

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Di era digital saat ini, semua sektor kegiatan dapat dijalankan secara *online* salah satunya adalah penyediaan berbagai macam jasa. Jasa yang paling banyak digunakan sehari-hari adalah jasa transportasi *online*. Jasa transportasi *online* selalu mendapat hati di masyarakat karena transportasi merupakan sarana penghubung antar manusia serta alat perpindahan tempat [1] sehingga transportasi merupakan salah satu kebutuhan utama manusia dari dulu hingga sekarang. Permintaan jasa transportasi juga semakin hari bertambah. Hal itu dikarenakan transportasi diperlukan untuk membantu berbagai macam sektor [1]. Kemudahan dalam penggunaannya serta manfaat yang disediakan juga menjadi salah satu daya tarik dari jasa transportasi *online*.

Keadaan tersebut memunculkan banyak perusahaan transportasi *online*. Kemunculan perusahaan tersebut di Indonesia sendiri diawali dengan kemunculan Uber pada tahun 2014, disusul oleh Gojek, Grab dan perusahaan yang lain [2]. Perusahaan jasa transportasi *online* ini umumnya menyediakan beberapa jasa yang sering digunakan yaitu jasa antar jemput penumpang, jasa antar barang cepat, dan jasa antar makanan. Hal tersebut yang menjadikan jasa transportasi *online* digandrungi oleh banyak orang [3].

Di Indonesia, terdapat aplikasi yang merajai jasa transportasi *online* sejak 2018 yaitu Gojek [4]. Gojek adalah *platform* layanan transportasi *online* asal Indonesia yang lahir pada tahun 2010 dan menjadi salah satu perusahaan transportasi *online* terkemuka di dunia. Dengan berbagai fitur dan kemudahannya, Gojek dengan mudah dicintai oleh banyak orang di Indonesia dari berbagai kalangan. Hal ini ditunjukkan dengan jumlah unduhan Gojek yang merajai pasar transportasi Indonesia dibandingkan dengan kompetitornya.



Gambar 1. 1 Statistik Unduhan Transportasi *Online* di Indonesia[5]

Meskipun Gojek masih menempati urutan teratas dalam jumlah unduhan, data menunjukkan penurunan signifikan dalam jumlah unduhan dari tahun 2022 ke 2023, sementara aplikasi transportasi online lainnya mengalami peningkatan. Penurunan ini dapat menjadi indikasi adanya tantangan dalam mempertahankan daya saing di pasar yang kompetitif, yang pada gilirannya dapat memengaruhi keberlanjutan perusahaan. Keberlanjutan tersebut sangat bergantung pada persepsi dan kepuasan pelanggan terhadap layanan yang diberikan. Oleh sebab itu, analisis sentimen menjadi alat penting untuk mengungkap pandangan masyarakat dan menemukan hal-hal yang membentuk pandangan tersebut [6], sehingga dapat memberikan gambaran lebih jelas tentang potensi keberlanjutan Gojek di masa yang akan datang.

Analisis sentimen merupakan metode guna mengidentifikasi serta mengklasifikasikan pandangan atau emosi yang terdapat pada sebuah tulisan. [7], terutama untuk mengidentifikasi pandangan terhadap suatu topik, apakah bersifat positif, negatif, atau netral. Pada konteks *big data*, analisis sentimen

menjadi semakin penting karena besar data yang dihasilkan dan disebarkan melalui internet, media sosial, dan platform digital lainnya sangat besar dan terus bertambah dengan cepat.

*Big data* memungkinkan analisis sentimen untuk dilakukan dengan cakupan yang jauh lebih luas dan mendalam [8]. Dengan menggunakan teknik *machine learning*, analisis sentimen dalam *big data* dapat dilakukan dengan lebih mudah karena memudahkan dalam melakukan polarisasi opini positif dan negatif dengan cepat [9] sehingga pemrosesan analisis sentimen dan prediksi menjadi lebih efektif. Hasil dari analisis sentimen dan prediksi ini kemudian dapat digunakan untuk *knowledge* kepada masyarakat akan apa yang akan terjadi pada suatu perusahaan.

Untuk melakukan analisis sentimen, diperlukan peran *Big Data Analytics* untuk pemrosesannya. Terdapat beberapa algoritma yang dapat dipakai pada pemrosesan analisis sentimen yaitu *Naive Bayes*, *K-Nearest Neighbor*(KNN), *Support Vector Machine* (SVM) [10], serta beberapa algoritma lainnya. Studi ini bertujuan membandingkan tiga algoritma, *Naive Bayes*, KNN, dan SVM. Ketiga algoritma tersebut dipilih karena sering digunakan dalam melakukan analisis sentimen dan dinilai efektif dalam melakukan analisis sentimen. Hal ini divalidasi dengan nilai akurasi yang cukup tinggi pada penelitian sebelumnya.

