BAB II

LANDASAN TEORI

2.1 Penelitian Terdahulu

Pada tahapan awal penelitian, dilakukan pencarian penelitian terlebih dahulu untuk memperdalam pengetahuan untuk mengatasi analisis sentimen berbahasa Indonesia. Hal ini dilakukan untuk mendapatkan penelitian yang memiliki fokus serupa dan dapat dijadikan sebagai acuan pada penelitian ini.

Tabel 2.1 Tabel Penelitian Terdahulu

Judul	Tools / Method	Hasil
Performance Comparison Of BERT Metrics and Classical Machine Learning Models (SVM, Naive Bayes) for Sentiment Analysis – Majid & Nuari (2024) [18].	Machine Learning, TF-IDF	Model klasifikasi berbasis machine learning dapat digunakan secara efektif untuk analisis sentimen pada ulasan produk.
Sentiment Analysis of Cyber Attacks in Bank Syariah Indonesia Using SVM and Indobert Method – Apriyadi & Styawati (2025) [19].	SVM, IndoBERT	IndoBERT mengungguli SVM dengan akurasi 85% dan F1-score 82%.
Sentiment Analysis of COVID-19 Public Activity Restriction (PPKM) Impact using BERT Method – Fransiscus and Girsang (2022) (2021) [20].	IndoBERT	Penelitian menunjukkan bahwa model IndoBERT mampu menghasilkan performa klasifikasi yang lebih baik dibandingkan Support Vector Machine (SVM) dan Multinomial Naïve Bayes pada data Twitter berbahasa Indonesia. Keunggulan ini diperoleh karena IndoBERT dapat memahami konteks kalimat secara dua arah (bidirectional context), sehingga lebih efektif dalam menangkap makna semantik dan pola sentimen pada teks yang bersifat informal. Model memperoleh nilai precision, recall, dan F1-score sekitar 84%, serta menunjukkan performa yang stabil pada seluruh kelas sentimen.
Komparasi Metode Naive Bayes dan SVM pada Sentimen Twitter	Naïve Bayes, SVM	Penelitian menunjukkan bahwa Support Vector

Judul	Tools / Method	Hasil
Mengenai Persoalan Perpu Cipta Kerja – Farhan & Setiaji (2023) [21].		Machine memiliki performa klasifikasi yang lebih tinggi dibandingkan Naïve Bayes karena kemampuannya dalam memisahkan data berdimensi tinggi secara optimal. SVM menghasilkan akurasi yang lebih baik dan nilai F1-score yang lebih stabil pada seluruh kelas sentimen.
Sentiment Analysis of ICT Service User Using Naïve Bayes Classifier and Support Vector Machine Methods with TF-IDF Text Weighting - Trisnawati and Wibowo (2024) [22].	Naïve Bayes, Support Vector Machine, TF-IDF	Penelitian menunjukkan bahwa kombinasi TF-IDF dengan Support Vector Machine menghasilkan performa klasifikasi yang lebih tinggi dibandingkan Naïve Bayes pada analisis sentimen layanan ICT. SVM lebih efektif dalam menangani data teks berdimensi tinggi dan menghasilkan nilai precision, recall, dan F1-score yang lebih stabil, sedangkan Naïve Bayes tetap menawarkan proses pelatihan yang lebih cepat dan implementasi yang sederhana.
Comparison of Pre-Trained BERT-Based Transformer Models for Regional Language Text Sentiment Analysis in Indonesia – Asri et al. (2024) [23].	BERT-based Transformer Models (IndoBERT)	Penelitian menunjukkan bahwa model BERT dan variannya mampu menghasilkan akurasi tinggi dalam analisis sentimen teks bahasa Indonesia, karena kemampuannya memahami konteks kalimat secara dua arah (bidirectional context).
Sentiment Analysis on Shopee Product Reviews Using IndoBERT – Widyananda et al. (2024) [24].	IndoBERT	Penelitian menunjukkan bahwa penggunaan model IndoBERT mampu meningkatkan performa klasifikasi sentimen ulasan produk Shopee, karena model transformer mampu memahami konteks bahasa Indonesia dengan lebih baik dibanding metode tradisional.
Analisis Sentimen Terhadap Ulasan Penggunaan Shopee Melalui Tweet Pada Twitter Menggunakan Algoritma Naïve Bayes – Gondohanindijo (2023) [25].	Naïve Bayes	Algoritma Naïve Bayes mampu melakukan klasifikasi sentimen terhadap ulasan pengguna Shopee pada Twitter secara cukup baik, sehingga dapat digunakan sebagai metode dasar dalam

