JAISEN: **Journal of Advanced Information Systems and Engineering**
Vol. 1, 1 (August, 2025), pp 1-8
ISSN: - EISSN: - , DOI: -

ANALYSIS OF GOOGLE CAPCUT APP USER REVIEWS USING NAIVE BAYES

Najmuddin Mubarak MR[1*], Ahmad Zamsuri[2], Susandri Susandri[3], Feldiansyah[4]

[1] Magister Ilmu Komputer, Universitas Lancang Kuning; mudinnajim85@gmail.com
[2] Magister Ilmu Komputer, Universitas Lancang Kuning; ahmadzamsuri@unilak.ac.id
[3] Magister Ilmu Komputer, Universitas Lancang Kuning; susandri@unilak.ac.id
[4] Magister Ilmu Komputer, Universitas Lancang Kuning; feldiansyah@unilak.ac.id

ARTICLE INFO

Keywords:

Sentimen Analysis;
Naive Bayes;
Capcut;
Natural Language Processing;
Knowledge Discovery Data .

Article history:

Received 2025-08-09
Revised 2025-08-10
Accepted 2025-08-13

ABSTRACT

In the current digital landscape, video editing platforms such as CapCut have emerged as vital tools for content creators across various social media channels. Feedback from users—whether positive, negative, or neutral—serves as a critical resource for guiding application improvement. This research seeks to examine user sentiment toward CapCut through the application of the Naive Bayes classification algorithm, recognized for its efficiency and straightforward implementation in processing textual data. User review data was obtained via web scraping from the Google Play Store, yielding a dataset of 100 entries. The study was conducted following the Knowledge Discovery in Databases (KDD) framework, which includes stages of data selection, preprocessing, text transformation, classification using Naive Bayes, and assessment through performance metrics such as accuracy, precision, recall, and F-measure. Results demonstrate that the algorithm successfully distinguishes between sentiment categories, while sentiment visualization through word clouds highlights the most frequently occurring terms. The outcomes of this study offer practical insights for developers seeking to enhance application quality, as well as scholarly contributions to the fields of sentiment analysis and machine learning–based text mining.

Corresponding Author:
Najmuddin Mubarak MR
Universitas Lancang Kuning ; mudinnajim85@gmail.com

1. INTRODUCTION

In the current digital landscape, video editing applications have become indispensable tools for content creators, particularly those producing material for social media platforms such as Instagram, TikTok, and YouTube. Among these tools, Google CapCut stands out as one of the most widely adopted applications due to its rich set of editing features, intuitive interface, and accessibility (Nurian, 2023). The application's popularity has generated a substantial volume of user reviews reflecting a broad spectrum of experiences—ranging from high levels of satisfaction with its creative functionalities to criticisms concerning its stability, speed, and technical performance (Hendra & Fitriyani, 2021). While this abundance of user feedback provides valuable information, its unstructured nature makes comprehensive interpretation challenging without a systematic analytical framework.

The urgency of conducting this research is rooted in the growing need for developers to obtain a nuanced understanding of user feedback (Gunawan, Fauzi, & Adikara, 2017). Reviews are not merely expressions of user satisfaction or dissatisfaction; they also serve as a strategic resource for identifying emerging consumer preferences, guiding feature development, and informing marketing strategies (Insan, Hayati, & Nurdiawan, 2023). Sentiment analysis has emerged as a robust methodological approach for addressing this need, enabling the classification of textual feedback into sentiment categories—positive, negative, or neutral—thereby transforming raw, unstructured data into actionable insights (Simanjuntak et al., 2023).

One of the core challenges in sentiment analysis lies in managing the scale and complexity of large textual datasets, which require algorithms capable of efficient and accurate processing (Muflih, Abdillah, & Hasan, 2023). The Naive Bayes classification algorithm is widely recognized for its computational simplicity, scalability, and high accuracy in text classification tasks. Its probabilistic approach allows it to deliver reliable sentiment predictions with relatively low processing time, making it an effective choice for analyzing extensive collections of CapCut user reviews (Irnawati & Solecha, 2022). Employing Naive Bayes in this context addresses the dual challenge of data volume and linguistic diversity while maintaining analytical efficiency (Indarwati & Februariyanti, 2023).

The objective of this study is to systematically evaluate the sentiment expressed in Google CapCut user reviews by applying the Naive Bayes algorithm (Santoso & Wiboowo, 2022). The anticipated outcomes are twofold: from a practical perspective, the findings are expected to assist application developers in refining product quality and enhancing user satisfaction; from an academic standpoint, the research will contribute to the body of knowledge in natural language processing and computational text analytics, offering a methodological reference for future studies in sentiment classification and user experience research (Indriyani, Fauzi, & Faisal, 2023; Mubarak et al., 2025).

