

BAB I

PENDAHULUAN

1.1. Latar Belakang

Dalam upaya untuk menjawab kebutuhan masyarakat akan layanan transportasi yang efisien dan berkualitas, khususnya dalam sektor transportasi kereta api wilayah daerah operasional 8 Surabaya, perlu dilakukan langkah-langkah yang mampu memenuhi harapan pengguna. Salah satu bentuk layanan yang memiliki peranan vital dalam mendukung mobilitas masyarakat adalah transportasi kereta api bersubsidi. Menurut peraturan menteri perhubungan (RI, 2010) tentang penyelenggaraan kewajiban pelayanan publik angkutan orang dengan kereta api kelas ekonomi tahun anggaran 2010, Program ini merupakan bagian dari *Public Service Obligation* (PSO) yang ditetapkan pemerintah untuk meningkatkan kualitas layanan dengan membayar selisih tarif antara penetapan pemerintah dan tarif yang dihitung oleh PT KAI. Hal ini seiring dengan pemberlakuan grafik perjalanan kereta api terbaru yang mulai dilaksanakan pada 1 Juni 2023 menjadikan banyak problematika di sarana transportasi kereta api khususnya kereta api lokal Daop 8. Untuk mendukung hal ini, maka perlu dilakukan sebuah analisis sentimen pengguna terhadap layanan kereta api subsidi dalam aturan terbaru tersebut. Analisis sentimen menjadi relevan karena dapat memberikan wawasan yang berharga mengenai bagaimana pengguna sebenarnya merasakan dan menanggapi kualitas layanan yang diberikan (Estika dkk, 2021). Penelitian ini akan fokus pada evaluasi kualitas layanan Transportasi Kereta Api Subsidi melalui pengumpulan data yang melibatkan tanggapan dan pandangan langsung dari para pengguna. Faktor-faktor kritis seperti kebersihan fasilitas, ketepatan waktu kedatangan dan keberangkatan, kenyamanan selama perjalanan, serta interaksi dengan layanan pelanggan akan menjadi pokok analisis. Pengumpulan data akan mencakup aspek-aspek positif, negatif, maupun pandangan netral dalam merespons pengalaman pengguna.

Analisis sentimen adalah suatu disiplin ilmu yang berfokus pada metode analisis opini, pandangan, dan penilaian dari suatu individu terhadap suatu objek (Husada & Paramita, 2021). Oleh karena itu dibutuhkan suatu metode yang dapat

mengolah data tersebut dengan cara cepat dan otomatis untuk mengategorikan ulasan-ulasan tersebut baik ulasan positif, negatif, maupun netral (Estika dkk, 2021). Analisis sentimen merupakan teknik pemrosesan bahasa alami (NLP) yang digunakan untuk mendeteksi dan mengukur emosi dalam teks. Analisis sentimen dapat dilakukan secara otomatis oleh mesin untuk memperoleh hasil prediksi peluang terkait hasil analisis. Analisis sentimen banyak digunakan untuk mengolah teks/komentar, salah satunya adalah untuk menganalisis *review* moda transportasi di Indonesia.

Ulasan kepuasan pelanggan menjadi kriteria utama yang dapat menjadi acuan dalam menilai sebuah fasilitas umum, karena pelanggan merupakan pihak yang memiliki peran secara langsung terhadap moda transportasi khususnya kereta api. Selain itu, ulasan juga dapat digunakan sebagai tolak ukur kualitas dari pengguna sebelumnya yang menggunakan moda transportasi ini. Menurut penelitian yang dilakukan oleh (Chen, 2020) ulasan online memiliki dampak signifikan terhadap persepsi dan keputusan konsumen dalam membeli suatu produk. Untuk membantu proses analisis sebuah sentimen, terdapat metode yang digunakan yakni *naïve bayes classifier* dengan salah satu metode didalamnya yakni *multinomial naïve bayes*.

