



PERBANDINGAN ALGORITMA *REGRESI LINEAR* DAN *REGRESI LOGISTIK* DALAM ANALISIS SENTIMEN KOMENTAR TIKTOK TERHADAP FILM SLINE

Nurhabsah¹, Nurdin²

^{1,2}Magister Teknologi Informasi, Universitas Malikussaleh

Email: nurhabsah.257110201023@mhs.unimal.ac.id¹, nurdin@unimal.ac.id²

Informasi Artikel	ABSTRACT
Riwayat artikel : Disubmit : 23 Mei 2026 Direvisi : 27 Mei 2026 Diterima : 1 Juni 2026 Dipublikasi : 30 Juni 2026	<i>This study aims to compare the performance of Linear Regression and Logistic Regression algorithms in sentiment analysis of TikTok comments related to the film Sline. The increasing use of social media as a platform for public opinion expression has encouraged the application of machine learning methods to classify user sentiment automatically. The dataset used in this research consists of TikTok comments that were processed through several stages, including text preprocessing, tokenization, cleaning, and TF-IDF feature extraction. Linear Regression and Logistic Regression models were then implemented and evaluated using accuracy, precision, recall, and F1-score metrics. The results indicate that Logistic Regression achieved better classification performance compared to Linear Regression in identifying positive and negative sentiments from TikTok comments. This study concludes that Logistic Regression is more effective for sentiment classification problems because it is specifically designed for categorical prediction tasks. The findings are expected to contribute to the development of sentiment analysis research on social media data.</i>
Keywords: Sentiment Analysis, Linear Regression, Logistic Regression, TikTok, Machine Learning	
	ABSTRAK
Kata Kunci: Analisis Sentimen, Regresi Linear, Regresi Logistik, TikTok, Machine Learning	Penelitian ini bertujuan membandingkan performa algoritma Regresi Linear dan Regresi Logistik dalam analisis sentimen komentar TikTok terhadap film Sline. Meningkatnya penggunaan media sosial sebagai sarana penyampaian opini mendorong penerapan metode <i>machine learning</i> untuk mengklasifikasikan sentimen pengguna secara otomatis. <i>Dataset</i> berupa komentar TikTok diproses melalui tahapan <i>preprocessing</i> teks, meliputi <i>cleaning</i>, tokenisasi, penghapusan <i>stopword</i>, dan ekstraksi fitur menggunakan TF-IDF. Model Regresi Linear dan Regresi Logistik kemudian diuji menggunakan metrik <i>accuracy</i>, <i>precision</i>, <i>recall</i>, dan <i>F1-score</i>. Hasil penelitian menunjukkan bahwa Regresi Logistik memiliki performa lebih baik dibandingkan Regresi Linear. Regresi Logistik memperoleh nilai <i>accuracy</i> 89%, <i>precision</i> 88%, <i>recall</i> 90%, dan <i>F1-score</i> 89%, sedangkan Regresi Linear memperoleh <i>accuracy</i> 76%, <i>precision</i> 74%, <i>recall</i> 77%, dan <i>F1-score</i> 75%. Hasil tersebut menunjukkan bahwa Regresi Logistik lebih efektif dalam mengklasifikasikan sentimen positif dan negatif karena dirancang khusus untuk klasifikasi data kategorikal. Dengan demikian, Regresi Logistik lebih tepat digunakan pada analisis sentimen berbasis data media sosial.





PENDAHULUAN

Perkembangan teknologi informasi dan internet telah mendorong pertumbuhan media sosial sebagai sarana komunikasi dan pertukaran informasi secara digital. Salah satu *platform* media sosial yang mengalami peningkatan penggunaan secara signifikan adalah TikTok. *Platform* ini memungkinkan pengguna untuk membuat, membagikan, serta memberikan komentar terhadap berbagai konten digital, termasuk film dan hiburan. Aktivitas pengguna tersebut menghasilkan data teks dalam jumlah besar yang mengandung opini, kritik, dan tanggapan publik terhadap suatu topik tertentu. Banyaknya data komentar pada media sosial menjadikan analisis sentimen sebagai salah satu pendekatan yang penting dalam bidang data *mining* dan *machine learning* untuk mengetahui kecenderungan opini masyarakat secara otomatis.

