



Tersedia Online : <http://e-journals.unmul.ac.id/>

ADOPSI TEKNOLOGI DAN SISTEM INFORMASI (ATASI)

Alamat Jurnal : <http://e-journals2.unmul.ac.id/index.php/atasi/index>



Analisis Sentimen Twitter Menggunakan Machine Learning untuk Identifikasi Konten Negatif

Amalia Kartika Sari ¹⁾, Akhmad Irsyad ²⁾*, Dinda Nur Aini ³⁾, Islamiyah ⁴⁾, Stephanie Elfriede Ginting ⁵⁾

Program Studi Sistem Informasi, Universitas Mulawarman

E-Mail : akhmadirsyad@ft.unmul.ac.id

ARTICLE INFO

Article history:

Received : 29 June 2024

Revised : 28 June 2024

Accepted : 09 June 2024

Available online : 30 June 2024

Keywords:

Twitter

Sentimen Analisis

TF-IDF

Logistic Regression

Naive Bayes Classifier

Kata Kunci:

Twitter

Sentimen Analisis

TF-IDF

Logistic Regression

Naive Bayes Classifier

APA style in citing this article:

Kartika Sari, A., Akhmad

Irsyad, Dinda Nur Aini,

Islamiyah, & Stephanie

Elfriede Ginting. Analisis

Sentimen Twitter

Menggunakan Machine

Learning untuk Identifikasi

Konten Negatif. Adopsi

Teknologi Dan Sistem

Informasi (ATASI), 3(1).

<https://doi.org/10.30872/atasi.v3i1.1373>

3i1.1373

ABSTRACT

Negative Content is electronic information and documents that contain disrespectful content, gambling, insults, defamation, extortion, threatening, and spreading false and misleading news that can harm users. In addition, negative content also includes things that trigger hatred or hostility based on ethnicity, religion, race, and class. Twitter is a social media platform that allows sending and reading messages used by individuals, organizations and companies to share information, comment on current events, and interact with other users. The analysis process includes data cleaning, data exploration, and application of machine learning models such as TF-IDF, Logistics Regression Model, and Naive Bayes Classifier for sentiment classification. The results show that machine learning models can classify the sentiment of tweets very accurately, providing important insights into public opinion on the Twitter platform. The best performance with an F1-Score of 68.4% was generated from the Logistic Regression trial.

ABSTRAK

Konten Negatif adalah informasi dan dokumen elektronik yang mengandung konten yang tidak sopan, perjudian, penghinaan, pencemaran nama baik, pemerasan, pengancaman, serta penyebaran berita bohong dan menyesatkan yang dapat merugikan pengguna. Selain itu, konten negatif juga mencakup hal-hal yang memicu kebencian atau permusuhan berdasarkan suku, agama, ras, dan golongan (Alita & Isnain, 2020). Twitter adalah Wadah media sosial yang memungkinkan mengirim dan membaca pesan yang digunakan oleh individu, organisasi dan perusahaan untuk berbagi informasi, mengomentari peristiwa terkini, dan berinteraksi dengan pengguna lain. Proses analisis mencakup pembersihan data, eksplorasi data, dan penerapan model pembelajaran mesin seperti TF-IDF, Logistics Regression Model, dan Naive Bayes Classifier untuk klasifikasi sentimen. Hasil penelitian menunjukkan bahwa model pembelajaran mesin dapat mengklasifikasikan sentimen tweet dengan sangat akurat, memberikan wawasan penting tentang opini publik di platform Twitter. Kinerja terbaik dengan skor F1 68,4% dihasilkan dari uji coba Logistic Regression.

2024 Adopsi Teknologi dan Sistem Informasi (ATASI) with CC BY SA license.

*) Corresponding Author

<https://doi.org/10.30872/atasi.v3i1.1373>

2024 Adopsi Teknologi dan Sistem Informasi (ATASI) with CC BY SA license.

1. PENDAHULUAN

Di era digital ini, media sosial telah menjadi platform utama bagi individu untuk berbagi pemikiran, pendapat, dan pengalaman secara cepat dan luas. Twitter, sebagai salah satu platform media sosial terdepan, memfasilitasi interaksi sosial dengan memungkinkan pengguna untuk mengirimkan pesan singkat yang dikenal sebagai "tweet". Namun, dengan kebebasan berekspresi yang dimiliki oleh pengguna, seringkali terjadi penyalahgunaan platform ini. Konten berisi kebencian, ujaran kebencian, dan sentimen negatif lainnya dapat tersebar dengan mudah, berpotensi menimbulkan dampak negatif bagi pengguna lain (Multi Fani & Santoso, 2023).

Upaya Twitter untuk memerangi konten negatif di platformnya membutuhkan solusi yang cerdas dan efektif. Salah satu langkah penting adalah dengan mengembangkan model klasifikasi berbasis Natural Language Processing (NLP) yang dapat membedakan tweet negatif dan memblokir konten tersebut. Model klasifikasi ini berperan penting dalam menjaga keamanan dan kenyamanan pengguna Twitter. Dengan mengidentifikasi dan memblokir konten negatif, Twitter dapat menciptakan lingkungan yang lebih positif dan kondusif bagi penggunaannya untuk berinteraksi dan berbagi informasi.

