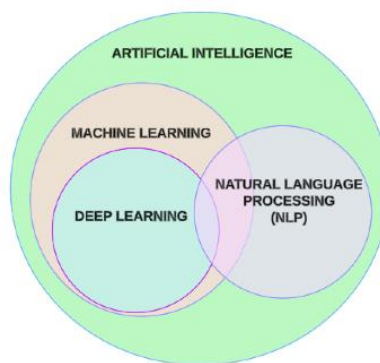


BAB II LANDASAN TEORI

2.1. *Artificial Intelligence (AI)*

Artificial Intelligence (AI) merupakan cabang ilmu komputer yang berfokus pada pengembangan sistem atau mesin yang mampu meniru kemampuan kognitif manusia, seperti belajar, menganalisis, mengambil keputusan, dan memecahkan masalah (Sardar et al., 2023). AI memanfaatkan algoritma dan model matematis untuk memproses data dalam jumlah besar sehingga dapat menemukan pola, membuat prediksi, dan menghasilkan keputusan secara otomatis. Teknologi AI mencakup berbagai bidang seperti *Machine Learning*, deep learning, computer vision, hingga *Natural Language Processing* yang memungkinkan komputer memahami bahasa manusia. Dengan kemampuannya yang terus berkembang, AI telah menjadi fondasi penting dalam berbagai aplikasi modern yang memerlukan analisis cerdas dan efisien.



Gambar 2. 1. Artificial Intelligence (AI)

AI banyak diterapkan dalam pengolahan data teks melalui algoritma pembelajaran mesin yang mampu mempelajari pola linguistik dari kumpulan data (Xu & Mishra, 2022). Algoritma seperti *Naïve Bayes*, *Support Vector Machine*, dan

jaringan saraf digunakan untuk memahami karakteristik teks, mengidentifikasi sentimen, serta mengelompokkan informasi ke dalam kategori tertentu. Penerapan AI dalam analisis teks membantu meningkatkan ketepatan dalam membaca opini atau kecenderungan pengguna melalui proses otomatis yang lebih cepat dibandingkan analisis manual. Dengan bantuan teknik pembobotan kata seperti *TF-IDF*, AI dapat bekerja lebih optimal dalam mengenali kata-kata penting yang memengaruhi sentimen dalam data ulasan.

Pemanfaatan AI dalam analisis data membawa efisiensi yang signifikan baik dalam proses pengolahan data maupun pengambilan keputusan yang berbasis informasi. Teknologi ini memungkinkan hasil analisis menjadi lebih akurat, sistematis, dan mudah dipahami melalui model-model yang terstruktur. Kemampuan AI untuk mengolah data dalam jumlah besar juga mendukung kejelasan interpretasi pola tertentu, sehingga hasil analisis yang dihasilkan dapat digunakan untuk memperoleh wawasan yang lebih mendalam. Integrasi AI dengan teknik pengolahan data teks menjadikan analisis yang dilakukan lebih relevan, cepat, dan dapat diterapkan untuk berbagai kebutuhan yang memerlukan pemahaman terhadap opini atau perilaku pengguna.

2.1.1. *Machine Learning*

Machine Learning adalah cabang dari kecerdasan buatan yang berfokus pada pengembangan algoritma dan model yang mampu belajar dari data untuk membuat prediksi atau mengambil keputusan secara otomatis (Lakhdari et al., 2023). Sistem dalam *Machine Learning* bekerja dengan mengenali pola yang terdapat pada data, kemudian menggunakannya untuk memproses informasi baru tanpa harus

diprogram secara eksplisit. Konsep ini mencakup berbagai teknik seperti klasifikasi, regresi, clustering, dan analisis data tekstual, yang menjadikannya salah satu pendekatan utama dalam pengolahan data modern.