Analisis sentimen merupakan bagian dari text mining yang bertujuan untuk mengidentifikasi dan mengklasifikasikan opini pengguna ke dalam kategori positif, negatif, maupun netral. Penerapan analisis sentimen telah digunakan dalam berbagai bidang seperti bisnis, politik, pendidikan, layanan publik, hingga industri hiburan. Menurut Nurdin dan Khaledy (2017), teknik data mining mampu menemukan pola tersembunyi dari kumpulan data yang besar melalui proses klasifikasi sehingga dapat membantu pengambilan keputusan secara efektif. Penelitian tersebut juga menjelaskan bahwa proses data mining dilakukan melalui tahapan *Knowledge Discovery in Database* (KDD) yang meliputi *selection*, *cleaning*, *transformation*, *data mining*, dan *interpretation*. Tahapan tersebut relevan dengan penelitian analisis sentimen karena klasifikasi data teks memerlukan *preprocessing* sebelum model *machine learning* diterapkan.

Perkembangan *machine learning* telah menghasilkan berbagai metode klasifikasi yang digunakan untuk analisis data teks. Indo Intan et al. (2021) menyatakan bahwa algoritma *machine learning* mampu menghasilkan tingkat akurasi yang tinggi dalam proses klasifikasi data. Penelitian tersebut membandingkan beberapa algoritma *machine learning* untuk memperoleh performa terbaik berdasarkan akurasi, error, dan waktu komputasi. Selain itu, penelitian tersebut menjelaskan bahwa regresi linear memiliki kelemahan terhadap data *outlier* dan membutuhkan jumlah data yang lebih besar agar performa model meningkat. Hal ini menunjukkan bahwa pemilihan algoritma sangat mempengaruhi hasil klasifikasi maupun prediksi pada suatu dataset. Penelitian lain oleh Maulita dan Nurdin (2023) menjelaskan bahwa regresi linear merupakan metode statistik yang digunakan untuk memodelkan hubungan antara variabel dependen dan independen dalam proses prediksi berbasis data mining. Penelitian tersebut menyebutkan bahwa regresi linear memiliki keunggulan dalam menghasilkan prediksi yang efektif dan efisien, khususnya pada dataset dengan jumlah variabel yang





relatif kecil. Selain itu, metode regresi mampu memberikan interpretasi hubungan antar variabel secara jelas sehingga sering digunakan pada penelitian berbasis prediksi dan klasifikasi.

Dalam penelitian analisis sentimen, berbagai algoritma *machine learning* telah banyak digunakan, seperti Naïve Bayes, Support Vector Machine (SVM), Random Forest, Decision Tree, Regresi Linear, dan Regresi Logistik. Ahmed et al. (2020) menyatakan bahwa Regresi Logistik memiliki tingkat akurasi yang tinggi pada klasifikasi sentimen media sosial karena mampu memprediksi data kategorikal secara efektif. Penelitian lain oleh Kurniawan dan Putra (2021) menunjukkan bahwa penggunaan TF-IDF sebagai metode ekstraksi fitur dapat meningkatkan performa klasifikasi sentimen berbasis machine learning. Selain itu, Sari et al. (2022) menjelaskan bahwa Regresi Logistik memiliki performa yang lebih stabil dibandingkan metode linear lainnya pada klasifikasi sentimen ulasan film. Penelitian oleh Pratama et al. (2021) membandingkan algoritma Support Vector Machine, Naïve Bayes, dan Logistic Regression pada data Twitter berbahasa Indonesia dan menunjukkan bahwa Logistic Regression memiliki nilai precision dan recall yang lebih baik dibandingkan algoritma lainnya. Penelitian Utami dan Kurnia (2023) juga menunjukkan bahwa Logistic Regression mampu menghasilkan akurasi yang tinggi pada analisis sentimen TikTok karena memiliki kemampuan klasifikasi yang baik terhadap data teks berdimensi besar. Selanjutnya, Ramadhan dan Fitria (2023) menjelaskan bahwa tahapan *preprocessing* seperti *case folding*, *tokenizing*, *stopword removal*, dan *stemming* sangat mempengaruhi hasil klasifikasi sentimen.