Dalam upaya mencapai tujuan tersebut, analisis sentimen memiliki peran krusial. Analisis sentimen, sering dikenal sebagai penambangan opini, melihat bagaimana perasaan individu terhadap suatu hal tertentu. Hal ini juga dikenal sebagai penambangan teks opini. Subbidang linguistik komputasi dan NLP ini berkaitan dengan teknik untuk mengekstraksi, mengkategorikan, memahami, dan menilai pendapat yang diungkapkan dalam publikasi online. Dengan mengintegrasikan analisis sentimen dalam strategi pemberantasan konten negatif di Twitter, platform ini dapat lebih efektif menciptakan lingkungan yang positif dan aman bagi interaksi penggunaannya.

2. TINJAUAN PUSTAKA

Analisis sentimen di media sosial, terutama di platform seperti Twitter, telah menjadi fokus penelitian yang signifikan dalam berbagai bidang seperti politik, ekonomi, dan lain-lain. Penelitian-penelitian ini sering mengadopsi algoritma machine learning untuk mengklasifikasikan sentimen pengguna terhadap berbagai topik. Di antara penelitian-penelitian yang relevan, beberapa yang menonjol adalah sebagai berikut:

(Drus & Khalid, 2019) mengidentifikasi dua pendekatan utama dalam analisis sentimen: pendekatan berbasis kamus dan pendekatan berbasis pembelajaran mesin. Pendekatan berbasis kamus menggunakan daftar kata yang telah ditentukan sebelumnya untuk menghitung kata-kata positif dan negatif, sementara pendekatan pembelajaran mesin memanfaatkan algoritma untuk mengekstraksi dan mendeteksi sentimen dari data

(Sahayak, Shete, & Pathan, 2015), dalam jurnal *International Journal of Innovative Research in Advanced Engineering (IJIRAE)*, mengeksplorasi analisis sentimen pada data Twitter dengan menggunakan berbagai algoritma machine learning seperti Naive Bayes, Maximum Entropy (MaxEnt), dan Support Vector Machines (SVM). Mereka menggunakan data tweet yang diklasifikasikan sebagai positif, negatif, atau netral, serta memanfaatkan dictionary yang dianotasi secara manual untuk emoticon dan akronim sebagai label yang bising. Penelitian ini menemukan bahwa algoritma SVM dan MaxEnt memiliki performa yang lebih baik dibandingkan Naive Bayes dalam menangani masalah overlapping fitur dan klasifikasi sentimen yang kompleks.

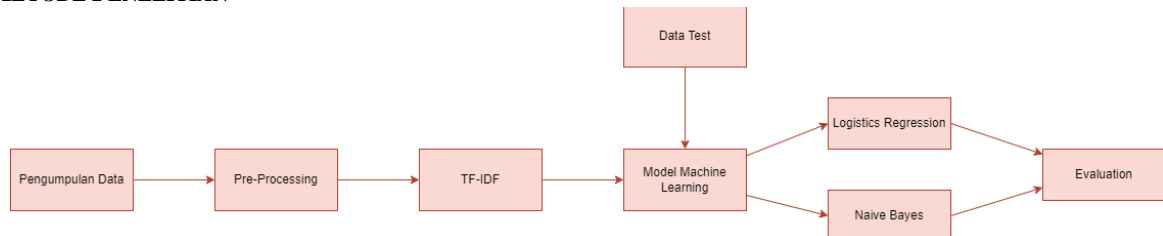
Agarwal et al. (2014) dari Columbia University memperkenalkan fitur prior polarity yang spesifik untuk Part-of-Speech (POS) dan mengusulkan penggunaan tree kernel untuk mengurangi kebutuhan akan rekayasa fitur yang melelahkan. Mereka mengembangkan model untuk mengklasifikasikan tweet menjadi sentimen positif, negatif, dan netral. Hasil penelitian ini menunjukkan bahwa model tree kernel dan model berbasis fitur mereka, yang menggunakan 100 fitur, dapat mengungguli baseline unigram yang menggunakan lebih dari 10.000 fitur. Mereka juga menyoroti pentingnya fitur yang menggabungkan polaritas awal kata dengan tag POS untuk tugas klasifikasi sentimen. Selain itu, penelitian ini menyediakan dataset yang dianotasi secara manual yang mewakili sampel acak dari streaming tweet, berbeda dengan dataset yang dikumpulkan berdasarkan kueri pencarian tertentu.

