

# BAB I

## PENDAHULUAN

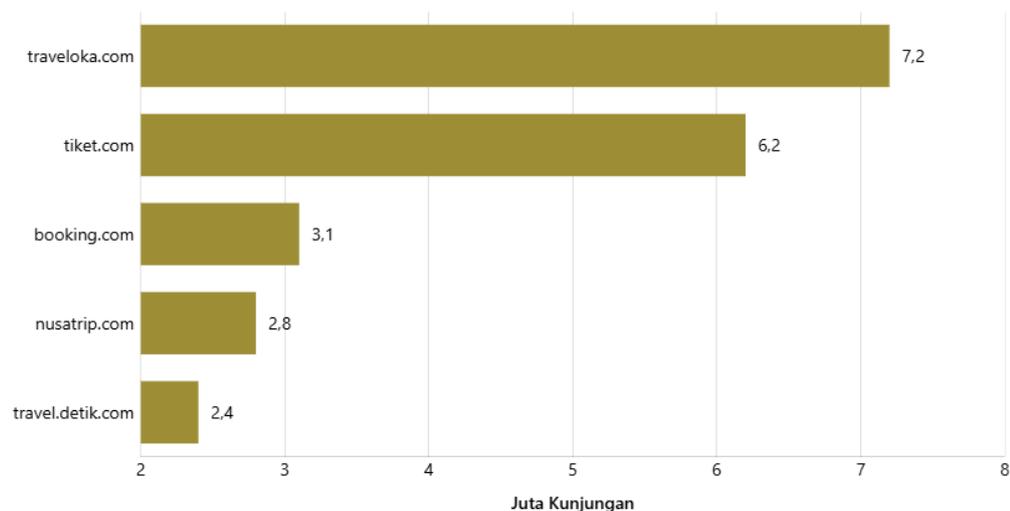
### 1.1 Latar Belakang

Perkembangan teknologi semakin menggeser sistem konvensional, dari yang sebelumnya kita harus melakukan transaksi atau pembelian langsung (*offline*) namun sekarang kita dapat mudah memesannya melalui platform digital online[1]. Hal ini juga berdampak terhadap tren reservasi fasilitas liburan dan berbagai tempat wisata. Keadaan tersebut memunculkan banyaknya *platform* situs perjalanan seperti Traveloka, Tiket.com, Agoda dan banyak yang lainnya.

*E-Commerce* dapat didefinisikan sebagai transaksi komersial yang menggunakan teknologi digital. Media *E-Commerce* melibatkan pengguna internet, aplikasi maupun browser pada setiap perangkat seluler[2]. Platform mobile merupakan pengembangan terbaru melalui jaringan nirkabel (wifi) atau layanan telepon seluler. Perkembangan teknologi memungkinkan para pengguna *e-commerce* untuk memenuhi kebutuhan Masyarakat dalam era digital[3].

Traveloka merupakan *online travel agency* terbesar di Indonesia yang memiliki banyak fitur dan lengkap untuk kebutuhan tiket *travelling*. Traveloka menjadi pilihan utama bagi para *traveller* karena Traveloka memiliki laman yang praktis, aman, terpercaya, cepat dan murah[4]. Traveloka menyediakan berbagai tiket perjalanan, seperti pesawat, kereta api, bus, booking hotel hingga travel yang dapat dipesan secara *online*. Traveloka menawarkan mekanisme perbandingan harga tiket pesawat antar maskapai setiap harinya, sistem pembayaran yang aman, metode pembayaran yang beragam, hingga *call center* yang selalu siap 24 jam, menjadikan Traveloka mampu memimpin pasar penjualan tiket pesawat *online* untuk konsumen Indonesia [5].