Metode *naïve bayes* merupakan algoritma yang mampu menerima input dalam bentuk apapun dan memiliki kecepatan dalam memproses suatu data baik itu berupa teks ataupun berita (Handoko dkk, 2021). Model pengklasifikasian *naïve bayes* dikenal sebagai model yang paling sederhana dan umum digunakan dengan tingkat akurasi yang tinggi. (Yulita, 2021). Algoritma *naïve bayes* memiliki beberapa model, diantaranya yakni *Gaussian Naïve Bayes*, *Bernoulli Naïve Bayes*, dan *Multinomial Naïve Bayes*. Dari ketiga jenis algoritma *naïve bayes* diatas, dapat disimpulkan bahwasannya algoritma ini memiliki kemampuan yang baik dalam pengolahan data khususnya dalam bentuk pengolahan teks atau *Natural Language Processing* (NLP).

Dalam penelitian terdahulu yang dilakukan oleh (Fahrudin dkk, 2022) tentang analisis sentimen berbasis algoritma *Naïve Bayes* terhadap topik yang diangkat yakni Pemodelan Teks Tweet pada Isu Pelecehan Seksual Berbasis Analisis Sentimen dan Leksikon Emosi, menunjukkan bahwa algoritma *Naïve Bayes* memiliki akurasi sebesar 74% dengan persentase proporsi sentimen 72% tweet

negatif, 18.8% tweet positif, dan 9.2% tweet netral. Selain itu, pada penelitian ini menggunakan *NRC Emotion Lexicon* sebagai analisis emosi apa saja yang terdapat pada topik yang dianalisis. Dari hasil penelitian terdahulu tersebut, akurasi algoritma *Naïve Bayes* masih tergolong mendapatkan akurasi yang rendah. Maka untuk meningkatkan akurasinya diperlukan sebuah solusi untuk meningkatkan nilai akurasi tersebut sehingga diharapkan memiliki tingkat akurasi yang mendekati, yaitu dengan menggunakan sebuah metode seleksi fitur.

Untuk dapat melakukan sebuah klasifikasi dan meningkatkan akurasi, diperlukan sebuah metode pemrosesan seleksi fitur. Yakni metode *Principal Component Analysis* yang mana pada pemrosesan ini akan dipilih beberapa dimensi ataupun fitur penting dari kumpulan data yang telah melalui proses pembobotan sebelumnya. Dari sini metode *Principal Component Analysis* digunakan untuk mereduksi beberapa fitur agar lebih efisien ketika digunakan untuk pembangunan model klasifikasi (Naibaho, 2021).

Dari penjabaran latar belakang pada penelitian dan mempertimbangkan dari analisa penelitian terdahulu, maka penelitian ini menggunakan *Principal Component Analysis* untuk pemilihan fitur dan untuk klasifikasi menggunakan algoritma naïve bayes jenis *Multinomial Naïve Bayes*. Algoritma Multinomial Naïve Bayes dipilih untuk melakukan klasifikasi dalam analisis sentimen karena efektif dalam mengelola data teks. Metode ini sangat cocok untuk penelitian analisis sentimen karena mampu mempertimbangkan frekuensi kemunculan kata-kata dalam dokumen, yang merupakan aspek kunci dalam analisis sentimen. Sentimen pada penilaian pengguna tersebut nantinya akan dibagi menjadi dua kategori, yaitu positif dan negatif. Adapun tahapan dalam menyelesaikan penelitian ini yakni analisa kebutuhan data, kemudian tahap kedua dilakukan persiapan data, tahap ketiga dilakukan tahap analisis sentimen, tahap keempat dilakukan evaluasi sistem, dan tahap terakhir yakni skenario pengujian untuk data latih dan data uji. Dengan memahami perspektif pengguna, penelitian ini membantu dalam memahami kinerja dan meningkatkan standar layanan. Selain itu, metode yang digunakan seperti *Principal Component Analysis* diharapkan dapat meningkatkan akurasi algoritma klasifikasi sentimen untuk menjadikan hasil analisis lebih dapat akurat. Dengan demikian, penelitian ini perlu dilakukan untuk memberikan

kontribusi dalam meningkatkan kualitas layanan transportasi kereta api bersubsidi dan memenuhi kebutuhan mobilitas masyarakat dengan lebih baik.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan baik itu bersumber dari penelitian terdahulu maupun penelitian yang akan dilakukan, maka rumusan masalah yang dapat diambil adalah sebagai berikut:

- 1) Bagaimana *Principal Component Analysis* dapat digunakan sebagai seleksi fitur untuk meningkatkan nilai akurasi pada *Multinomial Naïve Bayes* dalam sebuah analisis sentimen yang diangkat pada judul ini?
- 2) Bagaimana hasil analisis sentimen yang dilakukan terhadap komentar pengguna kereta api lokal DAOP 8?
- 3) Bagaimana perbedaan nilai akurasi dari metode *Multinomial Naïve Bayes* tanpa optimasi dan menggunakan optimasi *Principal Component Analysis*?
- 4) Bagaimana fitur yang dihasilkan dari proses seleksi fitur menggunakan *Principal Component Analysis*?

1.3. Batasan Masalah

Adapun batasan masalah pada penelitian ini bertujuan untuk menspesifikasikan pembahasan yang akan dilakukan dalam analisis agar tidak keluar dari batasan yang sudah ditetapkan, berikut merupakan batasan masalah dari penelitian “Analisis Sentimen Terhadap Kewajiban Pelayanan Publik Kereta Api Lokal Daop 8 Menggunakan Algoritma Naive Bayes”:

- 1) Data yang diambil berasal dari ulasan pengguna kereta api lokal Daop 8 Surabaya, yang didapatkan melalui survey kepuasan pelanggan
- 2) Data yang digunakan adalah data yang berbahasa Indonesia
- 3) Metode klasifikasi yang digunakan dalam analisis sentimen adalah *Multinomial Naïve Bayes*
- 4) Metode yang digunakan untuk seleksi fitur adalah *Principal Component Analysis*
- 5) Dalam pemilihan kelas digunakan 2 kategori, yakni positif dan negatif
- 6) Pengimplementasian algoritma serta seluruh metode yang digunakan yakni dengan bahasa pemrograman Python

1.4. Tujuan Penelitian

Berdasarkan latar belakang dan rumusan masalah yang sudah disampaikan, maka pada penelitian ini memiliki beberapa tujuan diantaranya:

- 1) Melakukan implementasi *Principal Component Analysis* guna meningkatkan akurasi algoritma *Multinomial Naïve Bayes* dalam konteks analisis sentimen.
- 2) Mengetahui hasil analisis sentimen pengguna kereta api lokal DAOP 8 dalam kategori positif dan negatif.
- 3) Menghitung nilai akurasi dari klasifikasi analisis sentiment menggunakan algoritma *Multinomial Naive Bayes* dan pemilihan fitur menggunakan *Principal Component Analysis*.
- 4) Mengidentifikasi fitur yang dihasilkan dari proses *Principal Component Analysis*

1.5. Manfaat Penelitian

Dalam penelitian yang dilakukan oleh penulis tentang analisis sentimen terhadap kewajiban pelayanan publik kereta api lokal daop 8 diharapkan dapat memberikan manfaat diantaranya:

- 1) Manfaat teoritis
Penelitian ini diharapkan dapat memberikan pemahaman yang lebih mendalam tentang penerapan dari algoritma *Multinomial Naïve Bayes* dengan menggunakan metode *Principal Component Analysis* sebagai seleksi fitur. Hal ini juga dapat meningkatkan pemahaman teoritis yang lebih baik dalam pengembangan pemahaman tentang cara algoritma tersebut digunakan dalam menganalisis sentimen.
- 2) Manfaat praktis
 - a. Bagi penulis penelitian ini diharapkan dapat menambah ilmu dan pemahaman dari proses analisis sentimen menggunakan metode *Multinomial Naïve Bayes* serta dapat memahami luaran hasil yang didapatkan
 - b. Bagi peneliti selanjutnya penelitian ini diharapkan dapat menjadi sumber inspirasi dalam melakukan penelitian atau pengembangan yang akan datang.