

## **BAB III**

### **METODOLOGI PENELITIAN**

#### **3.1 Gambaran Umum Objek Penelitian**

##### **3.1.1 Ulasan Pelanggan Fore Coffee**

Objek penelitian ini berfokus pada ulasan dan opini pelanggan terhadap Fore Coffee yang dipublikasikan melalui platform media sosial Twitter/X. Fore Coffee merupakan salah satu merek kopi lokal yang berkembang pesat di Indonesia dan memiliki basis pelanggan yang cukup aktif di dunia digital. Melalui Twitter, pelanggan dapat menyampaikan pengalaman, tanggapan, serta kritik terhadap produk dan layanan yang ditawarkan oleh Fore Coffee secara terbuka.

Data yang digunakan dalam penelitian ini bersumber dari unggahan publik (*tweet*) yang mengandung kata kunci terkait Fore Coffee, seperti “fore coffee”, “fore kopi”, atau tagar serupa. Twitter dipilih sebagai sumber data karena platform ini memiliki karakteristik teks pendek yang bersifat spontan dan aktual, sehingga dapat merepresentasikan opini pelanggan secara langsung dan tanpa intervensi pihak perusahaan. Opini yang muncul di Twitter juga mencerminkan persepsi publik terhadap kualitas produk, pelayanan, maupun citra merek Fore Coffee di mata konsumen.

Dalam penelitian ini, data tweet dikumpulkan menggunakan *Application Programming Interface (API)* yang disediakan oleh X Developer, dengan batasan waktu tertentu yaitu dari tahun 2022 hingga 2025. Proses pengumpulan dilakukan dengan memfilter tweet agar tidak mencakup unggahan dari akun resmi @ForeCoffeeID, *retweet*, maupun balasan promosi yang tidak mengandung opini personal. Hal ini dilakukan agar data yang digunakan benar-benar mewakili pendapat asli pelanggan.

Objek penelitian ini dipilih karena Fore Coffee merupakan brand yang aktif berinteraksi di media sosial dan sering menjadi bahan diskusi pengguna Twitter, baik dalam konteks positif maupun negatif. Selain itu, karakteristik bahasa yang digunakan pengguna Twitter Indonesia yang

sering kali informal, singkat, dan mengandung unsur emosional menjadikan dataset ini menarik untuk diteliti menggunakan pendekatan *Natural Language Processing (NLP)*[15].

Penelitian ini berupaya mengidentifikasi serta mengklasifikasikan sentimen pelanggan Fore Coffee menjadi tiga kategori, yaitu positif, negatif, dan netral, dengan menggunakan empat pendekatan algoritma yang berbeda: *Naïve Bayes*, *Logistic Regression*, *Support Vector Machine (SVM)*, dan *IndoBERTweet*. Hasil dari klasifikasi ini akan dibandingkan untuk mengetahui algoritma mana yang memiliki performa terbaik dalam mengolah teks berbahasa Indonesia pada konteks media sosial.

Secara keseluruhan, Fore Coffee dipilih sebagai objek penelitian karena merepresentasikan merek lokal yang memiliki interaksi tinggi di ranah digital serta relevan dengan perkembangan riset analisis sentimen di Indonesia. Melalui penelitian ini, diharapkan dapat diperoleh gambaran yang lebih jelas mengenai persepsi publik terhadap Fore Coffee, serta bagaimana penerapan teknologi *machine learning* dan *deep learning* dapat membantu dalam memahami opini pelanggan secara lebih sistematis[2].

### 3.2 Metode Penelitian

Penelitian ini menggunakan metode *CRISP-DM (Cross Industry Standard Process for Data Mining)* sebagai kerangka kerja utama dalam pelaksanaan penelitian. Metode ini dipilih karena memiliki tahapan yang sistematis dan sesuai untuk penelitian yang berfokus pada analisis data berbasis teks. Tahapan dalam *CRISP-DM* meliputi enam langkah utama, yaitu *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment*[31].

