

PENERAPAN METODE *NAÏVE BAYES* DAN *SUPPORT VECTOR MACHINE* PADA ANALISIS SENTIMEN NETIZEN DI TWITTER VOLLEY BALL INDONESIA

Wismo Ginanjar¹⁾, Alexius Endy Budianto²⁾, Moh Ahsan³⁾,
Muhammad Priyono Tri Sulistyanto⁴⁾

Universitas PGRI Kanjuruhan Malang^{1,2,3,4)}

email : wismoginanjar@gmail.com

Abstrak

Media sosial telah menjadi bagian integral dari kehidupan masyarakat modern, menawarkan platform untuk ekspresi opini publik. Di Indonesia, volley ball adalah olahraga yang sangat populer, dan Volley Ball Indonesia menjadi topik utama diskusi di media sosial, terutama Twitter. Penelitian ini bertujuan untuk menganalisis sentimen komentar netizen pada akun Twitter resmi Volley Ball Indonesia (@volleyball.indonesia) dengan menggunakan metode *Naïve Bayes* dan *Support Vector Machine* (SVM). Data yang digunakan berjumlah 2.920 komentar dari 50 postingan dalam rentang waktu 28 September 2023 – 10 Mei 2024, difokuskan pada pertandingan Timnas Putra U-23 dan Senior. *Naïve Bayes* dan SVM dipilih karena keduanya adalah metode yang efektif dalam klasifikasi sentimen. *Naïve Bayes* menggunakan pendekatan probabilistik, sementara SVM mencari hyperplane terbaik untuk memisahkan kelas-kelas data. Hasil penelitian menunjukkan bahwa kedua metode ini dapat digunakan untuk menganalisis sentimen dengan tingkat akurasi yang baik. Hasil pengujian pada masing-masing data *training* dan data *testing* dengan presentasi yang berbeda akan memberikan hasil akurasi yang berbeda. Hasil pengujian dari metode *Naive Bayes* didapatkan nilai akurasi tertinggi sebesar 71% dengan rasio 70:30 dan *Support Vector Machine* mendapatkan nilai akurasi tertinggi sebesar 76% dengan rasio 80:20. Maka dapat disimpulkan bahwa metode *Support Vector Machine* mendapatkan nilai akurasi lebih tinggi dari pada metode *Naive Bayes*.

Kata Kunci : *analisis sentiment; naive bayes; support vector machine; akurasi klasifikasi; Volley Ball Indonesia.*

Abstract

Social media has become an integral part of modern society, offering a platform for public opinion expression. In Indonesia, volleyball is a very popular sport, and Volley Ball Indonesia is the main topic of discussion on social media, especially Twitter. This study aims to analyze the sentiment of netizen comments on the official Twitter account of Volley Ball Indonesia (@volleyball.indonesia) using the *Naive Bayes* method and *Support Vector Machine* (SVM). The data used amounted to 2,920 comments from 50 posts in the period of September 28, 2023 - May 10, 2024, focused on the U-23 and Senior Men's National Team matches. *Naive Bayes* and SVM were chosen because both are effective methods in sentiment classification. *Naive Bayes* uses a probabilistic approach, while SVM looks for the best hyperplane to separate data classes. The results of the study show that both methods can be used to analyze sentiment with a good level of accuracy. The test results on each training data and testing data with different presentations will provide different accuracy results. The test results of the *Naive Bayes* method obtained the highest accuracy value of 71% with a ratio of 70:30 and the *Support Vector Machine* obtained the highest accuracy value of 76% with a ratio of 80:20. So it can be concluded that the *Support Vector Machine* method gets a higher accuracy value than the *Naive Bayes* method.

Keywords : *Sentiment analysis; naive bayes; support vector machine; classification accuracy; Volley Ball Indonesia*

1. PENDAHULUAN

Di Indonesia, Volley Ball merupakan olahraga yang sangat populer dengan basis penggemar yang besar. Volley Ball Indonesia, sebagai representasi negara dalam kompetisi volley ball internasional, tentunya menjadi sorotan dan perhatian masyarakat Indonesia. Prestasi dan pencapaian Volley Ball Indonesia seringkali menjadi topik perbincangan di media

massa dan masyarakat (Kemenpora, 2023). Melalui media sosial Twitter, para penggemar dan masyarakat umum dapat dengan leluasa menyampaikan pendapat, dukungan, atau kritik terhadap Volley Ball Indonesia.

