Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.698-708

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.35290 Link : https://dx.doi.org/10.29408/jit.v9i2.35290

Analisis Sentimen Komentar Masyarakat di Twitter Menggunakan Algoritma Support Vector Machine

Arnila Sandi[1*], Yahya[2], Muhammad Qusaeri[3]
[1,3] Program Studi Teknik Komputer, Universitas Hamzanwadi
[2] Program Studi Sistem Informasi, Universitas Hamzanwadi
*arnilasandi1878@gmail.com

Abstrak

Media sosial telah menjadi sarana utama bagi masyarakat untuk menyampaikan opini dalam berbagai macam isu, salah satu isu yang di bahas adalah mengenai produk smartphone. Media sosial yang di gunakan untuk mengapresiasikan opini antara lain FaceBook, Instagram, Twitter, WhatsApp, dan Youtube. Twitter merupakan salah satu platform yang banyak digunakan untuk mengekspresikan pendapat dalam bentuk komentar yang dapat bersifat positif, negatif, maupun netral. Penelitian ini bertujuan untuk mengklasifikasikan sentimen komentar pengguna Twitter terkait smartphone menggunakan algoritma Support Vector Machine (SVM). Metode dalam penelitian ini menggunakan metode kuantitatif. Data yang digunakan berupa komentar atau data teks dalam format CSV yang diproses melalui tahapan preprocessing meliputi pembersihan data, tokenisasi, stopword removal, dan stemming. Selanjutnya, dilakukan pemodelan sentimen dan evaluasi kinerja model. Hasil penelitian menunjukkan bahwa metode SVM menghasilkan akurasi sebesar 82,28%, recall 82,26%, presisi 82,25%, dan F1-score 82,25%. Berdasarkan hasil tersebut, dapat disimpulkan bahwa algoritma SVM memiliki kinerja yang baik dalam melakukan analisis sentimen pada data media sosial, khususnya Twitter.
.

Kata kunci : Analisis Sentimen, Klasifikasi, Support Vektor machine (SVM), Teks mining

Abstract

Social media has become the primary means for the public to express their opinions on a wide range of issues, one of which is smartphone products. The social media platforms used to express opinions include Facebook, Instagram, Twitter, WhatsApp and YouTube. Twitter is one of the most widely used platforms for expressing opinions in the form of comments, which may be positive, negative or neutral. This study aims to classify the sentiment of Twitter users' comments regarding smartphones using the Support Vector Machine (SVM) algorithm. The research employs a quantitative methodology. The data used consists of comments or text data in CSV format, which is processed through a series of pre-processing steps, including data cleaning, tokenisation, stop-word removal and stemming. This will be followed by sentiment modelling and an evaluation of the model's performance. The results of the study show that the SVM method achieved an accuracy of 82,28%, a recall of 82,26%, a precision of 82,25%, and an F1-score of 82,25%. Based on these results, it can be concluded that the SVM algorithm performs well in sentiment analysis of social media data, particularly Twitter.

Keywords : *Classification, Sentiment Analysis, Support Vector Machines (SVM), Text Mining*

1. Pendahuluan

Perkembangan teknologi saat ini telah membuat proses komunikasi dan interaksi antarindividu menjadi semakin mudah. Media sosial menjadi salah satu platform utama yang digunakan masyarakat untuk mengekspresikan aspirasi, berbagi informasi, serta berinteraksi dengan berbagai isu atau berita yang sedang berkembang di linimasa mereka[1][2][3]. Berbagai jenis media sosial yang umum digunakan antara lain Facebook, Instagram, Twitter, TikTok, dan lainnya. Selain sebagai sarana interaksi melalui

Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.698-708

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.35290 Link : https://dx.doi.org/10.29408/jit.v9i2.35290

pertukaran komentar, baik positif maupun negatif, media sosial juga dimanfaatkan sebagai media pendukung aktivitas ekonomi, seperti pemasaran produk.

Salah satu platform media sosial yang populer saat ini adalah Twitter. Twitter merupakan platform komunikasi yang memungkinkan para penggunanya untuk dapat mengekspresikan pendapatnya dan telah digunakan secara luas di seluruh dunia. Pada Twitter, pesan yang dibagikan oleh pengguna dikenal dengan istilah *tweet*, yaitu unggahan singkat yang digunakan untuk menyampaikan informasi terkini, mengekspresikan perasaan, menyampaikan aspirasi, serta memberikan opini terhadap isu-isu yang sedang hangat diperbincangkan [4][5][6]. Karakteristik tweet yang singkat dan padat menjadikan platform ini sebagai sumber data yang relevan untuk mengetahui opini publik.