Namun, dalam penelitian ini tahap *deployment* tidak dilaksanakan karena fokus penelitian adalah pada proses pengolahan data, pemodelan, serta evaluasi dan perbandingan kinerja model *machine learning* (*Naïve Bayes*, *Logistic Regression*, dan *Support Vector Machine*) dengan *deep learning* (*IndoBERTweet*). Dengan demikian, penelitian ini hanya

melaksanakan tahapan hingga evaluation untuk memperoleh hasil analisis performa model yang paling optimal dalam mengklasifikasikan sentimen ulasan pelanggan Fore Coffee di Twitter.

Tahapan-tahapan dalam metode *CRISP-DM* diterapkan secara berurutan untuk memastikan bahwa proses analisis data dilakukan secara terstruktur dan terarah, dimulai dari pemahaman konteks bisnis hingga evaluasi hasil model.

## 2. Business Understanding

Tahap ini dilakukan untuk memperoleh pemahaman yang jelas mengenai arah dan tujuan penelitian. Penelitian ini difokuskan pada evaluasi performa beberapa algoritma untuk menentukan model yang paling efektif dalam mengklasifikasikan sentimen pelanggan Fore Coffee berdasarkan tweet. Hasil dari pendekatan ini diharapkan dapat memberikan informasi kuantitatif mengenai opini publik terhadap Fore Coffee, sehingga dapat dimanfaatkan perusahaan dalam pengambilan keputusan terkait peningkatan layanan dan strategi pemasaran digital.

## 3. Data Understanding

Pada tahap ini dilakukan proses pengumpulan dan eksplorasi data awal. Data dikumpulkan dari platform Twitter (X) menggunakan *API* dari *X Developer*, dengan kata kunci seperti “fore coffee”, “fore kopi”, dan “fore”. Tweet yang diambil merupakan tweet berbahasa Indonesia dari tahun 2022 hingga 2025. Data tersebut dieksplorasi untuk mengetahui jumlah total tweet, distribusi kata, serta proporsi sentimen awal. Selain itu, dilakukan pemeriksaan terhadap kemungkinan adanya *noise* seperti *retweet*, tweet promosi, dan unggahan dari akun resmi Fore Coffee yang kemudian dihapus agar dataset lebih bersih dan relevan.

## 4. Data Preparation

Tahap ini merupakan salah satu bagian terpenting dalam CRISP-DM karena kualitas data sangat berpengaruh terhadap hasil model. Pada tahap ini dilakukan serangkaian proses *Natural Language Processing* (NLP) yang meliputi :

1. *Case folding*: mengubah seluruh huruf menjadi huruf kecil agar seragam.
2. *Tokenization*: memisahkan kalimat menjadi kata-kata individu.
3. *Stopword removal*: menghapus kata umum yang tidak memiliki makna penting seperti "yang", "dan", "di".
4. *Stemming*: mengubah kata berimbuhan menjadi bentuk dasarnya menggunakan library *Sastrawi*.
5. *Labeling*: memberikan label sentimen (positif, negatif, atau netral) berdasarkan isi teks.

Setelah itu, data hasil *preprocessing* dikonversi menjadi bentuk numerik menggunakan *TF-IDF vectorization* untuk model *machine learning*, sementara data untuk *IndoBERTweet* diubah menjadi *token embedding* sesuai dengan format input model *transformer*.

## 5. Modeling

Tahap ini berfokus pada pembangunan dan pelatihan model klasifikasi sentimen menggunakan empat algoritma berbeda, yaitu:

1. *Naïve Bayes*: algoritma probabilistik yang sederhana dan cepat dalam mengklasifikasikan teks pendek.
2. *Support Vector Machine (SVM)*: algoritma dengan kemampuan tinggi dalam menemukan *hyperplane* terbaik untuk memisahkan kelas sentimen.
3. *Logistic Regression* : algoritma dengan kemampuan mengklasifikasikan data dengan mengestimasi probabilitas berbasis fungsi logistic.
4. *IndoBERTweet*: model *transformer* berbahasa Indonesia yang dilatih menggunakan lebih dari 500 juta tweet, sehingga unggul dalam memahami konteks bahasa informal di media sosial.

