

BAB I

PENDAHULUAN

1.1. Latar Belakang

Perkembangan teknologi, khususnya dalam bidang *text mining*, telah membawa dampak yang signifikan dalam berbagai aspek kehidupan sehari-hari. *Text Mining* merupakan teknik yang digunakan untuk mengambil informasi penting dari data berbentuk teks. Tujuan dari *text mining* bisa berupa sumber akademik, situs jejaring sosial, postingan atau komentar tentang berita, umpan balik dari pelanggan, *speech to text*, dan masih banyak lagi (Jung & Lee, 2020). Dalam beberapa tahun terakhir, kemajuan dalam *text mining* telah memberikan kontribusi besar dalam berbagai bidang, seperti pengenalan suara, terjemahan mesin, serta khususnya analisis sentimen.

Analisis sentimen adalah suatu disiplin ilmu yang berfokus pada metode analisis opini, pandangan, dan penilaian dari suatu individu terhadap suatu objek (Husada & Paramita, 2021). Analisis sentimen digunakan untuk mengkategorikan polaritas teks dalam dokumen atau kalimat, dengan tujuan untuk menentukan sentimen positif, negatif, atau netral (Samsir et al., 2021). Analisis sentimen menjadi salah satu bidang dalam *text mining* yang mana bertujuan untuk menganalisis sentimen sebuah objek permasalahan secara otomatis oleh mesin dan memungkinkan untuk memperoleh hasil prediksi setiap peluang terkait hasil analisis. Analisis sentimen banyak digunakan pada kasus pengolahan teks/komentar salah satu kasusnya yaitu tentang analisa pada *review* sebuah hotel.

Review pelanggan menjadi kriteria penting dalam menilai sebuah hotel karena calon pelanggan dapat mengetahui kualitas fasilitas dan pelayanan hotel berdasarkan pengalaman pelanggan sebelumnya. Hotel di kota-kota wisata berusaha meningkatkan fasilitas dan pelayanan guna memperoleh penilaian positif dari pelanggan. Salah satu hotel pada kota destinasi wisata yang memerlukan *review* pelanggan guna meningkatkan fasilitas dan pelayanannya adalah Favehotel Kusumanegara Yogyakarta. Dalam penelitian yang dilakukan oleh (Chen et al., 2020) dijelaskan bahwa ulasan online memiliki pengaruh yang besar terhadap

pandangan dan keputusan seorang konsumen dalam membeli produk. Adapun beberapa metode yang dapat digunakan dalam menganalisis sebuah sentimen dari komentar/*review* sebuah hotel yaitu *Naïve Bayes Classifier* yang mana salah satunya adalah metode *Multinomial Naïve Bayes*.

Naïve Bayes adalah sebuah metode yang melakukan perhitungan probabilitas dan statistik untuk mengklasifikasikan data, baik itu berupa teks ataupun berita. Selain itu, metode ini juga dapat digunakan dalam ilmu medis untuk membuat diagnosis. *Naïve Bayes* mampu mengklasifikasikan data dalam jumlah besar dan menghasilkan hasil dengan tingkat akurasi yang tinggi (Trisna Lestari et al., 2017). Algoritma *Multinomial Naïve Bayes* adalah salah satu teknik pembelajaran probabilitas yang mengandalkan pada teorema *Bayes*, dan sering diterapkan dalam *Natural Language Processing* (NLP). Dalam algoritma ini, prinsip kerjanya didasarkan pada frekuensi kemunculan kata dalam sebuah dokumen yang dikenal dengan istilah *term frequency* (Yuyun et al., 2021). Namun, dalam beberapa kasus penerapannya tak jarang *Multinomial Naïve Bayes* mendapatkan akurasi yang lebih rendah daripada algoritma yang lain. Dalam penelitian yang dilakukan oleh (Puspita & Widodo, 2020) tentang perbandingan metode KNN, *Decision Tree*, dan *Naïve Bayes* untuk kasus analisis sentimen menunjukkan bahwa hasil akurasi *Naïve Bayes* mendapatkan nilai terendah daripada 2 metode lainnya. KNN mendapatkan hasil akurasi sebesar 96,01%. Kemudian, *Decision Tree* mendapatkan hasil akurasi 96,13%. Sedangkan, *Naïve Bayes* hanya mendapatkan nilai akurasi sebesar 89,14%. Mengetahui dimana akurasi apabila menggunakan algoritma *Naïve Bayes* mendapatkan akurasi yang rendah maka diperlukan sebuah solusi untuk meningkatkan nilai akurasi tersebut, yaitu dengan menggunakan sebuah metode optimasi.

Banyak sekali metode optimasi yang dapat diterapkan pada algoritma untuk kasus analisis sentimen, salah satunya yaitu metode *Particle Swarm Optimization* (PSO). *Particle Swarm Optimization* (PSO) merupakan sebuah metode yang dipakai untuk mengoptimalkan sebuah hasil, dengan mengaplikasikan PSO dapat mengontrol pemilihan sebuah subset untuk menghasilkan akurasi yang terbaik (Hayuningtias & Sari, 2019). Pada penelitian yang dilakukan (Hayuningtias & Sari, 2019) tentang analisis sentimen menggunakan *Naïve Bayes* dan *Particle Swarm*

Optimization terhadap tempat wisata TMII menghasilkan akurasi lebih tinggi daripada metode *Naïve Bayes* tanpa optimasi yaitu sebesar 94%. Namun, dalam penelitian tersebut pembuatan model hanya menggunakan 1 skenario pembagian data dan juga PSO hanya digunakan sebagai seleksi fitur saja. Sedangkan, dalam penelitian lain yang dilakukan oleh (Maulana, 2022) tentang analisis sentimen pengguna aplikasi peduli lindungi pada platform twitter dengan menggunakan algoritma CNN dan PSO menunjukkan bahwa PSO berpengaruh dalam meningkatkan akurasi pada kasus analisis sentimen dimana PSO digunakan untuk pencarian parameter terbaik/*hyperparameter tuning* untuk algoritma CNN. Parameter yang dioptimasi yaitu meliputi nilai *filter*, *kernel size*, *pool size*, dan *stride*. Diketahui hasil akurasi dengan menggunakan metode optimasi PSO mendapatkan hasil yang lebih tinggi daripada tanpa metode optimasi yaitu 81%. Sedangkan, untuk hasil akurasi tanpa menggunakan PSO hanya mendapatkan nilai sebesar 77%.