Proses untuk menganalisis opini atau pandangan masyarakat terhadap suatu isu yang beredar di media sosial dikenal dengan istilah analisis sentimen. Analisis sentimen banyak digunakan untuk mengetahui persepsi konsumen terhadap suatu produk, memahami respons masyarakat terhadap berita politik, serta berbagai keperluan lainnya. Metode ini berfungsi untuk mengklasifikasikan emosi atau opini ke dalam kategori tertentu, seperti positif, negatif, atau netral[7][8].

Salah satu penerapan analisis sentimen dilakukan pada data Twitter, di mana sistem secara otomatis mengumpulkan data dari tweet pengguna. Berdasarkan hasil klasifikasi tersebut, informasi yang diperoleh dapat digunakan untuk memahami serta mengevaluasi suatu topik secara lebih objektif, sehingga dapat mendukung pengambilan keputusan yang lebih tepat[9][10]. Analisis sentimen merupakan bagian dari bidang *Natural Language Processing* (NLP) yang digunakan untuk mengidentifikasi dan mengklasifikasikan opini atau pandangan dalam bentuk data teks yang berkaitan dengan suatu isu atau peristiwa ke dalam kategori positif, negatif, atau netral [11][12].

Dalam metode analisis sentimen, terdapat proses text mining yang berkaitan dengan ekstraksi pola serta pengetahuan dari data berbentuk teks. Sejak diperkenalkan pada tahun 1990-an, text mining dikenal dengan berbagai istilah, seperti Knowledge Discovery from Text (KDT), Intelligent Text Analysis, dan Data Mining. Sesuai dengan berbagai istilah tersebut, text mining merupakan bidang multidisiplin yang mencakup data mining, knowledge discovery, information retrieval, statistik, serta natural language processing (NLP)[13][14].

Penelitian yang relevan dengan penelitian ini telah dilakukan oleh Dian Siti Utami dan rekan-rekan, yang menganalisis ulasan masyarakat terhadap layanan pinjaman online melalui media sosial

Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.698-708

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.35290 Link : https://dx.doi.org/10.29408/jit.v9i2.35290

Twitter. Pada penelitian tersebut, proses pre-processing dilakukan menggunakan perangkat lunak RapidMiner, dengan hasil tingkat akurasi sebesar 62,00%. Selain itu, diperoleh distribusi sentimen berupa 59% ulasan negatif dan 41% ulasan positif. Hasil tersebut menunjukkan bahwa model yang digunakan memiliki tingkat akurasi yang cukup baik[15]. Perbedaan penelitian tersebut dengan penelitian yang dilakukan dalam studi ini terletak pada penggunaan perangkat pengolahan data, yaitu menggunakan Google Colab dengan bahasa pemrograman Python, serta perbedaan pada topik atau isu yang dianalisis.

Dalam penelitian ini, metode klasifikasi yang digunakan adalah algoritma Support Vector Machine (SVM). Sebelum tahap klasifikasi dilakukan, data komentar dari Twitter terlebih dahulu melalui tahap pre-processing untuk mengurangi dimensi ruang vektor model sehingga proses klasifikasi menjadi lebih efisien. Tahapan pre-processing bertujuan untuk menghilangkan kata-kata yang tidak relevan, menyeragamkan bentuk huruf (menjadi huruf kecil), serta mengurangi jumlah kata yang tidak memiliki makna signifikan. Adapun tahapan pre-processing dalam penelitian ini meliputi cleaning, case folding, tokenization, stopwords removal, dan stemming[16].

Algoritma SVM merupakan salah satu metode supervised machine learning yang didasarkan pada teori pembelajaran statistik. Algoritma ini bekerja dengan memilih subset karakteristik dari data pelatihan sehingga proses klasifikasi dapat dilakukan secara optimal terhadap keseluruhan dataset. Oleh karena itu, SVM banyak digunakan untuk menyelesaikan berbagai permasalahan klasifikasi[17][18]. Secara umum, SVM bertujuan untuk menemukan hyperplane optimal yang dapat memisahkan data ke dalam kelas-kelas yang berbeda di dalam ruang fitur dengan margin terbaik [19][20].

2. Tinjauan Pustaka

2.1. Penelitian Terkait

Peneltian ini didukung oleh beberapa penelitian sebelumnya yang menggunakan algoritma SVM.

1. Penelitian yang dilakukan oleh Muhammad Irvan Abidin dkk pada tahun 2025 dengan judul Analisis Sentimen Terhadap Timnas Indonesia di piala Asia 2023. Penelitian ini bertujuan untuk menganalisis sentimen publik terhadap Tim Nasional Indonesia selama Piala Asia 2023 melalui komentar Instagram, dengan memanfaatkan model Transformer berbahasa Indonesia. Sebanyak 21.045 komentar dikumpulkan dari akun resmi @timnasindonesia dan difilter menjadi 17.829 komentar layak analisis setelah melalui proses preprocessing seperti text cleaning, case folding, normalisasi, tokenisasi, stopword removal, dan stemming. Hasil menunjukkan bahwa IndoBERT (learning

Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.698-708

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.35290 Link : https://dx.doi.org/10.29408/jit.v9i2.35290

rate 5e-5) memberikan performa terbaik dengan akurasi 0.8897 dan F1-score 0.8859, mengungguli model lainnya[21].

