

# APLIKASI STREAMLIT UNTUK DETEKSI SENTIMEN PADA ULASAN FILM MENGGUNAKAN METODE LOGISTIC REGRESSION

Aqeela Fathya Najwa  
*Fakultas Informatika*  
Telkom University,  
Bandung, Indonesia  
aqeelafn@student.telkomu  
niversity.ac.id

Nisrina Hana Anindya  
*Fakultas Informatika*  
Telkom University,  
Bandung, Indonesia  
hanaanindy@student.telk  
omuniversity.ac.id

Muhammad Ikmal Akbar  
*Fakultas Informatika*  
Telkom University,  
Bandung, Indonesia  
lokololo@student.telkomu  
niversity.ac.id

Muhammad Baari Adli  
*Fakultas Informatika*  
Telkom University,  
Bandung, Indonesia  
adlibaari@student.telkomu  
niversity.ac.id

**Abstract**— Analisis sentimen terhadap ulasan film sangat penting dalam memahami opini pengguna dan meningkatkan pengalaman mereka. Penelitian ini menggunakan dataset IMDB yang terdiri dari 50.000 ulasan film yang seimbang antara ulasan positif dan negatif. Dalam eksperimen ini, dua model pembelajaran mesin, yaitu Naive Bayes multinomial dan regresi logistik, diterapkan untuk mengklasifikasikan sentimen ulasan tersebut. Data ulasan diolah menggunakan metode TF-IDF untuk mengubah teks menjadi representasi numerik. Hasil evaluasi menunjukkan bahwa regresi logistik mencapai akurasi 89%, sedangkan Naive Bayes multinomial mencapai 86%. Aplikasi berbasis Streamlit juga dikembangkan untuk memungkinkan pengguna memasukkan ulasan dan melihat prediksi sentimen secara real-time, dengan regresi logistik terbukti lebih efektif dalam konteks analisis sentimen ulasan film.

**Keywords**—, *Pembelajaran Mesin, Analisis Sentimen, Ulasan Film, Regresi Logistik, Streamlit.*

## I. INTRODUCTION

### A. Latar Belakang

Analisis sentimen ulasan film merupakan sebuah bagian menarik dalam topik NLP. Dalam era digital saat ini, internet, dan aplikasi online telah menjadi sumber utama informasi bagi banyak orang dalam berbagai aspek kehidupan, termasuk dalam mengevaluasi produk atau layanan sebelum memutuskan untuk mengonsumsinya. Salah satu produk yang sering dievaluasi adalah film. Situs web seperti IMDb menyediakan halaman di mana pengguna dapat memberikan ulasan dan penilaian terhadap film yang mereka tonton.

Dengan berkembangnya media sosial dan forum daring, jumlah ulasan film di platform seperti IMDb terus meningkat. Hal ini menimbulkan tantangan bagi seseorang untuk mengevaluasi dan menganalisis ulasan yang ada secara manual. Oleh karena itu, analisis sentimen otomatis menjadi bidang penelitian yang penting untuk memahami opini dan tanggapan publik terhadap berbagai produk dan layanan, termasuk film.

Analisis sentimen film bertujuan untuk mengekstraksi dan menganalisis sentimen yang terkandung dalam ulasan film, apakah cenderung positif, negatif, atau netral. Metode ini memungkinkan kita memahami pandangan yang lebih luas tentang bagaimana penonton merespons suatu film. Hal ini berguna bagi produsen film untuk meningkatkan kualitas

produksi dan memperbaiki aspek-aspek yang kurang disukai penonton.

Analisis sentimen ulasan film penting karena berbagai alasan. Pertama, dalam industri film, pemahaman terhadap ulasan dan reaksi penonton sangat vital. Dengan analisis sentimen, produsen film dapat memperoleh wawasan berharga tentang bagaimana penonton merespons karya mereka. Informasi ini memungkinkan mereka untuk melakukan perubahan yang diperlukan atau meningkatkan strategi pemasaran guna meningkatkan penerimaan film tersebut di pasar.