Selain menggunakan metode optimasi, nilai akurasi juga dapat dipengaruhi oleh pembagian data latih dan data uji. Hal ini ditunjukkan pada penelitian yang dilakukan oleh (Putri, 2022) tentang analisis sentimen berdasarkan aspek pada ulasan aplikasi BCA mobile dengan menggunakan *Naïve Bayes Classifier* dilakukan 12 skenario pengujian dengan pembagian data latih dan data uji yang berbeda dengan hasil akurasi yang berbeda pula. Hasil akurasi tertinggi pada penelitian tersebut yaitu mencapai 96% untuk aspek *Efficiency* dengan pembagian data latih 90% untuk data latih dan 10% untuk data uji, sedangkan akurasi terendah didapatkan untuk aspek *Learnability* dan *Satisfaction* dengan pembagian 80% untuk data latih dan 20% untuk data uji.

Oleh karena beberapa latar belakang di atas, maka pada penelitian skripsi ini, penulis berencana untuk melakukan penelitian guna mengetahui pengaruh optimasi metode *Particle Swarm Optimization* (PSO) pada nilai akurasi algoritma *Multinomial Naïve Bayes* untuk kasus analisis sentimen tentang penilaian pelanggan hotel di Favehotel Kusumanegara Yogyakarta. Penelitian ini akan menggunakan 4355 data yang diambil langsung dari pihak Favehotel Kusumanegara Yogyakarta. Sentimen pada penilaian pelanggan tersebut nantinya akan dibagi menjadi dua kategori, yaitu positif dan negatif. *Particle Swarm*

Optimization (PSO) nantinya digunakan untuk seleksi fitur dan juga *hyperparameter tuning* pada algoritma *Multinomial Naïve Bayes* serta terdapat beberapa skenario pengujian untuk pembagian data latih dan data uji.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka perumusan masalah yang dapat diambil adalah sebagai berikut :

1. Apakah *Particle Swarm Optimization* (PSO) dapat diimplementasikan sebagai *hyperparameter tuning* dan seleksi fitur untuk meningkatkan nilai akurasi pada *Multinomial Naïve Bayes* dalam kasus analisis sentimen?
2. Bagaimana pengaruh *Particle Swarm Optimization* (PSO) pada akurasi algoritma *Multinomial Naïve Bayes* untuk kasus analisis sentimen?
3. Apakah jumlah pembagian data latih dan data uji mempengaruhi nilai akurasi pada algoritma *Multinomial Naïve Bayes*?

1.3. Tujuan

Tujuan yang ingin dicapai berdasarkan rumusan masalah yang ada adalah sebagai berikut:

1. Mengimplementasikan algoritma *Particle Swarm Optimization* (PSO) sebagai *hyperparameter tuning* dan seleksi fitur untuk meningkatkan nilai akurasi pada *Multinomial Naïve Bayes* dalam kasus analisis sentimen.
2. Mengetahui pengaruh *Particle Swarm Optimization* (PSO) pada akurasi algoritma *Multinomial Naïve Bayes* untuk analisis sentimen.
3. Mengetahui pengaruh jumlah pembagian data latih dan data uji pada nilai akurasi algoritma *Multinomial Naïve Bayes*.

1.4. Manfaat

Dalam penelitian tentang analisis sentimen penilaian pelanggan hotel pada Favehotel Kusumanegara Yogyakarta diharapkan dapat memberikan manfaat antara lain :

a. Manfaat Bagi Penulis

Manfaat yang didapatkan oleh penulis adalah penulis dapat menambah ilmu yang dapat dipelajari dalam menganalisis sentimen dari suatu komentar/penilaian.

b. Manfaat Bagi Pihak Favehotel Kusumanegara Yogyakarta

Manfaat yang didapatkan oleh pihak hotel adalah dapat menghasilkan model algoritma yang dapat dikembangkan dan digunakan oleh pihak hotel pada website hotel untuk mengetahui penilaian pelanggan tentang pelayanan hotel guna meningkatkan kualitas pelayanan hotel.

c. Manfaat Bagi Pembaca

Manfaat yang didapatkan oleh pembaca adalah dapat memberikan inspirasi dalam melakukan penelitian atau pengembangan selanjutnya.

1.5. Batasan Masalah

Batasan masalah yang terdapat pada penelitian ini antara lain :

1. Penelitian ini menggunakan 4355 data penilaian pelanggan dalam bahasa Indonesia, yang didapatkan langsung oleh pihak Favehotel Kusumanegara Yogyakarta.
2. Metode yang digunakan dalam analisis sentimen adalah *Multinomial Naïve Bayes* dan *Particle Swarm Optimization (PSO)*.
3. Pengimplementasian algoritma menggunakan bahasa pemrograman Python.
4. Pelabelan data menggunakan metode *Vader Lexicon*.
5. Penilaian dibagi menjadi 2 kategori, yaitu positif dan negatif.
6. Luaran dari penelitian ini berupa model algoritma untuk mengolah penilaian/ulasan pelanggan Favehotel Kusumanegara Yogyakarta.