EVALUATION OF USER PERCEPTIONS AND SATISFACTION THROUGH SENTIMENT ANALYSIS NEWS APPLICATIONS WITH NAIVE BAYES

Aldiansyah Kusuma[1*]; Diaz Aditya Yudha[2] ; Muhammad Bahril Afwa[3]; Hanafi Eko Darono[4]

Information System[1,2,3]
Universitas Bina Sarana Informatika, Jakarta, Indonesia[1,2,3,4]
https://www.bsi.ac.id/[1,2,3,4]
aldiansyahkus83@gmail.com[1*], diaz.yudha29@gmail.com[2] , muhammadbahril19@gmail[3],
hanafi.haf@bsi.ac.id[4]
(*) Corresponding Author

Abstract—*The development of digital technology has driven the transformation of mass media into online news platforms such as Detikcom, Kompas.id, and CNN Indonesia. Competition among these news applications has created the need to evaluate user perceptions of service quality. This study aims to analyze user sentiment toward the three news applications based on reviews from the Google Play Store. The methods employed include web scraping, text pre-processing, labeling using the IndoBERT model, feature extraction with the TF-IDF method, and sentiment classification with the Naive Bayes algorithm. To address class imbalance in the dataset, the Synthetic Minority Over-sampling Technique (SMOTE) was applied. Model evaluation was conducted using accuracy, precision, recall, and F1-score metrics. The results show that the Naive Bayes model achieved high accuracy, namely 88.5% for Kompas.id, 88.8% for Detikcom, and 90.8% for CNN Indonesia. The analysis also revealed that positive reviews are more dominant, although recurring criticisms were identified regarding advertisements and technical performance of the applications. The use of Generative AI further assisted in automatically summarizing opinions and sentiment patterns. These findings provide valuable insights for developers in enhancing user experience and refining the features of digital news applications .*

Keywords: *IndoBERT, Naive Bayes, Online News Application, Sentiment Analysis, SMOTE.*

Intisari—Perkembangan teknologi digital telah mendorong transformasi media massa menjadi platform berita online seperti Detikcom, Kompas.id, dan CNN Indonesia. Persaingan antar aplikasi berita ini menimbulkan kebutuhan untuk mengevaluasi persepsi pengguna terhadap kualitas layanan. Penelitian ini bertujuan untuk menganalisis sentimen pengguna terhadap ketiga aplikasi berita tersebut berdasarkan ulasan di Google Play Store. Metode yang digunakan meliputi web scraping, pre-processing data teks, pelabelan menggunakan model IndoBERT, ekstraksi fitur dengan metode TF-IDF, serta klasifikasi sentimen dengan algoritma Naive Bayes. Untuk mengatasi ketidakseimbangan kelas dalam dataset, digunakan teknik SMOTE (Synthetic Minority Over-sampling Technique). Evaluasi model dilakukan menggunakan metrik accuracy, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa model Naive Bayes memiliki performa akurasi yang tinggi, yakni 88,5% untuk Kompas.id, 88,8% untuk Detikcom, dan 90,8% untuk CNN Indonesia. Analisis juga mengungkapkan bahwa ulasan positif lebih dominan, meskipun terdapat kritik berulang terkait iklan dan performa teknis aplikasi. Penggunaan Generative AI turut membantu dalam merangkum opini dan pola sentimen secara otomatis. Temuan ini memberikan wawasan berharga bagi pengembang dalam meningkatkan pengalaman pengguna serta menyempurnakan fitur aplikasi berita digital.

Kata Kunci: *IndoBERT, Naive Bayes, Aplikasi Berita Online, Analisis Sentimen, SMOTE.*

INTRODUCTION

Since its inception, journalism has always been closely linked with technology. The invention of the 15th-century printing press made news faster and cheaper, establishing newspapers as the primary information source. Subsequently, the telegraph accelerated news delivery from days to hours. Both breakthroughs strengthened journalism's role as a provider of fast and current information. (Arlovin, Kusrini, & Kusnawi, 2024).