2. Penelitian dengan judul Analisis Sentimen Terhadap Pemutusan Hubungan Kerja di Indonesia: Komparasi IndoBert dengan SVM, Random Forest dan Decission Tree dengan Optimasi TF-IDF ini dilakukan oleh Nida Nur Aini Aryanti, dan Ozzi Suria. Tujuan dari penelitian ini adalah melakukan analisis sentiment public pada fenomena PHK di Indonesia dengan menggunakan empat metode klasifikasi yaitu, IndoBERT, SVM, Random Forest dan Decission Tree**.** Adapun hasil dari penelitian tersebut adalah memperlihatkan bahwa IndoBERT menghasilkan akurasi tertinggi sebesar 89,6%, algoritma SVM menghasilkan akurasi yakni 88%, precision 90%, recal 95%, F1-score 92%, diikuti oleh Random Forest (78,04%) dan Decision Tree (70,40%). Hasil tersebut diperoleh dari data periode Januari – Mei 2025 dan menghasilkan 36.507 tweet[22].

3. Penelitian yang dilakukan oleh Alfin Nur Saputra dkk pada tahun 2025 yang berjudul Enhancing Sentiment Analysis Accuracy Using SVM and Slang Word Normalization on YouTube Comments. Penelitian ini mengkaji peningkatan akurasi analisis sentimen pada komentar YouTube dengan mengintegrasikan Support Vector Machine (SVM) bersama normalisasi kata-kata slang. Kumpulan data terdiri dari 3.375 komentar mengenai film Pengabdi Setan 2. Setelah penyetelan hiperparameter menggunakan Grid Search, model SVM mencapai kinerja optimal, dengan presisi, recall, dan skor F1 sebesar 100% untuk kelas positif dan negatif. Hasil ini menunjukkan bahwa penggabungan normalisasi teks dan model pembelajaran mesin dapat secara efektif menangani kompleksitas bahasa informal dalam konten yang dibuat pengguna, sehingga menghasilkan klasifikasi sentimen yang sangat akurat[23].

4. Pada tahun 2022 penelitian dilakukan oleh Melati Indah Petiwi dkk. Dengan judul Analisis Sentimen Gofood Berdasarkan Twitter Menggunakan Metode *Naïve Bayes* dan *Support Vector Machine.* Penelitian ini bertujuan untuk menganalisa opini masyarakat terhadap kinerja Gojek (Gofood) di Indonesia. Hasil dari penelitian ini analisis sentimen masyarakat mengenai Gofood pada twitter menghasilkan 92,8% bernilai netral, 5,2% bernilai positif dan 2,0% bernilai negatif. Perbandingan hasil akurasinya, metode Support Vector Machine akurasinya lebih besar dari metode Naïve Bayes, dengan nilai akurasi SVM sebesar 83% dan 98,5% sedangkan Naïve Bayes sebesar 74,6% dan 91,5% bernilai positif dan 2,0% bernilai negative[24].

5. Penelitian yang dilakukan oleh Styawati pada tahun 2021 dengan judul Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja pada Twitter dengan Metode Support vector

Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.698-708

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.35290 Link : https://dx.doi.org/10.29408/jit.v9i2.35290

Machine bertujuan untuk menganalisis sentiment Masyarakat Indonesia terhada program kartu prakerja. Sentimen yang digunakan adalah positif, negative dan netral. Hasil evaluasi yang dilakukan pada nilai akurasi kernel linear 98.67%, precission 98%, recall 99%, dan F1-Score 98%, sedangkan pada nilai akurasi kernel RBF 98.34%, precission 97%, recall 98%, F1-Score 98%, dapat disimpulkan bahwa sentimen masyarakat dari pengguna twitter terhadap program kartu prakerja dimasa pandemi lebih condong ke netral sebesar 98,34%. Berdasarkan hasil evaluasi yang dilakukan pada nilai akurasi kernel linear menghasilkan nilai akurasi 98.67%, sedangkan kernel RBF menghasilkan akurasi 98.34%. Maka dari sisi akurasi kernel linear lebih akurat dari pada kernel RBF[25].

2.2. Landasan Teori

1. Analisis Sentimen

Analisis sentimen merupakan pengukuran pendapat orang terhadap tingkat kesepakatan mengenai topik/isu tertentu, produk, atau layanan, atau bahkan suatu kejadian. Dua pendekatan yang telah digunakan untuk mempelajari analisis sentimen yaitu natural language processing, dan algoritma machine learning. Dalam metode klasifikasian machine learning pada sentiment analysis ada beberapa metode yang bisa dipakai, yaitu Na¨ıve Bayes Classifier (NBC), Maximum Entropy, dan Support Vector Machine (SVM)[26].

