E-ISSN : 2988-1986
https://ejournal.warunayama.org/kohesi

Kohesi: Jurnal Multidisiplin Saintek
Volume 6 No 11 Tahun 2025

ANALISIS SENTIMEN ULASAN PENGGUNA BINANCE DI GOOGLE PLAY MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN SUPPORT VECTOR MACHINE DENGAN TEKNIK SMOTE

Riza Fahlapi[1], Danang Abu Hafid[2], Abdurrazaq[3], Moh. Azam Abulkhoir[4], Bebi Kurniawan[5], Yafianus Garamba[6]
[1-6]Fakultas Teknik dan Informatika, Program Studi Sistem Informatika, Universitas Bina Sarana Informatika
[1]riza.rzf@bsi.ac.id, [2]danangabuhafid@icloud.com, [3]abdurrazzaq268@gmail.com, [4]mazamabulkhoir@gmail.com, [5]awanhura174@gmail.com, [6]yafianusgarambayafgan@gmail.com

Abstrak

Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna aplikasi **Binance** di Google Play menggunakan dua algoritma pembelajaran mesin, yaitu **Naïve Bayes (NB)** dan ***Support Vector Machine*** **(SVM)**. Untuk menangani masalah ketidakseimbangan kelas dalam data, diterapkan teknik **SMOTE** (***Synthetic Minority Over-sampling Technique***). Data yang digunakan dalam penelitian ini diambil dengan metode ***web scraping*** pada ulasan aplikasi Binance, dengan data yang disaring berdasarkan ulasan terbaru dan rating. Hasil eksperimen menunjukkan bahwa dengan penerapan SMOTE, baik algoritma Naïve Bayes maupun SVM memberikan peningkatan yang signifikan pada akurasi serta metrik evaluasi lainnya seperti precision, recall, dan F1-score. Secara keseluruhan, penelitian ini menunjukkan bahwa teknik SMOTE efektif dalam menangani ketidakseimbangan kelas pada analisis sentimen ulasan aplikasi.
Kata kunci: Analisis Sentimen, Ulasan Pengguna, Naïve Bayes, Support Vector Machine, SMOTE.

Abstract

This study aims to analyze the sentiment of Binance app user reviews on Google Play using two machine learning algorithms, namely Naïve Bayes (NB) and Support Vector Machine (SVM). To overcome the problem of class imbalance in the data, the SMOTE (Synthetic Minority Over-sampling Technique) technique is applied. The data used in this study was taken using the web scraping method on Binance app reviews, with data filtered based on the latest reviews and ratings. The experimental results show that with the application of SMOTE, both the Naïve Bayes and SVM algorithms provide significant improvements in accuracy and other evaluation metrics such as precision, recall, and F1-score. Overall, this study shows that the SMOTE analysis technique is effective in handling class synchrony in app review sentiment.
Keywords*: Sentiment Analysis, User Reviews, Naïve Bayes, Support Vector Machine, SMOTE.*

Article History
Received: January 2025
Reviewed: January 2025
Published: January 2025

Plagirism Checker No 234
Prefix DOI :
10.8734/Kohesi.v1i2.365
Copyright : Author
Publish by : Kohesi

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

PENDAHULUAN

Aplikasi Binance adalah salah satu platform pertukaran cryptocurrency terbesar dan paling populer di dunia. Sejak diluncurkan pada tahun 2017, Binance telah berkembang pesat dengan menyediakan berbagai layanan, termasuk perdagangan aset digital, token, serta fitur-fitur lainnya seperti staking, lending, dan futures trading. Dengan antarmuka pengguna yang ramah dan kemudahan dalam transaksi, Binance telah menarik jutaan pengguna global, termasuk di Indonesia, yang tertarik untuk berinvestasi atau bertransaksi menggunakan cryptocurrency.

Sebagai platform yang berkembang pesat, Binance tidak hanya fokus pada kemudahan penggunaan, tetapi juga pada inovasi dan keamanan. Melalui fitur-fitur canggih dan berbagai penawaran menarik, seperti diskon untuk pengguna dengan token Binance (BNB), aplikasi ini memberikan pengalaman yang optimal bagi para trader maupun investor pemula dan profesional. Dengan adanya berbagai layanan di dalamnya, Binance telah menarik perhatian tidak hanya individu, tetapi juga institusi yang tertarik untuk memanfaatkan potensi pasar crypto.

