

**Ranah Research**
Journal of Multidisciplinary Research and Development
082170743613 | ranahresearch@gmail.com | <https://jurnal.ranahresearch.com>

E-ISSN: 2655-0865



DOI: <https://doi.org/10.38035/rj.v7i6>
<https://creativecommons.org/licenses/by/4.0/>

Analisis Sentimen Pengguna Aplikasi *STEAM* dengan Algoritma *Naive Bayes*

Andy Kho¹, Felliks F. Tampinongkol²

¹ Universitas Bunda Mulia, Indonesia, s32210081@student.ubm.ac.id

² Universitas Bunda Mulia, Indonesia, l1884@lecturer.ubm.ac.id

Corresponding Author: s32210081@student.ubm.ac.id¹

Abstract: *Steam is a widely used digital game distribution platform with millions of users worldwide. The abundance of user reviews on the Google Play Store serves as a valuable source for analyzing user perceptions and satisfaction regarding the application. Therefore, this study performs sentiment analysis to understand user opinions about the Steam app. This research employs the Naive Bayes algorithm to classify user reviews into two sentiment categories: positive and negative. The process begins with collecting user reviews from the Google Play Store using web scraping techniques. The data then undergo preprocessing steps such as case folding, cleaning, tokenization, stopwords removal, and stemming to improve its quality. TF-IDF is used for feature extraction from the review texts, which are then used as input for the Naive Bayes Classifier model. The model is trained with training data and evaluated with testing data that has been previously split. Model performance is evaluated using a Confusion Matrix and metrics such as accuracy, precision, recall, and F1-score. The results show that the Naive Bayes model achieves an average accuracy of 84% in classifying sentiment. These findings indicate that the method is effective in understanding user opinions about the Steam application. This research is expected to provide insights for developers to improve application quality based on user feedback.*

Keyword: *Sentiment Analysis, Naive Bayes, Steam, Google Play Store, TF-IDF, Natural Language Processing, Text Classification.*

Abstrak: Steam merupakan platform distribusi digital game yang digunakan secara luas oleh jutaan pengguna di seluruh dunia. Banyaknya ulasan yang diberikan pengguna melalui Google Play Store menjadi sumber data yang sangat potensial untuk menganalisis persepsi dan kepuasan pengguna terhadap aplikasi ini. Oleh karena itu, dilakukan analisis sentimen untuk memahami opini pengguna terhadap aplikasi Steam. Penelitian ini menggunakan algoritma Naive Bayes untuk mengklasifikasikan ulasan pengguna ke dalam dua kategori sentimen, yaitu positif dan negatif. Proses dimulai dengan mengumpulkan data ulasan dari Google Play Store melalui teknik web scraping. Selanjutnya, dilakukan preprocessing data seperti case folding, cleaning, tokenisasi, stopwords removal, dan stemming untuk meningkatkan kualitas data sebelum dianalisis. Metode TF-IDF digunakan untuk ekstraksi fitur dari teks ulasan, yang kemudian digunakan sebagai input bagi model Naive Bayes Classifier. Model dilatih menggunakan data latih dan diuji menggunakan data uji yang telah

dipisahkan sebelumnya. Evaluasi model dilakukan menggunakan Confusion Matrix dan metrik seperti akurasi, presisi, recall, dan F1-score. Hasil penelitian menunjukkan bahwa model Naive Bayes dapat mengklasifikasikan sentimen ulasan dengan akurasi rata-rata sebesar 84%. Temuan ini menunjukkan bahwa metode ini cukup efektif dalam memahami opini pengguna terhadap aplikasi Steam. Penelitian ini diharapkan dapat memberikan wawasan bagi pengembang untuk meningkatkan kualitas aplikasi berdasarkan umpan balik pengguna.

Kata Kunci: Analisis Sentimen, Naive Bayes, Steam, Google Play Store, TF-IDF, Pemrosesan Bahasa Alami, Klasifikasi Teks.