Machine Learning banyak digunakan ketika proses analisis data memerlukan kemampuan komputasi untuk mengenali pola yang tidak mudah diamati secara manual (Assegie, 2021). Melalui algoritma tertentu, proses pengolahan data teks dapat dioptimalkan menjadi informasi yang lebih terstruktur sehingga hasil analisis dapat diperoleh dengan lebih akurat. Penerapan metode pembobotan kata dan model klasifikasi dalam teks menjadi salah satu contoh bagaimana *Machine Learning* bekerja dalam mengolah informasi yang berasal dari interaksi pengguna pada platform digital.

Dengan karakteristiknya yang fleksibel dan adaptif, *Machine Learning* mampu membantu menghasilkan pemahaman yang lebih mendalam terhadap data yang dianalisis (Esteban et al., 2022). Model-model yang digunakan dapat dilatih untuk mengenali kecenderungan sentimen, mengevaluasi hasil prediksi, dan meningkatkan performa klasifikasi melalui proses pembelajaran yang berkelanjutan. Kemampuannya untuk menangani data besar dan mendeteksi pola tersembunyi membuat *Machine Learning* menjadi pendekatan yang relevan dan efektif dalam berbagai penelitian berbasis data, termasuk analisis yang memerlukan tingkat ketepatan tinggi dalam memahami opini atau perilaku pengguna.

1. Model Klasifikasi

Model klasifikasi merupakan salah satu teknik utama dalam bidang *Data mining* dan pembelajaran mesin yang bertujuan untuk mengelompokkan data ke

dalam kategori atau label tertentu berdasarkan pola yang ditemukan dalam data tersebut (Gatto et al., 2023). Proses klasifikasi biasanya melibatkan dua tahap utama, yaitu pelatihan dan pengujian. Pada tahap pelatihan, model mempelajari hubungan antara fitur dan label berdasarkan dataset yang telah diberi kategori. Setelah itu, pada tahap pengujian, model digunakan untuk memprediksi label dari data baru yang belum pernah dilihat sebelumnya. Model klasifikasi dapat bekerja pada berbagai jenis data, termasuk data numerik, kategorikal, maupun teks, sehingga teknik ini banyak digunakan dalam berbagai bidang seperti kesehatan, keuangan, pemasaran, hingga analisis opini.

Model klasifikasi sering dimanfaatkan untuk mengolah data teks agar dapat menentukan sentimen atau kategori yang sesuai (Maruli Tua Silaen, 2023). Data teks yang sebelumnya tidak terstruktur diubah terlebih dahulu menjadi representasi numerik melalui proses pembobotan atau ekstraksi fitur, kemudian dimasukkan ke dalam model untuk dipelajari. Dengan demikian, model dapat mengenali pola kata atau frasa tertentu yang berhubungan dengan kategori tertentu sehingga mampu memprediksi hasil klasifikasi secara otomatis. Keberadaan model klasifikasi ini membantu menghasilkan analisis yang lebih terukur, terutama ketika data yang digunakan memiliki jumlah yang besar dan tidak memungkinkan untuk diproses secara manual.

Penerapan model klasifikasi memberikan banyak keuntungan, mulai dari kemampuannya bekerja secara cepat, efisien, hingga menghasilkan prediksi yang konsisten terhadap data yang dianalisis. Model ini juga fleksibel karena dapat dikombinasikan dengan berbagai teknik pengolahan data awal, seperti pembobotan

kata *TF-IDF*, untuk meningkatkan akurasi prediksi. Dengan dukungan algoritma yang tepat dan pemrosesan data yang sistematis, model klasifikasi dapat memberikan gambaran yang jelas mengenai pola yang muncul pada data, sehingga hasil analisis yang diperoleh lebih mudah dipahami dan dapat digunakan untuk mendukung pengambilan keputusan yang lebih informasional.

