

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi internet yang terus berkembang kini menjadi bagian tak terpisahkan dari kehidupan banyak orang, memungkinkan komunikasi dan berbagi informasi melalui media sosial. Salah satu bentuk ekspresi yang muncul di internet adalah ulasan film. Film, sebagai seni yang memadukan gambar, suara, dan cerita, sering digunakan untuk menyampaikan pesan tertentu sekaligus menjadi salah satu bentuk hiburan yang paling populer [1].

Industri film adalah salah satu sektor yang sangat dipengaruhi oleh opini publik. lebih dari 80% penonton mempertimbangkan ulasan sebelum memutuskan untuk menonton film [2]. Analisis ulasan film memberikan informasi penting mengenai preferensi penonton, yang dapat digunakan untuk mengidentifikasi tren pasar, mengembangkan strategi pemasaran, dan meningkatkan kualitas produksi. Namun, tantangan muncul ketika ulasan film harus diklasifikasikan ke dalam beberapa kategori, terutama ketika sebuah ulasan mencakup elemen yang ambigu, seperti campuran opini positif dan negatif. Hal ini memerlukan pendekatan yang lebih canggih untuk meningkatkan akurasi klasifikasi [3]. Ulasan film berfungsi sebagai sumber informasi dan referensi yang membantu penonton dalam membuat keputusan menonton, sekaligus menjadi cerminan opini publik terhadap sebuah karya film. Dengan menganalisis ulasan, dapat diperoleh wawasan mengenai kualitas produksi, daya tarik cerita, serta tingkat kepuasan penonton, yang semuanya berdampak langsung terhadap keberhasilan komersial sebuah film.

Klasifikasi teks merupakan salah satu teknik yang sangat berguna dalam berbagai aplikasi yang berkaitan dengan pemrosesan bahasa alami NLP dan penambangan teks. Teknik ini memainkan peran penting dalam proses pengambilan atau ekstraksi informasi, di mana informasi relevan diidentifikasi dan diambil dari sejumlah besar data teks. klasifikasi teks menjadi semakin penting dalam membantu pengguna menemukan informasi yang mereka butuhkan dengan lebih efisien [4]. Klasifikasi teks merupakan teknik dasar dalam pemrosesan bahasa alami, yang digunakan dalam berbagai aplikasi seperti analisis sentimen, penyaringan konten, dan pengelompokan dokumen secara otomatis. Contoh aplikasinya sangat beragam,

termasuk penggunaan penyaring spam pada email untuk memisahkan pesan yang tidak diinginkan dari yang penting, serta pengorganisasian teks untuk membentuk arsip dokumen yang terstruktur dan mudah diakses. Selain itu, klasifikasi teks juga dapat diterapkan dalam analisis sentimen, di mana opini atau emosi dari sebuah teks dievaluasi dan dikategorikan [5]. diperlukan metode yang dapat mengelompokkan informasi secara otomatis dan efisien.

Dalam konteks ini klasifikasi teks, metode berbasis machine learning seperti *Naive Bayes* menjadi salah satu pendekatan yang efektif. *Naive Bayes Classifier* adalah algoritma klasifikasi berbasis probabilistik yang sering digunakan dalam data mining dan pemrosesan bahasa alami (NLP). Metode ini bekerja berdasarkan Teorema Bayes, dengan asumsi bahwa setiap fitur dalam dataset saling independen atau tidak bergantung satu sama lain. Dalam proses klasifikasinya, menghitung probabilitas setiap kelas berdasarkan distribusi kata dalam teks, kemudian menentukan kelas dengan probabilitas tertinggi sebagai hasil akhir. Keunggulan utama metode ini adalah kesederhanaannya, efisiensi dalam menangani dataset besar, serta performa yang baik dalam berbagai skenario klasifikasi teks. Oleh karena itu, metode ini banyak diterapkan dalam analisis sentimen, seperti klasifikasi ulasan film menjadi positif atau negatif [6].

Berdasarkan penelitian sebelumnya berjudul Berdasarkan penelitian sebelumnya berjudul “Analisis Sentimen Film The Marvels Dari Aplikasi Twitter Menggunakan Metode *Lexicon Based*”, penelitian ini menggunakan metode *Lexicon-Based* untuk menganalisis sentimen ulasan film The Marvels yang diambil dari media sosial Twitter. Data yang digunakan dalam penelitian ini berjumlah 1656 tweet yang dikumpulkan berdasarkan kata kunci The Marvels. Hasil dari analisis sentimen menunjukkan bahwa 67,2% dari ulasan memiliki sentimen positif, 12,6% bersifat netral, dan 20,2% memiliki sentimen negatif. Hal ini menunjukkan bahwa film The Marvels diterima dengan baik dan menarik perhatian masyarakat [7].

Penelitian selanjutnya yang berjudul Penelitian yang dilakukan oleh Adinda Cahya Kamilla, Natasya Priyani, Ressa Priskila, dan Viktor Handrianus Pranatawijaya berjudul "Analisis Sentimen Film Agak Laen dengan Kecerdasan Buatan: Text Mining Metode *Naive Bayes Classifier*" bertujuan untuk menganalisis opini masyarakat terhadap film *Agak Laen* di media sosial Twitter menggunakan

metode *Naive Bayes Classifier*. Dataset yang digunakan terdiri dari 255 tweet yang dikumpulkan melalui teknik *web scraping* menggunakan API Twitter dalam rentang waktu 9 Maret hingga 17 Maret 2024. Setelah melalui proses seleksi, sebanyak 195 tweet dipilih untuk diproses lebih lanjut. Setelah dilakukan klasifikasi dengan *Naive Bayes Classifier*, hasil evaluasi menunjukkan bahwa model mencapai akurasi sebesar 78%, *precision* sebesar 85,51%, *recall* sebesar 85,57%, dan *F1-score* sebesar 82,71%. Hasil ini menunjukkan bahwa metode *Naive Bayes Classifier* cukup efektif dalam menganalisis sentimen ulasan film di Twitter. Penelitian ini berkontribusi dalam pemahaman terhadap respons dan penilaian masyarakat terhadap film Agak Laen, serta dapat menjadi referensi untuk pengembangan penelitian di bidang analisis sentimen dan text mining [8].

Penelitian yang dilakukan Fery Anuar Ramadhan Putra, Frido Firman Fadilah, dan Ultach Enri berjudul “*ANALISIS SENTIMEN ULASAN FILM OPPENHEIMER PADA SITUS IMDB MENGGUNAKAN METODE NAIVE BAYES*”, penelitian ini bertujuan untuk mengevaluasi tingkat akurasi analisis sentimen terhadap film Oppenheimer berdasarkan ulasan penonton yang dipublikasikan di situs Internet Movie Database (IMDb) dengan menerapkan metode *Naive Bayes*. Data yang digunakan sebanyak 41.078 data dibagi menjadi dua kelompok. Kelompok pertama adalah data pelatihan yang mencakup 80% dari total data, yaitu 32.682 data, sedangkan kelompok kedua adalah data pengujian yang terdiri dari 20% atau sebanyak 8.216 data. Data diperoleh dari ulasan pengguna terkait film Oppenheimer di platform IMDb. Setelah itu, data melalui proses pembersihan sebelum dikonversi menjadi dokumen dalam format CSV. Hasil penelitian ini mencapai tingkat akurasi sebesar 96%, dengan presisi rata-rata 98%, *recall* rata-rata 90%, dan *F1-score* rata-rata 93% [9].

Penelitian yang dilakukan oleh Muhammad Haidar Rifki, Yustina Retno Wahyu Utami, dan Paulus Harsadi berjudul "Text Mining Untuk Analisis Sentimen Review Film Menggunakan Algoritma *Naive Bayes*" bertujuan untuk menganalisis opini pengguna terhadap film berdasarkan ulasan di situs IMDb. Dalam penelitian ini, metode *Naive Bayes Classifier* diterapkan untuk mengklasifikasikan sentimen pengguna ke dalam kategori positif dan negatif. Data dikumpulkan melalui teknik *web scraping* dan diproses melalui tahapan *preprocessing* seperti cleansing,

tokenizing, *stopword removal*, dan *stemming*. Dataset yang digunakan terdiri dari 5.000 ulasan film yang diperoleh dari Stanford Sentiment Analysis Dataset, dengan pembagian 2.500 ulasan positif dan 2.500 ulasan negatif. Hasil pengujian menunjukkan bahwa model menghasilkan akurasi sebesar 87,2%, *recall* sebesar 92,4%, *precision* sebesar 80,9%, dan error rate sebesar 12,8% [10].

Berdasarkan penelitian-penelitian sebelumnya masih menggunakan dataset atau dokumen ulasan dalam bahasa Inggris, sementara studi mengenai analisis sentimen dalam ulasan film berbahasa Indonesia masih terbatas. Untuk mengatasi keterbatasan ini, penelitian ini menggunakan dataset berbahasa Indonesia yang diperoleh dari situs Kaskus, yang lebih representatif terhadap ulasan film yang lebih panjang. Selain itu, pendekatan *Lexicon-Based* diterapkan untuk dapat meningkatkan akurasi klasifikasi. Dengan demikian, penelitian ini berkontribusi dalam pengembangan metode klasifikasi teks ulasan film berbahasa Indonesia serta memperluas wawasan terkait sentimen penonton terhadap berbagai film.

Dari penelitian sebelumnya, metode *Naive Bayes Classifier* menunjukkan tingkat akurasi yang baik dalam klasifikasi sentimen. Oleh karena itu, penulis melakukan penelitian dengan judul Klasifikasi Teks Ulasan Film Berbahasa Indonesia Menggunakan Metode *Naive Bayes* dengan Pendekatan *Lexicon-Based*. Metode *Naive Bayes* dipilih karena dikenal sebagai teknik klasifikasi teks yang sederhana, mudah untuk diterapkan, dan memiliki performa akurasi yang cukup baik.

1.2 Perumusan Masalah

Setelah mengetahui latar belakang penelitian tentang Klasifikasi Teks Ulasan Film Berbahasa Indonesia Menggunakan *Naive Bayes* Dengan Pendekatan *Lexicon-Based*, maka ditemukan permasalahan sebagai berikut:

1.2.1 Permasalahan

Analisis ulasan film dalam era digital menjadi tantangan karena banyaknya data dan kompleksitas opini penonton. Ulasan yang semakin banyak dihasilkan melalui platform digital, seperti media sosial dan situs ulasan, menimbulkan tantangan dalam pengelompokan informasi secara otomatis dan efisien. Salah satu tantangan utama adalah klasifikasi, di mana ulasan harus dikelompokkan ke dalam beberapa kategori yang berbeda.

Salah satu tantangan utama dalam analisis ini adalah klasifikasi teks ulasan secara otomatis dan akurat. Permasalahan semakin kompleks ketika ulasan mengandung elemen ambigu atau opini campuran, yang dapat menyebabkan kesalahan dalam penentuan sentimen. Pendekatan manual tidak lagi efisien dalam menangani volume data yang besar, sehingga diperlukan metode berbasis machine learning. Metode *Naive Bayes* dengan pendekatan *Lexicon-Based* menjadi alternatif yang menarik untuk klasifikasi ulasan film, karena diharapkan mampu memanfaatkan informasi dari kamus kata untuk meningkatkan akurasi klasifikasi sentimen positif dan negatif. Namun, masih diperlukan penelitian lebih lanjut untuk mengevaluasi efektivitas pendekatan ini dalam konteks klasifikasi teks ulasan film.

1.2.2 Metode Usulan

Penelitian ini mengusulkan metode berbasis machine learning yaitu metode *Naive Bayes* dengan pendekatan *Lexicon-Based* untuk klasifikasi teks pada ulasan film. Proses penelitian meliputi:

1. Pengumpulan data ulasan film.
2. Pelabelan sentiment menggunakan *Lexicon-Based*.
3. Pra-pemrosesan data untuk meningkatkan kualitas input, seperti tokenisasi, normalisasi, dan penghapusan stopwords.
4. Implementasi model *Naive Bayes* untuk klasifikasi teks.
5. Evaluasi performa kedua metode menggunakan metrik seperti akurasi, *precision*, *recall*, dan *F1-score*.
6. Analisa hasil Klasifikasi Teks.

Hasil dari penelitian ini diharapkan dapat memberikan wawasan yang bermanfaat dalam penerapan metode *Naive Bayes* dengan pendekatan *Lexicon-Based* untuk klasifikasi teks ulasan film, baik untuk kebutuhan akademis maupun industri.

1.2.3 Pertanyaan Penelitian

Berdasarkan latar belakang yang telah diuraikan maka dapat diketahui pertanyaan penelitian sebagai berikut :

1. Bagaimana performa metode *Naive Bayes* dengan pendekatan *Lexicon-Based* dalam mengklasifikasikan sentimen ulasan film berdasarkan metrik evaluasi seperti akurasi, presisi, *recall*, dan *F1-score*?

- 2 Bagaimana pengaruh pembagian dataset terhadap akurasi metode *Naive Bayes* untuk klasifikasi teks pada ulasan film?
- 3 Bagaimana pengaruh perubahan nilai parameter alpha terhadap akurasi metode *Naive Bayes* untuk klasifikasi teks pada ulasan film? Tujuan dan Manfaat

1.3 Tujuan dan Manfaat

Penelitian dilakukan guna mencapai tujuan dan manfaat. Adapun beberapa tujuan dan manfaat pada penelitian ini dijelaskan sebagai berikut

1.3.1 Tujuan

Berdasarkan rumusan masalah yang ada, dapat ditentukan tujuan dari penelitian ini diantaranya adalah sebagai berikut :

1. Melakukan Klasifikasi Teks ulasan film menggunakan metode *Naive Bayes* dengan pendekatan *Lexicon-Based* untuk mengetahui apakah sebuah ulasan mengandung sentimen positif atau sentimen negatif.
2. Menguji tingkat akurasi metode *Naive Bayes* dengan pendekatan *Lexicon-Based* berdasarkan metrik evaluasi klasifikasi, yaitu akurasi, presisi, *recall*, dan *F1-score*.
3. Mempersingkat waktu untuk membaca dan mengetahui apakah suatu film yang diulas termasuk sentimen positif atau sentimen negatif.