Selain itu, perkembangan teknologi dalam pemrosesan bahasa alami (Natural Language Processing/NLP) turut mendorong pentingnya analisis sentimen dalam konteks ulasan film. Kemajuan ini telah memungkinkan pengembangan model analisis sentimen yang lebih canggih dan efektif. Dengan menggunakan teknik-teknik NLP yang canggih itulah kita dapat memperoleh pemahaman yang lebih mendalam tentang opini, tanggapan publik terhadap berbagai aspek film. Hal ini memungkinkan para pengambil keputusan dalam industri film untuk lebih tepat dalam mengantisipasi tren pasar dan mengoptimalkan strategi produksi serta pemasaran.

Ulasan film, dan rating film juga sering menjadi aspek utama bagi penonton sebelum mereka memutuskan untuk menonton suatu film. Oleh karena itu, penting bagi kita untuk menganalisis sentimen di balik ulasan tersebut. Dengan melakukan hal ini, kita dapat membantu penonton membuat keputusan dan memaksimalkan pengalaman menonton mereka. Dengan demikian, analisis sentimen ulasan film tidak hanya memberikan manfaat bagi industri film, tetapi juga bagi penonton secara individu, memungkinkan mereka untuk menemukan film yang sesuai dengan preferensi dan ekspektasi mereka.

Maka dari itu, penelitian ini bertujuan untuk mengusulkan dan menerapkan model dan teknik klasifikasi yang tepat untuk analisis sentimen pada ulasan film IMDb. Melalui pendekatan ini, diharapkan dapat diperoleh pemahaman yang lebih baik tentang opini dan sentimen yang terkandung dalam ulasan film, serta keterlibatan terhadap industri film secara keseluruhan.

## B. Rumusan Masalah

1. Bagaimana proses eksplorasi dilakukan?
2. Bagaimana cara melakukan penanganan pada volume yang besar pada data ulasan film di IMBD?
3. Apa saja preprocessing yang dilakukan terhadap data?
4. Bagaimana proses implementasi model?

## C. Tujuan

Tujuan yang dapat diukur dari analisis penelitian ini, yaitu:

1. Mengembangkan model klasifikasi sentimen yang akurat, efektif, dan dapat diandalkan.
2. Melatih model menggunakan dataset yang telah dikumpulkan untuk memperoleh hasil yang optimal.
3. Agar dapat memahami opini dan tanggapan penonton terhadap film.
4. Mengidentifikasi fitur-fitur penting dalam ulasan film.

## II. TINJAUAN PUSTAKA

### A. Penelitian Terkait

Penelitian terkait analisis sentimen yang dilakukan oleh P. Frangidis, K. Georgiou, dan S. Papadopoulos berjudul "Sentiment Analysis on Movie Scripts and Reviews" [1]. Membahas tentang kombinasi skrip film dan ulasan untuk analisis sentimen dan prediksi rating film menggunakan berbagai algoritma pembelajaran mesin. Model ini dirancang untuk menganalisis skrip dan ulasan film guna memberikan prediksi rating yang lebih akurat. Penelitian ini mengumpulkan data skrip dan ulasan film, yang kemudian dianalisis menggunakan algoritma seperti regresi, klasifikasi, RNN, dan CNN. Model ini dilatih dan dievaluasi menggunakan berbagai metrik seperti akurasi dan precision. Hasilnya menunjukkan bahwa kombinasi analisis skrip dan ulasan dapat meningkatkan akurasi prediksi rating film dan memberikan wawasan yang lebih kaya mengenai persepsi penonton.

Penelitian berikutnya masih terkait analisis sentimen yang dilakukan oleh Y. Wang, G. Shen, dan L. Hu yang berjudul "Importance Evaluation of Movie Aspect: Aspect-Based Sentiment Analysis" [2]. Membahas tentang evaluasi aspek-aspek film yang paling berpengaruh berdasarkan analisis sentimen ulasan film. Penelitian ini bertujuan untuk mengidentifikasi aspek mana yang paling mempengaruhi sentimen penonton terhadap film. Data ulasan film dikumpulkan dan dianalisis menggunakan model topic modeling dan analisis sentimen berbasis neural network. Arsitektur model melibatkan analisis aspek untuk mengidentifikasi dan mengevaluasi aspek penting. Hasil penelitian menunjukkan bahwa aspek seperti cerita, akting, dan efek visual memiliki pengaruh yang signifikan terhadap sentimen penonton, dan analisis berbasis aspek dapat

membantu mengidentifikasi kekuatan dan kelemahan film dengan lebih akurat.