Keempat model ini dilatih dan diuji menggunakan dataset yang sama agar hasil perbandingan performa menjadi objektif. Pelatihan model dilakukan di *Google Colab* dengan bantuan library seperti *Scikit-learn* untuk *machine learning* dan *Hugging Face Transformers* untuk *IndoBERTweet*.

## 6. Evaluation

Tahapan ini bertujuan untuk mengevaluasi hasil klasifikasi dari setiap algoritma berdasarkan sejumlah metrik evaluasi, yaitu:

1. *Accuracy*: mengukur proporsi prediksi yang tepat dibandingkan dengan seluruh data uji.
2. *Precision*: menilai sejauh mana model mampu memberikan prediksi yang akurat untuk suatu kelas tertentu.
3. *Recall*: menghitung kemampuan model dalam mendekripsi seluruh data yang benar-benar termasuk ke dalam suatu kelas.
4. *F1-score*: merupakan kombinasi harmonis antara precision dan recall yang memberikan gambaran keseimbangan performa model.

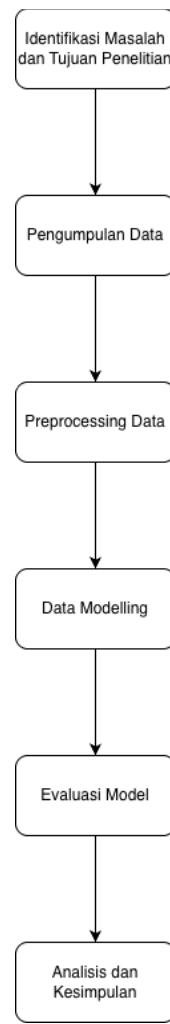
Selain metrik tersebut, digunakan juga *confusion matrix* sebagai alat visualisasi untuk menunjukkan jumlah prediksi benar dan salah pada tiap kategori sentimen. Evaluasi ini diterapkan pada semua algoritma dengan tujuan mengidentifikasi model yang menghasilkan performa paling optimal secara kuantitatif.

## 7. Deployment

Tahap terakhir dalam *CRISP-DM* adalah *deployment*, yaitu penyajian hasil penelitian ke dalam bentuk visualisasi grafik dan tabel agar mudah dipahami. Tahap ini tidak mencakup penerapan model ke sistem produksi karena penelitian bersifat akademik. Namun, hasil penelitian dapat dijadikan dasar untuk pengembangan sistem analisis sentimen otomatis di masa mendatang yang dapat membantu perusahaan dalam memantau opini publik secara real-time[31].

### 3.2.1 Alur Penelitian

Alur penelitian menjelaskan tahapan-tahapan yang dilakukan peneliti dari awal hingga akhir penelitian berdasarkan metodologi *CRISP-DM* (*Cross Industry Standard Process for Data Mining*). Setiap tahap saling berkaitan secara sistematis untuk mencapai tujuan utama penelitian, yaitu membandingkan kinerja algoritma *machine learning* dan *deep learning* dalam menganalisis sentimen pelanggan Fore Coffee di Twitter.



Gambar 3.1 Alur Penelitian

Dari gambar 3.1 Alur Penelitian tersebut dapat dijabarkan sebagai berikut:

1. Perumusan Masalah dan Tujuan Penelitian

Pada tahap ini dilakukan identifikasi masalah yang akan diteliti, yaitu bagaimana membandingkan tingkat akurasi dan *F1-score* antara algoritma *Naïve Bayes*, *Logistic Regression*, *Support Vector Machine (SVM)*, dan *IndoBERTweet* dalam analisis sentimen. Selain itu, ditetapkan pula tujuan penelitian, yaitu untuk mengetahui algoritma yang paling efektif dalam memahami opini pelanggan berbahasa Indonesia di media sosial.

2. Pengumpulan Data

Data diperoleh melalui proses scraping menggunakan Twitter API (*X Developer*) dengan kata kunci “fore coffee”, “fore kopi”, dan “fore”. Data dikumpulkan dalam rentang waktu tahun 2022 hingga 2025 dan disimpan dalam format *CSV*. Tweet dari akun resmi Fore Coffee serta *retweet* dihapus agar data yang digunakan hanya berisi opini asli dari pengguna.