Advancement continued with the arrival of radio and television. Radio allowed auditory news delivery, while television combined moving images and sound, making journalism more vivid and engaging. This era of electronic media dominated for decades. The next transformative milestone was the internet, which introduced online journalism. This format enabled instant reporting through text, images, audio, video, and live streaming, making journalism more interactive and accessible anytime and anywhere.

New media, particularly the internet, greatly influenced information preferences. Previously, sources were limited to newspapers, radio, and television. Today, digital platforms provide instant access to diverse sources, including portals, social media, and apps. As a result, audiences have become more selective and diverse in choosing sources aligned with their needs (Hapsari & Priliantini, 2025).

This shift encouraged media companies to develop digital platforms, especially smartphone-based news apps. Online media became one of the most consumed forms of mass media, offering accessibility beyond print and electronic predecessors. Consequently, competition among online platforms intensified, focusing on content quality, app features, and user experience (Kusnia & Kurniawan, 2022).

In Indonesia, popular platforms include Detikcom, Kompas.id, and CNN Indonesia. They compete not only in content but also in design, speed, and interactivity. New portals continuously emerge, creating challenges for developers to innovate and retain engagement. A key measure of user satisfaction lies in reviews available on app stores like Google Play (Larasati, Ratnawati, & Hanggara, 2022).

On Google Play, users provide ratings and text-based reviews, functioning as open forums for experiences, opinions, and criticisms. This study focuses on such reviews as direct reflections of public evaluation. Reviews generally fall into positive containing appreciation and suggestions or negative, expressing dissatisfaction (Kusnia & Kurniawan, 2022).

Reviews are common across industries, especially digital services, and many people rely on them before making decisions. These texts contain experiences and evaluations, making them valuable for understanding user needs, preferences, and complaints (Nadira, Setiawan, & Purnomo, 2023).

To analyze such data, sentiment analysis is often applied. Also called opinion mining, it identifies emotions, attitudes, and opinions expressed toward entities such as products or services. In this study, the entities are online news apps. Sources for sentiment analysis usually come from reviews, comments, or posts. With this technique, researchers can detect sentiment tendencies positive or negative (Ernianti Hasibuan & Elmo Allistair Heriyanto, 2022).

In recent years, sentiment analysis has grown essential for understanding user perceptions of digital applications. It classifies reviews into categories and provides insights for improvement. Its economic value is evident, as around 20–30 U.S. companies specialize in this service (Nurtikasari, Syariful Alam, & Teguh Iman Hermanto, 2022).

Previous studies show the effectiveness of sentiment analysis in online news media. For example, research on Tribunnews.com reviews applied Naive Bayes with Information Gain feature selection. Using 1,036 pre-processed and manually labeled reviews tested with K-Fold Cross Validation, the model achieved high performance: 96% accuracy, 97% precision, 98% recall, 88% specificity, and 0.93 AUC. Feature selection significantly boosted results (Samiaji, Hananto, & Kom, 2022).