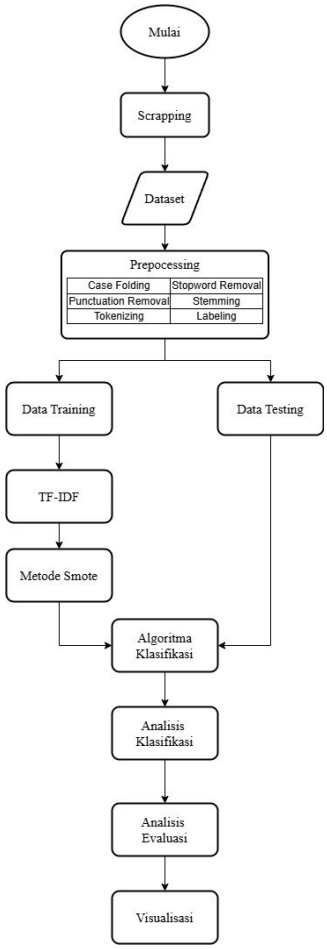
Namun, dengan meningkatnya jumlah pengguna, juga terdapat beragam ulasan mengenai pengalaman mereka saat menggunakan aplikasi Binance, baik yang positif maupun negatif. Ulasan-ulasan ini menjadi penting untuk dianalisis agar dapat memberikan gambaran yang lebih jelas mengenai kepuasan pengguna serta area-area yang masih perlu diperbaiki.

Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna terhadap aplikasi Binance yang terdapat di Google Play Store. Dengan menggunakan metode klasifikasi seperti Naïve Bayes dan Support Vector Machine (SVM), penelitian ini akan mengidentifikasi pola sentimen dari ulasan-ulasan pengguna, apakah cenderung positif atau negatif. Untuk menangani masalah ketidakseimbangan data pada ulasan pengguna, metode SMOTE (Synthetic Minority Over-sampling Technique) akan diterapkan, yang diharapkan dapat meningkatkan akurasi model dalam mengklasifikasikan sentimen.

Diharapkan penelitian ini dapat memberikan wawasan yang lebih mendalam mengenai persepsi pengguna terhadap aplikasi Binance serta dampaknya terhadap kinerja dan kepercayaan pengguna dalam bertransaksi di platform tersebut. Hasil analisis sentimen ini dapat menjadi acuan bagi pengembang aplikasi dalam meningkatkan kualitas layanan dan mengoptimalkan pengalaman pengguna.

METODOLOGI PENELITIAN

Penelitian ini dilakukan dengan pendekatan yang sistematis, dimulai dengan tahap pengumpulan data yang relevan. Setelah itu, data akan diproses melalui tahap *preprocessing*, yang mencakup pembersihan dan transformasi data untuk memastikan kualitasnya. Selanjutnya, data dibagi menjadi data pelatihan dan data pengujian. Untuk mengatasi masalah ketidakseimbangan data, metode SMOTE digunakan untuk menyeimbangkan distribusi kelas. Model klasifikasi kemudian dilatih menggunakan algoritma yang dipilih, dan hasilnya dianalisis untuk mengevaluasi performa model dalam mengklasifikasikan data.



Gambar 1. Tahapan Penelitian

1.1. Dataset

Mengumpulkan data teks dari sumber yang relevan untuk analisis klasifikasi menggunakan teknik *scraping* untuk mengambil data dari situs web atau platform yang sesuai. Data yang diperoleh berupa teks yang akan digunakan dalam analisis.

1.2. Preprocessing Data

Tahap *preprocessing* dilakukan untuk membersihkan dan mempersiapkan data agar siap digunakan dalam proses pemodelan. Langkah-langkah yang dilakukan meliputi:

a. *Case Folding*: Mengubah semua teks menjadi huruf kecil.
b. *Punctuation Removal*: Menghapus semua tanda baca dari teks untuk memfokuskan pada kata-kata relevan.
c. *Tokenizing*: Memecah teks menjadi kata-kata.
d. *Stopword Removal*: Menghapus kata umum yang tidak memiliki arti penting (misalnya: atau", "yang", "dan").
e. *Stemming*: Mengubah kata ke bentuk dasarnya.
f. *Labeling*: Memberikan label pada data sesuai kategori yang diinginkan. Misalnya, positif atau negatif.

1.3. Pembagian Data

Pembagian data bertujuan untuk memisahkan dataset menjadi dua kelompok utama: satu untuk proses pelatihan (training) dan yang lainnya untuk pengujian (testing). Umumnya, data dibagi dengan proporsi tertentu, seperti 80% untuk pelatihan dan 20% untuk pengujian.