2. METHODS

The Introduction section should also incorporate a concise description of the research method. This includes an explanation of the research approach, the characteristics of the study participants, the procedures undertaken during the research, the materials and instruments employed, as well as the techniques used for data collection and data analysis.

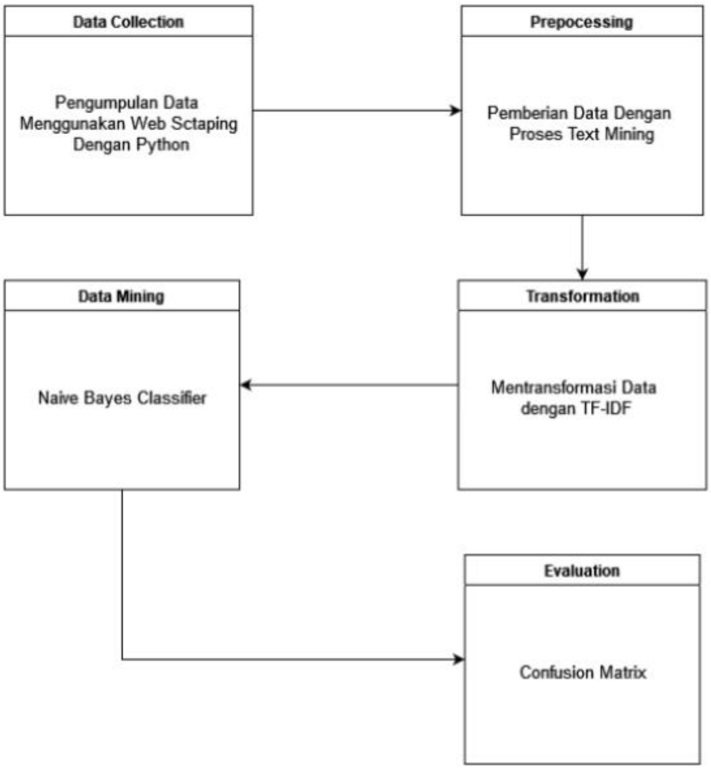

Figure 1. Research methodology

2.1 Data Selection

Data was collected from the Google Play Store regarding user reviews of the CapCut app. The collected dataset consisted of 100 data points.

2.2 Preprocessing

The text preprocessing stage is the initial stage of text mining. Text preprocessing is the process of selecting data to be processed in each document. The preprocessing stage is necessary to clean the data of unnecessary words (Hasibuan & Heriyanto, 2022). The preprocessing stage in this study consists of cleaning, normalization, and labeling (Zhafira, Rahayudi, & Indarti, 2021).

2.3 Text Transformation

Text transformation is the third stage in text mining, which aims to prepare text data for data mining (Nitamia & Februariyanti, 2022).

2.4 Data Mining

Naive Bayes Classifier: The classification stage using the Naive Bayes algorithm is divided into two processes: training and testing. First, training is performed, then testing is performed based on the probabilities from the training dataset. The Naive Bayes Classifier method is one method used to effectively classify opinions. Capable of categorizing comments into positive, neutral, or negative regarding a product or issue currently circulating among the public (Septiani, Aribbe, & Diansyah, 2020).

2.5 Evaluation

The evaluation stage is the stage carried out to assess the performance of the classification algorithm used in the research. The benchmarks used to measure this performance are accuracy, precision, recall, and f-measure (Tanggraeni & Sitokdana, 2020). To provide a clearer picture of performance, the confusion matrix allows analysis across four dimensions:

1. Accuracy: To determine the level of similarity between predicted values and actual values.
2. Precision: To determine how precise or accurate the model is in predicting positive values. Precision is also a good measure for determining when a model has a high False Positive rate.
3. Recall: To calculate how many of the Actual Positive values were successfully identified by the model by labeling them as True Positive. Recall will also be the model metric used to select the best model when there is a high value associated with False Negatives.
4. F-Measure: A weighted average comparison of precision and recall values.

3. FINDINGS AND DISCUSSION

The results and discussion of sentiment analysis on the Dana app using the Naive Bayes classifier processed in Jupyter. Examples of scraped and labeled data, preprocessing, and accuracy values for each are presented in the next subchapter (Dhery, Assyam, & Hasan, 2023).

