

## ANALISIS SENTIMEN TERHADAP ULASAN PENGGUNA APLIKASI RUMAH PENDIDIKAN DI PLAYSTORE MENGGUNAKAN ALGORITMA *NAIVE BAYES*

Abdul Razaq<sup>1</sup>, Rahmat Hidayat<sup>2</sup>, Ryan Fahreza Pasaribu<sup>3</sup>,

Rian Farta Wijaya<sup>4</sup>, Zulham Sitorus<sup>5</sup>

Universitas Pembangunan Panca Budi, Medan

e-mail: <sup>1</sup>abdoell.razzaq@gmail.com, <sup>2</sup>92.rahmathidayat@gmail.com,

<sup>3</sup>ryanfahreza16@gmail.com, <sup>4</sup>rianfartawijaya@gmail.com, <sup>5</sup>zulhamsitorus@gmail.com

**Abstract:** *The development of digital technology has led to the emergence of various educational applications that play an important role in supporting the teaching and learning process. One of the most widely used applications is Rumah Pendidikan, which provides a variety of online learning features. However, user perceptions and experiences of this application vary greatly, requiring a systematic analysis to determine public sentiment towards the application. This study aims to apply the Naive Bayes Classifier algorithm in analyzing user sentiment reviews of the Rumah Pendidikan application on the Google Playstore platform. The data, consisting of 284 reviews, was classified into three sentiment categories, namely positive, negative, and neutral. The analysis process included data cleaning, tokenization, stopword removal, and TF-IDF vectorization before classification. The results show that the Naive Bayes algorithm is capable of classifying sentiment with an accuracy of 78.95%, a precision value of 72%, a recall of 67%, and an F1-score of 70%. These findings indicate that the Naive Bayes approach is effective in identifying user sentiment towards the Rumah Pendidikan application. The findings of this analysis can be used by developers as a basis for consideration in improving the quality of the application and user satisfaction in the future.*

**Keyword:** *Sentiment Analysis, Naive Bayes, Rumah Pendidikan Application, Google Play Store*

**Abstrak:** Perkembangan teknologi digital telah mendorong munculnya berbagai aplikasi pendidikan yang berperan penting dalam mendukung proses belajar-mengajar. Salah satu aplikasi yang banyak digunakan adalah Rumah Pendidikan, yang menyediakan beragam fitur pembelajaran daring. Meskipun demikian, persepsi dan pengalaman pengguna terhadap aplikasi ini sangat beragam, sehingga diperlukan analisis yang sistematis untuk mengetahui sentimen masyarakat terhadap aplikasi tersebut. Penelitian ini bertujuan untuk menerapkan algoritma *Naive Bayes Classifier* dalam menganalisis sentimen ulasan pengguna terhadap aplikasi Rumah Pendidikan pada platform *Google Playstore*. Data yang terdiri atas 284 ulasan diklasifikasikan ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral. Proses analisis meliputi tahap *data cleaning*, *tokenization*, *stopword removal*, serta *TF-IDF vectorization* sebelum dilakukan klasifikasi. Hasil penelitian menunjukkan bahwa algoritma *Naive Bayes* mampu mengklasifikasikan sentimen dengan akurasi sebesar 78.95%, nilai *precision* sebesar 72%, *recall* sebesar 67%, dan *F1-score* sebesar 70%. Temuan ini menunjukkan bahwa pendekatan *Naive Bayes* mampu bekerja secara efektif dalam mengidentifikasi sentimen pengguna terhadap aplikasi Rumah Pendidikan. Temuan analisis tersebut dapat dimanfaatkan oleh pengembang sebagai dasar pertimbangan dalam meningkatkan kualitas aplikasi dan tingkat kepuasan pengguna di masa mendatang.

**Kata kunci:** Analisis Sentimen, *Naive Bayes*, Rumah Pendidikan, *Playstore*

## PENDAHULUAN

Perkembangan teknologi digital dan internet telah membawa perubahan signifikan dalam bidang pendidikan, terutama melalui hadirnya berbagai aplikasi pembelajaran daring (*e-learning*) yang memfasilitasi proses belajar-mengajar tanpa batasan ruang dan waktu. Aplikasi semacam ini menjadi bagian penting dalam upaya meningkatkan akses dan kualitas pendidikan, khususnya di era transformasi digital (Kurniawan et al., 2025). Di Indonesia, aplikasi pendidikan tidak hanya digunakan oleh siswa dan guru, tetapi juga oleh masyarakat umum yang mencari informasi pembelajaran tambahan atau remidi secara daring. Namun demikian, keberhasilan sebuah aplikasi tidak hanya diukur dari jumlah unduhan atau fitur yang tersedia, melainkan juga dari tanggapan, pengalaman, dan persepsi pengguna terhadap aplikasi tersebut. Ulasan yang diberikan pengguna melalui platform seperti *Google Play Store* menjadi salah satu sumber data penting untuk memahami kualitas, kegunaan, dan kepuasan pengguna terhadap aplikasi (Zebua et al., 2021).