Judul	Tools / Method	Hasil
		analisis sentimen berbasis teks.
Perbandingan Kinerja Pre-Trained IndoBERT-Base dan IndoBERT-Lite pada Klasifikasi Sentimen Ulasan TikTok Tokopedia Seller Center – Hidayat & Nastiti (2024) [26].	IndoBERT	Hasil penelitian menunjukkan bahwa model Indobert-base-p2 memiliki performa terbaik dengan akurasi 97%, precision 97%, recall 97%, dan F1-score 97%, sedangkan Indobert-lite-base-p2 memperoleh akurasi 94%, precision 94%, recall 94%, dan F1-score 94%.
Sentiment Analysis of Tokopedia Customer Reviews Using BiLSTM and IndoBERT with Comparative Analysis of Preprocessing and Labeling Methods –Anadra, Wijayanto & Sadik (2025) [27].	BiLSTM, IndoBERT	Penelitian menunjukkan bahwa IndoBERT memiliki performa lebih tinggi dibandingkan BiLSTM dalam klasifikasi sentimen ulasan Tokopedia dengan balanced accuracy hingga 0.85 dan F1-score 0.83, sedangkan BiLSTM mencapai balanced accuracy 0.78 dan F1-score 0.76. Namun BiLSTM memiliki waktu pelatihan lebih cepat. Penggunaan class weight dan focal loss meningkatkan performa model sekitar 2%–11%.

Berdasarkan penelitian terdahulu yang telah dirangkum pada Tabel 2.1, berbagai pendekatan telah digunakan dalam analisis sentimen teks berbahasa Indonesia, khususnya pada ulasan produk dan opini pengguna di platform digital. Pendekatan machine learning klasik berbasis representasi fitur TF-IDF terbukti mampu melakukan klasifikasi sentimen secara efektif pada dataset ulasan produk berbahasa Indonesia [18]. Metode ini umumnya memiliki keunggulan dalam hal kesederhanaan model, efisiensi komputasi, serta kemudahan implementasi pada dataset berukuran menengah. Sejalan dengan hal tersebut, algoritma Naïve Bayes juga terbukti mampu melakukan klasifikasi sentimen terhadap ulasan pengguna pada platform digital secara cukup baik, sehingga dapat digunakan sebagai metode dasar dalam analisis sentimen berbasis teks [25].

Selain Naïve Bayes, algoritma Support Vector Machine (SVM) juga banyak digunakan dalam penelitian analisis sentimen bahasa Indonesia. Penelitian yang

membandingkan Naïve Bayes dan SVM pada data Twitter menunjukkan bahwa SVM memiliki performa klasifikasi yang lebih tinggi karena kemampuannya dalam memisahkan data berdimensi tinggi secara optimal, dengan akurasi dan F1-score yang lebih stabil pada seluruh kelas sentimen [21]. Hal ini menunjukkan bahwa SVM merupakan pilihan yang lebih unggul dibandingkan Naïve Bayes dalam skenario teks berdimensi tinggi, meskipun keduanya sama-sama mengandalkan representasi fitur berbasis statistik yang tidak mempertimbangkan konteks semantik antar kata.

Perkembangan terbaru dalam bidang Natural Language Processing memperkenalkan model berbasis transformer, khususnya BERT dan variannya, yang mampu memahami konteks kalimat secara dua arah (bidirectional context) dan menghasilkan akurasi tinggi dalam analisis sentimen teks bahasa Indonesia [23]. Dalam konteks bahasa Indonesia, model BERT yang di-fine-tuning menunjukkan performa yang lebih baik dalam analisis sentimen pada ulasan aplikasi mobile dibandingkan metode pembelajaran mesin konvensional [20]. Lebih spesifik, IndoBERT yang dilatih pada korpus bahasa Indonesia terbukti mampu meningkatkan performa klasifikasi sentimen ulasan produk e-commerce karena kemampuannya memahami konteks bahasa Indonesia dengan lebih baik dibanding metode tradisional [24].