1.4. Penggunaan TF-IDF

TF-IDF digunakan untuk mengonversi data teks menjadi representasi numerik yang dapat diproses oleh algoritma klasifikasi. Proses ini menghasilkan sebuah matriks fitur, di mana setiap elemen menunjukkan tingkat kepentingan suatu kata dalam dokumen dibandingkan dengan keseluruhan dataset, sehingga memberikan representasi numerik yang lebih bermakna untuk keperluan klasifikasi.

1.5. Metode SMOTE

SMOTE digunakan untuk mengatasi ketidakseimbangan dalam dataset dengan menambahkan sampel sintetis pada kelas minoritas. Proses ini dimulai dengan mengidentifikasi adanya perbedaan signifikan antara jumlah sampel kelas mayoritas dan minoritas. Jika ketidakseimbangan terdeteksi, SMOTE akan membuat sampel baru pada kelas minoritas dengan cara menginterpolasi data yang sudah ada, sehingga jumlah sampel di kelas tersebut bertambah. Setelah itu, dataset diverifikasi untuk memastikan keseimbangannya sebelum digunakan dalam pelatihan model klasifikasi.

1.6. Algoritma Klasifikasi

Langkah ini bertujuan untuk membangun model klasifikasi menggunakan algoritma tertentu guna mengelompokkan data ke dalam kategori yang diinginkan. Algoritma seperti Naive Bayes, yang berbasis teorema Bayes dan efektif untuk klasifikasi teks, atau SVM (Support Vector Machine), yang mencari hyperplane optimal untuk memisahkan kelas-kelas dalam dataset, digunakan dalam proses ini. Model dilatih menggunakan data training yang telah diproses sebelumnya, termasuk representasi fitur dengan TF-IDF dan penyeimbangan kelas melalui SMOTE. Tahap ini memastikan model mampu mengenali pola dalam data dengan baik, sehingga dapat memberikan prediksi yang akurat pada data uji.

1.7. Analisis dan Evaluasi

Evaluasi bertujuan untuk mengukur kinerja model klasifikasi menggunakan metrik yang tepat. Langkah pertama melibatkan perhitungan akurasi, yaitu persentase prediksi yang benar, untuk memberikan gambaran umum tentang performa model. Selain itu, AUC (*Area Under the Curve*) dihitung untuk mengevaluasi kemampuan model dalam membedakan antara kelas positif dan negatif, yang menawarkan perspektif lebih mendalam dibandingkan akurasi saja. Selanjutnya, hasil kinerja model dibandingkan antara penggunaan SMOTE dan tanpa SMOTE untuk menilai dampak penyeimbangan data terhadap efektivitas klasifikasi.

1.8. Visualisasi

Visualisasi bertujuan untuk menampilkan kata-kata yang paling sering muncul dalam ulasan pengguna berdasarkan kategori sentimen. *Wordcloud* yang dihasilkan merepresentasikan setiap kategori sentimen, dengan ukuran kata mencerminkan frekuensi kemunculannya. Semakin sering sebuah kata muncul dalam ulasan, semakin besar tampilannya dalam *wordcloud*.

HASIL DAN PEMBAHASAN

1.9. Dataset

Penelitian ini menggunakan data ulasan aplikasi Binance dari Google Play, yang dikumpulkan melalui metode scraping menggunakan Python. Sebanyak 18000 ulasan terbaru berhasil diperoleh pada tanggal 20 Desember 2024, dan data tersebut disimpan dalam format CSV untuk analisis lebih lanjut.

1.10. Preprocessing Data

a. *Case Folding*

Sebelum	Sesudah
Salah satu aplikasi terbaik buat trading. Karena fiturnya yang lengkap. Good job binance.	salah satu aplikasi terbaik buat trading. Karena fiturnya yang lengkap. good job binance.
Tolong bug nya di perbaiki, saya berulang kali daftar akun tampilan layar warna putih terus.	tolong bug nya di perbaiki, saya berulang kali daftar akun tampilan layar warna putih terus.