3.1 Data Scraping

In this stage, data was scraped by including the Dana app link from the Play Store. The results of the data scraping, using the keyword Capcut and a search weight of 100 entries in the Capcut app, can be seen below using a script (Parastati, Bachtiar, & Setiawan, 2020).

Scarping
from google_play_scraper import app from google_play_scraper import Sort, reviews result, continuation_token = reviews('com.lemon.lvoverseas', lang='id', country='id', sort=Sort.NEWEST, count=100, filter_score_with=None)

The following is the output:

[25]:

	content	score	at
0	Minnn Watermark nya ilangin woi,, masa mau nyi...	1	2024-11-18 01:45:13
1	bagus banget si aplikasi nya cuman itu ga usah...	5	2024-11-18 01:25:10
2	Kurang nyaman sama iklan nya..	3	2024-11-18 01:22:35
3	sekarang capcut jelek banget masa mau mengeksp...	1	2024-11-18 01:22:20
4	Semakin banyak orang mengunakan aplikasi ini a...	1	2024-11-18 01:21:27
...	...	...	...
95	👍	5	2024-11-17 21:34:23
96	Bagus	5	2024-11-17 21:34:22
97	Ruangan saya penuh	5	2024-11-17 21:32:01
98	Kondisi mulus dan bersih	1	2024-11-17 21:29:50
99	udah bnyak iklan, hrus pro dlu buat hapus wate...	1	2024-11-17 21:29:15

100 rows × 3 columns

Figure 2. Data Crawling Results

3.2 Automatic Labeling

After collecting the data, the research conducted labeling in the Capcut app. The labeling output is shown in Figure 3.

	content	score	Label
0	Minnn Watermark nya ilangin woi,, masa mau nyi...	1	Negatif
1	bagus banget si aplikasi nya cuman itu ga usah...	5	Positif
2	Kurang nyaman sama iklan nya..	3	Netral
3	sekarang capcut jelek banget masa mau mengeksp...	1	Negatif
4	Semakin banyak orang mengunakan aplikasi ini a...	1	Negatif
5	Ok good	5	Positif
6	Kurang" In iklan nya, Bgus	5	Positif
7	Pencarian tamplate hilang 😭😭😭😭	1	Negatif
8	Gabisa di download	1	Negatif
9	Update wae euyy capcut na anu ayenamh beda jen...	1	Negatif

Figure 3. Data Labeling Results

3.3 Preprocessing

Next, the preprocessing stage was carried out. The following is the result of sampling text for preprocessing, taking one sample text. The process begins with the original data from the scraping process, followed by removing URLs, removing non-ASCII characters, removing excess spaces, removing numbers, removing punctuation, removing hashtags, removing non-alphanumeric characters, and converting the text to lowercase (Nurkumalawati & Rofli, 2023).

[254]:

	content	score	at	content_token
0	menurut ku apk nya bagus bgt dan berguna , apa...	5	2024-11-17 00:53:03	menurut ku apk nya bagus bgt dan berguna apala...
1	merepotkan	1	2024-11-17 00:52:36	merepotkan
2	aplikasi sampah fitur sampah udah bagus bagus ...	1	2024-11-17 00:48:50	aplikasi sampah fitur sampah udah bagus bagus ...
3	bagus untuk mengedit	5	2024-11-17 00:42:55	bagus untuk mengedit
4	Lamak bnr buka nyaa	1	2024-11-17 00:36:16	lamak bnr buka nyaa
...	...	...	...	...
95	Aplikasi nya ngelag-ngelag padahal sudah pake ...	1	2024-11-16 21:58:03	aplikasi nya ngelag ngelag padahal sudah pake ...
96	Ini kenpa ga bisa di download yaaaa pdhl jarin...	1	2024-11-16 21:57:49	ini kenpa ga bisa di download yaaaa pdhl jarin...
97	Maaf ya kalo aku login pasti ada tulisan *coba...	1	2024-11-16 21:55:43	maaf ya kalo aku login pasti ada tulisan coba ...
98	Baguss	5	2024-11-16 21:48:59	baguss
99	jelek, kadang atur foto nya ngebug trus suka t...	2	2024-11-16 21:47:43	jelek kadang atur foto nya ngebug trus suka te...