Hassan Saif dan rekan-rekannya (2012) dari Knowledge Media Institute, The Open University, memperkenalkan pendekatan baru dengan menambahkan konsep semantik sebagai fitur tambahan dalam analisis sentimen. Mereka menyelidiki korelasi antara konsep semantik entitas yang diekstrak dari tweet dengan sentimen negatif/positif. Hasil penelitian mereka menunjukkan peningkatan rata-rata skor akurasi F harmonic sekitar 6.5% untuk sentimen negatif dan 4.8% untuk sentimen positif dibandingkan dengan baseline yang menggunakan unigrams dan fitur part-of-speech.

(Shah, Akshata Bhat, Sumitra Singh, Arya Chavan, & Aryan Singh, 2024) mencatat bahwa analisis sentimen memiliki berbagai keuntungan, seperti untuk riset pasar, dukungan pelanggan, manajemen merek, analisis politik, dan sektor kesehatan. Namun, ada tantangan signifikan yang harus dihadapi, seperti kesulitan dalam mendeteksi nada atau sarkasme dalam teks, penanganan bahasa gaul dan informal, serta keterbatasan dalam menganalisis konten multibahasa dan media non-teks.

Penelitian-penelitian ini memberikan kontribusi yang signifikan dalam pemahaman kita tentang analisis sentimen di platform media sosial, dan memberikan dasar yang kuat untuk melakukan penelitian dalam bidang ini.

3. METODE PENELITIAN



Gambar 1. Alur Penelitian

Metode penelitian ini dimulai dengan pengumpulan data, yang kemudian dilakukan preprocessing untuk membersihkan dan mempersiapkan data. Setelah itu, data train diekstraksi fiturnya menggunakan metode TF-IDF. Metode ini menilai bobot kata dalam dokumen secara proporsional. Data yang telah diproses selanjutnya dianalisis menggunakan dua model machine learning, yakni Logistic Regression dan Naive Bayes. Langkah terakhir, model dievaluasi dengan menggunakan data uji, di mana akurasi, presisi, recall, dan F1-score digunakan sebagai metrik untuk menilai kinerja model dalam mengidentifikasi dan mengklasifikasikan konten negatif di Twitter.

A. Preprocessing

Preprocessing adalah tahap awal dalam analisis data teks yang bertujuan untuk membersihkan dan mempersiapkan data sehingga siap digunakan oleh model machine learning. Tahapan preprocessing meliputi case folding, normalisasi kata, cleaning, stopword removal, dan stemming (Irham, Adiwijaya, & Wisesty, 2019). Berikut penjelasan masing-masing tahapan preprocessing:

1. Case Folding:

Case folding adalah proses mengubah semua huruf dalam teks menjadi huruf kecil (lowercase) (Multi Fani & Santoso, 2023). Hal ini dilakukan untuk memastikan bahwa perbedaan dalam kapitalisasi tidak mempengaruhi analisis teks. Misalnya, "Data" dan "data" akan dianggap sama setelah case folding. Hal ini membantu dalam menyederhanakan data dan mengurangi variasi kata yang tidak diperlukan.

2. Normalisasi Kata:

Normalisasi kata adalah proses mengubah berbagai bentuk kata yang berbeda menjadi bentuk standar yang sama. Hal ini sering melibatkan penanganan kata-kata yang memiliki makna sama tetapi ditulis dengan cara berbeda. Contohnya adalah mengubah "gak" menjadi "tidak", "nggak" menjadi "tidak", atau "ya" menjadi "iya". Normalisasi membantu dalam konsistensi data dan meningkatkan akurasi analisis.

3. Cleaning:

Cleaning atau pembersihan adalah proses menghapus karakter yang tidak relevan atau tidak diinginkan dari teks. Ini termasuk penghapusan tanda baca, angka, simbol, URL, dan emotikon. Misalnya, teks "Hello!!! How are you? Visit http://example.com" akan dibersihkan menjadi "Hello How are you Visit". Cleaning bertujuan untuk menghilangkan noise dan memastikan bahwa hanya data yang bermakna yang dianalisis.

4. Stopword Removal:

Stopword removal adalah proses menghapus kata-kata umum yang sering muncul dalam teks tetapi biasanya tidak memberikan informasi penting. Contohnya termasuk kata-kata seperti "dan", "atau", "yang", "itu" dalam bahasa Indonesia, atau "the", "is", "in", "and" dalam bahasa Inggris. Menghapus stopwords membantu mengurangi jumlah fitur yang tidak penting dan fokus pada kata-kata yang lebih bermakna.

5. Stemming:

Stemming adalah proses mengurangi kata ke bentuk dasarnya (stem) dengan menghapus imbuhan seperti akhiran atau awalan. Misalnya, kata "running", "runner", dan "ran" dalam bahasa Inggris semua dikurangi menjadi "run". Dalam bahasa Indonesia, kata "bermain", "bermainan", "dimainkan" dikurangi menjadi "main". Stemming membantu mengurangi variasi kata dan mengkonsolidasikan kata-kata yang memiliki makna dasar yang sama.