Tabel 1. Contoh Hasil *Case Folding*

b. *Punctuation Removal*

Sebelum	Sesudah
salah satu aplikasi terbaik buat trading. Karena fiturnya yang lengkap. good job binance.	salah satu aplikasi terbaik buat trading Karena fiturnya yang lengkap good job binance
tolong bug nya di perbaiki, saya berulang kali daftar akun tampilan layar warna putih terus.	tolong bug nya di perbaiki saya berulang kali daftar akun tampilan layar warna putih terus

Tabel 2. Contoh *Punctuation Removal*

c. *Tokenizing*

Sebelum	Sesudah
sangat mudah biaya transfer murah	sangat
	mudah
	biaya
	transfer
	murah

Tabel 3. Contoh *Tokenizing*

d. *Stopword Removal*

Sebelum	Sesudah
tolong bug nya di perbaiki saya berulang kali daftar akun tampilan layar warna putih terus	tolong bug nya perbaiki berulang kali daftar akun tampilan layar warna putih

Tabel 4. Contoh *Stopword Removal*

e. *Stemming*

Sebelum	Sesudah
tolong bug nya perbaiki	tolong bug nya perbaiki
berulang kali daftar akun	ulang kali daftar akun
tampilan layar warna putih	tampil layar warna putih

Tabel 5. Contoh *Stemming*

f. *Labeling*

Setelah *preprocessing data,* data terkumpul menjadi 7.946 data, dan selanjutnya akan di berikan label. Proses pelabelan data dalam penelitian ini dilakukan secara otomatis berdasarkan nilai pada kolom skor. Data dengan skor tertentu diklasifikasikan menjadi dua kategori: positif dan negatif, sementara kategori netral tidak digunakan. Berikut distribusi label untuk dataset yang sudah melewati *preprocessing data:*

Label	Jumlah Ulasan
Positif	6106
Negatif	1840
Total	7946

Tabel 6. Distribusi Label Dataset

1.11. Pembagian Data

Setelah pelabelan, selanjutnya data akan dibagi 2, yaitu data *training* (80%) dan data *testing* (20%). Berikut ditribusi label untuk data *training* dan data *testing*:

Label	Jumlah Ulasan
Positif	4884
Negatif	1472
Total	6356

Tabel 7. Distribusi Label Data *Training*

Berdasarkan tabel 7, jumlah data *training* adalah 1923 yang merupakan 80% dari data ulasan dan terdiri dari 1106 ulasan positif dan 817 ulasan negatif.

Label	Jumlah Ulasan
Positif	1222
Negatif	368
Total	1590

Tabel 8. Distribusi Label Data *Testing*

Sedangkan jumlah data *testing* adalah 1590 ulasan, 1222 ulasan positif dan 368 ulasan negatif.

1.12. Algoritma Klasifikasi

Berdasarkan hasil analisis yang sudah dilakukan, berikut adalah evaluasi dan interpretasi hasil yang diperoleh untuk model Naïve Bayes dan SVM dengan TF-IDF serta SMOTE.

Hasil Model:

Algoritma	*Accuracy*	AUC
Naïve Bayes + TF-IDF	90.12%	94.61%
SVM + TF-IDF	91.38%	94.78%
Naïve Bayes + SMOTE + TF-IDF	88.86%	95.03%
SVM + SMOTE + TF-IDF	88.53%	93.22%

Tabel 9. Hasil Model

Evaluasi hasil:

a. Akurasi (*Accuracy*):
Berdasarkan hasil model, algoritma SVM + TF-IDF memberikan akurasi tertinggi sebesar 91.38% dan AUC 94.78% tanpa menggunakan SMOTE. Sementara itu, algoritma Naïve Bayes + TF-IDF memiliki akurasi sebesar 90.12% dengan AUC 94.61%. Dengan penerapan SMOTE, akurasi Naïve Bayes sedikit menurun dari 90.12% menjadi 88.86%, meskipun AUC meningkat dari 94.61% menjadi 95.03%. Pada sisi lain, akurasi SVM juga mengalami penurunan dari 91.38% menjadi 88.53%, disertai penurunan AUC dari 94.78% menjadi 93.22%. Penurunan akurasi ini menunjukkan bahwa penggunaan SMOTE tidak memberikan keuntungan signifikan pada performa kedua algoritma untuk dataset ini.

b. AUC (Area *Under the Curve*):
Naïve Bayes + SMOTE + TF-IDF memiliki AUC tertinggi, yaitu 95.03%, sedikit lebih tinggi dibandingkan Naïve Bayes + TF-IDF (94.61%). SVM + TF-IDF memiliki AUC sebesar 94.78%, yang sedikit lebih tinggi dibandingkan dengan SVM + SMOTE + TF-IDF (93.22%). Perbedaan ini menunjukkan bahwa SMOTE memberikan dampak yang lebih signifikan pada Naïve Bayes dalam meningkatkan AUC, sementara pada SVM dampaknya relatif kecil karena SVM sudah memiliki kemampuan yang baik dalam menangani ketidakseimbangan data.