100 rows × 4 columns

Figure 4. Data Preprocessing Results

3.4 Accuracy Results

The classification stage is carried out by creating a machine learning framework using training and testing data on all data in the dataset randomly to perform cross-validation and generate prediction values for accuracy. Below is an illustration of the results of the classification stage using a script from the Naïve Bayes algorithm (Giovani *et al*., 2020).

```
===== Multinomial Naive Bayes =====
Accuracy: 0.625
Classification Report:
              precision    recall  f1-score   support

           0       0.21      1.00      0.35         3
           1       1.00      0.77      0.87        13
           2       0.88      0.44      0.58        16

    accuracy                           0.62        32
   macro avg       0.70      0.74      0.60        32
weighted avg       0.86      0.62      0.68        32

=====================
```

Figure 5. Accuracy Results

The results of the sentiment analysis classification of Capcut app user reviews can be visualized using a word cloud to provide an overview or general information about the Capcut app user review data on the Google Play site. The following is a discussion of the word visualization for each sentiment class.

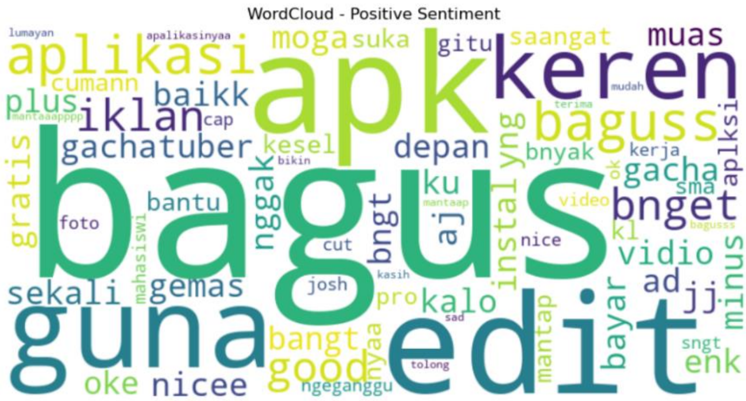

Figure 6. Positive WordCloud Results

Figure 6 shows that the words "good," "edit," "use," "apk," and "cool" are the most frequently used words in Capcut app reviews in this study. The larger the word size in the word cloud, the higher the frequency of the word.

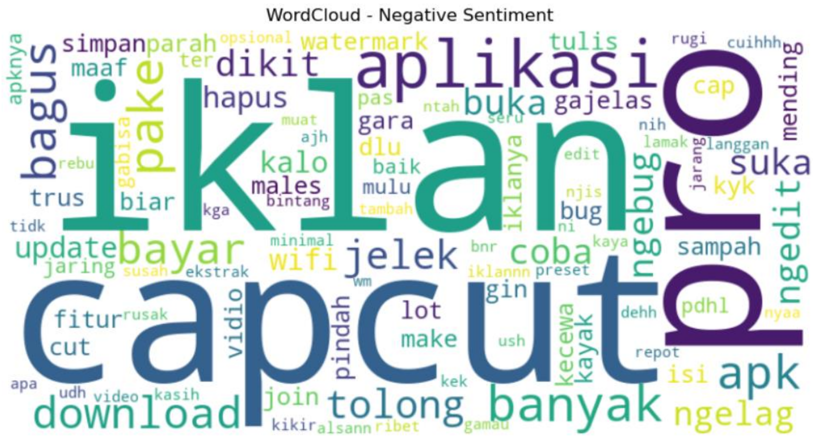

Figure 7. Negative WordCloud Results

Figure 7 shows that the words "advertisement," "capcut," "pro," "application," and "download" were the most frequently used words in reviews of the Capcut application in this study. The larger the word cloud, the higher the frequency.

Figure 8. Neutral WordCloud Results

Figure 8 shows that the words "pro," "ad," "edit," "application," and "mantap" were the most frequently used words in CapCut app reviews in this study. The larger the word size in the word cloud, the higher the frequency.

4. CONCLUSION

This study successfully identified CapCut user review sentiment using the Naive Bayes algorithm with a systematic approach through the KDD method. The analysis process, from review data collection to algorithm performance evaluation, yielded good accuracy in sentiment classification. The word cloud visualization shows the dominant keywords in each sentiment, such as "good" and "cool" for positive sentiment, and "ad" and "pro" for negative sentiment. This research provides practical benefits for app developers in understanding user needs and improving products, while also serving as an academic reference for similar research in the fields of text data analysis and natural language processing.

REFERENCES

Dhery, H., Assyam, A., & Hasan, F. N. (2023). Analisis sentimen Twitter terhadap perpindahan ibu kota negara ke IKN Nusantara menggunakan Orange Data Mining. *KLIK: Kajian Ilmiah Informatika dan Komputer*, 4(1), 341–349.