PENDAHULUAN

Steam adalah platform distribusi digital yang diciptakan oleh perusahaan Valve Corporation, pertama kali diluncurkan pada tahun 2003. Sejak saat itu, Steam telah berkembang menjadi salah satu platform game terbesar di dunia dan memiliki lebih dari 150 juta pengguna. Steam menyediakan layanan pembelian, pengunduhan, dan pemutaran game. Selain game, Steam juga menawarkan software, konten digital lainnya, dan berbagai fitur komunitas yang memungkinkan para penggunanya untuk berinteraksi melalui forum, komentar, dan fitur sosial lainnya. Dengan adanya fitur-fitur ini, Steam tidak hanya menjadi tempat untuk membeli dan memainkan game, tetapi juga sebagai platform sosial yang menghubungkan para gamer di seluruh dunia. Dalam penelitian ini, analisis sentimen pengguna pada aplikasi Steam akan dilakukan menggunakan algoritma Naive Bayes. Algoritma ini akan digunakan untuk mengklasifikasikan ulasan pengguna ke dalam kategori sentimen positif, atau negatif berdasarkan teks yang diberikan oleh pengguna. Dengan memanfaatkan teknik ini, diharapkan dapat memperoleh wawasan yang lebih dalam mengenai persepsi dan pengalaman pengguna terhadap game dan layanan yang tersedia di Steam.

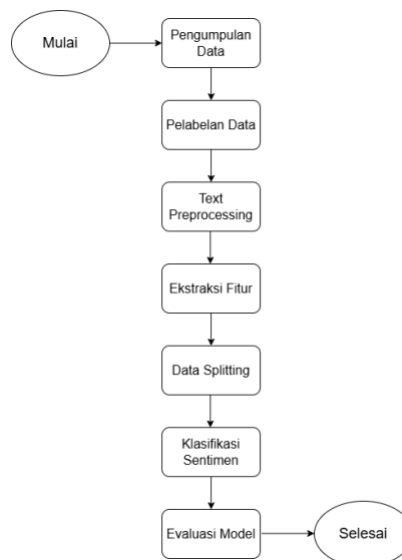
METODE

Pemilihan Algoritma

Dalam pengembangan perangkat lunak untuk analisis sentimen pengguna aplikasi Steam menggunakan algoritma Naive Bayes, beberapa langkah rekayasa kebutuhan dan perencanaan penelitian diperlukan agar sistem yang dihasilkan sesuai dengan kebutuhan pengguna dan tujuan penelitian. Algoritma Naive Bayes digunakan dalam analisis sentimen ini karena kemampuannya yang efisien dalam menangani data teks dalam jumlah besar dan melakukan klasifikasi berdasarkan probabilitas. Meskipun ulasan yang dianalisis sangat bervariasi, Naive Bayes tetap dapat memberikan hasil yang cepat dan akurat. Salah satu keunggulan utama Naive Bayes adalah kemampuannya untuk bekerja dengan baik bahkan dengan data pelatihan yang terbatas.

Perancangan Proses

Berikut adalah flowchart yang menggambarkan tahapan-tahapan yang dilakukan dalam penelitian ini yang meliputi pengambilan data, pengolahan data, pelatihan model, sampai evaluasi model.



Gambar 1. Flowchart Perancangan Proses

Pengumpulan Data

Tahap awal untuk mengumpulkan data berupa ulasan pengguna pada aplikasi Steam di Google Play Store. Data ulasan ini dikumpulkan dengan menggunakan metode web scraping dengan bahasa pemrograman Python. Dikumpulkan sebanyak 5.000 ulasan pengguna Steam di Google Play Store dari periode 4 April 2024 hingga 31 Desember 2024. Ulasan dari setiap pengguna mempunyai 11 informasi seperti reviewID, userName, userImage, content, score, thumbsUpCount, reviewCreatedVersion, at, replyContent, repliedAt, dan appVersion. Namun yang akan kita pakai dalam penelitian ini hanya userName, content, score, dan at.