2. *Pattern Recognition*

Pattern Recognition adalah bidang dalam ilmu komputer dan kecerdasan buatan yang berfokus pada pengenalan pola atau struktur tertentu dari sekumpulan data (Sekar Setyaningtyas et al., 2022). Bidang ini mempelajari bagaimana sistem dapat mengidentifikasi keteraturan, kesamaan, maupun karakteristik khusus dari data yang beragam, baik dalam bentuk teks, gambar, suara, maupun sinyal lainnya. Proses ini melibatkan tahapan seperti ekstraksi fitur, klasifikasi, dan pengelompokan, yang memungkinkan sistem untuk memahami data secara otomatis. Dengan menggunakan algoritma tertentu, *Pattern Recognition* mampu mengubah data mentah menjadi informasi yang lebih bermakna serta membantu dalam pengambilan keputusan berdasarkan pola yang ditemukan.

Pattern Recognition dapat diterapkan untuk mengolah dan menganalisis data teks agar dapat mengenali kategori sentimen dari ulasan yang diberikan pengguna (Lin et al., 2021). Teknik ini memungkinkan proses identifikasi pola kata, interpretasi frekuensi kemunculan istilah, dan pemetaan hubungan antar kata sehingga sentimen dapat diklasifikasikan ke dalam kategori tertentu secara otomatis. Ulasan yang semula bersifat tidak terstruktur dapat diubah menjadi pola

yang dapat dipelajari oleh algoritma, sehingga menghasilkan analisis yang lebih terarah dan terukur.

Dengan kemampuan untuk mendeteksi struktur dan kecenderungan dalam data, *Pattern Recognition* memberikan dasar yang kuat bagi pengembangan model analisis sentimen yang efektif. Proses pengenalan pola membantu memperjelas hubungan antara fitur teks dan kategori sentimen, sehingga model yang dibangun dapat menghasilkan prediksi yang lebih akurat. Melalui pendekatan ini, proses analisis menjadi lebih efisien, konsisten, dan mampu memberikan gambaran yang lebih jelas mengenai persepsi pengguna berdasarkan pola yang ditemukan dalam data ulasan.

3. Metode *Naïve Bayes*

Metode *Naïve Bayes* merupakan salah satu algoritma klasifikasi yang berlandaskan pada teori probabilitas dan Teorema Bayes (Pratama et al., 2023). Algoritma ini bekerja dengan asumsi bahwa setiap fitur dalam data saling bebas satu sama lain, meskipun dalam kenyataannya hubungan tersebut tidak selalu sepenuhnya independent (Anam et al., 2022). Meskipun asumsi tersebut tergolong sederhana, *Naïve Bayes* dikenal sangat efektif dalam menangani berbagai jenis permasalahan klasifikasi, termasuk pengolahan teks, spam filtering, deteksi topik, dan analisis sentimen. Keunggulan metode ini terletak pada kemampuannya memproses data berukuran besar dengan cepat serta menghasilkan model yang stabil meskipun data yang digunakan tidak terlalu banyak.

Naïve Bayes dapat bekerja optimal untuk klasifikasi teks karena algoritma ini mampu menghitung probabilitas kemunculan kata dalam setiap kategori sentimen

