



## Perbandingan Naive Bayes Classifier dan SVM untuk Analisis Sentimen Desain Seragam Atlet Indonesia pada Media Sosial X di Olimpiade Paris 2024

Azizah Salma Nida<sup>1</sup>, Irwansyah<sup>2\*</sup>, Ade Davy Wiranata<sup>3</sup>

<sup>1,2,3</sup>Teknik Informatika, Fakultas Teknologi Industri dan Informatika, Universitas Muhammadiyah Prof. Dr. Hamka, Jakarta, Indonesia

\*Corresponding Author: [irwansyah@uhamka.ac.id](mailto:irwansyah@uhamka.ac.id)

### ARTICLE INFO

Vol. 1, No. 2, April 2026, pp. 107-118.  
DOI: <https://doi.org/10.66217/digintel-ai.v1i2.13>

#### Article history:

Revised: 15 April 2026  
Accepted: 19 April 2026  
Published: 30 April 2026

### Abstract

*The Olympics is an international sporting event held every four years and serves as a platform for countries to showcase their athletic capabilities and national identity. One aspect that attracts public attention is the design of athletes' uniforms, which not only have aesthetic value but also support athletic performance. Differences in public perception of these designs generate various opinions expressed on social media X. This study aims to analyze public sentiment toward the design of Indonesian athletes' uniforms at the Paris 2024 Olympics on social media X and to compare the performance of Naive Bayes Classifier and Support Vector Machine algorithms. The dataset consists of textual data collected from social media X and processed through preprocessing stages and split into training and testing data with an 80:20 ratio. The results show that there are 1,014 positive and 728 negative sentiments. Model evaluation indicates that the Naive Bayes Classifier achieved an accuracy of 80.5%, while the Support Vector Machine achieved 94.2%, outperforming the former. These findings demonstrate that the Support Vector Machine is more effective than the Naive Bayes Classifier for sentiment analysis of social media text data related to the design of Indonesian athletes' uniforms at the Paris 2024 Olympics.*

**Keywords:** Sentiment Analysis; Naive Bayes Classifier; Support Vector Machine; Social Media X; Paris 2024 Olympics.

### Abstrak

Olimpiade merupakan ajang olahraga internasional yang diselenggarakan setiap empat tahun dan menjadi sarana bagi suatu negara untuk menunjukkan kemampuan atletik serta identitas nasional. Salah satu aspek yang mendapat perhatian publik adalah desain seragam atlet, yang tidak hanya memiliki nilai estetika, tetapi juga mendukung performa atlet. Perbedaan persepsi masyarakat terhadap desain tersebut memunculkan berbagai opini yang disampaikan melalui media sosial X. Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap desain seragam atlet Indonesia pada Olimpiade Paris 2024 di media sosial X serta membandingkan kinerja algoritma *Naive Bayes Classifier* dan *Support Vector Machine*. Dataset yang digunakan merupakan data teks yang diperoleh dari media sosial X dan diproses melalui tahapan *preprocessing* serta pembagian data latih dan data uji dengan rasio 80:20. Hasil penelitian menunjukkan bahwa Terdapat 1014 sentimen positif dan 728 sentimen negatif. Berdasarkan evaluasi model, algoritma *Naive Bayes Classifier* menghasilkan akurasi sebesar 80,5%, sedangkan *Support Vector Machine* menunjukkan akurasi sebesar 94,2%, yang lebih tinggi. Temuan ini menunjukkan bahwa *Support Vector Machine* lebih efektif daripada *Naive Bayes Classifier* dalam



Copyright: © 2026 by the authors. Licensee DIGINTEL AI, Jakarta,

Indonesia. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0)

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

menganalisis sentimen data teks media sosial yang berkaitan dengan desain seragam atlet Indonesia pada Olimpiade Paris 2024.

**Kata Kunci:** Analisis Sentimen; *Naive Bayes Classifier*; *Support Vector Machine*; Media Sosial X; Olimpiade Paris 2024.

## 1. Pendahuluan

Perkembangan internet dan teknologi digital telah mendorong peningkatan signifikan dalam penggunaan media sosial di berbagai lapisan masyarakat. Media sosial tidak hanya berperan sebagai sarana komunikasi, tetapi juga sebagai platform bagi individu untuk mengekspresikan opini, persepsi, serta respons terhadap berbagai fenomena social (Hu, 2025). Tingginya intensitas aktivitas pengguna menghasilkan data dalam jumlah besar yang bersifat tidak terstruktur, sehingga diperlukan metode analisis yang tepat untuk mengekstraksi informasi yang bermakna (Taha, 2025).

Salah satu platform yang digunakan untuk menyampaikan opini publik adalah media sosial X, yang memungkinkan pengguna membagikan pandangan secara terbuka dalam bentuk teks singkat. Karakteristik data pada platform ini yang bersifat real-time dan dinamis menjadikannya sebagai sumber yang potensial untuk analisis *sentiment* (Weng et al., 2025). Dalam konteks ini, analisis sentimen merupakan salah satu pendekatan dalam *text mining* yang digunakan untuk mengidentifikasi polaritas opini, seperti sentimen positif dan negatif, dari data teks (Ramezani, 2025).

Dalam ranah olahraga internasional, Olimpiade merupakan ajang prestisius yang tidak hanya menampilkan kemampuan atletik, tetapi juga berfungsi sebagai media representasi identitas nasional suatu negara. Salah satu elemen penting dalam representasi tersebut adalah desain seragam atlet, yang mencerminkan nilai estetika sekaligus identitas budaya. Pada Olimpiade Paris 2024, desain seragam atlet Indonesia menjadi perhatian publik dan memicu berbagai tanggapan di media sosial X, baik dalam bentuk sentimen positif maupun negative (Purba et al., 2025).

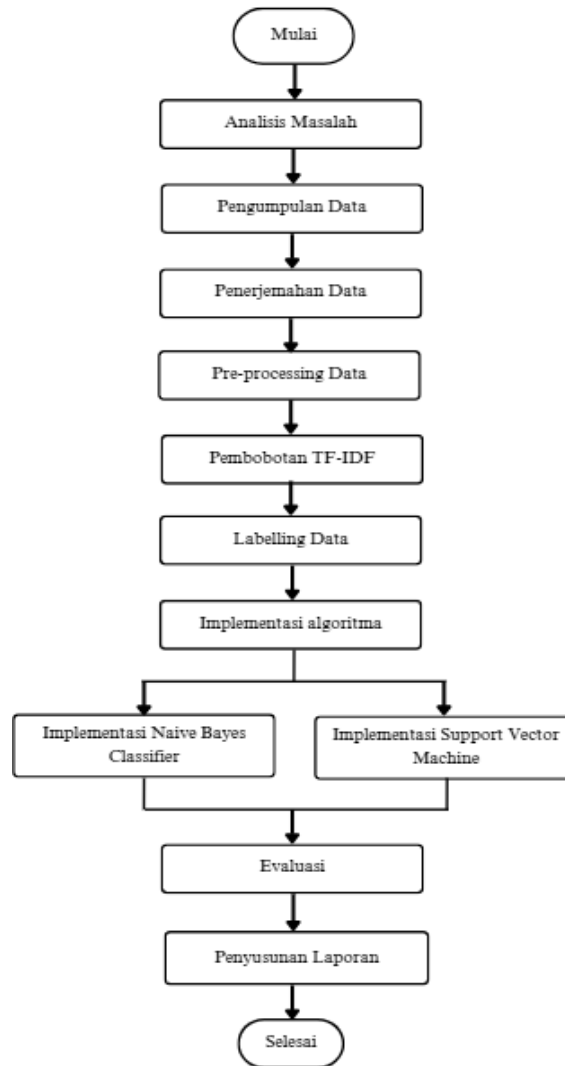
Berbagai penelitian sebelumnya menunjukkan bahwa algoritma *Naive Bayes Classifier* dan *Support Vector Machine* banyak digunakan dalam analisis sentimen berbasis data teks media sosial. Penelitian oleh Prastia & Asra, (2026) menunjukkan bahwa *Support Vector Machine* memiliki kinerja yang lebih unggul dibandingkan *Naive Bayes* dalam mengklasifikasikan ulasan pengguna, yang ditunjukkan melalui tingkat akurasi yang lebih tinggi. Selain itu, penelitian lain juga mengindikasikan bahwa kedua algoritma tersebut efektif dalam mengklasifikasikan opini publik pada berbagai domain, seperti kebijakan publik dan layanan digital (Anadas & Suryono, 2025; Purnomo et al., 2025). Lebih lanjut, analisis sentimen pada media sosial X semakin relevan karena kemampuannya dalam merepresentasikan opini masyarakat secara langsung dan real-time (Mahmudah & Yudhistira, 2025).

Meskipun demikian, sebagian besar penelitian terdahulu masih berfokus pada domain umum, seperti kebijakan publik, layanan digital, dan isu sosial lainnya (Chaudhuri et al., 2023). Kajian yang secara spesifik mengangkat analisis sentimen terhadap desain seragam atlet Indonesia pada Olimpiade Paris 2024 masih terbatas, khususnya yang membandingkan kinerja antara algoritma *Naive Bayes Classifier* dan *Support Vector Machine*.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap desain seragam atlet Indonesia pada Olimpiade Paris 2024 di media sosial X serta mengevaluasi dan membandingkan kinerja algoritma *Naive Bayes Classifier* dan *Support Vector Machine*. Penelitian ini memberikan kontribusi dalam memahami persepsi publik serta dalam menentukan metode yang lebih efektif untuk analisis sentimen pada data teks media sosial.

## 2. Metodologi

Gambar 1 menyajikan tahapan metodologi penelitian yang digunakan dalam penelitian ini.



Gambar 1. Tahapan Metodologi Penelitian

Gambar 1 menyajikan tahapan metodologi penelitian yang dimulai dari analisis masalah hingga penyusunan laporan.

### 2.1 Analisis Masalah

Pada tahap ini dilakukan identifikasi permasalahan serta penentuan kata kunci untuk mengamati respons masyarakat di media sosial X terhadap desain seragam atlet Indonesia pada Olimpiade Paris 2024. Tahap ini bertujuan untuk memperoleh pemahaman awal mengenai pola opini yang berkembang di masyarakat (Novryadi et al., 2026).

### 2.2 Pengumpulan Data

Data dikumpulkan dari media sosial X menggunakan *Google Colab* dengan kata kunci “Paris 2024”, “seragam timnas Paris 2024”, dan “jersey timnas Indonesia Paris 2024”. Dari proses ini diperoleh sebanyak

1.742 data komentar (Kurniawan et al., 2026). Selain itu, dilakukan studi literatur dari jurnal dan artikel ilmiah sebagai dasar teoritis penelitian.

### 2.3 Penerjemahan Data

Data yang diperoleh memiliki variasi bahasa, sehingga dilakukan proses penerjemahan untuk menyeragamkan bahasa. Bahasa Inggris dipilih sebagai standar untuk menjaga konsistensi dalam analisis sentimen tanpa mengubah makna emosional dari teks.

### 2.4 Pre-processing Data

Data yang telah dikumpulkan diproses melalui tahap *preprocessing* untuk menghilangkan noise (Anisah & Irwansyah, 2025). Tahapan ini meliputi:

- a. *Cleaning* untuk menghapus URL, mention, hashtag, dan simbol
- b. *Case folding* untuk mengubah teks menjadi huruf kecil
- c. *Tokenization* untuk memecah kalimat menjadi kata
- d. *Stopword removal* untuk menghapus kata tidak penting
- e. *Lemmatization* untuk mengubah kata ke bentuk dasar

### 2.5 Pembobotan TF-IDF

Metode TF-IDF digunakan untuk mengubah data teks menjadi representasi numerik. TF mengukur frekuensi kemunculan kata, sedangkan IDF memberikan bobot lebih pada kata yang jarang muncul, sehingga menghasilkan representasi fitur yang lebih informatif (Putri et al., 2026).

### 2.6 Labelling Data

Pelabelan data dilakukan menggunakan pendekatan leksikon dengan memberikan skor sentimen pada setiap kata. Skor total digunakan untuk menentukan label, yaitu sentimen positif jika nilai  $> 0$  dan negatif jika nilai  $< 0$ .

### 2.7 Implementasi Algoritma

Data diklasifikasikan menggunakan algoritma *Naive Bayes Classifier* dan *Support Vector Machine* dengan pendekatan *binary classification* (positif dan negatif) (Irwansyah et al., 2023). Proses dilakukan menggunakan Python pada Google Colab.

### 2.8 Evaluasi

Kinerja model dievaluasi menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score* untuk mengukur performa klasifikasi (Irwansyah et al., 2024).

### 2.9 Penyusunan Laporan

Tahap akhir adalah penyusunan laporan berdasarkan hasil analisis dan perbandingan kinerja kedua algoritma terhadap data sentimen desain seragam atlet Indonesia pada Olimpiade Paris 2024.

## 3. Hasil dan Pembahasan

### 3.1 Implementasi Algoritma Naive Bayes Classifier

Pembagian dataset dilakukan dengan perbandingan 80:20, dengan rincian 80% untuk memahami pola dan karakteristik dataset sedangkan 20% untuk menguji evaluasi kinerja dari data latih. Pemilihan data uji dan data latih dilakukan secara acak untuk mewakili distribusi data secara keseluruhan, sehingga bisa menampilkan

seberapa baik model diterapkan pada data yang baru. Dengan fungsi *train test split* dari *sklearn* untuk membagi data uji dan data latih pada dataset seperti proses dan hasil pada Gambar 2.

```
from sklearn.model_selection import train_test_split

X = tfidf_df
y = df['sentimen']
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42
)
# Menampilkan ukuran data
print("Ukuran data latih:", X_train.shape)
print("Ukuran data uji:", X_test.shape)

Ukuran data latih: (1393, 1653)
Ukuran data uji: (349, 1653)
```

Gambar 2. Pembagian Data latih dan Data uji

Implementasi algoritma *naïve bayes classifier* sebagai penghitung kemungkinan data uji masuk dalam sentimen positif dan negatif dengan melihat pola pada data latih berdasarkan fitur dalamnya dengan asumsi bahwa tiap fitur berbeda satu sama lain. Dengan kelas *multinomialNB* dari *sklearn naive\_bayes* untuk implementasi algoritma *naive bayes*, lalu dengan *sklearn.metrics* untuk menghitung nilai prediksi akurasi, dengan objek *nb=multinomialNB()* melakukan data latih *x\_train* (fitur) dan *y\_train* (label) dipelajari probabilitas syarat fitur kelas berdasarkan data latih dan dilakukan perhitungan pada data uji (*x\_test*) dengan hasil akurasi 0,805. Dengan proses implementasi algoritma dan hasil pada akurasi Gambar 3.

```
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import accuracy_score, precision_score

nb = MultinomialNB()
# Melatih model dengan data latih
nb.fit(X_train, y_train)

# Melakukan prediksi pada data uji
y_pred = nb.predict(X_test)
akurasi = accuracy_score(y_test, y_pred)
print("Akurasi Naive Bayes Classifier:", akurasi)

Akurasi Naive Bayes Classifier: 0.8051575931232091
```

Gambar 3. Implementasi Naive Bayes Classifier

### 3.2 Implementasi Algoritma Support Vector Machine

Implementasi algoritma Support Vector Machine dilakuakn dengan maksud mencari garis pemisah (*hyperplane*) untuk mengoptimalkan margin antar kelas dan efisiensi pemisahan data. Dengan penggunaan fungsi *svc* dari *library sklearn.svm* dan fungsi *accuracy\_score* dari *library sklearn.metrics*, akan dibuat model untuk melakukan proses data latih (*x\_train*) dan label target (*y\_train*) sebagai pencarian garis pemisah optimal

yang akan dilakukan pengujian dengan data uji ( $x_{test}$ ) menentukan titik pada tiap sisi kelas, hasilnya akan disimpan pada variabel ( $y_{pred}$ ) dan terakhir akan dilakukan perhitungan perbandingan label aktual ( $y_{test}$ ) dengan label prediksi ( $y_{pred}$ ) yang menghasilkan nilai akurasi sebesar 0,942. Seperti pada gambar 4.16.

```
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score

svm_model = SVC(kernel='linear', random_state=42)
svm_model.fit(X_train, y_train)
y_pred = svm_model.predict(X_test)
akurasi = accuracy_score(y_test, y_pred)
print("Akurasi SVM:", akurasi)
```

---

Akurasi SVM: 0.9426934097421203

Gambar 4. Implementasi Support Vector Machine

### 3.3 Evaluasi Hasil Implementasi Algoritma Naive Bayes

*Confusion matrix* diterapkan untuk menilai hasil implementasi algoritma *naive bayes classifier* dengan menampilkan jumlah prediksi dan aktual masing-masing kelas, membuat rinci kemampuan model untuk klasifikasikan data. Dengan menghitung perbandingan label sebenarnya dan hasil prediksi model seperti proses pada Gambar 5 dan hasil pada Gambar 6.

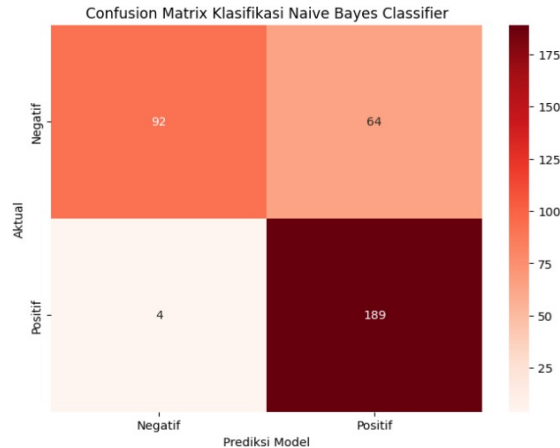
```
import seaborn as sns
import matplotlib.pyplot as plt

cm = confusion_matrix(y_test, y_pred, labels=['negatif', 'positif'])

plt.figure(figsize=(8, 6))
sns.heatmap(
    cm,
    annot=True,
    fmt='d',
    cmap='Reds',
    xticklabels=['Negatif', 'Positif'],
    yticklabels=['Negatif', 'Positif']
)
plt.title('Confusion Matrix Klasifikasi Naive Bayes Classifier')
plt.xlabel('Prediksi Model')
plt.ylabel('Aktual')
plt.show()
```

Gambar 5. Proses Evaluasi NBC

Dengan menggunakan *library* seaborn dan matplotlib.pyplot untuk menampilkan visualisasi data, dengan label aktual label aktual ( $y_{test}$ ) dan label prediksi model ( $y_{pred}$ ) yang akan menghasilkan variabel cm berisi nilai *True Negatif* (TN)=92, *True Positif* (TP)=189, *False Positif* (FP)=64 dan *False Negatif* (FN)=4. Dan diperoleh nilai akurasi 80,5%, presisi 74,4%, *recall* 97,9% dan *f1-Score* 84,5%. Dengan hasil visualisasi pada Gambar 6 dan rincian perhitungan sebagai berikut:



Gambar 6. *Confusion Matrix NBC*

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{189+92}{189+92+64+4} = \frac{281}{349} = 0.805 \times 100\% = 80.5\% \quad (1)$$

$$\text{Presisi} = \frac{TP}{TP+FP} = \frac{189}{189+64} = \frac{189}{254} = 0.744 \times 100\% = 74.4\% \quad (2)$$

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{189}{189+4} = \frac{189}{193} = 0.979 \times 100\% = 97.9\% \quad (3)$$

$$\text{F1 - Score} = 2 \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} = 2 \frac{0.744 \times 0.979}{0.744 + 0.979} = 2 \frac{0.728}{1.723} = 0.845 \times 100\% = 84.5\% \quad (4)$$

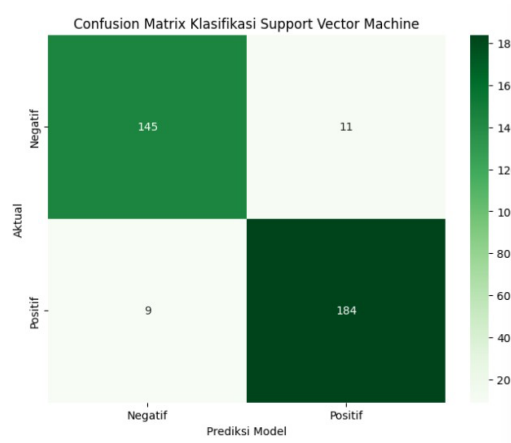
### 3.4 Evaluasi Hasil Implementasi Algoritma Support Vector Machine

Masih dengan penggunaan *library* seaborn dan matplotlib.pyplot untuk membuat visualisasi data, dengan label aktual label aktual ( $y_{test}$ ) dan label prediksi model ( $y_{pred}$ ) yang akan menghasilkan variabel cm dengan proses seperti pada Gambar 7.

```
cm = confusion_matrix(y_test, y_pred, labels=['negatif', 'positif'])

plt.figure(figsize=(8, 6))
sns.heatmap(
    cm,
    annot=True,
    fmt='d',
    cmap='Greens',
    xticklabels=['Negatif', 'Positif'],
    yticklabels=['Negatif', 'Positif']
)
plt.title('Confusion Matrix Klasifikasi Support Vector Machine')
plt.xlabel('Prediksi Model')
plt.ylabel('Aktual')
plt.show()
```

Gambar 7. Proses Evaluasi SVM



Gambar 8. *Confusion Matrix SVM*

Dari visualisasi diperoleh untuk *True Negatif* (TN) = 145, *True Positif* (TP) = 184, *False Positif* (FP) = 11 dan *False Negatif* (FN) = 9. Dan diperoleh nilai akurasi 94,2%, presisi 94,3%, recall 95,3% dan f1-score 94,7%. Dengan rincian perhitungan sebagai berikut:

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{184+145}{184+145+11+9} = \frac{329}{349} = 0.942 \times 100\% = 94.2\% \quad (5)$$

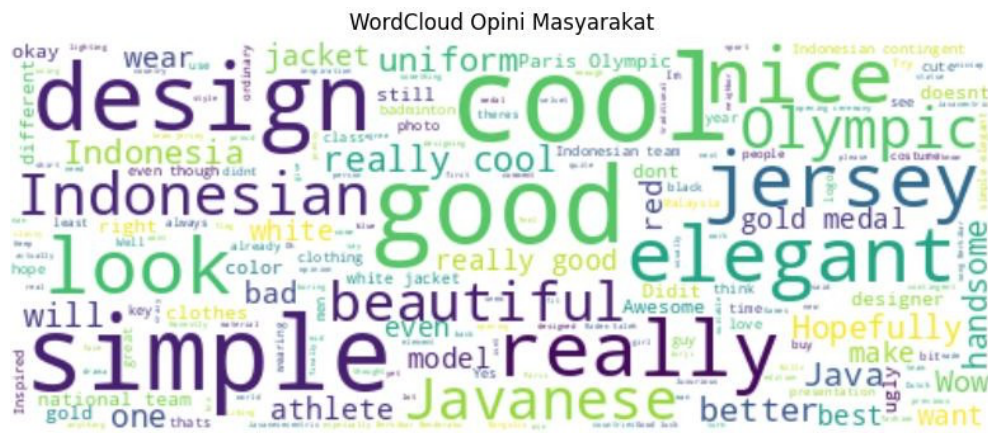
$$\text{Presisi} = \frac{TP}{TP+FP} = \frac{184}{184+11} = \frac{184}{195} = 0.943 \times 100\% = 94.3\% \quad (6)$$

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{184}{184+9} = \frac{184}{193} = 0.953 \times 100\% = 95.3\% \quad (7)$$

$$\text{F1 - Score} = 2 \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} = 2 \frac{0.943 \times 0.953}{0.943 + 0.953} = 2 \frac{0.898}{1.896} = 0.947 \times 100\% = 94.7\% \quad (8)$$

### 3.5 Visualisasi Word Cloud

Dari respon masyarakat terhadap desain seragam atlet Indonesia diperoleh kata dengan frekuensi sering muncul, makin besar kata yang terlihat makin besar juga frekuensi kata yang muncul pada hasil visualisasi. Dengan menggunakan fungsi wordcloud untuk menampilkan visualisasi kata berdasarkan frekuensi kemunculannya pada data *tweets* seperti hasil pada Gambar 9.



Gambar 9. Visualisasi Word Cloud

## Pembahasan

Hasil penelitian menunjukkan bahwa analisis sentimen terhadap desain seragam atlet Indonesia pada Olimpiade Paris 2024 didominasi oleh sentimen positif, dengan jumlah 1.014 data positif dan 728 data negatif. Dominasi sentimen positif ini mengindikasikan bahwa mayoritas masyarakat memberikan respons yang baik terhadap desain yang ditampilkan. Temuan ini juga diperkuat oleh visualisasi *word cloud* yang menunjukkan kemunculan kata-kata dominan seperti “cool”, “good”, “simple”, “elegant”, dan “beautiful”. Kata-kata tersebut merepresentasikan persepsi publik yang menilai desain seragam sebagai kombinasi antara kesederhanaan dan estetika yang menarik. Selain itu, kemunculan kata seperti “Indonesian” dan “jersey” menunjukkan bahwa masyarakat tidak hanya menilai dari aspek visual, tetapi juga mengaitkannya dengan identitas nasional. Dengan demikian, hasil analisis kualitatif melalui visualisasi data sejalan dengan hasil klasifikasi kuantitatif yang menunjukkan dominasi sentimen positif.

Dari sisi performa model, hasil evaluasi menunjukkan bahwa algoritma Naive Bayes Classifier menghasilkan akurasi sebesar 80,5%, sedangkan Support Vector Machine (SVM) menunjukkan performa yang lebih unggul dengan akurasi sebesar 94,2%. Meskipun Naive Bayes memiliki keunggulan dalam nilai recall yang tinggi, yaitu sebesar 97,9%, yang menunjukkan kemampuannya dalam mengidentifikasi sebagian besar data positif, namun nilai precision yang lebih rendah mengindikasikan adanya kesalahan klasifikasi, khususnya pada data negatif yang terklasifikasi sebagai positif. Sebaliknya, SVM menunjukkan performa yang lebih seimbang dengan nilai precision, recall, dan F1-score yang tinggi, sehingga mampu memberikan hasil klasifikasi yang lebih akurat dan stabil. Keunggulan SVM ini disebabkan oleh kemampuannya dalam membentuk *hyperplane* optimal yang memaksimalkan margin antar kelas, sehingga lebih efektif dalam menangani data teks yang kompleks dan berdimensi tinggi seperti data dari media sosial.

Jika dibandingkan dengan penelitian sebelumnya, hasil penelitian ini konsisten dengan temuan yang menyatakan bahwa SVM memiliki kinerja yang lebih unggul dibandingkan *Naive Bayes* dalam analisis sentimen. Beberapa penelitian terdahulu menunjukkan bahwa SVM mampu memberikan tingkat akurasi yang lebih tinggi serta performa yang lebih stabil dalam mengklasifikasikan data teks, terutama pada data yang memiliki kompleksitas tinggi dan tidak terstruktur. Dengan demikian, hasil penelitian ini tidak hanya memperkuat temuan sebelumnya, tetapi juga memberikan kontribusi baru dengan mengkaji secara spesifik sentimen masyarakat terhadap desain seragam atlet Indonesia pada Olimpiade Paris 2024, yang masih jarang dibahas dalam literatur.

Hasil penelitian ini juga dipengaruhi oleh beberapa faktor penting, antara lain kualitas data yang digunakan, proses *preprocessing*, metode pembobotan fitur, serta karakteristik algoritma yang diterapkan. Data yang berasal dari media sosial X memiliki sifat tidak terstruktur dan mengandung berbagai noise seperti slang, singkatan, dan simbol, sehingga memerlukan proses *preprocessing* yang optimal untuk meningkatkan kualitas data. Penggunaan metode TF-IDF dalam pembobotan kata terbukti mampu meningkatkan representasi fitur sehingga membantu model dalam membedakan sentimen secara lebih efektif. Selain itu, perbedaan karakteristik algoritma juga mempengaruhi hasil, di mana *Naive Bayes* menggunakan pendekatan probabilistik dengan asumsi independensi fitur, sedangkan SVM lebih unggul dalam menangani data berdimensi tinggi dengan membentuk batas keputusan yang optimal. Distribusi data yang didominasi oleh sentimen positif juga turut mempengaruhi performa model dalam mempelajari pola data.

Penelitian ini memiliki beberapa kelebihan, di antaranya penggunaan data *real-time* dari media sosial X yang mencerminkan kondisi aktual opini masyarakat, serta penerapan dua algoritma populer yang dibandingkan secara komprehensif menggunakan berbagai metrik evaluasi. Selain itu, kombinasi antara analisis kuantitatif dan visualisasi data memberikan pemahaman yang lebih mendalam terhadap pola sentimen yang terbentuk. Namun demikian, penelitian ini juga memiliki keterbatasan, seperti jumlah *dataset* yang relatif terbatas, penggunaan metode pelabelan berbasis leksikon yang berpotensi menghasilkan bias, serta belum mempertimbangkan konteks

kalimat secara mendalam seperti sarkasme atau ironi. Selain itu, penelitian ini belum memanfaatkan pendekatan berbasis *deep learning* yang berpotensi memberikan hasil yang lebih optimal.

Secara keseluruhan, penelitian ini memberikan implikasi yang signifikan baik secara teoritis maupun praktis. Secara teoritis, hasil penelitian ini memperkuat literatur yang menyatakan bahwa SVM merupakan metode yang lebih efektif dalam analisis sentimen berbasis teks, khususnya pada data media sosial. Secara praktis, hasil penelitian ini dapat digunakan sebagai referensi bagi pihak terkait dalam memahami persepsi publik terhadap desain seragam atlet, serta sebagai dasar dalam pengembangan sistem analisis sentimen otomatis. Selain itu, penelitian ini juga membuka peluang bagi penelitian selanjutnya untuk mengembangkan metode yang lebih canggih, seperti penggunaan *deep learning* dan perluasan *dataset*, guna meningkatkan akurasi dan kualitas analisis sentimen.

#### 4. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa tujuan penelitian untuk menganalisis sentimen masyarakat terhadap desain seragam atlet Indonesia pada Olimpiade Paris 2024 serta membandingkan kinerja algoritma *Naive Bayes Classifier* dan *Support Vector Machine* telah berhasil dicapai. Hasil analisis menunjukkan bahwa sentimen masyarakat cenderung didominasi oleh sentimen positif, yang tercermin dari jumlah data positif yang lebih tinggi dibandingkan data negatif. Hal ini menunjukkan bahwa secara umum desain seragam atlet Indonesia mendapatkan apresiasi yang baik dari publik, baik dari aspek estetika maupun representasi identitas nasional. Temuan ini juga diperkuat oleh visualisasi *word cloud* yang menunjukkan dominasi kata-kata bernuansa positif seperti “cool”, “good”, dan “elegant”.

Dari sisi performa model, hasil evaluasi menunjukkan bahwa *Support Vector Machine* memiliki kinerja yang lebih unggul dibandingkan *Naive Bayes Classifier* dalam melakukan klasifikasi sentimen. Hal ini dibuktikan dengan nilai akurasi SVM yang mencapai 94,2%, lebih tinggi dibandingkan *Naive Bayes* sebesar 80,5%. Selain itu, SVM juga menunjukkan keseimbangan yang lebih baik pada metrik *precision*, *recall*, dan *F1-Score*, sehingga lebih efektif dalam mengklasifikasikan data teks yang bersifat kompleks dan tidak terstruktur seperti pada media sosial X. Dengan demikian, SVM dapat direkomendasikan sebagai metode yang lebih optimal untuk analisis sentimen pada kasus serupa.

Meskipun demikian, penelitian ini masih memiliki beberapa keterbatasan, seperti jumlah dataset yang terbatas, penggunaan metode pelabelan berbasis leksikon, serta belum mempertimbangkan konteks bahasa yang lebih kompleks seperti sarkasme. Oleh karena itu, penelitian selanjutnya disarankan untuk menggunakan dataset yang lebih besar dan beragam, serta mengadopsi metode yang lebih canggih seperti *deep learning* (misalnya LSTM atau BERT) guna meningkatkan akurasi dan kemampuan pemahaman konteks. Selain itu, penelitian selanjutnya juga dapat mengembangkan analisis sentimen ke dalam kategori yang lebih spesifik (*multiclass*) serta mengintegrasikan aspek analisis visual dan *multimodal* untuk memperoleh hasil yang lebih komprehensif. Dengan pengembangan tersebut, diharapkan penelitian di bidang analisis sentimen dapat memberikan kontribusi yang lebih signifikan baik secara akademis maupun praktis.

#### Daftar Pustaka

Anadas, S., & Suryono, R. R. (2025). Analisis Sentimen Publik Terhadap Program Laporan Tahunan Wapres Periode 2024 di Media Sosial X Menggunakan Metode Naive Bayes dan Support Vector Machine (SVM). *Jurnal Pendidikan Dan Teknologi Indonesia*, 5(8), 2183–2192. <https://doi.org/10.52436/1.jpti.934>

- Anisah, S., & Irwansyah, I. (2025). Analisis Data Mining untuk Klasifikasi Kafe Populer di Jakarta Menggunakan Decision Tree dan Visualisasi dengan Tableau. *Jutisi : Jurnal Ilmiah Teknik Informatika Dan Sistem Informasi*, 14(2), 890. <https://doi.org/10.35889/jutisi.v14i2.2660>
- Ari Rama Novryadi, Irwansyah, & Moh Shidqon. (2026). Analisis Sentimen Masyarakat Terhadap Kinerja Presiden Indonesia Joko Widodo Periode Kedua Menggunakan Metode Naïve Bayes dan SVM. *DIGINTEL-AI : DIGital INnovation and inTElligence – AI*, 1(1), 11–24. <https://doi.org/10.66217/digintel-ai.v1i1.2>
- Chaudhuri, R., Deb, S., & Das, H. (2023). Noble Approach on Sensor Fused Bio Intelligent Path Optimisation and Single Stage Obstacle Recognition in Customized Mobile Agent. *Procedia Computer Science*, 218, 778–787. <https://doi.org/10.1016/j.procs.2023.01.058>
- Hu, Y. (2025). Research on the Application of Social Media Data Mining Based on Sentiment Analysis. *Applied and Computational Engineering*, 121(1), 147–153. <https://doi.org/10.54254/2755-2721/2025.19843>
- Ihza Kurniawan, D., Irwansyah, & Taufik, A. (2026). Analisis Sentimen Terhadap Komentar Video IShowSpeed Tour Indonesia Pada YouTube Menggunakan Metode SVM. *DIGINTEL-AI : DIGital INnovation and inTElligence – AI*, 1(1), 50–62. <https://doi.org/10.66217/digintel-ai.v1i1.5>
- Irwansyah, I., Dittyata, R., Rizal, R., & Wiyono, W. (2024). Optimalisasi Klasifikasi Uji Emisi Sepeda Motor Menggunakan Algoritma Naïve Bayes. *Infotech: Journal of Technology Information*, 10(2), 337–242. <https://doi.org/10.37365/jti.v10i2.327>
- Irwansyah, I., Wiranata, A. D., & M, T. T. (2023). Komparasi Algoritma Decision Tree, Naive Bayes Dan K-Nearest Neighbor Untuk Menentukan Kualitas Udara Di Provinsi Dki Jakarta. *Infotech: Journal of Technology Information*, 9(2), 193–198. <https://doi.org/10.37365/jti.v9i2.203>
- Lasuci Prastia, A., & Asra, T. (2026). Perbandingan Algoritma Random Forest, SVM, dan Naïve Bayes dalam Analisis Sentimen Ulasan Spotify di Play Store Berbasis SMOTE. *Jurnal Informatika Kaputama (JIK)*, 10(1), 17–27. <https://doi.org/10.59697/jik.v10i1.1157>
- Mahmudah, S. A., & Yudhistira, A. (2025). Analisis Sentimen Terhadap Cyberbullying pada Platform Media Sosial X Menggunakan Algoritma Naive Bayes. *Jurnal Pendidikan Dan Teknologi Indonesia*, 5(1), 189–200. <https://doi.org/10.52436/1.jpti.628>
- Purba, R., Sinaga, F. M., Jurnal Pipin, S., & Kelvin, K. (2025). Fine-Grained Sentiment Analysis On Big Data From Multi-Platform In Indonesia. *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, 11(1), 64–75. <https://doi.org/10.33480/jitk.v11i1.6549>
- Purnomo, D., Firgiawan, W., & Nur, N. (2025). Komparasi Algoritma Random Forest, Naïve Bayes, dan SVM pada Sentimen Kebijakan PPN 12%. *Jurnal Tekno Kompak*, 19(2), 155–167. <https://doi.org/10.33365/jtk.v19i2.122>
- Ramezani, E. B. (2025). Sentiment analysis applications using deep learning advancements in social networks: A systematic review. *Neurocomputing*, 634, 129862. <https://doi.org/10.1016/j.neucom.2025.129862>

- Salsa Billa Permana Putri, Irwansyah, & Tri M, T. (2026). Implementasi Algoritma K-NN Pada Sosial Media X Untuk Analisis Sentimen Pengalaman Warganet Tinggal Di Luar Negeri. *DIGINTEL-AI: DIGital INnovation and inTELLigence – AI*, 1(1), 37–49. <https://doi.org/10.66217/digintel-ai.v1i1.4>
- Taha, K. (2025). Big Data Analytics in IoT, social media, NLP, and information security: Trends, challenges, and applications. *Journal of Big Data*, 12(1), 150. <https://doi.org/10.1186/s40537-025-01192-9>
- Weng, Y., Isleem, H. F., Hindi, K. E., & Ezugwu, A. E. (2025). Natural language processing for extracting consumer sentiment dynamics through multimodal social media analysis to predict microeconomic consumption pattern shifts. *Journal of Big Data*, 12(1), 254. <https://doi.org/10.1186/s40537-025-01315-2>