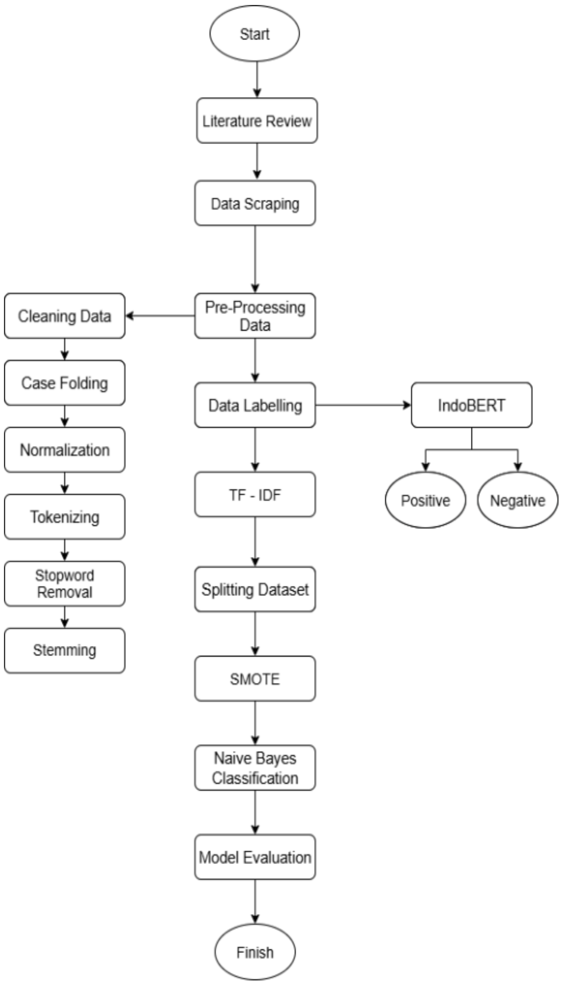
This study analyzes the sentiment of Mobile JKN application reviews on Google PlayStore using the IndoBERT model. From 5,404 reviews, IndoBERT successfully classified sentiments (positive, negative, neutral) with very high performance, achieving an average accuracy of 97.28%. This result demonstrates that IndoBERT is far superior to traditional methods like Naïve Bayes and SVM, proving its strong capability in understanding Indonesian language context. A word cloud analysis also provided insights into the dominant keywords for each sentiment, which can help Mobile JKN developers improve service quality(Tarwoto, Nugroho, Azka, & Graha, 2025).

Previous research on sentiment analysis of news application reviews, for example on Tribunnews.com, utilized the Naive Bayes algorithm with Information Gain feature selection and resampling techniques, which resulted in very high performance with 96% accuracy and significant improvement in other evaluation metrics. Another study used IndoBERT on 5,404 Mobile JKN application reviews and showed an average accuracy of 97.28%, confirming the superiority of deep learning-based models in understanding Indonesian language context compared to traditional methods.

This was also complemented by a word cloud analysis to reveal dominant keywords for each sentiment category. This current study takes a broader scope, covering three major news applications (Detikcom, Kompas.id, and CNN Indonesia) with review data from the Google Play Store labeled using IndoBERT. The focus of the study is not only on model performance but also on analyzing user perceptions, sentiment distribution, and the use of Generative AI to accelerate the identification of patterns and key issues. With this approach, the study provides a more contextual contribution, not only to the technical aspects of the analysis but also to understanding the quality of service across different applications and its implications for the evaluation and development of digital services.

MATERIALS AND METHODS

This research applies a quantitative approach with an experimental design to conduct sentiment analysis on user reviews of digital news applications. The research process consists of several stages, as illustrated in Figure1



Source : (Research result, 2025)
Figure 1. Research Methods

The research process began with a literature study, followed by data collection through web scraping and pre-processing to prepare the data for analysis. The cleaned data were then labeled using IndoBERT, which classified the reviews into positive and negative sentiments. Next, the text was converted into numerical form using TF-IDF, and the dataset was split into training and testing sets. To address class imbalance, the SMOTE method was applied. Finally, a Naive Bayes classification model was built, chosen because it is well-suited for text classification, lightweight in computation, fairly accurate, and commonly used as a baseline in initial sentiment analysis. Its performance was then evaluated using accuracy, precision, recall, and F1-score metrics, complemented by user perception evaluation with Generative AI.

Literature Review

In the literature study, the researcher searched for and obtained information from scientific journals using one of the journal indexing applications, namely Publish or Perish. This process was carried out to understand the latest developments in the field of study, to identify research gaps that have not been widely addressed, and to compare various perspectives from previous researchers.

Data Scraping

This research begins with the collection of primary data from the Google Play Store. The research objects are three leading digital news applications in Indonesia, namely Detikcom, CNN Indonesia, and Kompas.id. The data collection process uses a web scraping technique with the Python library google-play-scraper. Data retrieval was configured for reviews in the Indonesian language (lang='id') within the Indonesian region (country='id'), sorted by the newest reviews (sort=Sort.NEWEST). The time range for the reviews was limited to between January 1, 2022, and May 12, 2025. To ensure an efficient process, the scraping was performed asynchronously. The successfully collected data, which includes review IDs, usernames, ratings, review texts, and dates, were then exported into CSV file format. The dataset from the three applications, ready for analysis, totals 5,113 unique reviews.