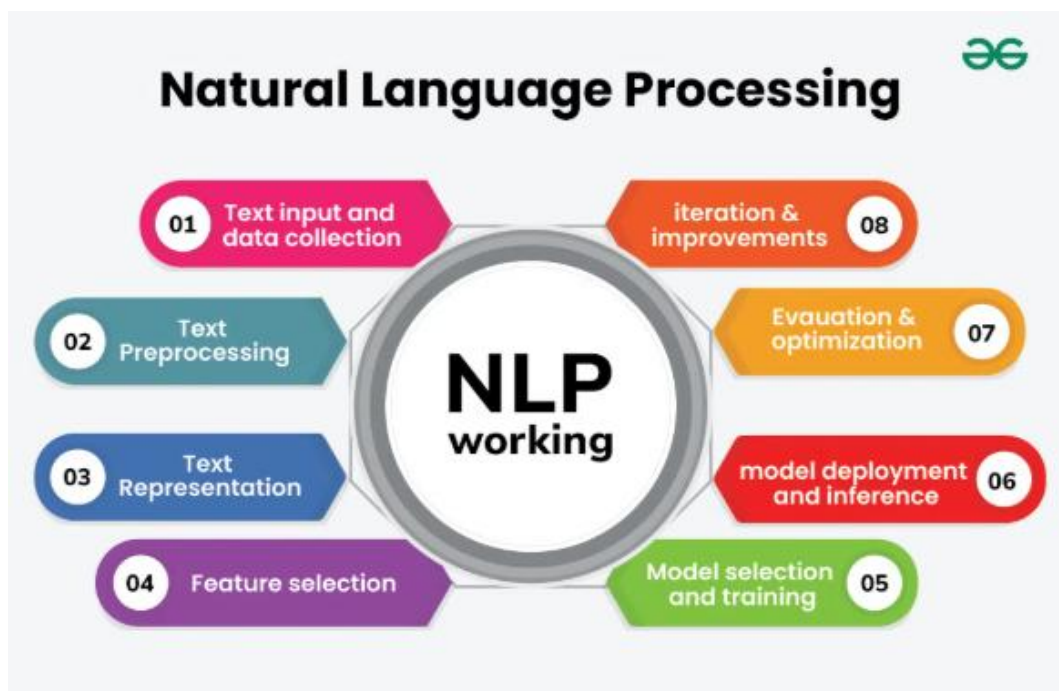
sehingga dapat menentukan kelas yang paling sesuai untuk suatu dokumen (Ayuningtyas et al., 2022). Cara kerja berbasis probabilitas membuatnya efisien dalam menganalisis ulasan yang memiliki struktur bahasa beragam, seperti yang ditemukan pada komentar pengguna aplikasi. Penggunaan pembobotan kata melalui *TF-IDF* dapat meningkatkan performa *Naïve Bayes* dengan memberikan representasi yang lebih akurat terhadap kata-kata yang memiliki pengaruh kuat dalam menentukan sentimen. Kombinasi ini menghasilkan proses pemodelan yang lebih terarah dalam mengelompokkan teks ke dalam kategori tertentu.

Metode *Naïve Bayes* sering dipilih karena kesederhanaannya dalam implementasi serta kemampuannya memberikan hasil klasifikasi yang kompetitif dibandingkan algoritma lain yang lebih kompleks (Rayuwati et al., 2022). Pendekatan probabilistik yang digunakan memungkinkan model untuk tetap bekerja dengan baik meskipun terdapat kata-kata baru dalam data yang belum pernah muncul dalam proses pelatihan. Dengan performa yang cepat, tingkat akurasi yang relatif baik, serta efektivitas ketika dikombinasikan dengan teknik pembobotan seperti *TF-IDF*, *Naïve Bayes* menjadi salah satu algoritma yang sangat relevan untuk proses analisis sentimen dan klasifikasi teks dalam penelitian berbasis data ulasan.

2.2. *Natural Language Processing (NLP)*

Natural Language Processing (NLP) adalah cabang dari kecerdasan buatan yang berfokus pada bagaimana komputer dapat memahami, memproses, dan menganalisis bahasa manusia (Kusuma & Cahyono, 2023). NLP memungkinkan mesin untuk menginterpretasikan teks atau ucapan dengan cara yang mendekati

pemahaman manusia melalui berbagai teknik seperti tokenisasi, *Stemming*, lemmatization, pembobotan kata, hingga klasifikasi teks. Teknologi ini digunakan dalam berbagai aplikasi modern seperti chatbot, sistem rekomendasi, pendeteksi spam, mesin penerjemah, hingga analisis sentimen. Dengan kemampuan mengolah data teks dalam jumlah besar, NLP menjadi salah satu bidang terpenting dalam perkembangan sistem berbasis data.



