

# **BAB I**

## **PENDAHULUAN**

### **1.1. Latar Belakang**

Transportasi memiliki peran penting dalam mendorong pertumbuhan ekonomi masyarakat dan menjadi salah satu pilar utama dalam usaha pembangunan ekonomi suatu negara. Industri penerbangan domestik adalah salah satu sektor yang tumbuh pesat. Setelah terdampak Covid-19, industri penerbangan di Indonesia kini memasuki fase pemulihan. Diprediksi bahwa Bisnis penerbangan akan secara bertahap pulih dan mengalami peningkatan permintaan yang signifikan (PURWANTO, 2023). Indonesia merupakan pasar terbesar dengan pertumbuhan tertinggi dalam industri penerbangan di ASEAN. Pada 2030, pasarnya diperkirakan 390 juta penumpang dari dan di dalam Indonesia (PUSPARISA, 2023). Menurut data dari Badan Pusat Statistik Provinsi Sulawesi Selatan, Total penumpang yang datang, berangkat, dan transit penerbangan domestik meningkat sebesar 3,56%, dari 832.971 orang pada November 2023 menjadi 862.584 orang pada Desember 2023. Dibandingkan Desember 2022 (871.301 orang), turun 1,00%. Namun sepanjang tahun 2023, jumlah penumpang domestik meningkat sebesar 15,45%. Karena kecepatannya yang lebih tinggi, kenyamanan fasilitas yang disediakan di dalam pesawat, kemampuannya untuk mencapai destinasi sulit dijangkau, efisiensi waktu banyak masyarakat yang memilih lebih menggunakan pesawat dibandingkan moda transportasi lain.

Banyaknya keunggulan yang ditawarkan dari transportasi pesawat tetapi masih ada persoalan di mana harga tiket pesawat domestik menjadi salah satu faktor krusial yang memengaruhi keputusan pembelian konsumen. Dengan meningkatnya jumlah penumpang pesawat domestik mengakibatkan opini yang bervariasi terkait harga tiket pesawat domestik. Terdapat beberapa pihak yang pro dan kontra. Namun, persepsi dan respon masyarakat terhadap harga tiket pesawat domestik tidak selalu mudah dipahami. Analisis sentimen, yang memanfaatkan teknologi pengolahan,

menjadi pendekatan yang tepat untuk menggali dan memahami opini publik secara luas terkait dengan harga tiket pesawat. Tujuannya adalah untuk memberikan informasi kepada pemerintah dan pemangku kepentingan agar dapat mengambil keputusan yang lebih bijaksana dengan memahami opini positif dan negatif.

Berdasarkan laporan *We Are Social* Januari 2022, jumlah pengguna aktif media sosial di Indonesia meningkat sebesar 12,35% lebih tinggi dibandingkan dengan tahun sebelumnya, mencapai total 191 juta orang. Platform media sosial yang banyak digunakan di Indonesia meliputi Whatsapp, Instagram, Facebook, Tiktok, Twitter, dan Telegram. Mengingat banyaknya pengguna media sosial di Indonesia saat ini, hal tersebut dapat digunakan untuk menganalisis opini masyarakat terkait dengan topik atau isu yang sedang hangat di masyarakat (Ketut & Arnawa, 2022.). Saat ini, topik yang banyak diperdebatkan di jejaring sosial adalah “Harga Tiket Pesawat Domestik”. Analisis sentimen adalah teknik yang digunakan untuk mengenali teks yang mengungkapkan pendapat, pandangan, pendapat, penilaian, dan emosi individu seseorang terhadap suatu objek (Pasek *et al.*, 2022). Tujuan utama dari analisis sentimen adalah untuk memprediksi apakah opini yang diungkapkan dalam suatu bahasa memiliki polaritas positif atau negatif (Alfi *et al.*, 2021). Media sosial Twitter atau X memfasilitasi penyampaian opini publik di ruang digital. Tweet yang tersedia secara publik menunjukkan mereka dapat menunjukkan reaksi dan keterlibatan terhadap sebuah tweet melalui *retweet*, *like*, *mention*, atau balasan terhadap tweet pengguna lain (Karami *et al.*, 2020). Ada banyak algoritma yang dapat membantu dalam proses klasifikasi dalam analisis sentimen, termasuk *Naïve Bayes Classifier*, *Support Vector Machine*, *Decision Tree*, *Random Forest*, *K-Nearest Neighbor*, *Rule-based Classification*, *K-Means*, dan masih banyak lagi.