Pre-processing Data

Following the scraping stage, the pre-processing stage is conducted to clean and standardize the raw text data. This process is crucial for eliminating noise that can degrade the model's performance.

1. Data Cleaning

The first step is to clean the text of irrelevant elements. This process includes the removal of URLs, HTML tags, emojis, non-alphanumeric symbols, and numbers, as well as trimming excess whitespace. At this stage, a data reduction occurred where the reviews for Kompas.id decreased from 1,039 to 801, Detikcom from 3,669 to 2,762, and CNN Indonesia from 2,211 to 1,656.

2. Case Folding
   All text in the reviews is converted to lowercase to standardize the corpus.

3. Normalization
   Non-standard words, abbreviations, and slang terms are normalized to standard words according to the KBBI (Official Dictionary of the Indonesian Language). This process utilizes the gemini-2.5-flash-preview-05-20 generative model from Google. Data is sent in batches to the API with specific instructions to convert words to their standard form, translate foreign or regional languages, and expand abbreviations without changing the original meaning of the sentence.

4. Tokenizing
   Each sentence in the reviews is split into a list of individual words (tokens) using the split() function.
5. Stopword Removal
   Common words that do not carry sentiment (e.g., 'yang', 'di', 'dan') are removed using the standard Indonesian stopword list from the NLTK library.

6. Stemming
   Each word is reduced to its base form using a stemmer from the Sastrawi library. After this process, data rows where the text has become empty are deleted to ensure quality. This reduces the final amount of data ready for labeling to 778 for Kompas.id, 2,713 for Detikcom, and 1,622 for CNN Indonesia.

Data Labelling

After the pre-processing stage, sentiment labeling is automatically performed using the *indonesia-bert-sentiment-classification* model from Hugging Face, which is a version of IndoBERT fine-tuned on Indonesian-language datasets such as the *Prosa Sentiment Dataset*. Through the *text-classification pipeline* in the *transformers library*, text is tokenized into vector representations and classified into two sentiment categories (positive and negative) based on the highest probability, thereby enabling large-scale data labeling that is efficient, consistent, and accurate.

Term Frequency-Inverse Document Frequency (TF-IDF)

The *Term Frequency–Inverse Document Frequency* (TF-IDF) method is applied to transform user reviews into vector representations by weighting words according to their frequency and importance. Implementation with TfidfVectorizer in *scikit-learn* employs key parameters *norm='l2'*, *ngram_range=(1,2)*, and .smooth_idf=True to prevent division by zero errors during IDF weight calculation.

1. Term Frequency (TF)
   is used to measure the frequency of a word's occurrence in a document by comparing the number of times the word appears with the total number of words in the document (Septiani & Isabela, 2022).

$$Tf\ (t,d) = \text{the number of occurences term t in document d} \quad (1)$$

Description :
t : Term or word frequency
d : Documents

2. Inverse Document Frequency (IDF)
   Inverse Document Frequency (IDF) assigns higher weights to words that appear less frequently across all documents. Its value is calculated by comparing the total number of documents with the number of documents containing the word, followed by applying a logarithmic function (Septiani & Isabela, 2022).

$$idf = In\left(\frac{N+1}{df+1}\right) + 1 \quad (2)$$

Description :
In : Natural Logarithm
N : Total Number of Documents
Df : Number of Documents Containing the Term / Word

3. Term Frequency-Inverse Document Frequency (TF-IDF)
   The final weight of a word is obtained by multiplying its TF and IDF values, which indicates the degree of its importance within the document (Br Sinulingga & Sitorus, 2024).

$$TF - IDF = TF\ x\ IDF \quad (3)$$

Description :
TF : Result of Term Frequency calculation
IDF : Result of Inverse Document Frequency calculation