Lalu penelitian lainnya terkait analisis sentimen yang dilakukan oleh S. Prathap, dan Sk. Moinuddin Ahmad yang berjudul "Sentiment Analysis on Movie Reviews" [3]. Membahas tentang penggunaan algoritma deep learning untuk analisis sentimen pada ulasan film guna menentukan apakah ulasan tersebut bersifat positif atau negatif. Model ini menggunakan pendekatan deep learning seperti LSTM dan GRU untuk pemrosesan teks ulasan. Data ulasan film dikumpulkan dari platform ulasan populer, kemudian diproses dan dilabeli. Arsitektur model melibatkan penggunaan LSTM dan GRU untuk menangkap hubungan temporal dalam teks. Model ini dilatih dan dievaluasi menggunakan metrik seperti F1-score, AUC, precision, dan recall. Hasil penelitian menunjukkan bahwa model deep learning seperti LSTM dan GRU dapat memberikan performa yang lebih baik dalam mengklasifikasikan sentimen ulasan film dibandingkan dengan metode tradisional.

### B. Definisi dan Konsep

#### 1. Analisis Sentimen dan Ulasan Film

Analisis sentimen merupakan proses ekstraksi untuk menentukan sikap, perasaan, atau opini seseorang terhadap suatu topik berdasarkan teks yang mereka tulis dan merupakan sub bidang dari Natural Language Processing & Text Mining yang sering digunakan untuk memahami opini public dari data seperti ulasan produk, ulasan film, postingan media sosial, artikel berita, dan lainnya. Ulasan film adalah suatu evaluasi kritis, tanggapan, opini yang ditulis oleh penonton yang menggambarkan pengalaman menonton mereka pada sebuah film. Ulasan ini mencakup berbagai aspek dari film, seperti alur film, akting para aktor dan aktris, sinematografi, background music, dan elemen dalam film lainnya. Tujuan ulasan film adalah memberikan informasi kepada calon penonton, pembaca dan pengguna lain tentang kualitas film karena ulasan film, dan rating film juga sering menjadi aspek utama bagi penonton sebelum mereka memutuskan untuk menonton suatu film dan.

Analisis sentimen pada ulasan film bertujuan untuk mengekstraksi dan menganalisis sentimen yang terkandung dalam ulasan film, apakah cenderung positif, negatif, atau netral. Metode ini memungkinkan kita memahami pandangan yang lebih luas tentang bagaimana penonton merespons suatu film. Hal ini berguna bagi produsen film untuk meningkatkan kualitas produksi dan memperbaiki aspek-aspek yang kurang disukai penonton.

#### 2. Logistic Regression

Logistic regression (regresi logistik) merupakan salah satu algoritma Machine Learning yang digunakan untuk memodelkan hubungan antara independent dan variable dependen biner, juga digunakan untuk mengklasifikasikan data ke dalam salah satu dari dua kategori (klasifikasi biner). Meskipun namanya

mengandung kata "regression," regresi logistik lebih sering digunakan untuk masalah klasifikasi. Pada analisis sentimen, logistic regression digunakan untuk memprediksi probabilitas bahwa sebuah teks, seperti ulasan film, termasuk dalam kategori tertentu berdasarkan fitur-fitur yang diekstraksi dari teks tersebut. Proses ini dimulai dengan pengumpulan dan pra-pemrosesan data ulasan, yang melibatkan pembersihan teks, tokenisasi, dan ekstraksi fitur menggunakan metode seperti TF-IDF, Bag of Words ataupun metode lainnya. Setelah model dilatih dan dievaluasi, dapat digunakan untuk memprediksi sentimen ulasan baru, membantu pengguna memahami apakah sebuah ulasan film bersifat positif atau negatif.