2. Klasifikasi

Klasifikasi merupakan sebuah proses untuk menganalisa data sehingga menciptakan model-model yang dapat mendeskripsikan kelas yang ada pada data dan merupakan salah satu teknik dari pengolahan data yang biasa digunakan dalam pembelajaran mesin[27].

3. Support Vektor Machine

Support Vector Machine (SVM) merupakan algoritma pendekatan yang sering digunakan untuk masalah Regresi dan Klasifikasi yang membutuhkan Supervised Learning[28].

4. Convusion Matriks

Confusion Matrix biasa digunakan sebagai alat untuk mengevaluasi kinerja dari suatu metode klasifikasi. Metode ini menggunakan empat istilah hasil klasifikasi, yaitu True Positif (TP), True Negatif (TN), False Positif (FP), dan False Negatif (FN)[29].

3. Metode Penelitian

3.1. Metode dan tahapan penelitian

Penelitiaan ini menggunakan metode kuantitatif dengan menerapkan metode klasifikasi untuk menganalisis data teks.

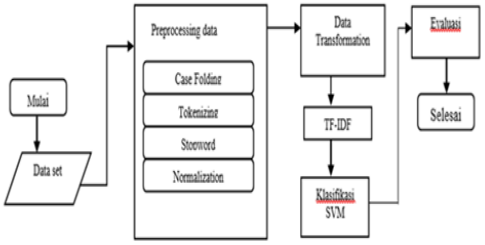
Metode kuantitatif merupakan rangkaian sistematika terhadap sebuah masalah dengan mengumpulkan data, lalu diolah dengan metode statistic matematika. Support Vector Machine adalah algoritma yang digunakan untuk menganalisis hasil evaluasi sentiment

Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.698-708

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.35290 Link : https://dx.doi.org/10.29408/jit.v9i2.35290

masyarakar di media sosial twitter, Adapun tahapan penelitian dapat dilihat pada gambar 1.



Gambar 1 Tahapan penelitian

3.2. Lokasi Penelitian

Penelitian dilakukan di Selong, Lombok Timur Nusa Tenggara Barat

3.3. Sumber data penelitian

Data yang digunakan berupa data teks dalam bentuk opini atau pendapat masyarakat, data diambil dari Kaggle. Objek pada penelitian ini adalah opini pengguna twiter yang maembahas tentang smartphone, dan data yang diambil sebanyak 3309 data, yang terdiri dari tiga kolom yaitu *sentence* (kalimat opini masyarakat), sentiment dan score

1	Good case, Excellent value.	Positif	1
2	Great for the jawbone.	Positif	1
3	Tied to charger for conversations lasting more...	Negatif	2
4	The mic is great.	Positif	1
...	...	...	...
3304	Online anonymity enables freedom of expression...	Negatif	2
3305	Smart cities leverage data and technology to i...	Positif	1
3306	Biometric identification technologies offer co...	Negatif	2
3307	Data ethics guidelines are essential for ensur...	Positif	1
3308	Decentralized technologies such as blockchain ...	Positif	1

3309 rows × 3 columns

Gambar 2 Tampilan data

3.4. Teknik Pemrosesan data

Penelitian ini menggunakan algoritma Support Vector Machine (SVM). Dan dilakukan beberapa tahapan processing data yaitu:

1. *Case folding*, dilakukan untuk menghilangkan redudansi data. Sebagai contoh “Excellent”, “excellent” dan “EXCELLENT”, hal ini akan mengganggu proses komputasi karena semua dianggap entitas yang berbeda. Dengan menerapkan case folding maka ketiga variasi penulisan tersebut akan diseragamkan menjadi “excellent”. Berikut adalah hasil proses case folding dapat dilihat pada gambar 2 dan 3.

	sentence	sentiment	score
0	So there is no way for me to plug it in here i...	negative	2
1	Good case, Excellent value.	positive	1
2	Great for the jawbone.	positive	1
3	Tied to charger for conversations lasting more...	negative	2
4	The mic is great.	positive	1

Gambar 3 Kalimat sebelum proses case folding

	sentence	sentiment	score
0	so there is no way for me to plug it in here i...	Negatif	2
1	good case excellent value	Positif	1
2	great for the jawbone	Positif	1
3	tied to charger for conversations lasting more...	Negatif	2
4	the mic is great	Positif	1

Gambar 4 Kalimat setelah proses case folding

2. Tokenization, tahap ini dilakukan untuk memisahkan tiap kata penyusun dengan proses pemotongan kalimat.
3. Sopword Removal, proses ini dilakukan untuk menghapus kata yang tidak perlu digunakan.
4. Normalizatoin, dilakukan untuk mengubah kata yang tidak baku menjadi kata baku
5. Steaming, pada proses ini dilakukan perubahan kata-kata yang berimbuhan menjadi kata dasar.
6. Metode TF-IDF, metode pembobotan kata yang sering digunakan untuk mengevaluasi seberapa penting sebuah kata dalam sebuah

Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.698-708

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.35290 Link : https://dx.doi.org/10.29408/jit.v9i2.35290

dokumen.