### 3. Preprocessing Data

Pada bagian ini, dataset awal diproses dan dibersihkan dengan metode *Natural Language Processing (NLP)*. Tahapan yang dilakukan meliputi *case folding*, *tokenization*, penghapusan *stopword*, serta *stemming* menggunakan library *Sastrawi*. Setelah proses tersebut, setiap tweet diberikan label sentimen ke dalam tiga kelas, yaitu positif, netral, dan negatif. Data yang telah diproses kemudian dikonversi ke bentuk numerik menggunakan *TF-IDF* untuk model machine learning, sedangkan *IndoBERTweet* menggunakan *token embedding* sebagai format inputnya.

### 4. Pemodelan (Modeling)

Setelah data siap, dilakukan pembangunan model menggunakan empat algoritma berbeda:

1. *Naïve Bayes* untuk klasifikasi berbasis probabilistik,
2. *SVM* untuk pemisahan data menggunakan *hyperplane*,
3. *Logistic Regression* untuk mengklasifikasikan data dengan mengestimasi probabilitas, dan
4. *IndoBERTweet* sebagai model *deep learning* berbasis *transformer*.

Semua proses pemodelan dilakukan menggunakan *Python* di *Google Colab* agar memanfaatkan dukungan GPU dan kemudahan integrasi dengan library *machine learning*.

### 5. Evaluasi Model

Hasil klasifikasi dari keempat algoritma kemudian dievaluasi menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Selain itu, *confusion matrix* juga digunakan untuk melihat distribusi

prediksi benar dan salah dari tiap kelas. Tahapan ini bertujuan untuk menentukan model mana yang paling baik dalam mengenali pola sentimen pelanggan Fore Coffee.

## 6. Analisis dan Kesimpulan

Setelah evaluasi, dilakukan analisis hasil untuk mengetahui algoritma dengan performa terbaik. Tahapan ini juga mencakup interpretasi hasil perbandingan dan penarikan kesimpulan yang didukung oleh nilai kuantitatif dari masing-masing metrik evaluasi.

### 3.3 Teknik Pengumpulan Data

Pengumpulan data pada penelitian ini dilakukan melalui pendekatan web scraping terhadap platform media sosial Twitter (X). Platform ini dipilih karena bersifat terbuka dan banyak digunakan oleh masyarakat Indonesia untuk mengekspresikan opini terhadap suatu produk atau layanan, termasuk Fore Coffee. Data yang dikumpulkan berupa tweet dari pengguna Twitter yang berisi ulasan, tanggapan, maupun pengalaman pribadi mereka terhadap Fore Coffee. Pendekatan ini memungkinkan peneliti memperoleh data opini publik secara langsung dan aktual tanpa harus melalui survei manual.

Proses pengumpulan data dilakukan menggunakan skrip *Python* yang dijalankan di lingkungan *Google Colab*. *Notebook* tersebut berfungsi untuk mengotomatisasi proses pengambilan data berdasarkan kata kunci tertentu dengan bantuan library seperti *sns scrape* dan *pandas*. Dengan metode scraping ini, peneliti dapat memperoleh data dalam jumlah besar secara efisien dan tetap sesuai dengan batasan pencarian yang ditentukan.

Adapun parameter dan ketentuan teknis dalam proses pengumpulan data dijelaskan sebagai berikut.

1. Bahasa tweet yang dikumpulkan dibatasi hanya pada bahasa Indonesia, agar hasil analisis dapat merepresentasikan persepsi pelanggan lokal terhadap Fore Coffee.
2. Data diambil dengan rentang waktu antara 1 Januari 2022 hingga 30 September 2025, yang mencakup periode pasca pandemi serta masa aktif promosi merek Fore Coffee. Rentang waktu ini dipilih agar dapat menggambarkan perubahan opini publik secara lebih komprehensif.