Analisis sentimen merupakan suatu metode yang memanfaatkan data teks ulasan untuk mengidentifikasi sikap, perasaan, serta opini pengguna terhadap suatu produk atau layanan. Teknik ini telah banyak digunakan dalam konteks aplikasi digital untuk mengungkap aspek-aspek yang mempengaruhi pengalaman pengguna, seperti kemudahan penggunaan, performa aplikasi, dan kepuasan layanan (Septiani Gumilar et al., 2024). Berdasarkan teori probabilitas Bayes serta asumsi kemandirian antarfitur, algoritma *Naive Bayes Classifier* (NBC) merupakan salah satu teknik yang banyak digunakan dalam analisis sentimen. Meski memiliki asumsi sederhana, metode ini terbukti efisien dan cukup efektif dalam klasifikasi teks, terutama pada dataset dengan dimensi

besar seperti ulasan pengguna (Septiani Gumilar et al., 2024).

Penelitian-penelitian terkini di Indonesia menunjukkan bahwa aplikasi berbasis mobile yang mendapatkan ulasan pengguna di *Play Store* dapat dianalisis menggunakan metode *Naive Bayes* untuk mengidentifikasi sentimen positif, netral, dan negatif. Sebagai contoh, studi pada aplikasi pendidikan menunjukkan bahwa mayoritas ulasan cenderung positif, namun masih terdapat keluhan terkait performa dan fitur aplikasi (Dhendra & Gayuh Utomo, 2025). Penelitian lain mengindikasikan bahwa klasifikasi sentimen ulasan aplikasi publik menggunakan *Naive Bayes* atau teknik serupa terus menjadi pilihan utama dalam mengevaluasi pengalaman pengguna (Tanggraeni & Sitokdana, 2022).

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna aplikasi Rumah Pendidikan di *Play Store* menggunakan algoritma *Naive Bayes*. Dengan menggunakan data sebanyak 284 ulasan, penelitian ini melakukan pra-pemrosesan teks dan klasifikasi sentimen ke dalam tiga kategori: positif, negatif, dan netral. Diharapkan bahwa hasil analisis ini dapat memberikan wawasan yang relevan bagi pengembang dalam upaya meningkatkan fungsionalitas, kualitas, serta tingkat kepuasan pengguna.

## METODE

Metode penelitian ini diterapkan untuk menjamin ketepatan dan keteraturan dalam pengumpulan serta analisis data. Tahapan metodologi yang ditampilkan pada Gambar 1 mencakup langkah-langkah krusial yang dilakukan untuk memperoleh hasil penelitian yang valid. Langkah-langkah alur kerja penelitian ditunjukkan pada Gambar 1, yang dimulai dari proses pengumpulan data ulasan aplikasi Rumah Pendidikan melalui *Google Play Store*, dilanjutkan dengan tahap pra-pemrosesan, pelabelan

```

4 # Install google-play-scraper into the environment
5 from google_play_scraper import Search, Review, All
6 import sys
7 import pandas as pd
8 import numpy as np
9 import re
10 import time
11 from datetime import datetime
12 from google_play_scraper.exceptions import InvalidPackageNameError
13
14 # Parameters
15 package_name = 'com.kemdiknas.go.id'
16 app_name = 'Aplikasi Rapor Pendidikan'
17 country_code = 'id'
18
19 # Convert local time to UTC
20 review_dt = pd.DataFrame(columns=['date'])
21
22 # Download data from file CSV
23 review_df = pd.read_csv('review_data.csv', index_col=0)
24
25 print('Data has been downloaded to file %s' % review_df)
26
27 # Requirements are satisfied: google-play-scraper is installed
28 # Data has been downloaded to file %s' % review_df)

```



Fase pertama penelitian terdiri atas pengumpulan data ulasan pengguna aplikasi Rumah Pendidikan dari *Google Play Store*. Proses ini dilakukan menggunakan teknik web *scraping*, yang memungkinkan peneliti mengekstraksi data ulasan secara otomatis dari halaman aplikasi pada platform tersebut. Dengan memanfaatkan *web scraping* menggunakan *library google-play-scraper* pada bahasa pemrograman *Python*, data ulasan secara otomatis diekstraksi dan diorganisir dalam format CSV". (Kurniawan et al., 2025).