Algoritma *Naive Bayes*, KNN, SVM dipilih karena memiliki kelebihan masing-masing. Kelebihan algoritma *Naive Bayes* yang menjadi acuan pemilihan algoritma ini adalah cepat dan efisien untuk data dalam jumlah besar, dapat bekerja dengan baik bahkan ketika jumlah data tidak banyak, dan mudah untuk diimplementasikan [11]. Keunggulan dari algoritma *K-Nearest Neighbor*(KNN) yang menjadi acuan pemilihan algoritma ini adalah kemampuannya untuk diterapkan berukuran besar dengan cara yang efisien [12], ketahanannya terhadap data pelatihan yang mengandung banyak noise [13], dan kemampuannya untuk memberikan rekomendasi yang akurat kepada pengguna dengan kualitas yang tinggi [14]. Keunggulan algoritma *Support*

*Vector Machine* (SVM) yang menjadi dasar pemilihannya adalah kemampuannya untuk menemukan *hyperplane* optimal dan menciptakan model klasifikasi yang efisien dan memiliki akurasi tinggi [15].

Algoritma *Naïve Bayes* adalah teknik pengelompokan yang berlandaskan pada teori peluang dan statistika [16]. Algoritma ini memanfaatkan klasifikasi berdasarkan probabilitas untuk memprediksi kelas dari suatu data dengan mempertimbangkan nilai atribut yang dimilikinya. Di sisi lain, KNN merupakan metode non-parametrik yang mengklasifikasikan objek berdasarkan kemiripan fitur dengan data latih. Sementara itu, algoritma SVM salah satu cara pada penambangan data yang dipakai untuk tujuan pengelompokan. SVM beroperasi dengan melakukan transformasi data latih awal ke dimensi yang lebih tinggi melalui pemetaan non-linear [8].

Dalam penelitian ini, data sentimen akan dikumpulkan dari *Review* yang ada di *Google Play Store*. Data ini akan diproses dan dikaji dengan membandingkan antara algoritma *Naïve Bayes*, KNN, dan SVM untuk menentukan sentimen publik terhadap Gojek. Hasil analisis ini akan memberikan wawasan mengenai tren sentimen masyarakat dan aspek-aspek yang berpengaruh terhadap rasa puas atau tidak puas pengguna terhadap layanan Gojek.

Penelitian ini bertujuan menyajikan pemahaman mengenai keberlanjutan Gojek dalam pasar transportasi *online* di Indonesia dengan menggunakan pendekatan analisis sentimen. Objek penelitian yang diangkat adalah sentimen publik serta faktor-faktor yang memengaruhi persepsi terhadap Gojek dari berbagai aspek mulai dari penyediaan aplikasi, emosi pengguna, dan juga faktor dari luar seperti penyediaan promo. Penelitian ini diharapkan mampu memberikan kontribusi dalam menyediakan informasi mengenai keberlanjutan Gojek di pasar transportasi *online* di Indonesia.

*Domain Problem* yang diangkat dalam penelitian ini adalah klasifikasi sentimen secara keseluruhan, tanpa membedah sentimen berdasarkan aspek tertentu. Studi ini memiliki tujuan untuk membuat perbandingan model dari

algoritma *Naïve Bayes*, KNN, dan SVM. Dengan pendekatan *opinion mining*, penelitian ini menggali opini publik dari ulasan, komentar media sosial, dan sumber teks lainnya untuk memahami persepsi masyarakat terhadap Gojek secara menyeluruh.

Penelitian ini juga bertujuan menambah literatur akademik dengan menyediakan model yang efektif dari algoritma *Naïve Bayes*, KNN, dan SVM. Penelitian sebelumnya memiliki nilai akurasi terhadap algoritma *Naïve Bayes* sebesar 97% [19], terhadap algoritma KNN sebesar 79,43% [18], dan terhadap algoritma SVM sebesar 85,54% [6]. Penelitian ini akan menyediakan model yang baik serta mendukung keberlanjutan Gojek di industri transportasi online Indonesia.

## 1.2 Rumusan Masalah

Merujuk pada latar belakang penelitian, berikut merupakan rumusan masalah yang dikaji lebih lanjut.

1. Bagaimana cara melakukan pengumpulan data dari *platform Google Play Store* terkait pengalaman pengguna Gojek?
2. Bagaimana membangun model klasifikasi sentimen menggunakan algoritma *Naïve Bayes*, *K-Nearest Neighbors* (KNN), dan *Support Vector Machine* (SVM) guna memprediksi keberlanjutan Gojek?
3. Seberapa akurat dan efektif model analisis sentimen *Naïve Bayes*, *K-Nearest Neighbors* (KNN), dan *Support Vector Machine* (SVM) dalam mengidentifikasi sentimen positif, negatif, dan netral pada data ulasan pengguna Gojek di *Google Play Store*?
4. Bagaimana keberlanjutan aplikasi Gojek di Indonesia?

## 1.3 Batasan Masalah

Berikut merupakan batasan masalah dari studi ini.

