

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Media sosial telah menjadi bagian penting dalam kehidupan masyarakat modern, di mana setiap individu dapat dengan bebas menyampaikan opini, pengalaman, dan penilaian terhadap produk maupun layanan yang mereka gunakan. Platform seperti Twitter memiliki peran signifikan dalam membentuk persepsi publik terhadap suatu merek, karena sifatnya yang terbuka, cepat, dan interaktif. Melalui tweet, pelanggan dapat memberikan masukan secara langsung, baik berupa pujian maupun keluhan, yang kemudian dapat dibaca dan disebarluaskan dengan cepat oleh pengguna lain. Kondisi ini menjadikan Twitter sebagai sumber data yang sangat kaya untuk memahami persepsi konsumen terhadap suatu brand atau perusahaan[1].

Dalam konteks industri kopi di Indonesia, Fore Coffee merupakan salah satu brand kopi lokal yang mengalami perkembangan pesat dan menjadi perbincangan luas di media sosial. Keberadaan Fore Coffee di Twitter menimbulkan beragam reaksi dari para pelanggan, mulai dari ulasan positif mengenai cita rasa kopi dan suasana gerai, hingga kritik terkait pelayanan atau sistem pemesanan. Jumlah percakapan yang besar ini memberikan peluang bagi perusahaan untuk memperoleh wawasan berharga mengenai tingkat kepuasan dan persepsi pelanggan terhadap produk mereka. Namun demikian, data dari Twitter memiliki sifat yang tidak terstruktur dan sulit untuk dianalisis secara manual karena terdiri dari teks pendek, penggunaan bahasa informal, singkatan, emoji, bahkan sarkasme yang mempersulit proses interpretasi secara konvensional[2].

Untuk mengatasi permasalahan tersebut, diperlukan pendekatan otomatis yang mampu mengekstraksi makna dan emosi dari teks dalam jumlah besar secara efisien. Salah satu pendekatan yang banyak digunakan adalah analisis sentimen, yaitu proses pengolahan teks untuk menentukan kecenderungan emosi dalam suatu pernyataan, apakah bersifat positif, negatif, atau netral[3].

Analisis sentimen telah banyak diterapkan dalam bidang bisnis untuk mengukur kepuasan pelanggan dan membantu perusahaan dalam mengambil keputusan berbasis data. Pendekatan ini umumnya menggunakan algoritma kecerdasan buatan seperti *machine learning* dan *deep learning* untuk membangun model yang dapat mempelajari pola linguistik dari data teks[4].

Penelitian terdahulu yang dilakukan oleh Agung Putra Pamungkas dan rekan-rekan (2025) berjudul “*Comparison of Machine Learning Models for Classifying Consumer Sentiment of Coffee Shops on Social Media X*” menunjukkan bahwa algoritma *Naïve Bayes*, *Logistic Regression*, dan *Support Vector Machine* mampu mengklasifikasikan sentimen pelanggan terhadap coffee shop di Twitter dengan akurasi *Naïve Bayes* sebesar 75 persen, akurasi *Logistic Regression* sebesar 79 persen, dan akurasi *Support Vector Machine* sebesar 78 persen[5]. Penelitian tersebut menginspirasi penulis untuk mengembangkan kajian serupa dalam konteks Indonesia, khususnya terhadap ulasan pelanggan Fore Coffee di Twitter, dengan melakukan perluasan pada penggunaan algoritma dan peningkatan akurasi. Jika penelitian sebelumnya hanya menggunakan *Naïve Bayes*, *Logistic Regression*, dan *Support Vector Machine* penelitian ini memperkenalkan pendekatan yang lebih luas melalui perbandingan antara algoritma *machine learning* klasik (*Naïve Bayes*, *SVM*, dan *Logistic Regression*) dengan algoritma *deep learning* berbasis *transformer*, yaitu *IndoBERTweet* yang dirancang khusus untuk teks berbahasa Indonesia di media social (twitter/X)[6].

Melalui perbandingan ini, penelitian diharapkan dapat menjawab sejauh mana perbedaan performa antara pendekatan klasik dan modern dalam menganalisis sentimen pelanggan. Dengan kata lain, penelitian ini dirancang untuk melakukan sentimen analisis terhadap ulasan dari pelanggan Fore Coffee di Twitter dengan pendekatan komparatif menggunakan algoritma *Naïve Bayes*, *SVM*, *Logistic Regression* dan *IndoBERTweet*. Penelitian ini tidak hanya bertujuan untuk mengukur akurasi dari masing-masing algoritma, tetapi juga untuk memahami kelebihan dan kelemahan setiap pendekatan dalam mengolah teks pendek dan tidak terstruktur seperti tweet. Selain itu, hasil dari penelitian ini diharapkan dapat memberikan kontribusi praktis bagi Fore Coffee dalam

memahami opini pelanggan dengan lebih mendalam, sehingga perusahaan dapat mengambil keputusan yang strategis untuk meningkatkan kualitas produk, pelayanan, dan strategi pemasaran[7].