3. Hanya tweet asli yang diikutsertakan dalam dataset, sedangkan retweet dan reply dikecualikan untuk menjaga keaslian opini. Tweet dari akun resmi @ForeCoffeeID juga dihapus agar data hanya berisi pandangan pelanggan, bukan tweet promosi perusahaan.
4. Kata kunci pencarian yang digunakan meliputi frasa “fore coffee”, “fore kopi”, dan “fore”, yang disusun untuk menangkap berbagai variasi penyebutan merek dalam percakapan pengguna Twitter.
5. Format data hasil pengumpulan disimpan dalam bentuk *CSV* (*Comma-Separated Values*) dengan atribut utama seperti tanggal unggahan (*date*), nama pengguna (*username*), isi tweet (*content*), dan tautan menuju tweet asli (*url*).

Proses pengambilan data dilakukan secara bertahap untuk memastikan kelancaran dan konsistensi hasil. Setelah proses scraping selesai, dilakukan pengecekan awal untuk memastikan data telah terkumpul dengan benar. Tahap ini mencakup penghapusan tweet ganda, pengecekan kelengkapan atribut, serta pemisahan tweet yang tidak relevan seperti iklan, spam, atau promosi yang tidak berhubungan dengan pengalaman pelanggan. Validasi ini dilakukan agar dataset yang digunakan benar-benar bersih dan sesuai dengan konteks penelitian.

Metode scraping menggunakan snscreape dipilih karena fleksibel dan efisien, walaupun tetap memiliki batasan rate limit per 100-200. Pendekatan ini juga dinilai sesuai dengan kebutuhan penelitian berbasis opini publik dalam jumlah besar. Hasil akhir dari proses pengumpulan data berupa kumpulan teks tidak terstruktur (*unstructured text corpus*) yang merepresentasikan opini pelanggan terhadap Fore Coffee di Twitter. Dataset tersebut kemudian digunakan dalam tahap *preprocessing* untuk pembersihan data dan analisis sentimen menggunakan algoritma *Naive Bayes*, *Logistic Regression*, *Support Vector Machine (SVM)*, dan *IndoBERTweet*.

### 3.4 Teknik Pengolahan Data

Data yang diperoleh melalui proses *web scraping* selanjutnya diproses agar dapat digunakan pada analisis sentimen. Seluruh proses pengolahan dilakukan di lingkungan *Google Colab* dengan bahasa pemrograman *Python* serta

sejumlah library pendukung seperti *Pandas*, *Sastrawi*, *NLTK*, *Scikit-learn*, dan *Hugging Face Transformers*. Pada tahap ini, tujuan utamanya adalah membersihkan data, menyiapkannya dalam format yang seragam, dan mengubah teks mentah yang tidak terstruktur menjadi data yang siap digunakan dalam pelatihan model *machine learning* maupun *deep learning*.

Tahapan awal yang dilakukan adalah preprocessing untuk menormalisasi teks. Proses ini dimulai dengan *case folding*, yaitu mengonversi seluruh karakter menjadi huruf kecil agar konsisten. Setelah itu dilakukan pembersihan elemen yang tidak diperlukan seperti tautan, angka, simbol, tanda baca, serta emoji agar isi teks lebih relevan terhadap konteks analisis. Kemudian dilanjutkan dengan *tokenization* untuk memecah kalimat menjadi unit kata agar lebih mudah diproses oleh sistem.

Langkah berikutnya adalah *stopword removal*, yaitu menghapus kata-kata umum yang tidak memberikan kontribusi penting dalam analisis sentimen seperti “dan”, “atau”, serta kata serupa lainnya. Tahapan ini dilakukan menggunakan library *NLTK* dengan tambahan daftar *stopword* yang disesuaikan untuk karakteristik bahasa Indonesia di media sosial.

Setelah itu dilakukan proses stemming menggunakan library *Sastrawi* untuk mengubah kata berimbahan menjadi bentuk dasar, misalnya “dibeli” menjadi “beli”. Langkah ini penting agar kata yang memiliki akar makna sama tidak dianggap sebagai entitas yang berbeda oleh model. Setelah seluruh proses pembersihan selesai, setiap data diberi label sentimen dalam tiga kategori, yaitu positif, netral, dan negatif melalui proses labeling manual untuk memastikan akurasi data sebelum masuk ke tahap pemodelan. Dataset yang telah siap kemudian diproses dalam dua alur berbeda yaitu *machine learning* klasik dan *deep learning*.