Pelabelan Data (Labelling)

Tahap Labelling adalah tahap dimana data yang sudah dikumpulkan, diberi label agar sifat dari suatu data dapat dikenali dan dipelajari oleh model. Pada penelitian ini, setiap teks ulasan dari pengguna diberi label klasifikasi sentimen positif atau negatif.

Text Preprocessing

Tahap Text Preprocessing adalah tahap dimana data berupa teks yang sudah dikumpulkan, ditingkatkan kualitasnya sebelum dilakukan analisis sentimen agar dapat memudahkan proses dan menghasilkan performa analisis sentimen yang lebih baik. Berikut adalah beberapa proses dari tahap Text Preprocessing mulai dari Casefolding, Text Cleaning, Tokenization, Stopwords Removal, dan Stemming.

Ekstraksi Fitur

TF-IDF digunakan untuk mengukur pentingnya sebuah kata dalam sebuah dokumen, dengan mempertimbangkan dua hal utama yaitu seberapa sering suatu kata muncul dalam dokumen, dan seberapa jarang kata tersebut muncul dalam keseluruhan dokumen.

Term Frequency (TF)

Term Frequency (TF) merupakan ukuran yang digunakan untuk mengukur seberapa pentingnya suatu kata dalam dokumen, dengan cara menghitung seberapa sering suatu kata muncul dalam sebuah dokumen. Semakin sering kata tersebut muncul, semakin tinggi nilai TF-nya, yang berarti kata tersebut lebih relevan atau penting dalam konteks suatu dokumen.

Contoh :

Pada dokumen 1, term “app” muncul sebanyak 3 kali.

$$tf = \frac{3}{11} = 0,272$$

Inverse Document Frequency (IDF)

Inverse Document Frequency (IDF) merupakan ukuran yang digunakan untuk mengukur seberapa penting suatu kata pada seluruh kata-kata dalam seluruh dokumen. IDF memberikan bobot yang lebih tinggi pada kata yang jarang muncul di dokumen lainnya. Sebaliknya, kata yang sering muncul di banyak dokumen akan memiliki IDF yang rendah, karena kata yang sering muncul tidak banyak membantu dalam membedakan dokumen.

Contoh :

Kata “app” terdapat pada 2 dari 3 dokumen.

$$idf(x) = \log\left(\frac{3}{2}\right) = 0,1761$$

Kata “buy” terdapat pada 1 dari 3 dokumen

$$idf(x) = \log\left(\frac{3}{1}\right) = 0,4771$$

Nilai TF-IDF

Term Frequency – Inverse Document Frequency (TF-IDF) adalah gabungan dari dua metrik penting dalam analisis teks, yaitu Term Frequency (TF) dan Inverse Document Frequency (IDF). Nilai dari TF-IDF ini digunakan untuk mengukur seberapa penting suatu kata dalam sebuah dokumen, dengan memperhitungkan frekuensi suatu kata dalam dokumen tertentu serta seberapa langka kata tersebut pada keseluruhan dokumen.

Contoh :

Nilai TF dari kata “app” = 0,272

Nilai IDF dari kata “app” = 0,1761

$$tf - idf(x, d) = tf_{x,d} \times idf_{x,d}$$

$$tf - idf(x, d) = 0,272 \times 0,1761$$

$$tf - idf(x, d) = 0,04789$$

Klasifikasi Sentimen

Setelah proses pembobotan kata dengan TF-IDF selesai, dilanjutkan dengan klasifikasi sentimen menggunakan Naïve Bayes Classifier yang merupakan penerapan dari algoritma Naïve Bayes, menggunakan teks yang sudah dibuat menjadi representasi numerik oleh TF-IDF, Naïve Bayes menggunakan fitur tersebut untuk menghitung probabilitas dan mengklasifikasikan teks ke sentimen yang sesuai. Naïve bayes bekerja berdasarkan Teorema Bayes yang berperan dalam memperbaiki hitungan probabilitas dengan memanfaatkan data-data yang ada sebagai informasi tambahan.