B. TF-IDF

TF-IDF (Term Frequency-Inverse Document Frequency) adalah teknik yang digunakan untuk mengubah teks menjadi representasi numerik yang dapat digunakan oleh model machine learning. Metode ini menilai pentingnya sebuah kata dalam sebuah dokumen relatif terhadap kumpulan dokumen. Term Frequency (TF) adalah kemunculan sebuah kata dalam suatu dokumen sedangkan Inverse Document Frequency (IDF) adalah jumlah seluruh dokumen yang mengandung kata tertentu (Irhama et al., 2019). Dengan menggabungkan TF dan IDF, TF-IDF dapat menghitung total bobot dari kata dalam sebuah dokumen, memberikan nilai yang lebih tinggi untuk kata-kata yang sering muncul dalam satu dokumen tetapi jarang muncul di dokumen lain.

1. Term Frequency (TF): TF mengukur frekuensi kemunculan sebuah kata (term) dalam dokumen. Rumusnya adalah:

$$TF(t, d) = \frac{\text{Jumlah kemunculan term } t \text{ dalam dokumen } d}{\text{total jumlah term dalam dokumen } d}$$

TF adalah rasio antara jumlah kemunculan term t dalam dokumen d dengan total jumlah term dalam dokumen tersebut, mengukur seberapa sering sebuah term muncul dalam dokumen.

2. Inverse Document Frequency (IDF): IDF mengukur seberapa banyak dokumen dalam koleksi D yang mengandung term t . Rumusnya adalah:

$$IDF(t, D) = \log \frac{\text{Total jumlah dokumen dalam koleksi } D}{\text{jumlah dokumen yang mengandung term } t + 1}$$

IDF adalah logaritma dari rasio total jumlah dokumen dalam koleksi D dengan jumlah dokumen yang mengandung term t , ditambah satu. Ini memberikan bobot lebih tinggi pada term yang jarang muncul di banyak dokumen, mengurangi pengaruh term yang umum.

3. TF-IDF: TF-IDF adalah hasil perkalian antara Term Frequency (TF) dan Inverse Document Frequency (IDF). Rumusnya adalah:

$$TF - IDF(t, d, D) = TF(t, d) \times IDF(t, D)$$

TF-IDF menghitung total bobot dari kata dalam sebuah dokumen, memberikan nilai yang lebih tinggi untuk kata-kata yang sering muncul dalam satu dokumen tetapi jarang muncul di dokumen lain.

Dalam rumus tersebut:

- t adalah term atau kata kunci tertentu.
- d adalah dokumen tertentu.
- D adalah koleksi seluruh dokumen.

C. Logistic Regression

Logistic Regression adalah metode statistik yang digunakan untuk memprediksi probabilitas kejadian biner, yaitu dua hasil yang mungkin, seperti "ya" atau "tidak". Metode ini digunakan dalam machine learning untuk masalah klasifikasi, di mana hasil variabel target adalah kategori yang berbeda. Logistic Regression menggunakan fungsi logistik untuk mengubah nilai output menjadi probabilitas antara 0 dan 1 (Alfyando, Anggraeny, & Sihananto, 2024). Fungsi ini memungkinkan model untuk memetakan setiap kombinasi input ke dalam rentang probabilitas. Misalnya, jika kita ingin mengklasifikasikan apakah sebuah tweet bersentimen positif atau negatif, Logistic Regression dapat digunakan untuk memberikan probabilitas bahwa tweet tersebut

termasuk ke dalam salah satu kategori tersebut. Model ini efektif untuk masalah klasifikasi sederhana dan memberikan interpretasi yang mudah dipahami mengenai pengaruh tiap fitur terhadap hasil prediksi.

D. Naive Bayes

Naive Bayes adalah metode klasifikasi berbasis teorema Bayes dengan asumsi bahwa setiap fitur dalam dataset adalah independen satu sama lain. Meskipun asumsi ini jarang berlaku sepenuhnya dalam praktik, metode ini sering memberikan hasil yang baik dan efisien untuk berbagai tugas klasifikasi teks, seperti analisis sentimen atau spam detection. Naive Bayes menghitung probabilitas bahwa sebuah data masuk ke dalam kelas tertentu berdasarkan nilai-nilai fitur yang ada. Misalnya, dalam analisis sentimen, Naive Bayes dapat digunakan untuk menghitung probabilitas bahwa sebuah tweet bersentimen positif atau negatif berdasarkan kemunculan kata-kata tertentu dalam tweet tersebut (Alfyando et al., 2024). Keuntungan utama dari Naive Bayes adalah kesederhanaannya dan kemampuannya untuk bekerja dengan baik bahkan dengan data terbatas.

E. Evaluation

Evaluasi kinerja model dalam analisis sentimen melibatkan penilaian hasil eksperimen sistem serta tanggapan dari para responden. Metode evaluasi ini menggunakan berbagai metrik standar seperti akurasi, presisi, recall, F1-Score, dan confusion matrix (NOR, MUSLIM, & ASWIN, 2022). Akurasi menggambarkan performa keseluruhan model, sedangkan presisi dan recall mengukur ketepatan prediksi dan kemampuan model dalam menemukan kasus positif. F1-Score memberikan keseimbangan antara kedua metrik tersebut. Confusion matrix memberikan wawasan tentang jenis kesalahan yang dibuat oleh model. Dengan menggunakan metrik-metrik evaluasi ini secara bersama-sama, evaluasi memberikan gambaran yang komprehensif tentang kinerja model dalam mengidentifikasi sentimen, mendukung pengambilan keputusan yang akurat dalam penggunaan model di dunia nyata.

