.5 Named Entity Recognition (NER)

Named Entity Recognition (NER) adalah tugas pemrosesan bahasa alami yang bertujuan mengenali dan mengklasifikasikan rentang teks yang merujuk pada entitas bernama ke dalam kategori yang telah ditetapkan. Berbeda dengan text classification yang memberikan satu label untuk keseluruhan dokumen, NER bekerja pada tingkat token sehingga mampu menunjukkan bagian mana dari sebuah kalimat yang memuat informasi penting (Jurafsky & H. Martin, 2026). Perbedaan granularitas inilah yang mendasari penggunaan NER pada teks pengaduan pelanggan, sebab satu ulasan dapat memuat aspek layanan, jenis masalah, serta identitas produk sekaligus, sehingga satu label pada tingkat dokumen tidak memadai untuk kebutuhan analisis.

Penelitian NER berbahasa Indonesia pada domain spesifik masih tergolong sedikit. Dave dan Chowanda (2024) menyatakan bahwa penelitian NER bahasa Indonesia

yang berfokus pada domain tertentu masih minim, lalu mengembangkan ekstraksi entitas keuangan personal menggunakan arsitektur IndoBERT-BiGRU-CRF dan menyoroti persoalan out-of-vocabulary pada entitas domain. Pada domain hukum, (Yulianti et al., 2024) membangun dataset legal entity recognition dari dokumen putusan pengadilan Indonesia dan membandingkan beberapa model berbasis transformer, di antaranya IndoBERT, M-BERT, RoBERTa, dan XLM-RoBERTa. Kedua penelitian tersebut memperlihatkan bahwa pembangunan korpus beranotasi entitas untuk domain tertentu masih menjadi kebutuhan nyata dalam penelitian NLP bahasa Indonesia, sekaligus memperkuat posisi penelitian ini pada domain pengaduan kelistrikan.

2.2.5.1 Skema Pelabelan BIO

Agar tugas NER dapat diselesaikan sebagai persoalan sequence labeling, anotasi entitas perlu direpresentasikan dalam bentuk label per token. Skema yang paling banyak digunakan adalah BIO, yang pertama kali diperkenalkan Ramshaw dan Marcus (1995) untuk tugas text chunking. Pada skema ini, token pertama sebuah entitas diberi label B- (begin), token lanjutan dalam entitas yang sama diberi label I- (inside), dan token yang bukan bagian dari entitas mana pun diberi label O (outside). Dengan demikian, untuk n tipe entitas akan terbentuk $2n + 1$ label.

Terdapat beberapa varian skema pelabelan selain BIO, yaitu IO yang meniadakan label B- serta BIOES yang menambahkan label E- untuk token akhir entitas dan S- untuk entitas satu token (Jurafsky & H. Martin, 2026). Skema IO tidak dipilih pada penelitian ini karena tidak mampu memisahkan dua entitas sejenis yang berdampingan, sedangkan BIOES tidak dipilih karena menggandakan jumlah kelas label dan berpotensi memperparah ketidakseimbangan data pada kelas entitas yang jarang muncul. Skema BIO dinilai paling sesuai karena memberikan keseimbangan antara ketepatan penentuan batas entitas dan jumlah kelas label yang harus dipelajari model.

2.2.6 Cohen's Kappa

Cohen's Kappa digunakan untuk mengukur tingkat kesepakatan antar-anotator dengan mempertimbangkan kemungkinan kesepakatan yang terjadi secara kebetulan.

Berbeda dengan persentase kesepakatan sederhana (*raw agreement*) yang hanya menghitung jumlah kesamaan label, Cohen's Kappa melakukan koreksi terhadap peluang kesepakatan acak sehingga memberikan gambaran reliabilitas yang lebih representatif. Studi terbaru mengenai kualitas anotasi dataset NLP juga menunjukkan bahwa pengukuran kesepakatan antar-anotator tidak cukup hanya menggunakan persentase kesesuaian karena dapat terpengaruh oleh ketidakseimbangan distribusi label dan jumlah kategori yang digunakan dalam anotasi (Paun et al., 2022) .

Secara matematis, Cohen's Kappa dirumuskan sebagai berikut:

$$\kappa = \frac{(P_o - P_e)}{(1 - P_o)} \quad (2.1)$$

κ merupakan nilai koefisien Cohen's Kappa, P_o merupakan proporsi kesepakatan aktual antara dua anotator (*observed agreement*), sedangkan P_e merupakan proporsi kesepakatan yang diperkirakan terjadi secara kebetulan (*expected agreement*). Nilai κ berada pada rentang -1 hingga 1. Nilai 1 menunjukkan kesepakatan sempurna antar-anotator, nilai 0 menunjukkan tingkat kesepakatan yang setara dengan peluang acak, sedangkan nilai negatif menunjukkan adanya ketidaksepakatan yang lebih besar dibandingkan kesepakatan yang diharapkan secara kebetulan. Dalam penelitian anotasi NLP terbaru, Cohen's Kappa masih banyak digunakan sebagai indikator reliabilitas, khususnya pada proses pembangunan dataset dengan dua anotator independen (Klie et al., 2024).

Interpretasi nilai Cohen's Kappa pada penelitian ini menggunakan kategori tingkat kesepakatan yang umum digunakan dalam penelitian anotasi, yaitu nilai kurang dari 0,00 menunjukkan *poor agreement*, 0,00–0,20 menunjukkan *slight agreement*, 0,21–0,40 menunjukkan *fair agreement*, 0,41–0,60 menunjukkan *moderate agreement*, 0,61–0,80 menunjukkan *substantial agreement*, dan 0,81–1,00 menunjukkan *almost perfect agreement*. Namun, penelitian terbaru menekankan bahwa interpretasi nilai kappa tidak dapat dilepaskan dari konteks tugas anotasi, jumlah kategori, serta distribusi setiap kelas label. Nilai kappa yang rendah tidak selalu menunjukkan kualitas anotasi yang buruk, terutama pada tugas dengan jumlah kelas besar atau distribusi label yang tidak seimbang.

Dalam bidang linguistik komputasional dan pembangunan dataset NLP, pengukuran reliabilitas anotasi juga perlu mempertimbangkan karakteristik tugas yang dilakukan. Beberapa penelitian menunjukkan bahwa perbedaan interpretasi antar-anotator sering muncul akibat batasan definisi kategori, kompleksitas konteks bahasa, serta ambiguitas makna teks. Oleh karena itu, nilai Cohen's Kappa sebaiknya dilaporkan bersama penjelasan mengenai pedoman anotasi, jumlah kategori, dan karakteristik data yang digunakan agar interpretasinya lebih objektif.

