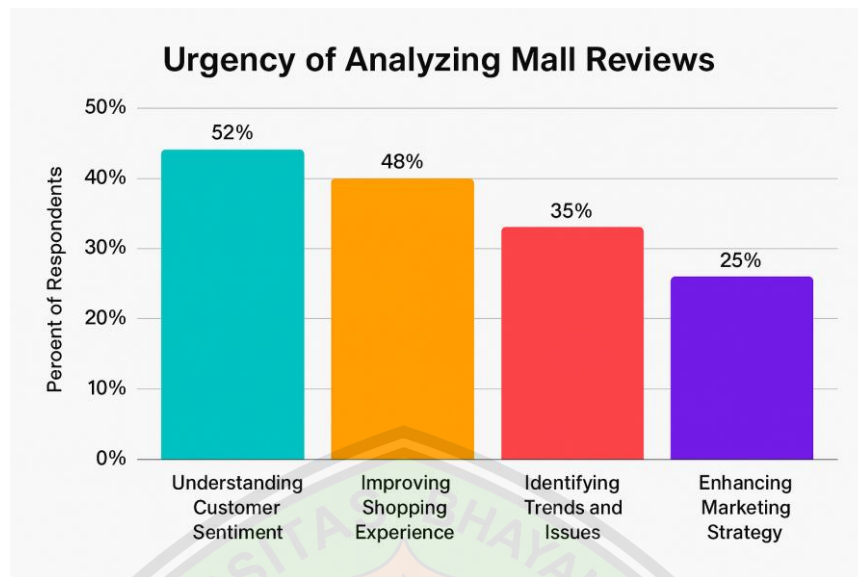


BAB I

PENDAHULUAN

1.1 Latar Belakang

Mall, sebagai destinasi publik, memainkan peran penting sebagai pusat perbelanjaan dan ruang sosial dan rekreasi. Menurut data dari BPS (Badan Pusat Statistik) dan Jakpat pada tahun 2016 dibandingkan dengan tempat wisata lainnya, dari rentang usia 18-40 tahun responden, 97% mengaku setidaknya pernah mengunjungi Mall setidaknya sekali, dengan 60% nya mengaku sering mengunjungi Mall.[1] Banyak pengunjung yang datang ke Mall menceritakan berbagai pengalaman mereka melalui ulasan online, terutama di *platform* seperti Google Maps.[2] Ulasan-ulasan ini memiliki informasi yang sangat penting karena menggambarkan pandangan dan kepuasan pengunjung terhadap layanan, fasilitas, dan suasana Mall tersebut. Oleh karena itu, sangat penting untuk menganalisis ulasan pengunjung Mall secara menyeluruh untuk memahami preferensi dan harapan masyarakat sekaligus untuk mengetahui bagaimana Mall tersebut memenuhi harapan masyarakat.[3]



Gambar 1. 1 Grafik Data Urgensi Analisis Mall

Sumber: Zeng, D., Chen, H., Lusch, R., & Li, S. (2010) [4].

Mall Pakuwon Bekasi resmi dibuka pada 22 November 2024, berlokasi di kawasan superblok 3,6 ha dengan *net leasable area* 45.300 m² dan tingkat okupansi mencapai 95 % saat pembukaan, karena itu dipilih sebagai fokus penelitian karena merupakan salah satu pusat belanja modern yang cepat berkembang di Bekasi.[5] Mall ini menarik perhatian karena menawarkan pengalaman berbelanja yang menyeluruh dan up-to-date, serta memperoleh ribuan ulasan dari pengunjung yang memiliki berbagai sentimen. Banyaknya ulasan yang diterima menjadi tantangan tersendiri, karena akan sulit untuk mengkategorikan dan menganalisisnya secara manual dengan efektif. Oleh karena itu, penting untuk menggunakan metode analisis sentimen otomatis untuk memproses data ulasan tersebut dengan cepat dan akurat [6].

Berbagai algoritma pembelajaran mesin sudah banyak dipakai untuk menganalisis sentimen publik, khususnya pada pemrosesan bahasa alami. *Naïve Bayes* dan *Support Vector Machine (SVM)* merupakan dua algoritma klasifikasi teks

yang populer. Dikenal sebagai algoritma yang sederhana dan efektif untuk mengklasifikasikan teks, *Naïve Bayes* cenderung tidak akurat jika data memiliki korelasi antar fitur. Di sisi lain, mesin pemisah varian (*SVM*) memiliki kemampuan untuk menghasilkan klasifikasi yang lebih akurat karena kemampuan mereka untuk memisahkan data dengan margin yang paling besar, tetapi membutuhkan proses pelatihan yang lebih kompleks dan parameter *tuning* yang diperhatikan [7].

Namun, kedua pendekatan memiliki kesulitan tersendiri saat diterapkan. *Naïve Bayes* sangat bergantung pada asumsi independensi antar fitur, yang seringkali tidak terpenuhi dalam praktiknya, yang dapat menyebabkan keakuratan yang lebih rendah. Namun, pemilihan parameter *Support Vector Machine* yang tepat seperti kernel, regularisasi, dan margin sangat penting. Jika parameter ini tidak dioptimalkan dengan benar, hal itu dapat menyebabkan *overfitting* atau *underfitting* serta memperpanjang waktu pelatihan. Selain itu, karena kebutuhan komputasi yang tinggi, *SVM* tidak efektif untuk dataset yang besar [7].

Meskipun masing-masing algoritma memiliki kelebihan dan kekurangannya sendiri, penting untuk melakukan perbandingan kinerja secara langsung untuk menentukan algoritma mana yang lebih baik dalam hal klasifikasi komentar pengguna di Mall Pakuwon Bekasi. Untuk melakukan perbandingan ini, metrik akurasi akan digunakan, selain menerapkan *preprocessing* teks yang tepat.

Pada penelitian terkait juga pernah dilakukan analisis sentimen pada sebuah pusat perbelanjaan di daerah Bekasi, yaitu Summarecon Mall Bekasi dengan mencapai 60.000 komentar. Untuk mengetahui sentimen dari komentar-komentar yang banyak tersebut perlu menggunakan alat bantu komputasi. Penelitian tersebut

menerapkan *sentiment analysis* menggunakan metode *Naïve Bayes* dan *Support Vector Machine*. Pengambilan data dilakukan dengan teknik *web scrapping* kemudian dilakukan *pre-processing* dan pelabelan menggunakan kamus *Lexicon*. Penerapan analisis sentimen dilakukan untuk menentukan komentar tersebut bersifat positif atau negatif. Dalam penelitian ini dapat diketahui akurasi metode *Naïve Bayes* dan *Support Vector Machine* dalam melakukan analisis sentimen pada ulasan Summarecon Mall Bekasi dengan data sebanyak 2.143 komentar dengan akurasi untuk *Naïve Bayes* dan *Support Vector Machine* berturut-turut 80,95% dan 100%. Sebuah prototipe berbasis Jason pada penelitian ini diimplementasikan dengan *framework flask* [8].

Pada penelitian terkait melakukan analisis dengan membandingkan Algoritma *Naïve Bayes* dan Algoritma *Support Vector Machine (SVM)* untuk mengklasifikasikan ulasan aplikasi *Netflix* pada *Google Play Store*. Menerapkan metode *scraping*, kemudian dipecah menjadi data latih 70% dan test data 30%

Penerapan algoritma *Naïve Bayes* mendapatkan hasil *accuracy* senilai 82%, lalu *Support Vector Machine (SVM)* memperoleh hasil *accuracy* senilai 85%. Sehingga, menunjukkan hasil penerapan algoritma *Naive Bayes* tidak lebih efektif dibandingkan *Support Vector Machine (SVM)* dengan menghasilkan akurasi 0,85 (85%) dan *Naive Bayes* menghasilkan akurasi 0.82 (82%) [9].

Penelitian lainnya juga membandingkan algoritma klasifikasi *Naïve Bayes* dan *Support Vector Machine (SVM)* dengan tujuan untuk mengetahui algoritma klasifikasi terbaik berdasarkan nilai *accuracy* untuk sentimen analisis penggunaan *platform market place*. Dengan menggunakan metode *Cross Validation* dan dari

hasil-hasil pengujian perbandingan tersebut bahwa algoritma *SVM* memiliki nilai *accuracy* yang terbaik dibandingkan dengan algoritma NB. Dimana nilai *accuracy* Tokopedia menggunakan algoritma NB sebesar 85.34%, dan nilai *accuracy* menggunakan *SVM* 86.82%, untuk nilai *accuracy* Shopee menggunakan NB sebesar 80.04%, dan nilai *accuracy* menggunakan *SVM* 80.91%, dan Lazada yang menggunakan algoritma NB mempunyai nilai *accuracy* 83.52%, sedangkan nilai *accuracy* menggunakan *SVM* 88.93% yang artinya penggunaan algoritma *SVM* memiliki tingkat akurasi terbaik [7].

Penelitian ini juga diharapkan dapat memberikan kontribusi dalam pengembangan sistem otomatis untuk analisis sentimen di platform media sosial, khususnya Google Maps. Selain itu, penelitian ini akan berfungsi sebagai referensi untuk penelitian dan pengembangan sistem klasifikasi ulasan secara otomatis di masa mendatang. Berdasarkan latar belakang permasalahan berikut dan penelitian berikut maka terbangun rasa keinginan untuk menyusun tugas akhir dengan judul **“Perbandingan Algoritma *Naïve Bayes* dan *Support Vector Machine (SVM)* pada Analisis Sentimen Mall Pakuwon di Google Maps”**.

1.2 Identifikasi Masalah

Berdasarkan latar belakang penelitian ini maka permasalahan dalam tugas akhir ini adalah: (samakan banyak point nya dengan rumusan, tujuan)

1. Belum diketahui algoritma mana yang memberikan hasil akurasi yang lebih tinggi antara *Support Vector Machine (SVM)* dan *Naive Bayes* dalam mengklasifikasikan sentimen ulasan Mall Pakuwon Bekasi.

2. Dibutuhkan metode otomatis untuk mengklasifikasikan sentimen dari ulasan Mall Pakuwon Bekasi menggunakan algoritma *Naive Bayes* dan *Support Vector Machine (SVM)* agar pihak manajemen dapat memahami persepsi publik.

1.3 Rumusan Masalah

Berdasarkan identifikasi masalah yang ada, maka masalah yang dapat dirumuskan adalah:

1. Bagaimana perbandingan akurasi dan kinerja antara algoritma *SVM* dan *Naive Bayes* dalam klasifikasi sentimen ulasan tersebut?
2. Bagaimana cara mengklasifikasikan sentimen ulasan pengunjung *Mall Pakuwon Bekasi* menggunakan algoritma *SVM* dan *Naive Bayes*?

1.4 Batasan Masalah

Batasan masalah yang ada pada penelitian ini:

1. Data yang digunakan berasal dari ulasan pengguna di Mall Pakuwon Bekasi, yang dapat diakses melalui Google Maps.
2. Ada tiga jenis perasaan: positif, negatif, dan netral.
3. Hanya Bahasa Indonesia yang digunakan sebagai bahasa ulasan yang dievaluasi.
4. Untuk perbandingan, hanya ada dua algoritma: *Naive Bayes* dan *Support Vector Machine (SVM)*.
5. Untuk mempertahankan efisiensi pemrosesan, data ulasan dibatasi pada jumlah tertentu, yaitu 1037 data ulasan dari November 2024 – Maret 2025.

1.5 Tujuan Penelitian

Dengan rumusan masalah tersebut, maka tujuan dilakukannya penelitian ini adalah:

1. Membangun model klasifikasi sentimen untuk menganalisis ulasan pengunjung Mall Pakuwon Bekasi menggunakan algoritma Support Vector Machine (*SVM*) dan *Naive Bayes*.
2. Membandingkan performa kedua algoritma tersebut berdasarkan akurasi, *precision*, *recall*, dan *f1-score*.
3. Menyediakan insight tentang persepsi publik terhadap Mall Pakuwon Bekasi berdasarkan hasil analisis sentimen, yang dapat digunakan untuk perbaikan pelayanan dan strategi promosi Mall.

1.6 Manfaat Penelitian

Dengan tercapainya tujuan penelitian, maka dari itu penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. **Manfaat Teoritis**
 - a Kontribusi pada kajian klasifikasi teks, penelitian ini memberikan kontribusi ilmiah dalam bidang data mining dan natural language processing (NLP), khususnya dalam pemanfaatan algoritma *Naive Bayes* dan *SVM* untuk klasifikasi teks dalam bahasa Indonesia.
 - b Pengayaan literatur perbandingan algoritma, studi ini menambah referensi akademik tentang perbandingan performa klasifikasi antara dua algoritma populer *Naive Bayes* dan Support Vector Machine pada domain ulasan

publik, yang sebelumnya lebih banyak diterapkan pada produk, politik, atau layanan publik.

- c Dasar pengembangan model klasifikasi berbasis sentimen, hasil penelitian ini dapat digunakan sebagai dasar untuk pengembangan model-model klasifikasi atau sistem rekomendasi berbasis ulasan (*review-based recommendation*) dalam konteks pusat perbelanjaan.

2. Manfaat Praktis

- a Membantu manajemen Mall memahami persepsi publik, dengan mengolah ribuan ulasan secara otomatis, manajemen Mall Pakuwon Bekasi dapat memahami tingkat kepuasan, keluhan, dan harapan pengunjung, sehingga dapat mengambil keputusan yang lebih tepat sasaran.
- b Mendukung pengambilan keputusan strategis berbasis data, hasil klasifikasi sentimen dapat digunakan untuk mendukung strategi pelayanan, promosi, atau pengembangan tenant, sehingga lebih sesuai dengan kebutuhan dan pengalaman pengunjung.
- c Memberikan gambaran bagi pengelola Mall lain, metodologi dan hasil penelitian ini dapat direplikasi oleh Mall lain atau pelaku usaha retail besar untuk memantau reputasi merek dan layanan mereka secara berkelanjutan melalui data ulasan online.

1.7 Sistematika Penulisan

Penyusunan sistematika ini berguna untuk memberikan gambar singkat yang jelas mengenai penelitian ini, maka dibuatlah sistematika penulisan sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisi latar belakang masalah, identifikasi masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penelitian.

BAB II LANDASAN TEORI

Pada bab ini membahas berbagai teori pendukung, landasan teoritis, konsep, model, metode, analisis, pustaka ilmiah yang relevan dengan topik yang diteliti, serta kajian pustaka penelitian terkait.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan secara men-detail tentang metode penelitian yang digunakan dalam penyelidikan materi pelajaran, algoritma yang digunakan, tahapan-tahapan penelitian, dan metode pengumpulan data.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan uraian hasil yang diperoleh dari pengujian yang dilakukan dan hasil pengujian dalam konteks teori penelitian sebelumnya

BAB V PENUTUP

Berisi kesimpulan penelitian dan saran penelitian.