Beberapa penelitian juga secara langsung membandingkan IndoBERT dengan metode machine learning konvensional. Pada domain keamanan siber, IndoBERT mengungguli SVM dengan akurasi 85% dan F1-score 82% [19]. Pada analisis sentimen kebijakan publik, IndoBERT juga menunjukkan performa lebih baik dibandingkan SVM dan Multinomial Naïve Bayes dengan nilai precision, recall, dan F1-score sekitar 84% [22]. Dalam konteks ulasan produk e-commerce Tokopedia secara spesifik, model IndoBERT-base bahkan mampu mencapai akurasi hingga 97% [26], dan terbukti memiliki performa lebih tinggi dibandingkan model deep learning lain seperti BiLSTM dengan balanced accuracy 0,85 dan F1-score 0,83, meskipun BiLSTM memiliki waktu pelatihan yang lebih cepat [27].

Berdasarkan kajian terhadap penelitian-penelitian terdahulu yang telah dipaparkan, dapat diidentifikasi beberapa celah penelitian yang belum terjawab secara komprehensif. Pertama, meskipun beberapa penelitian telah membandingkan IndoBERT dengan SVM [19], [22] atau IndoBERT dengan BiLSTM [27], belum ada penelitian yang secara serentak membandingkan ketiga pendekatan Naïve Bayes, SVM, dan IndoBERT dalam satu eksperimen yang terkontrol dengan dataset, preprocessing, dan metrik evaluasi yang seragam. Kedua, penelitian yang menggunakan domain ulasan produk *e-commerce* umumnya berfokus pada satu platform tertentu dengan kategori produk yang terbatas [24], [26], sementara penelitian ini menggunakan dataset multi-kategori dari Tokopedia yang mencakup elektronik, fashion, olahraga, handphone, dan pertukangan, sehingga lebih representatif terhadap variasi bahasa ulasan konsumen. Ketiga, penelitian terdahulu jarang membahas secara mendalam permasalahan klasifikasi pada kelas sentimen netral, padahal kelas ini secara konsisten menjadi tantangan terbesar dalam analisis sentimen tiga kelas berbahasa Indonesia. Oleh karena itu, penelitian ini hadir untuk mengisi celah tersebut melalui perbandingan empiris yang sistematis antara Naïve Bayes, SVM, dan IndoBERT, sekaligus menganalisis keterbatasan masing-masing model dalam menangani karakteristik bahasa informal dan ambiguitas sentimen pada ulasan produk e-commerce berbahasa Indonesia.

2.2 Teori yang berkaitan

2.2.1 Analisis Sentimen

Analisis sentimen merupakan cabang dari *Natural Language Processing* yang bertujuan untuk mengidentifikasi dan mengklasifikasikan opini atau sikap terhadap suatu objek atau topik tertentu. Hasil analisis sentimen umumnya dikelompokkan ke dalam beberapa kelas polaritas, seperti positif, negatif, dan netral [28]. Proses analisis sentimen melibatkan beberapa tahapan, mulai dari pengumpulan data teks, *preprocessing*, ekstraksi fitur, hingga proses klasifikasi menggunakan algoritma tertentu [29].

Dalam konteks ulasan produk, analisis sentimen berperan penting dalam memahami persepsi konsumen terhadap kualitas produk, layanan, dan

pengalaman pengguna. Informasi ini dapat dimanfaatkan untuk evaluasi produk, pengambilan keputusan bisnis, serta pengembangan strategi pemasaran berbasis data.

2.2.2 Machine Learning dan Deep Learning dalam NLP

Machine Learning dalam analisis sentimen mengacu pada penggunaan algoritma pembelajaran mesin untuk mempelajari pola dari data teks yang telah direpresentasikan dalam bentuk numerik. Metode klasik seperti *Naïve Bayes* dan *Support Vector Machine* mengandalkan fitur berbasis frekuensi kata, seperti *bag-of-words* dan TF-IDF, untuk melakukan klasifikasi sentimen [24].

Deep Learning merupakan pengembangan lanjutan dari Machine Learning yang menggunakan jaringan saraf dengan banyak lapisan untuk mempelajari representasi data secara otomatis. Dalam NLP, pendekatan Deep Learning memungkinkan model untuk memahami hubungan semantik dan konteks kata dalam kalimat [23]. Model berbasis *transformer* seperti IndoBERT mampu menghasilkan representasi kata yang kontekstual, sehingga lebih efektif dalam menangani ambiguitas bahasa dan variasi makna kata [22], [23].

2.2.3 *Support Vector Machine*

Support Vector Machine (SVM) adalah algoritma pembelajaran mesin yang bekerja dengan mencari hyperplane optimal yang memaksimalkan margin antara kelas-kelas dalam ruang fitur berdimensi tinggi. Dalam konteks analisis sentimen teks, SVM dikombinasikan dengan representasi TF-IDF untuk mengubah teks menjadi vektor numerik yang kemudian dipisahkan secara linear maupun non-linear menggunakan kernel tertentu [28].