Evaluasi kinerja model dalam analisis sentimen melibatkan penilaian hasil eksperimen sistem serta tanggapan dari para responden. Metode evaluasi ini menggunakan berbagai metrik standar seperti akurasi, presisi, recall, F1-Score, dan confusion matrix. Akurasi menggambarkan performa keseluruhan model, sedangkan presisi dan recall mengukur ketepatan prediksi dan kemampuan model dalam menemukan kasus positif. F1-Score memberikan keseimbangan antara kedua metrik tersebut. Confusion matrix memberikan wawasan tentang jenis kesalahan yang dibuat oleh model. Dengan menggunakan metrik-metrik evaluasi ini secara bersama-sama, evaluasi memberikan gambaran yang komprehensif tentang kinerja model dalam mengidentifikasi sentimen, mendukung pengambilan keputusan yang akurat dalam penggunaan model di dunia nyata.

- Akurasi (Accuracy): Persentase prediksi yang benar dari semua prediksi yang dibuat oleh model. Rumusnya:

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

- Presisi (Precision): Proporsi prediksi positif yang benar dari semua prediksi positif yang dibuat oleh model. Rumusnya:

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

- Recall (Recall): Proporsi prediksi positif yang benar dari semua kasus sebenarnya positif. Rumusnya:

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

- F1-Score: Harmonik rata-rata dari presisi dan recall, memberikan keseimbangan antara keduanya. Rumusnya:

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

4. HASIL DAN PEMBAHASAN

A. Dataset

Dataset yang saya digunakan dalam penelitian ini adalah "Twitter Tweets Sentiment Dataset" oleh M Yasser H. Dataset ini berasal dari Kaggle dan berfokus pada analisis sentimen tweet di platform media sosial Twitter. Setiap baris dalam dataset ini berisi teks dari tweet dan label sentimen yang terkait.

Dataset ini memiliki beberapa kolom penting, yaitu textID (ID unik untuk setiap teks), text (teks dari tweet), sentiment (sentimen umum dari tweet seperti positif, negatif, atau netral), dan selected_text (kata atau frasa dari tweet yang mencerminkan sentimen yang diberikan). Untuk memastikan kualitas data, saat parsing file CSV, tanda kutip pada awal dan akhir teks dihapus. Data juga dibersihkan dari duplikasi, nilai yang hilang, serta karakter atau simbol yang tidak diperlukan. Setelah pembersihan, data dibagi menjadi dua set: 90% untuk data pelatihan dan 10% untuk data uji, guna memastikan model dapat dilatih dengan baik dan diuji keakuratannya pada data yang tidak dilihat sebelumnya. Dataset ini bertujuan untuk memahami dan membersihkan data, serta membangun model klasifikasi yang efektif dalam memprediksi sentimen tweet.

B. Preprocessing

Pre-processing adalah proses transformasi data mentah menjadi format yang lebih cocok untuk analisis. Dalam konteks analisis teks dan Natural Language Processing (NLP), pre-processing melibatkan beberapa langkah penting, yaitu:

1. Case Folding

Case folding adalah proses mengubah semua huruf dalam teks menjadi huruf kecil. Proses ini penting untuk memastikan bahwa kata-kata yang sama dalam berbagai bentuk huruf (huruf besar atau kecil) diperlakukan sebagai entitas yang sama.

Tabel 1. Penerapan *Case Folding*

Sebelum Case Folding	Sesudah Case Folding
I`d have responded, if I were going	i`d have responded, if i were going

2. Cleansing

Cleansing adalah proses pembersihan teks dari karakter atau simbol yang tidak diperlukan. Langkah ini mencakup penghapusan tanda baca, angka, URL, karakter khusus, dan elemen-elemen lain yang tidak relevan.

Tabel 2. Penerapan *Cleansing*

Sebelum Cleansing	Sesudah Cleansing
my boss is bullying me...	my boss is bullying me

3. Stopword Removal

Stopword removal adalah proses menghapus kata-kata umum yang sering muncul tetapi tidak memiliki makna signifikan dalam analisis, seperti "the" "is" "in" dan sejenisnya.