Tiket.com, yang juga merupakan pemain besar di industri ini[6]. Tiket.com memiliki fitur unggulan seperti promo tiket, tampilan antarmuka yang *user-friendly*, dan berbagai metode pembayaran yang fleksibel[7]. Sementara itu, Agoda, sebagai platform global, menawarkan berbagai pilihan akomodasi dengan harga kompetitif serta fitur loyalty program yang menarik bagi pengguna setianya[8].



Gambar 1.1 Survey Top Situs Perjalanan di Indonesia[9].

Berdasarkan Gambar 1.1 menunjukkan hasil survey yang dilakukan oleh databoks mengenai situs perjalanan yang paling banyak dikunjungi oleh Masyarakat Indonesia, Traveloka menempati urutan pertama sebagai situs perjalanan yang banyak dikunjungi internet di Indonesia. Total kunjungan situs Traveloka mencapai hingga 7.2 juta per Maret 2022[9].

4	 Traveloka: Book Hotel & Flight Traveloka	4	=
5	 Gojek Driver PT. GoTo Gojek Tokopedia Tbk	5	=
6	 Access by KAI PT. Kereta Api Indonesia	6	=
7	 tiket.com - Hotels and Flights Tiket.com	7	=
8	 ShopeeFood Driver Shopee	8	=
9	 Agoda: Cheap Flights & Hotels agoda.com	9	=

Gambar 1.2 Top Apps Ranking[10]

Berdasarkan Gambar 1.2 top apps ranking dari similarweb yang berhubungan dengan *online travel agency*, Traveloka menduduki peringkat pertama, Tiket.com menempati peringkat kedua dan juga Agoda menempati peringkat ketiga[10].

Dalam hal yang berkaitan dengan evaluasi layanan, untuk memperbaiki dan meningkatkan kualitasnya, Aplikasi OTA tersebut tentu memerlukan *review* dan ulasan dari penggunanya. *Review* atau ulasan yang dapat diperoleh melalui rating dalam aplikasi Playstore. Lalu, *review* atau ulasan tersebut nantinya akan dipahami oleh Traveloka, Tiket.com dan juga Agoda dengan menggunakan analisis sentiment yang nantinya dapat dengan mudah mengetahui mana komentar positif, dan negatif terhadap penggunaan aplikasinya.

Analisis sentiment merupakan proses menganalisis teks digital untuk menentukan apakah nada emosional pesan tersebut positif, negative, atau netral [11]. Analisis sentiment dapat membantu pemilik aplikasi untuk mengetahui secara detail mengenai penggunaan aplikasi maupun layanannya. Dengan memahami analisis penggunaannya, pemilik aplikasi dapat memperbaiki dan meningkatkan kualitas penggunaan aplikasinya serta layanannya sehingga nanti dapat banyak memberikan respon positif dari penggunanya.

Untuk melakukan analisis sentiment, diperlukan peran *Big Data* didalamnya. Terdapat lima algoritma yang dipakai terhadap proses analisis sentimen pada penelitian ini, yaitu Naïve Bayes, KNN, Decision Tree, Random Forest dan SVM. Penggunaan algoritma Naïve Bayes merupakan salah satu metode klasifikasi yang menggunakan probabilitas dan Teorema Bayes untuk memprediksi kelas dari suatu data berdasarkan atribut atributnya. Algoritma ini dapat digunakan untuk melakukan klasifikasi teks dalam konteks analisis sentiment[12]. Penggunaan algoritma K-Nearest Neighbor (KNN) disini bertujuan untuk mengklasifikasikan opini negative dan opini positif dan perhitungan nilai akurasi dan algoritma K-Nearest Neighbors (KNN) juga memiliki keuntungan dalam fleksibilitas menangani data teks[13].