Representasi TF-IDF untuk setiap kata dihitung menggunakan rumus:

$$TF\text{-}IDF(t,d) = TF(t,d) \times \log(N / DF(t))$$

Rumus 2. 1 Rumus Representasi TF-IDF

dengan TF(t,d) adalah frekuensi kemunculan term t dalam dokumen d, N adalah jumlah total dokumen, dan DF(t) adalah jumlah dokumen yang memuat term t. Selanjutnya, SVM mencari hyperplane pemisah optimal dengan memformulasikan permasalahan optimisasi sebagai berikut:

$$min\ (1/2)||w||^2\ ,\ \text{s.t.}\ y_i(w \cdot x_i + b) \geq 1$$

Rumus 2. 2 Rumus Optimasi SVM

dengan w adalah vektor bobot normal terhadap *hyperplane*, b adalah bias, x_i adalah vektor fitur TF-IDF dari dokumen ke-i, dan y_i adalah label kelas sentimennya. Pada penelitian ini digunakan parameter class_weight='balanced' untuk mengurangi bias model terhadap kelas mayoritas akibat distribusi data yang tidak seimbang.

SVM merupakan salah satu algoritma yang paling banyak digunakan dan dibandingkan dalam literatur analisis sentimen teks, sehingga hasil yang diperoleh dapat dikontekstualisasikan secara langsung dengan penelitian-penelitian sebelumnya [21], [22]. Kedua, SVM memiliki fondasi matematis yang kuat dan terbukti efektif pada data teks berdimensi tinggi kondisi yang lazim terjadi ketika fitur TF-IDF digunakan dengan kosakata yang besar [8]. Ketiga, dibandingkan model *ensemble* seperti Random Forest atau XGBoost yang menggabungkan banyak *decision tree*, SVM lebih transparan dalam proses pengambilan keputusannya dan lebih ringan secara komputasi untuk dataset skala menengah, sehingga lebih sesuai dengan sumber daya yang tersedia dalam penelitian ini.

Pemilihan SVM juga sejalan dengan tujuan penelitian yang berfokus pada perbandingan paradigma yaitu antara *Machine Learning* konvensional berbasis fitur statistik versus Deep Learning berbasis konteks semantik bukan perbandingan antar model dalam satu paradigma. Dalam kerangka ini, SVM berfungsi sebagai representasi terkuat dari pendekatan *Machine Learning*

konvensional berbasis margin, yang memberikan baseline yang adil dan bermakna untuk dibandingkan dengan IndoBERT.

2.2.4 Naïve Bayes

Naïve Bayes adalah algoritma klasifikasi berbasis teorema Bayes yang mengasumsikan independensi kondisional antar fitur. Meskipun asumsi ini jarang terpenuhi secara sempurna dalam data teks nyata, Naïve Bayes terbukti secara empiris menghasilkan performa yang kompetitif dalam berbagai tugas klasifikasi teks, termasuk analisis sentimen, karena kemampuannya mengestimasi probabilitas posterior secara efisien dari data pelatihan [29], [30].

$$P(c|x) = P(x|c)P(c) / P(x)$$

Rumus 2. 3 Rumus Teorema Bayes

dengan P(c|x) adalah probabilitas posterior kelas sentimen c (positive, neutral, atau negative) diberikan fitur kata x, P(x|c) adalah likelihood, P(c) adalah probabilitas prior kelas c, dan P(x) adalah probabilitas evidence dari fitur x. Dengan asumsi independensi kondisional, P(x|c) untuk seluruh fitur dihitung sebagai hasil kali probabilitas masing-masing kata dalam dokumen, dan kelas dengan nilai posterior tertinggi dipilih sebagai prediksi akhir model.

Naïve Bayes mampu merepresentasikan pendekatan probabilistik yang secara konseptual berbeda dari SVM keduanya bersama-sama mencakup dua mazhab utama dalam Machine Learning konvensional untuk klasifikasi teks, yakni pendekatan berbasis margin dan berbasis probabilitas. Perbandingan keduanya dengan IndoBERT memberikan gambaran yang lebih komprehensif tentang keterbatasan Machine Learning konvensional secara keseluruhan, bukan hanya satu pendekatan saja. Kemudian, Naïve Bayes sering digunakan sebagai *baseline* standar dalam penelitian NLP karena kesederhanaan dan kecepatan pelatihannya [31]. Dengan menyertakan Naïve Bayes, penelitian ini

memiliki titik referensi terbawah (*lower bound*) yang jelas, sehingga kontribusi peningkatan performa dari SVM dan IndoBERT dapat diukur dan diinterpretasikan secara bertahap dan sistematis.