Tabel 3. Penerapan *Stopword Removal*

Sebelum Cleansing	Sesudah Cleansing
what interview! leave me alone	interview! leave alone

C. TF-IDF

Setelah data bersih, selanjutnya digunakan teknik TF-IDF (Term Frequency-Inverse Document Frequency) untuk mengubah teks mentah dari tweet menjadi fitur numerik yang dapat digunakan oleh model pembelajaran mesin. Dengan menerapkan `TfidfVectorizer` dari library `sklearn`, teks dalam kolom 'text' diubah menjadi fitur. Proses ini membantu dalam menilai pentingnya setiap kata dalam tweet relatif terhadap kumpulan tweet lainnya, sehingga kata-kata yang sering muncul namun tidak memiliki makna signifikan dapat diminimalkan. Fitur yang dihasilkan kemudian digunakan sebagai input untuk model klasifikasi, yang dibagi menjadi data pelatihan dan data pengujian dengan proporsi 80% untuk pelatihan dan 20% untuk pengujian. Transformasi ini bertujuan untuk meningkatkan efisiensi dan akurasi model dalam memprediksi sentimen tweet.

D. Model Machine Learning

Pada tahap ini, perhitungan hasil kinerja model Logistics Regression dan Naive Bayes Classifier dilakukan. Kinerja yang diperhitungkan meliputi Accuracy, Precision, Recall, F1-Score, dan Support. Accuracy digunakan untuk menghitung persentase tweet yang diklasifikasi dengan benar dari keseluruhan tweet. Precision digunakan untuk mengukur ketepatan model dalam memprediksi kelas (Negative, Neutral, dan Positive) dengan menunjukkan prediksi yang benar untuk masing-masing kelas. Untuk setiap kelas, Support adalah contoh aktual, dan F1-Score adalah ukuran gabungan dari Precision dan Recall (Irsyad, Geralda, & Wardhana, 2023).

1. Logistics Regression Model

Tabel 4. Hasil Evaluasi Logistics Regression Model

Accuracy	68.5%
F1 Score	68.4%

Hasil evaluasi model Logistic Regression (LR) menunjukkan bahwa akurasi model, yang merupakan persentase prediksi yang benar dari keseluruhan data, adalah sebesar 68,5%. Ini menunjukkan bahwa model mampu memprediksi sentimen tweet dengan benar sebesar 68,5% dari total data yang dievaluasi. Selain itu, nilai F1 model, yang merupakan rata-rata harmonik dari presisi dan recall, adalah sebesar 68,4%. Nilai F1 memberikan gambaran luas tentang kualitas prediksi model, dengan nilai yang baik.

Tabel 5. Classification Report Logistics Regression Model

	Precision	Recall	F1-Score	Support
Negative	0.73	0.56	0.63	1523
Neutral	0.62	0.78	0.69	2275
Positive	0.78	0.67	0.72	1698
Accuracy			0.69	5496
Macro Avg	0.71	0.67	0.68	5496
Weighted Avg	0.70	0.69	0.68	5496

Hasil dari classification report menunjukkan bahwa model LR memiliki tingkat presisi (precision) yang beragam untuk setiap kelas sentimen, dengan presisi tertinggi pada kelas positive (0.78) dan terendah pada kelas neutral (0.62). Selain itu, recall model LR juga bervariasi, dengan recall tertinggi pada kelas neutral (0.78) dan terendah pada kelas negative (0.56). F1-Score, yang merupakan rata-rata harmonik dari presisi dan recall, menunjukkan kinerja model dalam memprediksi setiap kelas sentimen. Meskipun model LR menunjukkan kinerja yang cukup baik dengan F1-Score di atas 0.5 untuk setiap kelas, terdapat perbedaan yang signifikan antara kelas sentimen negative, neutral, dan positive.

2. Naive Bayes Classifier

Tabel 6. Hasil Evaluasi Naive Bayes Classifier

Accuracy	63.1%
F1 Score	62.4%

Hasil evaluasi model Naive Bayes Classifier (NB) menunjukkan beberapa hal menarik. Nilai akurasi model NB adalah 63,1%, yang berarti model ini dapat memprediksi sentimen tweet dengan benar sebanyak 63,1% dari total data uji. Sementara itu, nilai F1 Score model NB adalah 62,4%, yang merepresentasikan kualitas prediksi model secara keseluruhan.

Tabel 7. Classification Report Naive Bayes Classifier

	Precision	Recall	F1-Score	Support
Negative	0.71	0.43	0.53	1523
Neutral	0.56	0.79	0.65	2275
Positive	0.75	0.60	0.67	1698
Accuracy			0.63	5496
Macro Avg	0.67	0.61	0.62	5496
Weighted Avg	0.66	0.63	0.62	5496

Dalam laporan klasifikasi, kinerja model NB berbeda untuk setiap kelas sentimen. Presisi (ketelitian) model NB paling tinggi pada kelas positif (0,75), diikuti oleh kelas negatif (0,71), dan kelas netral terendah (0,56). Kemampuan model untuk memprediksi tweet yang benar-benar termasuk ke dalam masing-masing kelas dengan akurat. Pada Recall, tweet kelas netral memiliki nilai tertinggi (0,79), sedangkan tweet kelas negatif memiliki nilai terendah (0,43). Ini menunjukkan bahwa model NB lebih mampu mengidentifikasi tweet kelas netral daripada tweet kelas negatif. Selain itu, nilai F1 Score untuk setiap kelas menunjukkan keseimbangan antara presisi dan recall. Nilai F1 tertinggi adalah untuk kelas positif (0,67), sedangkan nilai F1 terendah adalah untuk kelas negatif (0,53).