Pada penelitian ini, Cohen's Kappa digunakan untuk mengevaluasi reliabilitas anotasi manual yang dilakukan oleh dua anotator pakar terhadap data validasi. Pengukuran dilakukan secara terpisah berdasarkan tiga aspek anotasi, yaitu kelas utama, kode subclass, dan sentimen. Pemisahan pengukuran tersebut dilakukan karena setiap aspek memiliki tingkat granularitas dan kompleksitas kategori yang berbeda. Hasil Cohen's Kappa kemudian digunakan sebagai indikator kualitas kesepakatan anotator sebelum hasil adjudikasi ditetapkan sebagai *gold standard* untuk validasi eksternal model Named Entity Recognition.

2.2.7 Corpus

Corpus (korpus) merupakan kumpulan data bahasa yang dikumpulkan, disusun, dan disimpan secara sistematis sehingga dapat digunakan untuk analisis bahasa berbasis komputer. Dalam bidang Natural Language Processing (NLP), corpus menjadi sumber data utama yang digunakan untuk memahami pola bahasa, karakteristik teks, serta hubungan antar informasi yang terdapat dalam data. Berbeda dengan kumpulan dokumen biasa, corpus memiliki struktur yang terorganisasi sehingga memungkinkan proses analisis dilakukan secara lebih sistematis (Meyer, 2023).

Pembangunan sebuah corpus memerlukan perencanaan yang jelas mengenai tujuan penggunaan, jenis data yang dikumpulkan, bentuk teks, ukuran data, serta tingkat informasi tambahan yang akan diberikan pada corpus. Keputusan tersebut akan menentukan karakteristik corpus yang dibangun, termasuk jenis dokumen yang dimasukkan, format penyimpanan data, dan tingkat anotasi yang diperlukan untuk kebutuhan analisis.

Dalam pengembangan corpus, data tidak hanya terdiri atas teks mentah, tetapi juga dapat dilengkapi dengan informasi pendukung seperti metadata, struktur dokumen, dan anotasi. Metadata memberikan informasi mengenai karakteristik data yang terdapat dalam corpus, sedangkan anotasi memberikan informasi tambahan berupa label atau kategori tertentu yang memungkinkan data dipahami oleh sistem komputasi. Proses anotasi menjadi bagian penting karena memungkinkan corpus digunakan sebagai dataset terstruktur dalam pengembangan model NLP.

Anotasi pada corpus merupakan proses pemberian informasi tambahan terhadap data teks berdasarkan tujuan penelitian. Bentuk anotasi dapat berupa pemberian kategori linguistik, informasi semantik, maupun label tertentu sesuai kebutuhan analisis. Dalam penelitian berbasis pembelajaran mesin, anotasi memiliki fungsi penting karena menghasilkan data berlabel (*labeled dataset*) yang digunakan untuk proses pelatihan dan evaluasi model (Friginal, 2024).

Berdasarkan tujuan penggunaannya, corpus dapat dikembangkan menjadi corpus umum maupun *corpus* khusus (*specialized corpus*). *Corpus* umum biasanya mencakup berbagai jenis teks dan digunakan untuk menggambarkan penggunaan bahasa secara luas, sedangkan corpus khusus dikembangkan untuk merepresentasikan karakteristik bahasa pada bidang tertentu. *Corpus* khusus memiliki kemampuan lebih baik dalam menangkap istilah, pola komunikasi, serta konteks yang spesifik dibandingkan *corpus* umum karena data yang digunakan telah disesuaikan dengan domain penelitian.

2.2.7.1 *Corpus Domain Spesifik*

Corpus domain-spesifik merupakan corpus yang dibangun dengan fokus pada suatu bidang atau permasalahan tertentu. Corpus jenis ini berisi data teks yang memiliki karakteristik khusus sesuai dengan domain yang diteliti, sehingga mampu merepresentasikan kosakata, istilah teknis, dan pola bahasa yang muncul dalam bidang tersebut. Dalam penelitian NLP, keberadaan corpus domain-spesifik menjadi penting karena model bahasa membutuhkan data yang relevan agar mampu memahami konteks suatu domain secara lebih akurat (Chen et al., 2021). menjelaskan bahwa pembangunan corpus berbasis domain dapat meningkatkan kemampuan model dalam memahami

karakteristik bahasa khusus karena data yang digunakan lebih sesuai dengan konteks penerapan.

Pada penelitian ini, corpus domain-spesifik dikembangkan berdasarkan data pengaduan dan pelayanan pengguna aplikasi PLN Mobile. Data ulasan pelanggan memiliki karakteristik yang berbeda dibandingkan teks umum karena mengandung istilah layanan kelistrikan, permasalahan teknis, serta bentuk ekspresi pengguna yang bersifat informal. Oleh karena itu, pembangunan corpus khusus diperlukan agar model NLP mampu mengenali pola bahasa dan kategori keluhan yang terdapat pada domain layanan digital PLN.

Tahapan pembangunan corpus domain-spesifik meliputi proses pengumpulan data, seleksi data, preprocessing, anotasi, dan penyimpanan data dalam format yang terstruktur. Tahapan tersebut bertujuan memastikan bahwa corpus yang dihasilkan memiliki kualitas data yang baik serta dapat digunakan sebagai sumber pembelajaran bagi model NLP. Chen et al. menekankan bahwa kualitas corpus sangat dipengaruhi oleh proses konstruksi data, termasuk pemilihan sumber data, proses pembersihan teks, serta strategi anotasi yang digunakan (Gu et al., 2021).

Urgensi corpus domain-spesifik didukung oleh penelitian Zhong dan Goodfellow yang menunjukkan bahwa domain-specific pretraining dapat meningkatkan performa model BERT/RoBERTa dibanding model yang hanya dilatih dengan corpus umum (Zhong & Goodfellow, 2024). Dalam dua tugas NLP, yaitu text classification dan named entity recognition, model domain-spesifik pada corpus konstruksi meningkatkan F1-score sebesar 5,9% dan 8,5%. Temuan ini relevan dengan penelitian ini karena menunjukkan bahwa kualitas dan kekhususan corpus berpengaruh langsung terhadap hasil pemodelan (Wu et al., 2024).