Gambar 2. 2. Natural Language Processing (NLP)

NLP digunakan untuk mengubah kumpulan teks tidak terstruktur menjadi informasi yang dapat dianalisis lebih lanjut. Proses seperti pembersihan teks, pemilahan kata, penghapusan kata tidak penting, dan pembobotan kata menjadi langkah yang umum dalam pemrosesan data berbasis teks (Asrawi et al., 2023). Dengan dukungan teknik NLP, ulasan yang awalnya berupa teks mentah dapat dipetakan menjadi representasi numerik yang dibutuhkan untuk tahap klasifikasi.

Proses ini membuat analisis terhadap teks menjadi lebih terarah dan objektif karena setiap elemen bahasa dipecah dan diproses secara sistematis.

Peran NLP sangat penting dalam menghasilkan model analisis yang lebih akurat karena ia menyediakan fondasi awal berupa struktur data teks yang siap diolah oleh algoritma pembelajaran mesin. Dengan adanya teknik-teknik dalam NLP, model dapat bekerja lebih optimal dalam mengenali pola kata, menentukan bobot informasi, dan membedakan antara berbagai kategori sentimen. Integrasi NLP dengan algoritma klasifikasi memungkinkan pemahaman yang lebih mendalam terhadap opini pengguna yang sebelumnya sulit diukur secara manual. Dengan demikian, NLP menjadi komponen utama yang membantu seluruh proses analisis teks berjalan secara efektif, terstruktur, dan menghasilkan keluaran yang lebih menggambarkan kondisi sebenarnya.

2.2.1. Sentimen

Sentimen merupakan ekspresi perasaan, opini, atau tanggapan seseorang terhadap suatu objek, layanan, atau pengalaman tertentu (Putra & Ratih, 2021). Dalam analisis data, sentimen biasanya dikategorikan ke dalam kelompok positif, negatif, atau netral berdasarkan isi dan konteks kalimat yang disampaikan. Analisis sentimen digunakan untuk memahami persepsi publik terhadap suatu topik dengan memanfaatkan data teks seperti komentar, ulasan, atau postingan di media sosial. Melalui teknik *text mining* dan pembelajaran mesin, sentimen dapat diidentifikasi secara otomatis sehingga memberikan gambaran yang lebih objektif dan terukur tentang opini pengguna.

Sentimen yang diekspresikan dalam ulasan pengguna sering kali menggambarkan tingkat kepuasan, pengalaman, atau persepsi terhadap layanan yang mereka gunakan Sentimen (Kumar et al., 2021)(Widya Utami & Artana, 2022). Pemrosesan teks yang tepat dapat membantu mengklasifikasikan ulasan tersebut ke dalam kategori sentimen tertentu agar mudah dianalisis secara keseluruhan. Pengelompokan sentimen berdasarkan kata-kata yang muncul memungkinkan model untuk mengenali pola tertentu dalam opini pengguna sehingga menghasilkan interpretasi yang lebih terstruktur terhadap data teks.

Dengan memahami sentimen secara lebih mendalam, proses analisis menjadi lebih efektif karena setiap ulasan dapat diolah menjadi informasi yang bernilai untuk menggambarkan persepsi pengguna secara menyeluruh. Pendekatan ini tidak hanya membantu memetakan kecenderungan opini, tetapi juga dapat digunakan untuk mengevaluasi kualitas layanan, mengidentifikasi masalah, dan memberikan dasar bagi pengambilan keputusan yang lebih strategis. Sentimen yang telah terklasifikasi dapat menjadi acuan penting dalam melihat tren kepuasan dan membantu pengembangan sistem yang lebih responsif terhadap kebutuhan pengguna.

1. *Tokenizing*

Tokenizing adalah salah satu tahap dasar dalam pemrosesan bahasa alami (*Natural Language Processing*) yang bertujuan untuk memecah teks menjadi unit-unit yang lebih kecil, seperti kata, frasa, atau symbol (Widya Utami & Artana, 2022). Proses ini penting karena komputer tidak dapat memahami teks secara langsung tanpa memecahnya ke dalam bentuk yang dapat diolah. *Tokenizing*

biasanya dilakukan dengan memisahkan teks berdasarkan spasi, tanda baca, atau pola tertentu, sehingga menghasilkan daftar token yang dapat digunakan untuk analisis lebih lanjut. Tahap ini menjadi fondasi bagi proses-proses berikutnya seperti filtering, *Stemming*, lemmatization, dan pembobotan kata.