Beberapa penelitian lain juga memperlihatkan bahwa pemanfaatan *machine learning* dalam analisis sentimen terus berkembang. Fauzi et al. (2021) menggunakan TF-IDF dan Logistic Regression untuk analisis sentimen Twitter dan memperoleh hasil klasifikasi yang baik. Saputra et al. (2022) menerapkan Logistic Regression pada analisis sentimen komentar YouTube dan menunjukkan peningkatan performa klasifikasi setelah dilakukan *preprocessing* data. Selain itu, Maulana dan Hidayat (2022) membandingkan Random Forest, Logistic Regression, dan Decision Tree dalam klasifikasi sentimen dan menemukan bahwa Logistic Regression memiliki kestabilan performa yang cukup baik pada data teks berbahasa Indonesia.

Meskipun penelitian mengenai analisis sentimen telah banyak dilakukan, sebagian besar penelitian sebelumnya lebih berfokus pada penggunaan algoritma populer seperti Naïve Bayes, SVM, dan Random Forest. Penelitian yang secara khusus membandingkan Regresi Linear dan Regresi Logistik pada analisis sentimen media sosial masih relatif terbatas. Selain itu, sebagian besar penelitian terdahulu menggunakan dataset dari Twitter, Facebook, atau *marketplace*, sedangkan penelitian berbasis komentar TikTok masih belum banyak dieksplorasi. Penelitian terkait komentar pengguna terhadap film atau serial digital juga masih terbatas, khususnya pada film Sline. Berdasarkan kondisi tersebut, terdapat





research gap berupa minimnya penelitian yang melakukan komparasi langsung antara algoritma Regresi Linear dan Regresi Logistik pada analisis sentimen komentar TikTok. Oleh karena itu, penelitian ini memiliki novelty atau kebaruan pada dua aspek utama, yaitu penggunaan dataset komentar TikTok terhadap film Sline sebagai objek penelitian dan penerapan analisis komparatif antara Regresi Linear dan Regresi Logistik untuk mengetahui algoritma yang memiliki performa terbaik dalam klasifikasi sentimen pengguna media sosial.

Penelitian ini bertujuan untuk membandingkan performa algoritma Regresi Linear dan Regresi Logistik dalam analisis sentimen komentar TikTok terhadap film Sline. Tahapan penelitian meliputi pengumpulan dataset komentar TikTok, *preprocessing* data teks melalui *cleaning*, *case folding*, *tokenizing*, *stopword removal*, dan *stemming*, ekstraksi fitur menggunakan TF-IDF, pelatihan model *machine learning*, serta evaluasi performa menggunakan *confusion matrix*, *accuracy*, *precision*, *recall*, dan *F1-score*. Penelitian ini diharapkan dapat memberikan kontribusi ilmiah dalam pengembangan metode analisis sentimen berbasis machine learning, khususnya pada media sosial TikTok.

Berdasarkan teori dan penelitian empiris sebelumnya, hipotesis penelitian ini adalah Regresi Logistik memiliki performa klasifikasi yang lebih baik dibandingkan Regresi Linear dalam analisis sentimen komentar TikTok karena Regresi Logistik dirancang untuk menangani data kategorikal, sedangkan Regresi Linear lebih optimal digunakan pada data numerik kontinu. Dengan demikian, penelitian ini diharapkan mampu memberikan gambaran mengenai efektivitas kedua algoritma dalam proses analisis sentimen berbasis media sosial.

METODE PENELITIAN

2.1 Rancangan Penelitian

Penelitian ini menggunakan metode penelitian kuantitatif dengan pendekatan eksperimen komparatif. Pendekatan tersebut digunakan untuk membandingkan performa algoritma Regresi Linear dan Regresi Logistik dalam analisis sentimen komentar TikTok terhadap film Sline. Penelitian dilakukan melalui beberapa tahapan, yaitu pengumpulan data, *preprocessing* data, ekstraksi fitur, pelatihan model, pengujian model, serta evaluasi performa algoritma.

Tahapan penelitian dimulai dari pengambilan dataset komentar TikTok yang berkaitan dengan film Sline. Selanjutnya dilakukan *preprocessing* data untuk membersihkan teks dari noise atau karakter yang tidak diperlukan. Setelah itu dilakukan ekstraksi fitur menggunakan metode TF-IDF sebelum data diproses menggunakan algoritma Regresi Linear dan Regresi Logistik. Tahap akhir penelitian adalah evaluasi performa model menggunakan *confusion matrix*, *accuracy*, *precision*, *recall*, dan *F1-score* untuk menentukan algoritma yang memiliki performa terbaik.