Untuk jalur machine learning, teks terlebih dahulu diubah ke format numerik menggunakan metode *TF-IDF* (*Term Frequency–Inverse Document Frequency*). Metode ini berfungsi untuk mengidentifikasi kata-kata yang memiliki pengaruh paling besar terhadap suatu dokumen sehingga model dapat membedakan pola sentimen. Hasil representasi *TF-IDF* tersebut kemudian

digunakan untuk melatih tiga model *machine learning*, yaitu *Naive Bayes*, *Logistic Regression*, dan *Support Vector Machine (SVM)*[32].

Pada jalur *deep learning*, digunakan model *IndoBERTweet* yang tersedia melalui library *Hugging Face Transformers*. Sebelum dilakukan pelatihan, dataset diproses menggunakan *tokenizer* bawaan *IndoBERTweet* agar sesuai dengan format input *transformer*. Proses pelatihan dilakukan menggunakan parameter seperti *learning rate*, *batch size*, dan jumlah *epoch* yang telah disesuaikan untuk mendapatkan hasil pelatihan yang optimal tanpa penggunaan sumber daya yang berlebihan. Pelatihan dijalankan menggunakan GPU *Google Colab* untuk mempercepat proses komputasi.

Tahap selanjutnya adalah evaluasi performa prediksi dari setiap model. Evaluasi dilakukan menggunakan empat metrik, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. Selain itu, digunakan *confusion matrix* sebagai visualisasi untuk menunjukkan jumlah prediksi yang tepat maupun salah pada setiap kelas sentimen. Melalui evaluasi ini, performa masing-masing algoritma dapat dibandingkan sehingga dapat diketahui model yang menghasilkan klasifikasi paling akurat terhadap sentimen pelanggan Fore Coffee di Twitter.

Hasil akhir dari rangkaian proses ini berupa nilai evaluasi dari seluruh algoritma yang digunakan. Nilai-nilai tersebut kemudian dianalisis lebih lanjut untuk menentukan algoritma dengan performa terbaik serta mengevaluasi kelebihan dan keterbatasan masing-masing metode dalam konteks analisis sentimen berbahasa Indonesia pada media sosial.

### 3.5 Teknik Pengujian dan Evaluasi

Tahap pengujian dan evaluasi dilakukan untuk mengetahui kemampuan model dalam melakukan klasifikasi sentimen dengan tingkat akurasi yang optimal. Pada tahap ini, keempat algoritma yang digunakan dalam penelitian yaitu *Naïve Bayes*, *Logistic Regression*, *Support Vector Machine (SVM)*, dan *IndoBERTweet* diuji performanya. Tujuan utama pengujian ini adalah untuk membandingkan efektivitas masing-masing model dalam mengolah data sentimen pelanggan Fore Coffee berdasarkan tweet yang telah dikumpulkan dan diproses sebelumnya.

Dalam proses evaluasi, dataset dibagi menjadi dua subset, yaitu data latih (*training data*) dan data uji (*testing data*). Pembagian dilakukan dengan rasio 80:20, di mana 80% data digunakan sebagai bahan pelatihan model, sementara 20% sisanya digunakan sebagai data pengujian. Pembagian ini dipilih agar model dapat mempelajari pola sentimen dari sebagian besar dataset, lalu diuji menggunakan data yang belum pernah dilihat sebelumnya untuk menilai kemampuan generalisasinya.

Penilaian performa dilakukan dengan membandingkan prediksi model terhadap label sentimen yang sebenarnya. Empat metrik evaluasi digunakan untuk mengukur kualitas model, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. Sebagai tambahan, *confusion matrix* digunakan sebagai alat bantu visualisasi untuk melihat jumlah prediksi benar maupun salah pada setiap kelas sentimen (positif, netral, dan negatif).

### 1. Confusion Matrix

*Confusion matrix* merupakan tabel evaluasi yang menggambarkan hubungan antara hasil prediksi model dan label aktual pada data uji. Tabel ini terdiri dari empat komponen utama, yaitu :

1. True Positive (TP) : jumlah data berlabel positif yang berhasil diprediksi sebagai positif,
2. True Negative (TN) : jumlah data berlabel negatif yang berhasil diprediksi sebagai negatif,
3. False Positive (FP) : jumlah data negatif yang salah diklasifikasikan sebagai positif,
4. False Negative (FN) : jumlah data positif yang keliru diprediksi sebagai negatif.