Tokenizing digunakan untuk memecah setiap ulasan menjadi kumpulan kata yang dapat dianalisis oleh sistem. Dengan token-token tersebut, proses pembobotan menggunakan *TF-IDF* dapat berjalan karena setiap kata yang muncul akan dihitung frekuensinya. *Tokenizing* juga membantu model klasifikasi seperti *Naïve Bayes* mengenali distribusi kata dalam setiap kelas sentimen sehingga proses pengklasifikasian dapat dilakukan secara lebih terstruktur.

Dengan adanya *Tokenizing*, alur pemrosesan teks menjadi lebih sistematis dan terukur, karena setiap kata yang muncul dalam ulasan dapat diidentifikasi, dihitung, dan diproses lebih lanjut. *Tokenizing* juga meningkatkan akurasi dalam tahap ekstraksi fitur karena membantu mengurangi ambiguitas dalam teks. Tahapan ini memastikan bahwa seluruh proses analisis sentimen berjalan dengan baik, dimulai dari pengolahan teks mentah hingga pembentukan model klasifikasi yang optimal. *Tokenizing* menjadi bagian integral dari keseluruhan proses pengolahan data teks yang berperan penting dalam menghasilkan output analisis yang akurat dan dapat dipertanggungjawabkan.

2. *Stopword*

Stopword merupakan istilah dalam pengolahan teks yang merujuk pada kata-kata umum yang sering muncul tetapi tidak memiliki kontribusi signifikan terhadap pemaknaan suatu dokumen, seperti “dan”, “yang”, “atau”, “di”, dan “ke”. Kata-kata

tersebut biasanya dihapus selama proses *Preprocessing* karena tidak memberikan nilai informatif yang cukup untuk analisis lebih lanjut (Kusuma & Cahyono, 2023). Penghapusan *Stopword* membantu mengurangi noise dalam data, memperkecil dimensi kata, serta meningkatkan fokus analisis pada kata-kata yang lebih relevan dengan isi atau sentimen suatu teks. Dengan demikian, proses ini menjadi salah satu tahapan penting dalam *text mining* dan *Natural Language Processing*.

Stopword berperan dalam membantu memurnikan data teks agar analisis yang dilakukan lebih efektif. Ketika *Stopword* dihilangkan, kata-kata penting seperti “bagus”, “buruk”, “cepat”, atau “mengecewakan” menjadi lebih dominan dalam proses pembobotan maupun klasifikasi (Halim & Andri Safuwani, 2023). Tahap ini juga mempermudah pembentukan fitur karena hanya kata-kata dengan makna kuat yang dipertahankan untuk diproses oleh metode pembobotan seperti *TF-IDF* atau algoritma klasifikasi seperti *Naïve Bayes*. Hal tersebut membuat hasil analisis menjadi lebih terarah dan representatif.

Dengan diterapkannya proses penghapusan *Stopword*, model dapat bekerja lebih efisien karena hanya berfokus pada kata-kata yang benar-benar berkontribusi terhadap pengenalan pola sentimen. Data teks yang telah dibersihkan dari kata-kata tidak penting menghasilkan matriks fitur yang lebih padat makna dan mengurangi risiko misinterpretasi oleh model pembelajaran mesin. Tahapan ini juga membantu meningkatkan akurasi dan kualitas analisis secara keseluruhan, karena model tidak terbebani oleh kata-kata umum yang tidak relevan dan dapat memproses informasi penting dengan lebih optimal.