Untuk sisi akademik, penelitian ini juga memberikan kontribusi terhadap pengembangan studi analisis sentimen berbahasa Indonesia, yang sampai saat ini masih relatif terbatas dibandingkan dengan penelitian dalam bahasa Inggris. Dengan memanfaatkan model *IndoBERTweet*, penelitian ini berupaya menunjukkan potensi penerapan *deep learning* dalam memahami konteks linguistik khas media sosial Indonesia yang sering kali tidak baku. Sementara itu, algoritma klasik seperti *Naïve Bayes*, *SVM*, dan *Logistic Regression* tetap menjadi acuan penting untuk menilai sejauh mana model modern seperti *IndoBERTweet* mampu memberikan peningkatan kinerja yang signifikan[8].

Secara keseluruhan, penelitian ini bertujuan untuk melakukan analisis sentimen terhadap ulasan pelanggan Fore Coffee di Twitter menggunakan empat algoritma berbeda dan membandingkan hasil kinerjanya[9]. Penelitian ini akan berfokus pada analisis sentimen ulasan pelanggan Fore Coffee yang dipublikasikan melalui platform Twitter. Analisis dilakukan menggunakan pendekatan *machine learning* dan *deep learning* untuk membandingkan performa tiga algoritma klasik, yaitu *Naïve Bayes*, *Support Vector Machine*, dan *Logistic Regression*, dengan model *IndoBERTweet* yang dirancang khusus untuk bahasa Indonesia. Pemilihan Fore Coffee sebagai objek penelitian didasarkan pada tingginya interaksi publik di media sosial serta kurangnya studi komprehensif yang mengevaluasi sentimen pelanggan brand kopi lokal menggunakan pendekatan komparatif model klasifikasi teks[10].

Penelitian ini juga diharapkan dapat menjadi acuan bagi penelitian lanjutan dalam bidang analisis teks, *NLP*, serta penggunaan algoritma dalam melakukan analisis sentimen. Hasil akhir dari penelitian ini diharapkan tidak hanya memberikan manfaat bagi dunia akademik, tetapi juga bagi industri, khususnya bagi perusahaan yang ingin memahami persepsi pelanggan melalui data media sosial secara lebih efektif dan berbasis teknologi[11].

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut :

1. Bagaimana cara mengklasifikasikan sentimen dari tweet pelanggan Fore Coffee menjadi kategori positif, negatif, dan netral menggunakan algoritma *Naïve Bayes*, *Logistic Regression*, *Support Vector Machine (SVM)*, dan *IndoBERTweet*?
2. Bagaimana perbandingan kinerja antara algoritma *machine learning* (*Naïve Bayes*, *Logistic Regression*, dan *SVM*) dengan algoritma *deep learning* (*IndoBERTweet*) dalam analisis sentimen ulasan pelanggan Fore Coffee di Twitter?
3. Bagaimana hasil analisis sentimen tersebut dapat dimanfaatkan untuk memberikan wawasan dan rekomendasi strategis bagi Fore Coffee dalam meningkatkan pelayanan dan kepuasan pelanggan?

## 1.3 Batasan Masalah

Agar penelitian ini lebih terarah dan fokus, maka batasan masalah yang digunakan adalah sebagai berikut:

1. Data yang digunakan dalam penelitian ini merupakan tweet berbahasa Indonesia yang mengandung kata kunci seperti “fore coffee”, “fore kopi”, atau kata kunci serupa, serta tidak termasuk tweet dari akun resmi @ForeCoffeeID maupun retweet.
2. Pengumpulan data dilakukan menggunakan teknik web scraping dengan *tweet harvest*, terbatas pada periode tahun 2022 hingga 2025.
3. Penelitian hanya berfokus pada analisis teks dan tidak melibatkan data berupa gambar, video, atau bentuk media lainnya.
4. Algoritma yang digunakan terdiri dari dua kelompok, yaitu algoritma *machine learning* klasik (*Naïve Bayes*, *Logistic Regression*, dan *Support Vector Machine*) dan algoritma *deep learning* berbasis *transformer* (*IndoBERTweet*).
5. Evaluasi model dilakukan menggunakan metrik *accuracy* dan *F1-score (macro)* untuk membandingkan performa antar algoritma secara objektif.