3.5. Pemodelan data

Langkah berikutnya adalah pemodelan data menggunakan algoritma Support Vector Machine (SVM).

```
# 4. Model SVM
svm_model = SVC(kernel='linear')
svm_model.fit(X_train, y_train)
```

3.6. Evaluasi

Tahap terkahir yaitu evaluasi atau pengujian, pada tahap ini pengujian dilakukan menggunakan algoritma Support Vector Machine (SVM) untuk mendapatkan nilai akurasi, presisi, dan recall. Sebelum proses pencarian nilai tersebut dilakukan proses split data dengan jumlah data 3309, maka jumlah data terbagi menjadi 70% berbanding dengan 30%. Pembagian 70:30 dipilih karena memberikan keseimbangan antara jumlah data yang digunakan untuk melatih model dan jumlah data yang digunakan untuk mengevaluasi kinerja model. Sebanyak 70% data dianggap cukup untuk mempelajari pola dan karakteristik data sehingga model dapat dibangun dengan baik, sedangkan 30% sisanya digunakan sebagai data uji yang cukup representatif untuk mengukur kemampuan model dalam melakukan prediksi terhadap data yang belum pernah dilihat sebelumnya.

4. Hasil dan Pembahasan

4.1. Hasil Penelitian

Hasil yang telah dicapai dari penelitian yang dilakukan, antara lain: persiapan data set, pembersihan (cleaning), penyamaan huruf (case folding), pemisahan kata (tokenization), penghapusan kata-kata berhenti (stopwords removal), dan pengambilan akar kata (stemming), Normalization, dan metode TF-IDF. Kemudian dilakukan pemodelan data dan evaluasi data untuk mencari nilai akurasinya. Sebelum proses pencarian nilai tersebut maka dilakukan proses split data dengan jumlah data 3309, dimana data terbagi menjadi 70% berbanding 30% dengan perincian sebagai berikut, jumlah data train : 2316, dan jumlah data test :993. Pemilihan rasio 70:30 bertujuan untuk mengurangi risiko *overfitting*, yaitu kondisi ketika model terlalu menyesuaikan diri dengan data latih sehingga kurang mampu melakukan generalisasi terhadap data baru. Dengan menyediakan porsi data uji yang memadai, hasil evaluasi dapat lebih mencerminkan performa model pada kondisi nyata. Selain itu tahapan perhitungan juga dilakukan menggunakan confusion matrik, dengan hasil sebagai berikut :

1. Rincian untuk kelas Positif

- TP : 394 - FP : 105
- TN : 410 - FN : 84

2. Rincian untuk kelas Negatif

- TP : 410 - FP : 84

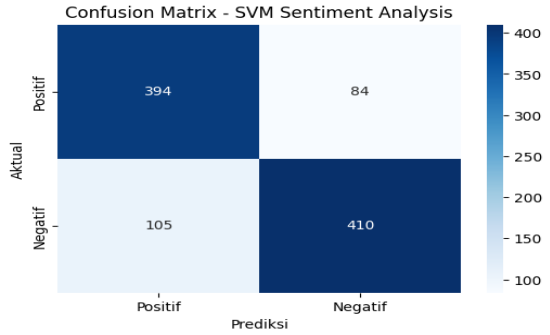
Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.698-708

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.35290 Link : https://dx.doi.org/10.29408/jit.v9i2.35290

- TN : 394 - FN : 105

Tampilan visual untuk nilai confusion matriks dapat dilihat pada gambar 6.



Gambar 5 Tampilan visual convusion matriks

Setelah itu proses selanjutnya adalah evaluasi data dengan menggunakan algoritma SVM, hasil dapat dilihat pada tabel 1 berikut:

Tabel 1 Hasil evauasi model SVM

No	Keterangan	Nilai
1	Akurasi	0,8228
2	Presisi	0,8225
3	Recall	0,8226
4	F1-Score	0,8225

	precision	recall	f1-score	support
Negatif	0.83	0.83	0.83	515
Positif	0.81	0.82	0.82	478
accuracy			0.82	993
macro avg	0.82	0.82	0.82	993
weighted avg	0.82	0.82	0.82	993

Gambar 6 hasil evaluasi data

Visualisasi data dilakukan untuk memperoleh gambaran awal mengenai karakteristik umpan balik pengguna. Visualisasi disajikan dalam bentuk *word cloud* untuk seluruh data, sentimen positif, dan sentimen negatif guna mengidentifikasi kata-kata yang dominan muncul pada masing-masing kategori. Selain itu, distribusi sentimen divisualisasikan menggunakan diagram batang dan diagram lingkaran untuk menunjukkan proporsi data pada setiap kelas sentimen.