3. *Stemming*

Stemming merupakan proses dalam pengolahan bahasa alami yang bertujuan mengubah kata berimbuhan menjadi bentuk dasarnya (Widya Utami & Artana, 2022). Teknik ini bekerja dengan cara menghilangkan awalan, akhiran, atau imbuhan lain yang menempel pada suatu kata sehingga menghasilkan bentuk kata yang lebih sederhana. Dengan mengembalikan kata ke bentuk dasar, jumlah variasi kata menjadi berkurang dan data teks menjadi lebih terstruktur, sehingga memudahkan proses analisis lebih lanjut. *Stemming* banyak digunakan dalam *text mining*, pencarian informasi, dan analisis sentimen untuk meningkatkan akurasi model yang bekerja berdasarkan pola kata.

Stemming membantu menyederhanakan dataset teks dengan mengurangi keragaman kata yang muncul akibat penggunaan imbuhan atau bentuk kata yang berbeda-beda. Kata seperti “membeli,” “dibeli,” dan “pembelian” dapat diubah menjadi bentuk dasar yang sama, sehingga pemodelan teks dapat lebih fokus pada makna inti dari kata tersebut. Penyederhanaan struktur kata ini mendukung pembentukan matriks fitur yang lebih rapi dan konsisten, terutama ketika pembobotan kata menggunakan metode *TF-IDF* diterapkan untuk menghasilkan representasi numerik dari setiap kata.

Dengan penerapan *Stemming* yang tepat, proses analisis sentimen menjadi lebih efisien karena model dapat bekerja dengan data kata yang lebih homogen dan representatif. Teknik ini meningkatkan peluang model klasifikasi dalam mengenali pola sentimen secara lebih akurat, mengingat setiap kata dalam bentuk dasar memiliki makna yang lebih langsung untuk dianalisis. Penggunaan *Stemming* juga

mampu mengurangi noise pada data teks sehingga hasil klasifikasi menjadi lebih stabil dan mudah diinterpretasikan.

2.2.2. TF-IDF

TF-IDF merupakan metode pembobotan kata yang digunakan dalam pengolahan data teks untuk menilai seberapa penting sebuah kata dalam sebuah dokumen dibandingkan dengan keseluruhan kumpulan dokumen (Rambe et al., 2023). Metode ini terdiri dari dua komponen utama, yaitu *Term Frequency (TF)* yang menggambarkan frekuensi kemunculan kata dalam sebuah dokumen, serta *Inverse Document Frequency (IDF)* yang menunjukkan tingkat keunikan kata tersebut dalam seluruh dokumen (Kusuma Dewi et al., 2023). Kombinasi keduanya menghasilkan nilai bobot yang dapat membedakan kata-kata yang benar-benar memiliki makna penting dari kata-kata umum yang sering muncul. Dengan demikian, *TF-IDF* mampu membantu sistem untuk memahami representasi numerik dari teks secara lebih efisien dan informatif.

TF-IDF digunakan untuk mengubah ulasan pengguna menjadi bentuk numerik agar dapat diolah oleh algoritma klasifikasi. Kata-kata yang memiliki pengaruh besar terhadap sentimen akan mendapatkan bobot lebih tinggi, sehingga model dapat mengenali pola opini dengan lebih tepat (Husein et al., 2023). Pembobotan yang dihasilkan memungkinkan proses klasifikasi berjalan lebih terarah karena kata yang memiliki makna kuat terhadap sentimen—seperti “bagus”, “cepat”, atau “mengecewakan”—menjadi lebih dominan dibandingkan kata yang kurang relevan. Representasi numerik ini menjadi dasar bagi algoritma untuk melakukan proses pengenalan pola pada data teks.

Penggunaan *TF-IDF* memungkinkan proses analisis menjadi lebih akurat karena membantu menonjolkan kata-kata yang memiliki karakteristik penting dalam pengambilan keputusan model (Rahayu et al., 2022). Bobot yang dihasilkan tidak hanya memperjelas struktur informasi pada teks, tetapi juga meminimalkan pengaruh kata-kata yang terlalu umum agar tidak mendominasi proses perhitungan. Kemampuan *TF-IDF* dalam menghasilkan representasi fitur yang jelas menjadikannya metode yang relevan untuk berbagai kebutuhan analisis teks, termasuk penilaian tingkat kepuasan pengguna berdasarkan ulasan yang mereka berikan. Dengan cara ini, *TF-IDF* memberikan fondasi yang kuat untuk membangun model klasifikasi yang lebih baik.