Penelitian ini melakukan pengumpulan data dengan memanfaatkan *Google Colab* untuk mengekstrak ulasan aplikasi Rumah Pendidikan dari *Google Play Store*, kemudian menyimpannya dalam format CSV. Selama proses pengumpulan, peneliti berhasil memperoleh 284 ulasan yang dianggap relevan, meliputi komentar dan skor yang diberikan pengguna. Data yang terkumpul kemudian diekspor menjadi file .csv. Gambar 2 di bawah ini menunjukkan proses penyaringan data tersebut.

Gambar 2 menampilkan program yang digunakan untuk mengumpulkan ulasan aplikasi Rumah Pendidikan dari *Google Play Store* dengan memanfaatkan *library google\_play\_scraper*, mengubahnya menjadi *DataFrame* menggunakan *pandas*, dan menyimpannya ke dalam file CSV bernama 'rumahpendidikan\_reviews.csv'."

Fase pelabelan data melibatkan pengklasifikasian setiap ulasan pengguna di aplikasi Rumah Pendidikan ke dalam sentimen positif, negatif, atau netral, yang digunakan untuk melatih model *Naive Bayes* dalam mengukur sentimen (Arifin et al., 2025).

	content	score	sentiment
0	sudah bagus, semua fitur berjalan dengan baik.	5	Positif
1	bagus	4	Positif
2	Aplikasi yg ini sudah bagus tetapi ketika kita...	5	Positif
3	aplikasi sangat membantu untuk proses pembelaj...	5	Positif
4	dengan rumah pendidikan mengakses pembelajaran...	5	Positif

Berdasarkan nilai *score* tersebut diberi label sebagai sentimen yang dimodifikasi, yaitu “negatif” untuk nilai *score* 1 dan 2, “netral” untuk nilai *score* 3, dan “positif” untuk nilai *score* lainnya.

Langkah pertama dalam *preprocessing* adalah normalisasi teks. Teks yang tidak terstruktur memerlukan proses transformasi menjadi bentuk tulisan yang teratur untuk mempersingkat proses eksplorasi data dalam analisis sentiment. Pada langkah ini dilakukan normalisasi teks pada data *frame* untuk membakukan kata singkatan, misalnya mengubah ("yg","dgn","aja","nih","sih","kalo","ga","

gak","tdk","dong") (Khoerunnisa et al., 2025).

```
Contoh hasil preprocessing:
content \
0 sudah bagus, semua fitur berjalan dengan baik.
1 bagus
2 Aplikasi yg ini sudah bagus tetapi ketika kita...
```

**Gambar 4 Normalisasi Teks**

Setelah melakukan normalisasi teks, tindakan berikutnya yaitu *case folding*, dimana semua huruf pada teks *review* pengguna diubah menjadi huruf kecil, kemudian dilakukan tahap *cleansing* atau pembersihan teks (Khoerunnisa et al., 2025).

	content	score	sentiment
0	sudah bagus, semua fitur berjalan dengan baik.	5	Positif
1	bagus	4	Positif
2	Aplikasi yg ini sudah bagus tetapi ketika kita...	5	Positif
3	aplikasi sangat membantu untuk proses pembelaj...	5	Positif
4	dengan rumah pendidikan mengakses pembelajaran...	5	Positif

**Gambar 5 Case Folding**

*Cleansing* pembersihan teks dilakukan dengan menggunakan berbagai teknik termasuk penghapusan data duplikat, penghapusan teks kosong, penghapusan karakter dan tanda baca, penghapusan *Uniform Resource Locator* (URL) dan *HyperText Markup Language* (HTML), serta penghapusan simbol dan referensi yang tidak terkait. Hasil tahap pembersihan teks seperti pada Gambar 6.

	content	score	sentiment
0	sudah bagus, semua fitur berjalan dengan baik.	5	Positif
1	bagus	4	Positif
2	Aplikasi yg ini sudah bagus tetapi ketika kita...	5	Positif
3	aplikasi sangat membantu untuk proses pembelaj...	5	Positif
4	dengan rumah pendidikan mengakses pembelajaran...	5	Positif