Beberapa studi sebelumnya juga telah menjadikan analisis sentimen sebagai topik penelitian. Apriani dan Gustian telah melakukan penelitian pada 2019 mengenai analisis sentimen terhadap komentar aplikasi

Tokopedia menggunakan Naïve Bayes. Mereka menggunakan dataset dari komentar pengguna di Google Playstore, mengklasifikasikannya dengan algoritma tersebut, dan mengukur akurasinya di Rapidminer. Data dikumpulkan dalam dua tahap: tahap pertama untuk data latih/training dan tahap kedua untuk data uji. Hasilnya menunjukkan tingkat akurasi yang tinggi, mencapai 97,13% dengan nilai *precision* sebesar 1 dan *recall* sebesar 95,49%. AUC juga mencapai 0,980 (Apriani *et al.*, 2019).

Kemudian penelitian yang telaah dilakukan Evita Fitri dan rekannya mengenai Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma *Naive Bayes*, *Random Forest* Dan *Support Vector Machine*, penelitian tersebut menggunakan metode klasifikasi *Naive Bayes*, *Random Forest*, dan *Support Vector Machine*. Tahapan analisis sentimen meliputi pengumpulan dataset, *preprocessing*, dan penerapan model klasifikasi untuk menentukan apakah sentimen tersebut termasuk dalam kategori positif atau negatif. Penelitian juga mengevaluasi tingkat akurasi dan AUC (Area Under Curve) dari setiap model klasifikasi yang digunakan. Dataset yang diambil dari *Playstore* terdiri dari 1629 teks dengan tengang waktu dari 01 Maret sampai 30 Maret 2020. Karena ketidak seimbangan kelas, metode *SMOTE* diterapkan, menghasilkan model klasifikasi *Random Forest* dengan akurasi 97,16% dan *Area Under Curve* 0,996. Model *Support Vector Machine* mencapai akurasi 96,01% dan AUC 0,543, sementara model *Naive Bayes* mencapai akurasi 94,16% dengan *Area Under Curve* 0,999. Ditemukan peningkatan akurasi sebesar 7,16% dari penelitian sebelumnya, dengan model *Random Forest* sebagai yang terbaik. (Fitri *et al.*, 2020). Selain itu, untuk tugas penelitian akhir ini peneliti akan menggunakan beberapa algoritma klasifikasi. Algoritma yang akan digunakan pada penelitian ini adalah *Naive Bayes* dan *Random Forest*. Kesimpulannya, *Naive Bayes* digunakan dalam penelitian ini karena mampu mengolah data dalam jumlah besar dengan akurasi yang tinggi. *Random forest* merupakan teknik pembelajaran ansambel yang menggunakan pohon keputusan sebagai pengklasifikasi dasar yang akan dibangun dan digabungkan. Beberapa aspek penting dari metode *Random Forest* antara lain penggunaan *bootstrap*

*sampling* untuk membuat pohon prediksi, setiap pohon keputusan melakukan prediksi menggunakan prediktor acak, dan *Convolutional Random Forests* hasil prediksi dari seluruh pohon keputusan dengan suara terbanyak untuk klasifikasi serta rata-rata untuk regresi (Primajaya & Sari, 2018). Algoritma ini dipilih karena kemampuannya untuk mempertimbangkan semua kemungkinan dan menghasilkan kesimpulan yang akurat. Selanjutnya, digunakan metode TF-IDF untuk pembobotan kata.