Pada penelitian terdahulu yang menggunakan regresi logistik berjudul "Aspect-Based Sentiment Analysis on Movie Reviews 2023" oleh Brentton Wong Swee Kit dan M. Joseph [4]. Dalam penelitian ini, dua tugas utama dilakukan: prediksi aspek dan prediksi sentimen. Prediksi aspek dilakukan menggunakan model Logistic Regression dan Decision Tree, sedangkan prediksi sentimen dilakukan menggunakan model Logistic Regression dan Multinomial Naïve Bayes. Pada regresi logistik digunakan fungsi sigmoid untuk memprediksi probabilitas kategori, dengan akurasi mencapai 92% untuk prediksi aspek dan 93% untuk prediksi sentimen. Untuk Multinomial Naïve Bayes mendapatkan akurasi sebesar 91% untuk memprediksi sentimen. Selain itu untuk prediksi aspek pada penelitian ini Naïve Bayes memiliki akurasi yang lebih tinggi, yaitu 98%

Pada penelitian ini walaupun regresi logistik memberikan hasil yang baik untuk prediksi aspek, namun Decision Tree lebih unggul untuk memprediksi berdasarkan aspek. Namun regresi logistik juga menunjukkan performa yang sangat baik dalam prediksi sentimen, mengungguli Multinomial Naïve Bayes. Hasil penelitian ini tidak hanya membantu dalam memahami opini penonton film secara lebih mendalam tetapi juga memberikan wawasan berharga bagi industri film dalam mengevaluasi elemen-elemen penting dari film.

Jadi dalam konteks analisis sentimen pada ulasan film regresi logistik akan mempelajari pola dari dataset dan dapat mengkategorikan teks ulasan menjadi sentimen positif, negatif, atau netral berdasarkan fitur-fitur yang diekstraksi dari teks ulasan tersebut

### III. METODE

#### A. Dataset

Dataset yang kami gunakan, yaitu "IMDB Dataset" terdiri dari 50.000 baris data yang masing-masing berisi ulasan film dan sentimen yang terkait dengan ulasan tersebut. Dataset ini memiliki dua kolom utama: review yang berisi teks ulasan film dari pengguna, dan sentiment yang menunjukkan label sentimen dari ulasan tersebut, dengan dua kemungkinan nilai yaitu positive (positif) dan negative

(negatif). Distribusi sentimen dalam dataset ini sangat seimbang, dengan 25.000 ulasan positif dan 25.000 ulasan negatif.

#### B. Tahap Pembangunan Sistem

##### 1. *Processing Data*

Setelah data kami dapatkan, tahap selanjutnya yaitu data tersebut akan di lakukan analisis (EDA) pada data serta pengecekan apakah ada missing value, dan sebagainya. Pada tahap praproses kami mengubah semua huruf dalam teks ulasan menjadi huruf kecil. Tujuannya adalah untuk memastikan konsistensi dan menghindari perbedaan antara huruf besar dan huruf kecil. Selain itu kami juga melakukan penghapusan step berlebih diantara kata-kata dalam teks ulasan, menghapus stop word, dan stemming (mengubah kata-kata menjadi bentuk dasarnya).

##### 2. *Pembangunan Model Prediksi*

Data yang sudah siap untuk dijadikan model, kami lakukan pembagian dataset menjadi data latih dan data testing. Dalam proses ini, ulasan film (review) digunakan sebagai fitur (X) dan label sentimen (sentiment) digunakan sebagai target (y). Data dibagi dengan rasio 80:20, dimana 80% data digunakan untuk melatih model dan 20% untuk menguji model. Selanjutnya kami menggunakan TF-IDF untuk mengubah teks ulasan menjadi representasi numerik agar dapat di proses oleh algoritma ML. TF-IDF (Term Frequency-Inverse Document Frequency) adalah teknik yang mengubah teks menjadi fitur numerik berdasarkan seberapa sering kata muncul dalam dokumen dan seberapa penting kata tersebut dalam keseluruhan korpus.

Dengan data yang sudah dalam bentuk vector, selanjutnya model Naïve Bayes Multinomial dilakukan train-test pada data yang tadi sudah dibagi dengan rasio 80:20, dimana 80% data digunakan untuk melatih model dan 20% untuk menguji model. Prediksi yang dihasilkan oleh model dievaluasi menggunakan metrik seperti akurasi dan laporan klasifikasi, yang memberikan informasi tentang precision, recall, dan F1-score untuk setiap kategori sentimen (positif dan negatif).