Model seperti Random Forest dan XGBoost, meskipun umumnya menghasilkan performa yang lebih tinggi dari Naïve Bayes maupun SVM, memiliki kompleksitas model yang jauh lebih besar dan lebih sulit diinterpretasikan secara linguistik. Penggunaan model *ensemble* tersebut akan mengaburkan garis pembanding yang ingin ditarik penelitian ini, yaitu antara model yang mengandalkan statistik frekuensi kata semata versus model yang memahami konteks semantik secara mendalam. Oleh karena itu, Naïve Bayes dan SVM dipilih secara deliberatif sebagai representasi yang paling tepat untuk mewakili paradigma Machine Learning konvensional dalam kerangka perbandingan ini.

2.2.5 IndoBERT

IndoBERT merupakan model bahasa pra-latih (*pre-trained language model*) berbasis arsitektur BERT (Bidirectional Encoder Representations from Transformers) yang dilatih khusus pada korpus berskala besar berbahasa Indonesia, mencakup teks dari Wikipedia, berita daring, dan dokumen formal maupun informal lainnya. Sebagaimana model BERT pada umumnya, IndoBERT dibangun atas mekanisme self-attention pada arsitektur Transformer encoder, yang memungkinkan model memperhitungkan hubungan antara seluruh token dalam suatu kalimat secara simultan dari kedua arah (bidirectional context), berbeda dengan model sekuensial seperti RNN atau LSTM yang membaca teks secara satu arah [42]. Representasi vektor untuk setiap token dihasilkan melalui mekanisme attention berikut:

$$Attention(Q, K, V) = softmax(QK^T / \sqrt{d_k})\, V$$

Rumus 2. 4 Rumus *Scaled Dot-Product Attention*

dengan Q (query), K (key), dan V (value) merupakan hasil transformasi linear dari representasi embedding token, serta d_k adalah dimensi vektor key yang berfungsi sebagai faktor penskalaan agar nilai dot-product tidak terlalu besar. Dalam penelitian ini, model pra-latih yang digunakan adalah indobenchmark/indobert-base-p1, yang kemudian dilakukan proses fine-tuning pada dataset ulasan produk e-commerce berbahasa Indonesia. Proses fine-tuning menyesuaikan parameter model pra-latih agar optimal untuk tugas klasifikasi sentimen tiga kelas (positive, neutral, negative), dengan menambahkan satu lapisan klasifikasi (classification head) di atas representasi token [CLS] sebagai keluaran akhir. Keunggulan utama IndoBERT terletak pada representasi kontekstual yang dinamis, di mana makna suatu kata dapat berubah sesuai dengan kalimat tempatnya berada, sehingga model lebih mampu menangani negasi, ironi, dan ambiguitas makna yang umum ditemukan pada ulasan e-commerce berbahasa Indonesia yang bersifat informal [35], [42].

Secara keseluruhan, pemilihan Naïve Bayes, SVM, dan IndoBERT sebagai tiga model yang dibandingkan dalam penelitian ini didasarkan pada pertimbangan bahwa ketiganya secara kolektif merepresentasikan tiga tingkat kompleksitas pemodelan teks yang berbeda namun saling melengkapi: Naïve Bayes sebagai baseline probabilistik yang sederhana dan ringan secara komputasi; SVM sebagai model *Machine Learning* konvensional yang lebih kompleks dengan kemampuan membentuk batas keputusan non-linear melalui kernel; serta IndoBERT sebagai model Deep Learning berbasis transformer yang merepresentasikan state-of-the-art untuk pemahaman bahasa kontekstual. Ketiga model ini juga dipilih karena masing-masing telah banyak digunakan dan terbukti efektif dalam literatur analisis sentimen Bahasa Indonesia maupun bahasa lain, sehingga hasil perbandingan dalam penelitian ini dapat dikontekstualisasikan secara langsung dengan temuan-temuan sebelumnya, sekaligus memberikan gambaran progresif mengenai sejauh mana peningkatan

kompleksitas model berbanding lurus dengan peningkatan performa klasifikasi sentimen pada domain ulasan *e-commerce* berbahasa Indonesia.