Terdapat banyak cara untuk mengkombinasikan algoritma klasifikasi saat melakukan analisis sentimen. Dalam penelitian ini menggunakan implementasi metode *ensemble majority vote* sebagai penggabungan algoritma klasifikasi. Karena itu, penelitian ini difokuskan pada pembangunan model klasifikasi yang paling efektif dengan mengintegrasikan algoritma *Naïve Bayes* dan *Random Forest*. Kemudian, tingkat akurasi yang didapat dari implementasi metode *ensemble majority vote* akan di bandingkan dengan hasil akurasi dari masing-masing algoritma itu sendiri. Dari penelitian yang dilakukan Mochammad Alfi Rizky Reynaldhi dan Yuliant Sibaroni pada 2021 mendapat hasil penelitian tersebut menunjukkan bahwa algoritma klasifikasi Hybrid dengan implementasi metode *ensemble vote* dapat membantu dalam mengklasifikasikan ulasan film dan akurasi yang didapat optimal menggunakan metode hybrid (Alfi *et al.*, 2021). Dalam penelitian ini penulis berencana akan menggunakan implementasi metode *ensemble majority vote* dalam pengotimalkan algoritma klasifikas yang digunakan dalam analisis sentimen harga tiket pesawat domestik pada media sosial X atau *twitter*.

Berdasarkan latar belakang tersebut, penelitian ini akan mengusulkan judul sebagai berikut “IMPLEMENTASI METODE ENSEMBLE MAJORITY VOTE PADA ALGORITMA NAIVE BAYES DAN RANDOM FOREST UNTUK ANALISIS SENTIMEN TWITTER HARGA TIKET PESAWAT DOMESTIK”

## **1.2. Rumusan Masalah**

Dengan mengacu pada penjelasan latar belakang tersebut, permasalahan dapat dirumuskan sebagai berikut:

1. Bagaimana arah opini masyarakat terkait harga tiket pesawat domestik.
2. Bagaimana perbandingan tingkat akurasi antara algoritma *Naive Bayes*, *Random Forest* secara individu, dan setelah menggunakan implementasi metode *ensemble majority vote* pada algoritma dalam analisis sentimen terhadap harga tiket pesawat domestik pada media sosial X?
3. Apakah efektif mengimplementasi metode *ensemble majority vote* dalam meningkatkan akurasi analisis sentimen terhadap harga tiket pesawat domestik pada media sosial X?

### **1.3. Tujuan**

Berdasarkan perumusan masalah yang telah disusun, tujuan dari penelitian ini adalah:

1. Menganalisis arah opini masyarakat terkait harga tiket pesawat domestik.
2. Membandingkan tingkat akurasi antara algoritma *Naive Bayes*, *Random Forest* secara individu, dan setelah menggunakan implementasi metode *ensemble majority vote* pada algoritma dalam analisis sentimen terhadap harga tiket pesawat domestik pada media sosial X.
3. Menilai efektivitas implementasi metode *ensemble majority vote* dalam meningkatkan akurasi analisis sentimen terhadap harga tiket pesawat domestik pada media sosial X.

### **1.4. Manfaat**

Penelitian ini diharapkan akan memberikan manfaat sebagai berikut:

1. Bagi Peneliti, manfaat yang didapatkan adalah tambahan ilmu baru mengenai algoritma klasifikasi dengan metode pengotimasiannya.
2. Bagi Pembaca, diharapkan dapat memahami bagaimana cara

mengimplementasikan masing-masing metode yang digunakan, memberikan panduan bagi pembaca yang tertarik pada analisis sentimen khususnya topik data mining.

3. Bagi Pemerintah dan jajarannya, Pemerintah dapat menggunakan hasil penelitian ini sebagai dasar untuk merumuskan kebijakan yang berkaitan dengan industri penerbangan domestik, terutama dalam hal melindungi konsumen dan meningkatkan kualitas layanan.

### **1.5. Batasan Masalah**

Adapun batas-batas masalah dalam penelitian ditetapkan untuk memastikan agar penelitian tetap berpegang pada tujuan yang sebenarnya:

1. Penelitian hanya memfokuskan pada analisis sentimen terhadap harga tiket pesawat domestik.
2. Klasifikasi sentimen tweet dilakukan pada 2 kelas yaitu Positif dan Negatif.
3. Algoritma yang digunakan dalam analisis sentimen adalah *Naive Bayes*, *Random Forest*, dan metode *ensemble majority vote*.
4. Hasil Penelitian adalah perbandingan tingkat akurasi antara algoritma *Naive Bayes*, *Random Forest*, dan implementasi metode *ensemble majority vote* pada *Naive Bayes-Random Forest* dalam analisis sentimen terhadap harga tiket pesawat domestik pada media sosial X.