Analisis sentimen pada media sosial telah menjadi topik penelitian yang banyak dieksplorasi dalam beberapa tahun terakhir. Namun, sebagian besar penelitian sebelumnya berfokus pada analisis sentimen di platform media sosial seperti Twitter atau ulasan produk/layanan. Masih sedikit penelitian yang secara khusus menganalisis sentimen pada komentar *netizen* di media sosial Twitter, terutama terkait dengan lingkup olahraga dan tim nasional volley ball seperti Volley Ball Indonesia (Samsir et al., 2021).

Penelitian ini berusaha mengisi celah tersebut dengan berfokus pada analisis sentimen komentar *netizen* di akun Twitter resmi Volley Ball Indonesia (@volleyball.indonesia). Akun Twitter resmi Volley Ball Indonesia (@volleyball.indonesia) memiliki lebih dari 4,2 juta pengikut, yang menunjukkan besarnya basis penggemar Volley Ball Indonesia di media sosial. Komentar-komentar yang diunggah pada postingan di akun ini mencerminkan reaksi *netizen* terhadap berbagai aspek terkait Volley Ball Indonesia, seperti kinerja pemain, hasil pertandingan, strategi pelatih, kebijakan manajemen, dan bahkan isu-isu di luar lapangan (Aditia Rakhmat Sentiaji et al., 2014).

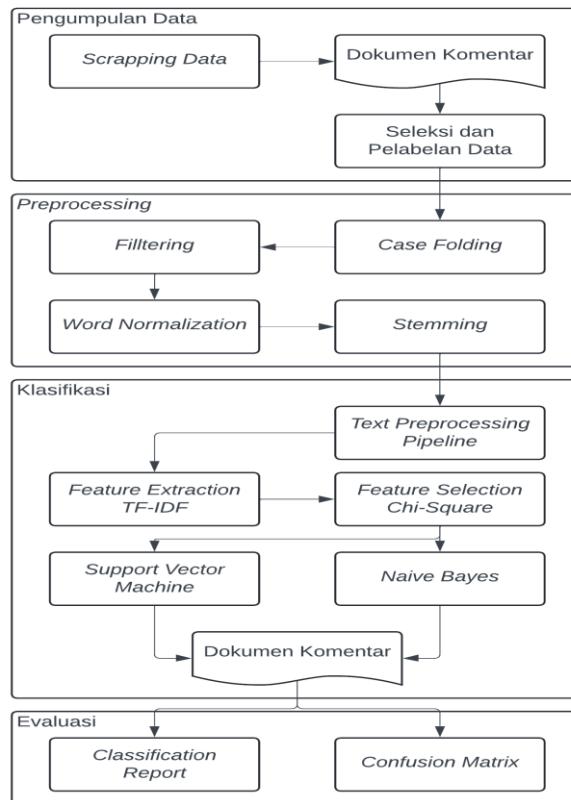
Analisis sentimen terhadap komentar-komentar ini dapat memberikan gambaran tentang persepsi publik terhadap Volley Ball Indonesia, baik sentimen positif maupun negatif. Dalam penelitian ini, metode *Naive Bayes* dan *Support Vector Machine* (SVM) akan digunakan untuk mengklasifikasikan sentimen komentar *netizen* di Twitter Volley Ball Indonesia. *Naive Bayes* adalah metode klasifikasi statistik yang menggunakan probabilitas untuk membuat model prediksi (Handayani & Sulistiyawati, 2021), sementara SVM adalah metode pembelajaran mesin yang mencari *hyperplane* terbaik untuk memisahkan kelas-kelas data (Wati & Ernawati, 2021). Perbandingan kedua metode ini dapat memberikan perspektif yang lebih komprehensif dalam menganalisis sentimen komentar *netizen*.