4. L2 Normalization
   L2 normalization is applied to normalize TF-IDF vectors so that each document has the same length. This ensures that document comparisons are more consistent and are not influenced by text length (Haas, Yolland, & Rabus, 2022).

$$V_{Norm} = \frac{v}{||v||2} = \frac{v}{\sqrt{v_1^2 + +v_2^2 + \cdots + v_n^2}} \quad (4)$$

Description :

v : The original vector (unnormalized) obtained from TF-IDF.

||v||2 : The length or L2 norm of the TF-IDF vector, calculated as the square root of the sum of the squares of all elements in the vector.

Splitting Data

The dataset was split into 80% training data and 20% testing data using the train-test split method. The training set was used to build the model, while the testing set was applied to assess its performance on unseen data, ensuring that the model is not only effective on familiar data but also reliable in making predictions(Tanggraeni & Sitokdana, 2022).

The vectorized dataset is divided into 80% training data and 20% testing data using the train_test_split function from scikit-learn. The random_state=42 parameter ensures reproducible results, while stratify=y preserves the class distribution of sentiment labels in both training and testing sets. From the splitting process, the Kompas dataset produces 622 training data and 156 testing data, the Detik dataset generates 2,170 training data and 543 testing data, and the CNN dataset results in 1,297 training data and 325 testing data..

Synthetic Minority Over-Sampling Technique (SMOTE)

At this stage, SMOTE is applied to handle class imbalance by adding synthetic data to the minority class. This technique generates new samples from combinations of existing data, thereby creating a more balanced class distribution and enabling the model to learn more effectively (Fatkhudin, Artanto, & Safli, 2024).

To address the issue of imbalanced class distribution in the training data, an oversampling technique is applied using the SMOTE class from the imbalanced-learn library. This method is applied only to the training data to create synthetic samples for the minority class until it is balanced with the majority class. The random_state=42 parameter is also used in this process to ensure consistency in the creation of synthetic samples.

Naïve Bayes Classification

This process applies the Naive Bayes Classifier, consisting of two phases: training and classification. In the training phase, sample data are used to compute the prior probability of each class, while in the classification phase, the algorithm predicts new data by calculating term frequencies in the text(Rina Noviana & Isram Rasal, 2023).

The specific variant used is MultinomialNB from scikit-learn, initialized with default settings (alpha=1.0) for Laplace smoothing to handle unseen words in the training data. The model was tested on separate test sets: 156 reviews for Kompas.id, 543 for Detikcom, and 325 for CNN Indonesia.

Model Evaluation

The model evaluation was conducted using the Confusion Matrix to assess the extent to which the model could perform the tasks with sufficient accuracy and efficiency. Additionally, accuracy, precision, recall, and F1-score metrics were applied to provide a comprehensive overview of the model's performance.

In the process, the Confusion Matrix is a table used to illustrate the performance of a classification model by showing the number of correct and incorrect predictions on the test data. This table presents a comparison between the actual labels and the model's predicted results, making it easier to evaluate how accurately an algorithm classifies data (Normawati & Prayogi, 2021).

Table 1. Confusion Matrix

Confusion Matrix		Class Prediction	
		Positive	*Negative*
True Class	*Positive*	*TP*	*FN*
	Negative	*FP*	*TN*

Source: (Normawati & Prayogi, 2021)

Description :

1. True Positive (TP)
   True Positive is the number of positive documents that are correctly classified as positive class.
2. True Negative (TN)
   True Negative is the number of negative documents that are correctly classified as negative class.
3. False Positive (FP)
   False Positive is the number of negative documents that are incorrectly classified as positive class.
4. False Negative (FN)
   False Negative is the number of positive documents that are incorrectly classified as negative class

RESULTS AND DISCUSSION

Data Collecting

The research gathered user reviews from three news applications on the Google Play Store Kompas.id, Detikcom, and CNN Indonesia employing a web scraping approach using Python in Google Colab. The collected dataset, covering the

Techno Nusa Mandiri: Journal of Computing and Information Technology
Vol. 22, No. 2 September 2025 | DOI https://doi.org/10.33480/techno.v22i2.7356

period from January 1, 2022, to May 12, 2025, was stored in Excel format, comprising 1,039 reviews from Kompas.id, 3,669 reviews from Detikcom, and 2,211 reviews from CNN Indonesia.