1.3.2 Manfaat

Manfaat yang di harapkan dari hasil penelitian ini adalah sebagai berikut :

1. Dapat membantu seseorang dalam mengambil keputusan saat ingin menonton sebuah film agar lebih efisien dibandingkan jika harus membaca semua ulasan yang dibuat oleh orang lain.
2. Dapat digunakan sebagai referensi untuk penelitian-penelitian selanjutnya dengan topik analisis sentimen.

Dengan demikian, penelitian ini dapat mendukung pengambilan keputusan bagi penonton dan kritikus dalam menilai kualitas suatu film secara lebih objektif.

1.4 Batasan Masalah

Penelitian ini menetapkan batasan masalah untuk memastikan fokus dan kejelasan dalam pembahasan yang dilakukan. Adapun batasan masalah dari penelitian ini adalah sebagai berikut:

1. Penelitian ini hanya membahas performa metode *Naive Bayes* dengan pendekatan *Lexicon-Based* dalam mengklasifikasikan teks ulasan film berbahasa Indonesia ke dalam dua kelas sentimen, yaitu positif dan negatif.
2. Dataset yang digunakan hanya pada ulasan film berbahasa Indonesia yang didapat dari situs kaskus.com pada forum movie, sehingga penelitian ini tidak mencakup ulasan dalam bahasa lain atau dari sumber selain Kaskus.
3. Dataset yang digunakan terdiri dari 3.250 data, di mana data tersebut mencakup total 17 film yang diambil dari film yang dirilis tahun 2021 hingga 2025.
4. Fokus penelitian adalah pada evaluasi performa metode *Naive Bayes* dengan pendekatan *Lexicon-Based* dalam klasifikasi teks ulasan film berdasarkan metrik akurasi, presisi, *recall*, dan *F1-score*

1.5 Sistematika Penulisan

Proposal skripsi ini disusun dengan mengikuti aturan penulisan yang telah ditetapkan oleh Program Studi Sistem Informasi, Universitas Trunojoyo Madura. Sistematika penulisan ini bertujuan untuk memberikan gambaran yang sistematis mengenai penelitian yang akan dilakukan, mencakup latar belakang masalah, perumusan masalah, tujuan penelitian, serta metode yang digunakan dalam penelitian ini.

BAB I PENDAHULUAN

Menjelaskan mengenai latar belakang, perumusan masalah, permasalahan, metode usulan, pertanyaan penelitian, tujuan dan manfaat penelitian, batasan masalah, serta sistematika penulisan terkait penelitian yang akan dilakukan.

BAB II TINJAUAN PUSTAKA

Pada bab ini berisi penjelasan kajian Pustaka yang mendukung penerapan metode yang akan digunakan dalam penelitian. menguraikan kajian literatur yang mendukung

penelitian, seperti konsep dasar klasifikasi teks, algoritma *Naive Bayes*,

BAB III METODE PENELITIAN

Pada bab ini memaparkan metode penelitian, meliputi desain penelitian, objek penelitian, langkah-langkah penelitian, serta alur proses klasifikasi teks menggunakan metode *Naive Bayes*

BAB IV HASIL DAN PEMBAHASAN

Bab ini menjelaskan hasil dari penerapan metode *Naive Bayes* dengan pendekatan *Lexicon-Based* dalam melakukan klasifikasi teks ulasan film berbahasa Indonesia serta pembahasan atas temuan-temuan yang diperoleh. Analisis dimulai dari proses *preprocessing* data teks, pelatihan dan pengujian model klasifikasi, evaluasi performa model berdasarkan metrik *akurasi*, *presisi*, *recall*, dan *F1-score*, hingga interpretasi terhadap hasil klasifikasi yang dihasilkan.

BAB V PENUTUP

Bab ini menjelaskan kesimpulan dari hasil penelitian yang telah dilakukan serta saran untuk pengembangan dan penelitian lanjutan. Kesimpulan disusun berdasarkan hasil analisis klasifikasi teks ulasan film yang diperoleh melalui penerapan metode *Naive Bayes* dengan pendekatan *Lexicon-Based*. Selain itu, bab ini juga memberikan masukan terkait potensi pengembangan metode, peningkatan kualitas data, maupun pendekatan alternatif yang dapat digunakan dalam penelitian sejenis di masa mendatang.