2.2.8 Rule-Based Annotation dan Entity Ruler

Pelabelan manual pada korpus berskala besar menuntut biaya dan waktu yang tinggi, sehingga banyak penelitian memanfaatkan *weak supervision*, yaitu pembentukan label awal secara otomatis menggunakan aturan, kamus istilah, atau *gazetteer*. (Lison et

al., 2021) mengembangkan kerangka kerja *skweak* yang menyusun sejumlah fungsi pelabelan heuristik dan menggabungkan keluarannya menjadi satu label probabilistik, dan menunjukkan bahwa pendekatan ini mampu menghasilkan data latih NER dalam skala besar tanpa anotasi manual. (Nastou et al., 2024) melaporkan bahwa pelabelan berbasis kamus tetap menjadi tulang punggung pembangunan korpus domain karena sifatnya yang deterministik dan dapat direplikasi, meskipun kelengkapan kamus menjadi penentu utama kualitas hasilnya.

Entity Ruler merupakan komponen pelabelan berbasis aturan pada pustaka spaCy yang bekerja dengan mencocokkan pola yang telah didefinisikan sebelumnya terhadap teks, kemudian menandai rentang yang cocok sebagai entitas beserta label tipenya. Pola dinyatakan dalam dua bentuk. Bentuk pertama adalah *phrase pattern*, yaitu pencocokan langsung terhadap untaian kata yang sudah pasti, yang sesuai untuk kamus istilah teknis dan nama entitas. Bentuk kedua adalah *token pattern*, yaitu pencocokan berbasis atribut token seperti bentuk kata dasar atau kelas kata, yang memungkinkan satu pola menjangkau beberapa variasi permukaan. Pembedaan ini relevan pada penelitian ini karena kamus domain yang disusun terdiri atas 85 istilah tunggal dan 333 istilah frasa, sehingga sebagian besar entri harus ditangani sebagai pola frasa multitoken, bukan sebagai kata tunggal.

Keunggulan pendekatan berbasis aturan adalah konsistensi penuh dan kemampuan menjangkau seluruh korpus dalam waktu singkat, sehingga sesuai untuk membentuk label awal pada puluhan ribu dokumen. Kelemahannya bersifat struktural: entitas yang belum terdaftar dalam kamus tidak akan pernah terdeteksi, sehingga aturan menghasilkan anotasi yang tidak lengkap. Persoalan ini dikenal sebagai *incomplete annotation* pada supervisi jarak jauh, dan (Jiang et al., 2021) menunjukkan bahwa kombinasi data berlabel kuat berjumlah kecil dengan data berlabel lemah berjumlah besar memberikan hasil yang lebih baik daripada mengandalkan salah satunya saja.

Karena keterbatasan tersebut, label hasil aturan tidak dapat diperlakukan sebagai *ground truth*. Penelitian ini menempatkan hasil pelabelan Entity Ruler sebagai anotasi awal, kemudian memvalidasinya melalui anotasi pakar pada sebagian data untuk

membentuk *gold standard* yang menjadi acuan pengukuran. Konsekuensi metodologis dari rancangan ini adalah bahwa kinerja model terhadap label aturan dan kesesuaian model terhadap *gold standard* merupakan dua besaran yang berbeda dan harus dilaporkan secara terpisah.

2.2.9 Ketidakseimbangan Kelas dan Cross Validation

Banyak tugas pemrosesan bahasa alami bersifat tidak seimbang secara alamiah, karena sebagian kategori jauh lebih sering muncul dibanding kategori lain, dan pada kondisi tersebut model cenderung berkinerja buruk pada kelas minoritas (Henning et al., 2022). Ketidakseimbangan pada penelitian ini terjadi pada dua tingkat sekaligus. Pada tingkat kelas utama, kelas terbesar memuat 34,54% data sedangkan kelas terkecil hanya 0,27%, sehingga rasio antara kelas terbesar dan terkecil mencapai 128,5 banding 1, dan dua kelas teratas saja sudah menguasai 64,11% data. Pada tingkat *subclass*, dari 64 kategori yang tersedia hanya 45 kategori yang memiliki sedikitnya satu contoh, sedangkan 19 kategori lainnya tidak terwakili sama sekali. Rasio antara *subclass* terbesar dan *subclass* terkecil mencapai 100 banding 1.

Selain jumlah, struktur taksonomi turut memperberat persoalan. Sejumlah topik yang sama muncul pada lebih dari satu kelas dan hanya dibedakan oleh maksud penutur, misalnya persoalan pembayaran yang dapat masuk ke kelas *Feature Error* bila dilaporkan sebagai kegagalan sistem atau ke kelas *Feature Design* bila disampaikan sebagai usulan perbaikan. Konsekuensinya, pembeda antarkelas bukan topik melainkan intensi, sehingga kategori-kategori tersebut sulit dipisahkan baik oleh anotator manusia maupun oleh model. Kondisi inilah yang menjelaskan rendahnya kesepakatan antarpakar pada tingkat *subclass*.

Tiga strategi digunakan untuk menanganinya. Pertama, pembobotan kelas pada fungsi kerugian, yaitu pemberian bobot yang berbanding terbalik dengan frekuensi kemunculan setiap kelas, sehingga kesalahan pada kelas minoritas memberi penalti lebih besar. Untuk tugas NER, (Nemoto et al., 2025) menunjukkan bahwa penanganan ketidakseimbangan pada tugas pelabelan token perlu memperhatikan dominasi label O yang jumlahnya jauh melampaui seluruh label entitas, dan bahwa perlakuan khusus

terhadap label mayoritas tersebut dapat meningkatkan kinerja kelas minoritas tanpa mengorbankan kelas mayoritas. Kedua, pemilihan metrik evaluasi yang tepat. Macro-F1 menghitung rata-rata F1 setiap kelas dengan bobot yang sama sehingga tidak tertutupi oleh dominasi kelas mayoritas, berbeda dengan *accuracy* maupun weighted-F1. (Opitz, 2024) mengingatkan bahwa istilah macro-F1 sendiri memiliki lebih dari satu definisi yang dapat menghasilkan selisih nilai cukup besar, sehingga rumus yang digunakan perlu dinyatakan secara eksplisit. Penelitian ini melaporkan micro-F1 dan macro-F1 secara berdampingan, karena selisih keduanya justru merupakan informasi mengenai perilaku model pada kategori berekor panjang.

Ketiga, penyesuaian skema validasi. Karena jumlah data pada beberapa kategori sangat terbatas, evaluasi dengan satu kali pembagian data akan menyisakan data uji yang terlalu sedikit untuk ditafsirkan, sehingga digunakan *stratified k-fold cross validation* yang mempertahankan proporsi kelas pada setiap lipatan dan melaporkan kinerja sebagai rata-rata beserta simpangan bakunya. stratifikasi hanya dapat dijalankan bila setiap kelas memiliki anggota tidak kurang dari jumlah lipatan. Kategori yang jumlah anggotanya di bawah ambang tersebut karena itu tidak dievaluasi sebagai kelas terpisah, melainkan digabungkan ke kategori induknya atau dilaporkan secara deskriptif tanpa nilai F1, dan keputusan penggabungan ini dinyatakan secara terbuka agar hasil dapat direplikasi.