Data Splitting

Data Splitting adalah proses memisahkan data menjadi bagian-bagian, dalam analisis sentiment, data dibagi menjadi 2 kelompok data yaitu data pelatihan dan data uji. Pada penelitian ini, sentimen ulasan dibagi berupa 20% dari data akan digunakan untuk data uji sedangkan sisanya yaitu 80% untuk data pelatihan.

Evaluasi Model

Hasil dari analisis sentimen yang dihasilkan menggunakan model algoritma Naïve Bayes Classifier kemudian di evaluasi menggunakan Confusion Matrix. Confusion Matrix mengevaluasi performa model klasifikasi Naïve Bayes Classifier dengan membandingkan hasil prediksi model dengan nilai aktual (nilai dari label)

	Positif (aktual)		Negatif (aktual)	
Prediksi Positif	True (TP)	Positive (FP)	False (FN)	Positive (TN)
Prediksi Negatif	False (FN)	Negative (TN)	True (TP)	Negative (FP)

Tabel 1. Confusion Matrix

HASIL DAN PEMBAHASAN

Implementasi Data Scraping

Bagian ini berfungsi untuk melakukan data scraping berupa ulasan pengguna aplikasi Steam di Google Play Store dengan menggunakan Python., library pandas, time, dan juga google_play_scraper. Data yang diambil berupa nama pengguna yang memberikan ulasan, isi ulasan, rating ulasan (1-5), dan tanggal ulasan. Data-data ini disimpan dalam kolom 'userName', 'content', 'score', dan 'at'.

Implementasi Text Preprocessing

Tahap text preprocessing mencakup Casefolding, Text Cleaning, Stopwords Removal, Tokenisasi, dan Stemming. Hasil dari text preprocessing dapat dilihat pada Gambar 2

	content
0	cant play game app plataform made play game ne...
1	nice experi gcash account cash money
2	play game steam anywhere
3	app total garbag tri activ hour knowledg type ...
4	remov share button longer link promis game fri...

Gambar 2. Hasil Text Preprocessing

Implementasi TF-IDF

Untuk menerapkan ekstraksi fitur TF-IDF, fungsi TfidfVectorizer() mengubah teks dari data training X_train dan data testing X_test menjadi matriks angka yang mewakili frekuensi kata-kata dalam dokumen. Langkah pertama adalah menerapkan fit_transform() pada data training menjadi bentuk numerik, kemudian transform() diterapkan pada data testing untuk mengubahnya ke dalam bentuk yang sama dengan data training. Kedua matriks X_train dan X_test kemudian diubah menjadi array menggunakan .toarray().

Term	1	2	3	TF(1)	TF(2)	TF(3)	IDF	TF-IDF(1)	TF-IDF(2)	TF-IDF(3)
thought	1	0	0	0.09	0	0	0.47	0.0423	0	0
app	3	1	0	0.27	0.09	0	0.17	0.0459	0.0153	0
give	1	0	0	0.09	0	0	0.47	0.0423	0	0
game	1	2	1	0.09	0.18	0.12	0	0	0	0
even	1	0	0	0.09	0	0	0.47	0.0423	0	0
start	1	0	0	0.09	0	0	0.47	0.0423	0	0
don't	1	0	0	0.09	0	0	0.47	0.0423	0	0
think	1	0	0	0.09	0	0	0.47	0.0423	0	0
good	1	1	0	0.09	0.09	0	0.17	0.0153	0.0153	0
pretty	1	0	0	0.09	0	0	0.47	0.0423	0	0
doesn't	0	1	0	0	0.09	0	0.47	0	0.0423	0
show	0	2	0	0	0.18	0	0.47	0	0.0423	0
library	0	1	0	0	0.09	0	0.47	0	0.0423	0
incorrect	0	1	0	0	0.09	0	0.47	0	0.0423	0
playtime	0	1	0	0	0.09	0	0.47	0	0.0423	0
many	0	0	1	0	0	0.12	0.47	0	0	0.0564
bug	0	0	1	0	0	0.12	0.47	0	0	0.0564
cant	0	0	1	0	0	0.12	0.47	0	0	0.0564
buy	0	0	2	0	0	0.25	0.47	0	0	0.1175
option	0	0	1	0	0	0.12	0.47	0	0	0.0564
miss	0	0	1	0	0	0.12	0.47	0	0	0.0564