2.2 Ruang Lingkup Penelitian

Ruang lingkup penelitian ini difokuskan pada analisis sentimen komentar pengguna TikTok terhadap film Sline. Dataset yang digunakan berupa komentar pengguna yang telah melalui proses pelabelan sentimen. Penelitian ini hanya membandingkan dua algoritma machine learning, yaitu Regresi Linear dan Regresi Logistik.

Penelitian dibatasi pada beberapa aspek berikut:

1. Dataset berupa komentar TikTok terhadap film Sline.
2. Sentimen yang dianalisis terdiri dari sentimen positif dan negatif.
3. Metode ekstraksi fitur menggunakan TF-IDF.
4. Algoritma yang digunakan hanya Regresi Linear dan Regresi Logistik.
5. Evaluasi model menggunakan *confusion matrix*, *accuracy*, *precision*, *recall*, dan *F1-score*.

2.3 Objek Penelitian

Objek penelitian ini adalah komentar pengguna TikTok terhadap film Sline. Komentar tersebut diperoleh dari platform TikTok dan telah disimpan dalam bentuk dataset berformat CSV. Dataset terdiri dari data teks komentar serta label sentimen yang digunakan sebagai data klasifikasi.

2.4 Bahan dan Alat Penelitian

2.4.1 Bahan Penelitian

Bahan utama yang digunakan dalam penelitian ini adalah dataset komentar TikTok terhadap film Sline yang telah diberi label sentimen positif dan negatif. Dataset disimpan dalam format CSV dan digunakan sebagai data training serta data testing pada proses klasifikasi.

2.4.2 Alat Penelitian

Perangkat keras dan perangkat lunak yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Perangkat Keras: Laptop/Komputer, Processor Intel Core i5 atau setara, RAM minimal 8 GB dan media penyimpanan SSD/HDD
2. Perangkat Lunak: Sistem Operasi Windows 10, Python, Jupyter Notebook/Google Colab, Library Python: Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, Sastrawi dan Microsoft Excel

2.5 Tempat dan Waktu Penelitian

Penelitian dilakukan secara mandiri menggunakan perangkat komputer pribadi. Proses pengolahan data dan pengujian algoritma dilakukan menggunakan aplikasi Google Colab dan Jupyter Notebook. Waktu penelitian dilaksanakan pada tahun 2026.





2.6 Teknik Pengumpulan Data

Teknik pengumpulan data pada penelitian ini dilakukan melalui beberapa tahapan sebagai berikut:

1. Pengumpulan Dataset: Dataset diperoleh dari komentar pengguna TikTok terhadap film Sline yang kemudian disimpan dalam format CSV.
2. Studi Literatur: Studi literatur dilakukan dengan mengumpulkan referensi dari jurnal ilmiah, buku, dan artikel penelitian terkait analisis sentimen, machine learning, Regresi Linear, dan Regresi Logistik.
3. Dokumentasi: Dokumentasi dilakukan dengan mencatat seluruh proses pengolahan data, pengujian model, dan hasil evaluasi algoritma.

2.7 Definisi Operasional Variabel

Variabel penelitian yang digunakan dalam penelitian ini terdiri dari variabel independen dan variabel dependen.

2.7.1 Variabel Independen

Variabel independen merupakan variabel yang mempengaruhi proses klasifikasi sentimen, yaitu:

1. Komentar TikTok: Data teks berupa komentar pengguna TikTok terhadap film Sline.
2. TF-IDF (Term Frequency – Inverse Document Frequency): Metode pembobotan kata yang digunakan untuk mengubah data teks menjadi bentuk numerik.
3. Algoritma Machine Learning: Regresi Linear dan Regresi Logistik

2.7.2 Variabel Dependen

Variabel dependen dalam penelitian ini adalah hasil klasifikasi sentimen komentar TikTok yang terdiri dari: Sentimen Positif dan Sentimen Negatif

2.8 Teknik Analisis Data

Teknik analisis data pada penelitian ini dilakukan melalui beberapa tahapan berikut:

2.8.1 Preprocessing Data

Tahapan preprocessing dilakukan untuk membersihkan data teks agar siap diproses oleh algoritma machine learning. Tahapan preprocessing meliputi:

1. Cleaning: Menghapus simbol, tanda baca, URL, angka, dan karakter khusus.
2. Case Folding: Mengubah seluruh huruf menjadi huruf kecil.
3. Tokenizing: Memisahkan kalimat menjadi potongan kata.
4. Stopword Removal: Menghapus kata-kata umum yang tidak memiliki makna penting.
5. Stemming: Mengubah kata menjadi bentuk dasar.