2.2.10 Unified Modelling Language (UML)

Unified Modeling Language (UML) merupakan bahasa pemodelan standar yang digunakan untuk memvisualisasikan, menspesifikasikan, membangun, dan mendokumentasikan artefak sistem perangkat lunak. UML menyediakan sejumlah diagram yang merepresentasikan sudut pandang berbeda terhadap sistem, sehingga kebutuhan fungsional, struktur, interaksi, alur aktivitas, dan perubahan keadaan dapat dianalisis secara lebih terarah. Dalam model-driven software engineering, keberadaan beberapa diagram membantu pengembang menjaga keterlacakan antara requirement dan desain, tetapi juga menuntut konsistensi antarmodel karena perubahan pada satu diagram dapat memengaruhi diagram lainnya (Torre et al., 2022).

Dalam pengembangan sistem informasi, penggunaan UML membantu menyelaraskan kebutuhan pengguna dengan rancangan sistem. Diagram UML dapat digunakan sebagai media komunikasi antara peneliti, pengembang, dan pemangku kepentingan karena setiap elemen memiliki notasi dan makna yang terdefinisi. Pada penelitian yang menghasilkan dashboard, UML relevan untuk menjelaskan fungsi yang dapat diakses oleh pengguna sebelum implementasi dilakukan.

2.2.10.1 *Use Case Diagram*

Dalam penelitian ini, Use Case Diagram ditempatkan sebagai bagian dari pemodelan UML untuk menggambarkan kebutuhan fungsional dashboard dari perspektif pengguna. Diagram ini menunjukkan aktor yang berinteraksi dengan sistem, fungsi atau layanan yang dapat digunakan, batas sistem, serta hubungan antara aktor dan fungsi tersebut. Use Case Diagram berfokus pada apa yang disediakan sistem dan siapa yang menggunakan fungsi tersebut, bukan pada detail algoritme atau mekanisme internal.

(Alyami et al., 2023) menunjukkan bahwa Use Case Diagram dapat menjadi titik awal yang menjaga keterlacakan dari requirement menuju desain perilaku yang lebih rinci. Elemen pada Use Case Diagram dapat disejajarkan dengan elemen Sequence Diagram melalui concept alignment, sehingga kebutuhan fungsional yang didefinisikan pada tahap analisis dapat diterjemahkan secara konsisten ke dalam urutan interaksi sistem. Pendekatan tersebut memperkuat peran Use Case Diagram sebagai representasi awal kebutuhan sistem yang tetap terhubung dengan artefak desain berikutnya.

Tabel 2. 11 Elemen utama pada Use Case Diagram

Elemen	Representasi	Fungsi
Actor	Figur pengguna atau sistem eksternal	Menunjukkan pihak yang berinteraksi dengan sistem.
Use Case	Oval	Merepresentasikan fungsi atau layanan yang disediakan sistem.

Association	Garis penghubung	Menunjukkan keterlibatan aktor terhadap suatu use case.
System Boundary	Kotak batas sistem	Membedakan fungsi internal sistem dari aktor atau lingkungan eksternal.
Include / Extend	Relasi antarpenggunaan	Menunjukkan penggunaan kembali perilaku atau perluasan skenario tertentu.

Pada dashboard pengaduan PLN Mobile, Use Case Diagram digunakan untuk memetakan interaksi pengguna terhadap fungsi utama sistem, seperti melihat ringkasan data, hasil preprocessing, karakteristik corpus domain-spesifik, hasil Named Entity Recognition, analisis sentimen, validasi pakar, dan visualisasi informasi pengaduan. Dengan demikian, UML dan Use Case Diagram diposisikan sebagai satu kesatuan landasan perancangan yang menghubungkan kebutuhan pengguna dengan struktur fungsi dashboard.

2.2.11 Dashboard

Dashboard merupakan komponen dalam sistem informasi yang digunakan untuk menyajikan data dan informasi dalam bentuk visual sehingga pengguna dapat memahami kondisi tertentu secara cepat dan efektif. Dashboard menggabungkan elemen visual seperti grafik, tabel, indikator, dan ringkasan statistik untuk membantu proses pemantauan dan pengambilan keputusan (Conrow et al., 2023). Menurut (Bach et al., 2023), dashboard adalah tampilan visual dari informasi penting yang diperlukan untuk mencapai suatu tujuan tertentu dengan menggabungkan berbagai sumber data ke dalam satu layar sehingga informasi dapat dipahami secara cepat. Konsep utama dashboard terdiri dari beberapa karakteristik:

- 1) Visualisasi informasi

Dashboard mengubah data mentah menjadi bentuk visual seperti diagram batang, diagram lingkaran, grafik tren, maupun tabel sehingga pola tertentu lebih mudah ditemukan.

2) Monitoring

Dashboard memungkinkan pengguna melakukan pemantauan terhadap perubahan data berdasarkan periode tertentu.

3) Pengambilan Keputusan

Informasi yang ditampilkan dashboard dapat membantu pengguna dalam menentukan prioritas tindakan berdasarkan kondisi yang ditemukan.

Pada penelitian ini, dashboard dikembangkan sebagai bagian dari implementasi sistem informasi analisis pengaduan PLN Mobile. Dashboard memanfaatkan hasil pengolahan corpus domain-spesifik dan model IndoBERT untuk menyajikan informasi mengenai pola pengaduan pengguna secara visual.

2.2.12 Pseudo-Labeling

Pseudo-labeling adalah teknik pembelajaran semi-terawasi yang memanfaatkan model terlatih untuk memberikan label pada data yang belum berlabel, kemudian menerima label tersebut sebagai tambahan data latih apabila keyakinan prediksinya melampaui ambang tertentu. Teknik ini relevan pada penelitian ini karena korpus yang dibangun berjumlah puluhan ribu ulasan.