Gambar 3. Hasil TF-IDF

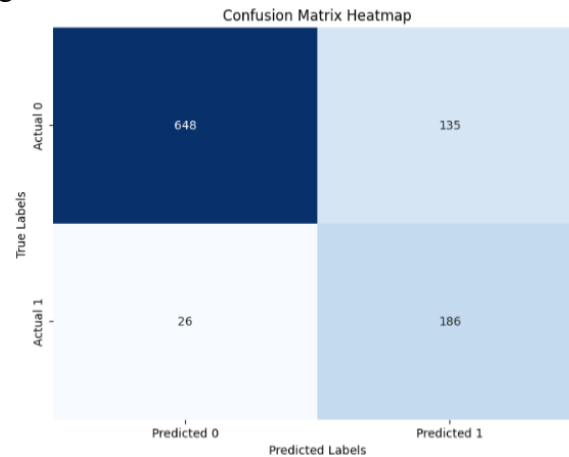
Hasil Evaluasi

Setelah dilakukan pelatihan dan pengujian pada model dengan rasio pembagian 80:20 yaitu pembagian dengan 80% data ke pelatihan dan 20% ke pengujian, kinerja model menunjukkan hasil seperti pada Gambar 4.

Classification Report:				
	precision	recall	f1-score	support
0.0	0.96	0.83	0.89	783
1.0	0.58	0.88	0.70	212
accuracy	nan	nan	0.84	995
macro avg	0.77	0.85	0.79	995
weighted avg	0.88	0.84	0.85	995

Gambar 4. Hasil Evaluasi Model

Dari hasil tersebut, terlihat bahwa model memiliki performa yang cukup bagus yaitu dengan akurasi 84%, dan precision untuk kelas negative sebesar 96%, sedangkan untuk kelas positif sebesar 58% dengan nilai rata-rata sebesar 77%.



Gambar 5. Confusion Matrix

Performa dari model juga menghasilkan Confusion Matrix seperti pada Gambar 5 dengan nilai TP, TN, FP, dan FN sebagai berikut:

$$TP = 186$$

$$TN = 648$$

$$FN = 26$$

$$TP = 186$$

Tingkat *accuracy* model:

$$\begin{aligned}
 Accuracy &= \frac{TP + TN}{TP + TN + FP + FN} \\
 &= \frac{186 + 648}{186 + 648 + 135 + 26} \\
 &= \frac{834}{995} = 0.8382 \approx 84\%
 \end{aligned}$$

Tingkat *precision* model:

$$\begin{aligned}
 Precision &= \frac{TP}{TP + FP} \\
 &= \frac{186}{321} = 0.5794 \approx 58\%
 \end{aligned}$$

Tingkat *recall* model:

$$\begin{aligned}
 Recall &= \frac{TP}{TP + FN} \\
 &= \frac{186}{212} = 0.8774 \approx 88\%
 \end{aligned}$$

Tingkat F1-score model:

$$\begin{aligned} F1 - Score &= 2 \times \frac{Precision \times Recall}{Precision + Recall} \\ &= 2 \times \frac{0.5794 \times 0.8774}{0.5794 + 0.8774} \\ &= 0.6996 \approx 70\%. \end{aligned}$$