Penelitian ini juga menggunakan menggunakan algoritma Decision Tree, Random Forest, dan Support Vector Machine (SVM) dalam analisis sentimen. Decision Tree adalah algoritma berbasis pohon keputusan yang membagi data berdasarkan aturan tertentu hingga mencapai hasil klasifikasi. Random Forest, sebagai pengembangan dari Decision Tree, menggunakan metode ensemble

learning dengan menggabungkan beberapa pohon keputusan untuk meningkatkan akurasi dan mengurangi overfitting[14]. Sementara itu, SVM bekerja dengan mencari untuk memisahkan kelas data secara maksimal, sehingga efektif dalam menangani data *hyperplane optimal* teks berdimensi tinggi[15]. Penggunaan kelima algoritma ini bertujuan untuk membandingkan performa masing-masing dalam analisis sentimen terhadap aplikasi Traveloka, Tiket.com, dan Agoda di Indonesia guna menentukan model yang paling akurat dalam mengklasifikasikan opini pengguna.

Berdasarkan penelitian terdahulu yang menerapkan algoritma Naïve Bayes dan *K-Nearest Neighbor* (KNN) untuk analisis sentiment, Analisis Sentimen Pada Ulasan pengguna Aplikasi Bibit dan Bareksa Dengan Algoritma KNN menunjukkan akurasi yang didapat sebesar 85,14% dan 81,70% [16]. Pada Sentimen Analisis Terhadap Aplikasi pada Google Playstore Menggunakan Algoritma Naive Bayes dan Algoritma Genetika, menunjukkan bahwa akurasi Naïve bayes yang didapat sebesar 96,53%, 95,54%, 96,87%, dan 96,54% [17]. Pada Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest menunjukkan bahwa akurasi Random Forest yang didapat sebesar 84%[18]. Pada Analisis Sentimen Aplikasi Tiket Online di Play Store Menggunakan Metode SVM menunjukkan bahwa akurasi yang didapat sebesar 78.21%[19] . Lalu pada Analisis Sentimen Pada Rating Aplikasi Shopee Menggunakan Metode Decision Tree Berbasis SMOTE menunjukkan bahwa akurasi Decision Tree sebesar 99.89%[20] .Hasil dari penelitian terdahulu menjelaskan bahwa supervised learning merupakan algoritma yang cukup baik untuk analisis sentiment, karena masih menunjukkan tingkat akurasi yang sangat tinggi.

Oleh karena itu, penelitian ini tidak hanya membandingkan performa lima algoritma klasifikasi dalam menganalisis sentimen ulasan pengguna terhadap aplikasi OTA (Traveloka, Tiket.com, dan Agoda), tetapi juga mengeksplorasi pengaruh tuning hyperparameter dan teknik SMOTE dalam meningkatkan akurasi klasifikasi. Pendekatan ini diharapkan dapat memberikan kontribusi dalam pengembangan model analisis sentimen yang lebih optimal serta menjadi sumber

informasi yang berguna bagi pengembang aplikasi OTA dalam memahami persepsi pengguna dan meningkatkan kualitas layanan mereka.

### **1.2 Rumusan Masalah**

Berdasarkan latar belakang , rumusan masalah dari penelitian yaitu:

1. Bagaimana perbandingan akurasi baseline dari algoritma Naive Bayes, K-Nearest Neighbor (KNN), Decision Tree, Random Forest, dan Support Vector Machine (SVM) dalam menganalisis sentimen pengguna aplikasi OTA (Traveloka, Tiket.com, dan Agoda)?
2. Bagaimana pengaruh tuning hyperparameter dan penanganan data imbalance (menggunakan SMOTE) terhadap akurasi model klasifikasi dalam analisis sentimen?
3. Aplikasi OTA (Traveloka, Tiket.com, atau Agoda) manakah yang memiliki sentimen pengguna paling positif berdasarkan hasil analisis sentimen?