**Gambar 6. Cleansing**

Setelah tahap pembersihan teks, langkah selanjutnya adalah proses *tokenizing*, memecah teks sehingga jadi lebih kecil untuk memudahkan analisis.

	filtered_content
0	aplikasi mempermudah mencari pekerjaan fiturfitur mudah cont
1	aplikasi membantu fitur lengkap antarmuka intuitif kitalulus r
2	aplikasi mempermudah mencari pekerjaan sesuai pengalaman bi
3	aplikasinya membantu mencari pekerjaan diaplikasi update pilih
4	aplikasi membantu mencari pekerjaan pilihan jenis pekerjaan i

**Gambar 7 Tokenizing Data**

Berikutnya adalah fase *stopword removal*, dimana kata-kata umum yang biasanya tidak mempunyai arti khusus dalam analisis teks dihilangkan seperti terlihat pada Gambar 8.

	filtered_content
0	aplikasi mempermudah mencari pekerjaan fiturfitur mudah cont
1	aplikasi membantu fitur lengkap antarmuka intuitif kitalulus r
2	aplikasi mempermudah mencari pekerjaan sesuai pengalaman bi
3	aplikasinya membantu mencari pekerjaan diaplikasi update pilih
4	aplikasi membantu mencari pekerjaan pilihan jenis pekerjaan i

**Gambar 8 Stopword Removal**

Setelah proses penghapusan *stopwords* selesai, langkah terakhir dalam *preprocessing* adalah *stemming*. *Stemming* adalah tahap perubahan kata yang mempunyai sufiks ke bentuk dasarnya. Proses ini penting untuk memastikan bahwa semua variasi kata dianggap sebagai entitas yang sama oleh model [15].

	stemmed_content
0	aplikasi mudah cari kerja fiturfitur mudah contoh pilih jenis
1	aplikasi bantu fitur lengkap antarmuka intuitif kitalulus bantu
2	aplikasi mudah cari kerja sesuai alam bidang apa admin
3	aplikasi bantu cari kerja aplikasi update pilih jenis kerja c
4	aplikasi bantu cari kerja pilih jenis kerja tipe kerja cv k

**Gambar 9 Stemming**

## Pembobotan TF-IDF

Setelah data telah melalui seluruh proses *preprocessing* yang meliputi normalisasi teks, *case folding*, pembersihan teks, *tokenizing*, *stopword removal* dan *stemming*, langkah selanjutnya adalah menghitung bobot TF - IDF.

```
y_pred = model.predict(X_test_vec)
print("\n== HASIL EVALUASI MODEL ==")
print("Akurasi:", round(accuracy_score(y_test, y_pred)*100,2), "%")
print("\nClassification Report:\n", classification_report(y_test, y_pred))
cm = confusion_matrix(y_test, y_pred, labels=["Positif","Netral","Negatif"])
print("\nConfusion Matrix:\n", cm)
```

**Gambar 10 Proses Pembobotan TF-IDF**

## Pembuatan Model

Tahap selanjutnya adalah pembuatan model, dimana model machine learning dibangun untuk tindakan analisis sentimen dari ulasan pengguna. Dari tahap ini, model *Naive Bayes* digunakan untuk klasifikasi teks. Model ini dilatih

Berdasarkan hasil klasifikasi, terlihat bahwa model *Naive Bayes* mendapat nilai akurasi 78.95%. Menggali lebih dalam hasil untuk setiap kategori, kami

menemukan bahwa model ini unggul dalam mengklasifikasikan kelas positif secara akurat, dengan *precision* 0,78, *recall* 1,00 dan *f1-score* sebesar 0,88. Namun, kinerjanya relatif lebih rendah ketika mengklasifikasikan kelas negatif dan netral. Kelas negatif menunjukkan *precision*, *recall*, dan *f1-score* masing-masing 0,31, sedangkan kelas netral menunjukkan *precision* 0,00, *recall* 0,00, dan *f1-score* 0,31. Dengan mempertimbangkan macro avg, menunjukkan *precision* (0,59), *recall* (0,39), dan *f1-score* (0,40). Di sisi lain, weighted avg menunjukkan bahwa model ini mempertahankan tingkat kinerjanya yang tinggi (*precision* 0.78, *recall* 0.79, *f1-score* 0.72), terutama karena jumlah sampel yang dominan di kelas positif. Rata-rata akurasi cross-val: 75.33 %

## SIMPULAN

Hasil penelitian memperlihatkan bahwa model klasifikasi *Naive Bayes* bekerja secara efektif dalam menilai sentimen pengguna terhadap aplikasi Rumah Pendidikan berdasarkan ulasan terbaru, dapat disimpulkan bahwa mayoritas ulasan bersifat positif (75,4%), diikuti ulasan negatif (20,1%) dan netral (4,6%) dari total 284 data. Hasil evaluasi model menunjukkan bahwa model mampu mengklasifikasikan ulasan positif dengan baik, terbukti dari *precision* 0,78, *recall* 1,00, dan *F1-score* 0,88. Namun, performa model untuk ulasan netral dan negatif masih terbatas; ulasan netral tidak berhasil dikenali (*precision*, *recall*, *F1-score* 0,00), sedangkan ulasan negatif meskipun memiliki *precision* 1,00, *recall* yang rendah 0,18 menghasilkan *F1-score* 0,31, menandakan banyak ulasan negatif yang tidak terdeteksi. Secara keseluruhan, akurasi model mencapai 0,79, dengan rata-rata *F1-score* makro 0,40 dan rata-rata *F1-score* tertimbang 0,72, menunjukkan bahwa model efektif untuk kelas dominan (positif), namun memerlukan peningkatan performa dalam mendeteksi ulasan minor (negatif dan

netral). Aplikasi KitaLulus terbukti memberikan efek positif bagi komunitas, karena mampu mengidentifikasi dan menampilkan mayoritas ulasan yang bersifat positif secara efektif."

## DAFTAR PUSTAKA

- Arifin, M. N., Amir Hamzah, Huda, Moh. A., & Hasanah, N. (2025). Analysis of Google Play Store User Sentiment Towards Application X Using the SVM Algorithm. *Brilliance: Research of Artificial Intelligence*, 5(1), 249–258. <https://doi.org/10.47709/brilliance.v5i1.6024>
- Dhendra, & Gayuh Utomo, V. (2025). Benchmarking IndoBERT and Transformer Models for Sentiment Classification on Indonesian E-Government Service Reviews. *Jurnal Transformatika*, 23(1), 86–95. <https://doi.org/10.26623/transformatika.v23i1.12095>
- Khoerunnisa, S., Shiddieq, D. F., & Nurhayati, D. (2025). Penerapan Algoritma Naive Bayes dengan Teknik TF-IDF dan Cross Validation untuk Analisis Sentimen Terhadap Starlink. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 5(2), 566–577. <https://doi.org/10.57152/malcom.v5i2.1852>
- Kurniawan, Y. I., Sidiq, R. H., Widadi, A. D. U., Yubiharto, A. R. A., Gibran, A. K., Aditama, M. R., & Laksono, F. A. T. (2025). Naive Bayes Classifier with SMOTE for Sentiment Analysis of Bilibili App Reviews on The Google Play Store. *Jurnal Penelitian Inovatif*, 5(3), 2675–2688. <https://doi.org/10.54082/jupin.1842>
- Septiani Gumilar, T., Astuti, R., & Wijaya, Y. A. (2024). ANALISIS SENTIMEN ULASAN APLIKASI LITA DI PLAY STORE MENGGUNAKAN ALGORITMA NAIVE BAYES. *Jurnal Mahasiswa*

- 
- Teknik Informatika*, 8(1), 543.
- Tanggraeni, A. I., & Sitokdana, M. N. N. (2022). Analisis Sentimen Aplikasi E-Government Pada Google Play Menggunakan Algoritma Naïve Bayes. *Jurnal Teknik Informatika Dan Sistem Informasi*, 9(2), 785–795.
- Tsani, M. R., Prima, A., Rupaka, G., Asmoro, L., & Pradana, B. (2020). ANALISIS SENTIMEN REVIEW TRANSPORTASI MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE BERBASIS CHI SQUARE. *Smart Comp*, 9(1). <http://www.yelp.com/nyc>,
- Zebua, F. J., Br Manalu, R. P., & Nababan, M. N. K. (2021). PREDIKSI KELULUSAN MAHASISWA MENGGUNAKAN PERBANDINGAN ALGORITMA C5.0 DENGAN REGRESSION LINEAR. *Jurnal Teknik Informasi Dan Komputer (Tekinkom)*, 4(2), 230. <https://doi.org/10.37600/tekinkom.v4i2.400>