Setelah itu, proses yang sama diterapkan pada model regresi logistik. Model ini juga dilatih menggunakan data latih (80%) yang sudah vektorisasi dan diuji dengan data uji (20%). Hasil prediksi dari model regresi logistik dievaluasi dengan cara yang sama seperti model Naive Bayes.

##### 3. *Implementasi Streamlit*

Langkah pertama dalam implementasi adalah mengintegrasikan model yang telah dilatih ke dalam aplikasi Streamlit. Model disimpan dalam format yang dapat diakses oleh aplikasi, dan Streamlit digunakan untuk membuat antarmuka pengguna yang interaktif. Dengan menggunakan Streamlit, pengguna dapat memasukkan data ulasan yang ingin mereka gunakan

dan melihat prediksi sentiment tersebut secara real-time. Streamlit adalah framework open-source yang memungkinkan pembuatan aplikasi web interaktif dengan cepat dan mudah menggunakan kode Python [5].

#### IV. HASIL DAN PEMBAHASAN

##### A. Hasil Eksperimen

Bagian ini menjelaskan hasil eksperimen dengan menggunakan dua model yang berbeda: Regresi Logistik dan Multinomial Naive Bayes. Kedua model ini dievaluasi pada kumpulan data film untuk menentukan kemampuannya dalam mengklasifikasikan emosi sebagai positif atau negatif. Tujuan dari percobaan ini adalah untuk mengidentifikasi model yang memberikan performa terbaik dan dapat diterapkan pada aplikasi analisis sentimen berbasis web menggunakan Streamlit.

##### 1. Eksperimen Model Naive Bayes Multinomial

Model Multinomial Naive Bayes dipilih karena kesederhanaannya dan kemampuannya dalam menangani data teks. Pada proses eksperimen, data ulasan film diubah ke dalam bentuk numerik dengan menggunakan teknik TF-IDF (Term Frequency-Inverse Document Frequency). TF-IDF membantu memberi bobot pada kata-kata yang lebih bermakna dan mengurangi pengaruh kata-kata umum.

Model Multinomial Naive Bayes menunjukkan akurasi 86%, dengan skor presisi, recall, dan F1 seimbang antara kategori negatif dan positif. Akurasi sebesar 0,85 untuk kelas negatif dan 0,87 untuk kelas positif menunjukkan bahwa model tersebut cukup akurat dalam mengklasifikasikan ulasan ke dalam kategori yang benar. Recall sebesar 0,87 untuk kelas negatif dan 0,85 untuk kelas positif menunjukkan bahwa model tersebut menangkap ulasan yang termasuk dalam setiap kategori dengan sangat baik. Skor F1 seimbang sebesar 0,86 menunjukkan bahwa model memiliki keseimbangan yang baik antara presisi dan recall.

##### 2. Eksperimen Model Logistic Regression

Selain model Multinomial Naive Bayes, kami juga menyelidiki model Regresi Logistik untuk melihat apakah model tersebut akan memberikan hasil yang lebih baik. Regresi logistik umumnya digunakan untuk klasifikasi biner dan dikenal karena kemampuan interpretasinya yang tinggi. Proses eksperimen regresi logistik mirip dengan Multinomial Naive Bayes. Data diubah ke bentuk numerik menggunakan TF-IDF dan dipisahkan menjadi data pelatihan dan pengujian, kemudian model dilatih pada data pelatihan dan diuji pada data pengujian.

Model Regresi Logistik menunjukkan hasil yang lebih unggul dengan akurasi 89% dibandingkan Multinomial Naive Bayes. Presisi untuk kelas negatif adalah 0,90 dan recall untuk kelas positif adalah 0,91, yang menunjukkan bahwa model tersebut sangat baik

dalam mengidentifikasi ulasan negatif dan juga sangat efektif dalam mendeteksi ulasan positif. Skor F1 sebesar 0,89 untuk kedua kelas menunjukkan bahwa regresi logistik memiliki keseimbangan yang sangat baik antara presisi dan recall, yang sangat penting dalam aplikasi analisis sentimen.