E. Evaluation

Pada tahap evaluation ini, hasil perhitungan dari Logistic Regression (LR) dan Naive Bayes Classifier akan dibandingkan satu sama lain berdasarkan Accuracy, Precision, Recall, dan F1-Score. Hasil evaluasi model Logistic Regression (LR) dan Naive Bayes Classifier (NB) dalam melakukan prediksi sentimen tweet dapat dikatakan bagus berdasarkan beberapa pertimbangan dan juga dibuktikan melalui matriks konfusi. Matriks konfusi tersebut yang nantinya akan divisualisasikan menggunakan heatmap.

Tabel 8. Kinerja Model Machine Learning

	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Logistic Regression (LR)	68.5	69.9	68.5	68.4
Naive Bayes Classifier (NB)	63.1	65.8	63.1	62.4

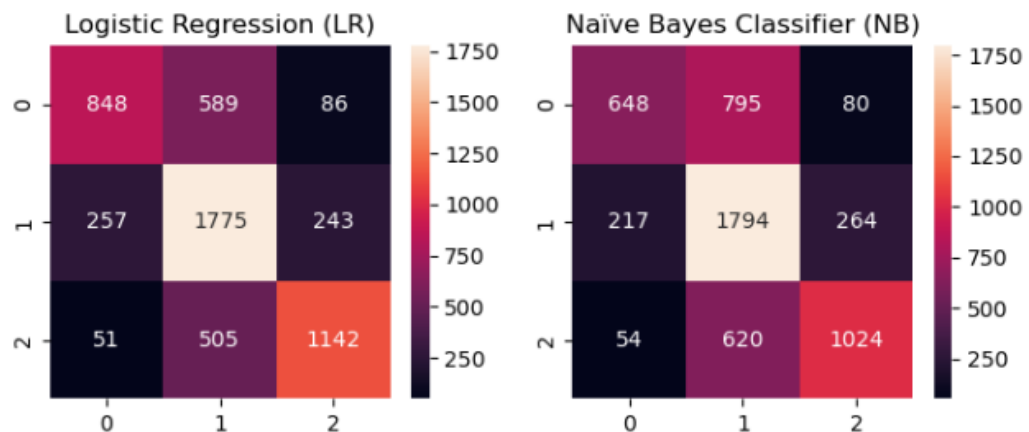
Dengan akurasi sebesar 68,5%, model regresi logistik dapat memprediksi sentimen tweet dengan tingkat kebenaran yang sangat tinggi, yang merupakan 68,5% dari total data uji. Ini menunjukkan bahwa model LR memiliki kemampuan klasifikasi yang kuat, di mana lebih dari dua tweet dari tiga dapat diklasifikasikan dengan benar. Sementara itu, akurasi model Naive Bayes Classifier lebih rendah, 63,1%, tetapi masih cukup baik. Akurasi model NB lebih rendah 5,4 poin persentase dibandingkan model LR, perbedaan ini menunjukkan bahwa model LR secara keseluruhan memiliki kemampuan prediksi sentimen yang lebih baik daripada model NB.

Ketika model LR memperkirakan bahwa suatu tweet akan masuk ke dalam kelas tertentu, 69,9% dari perkiraannya benar. Angka presisi yang tinggi ini menunjukkan bahwa model LR sangat akurat dalam menentukan kelas sentimen yang tepat untuk tweet. Sebaliknya, model Naive Bayes Classifier memiliki presisi yang lebih rendah, 65,8%, tetapi masih dianggap baik. Perbedaan presisi 4,1 poin persentase antara LR dan NB menunjukkan bahwa model LR lebih akurat dalam menentukan tweet yang benar-benar termasuk ke dalam kelas yang diprediksi.

Logistic Regression dan Naive Bayes Classifier masing-masing memiliki nilai recall yang sama, 68,5% dan 63,1%, masing-masing. Nilai recall menunjukkan seberapa banyak tweet yang benar-benar termasuk ke dalam suatu kelas yang dapat ditemukan oleh model. Oleh karena itu, dapat dikatakan bahwa

kemampuan kedua model untuk menemukan tweet yang sesuai dengan kelasnya adalah sama dan sangat baik.

Model Logistic Regression memiliki skor F1 yang lebih tinggi, 68,4%, dibandingkan dengan model Naive Bayes Classifier, yang hanya 62,4%. Perbedaan yang signifikan ini, hingga 6 poin persentase, menunjukkan bahwa model LR secara keseluruhan lebih baik dan lebih seimbang dalam memprediksi sentimen tweet.