6. Penelitian ini tidak membahas implementasi aplikasi, melainkan berfokus pada analisis performa dan perbandingan algoritma terhadap dataset yang sama.

## 1.4 Tujuan dan Manfaat Penelitian

### 1.4.1 Tujuan Penelitian

1. Melakukan analisis sentimen terhadap ulasan pelanggan Fore Coffee di Twitter menggunakan algoritma *Naïve Bayes*, *SVM*, *Logistic Regression*, dan *IndoBERTweet*.
2. Membandingkan kinerja antara algoritma *machine learning* dan *deep learning* dalam melakukan klasifikasi sentimen teks berbahasa Indonesia.
3. Memberikan rekomendasi dan insight bagi Fore Coffee berdasarkan hasil analisis sentimen pelanggan untuk mendukung pengambilan keputusan yang lebih efektif.

### 1.4.2 Manfaat Penelitian

- 1 Kontribusi terhadap pengembangan ilmu pengetahuan, khususnya dalam bidang *Natural Language Processing (NLP)* dan *machine learning* berbahasa Indonesia, melalui penerapan dan evaluasi algoritma klasifikasi teks dalam konteks data media sosial yang tidak terstruktur.
- 2 Pengayaan literatur ilmiah terkait analisis sentimen pada teks berbahasa Indonesia, yang selama ini masih lebih banyak berfokus pada studi berbahasa Inggris dan konteks global.
- 3 Penyediaan dasar konseptual bagi penelitian lanjutan dalam pengembangan model analisis teks dan perbandingan algoritma *machine learning* dan *deep learning*, sehingga dapat menjadi acuan dalam pengembangan sistem evaluasi opini publik berbasis data di masa mendatang.

## 1.5 Sistematika Penulisan

Sistematika penulisan dalam penelitian ini disusun berdasarkan tahapan dari metode *CRISP-DM (Cross Industry Standard Process for Data Mining)* yang

digunakan sebagai kerangka kerja penelitian. Tahapan tersebut terdiri dari *business understanding, data understanding, data preparation, modeling, evaluation, dan deployment*. Namun, dalam penelitian ini tahap *deployment* tidak dilakukan karena penelitian berfokus pada perbandingan kinerja model dan evaluasi hasil klasifikasi, bukan pada penerapan model ke sistem operasional.

Adapun sistematika penulisan skripsi ini terdiri atas lima bab yang saling berkaitan sebagai berikut:

### 1. Bab I Pendahuluan

Menjelaskan latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan. Selain itu, bab ini juga menggambarkan alasan pemilihan topik dan relevansi penelitian terhadap pengembangan analisis sentimen pada ulasan pelanggan Fore Coffee di Twitter.

### 2. Bab II Landasan Teori

Berisi teori-teori pendukung yang digunakan dalam penelitian, seperti konsep dasar text mining dan *Natural Language Processing* (NLP), algoritma *machine learning* (*Naïve Bayes, Logistic Regression, dan Support Vector Machine*), serta model *deep learning IndoBERTweet*. Bab ini juga menjelaskan metodologi *CRISP-DM* dan penelitian terdahulu yang relevan sebagai acuan dan pembeda terhadap penelitian ini.

### 3. Bab III Metodologi Penelitian

Menjelaskan tahapan penelitian yang dilakukan sesuai metode *CRISP-DM*, dimulai dari pemahaman konteks bisnis, pengumpulan dan pengolahan data, hingga pemodelan dan evaluasi hasil. Bab ini juga memaparkan pendekatan penelitian yang digunakan (kuantitatif), teknik pengumpulan data melalui *Tweet Harvest* dengan *XDeveloper*, serta perangkat dan pustaka pendukung yang digunakan dalam *Google Colab*.

### 4. Bab IV Analisis dan Hasil Penelitian

Membahas hasil dari setiap tahapan dalam *CRISP-DM*, mulai dari *business understanding, data understanding, data preparation, hingga modeling dan evaluation*. Dalam bab ini dijelaskan pula hasil

evaluasi performa empat algoritma yang digunakan (*Naïve Bayes*, *Logistic Regression*, *SVM*, dan *IndoBERTweet*), serta perbandingan hasil antar model berdasarkan metrik akurasi dan *F1-macro*.

## 5. Bab V Simpulan dan Saran

Berisi kesimpulan dari hasil penelitian yang telah dilakukan serta saran untuk pengembangan penelitian berikutnya. Kesimpulan dirumuskan berdasarkan hasil perbandingan kinerja model, sedangkan saran diberikan untuk penelitian lanjutan, seperti pengembangan sistem otomatis analisis sentimen pelanggan Fore Coffee menggunakan model terbaik.