##### B. Evaluasi Model

Prediksi yang dihasilkan oleh model Naïve Bayes Multinomial dievaluasi menggunakan metrik seperti akurasi dan laporan klasifikasi, yang memberikan informasi tentang precision, recall, dan F1-score untuk setiap kategori sentimen (positif dan negatif).

Tabel 4.1 laporan klasifikasi Naïve Bayes Multinomial untuk setiap kelas

<i>Class</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
<i>Negative</i>	0.85	0.87	0.86	4961
<i>Positive</i>	0.87	0.85	0.86	5039
<i>Accuracy</i>	-	-	0.86	10000
<i>Macro avg</i>	0.86	0.86	0.86	10000
<i>Weighted avg</i>	0.86	0.86	0.86	10000

Kami juga melakukan evaluasi pada model lain yaitu logistic regression untuk memastikan performa model mana yang lebih baik digunakan dalam melakukan proses analisis sentiment.

Tabel 4.2 laporan klasifikasi Logistic Regression untuk setiap kelas

<i>Class</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
<i>Negative</i>	0.90	0.87	0.89	4961
<i>Positive</i>	0.88	0.91	0.89	5039
<i>Accuracy</i>	-	-	0.89	10000
<i>Macro avg</i>	0.89	0.89	0.89	10000
<i>Weighted avg</i>	0.89	0.89	0.89	10000

Dari hasil evaluasi di atas, dapat kita lihat bahwa model Logistic Regression memiliki akurasi yang sedikit lebih tinggi (0.89) dibandingkan dengan model Naive Bayes Multinomial (0.86). Precision, recall, dan f1-score juga lebih tinggi untuk Logistic Regression. Hal ini menunjukkan bahwa Logistic Regression memberikan performa yang lebih baik dalam mengklasifikasikan sentimen ulasan film dalam dataset ini dibandingkan dengan Naive Bayes Multinomial.

### C. Pembahasan

Hasil percobaan menunjukkan bahwa aplikasi analisis sentimen berbasis Streamlit yang dikembangkan memiliki kinerja yang baik dalam memprediksi sentimen berdasarkan masukan data review. Model regresi logistik dipilih untuk diterapkan pada aplikasi karena memiliki akurasi yang lebih tinggi dan memberikan metrik evaluasi yang lebih baik dibandingkan dengan model Multinomial Naive Bayes. Streamlit memungkinkan pengguna memasukkan teks ulasan dengan mudah dan menerima hasil prediksi sentimen secara real-time.

Keberhasilan aplikasi ini dapat dikaitkan dengan beberapa faktor penting. Pertama, proses pra-pemrosesan data yang menyeluruh memastikan bahwa data ulasan film berada dalam format terbaik untuk dianalisis. Tahap ini meliputi pembersihan data, normalisasi teks, dan penghapusan kata-kata umum yang tidak bermakna (stopwords). Hal ini membantu dalam menghasilkan data yang lebih bersih dan relevan untuk pelatihan model.

Kedua, penggunaan TF-IDF sebagai metode vektorisasi teks membantu menangkap makna kata dalam konteks keseluruhan dokumen dan korpus. TF-IDF memberikan bobot yang lebih tinggi pada kata-kata yang lebih bermakna dan sering muncul dalam dokumen tertentu tetapi jarang muncul dalam dokumen lainnya, sehingga meningkatkan kemampuan model dalam memahami konten teks.

Ketiga, dengan memilih model regresi logistik yang sesuai, kami menemukan bahwa model ini dapat mengatasi kompleksitas data teks dan memberikan prediksi yang akurat. Regresi logistik terkenal karena kemampuannya dalam menangani klasifikasi biner dengan baik, serta memberikan interpretasi yang jelas dari koefisien model.

Implementasi aplikasi menggunakan Streamlit menyediakan antarmuka pengguna yang intuitif dan responsif yang memungkinkan pengguna berinteraksi langsung dengan model. Aplikasi ini tidak hanya menampilkan prediksi sentimen tetapi juga menambah nilai dengan membantu pengguna membuat keputusan yang lebih baik berdasarkan analisis sentimen ulasan film. Pengguna dapat dengan mudah memasukkan teks ulasan dan melihat hasil prediksi secara instan, yang meningkatkan pengalaman pengguna.