Gambar 7 Visualisasi data

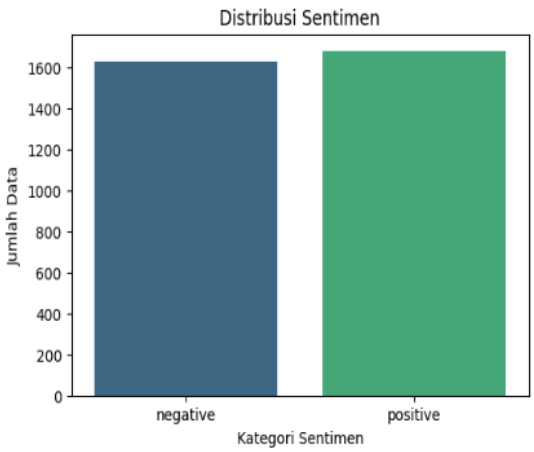
Analisis data antara ulasan positif dan negatif terlihat berbeda, ulasan positif sebanyak 1679 ulasan dan negatif mendapatkan sebanyak 1630 ulasan

label	count
Positif	1679
Negatif	1630

dtype: int64

Gambar 8 Hasil ulasan

Berikut tampilan grafik dari hasil ulasan data



Gambar 9 Grafik ulasan

Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.698-708

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.35290 Link : https://dx.doi.org/10.29408/jit.v9i2.35290

4.2. Pembahasan

Proses awal yang dilakukan sebelum evaluasi seperti case folding, tokenizing, stopword removal dan stemming terbukti memiliki pengaruh positif terhadap performa model. Case folding membantu menormalisasi teks menjadi sederhana agar proses pengolahan data tidak memiliki banyak noise. Tokenizing Adalah proses memecah teks menjadi token-token agar mudah diproses dan dipahami oleh komputer. Sedangkan Stopword removal membantu menghilangkan kata-kata yang tidak memiliki makna signifikan, sehingga model fokus pada kata kunci yang penting. Begitu juga dengan stemming menyederhanakan variasi kata ke bentuk dasar, yang mengurangi jumlah fitur dan memperkuat konsistensi data.

Penelitian kali ini menggunakan dataset yang hamper seimbang antara label positif dan negatif, sehingga model tidak mengalami masalah terhadap salah satu kelas. Namun, ada beberapa komentar yang sifatnya masih ambigu sehingga tetap menimbulkan tantangan bagi model, karena kata-kata sangat sulit ditangkap dengan representasi kata dasar. Hal ini menunjukkan bahwa meskipun preprocessing efektif, kompleksitas bahasa alami tetap menjadi faktor pembatas akurasi.

Berdasarkan hasil pengujian yang telah dilakukan, analisis hasil yang kami dapatkan bahwa model SVM menunjukkan kemampuan yang cukup baik dalam membedakan ulasan positif dan negatif. Nilai akurasi yang diperoleh 82,28% mencerminkan bahwa sebagian besar komentar berhasil diklasifikasikan dengan tepat. Hal ini mempertegas bahwa algoritma SVM efektif sebagai algoritma klasifikasi teks untuk dua label pada dataset yang relatif seimbang

5. Kesimpulan

Penelitian ini menunjukkan bahwa kombinasi pra-pemrosesan teks (case folding, tokenaizing, stopword removal dan stemming) dan model SVM sangat efektif untuk klasifikasi sentimen dua label, dengan nilai akurasi mencapai 82,28%, recall 82,26%, presisi 82,25%, dan F1-score 82,25%. Tahap prapemrosesan terbukti menyederhanakan teks dan meningkatkan kualitas fitur, sementara algoritma SVM mampu membedakan komentar positif dan negatif dengan baik. Meskipun demikian, beberapa komentar ambigu dan keterbatasan dataset masih menjadi tantangan, sehingga disarankan untuk menambah data, mengoptimalkan parameter SVM, serta mempertimbangkan representasi kata lanjutan seperti n-gram atau word embedding untuk peningkatan performa di masa mendatang.

6. Daftar Pustaka

[1] A. D. Andriani *et al.*, “Model Komunikasi Literasi Digital Dalam,” *Interak. J. Ilmu Komun.*, vol. 13, no. 2, pp. 439–464, 2024.