Berdasarkan hasil yang diperoleh, SVM menunjukkan keunggulan dalam hal akurasi, dengan nilai tertinggi sebesar 91.38%, meskipun perbedaan antara SVM dan Naïve Bayes tidak terlalu signifikan. AUC menunjukkan bahwa Naïve Bayes dengan SMOTE sedikit lebih baik dalam membedakan kelas, dengan AUC tertinggi sebesar 95.03%, yang menunjukkan bahwa model ini lebih sensitif terhadap ketidakseimbangan kelas dibandingkan SVM. Penggunaan SMOTE pada Naïve Bayes memberikan peningkatan dalam AUC, meskipun akurasi sedikit menurun. Sementara itu, pada SVM, dampaknya tidak terlalu besar dengan penurunan akurasi dan AUC. Hal ini bisa disebabkan oleh kemampuan SVM yang sudah baik dalam menangani data tidak seimbang tanpa perlu penyesuaian lebih lanjut.

Secara keseluruhan, baik Naïve Bayes maupun SVM memberikan performa yang sangat baik dalam pengklasifikasian data, terutama dengan pendekatan SMOTE + TF-IDF. Semua model menunjukkan nilai AUC di atas 90%, menandakan kemampuan yang sangat baik dalam membedakan antara kelas positif dan negatif. Penggunaan SMOTE membantu menangani ketidakseimbangan kelas, yang terbukti dari nilai AUC yang tetap tinggi bahkan setelah diterapkan. Namun, model terbaik dalam hal akurasi dan AUC adalah

SVM + TF-IDF. Model ini mencapai akurasi tertinggi sebesar 91.38% dengan AUC 94.78%, menjadikannya pilihan yang optimal untuk digunakan pada dataset ini. Sementara itu, kombinasi Naïve Bayes + SMOTE + TF-IDF menghasilkan AUC tertinggi sebesar 95.03%, meskipun akurasinya sedikit lebih rendah. Dari hasil ini, dapat disimpulkan bahwa pendekatan TF-IDF sebagai representasi fitur, dikombinasikan dengan algoritma SVM, memberikan performa klasifikasi terbaik tanpa perlu penyesuaian tambahan seperti SMOTE.

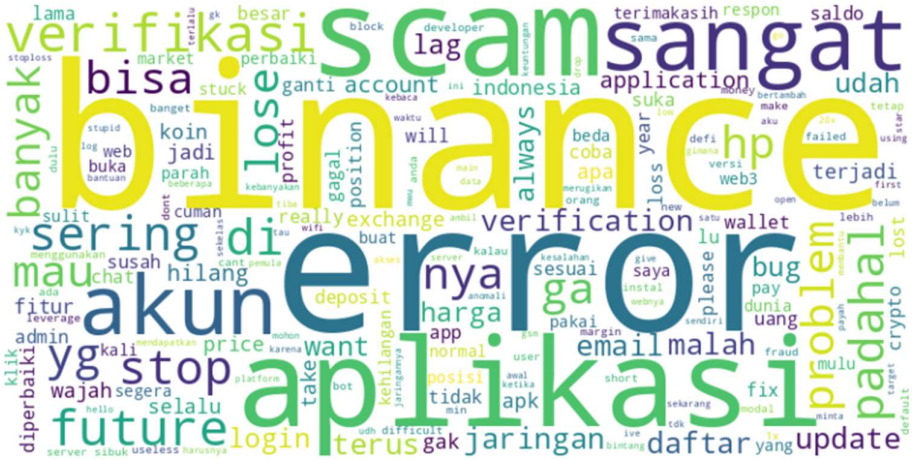
1.13. Visualisasi

Berikut gambar yang menampilkan hasil visualisasi sentimen positif:

Gambar 2. Visualisasi Sentimen Positif

Berdasarkan hasil visualisasi WordCloud untuk sentimen positif, kata-kata seperti "best", "good", "best exchange", "great" mendominasi ulasan positif. Hal ini menunjukkan bahwa pengguna cenderung mengapresiasi pengalaman yang berkualitas tinggi, menyenangkan, dan memuaskan saat menggunakan aplikasi. Selain itu, kata-kata terkait seperti "profit", "reward", "bonus", "amazing", dan "love" juga muncul dengan signifikan, yang mengindikasikan bahwa insentif dan penghargaan menarik menjadi faktor utama yang menarik perhatian pengguna.