Gambar 2. Heatmap Koefisien Matriks

Heatmap di atas menunjukkan bagaimana kedua model klasifikasi Logistic Regression (LR) dan Naive Bayes Classifier (NB) bekerja saat mengkategorikan tweet ke dalam tiga kategori sentimen: negatif (0), netral (1), dan positif (2). Pada model Logistic Regression, 848 tweet diklasifikasikan dengan benar sebagai negatif, 1775 sebagai netral, dan 1142 sebagai positif. Namun, terdapat beberapa kesalahan dalam mengklasifikasikan, seperti 589 tweet negatif diklasifikasikan sebagai netral dan 243 tweet netral diklasifikasikan sebagai positif. Pada model Naive Bayes, terdapat lebih banyak kesalahan dalam mengklasifikasikan, terutama dalam kategori negatif dan netral hanya 648 tweet negatif diklasifikasikan dengan benar sebagai negatif dan 243 tweet netral diklasifikasikan sebagai positif. Secara keseluruhan, dibandingkan dengan Naive Bayes, Logistic Regression bekerja lebih baik dalam mengklasifikasikan ketiga kelas sentimen.

5. KESIMPULAN

Berdasarkan hasil dan pembahasan jurnal sebelumnya, dapat disimpulkan bahwa model Logistic Regression (LR) lebih unggul dalam memprediksi sentimen tweet dibandingkan dengan Naive Bayes Classifier (NB). Meskipun keduanya memiliki performa yang cukup baik, model LR menunjukkan akurasi, presisi, dan F1-score yang lebih tinggi. Oleh karena itu, penggunaan model Logistic Regression (LR) dalam analisis sentimen tweet dapat dianggap efektif dan sesuai dengan tujuan penelitian. Dengan demikian, hasil analisis data mendukung kesimpulan bahwa model LR efektif dalam memprediksi sentimen tweet, sesuai dengan tujuan penelitian yang telah ditetapkan (Giovani, Ardiansyah, Haryanti, Kurniawati, & Gata, 2020).

6. DAFTAR PUSTAKA

- Alfyando, M., Anggraeny, F. T., & Sihananto, A. N. (2024). Perbandingan Algoritma Random Forest dan Logistic Regression Untuk Analisis Sentimen Ulasan Aplikasi Tumbuh Kembang Anak Di Play Store. *Jurnal Sistem Informasi Dan Ilmu Komputer*, 2(1), 77–86. <https://doi.org/10.59581/jusiik-widyakarya.v2i1.2262>
- Alita, D., & Isnain, A. R. (2020). Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier. *Jurnal Komputasi*, 8(2). <https://doi.org/10.23960/komputasi.v8i2.2615>
- Drus, Z., & Khalid, H. (2019). Sentiment Analysis in Social Media and Its Application: Systematic Literature Review. *Procedia Computer Science*, 161, 707–714. <https://doi.org/10.1016/j.procs.2019.11.174>
- Giovani, A. P., Ardiansyah, A., Haryanti, T., Kurniawati, L., & Gata, W. (2020). ANALISIS SENTIMEN APLIKASI RUANG GURU DI TWITTER MENGGUNAKAN ALGORITMA KLASIFIKASI. *Jurnal Teknoinfo*, 14(2), 115. <https://doi.org/10.33365/jti.v14i2.679>
- Irham, L. G., Adiwijaya, A., & Wisesty, U. N. (2019). Klasifikasi Berita Bahasa Indonesia Menggunakan Mutual Information dan Support Vector Machine. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 3(4), 284–292. <https://doi.org/10.30865/mib.v3i4.1410>
- Irsyad, A., Geralda, R. D., & Wardhana, R. (2023). Analisis Sentimen SEA Games 2023 di Twitter Metode dengan Machine Learning. *Adopsi Teknologi Dan Sistem Informasi (ATASI)*, 2(2), 126–131. <https://doi.org/10.30872/atasi.v2i2.1138>

- Multi Fani, S., & Santoso, R. (2023). PENERAPAN TEXT MINING UNTUK MELAKUKAN CLUSTERING DATA TWEET AKUN BLIBLI PADA MEDIA SOSIAL TWITTER MENGGUNAKAN K-MEANS CLUSTERING. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 10, 583–593. Retrieved from <https://ejournal3.undip.ac.id/index.php/gaussian/>
- NOR, S. N., MUSLIM, M. A., & ASWIN, M. (2022). Pengenalan Pola Dasar Angka berdasarkan Gerakan Tangan menggunakan Machine Learning. *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*, 10, 596–608.
- Sahayak, V., Shete, V., & Pathan, A. (2015). Sentiment Analysis on Twitter Data. *International Journal of Innovative Research in Advanced Engineering (IJIRAE)*, 2(1), 178–179. Retrieved from www.ijirae.com
- Shah, S., Akshata Bhat, M., Sumitra Singh, M., Arya Chavan, M., & Aryan Singh, M. (2024). INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS) SENTIMENT ANALYSIS. *SENTIMENT ANALYSIS*, 4(4), 1542–1546. <https://doi.org/10.58257/IJPREMS33384>