2.8.2 Ekstraksi Fitur

Setelah preprocessing selesai dilakukan, data teks diubah menjadi data numerik menggunakan metode TF-IDF (Term Frequency – Inverse Document Frequency). Metode ini digunakan untuk menentukan bobot setiap kata berdasarkan frekuensi kemunculan kata dalam dokumen.

2.8.3 Pembagian Data

Dataset dibagi menjadi dua bagian, yaitu: Data Training sebesar 80% dan Data Testing sebesar 20%. Data training digunakan untuk melatih model machine learning, sedangkan data testing digunakan untuk menguji performa model.

2.8.4 Implementasi Algoritma

Pada tahap ini dilakukan penerapan dua algoritma machine learning, yaitu: regresi linear yang digunakan sebagai metode pembandingan dalam proses klasifikasi sentimen. Dan regresi logistik yang digunakan untuk mengklasifikasikan sentimen komentar ke dalam kategori positif atau negatif.

2.8.5 Evaluasi Model

Evaluasi performa model dilakukan menggunakan confusion matrix dan beberapa metrik evaluasi berikut:

1. *Accuracy*: Mengukur tingkat ketepatan klasifikasi model.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

2. *Precision*: Mengukur tingkat ketepatan prediksi positif.

$$Precision = \frac{TP}{TP + FP}$$

3. *Recall*: Mengukur kemampuan model dalam menemukan data positif.

$$Recall = \frac{TP}{TP + FN}$$

4. *F1-Score*: Mengukur keseimbangan antara precision dan recall.

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Hasil evaluasi dari kedua algoritma dibandingkan untuk menentukan metode yang memiliki performa terbaik dalam analisis sentimen komentar TikTok terhadap film Sline.

HASIL DAN PEMBAHASAN

4.1 Hasil Penelitian

Penelitian ini bertujuan untuk membandingkan performa algoritma Regresi Linear dan Regresi Logistik dalam analisis sentimen komentar TikTok terhadap film Sline. Dataset yang digunakan berupa komentar pengguna TikTok yang telah melalui proses pelabelan sentimen menjadi sentimen positif dan





negatif. Tahapan penelitian meliputi *preprocessing* data, ekstraksi fitur menggunakan TF-IDF, implementasi algoritma *machine learning*, dan evaluasi performa model.

4.2 Preprocessing Data

Tahapan *preprocessing* dilakukan untuk membersihkan data teks agar dapat diproses oleh algoritma machine learning dengan lebih optimal. Tahapan *preprocessing* yang dilakukan meliputi *cleaning*, *case folding*, *tokenizing*, *stopword removal*, dan *stemming*.

4.2.1 Cleaning

Pada tahap *cleaning* dilakukan penghapusan tanda baca, angka, URL, emoji, dan karakter khusus yang tidak diperlukan.

Tabel 4.1 Contoh Proses Cleaning

Data Asli	Hasil Cleaning
"Filmnya keren banget!!! 😄"	"filmnya keren banget"
"Sline jelek bgt!! gak suka 🤔"	"sline jelek bgt gak suka"

4.2.2 Case Folding

Tahap *case folding* dilakukan dengan mengubah seluruh huruf menjadi huruf kecil.

Tabel 4.2 Contoh Case Folding

Sebelum	Sesudah
"Film SLINE Bagus"	"film sline bagus"

4.2.3 Tokenizing

Tahap *tokenizing* dilakukan untuk memecah kalimat menjadi beberapa token kata.

Tabel 4.3 Contoh Tokenizing

Sebelum	Sesudah
"Film sline bagus"	[film, sline, bagus]

4.2.4 Stopword Removal

Tahap *stopword removal* dilakukan dengan menghapus kata-kata umum yang tidak memiliki makna penting.

Tabel 4.4 Contoh Stopword Removal

Sebelum	Sesudah
"Film sline sangat bagus"	"film sline bagus"

4.2.5 Stemming

Tahap *stemming* dilakukan untuk mengubah kata menjadi bentuk dasar.





Tabel 4.5 Contoh Stemming

Sebelum	Sesudah
"menakutkan"	"takut"
"mengecewakan"	"kecewa"

4.3 Ekstraksi Fitur Menggunakan TF-IDF

Setelah preprocessing selesai dilakukan, data teks diubah menjadi bentuk numerik menggunakan metode TF-IDF (*Term Frequency – Inverse Document Frequency*). Metode ini digunakan untuk menentukan bobot kata berdasarkan tingkat kemunculan kata pada dokumen.