Persoalan pertama terletak pada pemilihan ambang keyakinan. Ambang yang terlalu rendah memasukkan label keliru ke dalam data latih dan kesalahan tersebut dapat terakumulasi, sedangkan ambang yang terlalu tinggi menyisakan terlalu sedikit data untuk dimanfaatkan. (Mekala et al., 2022) menunjukkan bahwa penyaringan berdasarkan skor probabilitas semata kurang memadai karena jaringan saraf dalam cenderung menghasilkan keyakinan yang terlalu tinggi. (Yang et al., 2024) mengatasi hal ini dengan memasukkan estimasi ketidakpastian ke dalam proses seleksi, dan melaporkan bahwa

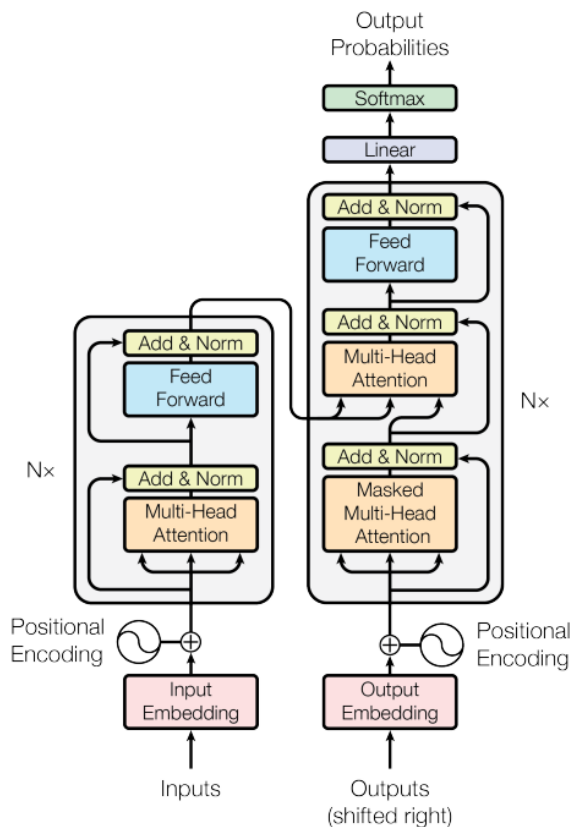
penyaringan berbasis ketidakpastian menghasilkan NER semi-terawasi yang lebih andal daripada penyaringan berbasis ambang probabilitas.

Persoalan kedua bersifat khusus pada rancangan penelitian ini dan perlu dinyatakan secara terbuka. Model dilatih pada label yang dihasilkan oleh aturan, sehingga apa yang dipelajari model adalah keteraturan aturan tersebut, bukan keragaman bahasa pada domain. Konsekuensinya, nilai F1 yang tinggi terhadap data uji berlabel lemah tidak dapat ditafsirkan sebagai kualitas pelabelan, melainkan sebagai ukuran seberapa setia model mereproduksi aturan. Bila *pseudo-label* kemudian dihasilkan oleh model yang sama dan dikembalikan sebagai data latih, kekeliruan sistematis aturan akan diperkuat alih-alih diperbaiki, karena kesalahan yang konsisten justru muncul dengan keyakinan tinggi dan lolos penyaringan ambang. Persoalan ini dikenal sebagai *confirmation bias* pada *self-training*.

Oleh sebab itu, penelitian ini menempatkan *gold standard* pakar, bukan data berlabel lemah, sebagai satu-satunya acuan penilaian kualitas pelabelan. Ambang keyakinan ditetapkan sebesar 0,90 sebagai titik awal dengan beberapa nilai alternatif yang dievaluasi, ulasan yang identik dengan data pakar dikecualikan agar tidak terjadi kebocoran data, dan setiap penambahan *pseudo-label* dinilai ulang terhadap *gold standard* untuk memastikan bahwa penambahan data tidak justru menurunkan kesesuaian terhadap anotasi manusia.

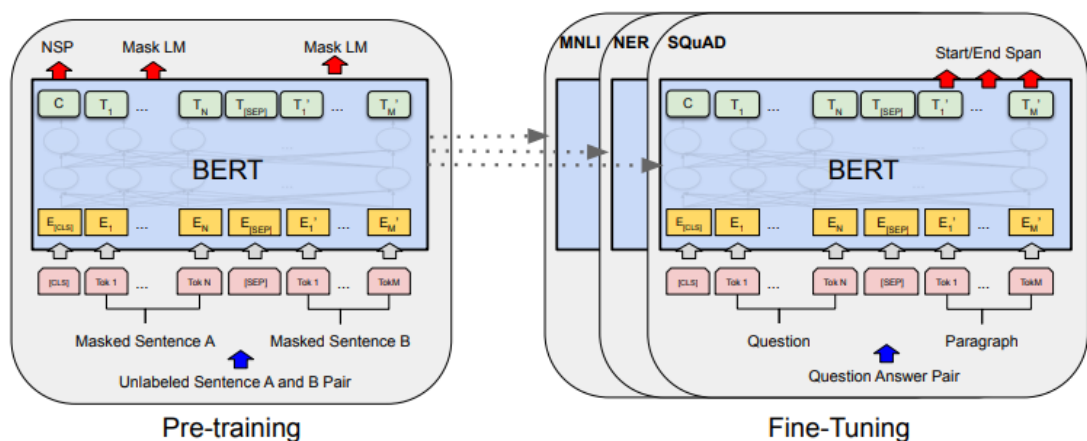
2.2.13 Bidirectional Encoder Representations from Transformers (BERT)

BERT (*Bidirectional Encoder Representations from Transformers*) adalah model bahasa pralatih yang diperkenalkan oleh (Devlin et al., 2019) dan dibangun di atas arsitektur Transformer. Berbeda dengan Transformer asli yang terdiri atas tumpukan *encoder* dan tumpukan *decoder*, BERT hanya menggunakan tumpukan *encoder*. Pilihan ini yang memungkinkan keunggulan utamanya, yaitu representasi kata yang dibangun secara dua arah, sehingga makna sebuah token dipahami berdasarkan konteks kiri dan kanan secara bersamaan. Tumpukan *decoder* tidak digunakan karena sifatnya kausal, yakni hanya boleh mengakses token sebelumnya, sehingga tidak sesuai dengan tujuan representasi dua arah.