Secara umum, struktur *confusion matrix* dapat digambarkan sebagai berikut:

	Prediksi Positif	Prediksi Negatif
Aktual Positif	True Positive (TP)	False Negative (FN)

	Prediksi Positif	Prediksi Negatif
Aktual Negatif	False Positive (FP)	True Negative (TN)

Dari *confusion matrix* ini, diperoleh nilai-nilai metrik evaluasi berikut.

## 2. Accuracy

*Accuracy* mengukur tingkat ketepatan keseluruhan model dalam melakukan klasifikasi terhadap data uji. Semakin tinggi nilai akurasi, semakin baik model dalam mengenali kelas sentimen yang benar. Rumus akurasi dinyatakan sebagai berikut:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Keterangan:

1.  $TP$  = jumlah data berlabel positif yang diklasifikasikan dengan benar,
2.  $TN$  = jumlah data berlabel negatif yang diklasifikasikan dengan benar,
3.  $FP$  = jumlah data berlabel negatif yang diklasifikasikan salah,
4.  $FN$  = jumlah data berlabel positif yang diklasifikasikan salah.

Nilai akurasi memberikan gambaran umum mengenai performa model secara keseluruhan, tetapi belum cukup untuk menilai kemampuan model dalam menangani ketidakseimbangan kelas.

## 3. Precision

*Precision* berfungsi untuk menilai tingkat ketepatan model dalam menghasilkan prediksi untuk kelas positif. Semakin tinggi nilai precision, semakin besar proporsi prediksi positif yang benar-benar berasal dari kelas positif. Rumus perhitungannya dituliskan sebagai berikut:

$$Precision = \frac{TP}{TP + FP}$$

Nilai *precision* yang tinggi menunjukkan bahwa model jarang keliru dalam memberi label positif pada data yang seharusnya termasuk kategori negatif.

#### 4. Recall

*Recall*, yang juga dikenal sebagai *Sensitivity* atau *True Positive Rate*, digunakan untuk mengukur sejauh mana model mampu mendeteksi seluruh data yang benar-benar termasuk dalam kelas positif. Rumus perhitungannya adalah:

$$\text{Recall} = \frac{TP}{TP + FN}$$

Nilai *recall* yang tinggi menandakan bahwa model berhasil mengidentifikasi sebagian besar data positif. Namun demikian, tingginya nilai *recall* tidak selalu berbanding lurus dengan *precision*, sehingga kedua metrik ini perlu dianalisis secara bersamaan untuk mendapatkan gambaran performa model yang lebih seimbang.

#### 5. F1-Score

Untuk mendapatkan ukuran performa yang seimbang antara *precision* dan *recall*, digunakan metrik *F1-score* yang dihitung sebagai rata-rata harmonik dari kedua nilai tersebut. Rumus perhitungannya adalah:

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Nilai F1-score yang tinggi menunjukkan bahwa model mampu menjaga keseimbangan yang baik antara *precision* dan *recall*. Metrik ini sering menjadi indikator utama dalam penelitian analisis sentimen, terutama pada kondisi ketika distribusi kelas tidak merata.

Tahap evaluasi dilakukan pada seluruh algoritma menggunakan metrik yang telah dijelaskan sebelumnya. Selain melihat nilai numerik seperti *accuracy* dan *F1-score*, hasil evaluasi juga ditampilkan melalui *confusion matrix* untuk mempermudah analisis performa masing-masing model. Hasil pengujian tersebut menunjukkan kemampuan setiap algoritma dalam mengelompokkan sentimen pelanggan Fore Coffee ke dalam tiga kelas, yaitu positif, negatif, dan netral.

Melalui proses ini, dapat dipastikan bahwa model yang digunakan tidak hanya menunjukkan akurasi yang baik, tetapi juga menghasilkan klasifikasi yang stabil dan sesuai dengan karakteristik opini publik yang dianalisis.