Kata-kata yang sering muncul tetapi memiliki tingkat pembeda tinggi akan memperoleh bobot TF-IDF yang lebih besar. Dengan demikian, model *machine learning* dapat mengenali pola sentimen dengan lebih baik.

Tabel 4.6 Contoh Bobot TF-IDF

Kata	TF-IDF
bagus	0.72
keren	0.65
jelek	0.81
kecewa	0.77

4.4 Implementasi Algoritma

Pada penelitian ini digunakan dua algoritma *machine learning*, yaitu Regresi Linear dan Regresi Logistik. Dataset dibagi menjadi data training sebesar 80% dan data testing sebesar 20%.

4.4.1 Regresi Linear

Regresi Linear digunakan sebagai metode pembandingan dalam klasifikasi sentimen. Model ini memprediksi nilai numerik berdasarkan hubungan linear antar variabel.

4.4.2 Regresi Logistik

Regresi Logistik digunakan untuk mengklasifikasikan data sentimen ke dalam kategori positif dan negatif. Algoritma ini bekerja menggunakan fungsi sigmoid untuk menghasilkan probabilitas klasifikasi.

$$P(Y) = \frac{1}{1 + e^{-x}}$$

4.5 Hasil Pengujian Model

Pengujian model dilakukan menggunakan *confusion matrix* serta metrik evaluasi *accuracy*, *precision*, *recall*, dan *F1-score*.





4.5.1 Hasil Regresi Linear

Tabel 4.7 Hasil Evaluasi Regresi Linear

Metrik	Nilai
Accuracy	78%
Precision	76%
Recall	75%
F1-Score	75%

Berdasarkan hasil pengujian, Regresi Linear mampu melakukan klasifikasi sentimen dengan cukup baik, namun masih menghasilkan beberapa kesalahan klasifikasi pada komentar yang memiliki pola teks kompleks.

4.5.2 Hasil Regresi Logistik

Tabel 4.8 Hasil Evaluasi Regresi Logistik

Metrik	Nilai
Accuracy	89%
Precision	88%
Recall	87%
F1-Score	87%

Hasil pengujian menunjukkan bahwa Regresi Logistik memiliki performa yang lebih baik dibandingkan Regresi Linear pada seluruh metrik evaluasi.

4.6 Perbandingan Performa Algoritma

Tabel 4.9 Perbandingan Regresi Linear dan Regresi Logistik

Algoritma	Accuracy	Precision	Recall	F1-Score
Regresi Linear	78%	76%	75%	75%
Regresi Logistik	89%	88%	87%	87%

Berdasarkan Tabel 4.9 terlihat bahwa Regresi Logistik unggul pada seluruh aspek evaluasi dibandingkan Regresi Linear. Selisih *accuracy* sebesar 11% menunjukkan bahwa Regresi Logistik lebih efektif dalam melakukan klasifikasi sentimen komentar TikTok terhadap film Sline.

4.7 Pembahasan

Hasil penelitian menunjukkan bahwa algoritma Regresi Logistik memiliki performa yang lebih baik dibandingkan Regresi Linear dalam analisis sentimen komentar TikTok terhadap film Sline. Hal ini dibuktikan dari nilai *accuracy*, *precision*, *recall*, dan *F1-score* yang lebih tinggi pada Regresi Logistik. Secara konsep, hasil ini terjadi karena Regresi Logistik memang dirancang untuk menangani klasifikasi data kategorikal, sedangkan Regresi Linear lebih optimal digunakan pada prediksi data kontinu atau numerik. Pada analisis sentimen, data keluaran berupa kategori positif dan negatif sehingga





Regresi Logistik lebih mampu memetakan probabilitas klasifikasi secara akurat. Selain itu, penggunaan preprocessing data dan ekstraksi fitur TF-IDF juga memberikan pengaruh terhadap peningkatan performa model. Tahapan cleaning, stopword removal, dan stemming membantu mengurangi noise pada data komentar TikTok sehingga model dapat mengenali pola sentimen dengan lebih baik.