Dengan kombinasi teknik pembelajaran mesin yang tepat dan antarmuka yang ramah pengguna, aplikasi ini berpotensi digunakan secara luas dalam berbagai konteks analisis sentimen. Misalnya, aplikasi ini dapat digunakan oleh perusahaan untuk menganalisis ulasan pelanggan, oleh peneliti untuk studi sentimen sosial, atau oleh individu untuk memahami tren sentimen pada platform media sosial.

Secara keseluruhan, hasil eksperimen dan evaluasi model menunjukkan bahwa aplikasi ini merupakan alat yang efektif dan efisien untuk analisis sentimen ulasan film. Dengan kemampuannya memberikan prediksi dengan akurasi tinggi

dan andal, aplikasi ini membantu pengguna lebih memahami sentimen ulasan film dan membuat keputusan yang lebih tepat. Aplikasi ini mencerminkan bagaimana teknologi pembelajaran mesin dapat diintegrasikan dengan antarmuka pengguna yang canggih untuk memberikan solusi yang praktis dan bermanfaat dalam analisis data teks.

### V. KESIMPULAN

Studi ini menggunakan dataset IMDB untuk mengembangkan dan mengevaluasi dua model pembelajaran mesin untuk analisis sentimen ulasan film: Multinomial Naive Bayes dan Regresi Logistik. Hasil eksperimen menunjukkan bahwa model regresi logistik mengungguli klasifikasi sentimen dengan akurasi 89%, sedangkan model Multinomial Naive Bayes mencapai akurasi 86%.

Implementasi aplikasi menggunakan Streamlit memberikan antarmuka yang intuitif dan responsif bagi pengguna untuk memasukkan ulasan dan melihat prediksi sentimen secara real-time. Keberhasilan aplikasi ini didukung oleh prapemrosesan data yang menyeluruh, penggunaan metode TF-IDF untuk vektorisasi teks, dan pemilihan model regresi logistik yang sesuai.

Selain memberikan prediksi sentimen yang akurat, aplikasi ini juga membantu pengguna membuat keputusan yang lebih baik berdasarkan analisis sentimen ulasan film. Potensi aplikasi ini sangat luas, dapat digunakan oleh bisnis dalam menganalisis ulasan pelanggan, oleh peneliti untuk studi sentimen sosial, dan oleh individu untuk memahami tren emosional di platform media sosial.

Studi ini menegaskan bahwa integrasi teknologi pembelajaran mesin ke dalam antarmuka pengguna tingkat lanjut dapat memberikan solusi praktis dan berguna untuk analisis data teks, dengan fokus pada pemahaman sentimen dalam konteks ulasan film.

### DAFTAR PUSTAKA

- [1] Frangidis, P., Georgiou, K., & Papadopoulos, S. (2020). Sentiment Analysis on Movie Scripts and Reviews. *Artificial Intelligence Applications and Innovations*, 583, 430 - 438. [https://doi.org/10.1007/978-3-030-49161-1\\_36](https://doi.org/10.1007/978-3-030-49161-1_36).
- [2] Wang, Y., Shen, G., & Hu, L. (2020). Importance Evaluation of Movie Aspects: Aspect-Based Sentiment Analysis. *2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE)*, 2444-2448. <https://doi.org/10.1109/ICMCCE51767.2020.00527>.
- [3] Prathap, S., & Ahmad, S. (2018). Sentiment Analysis on Movie Reviews. *International journal of scientific research in science, engineering and technology*, 4, 491-495. <https://doi.org/10.32628/IJSRSET18489>.
- [4] B. W. S. Kit and M. Joseph, "Aspect-Based Sentiment Analysis on Movie Reviews," in *2023 15th International Conference on Developments in eSystems Engineering (DeSE)*, 2023, pp. 237-243. doi: 10.1109/DeSE58274.2023.10099815.
- [5] A. Kumar, "Web-based applications for interactive user interfaces," *Journal of Open Source Software*, vol. 5, no. 50, pp. 2804-2812, 2020.