2.3. Aplikasi *Shopee*

Aplikasi *Shopee* merupakan salah satu platform perdagangan elektronik (*e-commerce*) yang banyak digunakan oleh masyarakat untuk melakukan kegiatan jual beli secara online. *Shopee* menyediakan berbagai fitur yang memudahkan pengguna dalam mencari produk, melakukan transaksi, serta berinteraksi dengan penjual melalui sistem yang terintegrasi dalam satu aplikasi. Selain itu, *Shopee* juga menawarkan berbagai layanan pendukung seperti promo gratis ongkir, sistem pembayaran digital, serta fitur ulasan dan penilaian dari pengguna setelah melakukan pembelian. Fitur ulasan tersebut memungkinkan pengguna untuk memberikan pendapat, pengalaman, maupun tingkat kepuasan terhadap produk dan layanan yang diterima, sehingga dapat menjadi referensi bagi pengguna lain sebelum melakukan pembelian.

Dalam penelitian ini, aplikasi *Shopee* digunakan sebagai sumber data ulasan pengguna yang dianalisis untuk mengetahui tingkat kepuasan pengguna terhadap layanan yang diberikan. Data ulasan yang diperoleh kemudian diolah menggunakan teknik analisis sentimen dengan metode *TF-IDF* sebagai pembobotan kata dan algoritma *Naive Bayes* sebagai metode klasifikasi. Melalui analisis sentimen tersebut, ulasan pengguna dapat dikategorikan ke dalam sentimen positif maupun negatif sehingga dapat memberikan gambaran mengenai persepsi pengguna terhadap aplikasi *Shopee*. Hasil analisis ini diharapkan dapat memberikan informasi yang bermanfaat dalam memahami pengalaman pengguna serta menjadi bahan evaluasi untuk meningkatkan kualitas layanan aplikasi *Shopee* di masa yang akan datang.

2.4. Google Colab

Google Colaboratory (Google Colab) merupakan sebuah layanan berbasis cloud yang disediakan oleh *Google* untuk menjalankan kode pemrograman Python secara langsung melalui browser tanpa perlu melakukan instalasi perangkat lunak tambahan pada komputer. Platform ini sangat populer dalam bidang analisis data, machine learning, dan penelitian berbasis pemrograman karena menyediakan berbagai library yang telah terintegrasi seperti *NumPy*, *Pandas*, *NLTK*, *Scikit-learn*, serta dukungan komputasi yang cukup baik. Selain itu, *Google Colab* juga memungkinkan pengguna untuk mengakses dan menyimpan *file* melalui *Google Drive*, sehingga memudahkan dalam pengelolaan dataset serta proses kolaborasi dalam penelitian.

Dalam penelitian ini, *Google Colab* digunakan sebagai lingkungan pengembangan untuk melakukan analisis sentimen terhadap ulasan pengguna aplikasi Shopee. Melalui *Google Colab*, proses pengolahan data dimulai dari tahap upload dataset, *preprocessing* teks seperti *Cleaning*, tokenisasi, *Stopword removal*, dan stemming, hingga proses pembobotan kata menggunakan metode *TF-IDF*. Selanjutnya, data yang telah diproses digunakan untuk membangun model klasifikasi menggunakan algoritma Naive Bayes, serta melakukan evaluasi model untuk mengetahui tingkat akurasi klasifikasi. Penggunaan *Google Colab* dalam penelitian ini memberikan kemudahan dalam menjalankan berbagai tahapan analisis data secara terintegrasi serta memungkinkan proses eksperimen dilakukan dengan lebih efisien dan fleksibel.