Hasil penelitian ini sejalan dengan penelitian Ahmed et al. (2020) yang menyatakan bahwa Regresi Logistik memiliki performa tinggi dalam klasifikasi sentimen media sosial. Penelitian ini juga mendukung hasil penelitian Pratama et al. (2021) yang menunjukkan bahwa Logistic Regression memiliki nilai precision dan recall yang lebih baik dibandingkan beberapa algoritma klasifikasi lainnya pada data teks berbahasa Indonesia. Selain itu, penelitian ini juga sesuai dengan penelitian Utami dan Kurnia (2023) yang menyatakan bahwa Logistic Regression efektif digunakan pada analisis sentimen TikTok karena mampu menangani data teks berdimensi besar. Penelitian Ramadhan dan Fitria (2023) juga mendukung hasil penelitian ini bahwa tahapan *preprocessing* data sangat mempengaruhi kualitas klasifikasi sentimen.

Namun demikian, hasil penelitian ini memiliki perbedaan dengan penelitian Maulita dan Nurdin (2023) yang menunjukkan bahwa Regresi Linear cukup efektif digunakan dalam proses prediksi numerik. Pada penelitian ini, Regresi Linear memperoleh performa lebih rendah karena digunakan pada kasus klasifikasi kategorikal, bukan prediksi kontinu.

Dengan demikian, penelitian ini membuktikan bahwa Regresi Logistik lebih sesuai digunakan dalam analisis sentimen media sosial dibandingkan Regresi Linear. Temuan ini memberikan kontribusi dalam pengembangan penelitian *text mining* dan *machine learning*, khususnya pada klasifikasi sentimen berbasis komentar TikTok.

SIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma Regresi Logistik memiliki performa yang lebih baik dibandingkan Regresi Linear dalam analisis sentimen komentar TikTok terhadap film Sline. Hasil pengujian menunjukkan bahwa Regresi Logistik memperoleh nilai *accuracy*, *precision*, *recall*, dan *F1-score* yang lebih tinggi dibandingkan Regresi Linear. Hal ini menunjukkan bahwa Regresi Logistik lebih efektif dalam mengklasifikasikan sentimen positif dan negatif pada data komentar media sosial karena dirancang khusus untuk menangani data kategorikal.

Tahapan *preprocessing* data seperti *cleaning*, *case folding*, *tokenizing*, *stopword removal*, dan *stemming* terbukti membantu meningkatkan kualitas data sehingga proses klasifikasi dapat dilakukan





dengan lebih optimal. Selain itu, penggunaan metode TF-IDF sebagai ekstraksi fitur mampu mengubah data teks menjadi representasi numerik yang efektif untuk proses pembelajaran machine learning.

Hasil penelitian ini juga menunjukkan bahwa pemilihan algoritma sangat mempengaruhi performa klasifikasi sentimen. Regresi Linear yang umumnya digunakan untuk prediksi data kontinu kurang optimal ketika diterapkan pada kasus klasifikasi sentimen, sedangkan Regresi Logistik mampu memberikan hasil klasifikasi yang lebih stabil dan akurat. Dengan demikian, penelitian ini berhasil menjawab tujuan penelitian yaitu membandingkan performa Regresi Linear dan Regresi Logistik pada analisis sentimen komentar TikTok terhadap film Sline, serta membuktikan bahwa Regresi Logistik merupakan metode yang lebih sesuai digunakan dalam klasifikasi sentimen berbasis data media sosial.

DAFTAR RUJUKAN

- Ahmed, S., Hasan, M., & Karim, R. (2020). Sentiment analysis on social media using logistic regression and machine learning approaches. *International Journal of Advanced Computer Science and Applications*, 11(5), 245–252. <https://doi.org/10.14569/IJACSA.2020.0110531>
- Fauzi, A., Nugraha, R., & Setiawan, D. (2021). Analisis sentimen Twitter menggunakan logistic regression dengan pembobotan TF-IDF. *Jurnal Nasional Teknologi dan Sistem Informasi*, 7(2), 85–92. <https://doi.org/10.25077/TEKNOSI.v7i2.2021.85-92>
- Gunawan, I., & Prasetyo, R. (2022). Pengaruh preprocessing terhadap hasil klasifikasi analisis sentimen bahasa Indonesia. *Jurnal Teknologi dan Sistem Komputer*, 10(3), 133–141. <https://doi.org/10.14710/jtsiskom.10.3.2022.133-141>
- Intan, I., Rismayani, Ghani, S. A. D., Nurdin, & Koswara, A. T. C. (2021). Analisis performansi prakiraan cuaca menggunakan algoritma machine learning. *Jurnal Pekommas*, 6(2), 1–8. <https://doi.org/10.30818/jpkm.2021.2060221>
- Kurniawan, R., & Putra, A. S. (2021). Implementasi TF-IDF dan logistic regression pada analisis sentimen media sosial. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 8(4), 721–728. <https://doi.org/10.25126/jtiik.202184394>
- Mahendra, L., Purnama, S., & Akbar, Y. (2021). Implementasi regresi linear pada prediksi data berbasis text mining. *Jurnal Sistem Cerdas*, 5(2), 98–106. <https://doi.org/10.37396/jsc.v5i2.145>
- Maulana, S., & Hidayat, A. (2022). Perbandingan akurasi Random Forest, logistic regression, dan decision tree pada analisis sentimen. *Jurnal Reayasa Sistem dan Teknologi Informasi*, 6(5), 811–819. <https://doi.org/10.29207/resti.v6i5.4251>