Berikut gambar yang menyajikan hasil visualisasi untuk sentiment negatif:



Gambar 3. Visualisasi Sentimen Negatif

Berdasarkan hasil visualisasi wordcloud untuk sentimen negatif, kata-kata seperti "binance", "scam", "error", "aplikasi", dan "verifikasi", "aplikasi" mendominasi. Hal ini mengindikasikan bahwa ulasan negatif pengguna umumnya berkaitan dengan masalah akun, proses verifikasi, server jaringan, dan penggunaan aplikasi. Kata-kata seperti "versi", "parah", "exchange", "bug", dan "lag" juga muncul cukup besar, yang terkait penggunaan aplikasi serta fitur.

KESIMPULAN

Berdasarkan hasil analisis, **SVM + TF-IDF tanpa SMOTE** mencatatkan akurasi tertinggi sebesar **91,38%**, menjadikannya model yang paling unggul dalam hal prediksi akurat dibandingkan dengan **Naïve Bayes + TF-IDF**, yang memiliki akurasi sebesar **90,12%**. Meskipun demikian, **Naïve Bayes + TF-IDF** memiliki nilai AUC sedikit lebih rendah (94,61%) dibandingkan dengan **SVM + TF-IDF** (94,78%). Hal ini menunjukkan bahwa kedua algoritma ini sangat baik dalam membedakan kelas positif dan negatif, meskipun SVM memberikan prediksi yang lebih akurat.

Ketika menggunakan SMOTE, kinerja **Naïve Bayes + SMOTE + TF-IDF** menunjukkan sedikit penurunan akurasi menjadi **88,86%**, namun nilai AUC meningkat menjadi **95,03%**, menunjukkan kemampuan yang lebih baik dalam membedakan kelas positif dan negatif. Di sisi lain, **SVM + SMOTE + TF-IDF** mencatatkan akurasi sebesar **88,53%** dan AUC sebesar **93,22%**.

Meskipun SVM + SMOTE + TF-IDF memiliki AUC yang sedikit lebih rendah dibandingkan Naïve Bayes + SMOTE + TF-IDF, pengaruh SMOTE terlihat lebih signifikan dalam meningkatkan AUC pada model Naïve Bayes. Namun, secara keseluruhan, SVM + TF-IDF tanpa SMOTE tetap menjadi model terbaik karena memberikan akurasi tertinggi dan AUC yang sangat baik.

REFERENSI

Hadi, A., Bernard, B., Hulu, L.N. and Hulu, W.N. 2025. SENTIMEN ULASAN PENGGUNA E-COMMERCE BLIBLI MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN SUPPORT VECTOR MACHINE DENGAN METODE SMOTE. *Kohesi: Jurnal Sains dan Teknologi*. 6, 5 (Jan. 2025), 41–50. DOI: https://doi.org/10.3785/kohesi.v6i5.9791.

Esti Hermanto, Kuntoro, A. Y., Asra, T., Pratama, E. B., Effendi, L., & Ocanitra, R. (2020). Gojek and Grab User Sentiment Analysis on Google Play Using Naive Bayes Algorithm And Support Vector Machine Based Smote Technique. *Journal of Physics: Conference Series*, *1641*(1), 012102. https://doi.org/10.1088/1742-6596/1641/1/012102

Syaripah, I., Martanto, M., & Suprapti, T. (2024). ANALISIS SENTIMEN PENGGUNA TERHADAP APLIKASI BINANCE PADA ULASAN GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA NAÏVE BAYES. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(6), 3714–3724. https://doi.org/10.36040/jati.v7i6.8289

Sakti Alamsyah, A., Nur Budiman, S., & Dwi Romadhona, R. (2024). Penerapan Algoritma Support Vector Machine Untuk Analisis Sentimen Data Ulasan Aplikasi Binance Pada Google Play Store. *Innovative : Journal Of Social Science Research*, *4*(5). https://doi.org/10.31004/innovative.v4i5.15721

Fahlapi, R., Hermanto, H., Asra, T., Kuntoro, A. Y., Ocanitra, R., Effendi, L., & Syukmana, F. (2022). ANALISA SENTIMEN VAKSINASI COVID-19 DENGAN METODE SUPPORT VECTOR MACHINE DAN NAÏVE BAYES BERBASIS TEKNIK SMOTE. *Jurnal Informatika Kaputama (JIK)*, *6*(1), 57–63. https://doi.org/10.59697/jik.v6i1.136