2. METODE / ALGORITMA

2.1 Prosedur Penelitian

Prosedur penelitian penting dalam sebuah penelitian karena membantu peneliti untuk melakukan penelitian secara terstruktur dan sistematis, sehingga dapat menghasilkan data yang akurat dan valid serta dapat diandalkan dalam mengambil kesimpulan. Dengan memahami prosedur penelitian dengan baik, peneliti dapat merencanakan dan melakukan penelitian dengan lebih efektif dan efisien. Prosedur penelitian biasanya mencakup tahap-tahap seperti perumusan masalah, pengumpulan data, analisis data, dan kesimpulan.

Tujuan dari penelitian yang berjudul “Penerapan Metode *Naive Bayes* Dan *Support Vector Machine* pada Analisis Sentimen *Netizen* di Twitter Volley Ball Indonesia” ini adalah untuk menerapkan dan menentukan model klasifikasi terbaik untuk mendapatkan hasil analisis sentimen komentar yang lebih akurat pada media sosial Twitter Volley Ball Indonesia. Berikut adalah tahapan penelitian ini :



Gambar .1 Tahapan Penelitian

Dari hasil *scrapping* data komentar yaitu 5.000 data komentar, dilakukan pelabelan data ulasan secara manual dan menghasilkan 2920 data komentar, hasilnya didapatkan sentimen positif sebanyak 1.750 data komentar dan sentimen negatif sebanyak 1.170 data komentar.

2.2 Preprocessing

Tahap preprocessing data akan menggunakan data yang dikumpulkan dari komentar netizen di media sosial Twitter Volley Ball Indonesia.



Gambar 2 Tahap Preprocessing

Berdasarkan Gambar 2. menggambarkan tentang analisis proses. Pada tahap *processing* akan menjelaskan langkah awal terhadap teks bagaimana teks tersebut dapat dipersiapkan menjadi data yang diproses pada tahap lebih lanjut. Pada tahap *preprocessing* ada 4 tahap yang harus berurutan yaitu :

1. *Case Folding*

Merubah kata menjadi format yang sama yaitu diubah menjadi huruf kecil. Oleh karena itu, diperlukan proses *case folding* supaya bisa meratakan menjadi huruf kecil.

2. *Filtering*

proses pemilihan kata pada dokumen atau pengurangan dimensi kata di dalam *corpus* yang disebut *stopwords*. *Stopwords* merupakan proses untuk menghilangkan kata yang dianggap tidak memiliki makna.

3. *Word Normalization*

Kata yang tidak ada pengaruhnya pada proses perolehan informasi akan dihapus dan kata yang tidak penting atau tidak berbobot akan dibuang.

4. *Stemming*

Stemming adalah tahap mengubah sebuah kata ke dalam bentuk kata dasarnya dengan menghapus kata imbuhan di depan maupun imbuhan di belakang kata.

2.3 Ekstraksi Fitur dan Seleksi Fitur

Pengambilan dan pemilihan fitur dilakukan setelah proses praproses data selesai. TF-IDF (Term Frequency-Inverse Document Frequency) adalah teknik ekstraksi fitur yang digunakan untuk mengekstraksi fitur yang diinginkan dari kumpulan data. Tahap ekstraksi fitur berikutnya bertujuan untuk mempercepat pemrosesan data serta mengurangi fitur atau atribut yang tidak relevan. Dalam penelitian ini, metode pemilihan fitur adalah Chi-Square.

2.4 Klasifikasi Menggunakan Support Vector Machine dan Naïve Bayes

Tahapan ini dalam penelitian ini melibatkan penggunaan algoritma *Support Vector Machine* dan *Naïve Bayes*. Pada kumpulan komentar netizen di media sosial Twitter Volley Ball Indonesia, yang telah melewati proses ekstraksi dan seleksi fitur. Uji akurasi dilakukan dengan menggunakan tiga presentase rasio pengujian, yaitu 70:30, 80:20, dan 90:10 untuk data pelatihan dan uji, untuk menentukan model klasifikasi terbaik.

2.5 Evaluasi

Tahapan evaluasi dilakukan diakhir proses penelitian. Perhitungan akurasi dan evaluasi model klasifikasi dilakukan dengan metode *Confusion Matrix* dan *Classification Report*. Setelah data uji diujikan dengan data pelatihan, akan dihasilkan daftar kelas-kelas dari data uji, yang dikenal sebagai prediksi kelas. Kemudian prediksi kelas dibandingkan dengan kelas yang sebenarnya dari data uji yang sebelumnya disembunyikan. untuk menunjukkan dan menghitung nilai *accuration*, *precision*, dan *recall*.

3. HASIL DAN PEMBAHASAN

Proses Encoder Label mengubah data label Sentiment menjadi 1 dan 0 negatif. Kemudian data ditampilkan terlebih dahulu dan cek apakah ada *null* didalam data tersebut. Berikut *source code* proses *Label Encoder* :

```
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
data.Sentiment = le.fit_transform(data.Sentiment.values)
```

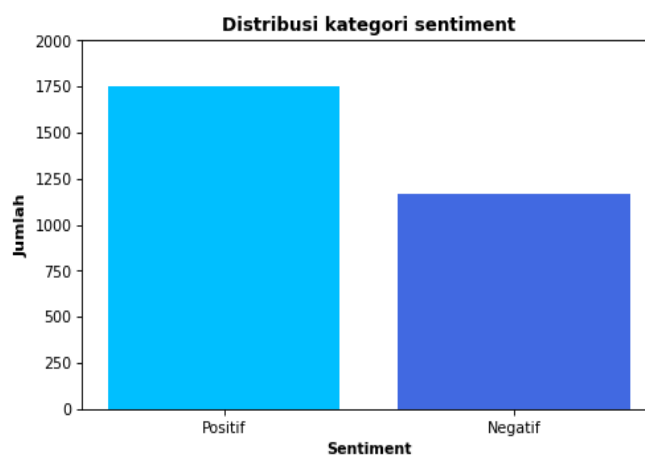
Gambar 3. Source Code tahap Label Encoder

Hasil proses *replace value* ini terdapat pada Gambar 3 berikut.

Tabel .1 Data Hasil Label Encoder

	Username	Twitter Comment Text	Sentiment
0	anditenri20	Bangga bngtðŸŸ³ðŸŸ³ðŸŸ³ðŸŸ³ðŸŸ³	1
1	hindra_pay	Liat 2 kali kebobolan dari sisi kiri,main kanyak apa parah	0
2	rendybrostito	Sudah bagus bisa sampai 16 besar. Maju terus. Lain kali kalau lawan tim besar jangan terlalu idealis cara mainnya. Lebih ke pragmatis aja.	1
3	rizalumami	pak @erickthohir perpanjangan kontrak @shintayong7777 ðŸ™ ðŸ™ ðŸ™ ðŸ™	1
4	ln.rohman18	@hindra_pay et kaga pinter pinter ðŸ˜,	0

Model dibuat untuk menghitung jumlah ulasan positif dan negatif dari 2920 data ulasan.



Gambar 4. Data Jumlah Sentimen Negatif dan Positif

3.1.1 Preprocessing Data

Dataset dari hasil *scrapping* komentar netizen pada media sosial Twitter Volley Ball Indonesia yang telah melewati beberapa proses seperti mengambil atribut yang diperlukan saja, pemberian label secara manual, melakukan *label encoder* pada atribut *Sentiment*. Dari 2920 data komentar, ada 1750 data komentar yang memiliki sentimen positif dan 1170 data yang memiliki sentimen negatif. Preprocessing data penelitian ini meliputi case folding, filtering (penghapusan kata kunci), dan langkah-langkah preprocessing lainnya, *word normalization* dan *stemming*. Raw data yang digunakan untuk sampel adalah data ke 20.

3.1.2 Case Folding

Pada tahap ini, data yang ada akan diproses, seperti menghapus angka, URL, dan karakter tanda baca, serta emoji.

Tabel 2. Hasil Case Folding

Raw Data	ðŸŒ¸ðŸŒ¸ðŸŒ¸ðŸŒ¸ðŸŒ¸ semangat Buat Volley Ball Indonesia
----------	--

Case	semangat buat Volley Ball Indonesia
Folding	

3.1.3 Filtering (Stopword Removal)

Untuk menghilangkan kata atau karakter tanda baca yang memiliki jumlah informasi yang rendah atau ganda, tahap filtering, juga dikenal sebagai penghapusan stopwords, dilakukan. Tahapan ini memerlukan *library nltk* sebagai penyedia data.

Tabel 3. Hasil Filtering (Stopword Removal)

Raw Data	semangat Buat Volley Ball Indonesia
Case Folding	semangat buat Volley Ball Indonesia
Filtering (Stopword Removal)	semangat Volley Ball Indonesia

3.1.4 Word Normalization

Word normalization digunakan untuk mengubah slang menjadi kosa kata baku atau dasar, Peneliti menggunakan data *key norm* atau *slang word* dari GitHub https://github.com/ksnugroho/klasifikasipamsms/blob/master/data/key_norm.csv.

Tabel 4. Hasil Word Normalization

Raw Data	semangat Buat Volley Ball Indonesia
Case Folding	semangat buat Volley Ball Indonesia
Filtering (Stopword Removal)	semangat Volley Ball Indonesia
Word Normalization	semangat Volley Ball Indonesia

3.1.5 Stemming

Tahap *Stemming* berfungsi untuk mencari kata baku atau kata dasar dari sebuah kata yang mengandung atau memiliki imbuhan. Proses ini membutuhkan *library* sastrawi untuk dapat melakukan *stemming*.

Tabel 5. Hasil Stemming

Raw Data	semangat Buat Volley Ball Indonesia
Case Folding	semangat buat Volley Ball Indonesia
Filtering (Stopword Removal)	semangat Volley Ball Indonesia
Word Normalization	semangat Volley Ball Indonesia
Stemming	semangat Volley Ball Indonesia

Selanjutnya, semua fungsi yang telah dibuat pada tahap preprocessing dijalankan pada satu fungsi, Pipeline Pengolahan Teks, yang menghasilkan kolom baru, kolom “*clean teks*” yang merupakan hasil dari *Text Processing Pipeline*. Sebelum masuk ke tahap ekstraksi fitur dan seleksi fitur, dilakukan pendefinisian atribut, dimana X adalah kolom *clean_teks*, dan y kolom *Sentiment*. Hal ini dilakukan untuk memudahkan proses pemanggilan.

3.2 Ekstraksi Fitur dan Seleksi Ftur

Pada tahap ekstraksi fitur, pemetaan fungsional membantu menggabungkan fitur baru dengan fitur asli. Selanjutnya, nilai-nilai yang dihasilkan akan dianalisis pada tahap berikutnya. Untuk mempercepat pemrosesan data, seleksi fitur mengurangi fitur atau atribut yang tidak penting dari dataset. Studi ini menggunakan metode TF-IDF untuk ekstraksi fitur dan metode Chi-Square untuk seleksi fitur.

Ekstraksi fitur bertujuan untuk mendapatkan *Original Feature Number* atau jumlah dan nilai dari yang fitur terseleksi pada ekstraksi fitur. Hasilnya didapatkan *Original Feature Number* pada TF-IDF dengan N-Gram 1,2 yaitu 22714. Sedangkan seleksi fitur bertujuan untuk mendapatkan *Reduced Feature Number* atau jumlah dan nilai dari fitur yang relevan dan terpilih pada seleksi fitur. Hasilnya didapatkan *Reduced Feature Number* pada *ChiSquare* yaitu 1000. Berikut adalah fitur dan nilai dari kedua fitur tersebut.

	aaa	Aaa menang	aaaa	Aaaa nang	aaaaa	Aaaaa senang	aaaaaay	aaddriant	Aaddriant pantesan	aamiin	Aamiin aamiin	Aamiin allah	Aamiin robbal	Aamiin volleyball	aamiinkan
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2631	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2632	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2633	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2634	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2635	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2636	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2637	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2638	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2639	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2640	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Gambar .5 Nilai Hasil TF-IDF

	nilai
0	0.451085
1	0.300723
2	0.571918
3	0.571918
4	0.310682
...	...
22709	0.215473
22710	1.070046
22711	1.070046
22712	0.177011
22713	0.177011

22714 rows × 1 columns

	nilai	fitur
0	0.451085	aaa
1	0.300723	aaa menang
2	0.571918	aaaa
3	0.571918	aaaa nang
4	0.310682	aaaaa
...	...	...
22709	0.215473	zuki harau
22710	1.070046	zulvaaa
22711	1.070046	zulvaaa na
22712	0.177011	zzahro
22713	0.177011	zzahro tuju

22714 rows × 2 columns

Gambar 6. Nilai Hasil *Chi-square* dan Nilai Hasil TF-IDF dan *Chi-Square*

	nilai	fitur
18506	2.437962e+01	semangat
6434	1.460147e+01	garuda
15751	1.116332e+01	out
21303	9.520753e+00	towel
1714	9.470449e+00	bacot
...	...	...
9506	3.786384e-06	kartu merah
16836	3.391677e-06	potong
13886	1.463736e-06	mudah
9638	4.549238e-07	kawan
2380	4.280903e-09	bawa nama

22714 rows × 2 columns

Gambar 7. Nilai Hasil Fitur Terbaik

	aamiin	aapaan	aapaan blunder	abang	acara	adam	adam alis	adek	admin	aduh nadeo	aduhhh	afidid	afidid mlempe	afiffprazetyo	afiffprazetyo hahaha	afk	agam
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
2915	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2916	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2917	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2918	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2919	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

2920 rows × 1000 columns

Gambar 8. Hasil Dari Nilai Fitur-fitur Yang Sudah Di Seleksi

3.3 Machine Learning Model

Fase pelatihan, di mana sistem belajar menemukan pola dalam data, menciptakan model pembelajaran mesin. Salah satu cara untuk menilai model pembelajaran mesin adalah dengan membagi tes latihan. Dataset dibagi menjadi dua bagian dengan metode ini: pelatihan data dan pengujian data. Sementara pengujian data digunakan untuk mengevaluasi kinerja model pembelajaran mesin, pelatihan data digunakan untuk melatihnya. Pada penelitian ini, prediksi pembagian antara pelatihan data dan pengujian data adalah 70:30, 80:20, dan 90:10.

Train-test split digunakan untuk mengukur performa model pada data yang belum pernah diuji sebelumnya, sehingga dapat memberikan estimasi kinerja model pada data yang belum diketahui. Sebuah model dikatakan baik atau optimal jika mampu memberikan hasil prediksi yang akurat pada data baru (*out-of-sample data*), karena tujuan utama dari model *machine learning* adalah untuk dapat melakukan prediksi yang akurat pada data baru yang belum pernah diuji sebelumnya.

Scikit-Learn merupakan salah satu *library* populer untuk *machine learning* di *Python* yang menyediakan berbagai algoritma pembelajaran mesin dan juga berbagai utilitas seperti *train_test_split*. Sementara untuk menyimpan model yang telah dilatih, peneliti menggunakan *library joblib dump*.

4. KESIMPULAN

Dari hasil dari penelitian ini, dapat disimpulkan :

1. Hasil pengujian pada masing-masing data *training* dan data *testing* dengan presentasi yang berbeda akan memberikan hasil akurasi yang berbeda.
2. Hasil pengujian dari metode *Naive Bayes* didapatkan nilai akurasi tertinggi sebesar 71% dengan rasio 70:30 dan *Support Vector Machine* mendapatkan nilai akurasi tertinggi sebesar 76% dengan rasio 80:20. Maka dapat disimpulkan bahwa metode *Support Vector Machine* mendapatkan nilai akurasi lebih tinggi dari pada metode *Naive Bayes*.

REFERENSI

- [1] Aditia Rakhmat Sentiaji, A. M. B., Sarjana, P. S., Statistika, D., Matematika, F., Ilmu, D. A. N., & Alam, P. (2014). Analisis Sentimen Terhadap Acara Televisi Berdasarkan Opini Publik. *Jurnal Ilmiah Komputer Dan Informatika (KOMPUTA)*.
- [2] Akbar, Y., & Ihsan, A. N. (2023). Analisis Sentimen Twitter Terhadap Opini Masyarakat Pada Sea Games Kamboja 2023 Menggunakan Algoritma Support Vector Machine. *INTECOMS: Journal of Information Technology and Computer Science*, 6(2). <https://doi.org/10.31539/intecom.s.v6i2.7670>
- [3] Amanah, D., & Harahap, D. A. (2018). Examining The Effect Of Product Assortment And Price Discount Toward Online Purchase Decision Of University Student In Indonesia. *Jurnal Manajemen Dan Kewirausahaan*, 20(2). <https://doi.org/10.9744/jmk.20.2.99-104>
- [4] Amrizal, V. (2018). Penerapan Metode Term Frequency Inverse Document Frequency (Tf-Idf) Dan Cosine Similarity Pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus: Hadits Shahih Bukhari-Muslim). *Jurnal Teknik Informatika*, 11(2). <https://doi.org/10.15408/jti.v11i2.8623>
- [5] Andika, L. A., Azizah, P. A. N., & Respatiwan, R. (2019). Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier. *Indonesian Journal of Applied Statistics*, 2(1). <https://doi.org/10.13057/ijas.v2i1.29998>
- [6] Arum, D. S., Butsianto, S., Astuti, R., & Pelita Bangsa, U. (2023). Analisis Sentimen Masyarakat Indonesia Terhadap Sea Games 2023 Di Twitter Dengan Metode Naïve Bayes. *Journal of Information System, Applied, Management, Accounting and Research*, 7(3).
- [7] Fattah, St. F., & Purnawansyah. (2022). Analisis sentimen terhadap Body Shaming pada Twitter menggunakan Metode Naïve Bayes Classifier. *Indonesian Journal of Data and Science*, 3(2). <https://doi.org/10.56705/ijodas.v3i2.46>
- [8] Gunawan, B., Pratiwi, H. S., & Pratama, E. E. (2018). Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 4(2). <https://doi.org/10.26418/jp.v4i2.27526>
- [9] Handayani, E. T., & Sulistiyawati, A. (2021). Analisis Sentimen Respon Masyarakat Terhadap Kabar Harian Covid-19 Pada Twitter Kementerian Kesehatan Dengan Metode Klasifikasi Naive Bayes. *Jurnal Teknologi Dan Sistem Informasi (JTISI)*, 2(3), 32–37. <http://jim.teknokrat.ac.id/index.php/JTISI>
- [10] Hardiani, T., Sulistyo, S., & Hartanto, R. (2014). Kajian Data Mining Customer Relationship Management pada Lembaga Keuangan Mikro. *JURNAL SISTEM INFORMASI BISNIS*, 4(3). <https://doi.org/10.21456/vol4iss3pp181-186>
- [11] Hudaya, C. S., Fakhrurroja, H., & Alamsyah, A. (2019). Analisis Persepsi Konsumen Terhadap Brand Go-Jek Pada Media Sosial Twitter Menggunakan Metode Sentiment Analysis Dan Topic Modelling. *Jurnal Mitra Manajemen*, 3(6).

- <https://doi.org/10.52160/ejmm.v3i6.244>
- [12] Kemenpora. (2023, March 10). *Sejarah Tembus Babak 16 Besar Piala Asia 2023, Menpora Dito Sebut Prestasi Volley Ball Indonesia Alami Peningkatan.*
- [13] Kurniawan, S., Gata, W., Puspitawati, D. A., -, N., Tabrani, M., & Novel, K. (2019). Perbandingan Metode Klasifikasi Analisis Sentimen Tokoh Politik Pada Komentar Media Berita Online. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 3(2). <https://doi.org/10.29207/resti.v3i2.935>
- [14] Nardilasari, A. P., Hananto, A. L., Hilabi, S. S., Tukino, T., & Priyatna, B. (2023). Analisis Sentimen Calon Presiden 2024 Menggunakan Algoritma SVM Pada Media Sosial Twitter. *JOINTECS (Journal of Information Technology and Computer Science)*, 8(1), 11. <https://doi.org/10.31328/jointecs.v8i1.4265>
- [15] Noor, L. N. (2015). Peran Media Sosial Dalam Proses Sosio-Edukasi Empat Mahasiswa Pendidikan Sosiologi Universitas Negeri Jakarta. *Universitas Negeri Jakarta.*
- [16] Nurjannah, M., & Fitri Astuti, I. (2013). Penerapan Algoritma Term Frequency-Inverse Document Frequency (Tf-Idf) Untuk Text Mining Mahasiswa S1 Program Studi Ilmu Komputer FMIPA Universitas Mulawarman Dosen Program Studi Ilmu Komputer FMIPA Universitas Mulawarman. *Jurnal Informatika Mulawarman*, 8(3).
- [17] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12.
- [18] Pertiwi, M. W. (2019). Analisis Sentimen Opini Publik Mengenai Sarana Dan Transportasi Mudik Tahun 2019 Pada Twitter Menggunakan Algoritma Naïve Bayes, Neural Network, Knn Dan Svm. *Inti Nusa Mandiri*, 14(1).
- [19] Prabowo, D. A., Fhadli, M., Najib, M. A., Fauzi, H. A., & Cholissodin, I. (2016). TF-IDF-Enhanced Genetic Algorithm Untuk Extractive Automatic Text Summarization. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 3(3). <https://doi.org/10.25126/jtiik.201633217>
- [20] Pramudita, Y. D., Putro, S. S., & Makhmud, N. (2018). Klasifikasi Berita Olahraga Menggunakan Metode Naïve Bayes dengan Enhanced Confix Stripping Stemmer. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 5(3). <https://doi.org/10.25126/jtiik.201853810>
- [21] Pratama, S. F., Andrean, R., & Nugroho, A. (2019). Analisis Sentimen Twitter Debat Calon Presiden Indonesia Menggunakan Metode Fined-Grained Sentiment Analysis. *JOINTECS (Journal of Information Technology and Computer Science)*, 4(2). <https://doi.org/10.31328/jointecs.v4i2.1004>
- [22] Pratiwi, S. N. D., & Ulama, B. S. S. (2016). Klasifikasi Email Spam dengan Menggunakan Metode Support Vector Machine dan k-Nearest Neighbor. *Jurnal Sains Dan Seni ITS*, 5(2).
- [23] Primartha, R., & Wahono, R. S. (2021). *Algoritma machine learning*. Informatika Bandung.
- [24] Salsabila, S. M., Alim Murtopo, A., & Fadhilah, N. (2022). Analisis Sentimen Pelanggan Tokopedia Menggunakan Metode Naïve Bayes Classifier. *Jurnal Minfo Polgan*, 11(2). <https://doi.org/10.33395/jmp.v11i2.11640>
- [25] Samsir, S., Ambiyar, A., Verawardina, U., Edi, F., & Watrianthos, R. (2021). Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes. *Jurnal Media Informatika Budidarma*, 5(1). <https://doi.org/10.30865/mib.v5i1.2580>
- [26] Santoso, H., Hariyadi, P., Teknik,), Stmik, I., Mataram, B., Ismail, J., & Mataram, M. *Penerapan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Netizen di Twitter Volley Ball Indonesia*

- (2016). Data Mining Analisa Pola Pembelian Produk Dengan Menggunakan Metode Algoritma Apriori. *Semnasteknomedia Online*, 4(1).
- [27] Saputra, A. (2023). Perbandingan Metode Naïve Bayes Classifier Dan Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter Mengenai Piala Dunia Fifa 2022. *Teknomatika*.
- [28] Setiawan, H., & Zufria, I. (2023). Analisis Sentimen Pembatalan Indonesia Sebagai Tuan Rumah Piala Dunia FIFA U-20 Menggunakan Naïve Bayes. *Jurnal Media Informatika Budidarma*, 7(3).
- [29] Suntoro, J., Christanto, F. W., & Indriyawati, H. (2018). Software Defect Prediction Using AWEIG+ADACOST Bayesian Algorithm for Handling High Dimensional Data and Class Imbalance Problem. *International Journal of Information Technology and Business*, 1(1). <https://doi.org/10.24246/ijiteb.112018.36-41>
- [30] Wati, R., & Ernawati, S. (2021). Analisis Sentimen Persepsi Publik Mengenai PPKM Pada Twitter Berbasis SVM Menggunakan Python. *Jurnal Teknik Informatika UNIKA Santo Thomas*, 240–247. <https://doi.org/10.54367/jtiust.v6i2.1465>
- [31] Werdiningsih, I., Nurqoba, B., & Mohammadun. (2020). *Data Mining Menggunakan Android, Weka, dan SPSS*. Airlangga University Press.