Implementation Pre-Processing Data

The data obtained through the scraping process were raw and thus required preprocessing to ensure analytical suitability. In this study, the preprocessing steps included data cleaning, case folding, normalization, tokenization, stopword removal, and stemming. Three data samples were utilized to illustrate these procedures.

Table 2. Result Pre-Processing (using Indonesian)

Stage	Text 1	Text 2	Text 3
Review Text	Bagus sekali. Andai ada paket murah untuk guru honorer.	support ♡	lansia sangat membutuhkan berita terimakasih Kompas
Cleaning	Bagus sekali Andai ada paket murah untuk guru honorer	support	lansia sangat membutuhkan berita terimakasih Kompas
Case Folding	bagus sekali andai ada paket murah untuk guru honorer	support	lansia sangat membutuhkan berita terimakasih kompas
Normalization	bagus sekali andai ada paket murah untuk guru honorer	dukungan	lansia sangat membutuhkan berita terima kasih kompas
Tokenizing	['bagus', 'sekali', 'andai', 'ada', 'paket', 'murah', 'untuk', 'guru', 'honorer']	['dukungan']	['lansia', 'sangat', 'membutuhkan', 'berita', 'terima', 'kasih', 'kompas']
Stopword Removal	['bagus', 'andai', 'paket', 'murah', 'guru', 'honorer']	['dukungan']	['lansia', 'membutuhkan', 'berita', 'terima', 'kasih', 'kompas']
Stemming	bagus andai paket murah guru honorer	dukung	lansia butuh berita terima kasih kompas

Source : (Research result, 2025)

Labelling Sentiment

Sentiment labeling was carried out on three preprocessed review samples using the IndoBERT model. Each review was assigned a sentiment label along with a varying confidence score, reflecting the model's certainty and the reliability of the sentiment classification for the analyzed reviews.

steming_data	Sentiment	Akurasi
bagus andai paket murah guru honorer	positive	0.733462
dukung	positive	0.962695
lansia butuh berita terima kasih kompas	positive	0.988291

Source : (Research result, 2025)

Figure 2. Labelling Result

Calculation TF-IDF

The figure below presents the list of words appearing in a document at index 0 from the Kompas.id application, along with detailed information for each word.

```
==================================================
     Analisis untuk Dokumen ke-1 (Index: 0)
     Teks Asli: bagus andai paket murah guru honorer
==================================================
     Kata        TF       IDF  TF-IDF (No Norm)  TF-IDF (L2 Norm)
    andai         1  6.559399          6.559399          0.430094
    bagus         1  2.730757          2.730757          0.179053
     guru         1  6.964864          6.964864          0.456680
  honorer         1  6.964864          6.964864          0.456680
    murah         1  6.048573          6.048573          0.396599
    paket         1  6.964864          6.964864          0.456680
```

Source : (Research result, 2025)

Figure 3. TF-IDF Result

The sentence was extracted from document 1: "Bagus Andai Paket Murah Guru Honorer." The TF (Term Frequency) column indicates the frequency of each word in the document, while the IDF (Inverse Document Frequency) column reflects the importance of the word across documents, lower IDF values indicate words that appear more frequently. TF-IDF (No Norm) represents the product of TF and IDF without normalization, whereas TF-IDF (L2 Norm) is the normalized version, allowing for balanced comparison of word weights.

1. Calculation Term Frequency (TF) Kompas.id

Table 3. Term Frequency Result (Using Indonesian)

Word	Calculation Results
Andai	1
Bagus	1
Guru	1
Honorer	1
Murah	1
Paket	1

Source : (Research result, 2025)

This table displays the frequency of each word across the document collection from Kompas.id. The "Word" column lists the analyzed words, while the "Frequency" column indicates the number of occurrences for each word. For example, the words "Andai," "Bagus," "Guru," "Honorer," "Murah," and "Paket" each appear once. Term Frequency (TF) is used to measure how often a word occurs within a document. kata tersebut muncul dalam dokumen tertentu.