Gambar 2. 2 Arsitektur Transformer (Devlin et al., 2019)

Pada Gambar 2.2 menunjukkan arsitektur Transformer secara umum yang menjadi dasar pengembangan BERT. Berbeda dengan Transformer asli yang memiliki komponen encoder dan decoder, BERT hanya memanfaatkan tumpukan Transformer encoder untuk membentuk representasi kontekstual setiap token. Setiap lapisan encoder terdiri atas mekanisme multi-head self-attention dan feed-forward network yang memungkinkan model mempelajari hubungan antartoken berdasarkan konteks di sebelah kiri maupun kanan. Encoder digunakan untuk membaca input data yang diberikan, sedangkan decoder digunakan untuk memprediksi atau menghasilkan output dari data tersebut. Setiap layer encoder memiliki dua sub-layer. Sub-layer pertama merupakan multi-head self-attention mechanism, yang memungkinkan model memperhatikan konteks kata lain dalam kalimat secara simultan, sehingga memahami hubungan antar token. Sub-layer kedua adalah fully-connected feed-forward network, yaitu jaringan saraf tiruan di mana koneksi antar node tidak membentuk siklus. Output dari self-attention layer digunakan sebagai input pada feed-forward network. Dengan mekanisme ini, BERT dapat menangkap konteks dua arah, yaitu dari sisi kiri dan kanan token dalam kalimat. Hal ini membuat BERT efektif dalam memahami nuansa bahasa, mengenali pola domain-spesifik, dan mendukung berbagai tugas NLP seperti klasifikasi teks, pelabelan entitas, dan analisis sentiment (Gardazi et al., 2025).



Gambar 2. 3 Pre-training dan Fine-Tuning BERT (Devlin et al., 2019)

Pada Gambar 2.3 ditampilkan tahapan kerja model BERT yang terdiri dari pre-training dan fine-tuning. Pre-training dilakukan pada data teks besar yang belum diberi label melalui dua tugas utama, yaitu Masked Language Model (MLM) yang menyembunyikan sebagian token untuk diprediksi sehingga model memahami konteks kata dari dua arah, dan Next Sentence Prediction (NSP) yang melatih model menentukan hubungan logis antar kalimat. Setelah itu, fine-tuning dilakukan dengan menyesuaikan model BERT terhadap tugas spesifik seperti Named Entity Recognition, Question Answering, atau klasifikasi teks, dengan menambahkan lapisan output sesuai kebutuhan dan melatih menggunakan data berlabel. Dalam konteks penelitian ini, konsep pre-training dan fine-tuning diterapkan pada IndoBERT, yang telah dilatih pada bahasa Indonesia, kemudian di-fine-tune menggunakan corpus domain-spesifik keluhan pelanggan PLN Mobile agar model mampu mengenali kategori keluhan seperti transaksi, layanan, verifikasi, fitur aplikasi, atau gangguan penggunaan aplikasi dengan akurasi yang tinggi.

Secara konseptual, BERT berbeda dari representasi kata statis karena makna kata yang sama bisa berubah tergantung konteks kalimat. Hal ini sangat penting untuk teks keluhan pelanggan, sebab istilah seperti “gagal”, “masuk”, “verifikasi”, atau “tagihan” bisa memiliki arti yang lebih spesifik jika dibaca bersama konteks ulasannya (Devlin et al., 2019).

2.2.13.1 *IndoBERT*

IndoBERT adalah model bahasa pra-latih untuk bahasa Indonesia yang diperkenalkan dalam kerangka IndoLEM. (Koto et al., 2020) mengembangkan IndoBERT untuk menjawab keterbatasan sumber daya NLP berbahasa Indonesia, seperti kurangnya dataset terannotasi, kelangkaan resource, dan kurangnya standardisasi. Hasil eksperimen mereka menunjukkan bahwa IndoBERT mencapai performa *state-of-the-art* pada sebagian besar tugas di benchmark IndoLEM.

IndoBERT relevan karena dua alasan. Pertama, data penelitian berupa ulasan pelanggan berbahasa Indonesia, sehingga model pra-latih yang memang dibangun untuk bahasa Indonesia lebih sesuai dibanding model multilingual umum. Kedua, IndoBERT dapat di-*fine-tune* untuk tugas downstream seperti sequence labeling atau klasifikasi domain (Dwi Purnomo & Sutopo, 2024). Walaupun demikian, penelitian sebelumnya juga menunjukkan bahwa kinerja model bahasa Indonesia dapat ditingkatkan lagi melalui penyesuaian kosakata dan data domain-spesifik, seperti pada IndoBERTtweet, sehingga keberadaan corpus domain-spesifik tetap menjadi komponen penting dalam penelitian ini (Koto et al., 2021).

2.3 Metrics Evaluation

Metrics evaluation merupakan seperangkat ukuran kuantitatif yang digunakan untuk menilai kesesuaian prediksi model terhadap label acuan. Pemilihan metrik perlu disesuaikan dengan jenis tugas dan distribusi kelas. Pada Named Entity Recognition (NER), evaluasi tidak cukup hanya menggunakan accuracy karena label non-entitas atau O dapat mendominasi token. Oleh karena itu, precision, recall, F1-score, micro-F1, macro-F1, dan evaluasi berbasis rentang entitas menjadi lebih informatif (Goel et al., 2026).

2.3.1.1 Confusion Matrix

Confusion matrix merupakan tabel yang memperlihatkan hubungan antara kelas aktual dan kelas hasil prediksi. Untuk penjelasan biner, confusion matrix terdiri atas True Positive (TP), False Positive (FP), False Negative (FN), dan True Negative (TN). Keempat komponen tersebut menjadi dasar perhitungan accuracy, precision, recall, dan F1-score (Park et al., 2023).

Tabel 2. 12 Struktur Dasar *Confusion Matrix*

	Prediksi Positif	Prediksi Negatif
Aktual Positif	<i>True Positive (TP)</i>	<i>False Negative (FN)</i>
Aktual Negatif	<i>False Positive (FP)</i>	<i>True Negative (TN)</i>

Keterangan:

- 1) *True Positive (TP)*: Kasus positif yang diprediksi positif oleh model.
- 2) *True Negative (TN)*: Kasus negatif yang diprediksi negatif.
- 3) *False Positive (FP)*: Kasus negatif yang diprediksi sebagai positif (*Type I Error*).
- 4) *False Negative (FN)*: Kasus positif yang diprediksi sebagai negatif (*Type II Error*).

Berdasarkan *Confusion Matrix*, sejumlah metrik evaluasi penting dapat dihitung untuk menilai kinerja model secara lebih komprehensif, khususnya dalam konteks klasifikasi multikelas seperti *Named Entity Recognition (NER)*. Metrik-metrik tersebut antara lain:

- 1) Akurasi (*Accuracy*):

Mengukur proporsi prediksi yang benar terhadap seluruh jumlah data yang diuji.

Rumus:

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (2.2)$$

- 2) Presisi (*Precision*):

Mengukur sejauh mana prediksi positif yang dihasilkan oleh model benar-benar relevan.