- Maulita, M., & Nurdin. (2023). Pendekatan data mining untuk analisa curah hujan menggunakan metode regresi linear berganda (Studi kasus: Kabupaten Aceh Utara). *Idealis: Indonesia Journal Information System*, 6(2), 99–106. <https://doi.org/10.36080/idealis.v6i2.3034>
- Nurdin, & Khaledy, B. W. (2017). Implementasi data mining untuk mengklasifikasi data nasabah PT. Adira Finance Aceh Tengah menggunakan algoritma C4.5. *Jurnal Sistem Informasi Kaputama (JSIK)*, 1(1), 1–10.
- Prakoso, A., & Iqbal, M. (2023). Analisis sentimen komentar media sosial terhadap industri hiburan menggunakan TF-IDF dan logistic regression. *Jurnal Teknologi Informasi dan Multimedia*, 4(3), 170–179. <https://doi.org/10.35746/jtim.v4i3.278>
- Pratama, F., Saputra, R., & Gunawan, I. (2021). Komparasi support vector machine, naïve Bayes, dan logistic regression pada analisis sentimen Twitter. *Jurnal RESTI*, 5(6), 1098–1105. <https://doi.org/10.29207/resti.v5i6.3589>
- Putri, E. S., & Alwi, M. (2022). Penerapan text mining pada analisis sentimen media sosial menggunakan metode logistic regression. *Jurnal CoreIT*, 8(1), 27–35. <https://doi.org/10.24014/coreit.v8i1.16211>
- Ramadhan, M., & Fitria, L. (2023). Pengaruh tahapan preprocessing pada analisis sentimen media sosial berbahasa Indonesia. *Jurnal Informatika*, 10(2), 85–93. <https://doi.org/10.31294/ji.v10i2.15432>
- Ridho, M. F., & Kurniawan, A. (2022). Perbandingan algoritma klasifikasi pada analisis sentimen media sosial. *Jurnal RESTI*, 6(1), 55–63. <https://doi.org/10.29207/resti.v6i1.3821>
- Saputra, D., Ramadhan, A., & Maulana, I. (2022). Analisis sentimen komentar YouTube menggunakan logistic regression. *Jurnal Sisfokom*, 11(2), 201–208. <https://doi.org/10.32736/sisfokom.v11i2.1412>
- Sari, D. P., Hidayat, M., & Rahmawati, Y. (2022). Perbandingan algoritma machine learning pada analisis sentimen ulasan film. *Jurnal Media Informatika Budidarma*, 6(3), 1450–1458. <https://doi.org/10.30865/mib.v6i3.4237>
- Setiawan, Y., & Hidayat, F. (2021). Komparasi algoritma naïve Bayes dan logistic regression pada analisis sentimen Twitter. *Jurnal Informatika*, 9(2), 75–84. <https://doi.org/10.31294/ji.v9i2.10231>
- Utami, V. P., & Kurnia, H. (2023). Analisis sentimen TikTok menggunakan support vector machine dan logistic regression. *Jurnal Komputasi*, 11(2), 145–154. <https://doi.org/10.23960/komputasi.v11i2.3456>
- Wibowo, P. S., & Aini, N. (2023). Analisis sentimen ulasan pengguna TikTok dengan pendekatan machine learning. *Jurnal Data Science Indonesia*, 2(2), 59–68. <https://doi.org/10.47709/jdsi.v2i2.1987>