Jurnal Informatika dan Teknologi

Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.698-708

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.35290 **Link : https://dx.doi.org/10.29408/jit.v9i2.35290**

[2] M. B. S. I Kadek Sista Priyanata1, "Framing dan Isu Sosial dalam Jurnalisme Media Sosial : Tinjauan Literatur atas Praktik Representasi Komunitas Pendatang," *J. Interak. J. Ilmu Komun.*, vol. 10, no. 1, pp. 77–91, 2026, doi: https://doi.org/10.30596/ji.v10i1.26534.

[3] T. W. Rahmania Mustaqlilah1, Okky Widyaningtyas2, "Efektivitas Penggunaan Twitter Sebagai Sarana Peningkatan Berpikir Kritis Mahasiswa Ilmu Komunikasi," *J. Ilmu Komun.*, vol. 2, no. 1, pp. 18–28, 2024, doi: 10.54259/mukasi.v2i1.1346.

[4] R. Mustaqlilah, O. Widyaningtyas, and T. Wantoro, "Efektivitas Penggunaan Twitter Sebagai Sarana Peningkatan Berpikir Kritis Mahasiswa Ilmu Komunikasi," vol. 2, no. 1, pp. 18–28, 2023, doi: 10.54259/mukasi.v2i1.1346.

[5] N. A. Tsurayya, P. Haniza, and R. Annisa, "Fungsi bahasa dalam jejaring media sosial twitter," *Kaji. Linguist. dan Sastra*, vol. 8, no. 2, pp. 142-160., 2023, doi: 10.23917/kls.v8i2.18463.

[6] D. Darwis, E. S. Pratiwi, and A. F. O. Pasaribu, "Penerapan Algoritma SVM Uuntuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia", *J. Ilm. Edutic*, vol. 7, no. 1, pp. 1–11, 2020, doi: 10.21107/edutic.v7i1.8779.

[7] A. Hendrawan and E. I. Sela, "Jurnal Indonesia : Manajemen Informatika dan Komunikasi Analisis Sentimen Komentar Youtube Tentang Resesi Global 2023 Menggunakan LSTM Abstrak Jurnal Indonesia : Manajemen Informatika dan Komunikasi," *J. Indones. Manaj. Inform. dan Komun.*, vol. 5, no. 1, pp. 587–593, 2024, doi: https://doi.org/10.35870/jimik.v5i1.526.

[8] A. Y. Pratama, G. A. Sanjaya, N. K. Lubis, and M. R. Aditya, "Analisis Sentimen Publik Terkait Danantara Menggunakan Algoritma IndoBERT pada Platform Media Sosial," vol. 9, no. 1, 2025, doi: 10.47002/metik.v9i1.1055.

[9] A. W. Muh. Hasbi Asshiddiq1*, "Analisis Sentimen tentang Penundaan Pengangkatan CPNS 2025 pada Platform X Menggunakan Metode IndoBERT," *J. Fak. Tek. Univ. Islam Lamongan*, vol. 17, no. 2, 2025, doi: 10.30736/jt.v17i2.1448.

[10] A. W. V Hutabarat, N. Luh, S. Saraswati, and K. Suryadi, "Analisis Sentimen Data Ulasan Pengguna MyPertamina di Twitter dengan Metode Machine Learning dan Deep Learning," vol. 13, no. 1, pp. 145–154, 2024, doi: https://doi.org/10.26593/jrsi.v13i1.6958.145-154.

[11] M. R. Novita Malasari a, 1, "Analisis Sentimen Media Sosial Menggunakan Algoritma BERT dan LSTM," *J. Comput. Sci. Inf. Technol.*, vol. 1, no. 3, pp. 85–92, 2025.

[12] M. I. Fikri, T. S. Sabrila, and Y. Azhar, "Comparison of Naïve Bayes and Support Vector Machine Methods in Twitter Sentiment Analysis," *Smatika J.*, vol. 10, no. 02, pp. 71–76, 2020.

[13] M. Qorib, T. Oladunni, M. Denis, E. Ososanya, and P. Cotae, "Covid-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on COVID-19 vaccination Twitter dataset," *Expert Syst. Appl.*, vol. 212, no. January 2022, p. 118715, 2023, doi: 10.1016/j.eswa.2022.118715.

[14] M. Jelita, "Text Mining dengan Topic Modelling LDA dari Pertanyaan Gelar Wicara Literasi Perpustakaan Nasional RI," *Media Pustak.*, vol. 31, no. 3, pp. 253–265, 2024, doi: 10.37014/medpus.v31i3.5237.

[15] D. S. Utami and A. Erfina, "Analisis Sentimen Pinjaman Online di Twitter Menggunakan Algoritma Support Vector Machine (SVM)," *SISMATIK (Seminar Nas. Sist. Inf. dan Manaj. Inform.*, vol. 1, no. 1, pp. 299–305, 2021.

Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.698-708

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.35290 Link : https://dx.doi.org/10.29408/jit.v9i2.35290

[16] H. Ma, A. P. Wibawa, and M. I. Akbar, "Klasifikasi artikel ilmiah dengan berbagai skenario preprocessing," *Sains, Apl. Komputasi dan Teknol. Inf.*, vol. 2, no. 2, pp. 70–78, 2020.

[17] D. R. Nurqotimah, A. N. Khudori, and R. S. Pradini, "Implementasi Algoritma Support Vector Machine (SVM) Untuk Klasifikasi Penyakit Stroke," *J. Appl. Comput. Sci. Technol. (JACOST)*, vol. 5, no. 2, pp. 179–185, 2024, doi: https://doi.org/10.52158/jacost.v5i2.817.

[18] S. Muawanah, U. Muzayanah, M. G. R. Pandin, M. D. S. Alam, and J. P. N. Trisnaningtyas, "Stress and Coping Strategies of Madrasah's Teachers on Applying Distance Learning During COVID-19 Pandemic in Indonesia," *Qubahan Acad. J.*, vol. 3, no. 4, pp. 206–218, 2023, doi: 10.48161/Issn.2709-8206.

[19] M. Z. Abidin, R. Makhfuddin, A. Syaifuddin, U. I. Majapahit, T. Rejo, and K. Mojokerto, "Analisis Sentimen Ulasan Aplikasi Sirekap Menggunakan Algoritma Support Vector Machine," *JITET (Jurnal Inform. dan Tek. Elektro Ter.*, vol. 13, no. 3, pp. 679–691, 2024, doi: http://dx.doi.org/10.23960/jitet.v13i3S1.7768.

[20] H. C. Husada and A. S. Paramita, "Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM)," *Teknika*, vol. 10, no. 1, pp. 18–26, 2021, doi: 10.34148/teknika.v10i1.311.

[21] M. I. Abidin and E. W. Pamungkas, "Analisis sentimen terhadap timnas indonesia di piala asia 2023 dengan model transformer berbahasa indonesia," *Rabit*, vol. 10, no. 2, pp. 482–496, 2025, doi: tps://doi.org/ 10.36341/rabit.v10i2.6142.

[22] N. Nur, A. Aryanti, and O. Suria, "Analisis Sentimen Terhadap Pemutusan Hubungan Kerja di Indonesia : Komparasi IndoBERT Dengan SVM , Random Forest , dan Decision Tree Dengan Optimasi TF - IDF", *J. Teknol. dan Sist. Inf. Univrab*, vol. 10, no. 2, pp. 1158–1176, 2025, doi: https://doi.org/10.36341/rabit.v10i2.6364.

[23] A. Nur, A. Saputra, R. E. Saputro, D. Intan, and S. Saputra, "Enhancing Sentiment Analysis Accuracy Using SVM and Slang Word Normalization on YouTube Comments," vol. 9, no. 2, pp. 687–699, 2025.

[24] M. I. Petiwi, A. Triayudi, and I. D. Sholihati, "Analisis Sentimen Gofood Berdasarkan Twitter Menggunakan Metode Naïve Bayes dan Support Vector Machine," vol. 6, pp. 542–550, 2022, doi: 10.30865/mib.v6i1.3530.

[25] N. Hendrastuty *et al.*, "Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine," vol. 6, no. 3, pp. 150–155, 2021.

[26] R. Ramlan, N. Satyahadewi, and W. Andani, "Analisis Sentimen Pengguna Twitter Menggunakan Support Vector Machine Pada Kasus Kenaikan Harga BBM," *Jambura J. Math.*, vol. 5, no. 2, pp. 431–445, 2023, doi: 10.34312/jjom.v5i2.20860.

[27] S. L. M. Yahya, Nurhidayati, "Pengaruh Pendidikan Terhadap Tingkat Kesehatan Masyarakat Kecamatan Suragala Kabupaten Lombok Timur Menggunakan Algoritma Random Forest," *Teknol. dan Inf.*, vol. 7, no. 2, 2024, doi: 10.29408/jit.v7i2.26352 Infotek.

[28] R. Ahmad, M. Saiful, and V. No, "Analisis Komparatif Feture Selection dan Metode Klasifikasi Intrution Detection System (IDS)," *Inform. dan Teknol.*, vol. 8, no. 2, pp. 673–685, 2025, doi: 10.29408/jit.v8i2.31103.

[29] A. Sandi and M. Qusyairi, "Analisis Prediksi Tingkat Kesejahteraan Masyarakat Nelayan Lombok Timur Dengan Algoritma Naïve Bayes Infotek," *Inform. dan Teknol.*, vol. 7, no. 2, pp. 563–574, 2024, doi: 10.29408/jit.v7i2.26543 Infotek.