2.3 Framework/Algoritma yang digunakan

CRISP-DM (*Cross-Industry Standard Process for Data Mining*) adalah framework yang sangat populer untuk pelaksanaan proses *data mining* [32]. *Framework* ini menawarkan pendekatan yang terstruktur dan fleksibel untuk mengelola proyek data mining, termasuk analisis sentimen. *Framework* ini terdiri dari enam tahap utama, yang dirancang untuk memastikan proyek berjalan dengan baik dan menghasilkan *output* yang sesuai dengan tujuan bisnis. Berikut adalah penjelasan tiap tahap berdasarkan referensi yang relevan:

1) *Business Understanding*

Tahap ini adalah langkah awal untuk memahami tujuan bisnis dan masalah yang ingin diselesaikan melalui data mining. Dalam konteks analisis sentimen, tujuan bisa berupa menganalisis opini publik terhadap suatu isu politik di media sosial, seperti Twitter. Dengan mendefinisikan tujuan secara jelas, langkah-langkah selanjutnya dapat disesuaikan untuk memenuhi kebutuhan tersebut.

2) *Data Understanding*

Pada tahap ini, data yang relevan dikumpulkan dan dianalisis untuk memahami pola awal. Data yang digunakan bisa berupa teks dari media sosial, seperti tweet, yang kemudian dijelajahi untuk mengenali karakteristik utama, pola distribusi, atau keberadaan *outlier.*

3) *Data Preparation*

Tahap ini melibatkan proses pembersihan data dan transformasi untuk membuat data siap diproses oleh algoritma. Dalam analisis sentimen, ini mencakup penghapusan *noise* (seperti simbol atau URL) dan representasi teks menjadi fitur numerik seperti *Bag of Words* atau TF-IDF. Data yang berkualitas baik pada tahap ini akan meningkatkan performa model.

4) Modeling

Pada tahap ini, algoritma pembelajaran mesin diterapkan. Contohnya adalah penerapan SVM atau *Naïve Bayes* pada data untuk membangun model analisis sentimen. Pemilihan parameter dilakukan secara hati-hati untuk menghasilkan model yang optimal.

5) *Evaluation*

Model yang dihasilkan dievaluasi menggunakan metrik performa seperti akurasi, presisi, *recall*, dan *F1-score* untuk memastikan hasilnya sesuai dengan tujuan awal. Jika performa model tidak memadai, tahap ini memungkinkan untuk dilakukan iterasi ulang di tahap sebelumnya.

6) *Deployment*

Hasil akhir dari analisis disajikan dalam bentuk laporan atau sistem yang dapat digunakan untuk mendukung pengambilan keputusan. Dalam analisis sentimen, hasil ini dapat membantu memahami pola opini publik dan menghasilkan wawasan strategis

2.4 Tools & software

Dalam penelitian ini, berbagai perangkat lunak dan pustaka digunakan untuk mendukung seluruh tahapan proses analisis sentimen, mulai dari pengumpulan data, pengolahan awal (*preprocessing*), pemodelan, hingga evaluasi hasil. Pemilihan *tools* dan *software* dilakukan dengan mempertimbangkan stabilitas, fleksibilitas, serta dukungan komunitas yang luas, sehingga dapat menunjang replikasi dan pengembangan penelitian di masa mendatang [33].

2.4.1 Python

Python digunakan sebagai bahasa pemrograman utama dalam penelitian ini karena kemudahan sintaks, fleksibilitas, serta ketersediaan pustaka yang sangat kaya untuk keperluan data mining dan machine learning. Python memungkinkan integrasi yang efisien antara proses pengolahan data, pemodelan algoritma, dan evaluasi hasil dalam satu lingkungan kerja yang terintegrasi [34].

2.4.2 Hugging Face Transformers

Hugging Face Transformers digunakan untuk mengimplementasikan model Deep Learning berbasis *transformer*, khususnya IndoBERT. Pustaka ini menyediakan model pra-latih yang telah dioptimalkan untuk Bahasa Indonesia, sehingga memungkinkan proses *fine-tuning* dan inferensi model dilakukan secara efisien. Dengan memanfaatkan Transformers, penelitian ini dapat mengevaluasi kemampuan model IndoBERT dalam memahami konteks semantik teks ulasan produk secara lebih mendalam dibandingkan pendekatan Machine Learning klasik [35].