1. Pengumpulan data dilakukan pada periode Januari 2024 - Oktober 2024.

2. Pengumpulan data hanya terbatas pada platform *Google Play Store*.
3. Data yang dikumpulkan hanya terkait dengan opini pengguna Gojek dalam menggunakan layanan yang ditawarkan oleh platform Gojek.
4. Penelitian ini hanya akan melibatkan analisis sentimen pada data berbahasa Indonesia, sehingga bahasa asing tidak akan diikutsertakan dalam analisis.
5. Keberlanjutan aplikasi Gojek di Indonesia dilihat dari aspek komentar positif dan negatif dari pengguna.

## **1.4 Tujuan dan Manfaat Penelitian**

### **1.4.1 Tujuan Penelitian**

Berikut merupakan tujuan dari penelitian ini.

1. Mengumpulkan data sentimen pengguna Gojek yang tersebar di *Google Play Store* terkait dengan pengalaman pengguna dengan cara yang efektif menggunakan pendekatan tepat.
2. Membangun model klasifikasi sentimen menggunakan algoritma *Naïve Bayes*, *K-Nearest Neighbors (KNN)*, dan *Support Vector Machine (SVM)* guna menganalisis sentimen pengguna Gojek yang terkumpul.
3. Mengukur keberhasilan performa algoritma dengan membandingkan dan mengukur akurasi dan efektivitas model klasifikasi sentimen *Naïve Bayes*, *K-Nearest Neighbors (KNN)*, dan *Support Vector Machine (SVM)* dalam mengidentifikasi sentimen positif, negatif, dan netral pada data pengguna Gojek di *Google Play Store*.
4. Melakukan prediksi mengenai keberlanjutan Gojek di pasar transportasi *online* berdasarkan hasil analisis sentimen publik.

#### 1.4.2 Manfaat Penelitian

Terdapat manfaat teoritis dari studi ini yaitu sebagai berikut.

1. Memperluas literatur akademik mengenai penggunaan algoritma *Naïve Bayes*, KNN, dan SVM dalam analisis sentimen, khususnya dalam konteks industri transportasi *online* di Indonesia.
2. Mendorong pengembangan lebih lanjut dalam penggunaan teknologi analitik dan kecerdasan buatan untuk meningkatkan kualitas dan efisiensi analisis sentimen di berbagai sektor industri.

Terdapat manfaat praktis dari studi ini adalah sebagai berikut.

1. Menyediakan pengetahuan kepada publik mengenai perkembangan pasar transportasi *online* di Indonesia.
2. Memberikan gambaran mengenai preferensi serta keperluan pengguna, yang dapat digunakan sebagai dasar untuk membuat strategi bisnis yang lebih efektif.
3. Menyediakan data dan analisis yang dapat digunakan untuk merumuskan kebijakan yang lebih baik dalam mendukung keberlanjutan dan persaingan sehat di pasar transportasi *online*.

#### 1.5 Sistematika Penulisan

Penelitian ini terdiri dari 5 bagian, yaitu:

1. BAB I Pendahuluan  
Bagian ini mencakup latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.
2. BAB II Landasan Teori  
Bagian ini membahas mengenai teori yang menjadi dasar prediksi pada penelitian ini, serta tinjauan pustaka yang meliputi penelitian-penelitian terdahulu mengenai analisis sentimen.
3. BAB III Metodologi Penelitian

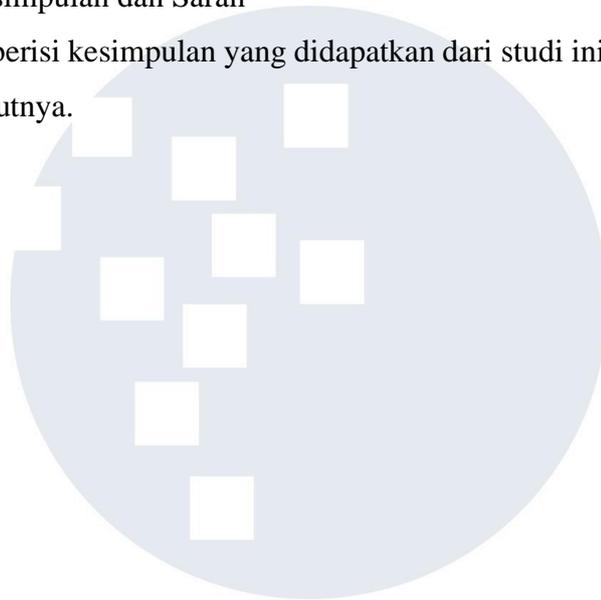
Bagian ini menyediakan ikhtisar mengenai objek penelitian, metode penelitian, variabel-variabel yang dilibatkan, serta teknik pengumpulan data yang diterapkan.

4. BAB IV Analisis dan Hasil Penelitian

Pada bagian ini, dijelaskan hasil implementasi pada studi ini. Hasil output dari penelitian ini akan ditulis dan dianalisis.

5. BAB V Kesimpulan dan Saran

Bagian ini berisi kesimpulan yang didapatkan dari studi ini dan saran untuk studi selanjutnya.



UMMN

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA