

# Analisis Sentimen Pengguna Aplikasi Tantan: Perbandingan Kinerja Metode Naive Bayes dan SVM

Oleh:

Serlindha Tri Andini

Ade Eviyanti

Progam Studi Informatika

Universitas Muhammadiyah Sidoarjo

Januari, 2025



# Pendahuluan

Tantan merupakan salah satu aplikasi kencan daring yang populer di Indonesia, dengan fitur-fitur seperti geser profil, pesan instan, dan verifikasi akun untuk keamanan pengguna. Ulasan pengguna di Google Play Store mencerminkan pengalaman mereka, baik yang positif maupun negatif. Ulasan positif biasanya menyoroti fitur yang memuaskan, sedangkan ulasan negatif sering kali mengkritisi masalah teknis atau biaya premium yang tinggi. Analisis sentimen menjadi langkah penting untuk memahami opini pengguna, yang dapat digunakan pengembang aplikasi dalam meningkatkan kualitas layanan. Penelitian ini bertujuan membandingkan kinerja algoritma Naive Bayes dan Support Vector Machine (SVM) dalam klasifikasi sentimen, sekaligus mengatasi tantangan data yang tidak seimbang menggunakan teknik SMOTE.

# Pertanyaan Penelitian (Rumusan Masalah)

Tantan, sebagai aplikasi kencan populer, menerima ulasan positif dan negatif yang mencerminkan pengalaman pengguna. Ketidakseimbangan data ulasan dan perbedaan kemampuan algoritma Naive Bayes dan SVM dalam klasifikasi sentimen menjadi tantangan utama. Oleh karena itu, diperlukan evaluasi untuk menentukan metode terbaik dalam memahami opini pengguna dan meningkatkan kualitas aplikasi.

# Metode

## 1. Pengumpulan Data:

Data ulasan dari Google Play Store menggunakan web scraping.

## 2. Pra-pemrosesan Data:

1. Pembersihan simbol, angka, dan karakter khusus.
2. Tokenisasi, stopwords removal, dan ekstraksi fitur menggunakan TF-IDF.

## 3. Penyeimbangan Data:

Menggunakan SMOTE untuk meningkatkan proporsi ulasan negatif.

## 4. Evaluasi Model:

Menggunakan akurasi, precision, recall, dan F1-score untuk menilai kinerja.

# Hasil

- Sampel Dataset

Data ulasan pengguna aplikasi Tantan didapatkan dari Google Play Store dengan cara web scraping sebanyak 1195 pada gambar dibawah ini.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	userName	score	at	content												
2	Lutfi Jauhari	5	28/05/2024 08.49	Nice												
3	Roy Bersepeda	3	28/05/2024 08.30	Saya masih baru, jdi belum bnyak tau, ini saya masih mempelajari nya												
4	Mulnas ry	5	28/05/2024 07.10	tantan sangat luar biasa seneng bermain nya												
5	Faris Maulana	5	28/05/2024 06.59	Sangat Bagus												
6	Andri Sutansyah	5	28/05/2024 04.18	Ok												
7	Ghly Rmdhni	1	28/05/2024 03.20	Kok gni sh?												
8	Dito Bm	4	28/05/2024 03.11	ok												
9	Febry Negara	2	27/05/2024 22.48	Gajeas												
10	Dedi Juhana	5	27/05/2024 22.11	Sangat membantu dalam rangka silaturahmi												
11	MC Channel (Miss Chatrine)	5	27/05/2024 19.19	Di antara semua apk dating, tantan ini paling buat aku nyaman ðŸ™ verified nya juga cuma itungan menit, ah mantap ðŸ™												
12	Almouzt 07	5	27/05/2024 18.57	Jos												
13	Americans Densamericans	5	27/05/2024 18.28	Keren												
14	nu sa	5	27/05/2024 17.28	Bagus dan terhibur ðŸ«°ðŸ%ðŸ'ðŸ%ðŸ«°ðŸ%												
15	ff erwin	5	27/05/2024 16.24	Okee sekali												
16	Bung Vii	5	27/05/2024 15.38	Baik												
17	jogja Prima	4	27/05/2024 14.56	good												
18	alexander giryar	5	27/05/2024 14.30	Love tantan												
19	Yuziruddin 1991	5	27/05/2024 12.55	Mantap												
20	Aldo Guswan Hendris	1	27/05/2024 12.19	Udah berlangganan tapi tetep gabisa.. Padahal pas di cek langganannya udah aktif.. Tapi error terus.. Perbaiki lah, masa mau untungnya												
21	terong ungu	1	27/05/2024 11.39	Banyak biodata kosong dan tak terverifikasi , rugi di biaya top up .												
22	Firda Wati	5	27/05/2024 11.20	Sangat baik is the best												
23	Afidatun Nisak	1	27/05/2024 10.50	Makin lama makin aneh												

# Hasil

- Pra-prosesan Data

Pra-pemrosesan Data yang sudah dikumpulkan akan di proses dengan langkah prapemrosesan yaitu Pembersihan Data menghapus simbol, angka, dan karakter khusus. Tokenizing: memecahkan teks berupa kalimat berdasarkan setiap kata. Stopword: Menyingkirkan kata yang tidak memberikan kontribusi berarti pada analisis.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	userName	score	at	content	cleaned_content											
2	Lutfi Jauhari	5	28/05/2024 08.49	Nice	nice											
3	Roy Bersepeda	3	28/05/2024 08.30	Saya masih baru, jdi be	saya masih baru jdi belum bnyak tau ini saya masih mempelajari nya											
4	Mulnas ry	5	28/05/2024 07.10	tantan sangat luar bia	tantan sangat luar biasa seneng bermain nya											
5	Faris Maulana	5	28/05/2024 06.59	Sangat Bagus	sangat bagus											
6	Andri Sutansyah	5	28/05/2024 04.18	Ok	ok											
7	Ghly Rmdhni	1	28/05/2024 03.20	Kok gni sh?	kok gni sh											
8	Dito Bm	4	28/05/2024 03.11	ok	ok											
9	Febry Negara	2	27/05/2024 22.48	Gajeas	gajeas											
10	Dedi Juhana	5	27/05/2024 22.11	Sangat membantu dala	sangat membantu dalam rangka silaturahmi											
11	MC Channel (Miss Chatrine)	5	27/05/2024 19.19	Di antara semua apk d	di antara semua apk dating tantan ini paling buat aku nyaman verifed nya juga cuma itungan menit ah mantap											
12	Almouzt 07	5	27/05/2024 18.57	Jos	jos											
13	Americans Densamericans	5	27/05/2024 18.28	Keren	keren											
14	nu sa	5	27/05/2024 17.28	Bagus dan terhibur d	bagus dan terhibur											
15	ff erwin	5	27/05/2024 16.24	Okee sekali	okee sekali											
16	Bung Vii	5	27/05/2024 15.38	Baik	baik											
17	jogja Prima	4	27/05/2024 14.56	good	good											
18	alexander giryar	5	27/05/2024 14.30	Love tantan	love tantan											
19	Yuziruddin 1991	5	27/05/2024 12.55	Mantap	mantap											
20	Aldo Guswan Hendris	1	27/05/2024 12.19	Udah berlangganan tap	udah berlangganan tapi tetep gabisa padahal pas di cek langganannya udah aktif tapi error terus perbaiki lah masa mau untun											
21	terong ungu	1	27/05/2024 11.39	Banyak biodata kosong	banyak biodata kosong dan tak terverifikasi rugi di biaya											
22	Firda Wati	5	27/05/2024 11.20	Sangat baik is the best	sangat baik best											
23	Afidatun Nisak	1	27/05/2024 10.50	Makin lama makin ane	makin lama makin aneh											

# Hasil

- Penyeimbangan Data

Data yang digunakan kurang seimbang maka menggunakan metode oversampling pada kelas minoritas (ulasan negatif) dengan SMOTE. Teknik ini menghasilkan sampel sintetis untuk kelas minoritas guna meningkatkan distribusi data dan memastikan model dapat belajar lebih baik dari ulasan negatif

```
Distribusi label sebelum SMOTE:
```

```
sentiment
```

```
positive      704
```

```
negative      251
```

```
Name: count, dtype: int64
```

```
Distribusi label setelah SMOTE:
```

```
sentiment
```

```
negative      704
```

```
positive      704
```

```
Name: count, dtype: int64
```



## Hasil

- Wordcloud

visualisasi data set yang di ambil dari beberapa kata yang sering muncul pada ulasan aplikasi tantan.





# Hasil

- Naive Bayes

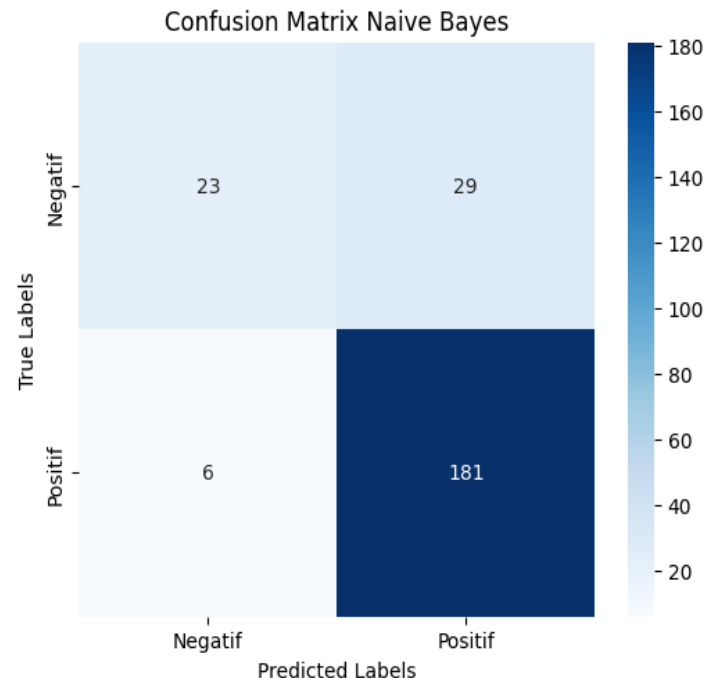
Tahap berikutnya yaitu menganalisis performa metode Naive Bayes dengan menerapkan metrik evaluasi yaitu *precision*, *recall*, *F1-score*, dan *accuracy*. yang menunjukkan hasil pengujian model berdasarkan klasifikasi pada data uji

Class	Precision	Recall	F1-Score	Support
Negative	0.47	0.87	0.61	52
Positive	0.95	0.73	0.83	187
Accuracy			0.76	239
macro avg	0.71	0.80	0.72	239
weighted avg	0.85	0.76	0.78	239

# Hasil

- Confusion Matrix Naive Bayes

Model Naive Bayes menunjukkan precision 95% dan recall 73% untuk kelas positif, menandakan akurasi tinggi dalam memprediksi ulasan positif meskipun ada beberapa yang terlewat. Untuk kelas negatif, recall mencapai 87%, namun precision rendah di 47%, menyebabkan banyak ulasan negatif salah diklasifikasikan. Akurasi total model adalah 76,15%, dengan kelemahan pada klasifikasi negatif.



# Hasil

- SVM

Klasifikasi Menggunakan SVM dilakukan untuk menganalisis performa model berdasarkan berbagai kernel yaitu Linear, RBF, Polynominal, dan Sigmoid yang ditunjukkan pada gambar di bawah .

Kernel	Precision (Negative)	Recall (Negative)	F1-Score (Negative)	Precision (Positive)	Recall (Positive)	F1-Score (Positive)	Accuracy (%)
Linear	0.63	0.73	0.68	0.92	0.88	0.90	84.94
RBF	0.46	0.60	0.52	0.88	0.81	0.84	76.15
Poly	0.42	0.87	0.57	0.95	0.67	0.79	71.55
Sigmoid	0.63	0.77	0.70	0.93	0.88	0.90	85.36

# Hasil

- Perbandingan Metode Naive Bayes dan SVM

Setelah data terkumpul dengan mengimplementasikan algoritma Naive Bayes dan Support Vector Machine (SVM), tahap berikutnya adalah melakukan langkah-langkah analisis lanjutan membandingkan di antara kernel svm hasil lebih baik adalah sigmoid maka naive bayes akan saya bandingkan dengan kernel sigmoid. Berikut hasilnya:

Metric	Naive Bayes	SVM (Kernel = Sigmoid)
Precision (Negative)	0.47	0.63
Recall (Negative)	0.87	0.77
F1-Score (Negative)	0.61	0.70
Precision (Positive)	0.95	0.93
Recall (Positive)	0.73	0.88
F1-Score (Positive)	0.83	0.90
Accuracy (%)	76.15	85.36

# Pembahasan

Model Naive Bayes menunjukkan akurasi sebesar 76,15% dengan precision rendah (0,47) namun recall tinggi (0,87) untuk kelas negatif. Hal ini menunjukkan bahwa metode mampu mendeteksi ulasan negatif dengan baik, tetapi sering salah dalam memprediksi ulasan ambigu sebagai positif. Keterbatasan ini sejalan dengan penelitian sebelumnya, yang menyatakan bahwa Naive Bayes memiliki kelemahan dalam menangani fitur yang saling berkorelasi. Selain itu, distribusi data yang tidak seimbang memperburuk performa model pada ulasan negatif. Sementara itu, model SVM dengan kernel sigmoid menunjukkan performa yang lebih baik dengan akurasi tertinggi mencapai 85,36%. Hasil ini konsisten dengan studi terdahulu, yang menunjukkan keunggulan SVM dalam menangani data berdimensi tinggi, terutama dengan pemilihan kernel yang tepat. SVM lebih seimbang dalam mendeteksi ulasan positif dan negatif dengan F1-score sebesar 0,90 untuk kelas positif dan 0,70 untuk kelas negatif, meskipun recall untuk kelas negatif sedikit lebih rendah dibandingkan Naive Bayes.

# Pembahasan

Penerapan SMOTE untuk menangani ketidakseimbangan data terbukti efektif dalam meningkatkan performa model, terutama dalam mendeteksi kelas minoritas (negatif). Hasil ini mendukung penelitian sebelumnya, yang menunjukkan bahwa teknik oversampling seperti SMOTE dapat mempengaruhi peningkatan kinerja model pada data tidak seimbang. Meskipun demikian, kesalahan klasifikasi masih ditemukan, terutama pada ulasan pendek dan ambigu seperti "OK" atau "Bagus" yang sulit dikenali oleh model. Hal ini sesuai dengan temuan penelitian terdahulu, yang menyebutkan bahwa ulasan singkat sering tidak memiliki fitur signifikan untuk analisis sentimen. Secara keseluruhan, penelitian ini memiliki hasil yang menunjukkan bahwa metode SVM dengan kernel sigmoid lebih unggul dibandingkan Naive Bayes dalam semua metrik evaluasi. Temuan ini konsisten dengan penelitian sebelumnya, yang menegaskan bahwa SVM lebih stabil serta efektif dalam menangani tantangan analisis sentimen pada dataset yang tidak seimbang.

# Pembahasan

- Analisis Kesalahan Klasifikasi

Kesalahan klasifikasi terutama terjadi pada ulasan pendek dengan kata-kata ambigu seperti "OK" atau "Bagus", yang sulit dikenali oleh model. Selain itu, ulasan negatif yang menggunakan kata-kata netral sering kali terklasifikasi sebagai positif. Faktor-faktor ini mempengaruhi performa model, menyebabkan peningkatan jumlah kesalahan prediksi. Metode Naive Bayes juga sering salah memprediksi ulasan negatif sebagai positif karena asumsi independensi yang tidak selalu terpenuhi dalam teks. Metode SVM, meskipun lebih baik dalam menangani ulasan negatif, masih memiliki kelemahan dalam menangani ulasan dengan panjang teks yang sangat pendek. Panjang ulasan dan intensitas kata-kata negatif terbukti memengaruhi performa model. SVM lebih responsif terhadap fitur-fitur ini dibandingkan Naive Bayes.



# Manfaat Penelitian

Penelitian ini menggunakan Model SVM dengan kernel sigmoid direkomendasikan karena performanya yang lebih stabil dibandingkan Naïve Bayes. Penelitian mendatang dapat mempertimbangkan eksplorasi algoritma lain, seperti Random Forest atau Transformer, yang dikenal unggul dalam klasifikasi teks pada dataset berukuran besar. Selain itu, penggunaan teknik ensemble model dapat meningkatkan kinerja dengan memadukan keunggulan dari berbagai algoritma. Selain itu, penelitian dapat diperluas dengan mempertimbangkan atribut tambahan, seperti panjang ulasan atau intensitas kata negatif. Penelitian ini memberikan kontribusi dengan menyediakan evaluasi komprehensif terhadap performa dua algoritma utama untuk analisis sentimen, yang berpotensi untuk menjadi acuan terhadap pengembang aplikasi dalam menyempurnakan kualitas layanan berdasarkan ulasan pengguna.

# Referensi

- [1] Y. Khai Shin BCS, "Chinese Culture, Contemporary Dating and Tantan: Exploring self-presentation in the Age of Mobile Dating Apps."
- [2] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," 2021.
- [3] "7099-25292-1-PB".
- [4] M. Rezki, D. N. Kholifah, M. Faisal, R. Suryadithia, U. Bina, and S. Informatika, "Analisis Review Pengguna Google Meet dan Zoom Cloud Meeting Menggunakan Algoritma Naïve Bayes." [Online]. Available: <http://ejournal.bsi.ac.id/ejurnal/index.php/infortech264>
- [5] M. Khoirul, U. Hayati, and O. Nurdiawan, "ANALISIS SENTIMEN APLIKASI BRIMO PADA ULASAN PENGGUNA DI GOOGLE PLAY MENGGUNAKAN ALGORITMA NAIVE BAYES," 2023.
- [6] H. Dwi Putra, L. Khairani, D. Hastari, P. Studi Sistem Informasi, F. Sains dan Teknologi, and C. Author, "SENTIMAS: Seminar Nasional Penelitian dan Pengabdian Masyarakat Comparison of Naive Bayes Classifier and Support Vector Machine Algorithms for Classifying Student Mental Health Data Perbandingan Algoritma Naive Bayes Classifier dan Support Vector Machine untuk Klasifikasi Data Kesehatan Mental Mahasiswa." [Online]. Available: <https://journal.irpi.or.id/index.php/sentimas>
- [7] S. Alpin Rizaldi, S. Alam, and I. Kurniawan, "ANALISIS SENTIMEN PENGGUNA APLIKASI JMO (JAMSOSTEK MOBILE) PADA GOOGLE PLAY STORE MENGGUNAKAN METODE NAIVE BAYES 1)," vol. 2, no. 3, pp. 109–117, 2023, doi: 10.55123.
- [8] A. Rifa, R. Ardhani, and D. Pratama, "ANALISIS SENTIMEN TERHADAP LAYANAN APLIKASI GRAB INDONESIA MENGGUNAKAN METODE NAÏVE BAYES," 2024.
- [9] I. Kurniawan *et al.*, "Perbandingan Algoritma Naive Bayes Dan SVM Dalam Sentimen Analisis Marketplace Pada Twitter," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 10, no. 1, 2023, [Online]. Available: <http://jurnal.mdp.ac.id>
- [10] J. H. Joloudari, A. Marefat, M. A. Nematollahi, S. S. Oyelere, and S. Hussain, "Effective Class-Imbalance Learning Based on SMOTE and Convolutional Neural Networks," *Applied Sciences (Switzerland)*, vol. 13, no. 6, Mar. 2023, doi: 10.3390/app13064006.

# Referensi

- [11] P. Manurung, "Parintan Manurung dan Guntoro, Analisis Sentimen Layanan Aplikasi Kesehatan Pada Ulasan Play Store: Systematic Literature Review Analisis Sentimen Layanan Aplikasi Kesehatan Pada Ulasan Play Store: Systematic Literature Review".
- [12] M. F. El Firdaus, N. Nurfaizah, and S. Sarmini, "Analisis Sentimen Tokopedia Pada Ulasan di Google Playstore Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor," *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 5, p. 1329, Oct. 2022, doi: 10.30865/jurikom.v9i5.4774.
- [13] A. P. Wibowo, W. Darmawan, A. Stmik, and W. Pratama, "KOMPARASI METODE NAÏVE BAYES DAN K-NEAREST NEIGHBOR TERHADAP ANALISIS SENTIMEN PENGGUNA APLIKASI PEDULILINDUNGI," 2022, [Online]. Available: <http://ejournal.stmik-wp.ac.id>
- [14] K. Nugroho and F. N. Hasan, "Analisis Sentimen Masyarakat Mengenai RUU Perampasan Aset Di Twitter Menggunakan Metode Naïve Bayes," *SMATIKA JURNAL*, vol. 13, no. 02, pp. 273–283, Dec. 2023, doi: 10.32664/smatika.v13i02.899.
- [15] B. Irawan and A. Bahtiar, "ANALISIS SENTIMEN PENGGUNA APLIKASI SHEJEK BERDASARKAN ULASAN DI GOOGLE PLAY STORE MENGGUNAKAN METODE NAÏVE BAYES," 2024.
- [16] A. Syaputra *et al.*, "Klasifikasi Penyakit Daun pada Tebu dengan Pendekatan Algoritma K-Nearest Neighbors, Multilayer Perceptron dan Support Vector Machine".
- [17] M. I. Fikri, T. S. Sabrila, Y. Azhar, and U. M. Malang, "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter".
- [18] A. P. Wibowo, W. Darmawan, A. Stmik, and W. Pratama, "KOMPARASI METODE NAÏVE BAYES DAN K-NEAREST NEIGHBOR TERHADAP ANALISIS SENTIMEN PENGGUNA APLIKASI PEDULILINDUNGI," 2022, [Online]. Available: <http://ejournal.stmik-wp.ac.id>
- [19] S. Rabbani, D. Safitri, N. Rahmadhani, A. A. F. Sani, and M. K. Anam, "Perbandingan Evaluasi Kernel SVM untuk Klasifikasi Sentimen dalam Analisis Kenaikan Harga BBM," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 2, pp. 153–160, Oct. 2023, doi: 10.57152/malcom.v3i2.897.
- [20] M. F. El Firdaus, N. Nurfaizah, and S. Sarmini, "Analisis Sentimen Tokopedia Pada Ulasan di Google Playstore Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor," *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 5, p. 1329, Oct. 2022, doi: 10.30865/jurikom.v9i5.4774.