KESIMPULAN

Klasifikasi sentimen ulasan pengguna aplikasi Steam dilakukan melalui tahapan text preprocessing, ekstraksi fitur, dan modelling. Dalam penelitian ini, selain data ulasan dan algoritma Naive Bayes, ekstraksi fitur menggunakan TF-IDF juga mempengaruhi akurasi klasifikasi sentimen. TF-IDF memberikan bobot pada kata-kata berdasarkan frekuensi kemunculannya di dokumen dan distribusinya di seluruh data, sehingga kata-kata yang lebih relevan dan unik diberi perhatian lebih besar. Dengan cara ini, model Naive Bayes dapat lebih akurat dalam membedakan sentimen positif dan negatif, karena dapat mengidentifikasi kata-kata yang memiliki pengaruh signifikan dalam menentukan sentimen. Berdasarkan evaluasi model dengan menggunakan Confusion Matrix, model yang dikembangkan mampu mencapai akurasi rata-rata sebesar 84%. Hasil ini menunjukkan bahwa model ini mampu membedakan dan mengklasifikasikan sentimen berdasarkan isi ulasan pengguna.

REFERENSI

- Chandra, R., & Sipayung, E. M. (2024). Analisis sentimen ulasan aplikasi Samsat Digital Nasional menggunakan algoritma Naive Bayes Classifier. *Jurnal Nasional Teknologi dan Sistem Informasi*, 10(3), 156–164. <https://doi.org/10.25077/teknosi.v10i3.2024.156-164>
- Ekrut. (2025, Januari 7). Apa itu Steam? *Ekrut Media*. <https://www.ekrut.com/media/apa-itu-steam>
- Esairina, A. (2021). Memahami confusion matrix, accuracy, precision, recall, specificity, dan F1-score. *Medium*. <https://esairina.medium.com/memahami-confusion-matrix-accuracy-precision-recall-specificity-dan-f1-score-610d4f0db7cf>
- Gunawan, B., Pratiwi, H. S., & Pratama, E. E. (2018). Sistem analisis sentimen pada ulasan produk menggunakan metode Naive Bayes. *Jurnal Edukasi dan Penelitian Informatika*, 4(2), 113–118.
- Hasibuan, E., & Heriyanto, E. A. (2022). Analisis sentimen pada ulasan aplikasi Amazon Shopping di Google Play Store menggunakan Naive Bayes Classifier. *Jurnal Teknik dan Sains*, 1(3), 13–24. <https://doi.org/10.56127/jts.v1i3.434>
- Mastan, I. A., & Toni, Y. (2020). Analisis sentimen terhadap tempat kuliner Ayam Gedebuk dari komentar pengunjung dengan menggunakan metode Naive Bayes Classifier. *JBASE - Journal of Business and Audit Information Systems*, 3(1). <https://doi.org/10.30813/jbase.v3i1.2062>
- Nadia, P. H., Wahyuni, S. S., & Sari, A. P. (2024). Penerapan Naive Bayes Classifier untuk analisis sentimen ulasan pelanggan pada Frenz Accessories Handphone. *JAMASTIKA*, 3(2), 19–29. <https://doi.org/10.35134/jamastika.v3i2.330>
- Putri, K. S., Setiawan, I. R., & Pambudi, A. (2023). Analisis sentimen terhadap brand skincare lokal menggunakan Naive Bayes Classifier. *Technologia*, 14(3), 227–232.
- Sudrajat, M. R., & Zakariyah, M. (2024). Penerapan natural language processing dan machine learning untuk prediksi stres siswa SMA berdasarkan analisis teks. *BISNIS & Teknologi Informasi (BiTS)*, 6(3). <https://doi.org/10.47065/bits.v6i3.6180>
- Takdirillah, R. (n.d.). Apa itu machine learning? Beserta pengertian dan cara kerjanya. *Dicoding*. <https://www.dicoding.com/blog/machine-learning-adalah/>

- Tannady, H., Andry, J. F., & Honni, H. (2024). Analisis big data Spotify dengan metode data mining. *JBASE - Journal of Business and Audit Information Systems*, 7(2).
<https://doi.org/10.30813/jbase.v7i2.6261>
- Tempo. (n.d.). Apa itu platform Steam? *Tempo.co*. <https://www.tempo.co/digital/apa-itu-platform-steam--314411>