Rumus:

$$Presisi = \frac{TP}{TP + FP} \times 100\% \quad (2.3)$$

- 3) *Recall (Sensitivity atau True Positive Rate)*:

Mengukur seberapa baik model dalam menangkap semua data positif yang ada.

Rumus:

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (2.4)$$

4) *F1-Score*:

Merupakan rata-rata harmonis antara *precision* dan *recall*. Metrik ini sangat berguna ketika dibutuhkan keseimbangan antara kemampuan model dalam mengenali entitas (*recall*) dan akurasi prediksinya (*precision*).

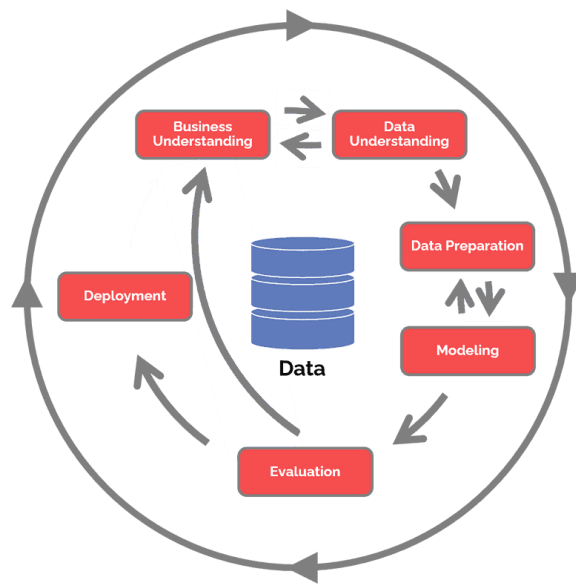
Rumus:

$$f1 - Score = \frac{2 \times Precision \times Sensitivity}{Precision + Sensitivity} \quad (2.5)$$

Keempat metrik ini akan digunakan secara bersamaan untuk mengevaluasi performa model dalam mengenali entitas dan sub-class dalam korpus domain spesifik.

2.4 Cross Industry Standard Process for Data Mining (CRISP-DM)

Cross Industry Standard Process for Data Mining (CRISP-DM) merupakan metodologi standar yang digunakan untuk mengembangkan proyek berbasis data mining secara sistematis. Metode ini menyediakan kerangka kerja yang terstruktur mulai dari pemahaman masalah bisnis hingga implementasi hasil analisis dalam bentuk solusi yang dapat digunakan. CRISP-DM banyak digunakan dalam penelitian berbasis data karena memiliki tahapan yang fleksibel dan dapat diterapkan pada berbagai bidang, termasuk analisis teks, machine learning, dan pengembangan sistem informasi. Metode ini membantu peneliti memastikan bahwa proses pengolahan data tidak hanya berfokus pada pembangunan model, tetapi juga mempertimbangkan kebutuhan pengguna, karakteristik data, evaluasi hasil, serta penerapan solusi (Schröer et al., 2021).



Gambar 2. 4 Tahapan CRISP-DM
(Kannengiesser & Gero, 2023)

Gambar 2.4 menunjukkan tahapan-tahapan dalam proses penambangan data. Panah-panah tersebut menunjukkan ketergantungan yang paling penting dan sering terjadi di antara tahapan-tahapan tersebut, sedangkan lingkaran luar melambangkan sifat siklikal dari penambangan data itu sendiri dan menggambarkan bahwa pelajaran yang dipetik selama proses penambangan data dan dari solusi yang diterapkan dapat memicu pertanyaan-pertanyaan bisnis baru yang seringkali lebih terfokus.

1. *Business Understanding*

Mungkin fase terpenting dalam setiap proyek data mining, fase pemahaman bisnis awal berfokus pada pemahaman tujuan proyek dari sudut pandang bisnis, mengubah pengetahuan ini menjadi definisi masalah data mining, dan kemudian menyusun rencana awal yang dirancang untuk mencapai tujuan tersebut. Agar dapat memahami data mana yang nantinya harus dianalisis, dan bagaimana caranya, sangatlah penting bagi para praktisi penambangan data untuk sepenuhnya memahami bisnis yang menjadi sasaran solusi yang mereka cari. Fase pemahaman bisnis melibatkan beberapa langkah utama, termasuk

menentukan tujuan bisnis, menilai situasi, menetapkan sasaran penambangan data, dan menyusun rencana proyek.

2. *Data Understanding*

Fase pemahaman data dimulai dengan pengumpulan data awal. Analisis kemudian dilanjutkan untuk memperdalam pemahaman terhadap data, guna mengidentifikasi masalah kualitas data, menemukan wawasan awal mengenai data tersebut, atau mendeteksi subset yang menarik untuk membentuk hipotesis tentang informasi tersembunyi. Fase pemahaman data mencakup empat langkah, yaitu pengumpulan data awal, deskripsi data, eksplorasi data, dan verifikasi kualitas data.

3. *Data Preparation*

Fase persiapan data mencakup semua kegiatan untuk menyusun kumpulan data akhir atau data yang akan dimasukkan ke dalam alat pemodelan dari data mentah awal. Tugas-tugas tersebut meliputi pemilihan tabel, catatan, dan atribut, serta transformasi dan pembersihan data untuk alat pemodelan. Lima langkah dalam persiapan data adalah pemilihan data, pembersihan data, penyusunan data, integrasi data, dan pemformatan data.

4. *Modeling*

Pada fase ini, berbagai teknik pemodelan dipilih dan diterapkan, serta parameternya dikalibrasi ke nilai-nilai optimal. Biasanya, terdapat beberapa teknik untuk jenis masalah penambangan data yang sama. Beberapa teknik memiliki persyaratan khusus terkait bentuk data. Oleh karena itu, mungkin perlu kembali ke fase persiapan data. Langkah-langkah pemodelan meliputi pemilihan teknik pemodelan, penyusunan rancangan pengujian, pembuatan model, dan penilaian model.

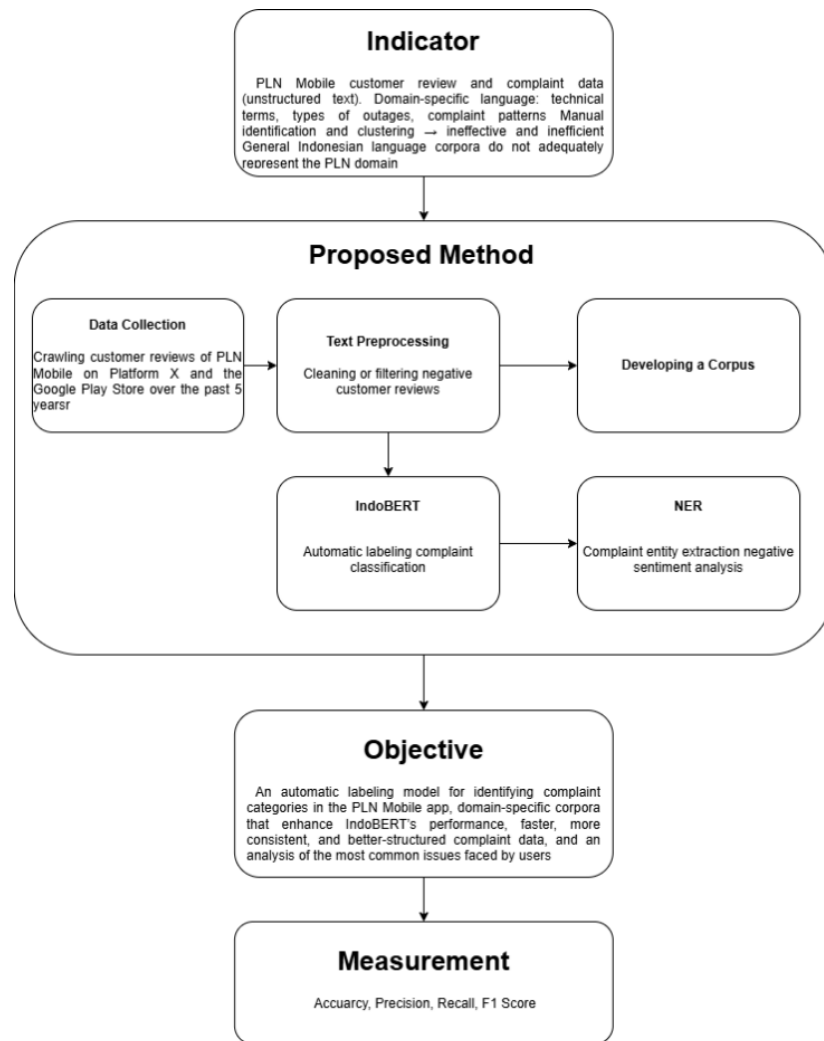
5. *Deployment*

Pembuatan model umumnya bukanlah akhir dari proyek. Pengetahuan yang diperoleh harus diorganisir dan disajikan sedemikian rupa sehingga pelanggan dapat memanfaatkannya, yang sering kali melibatkan penerapan model “aktif” dalam proses pengambilan

keputusan suatu organisasi, seperti personalisasi halaman web secara real-time atau penilaian berulang terhadap basis data pemasaran. Tergantung pada persyaratannya, fase penerapan bisa sesederhana membuat laporan atau serumit mengimplementasikan proses penambangan data yang dapat diulang di seluruh perusahaan. Meskipun seringkali pelanggan, bukan analis data, yang melaksanakan langkah-langkah penerapan tersebut, penting bagi pelanggan untuk memahami sejak awal tindakan apa saja yang harus diambil agar benar-benar dapat memanfaatkan model yang telah dibuat. Langkah-langkah utama di sini adalah perencanaan penerapan, perencanaan pemantauan dan pemeliharaan, pembuatan laporan akhir, serta tinjauan proyek.

2.5 Kerangka Pemikiran

Kerangka pemikiran yang disusun untuk menggambarkan langkah-langkah proses analisis data dari awal hingga akhir, dengan tujuan untuk mendapatkan kesimpulan dari data yang sedang dianalisis. Terdapat empat komponen yang menyusun kerangka pemikiran penelitian yaitu Indikator (Indicators), Metode yang diusulkan (Proposed Method), Tujuan (Objective), dan Pengukuran (Measurement). Pada penelitian ini, langkah yang diimplementasikan ke dalam sistem digambarkan melalui perancangan sistem sebagai berikut :



Gambar 2. 5 Kerangka Pemikiran

Pada gambar 2. 5, Berikut ini disajikan kerangka pemikiran untuk penelitian ini. Setiap aliran menunjukkan hubungan yang menarik antara komponen-komponen yang berbeda. Selanjutnya, penjelasan untuk masing-masing komponen akan diberikan di bawah ini.

1. Indikator

Dalam penelitian ini, indikator yang digunakan berasal dari data ulasan dan keluhan pelanggan pada aplikasi PLN Mobile. Data tersebut menjadi dasar untuk mengidentifikasi permasalahan utama, yaitu banyaknya informasi keluhan pelanggan

yang masih berbentuk teks tidak terstruktur, sehingga proses identifikasi dan pengelompokan keluhan secara manual menjadi kurang efektif dan efisien. Selain itu, penggunaan data teks yang berasal dari domain layanan kelistrikan juga memiliki karakteristik bahasa yang khas, seperti istilah teknis, jenis gangguan, dan bentuk keluhan tertentu yang tidak selalu terwakili dengan baik pada corpus umum bahasa Indonesia.

2. Proposed method

Berdasarkan permasalahan tersebut, metode yang diusulkan dalam penelitian ini adalah pembangunan corpus domain-spesifik yang merepresentasikan karakteristik bahasa pada keluhan pelanggan PLN Mobile, kemudian dilanjutkan dengan implementasi IndoBERT untuk melakukan pelabelan otomatis pada data keluhan. Corpus domain-spesifik dikembangkan agar model memiliki pemahaman yang lebih relevan terhadap istilah, konteks, dan pola bahasa yang muncul pada ulasan pelanggan PLN Mobile. Selanjutnya, IndoBERT digunakan sebagai model utama karena memiliki kemampuan memahami konteks bahasa Indonesia secara lebih baik melalui pendekatan *pre-trained language model*.

3. Objectives

Tujuan dari penelitian ini adalah menghasilkan model pelabelan otomatis yang mampu mengidentifikasi kategori keluhan pelanggan PLN Mobile secara lebih akurat. Dengan adanya sistem pelabelan otomatis, data keluhan pelanggan dapat diolah secara lebih cepat, konsisten, dan terstruktur, sehingga dapat membantu proses analisis terhadap jenis permasalahan yang paling sering dialami pengguna aplikasi PLN Mobile.

4. Measurement

Adapun pengukuran dalam penelitian ini dilakukan untuk menilai kinerja model IndoBERT dalam pelabelan otomatis keluhan pelanggan PLN Mobile. Pengukuran dilakukan menggunakan metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil pengukuran tersebut kemudian digunakan untuk mengetahui sejauh mana corpus domain-spesifik yang dikembangkan mampu mendukung performa IndoBERT dalam menghasilkan label keluhan secara otomatis,