Giovani, A. P., Ardiansyah, A., Haryanti, T., Kurniawati, L., & Gata, W. (2020). Analisis sentimen aplikasi Ruang Guru di Twitter menggunakan algoritma klasifikasi. *Jurnal Teknoinfo*, 14(2), 115.

Gunawan, F., Fauzi, M. A., & Adikara, P. P. (2017). Analisis sentimen pada ulasan aplikasi mobile menggunakan Naive Bayes dan normalisasi kata berbasis Levenshtein distance (Studi kasus aplikasi BCA Mobile). *Systemic: Information System and Informatics Journal*, 3(2), 1–6.

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.

Hendra, A., & Fitriyani, F. (2021). Analisis sentimen review Halodoc menggunakan Naïve Bayes classifier. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 6(2), 78–89.

Indarwati, K. D., & Februariyanti, H. (2023). Analisis sentimen terhadap kualitas pelayanan aplikasi Gojek menggunakan metode Naive Bayes classifier. *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, 10(1).

Indriyani, F. A., Fauzi, A., & Faisal, S. (2023). Analisis sentimen aplikasi TikTok menggunakan algoritma Naïve Bayes dan support vector machine. *Teknosains: Jurnal Sains, Teknologi dan Informasi*, 10(2), 176–184.

Insan, M. K., Hayati, U., & Nurdiawan, O. (2023). Analisis sentimen aplikasi Brimo pada ulasan pengguna. *Jurnal Mahasiswa Teknik Informatika*, 7(1), 478–483.

Irnawati, O., & Solecha, K. (2022). Analisis sentimen ulasan aplikasi Flip menggunakan Naïve Bayes dengan seleksi fitur PSO. *Jurnal Ilmiah Intech: Information Technology Journal UMUS*, 4(2), 189–199.

Mubarak, N., Salisah, F. N., Megawati, M., & Muttakin, F. (2025). Perancangan arsitektur sistem informasi enterprise menggunakan TOGAF pada Ma'Had Al Husna Indragiri. *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 10(1), 381–395.

Muflih, H. Z., Abdillah, A. R., & Hasan, F. N. (2023). Analisis sentimen ulasan pengguna aplikasi Ajaib menggunakan metode Naïve Bayes. *KLIK: Kajian Ilmiah Informatika dan Komputer*, 4(3), 1613–1621.

Nitamia, M. T., & Februariyanti, H. (2022). Analisis sentimen ulasan ekspedisi J&T Express menggunakan algoritma Naive Bayes. *Jurnal Manajemen Informatika dan Sistem Informasi*, 5(1), 20–29.

Nurian, A. (2023). Analisis sentimen ulasan pengguna aplikasi Google Play menggunakan Naïve Bayes. *Jurnal Informatika dan Teknik Elektro Terapan*, 11(3s1), 829–835.

Nurkumalawati, I., & Rofii, M. S. (2023). Public review of M-Paspor application in Indonesia: Mobile government, digital resilience, cyber security. In *Proceedings of the International Conference on Mobile Government* (pp. 1–8). Atlantis Press.

Parasati, W., Bachtiar, F. A., & Setiawan, N. Y. (2020). Analisis sentimen berbasis aspek pada ulasan pelanggan Restoran Bakso President Malang dengan metode Naïve Bayes classifier. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 4(4), 1–10.

Santoso, D. P., & Wibowo, W. (2022). Analisis sentimen ulasan aplikasi Buzzbreak menggunakan metode Naïve Bayes classifier pada situs Google Play Store. *Jurnal Sains dan Seni ITS*, 11(2).

Septiani, Y., Aribbe, E., & Diansyah, R. (2020). Analisis kualitas layanan sistem informasi akademik Universitas Abdurrab terhadap kepuasan pengguna menggunakan metode Servqual (Studi kasus: Mahasiswa Universitas Abdurrab Pekanbaru). *Jurnal Teknologi dan Open Source*, 3(1), 131–143.

Simanjuntak, R. L., Siagian, T. R., Anggriani, V., & Arnita, A. (2023). Analisis sentimen ulasan pada aplikasi e-commerce Shopee dengan menggunakan algoritma Naïve Bayes. *Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, 3(3), 23–39.

Tanggraeni, A. I., & Sitokdana, M. N. N. (2022). Analisis sentimen aplikasi e-government pada Google Play menggunakan algoritma Naïve Bayes. *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, 9(2), 785–795.

Zhafira, D. F., Rahayudi, B., & Indriati, I. (2021). Analisis sentimen. *Jurnal Sistem Informasi, Teknologi Informasi, dan Edukasi Sistem Informasi*, 2(1), 55–63.