### **1.3 Batasan Masalah**

Berikut merupakan Batasan masalah yang diterapkan dalam penelitian ini:

1. Proses pengumpulan data dilakukan melalui web scraping pada periode Januari hingga Februari 2025. Data yang diperoleh berjumlah 1.000 ulasan, yang berasal dari komentar pengguna aplikasi OTA di Google Play Store, dengan rentang waktu ulasan mulai dari Desember 2018 hingga Februari 2025.
2. Data dikumpulkan melalui teknik web scraping dari Google Play Store, mencakup username, waktu (at), rating dan isi ulasan.
3. Hanya ulasan yang mengandung teks yang digunakan dalam penelitian ini.
4. Penelitian ini hanya berfokus pada analisis sentimen terkait kepuasan penggunaan aplikasi Traveloka, Tiket.com dan Agoda
5. Studi ini hanya mencakup pengguna di Indonesia.

## 1.4 Tujuan dan Manfaat Penelitian

### 1.4.1 Tujuan Penelitian

1. Membandingkan akurasi baseline dari lima algoritma klasifikasi (Naive Bayes, KNN, Decision Tree, Random Forest, dan SVM) dalam analisis sentimen pada ulasan pengguna aplikasi OTA (Traveloka, Tiket.com, dan Agoda).
2. Menganalisis dampak penerapan hyperparameter tuning dan teknik penanganan ketidakseimbangan data menggunakan SMOTE untuk mengurangi bias terhadap kelas mayoritas dan meningkatkan akurasi serta metrik evaluasi lainnya (presisi, recall, F1-score) dari masing-masing algoritma klasifikasi.
3. Mengetahui aplikasi OTA yang memiliki sentimen pengguna paling positif, berdasarkan hasil analisis sentimen yang diperoleh dari ulasan pengguna di masing-masing aplikasi..

### 1.4.2 Manfaat Penelitian

Berikut ini merupakan manfaat dari penelitian yang berjudul “Sentimen Analisis Kepuasan Pengguna Aplikasi *Travel Agency* Menggunakan Algoritma *Machine Learning* dan Optimasi *Hyperparameter Tuning*”:

1. Memberikan wawasan mengenai judul baik dalam ruang lingkup Universitas Multimedia Nusantara ataupun masyarakat umum.
2. Dapat dijadikan sebagai acuan dalam penelitian sejenis di masa yang akan datang.

## 1.5 Sistematika Penulisan

Penelitian ini terdiri dari 5 bagian, yaitu:

### 1. BAB I Pendahuluan

Pada bab satu pendahuluan yang ingin dilakukan berisikan tentang latar belakang, rumusan masalah yang berisi beberapa pertanyaan permasalahan yang ingin dilakukan, Batasan masalah, tujuan dari penelitian, manfaat penelitian dan yang terakhir terdapat sistematika penulisan

### 2. BAB II Landasan Teori

Pada bab dua landasan teori yang ingin dilakukan berisikan teori teori tentang penelitian yang perlu dilakukan, selain itu, landasan teori yang ingin dipakai yaitu *Sentiment analysis, teorema bayes, K-Nearest Neighbor, Decision Tree, SVM, Random Forest dan Naïve Bayes*. Selain berisikan teori teori yang akan dijelaskan, bab dua juga akan berisikan penelitian terdahulu yang diperoleh dari jurnal jurnal sebagai tambahan.

### 3. BAB III Metodologi Penelitian

Pada bab tiga metodologi penelitian ini meliputi penejelasan mengenai objek penelitian secara keseluruhan dan penjelasan mengenai alur penelitian yang akan digunakan untuk mengumpulkan data, mengolah data dan mengevaluasi data yang diolah dengan menggunakan metode yang dilakukan dalam penelitian penulis.

### 4. BAB IV Analisis dan Hasil Penelitian

Pada bab empat analisis dan hasil penelitian ini berisikan tentang proses dan hasil penelitian sesuai dengan alur *framework* yang digunakan.

### 5. BAB V Simpulan dan Saran

Pada bab lima simpulan dan saran yang ingin dilakukan berisikan mengenai ringkasan dari penjelasan yang sudah dilakukan dari bab-bab sebelumnya, serta berisikan saran dari penulis kepada pembaca.