2. Calculation Inverse Document Frequency (IDF) Kompas.id

Table 4. Inverse Document Frequency Result (Using Indonesian)

Word	Calculation Results
Andai	$\ln\left(\frac{778+1}{2+1}\right)+1 = 6{,}559$
Bagus	$\ln\left(\frac{778+1}{137+1}\right)+1 = 2{,}730$
Guru	$\ln\left(\frac{778+1}{1+1}\right)+1 = 6{,}964$
Honorer	$\ln\left(\frac{778+1}{1+1}\right)+1 = 6{,}964$
Murah	$\ln\left(\frac{778+1}{4+1}\right)+1 = 6{,}048$
Paket	$\ln\left(\frac{778+1}{1+1}\right)+1 = 6{,}964$

Source : (Research result, 2025)

This table presents the IDF (Inverse Document Frequency) calculations for the Kompas.id application. IDF measures the importance of a word within the document collection, lower values indicate that the word appears more frequently across documents. The "Word" column lists the analyzed words, while the "Calculation Result" column shows the applied IDF formula. For example, the word "Bagus" appears in many documents, resulting in a lower IDF value (2.730), indicating that it is a common term.

3. Calculation TF-IDF Kompas.id

Table 5. TF-IDF Result (using Indonesian)

Word	Calculation Results
Andai	1x6,559 = 6,559
Bagus	1x2,730 = 2,730
Guru	1x6,964 = 6,964
Honorer	1x6,964 = 6,964
Murah	1x6,048 = 6,048
Paket	1x6,964 = 6,964

Source : (Research result, 2025)

This table presents the TF-IDF results for the Kompas.id application, obtained by multiplying the TF (term frequency within a document) by the IDF (word specificity across the entire document collection). The purpose is to determine the significance of a word in a particular document relative to the entire corpus. For example, the word "Bagus" has a TF-IDF value of 2.730, indicating that it is a common term frequently appearing in the documents.

4. Calculation TF-IDF L2 Normalization Kompas.id

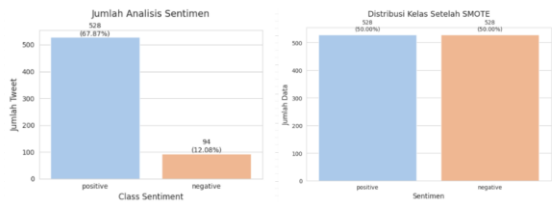
Table 6. TF-IDF L2 Result (Using Indonesian)

Word	Calculation Results
Andai	1x6,559 = 6,559
Bagus	1x2,730 = 2,730
Guru	1x6,964 = 6,964
Honorer	1x6,964 = 6,964
Murah	1x6,048 = 6,048
Paket	1x6,964 = 6,964

Source : (Research result, 2025)

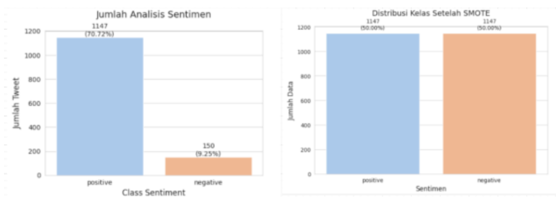
This table presents the L2-normalized TF-IDF results for the Kompas.id application. The L2 value is obtained by dividing each TF-IDF score by the vector magnitude of all words in the document. For example, a word like "Guru," which has a high TF-IDF due to its rarity in other documents, retains a relatively high value after L2 normalization, whereas common words such as "Bagus" see their TF-IDF values reduced after normalization.

Implementation Synthetic Minority Over-sampling Technique (SMOTE)



Source : (Research result, 2025)
Figure 4. SMOTE Result Kompas.id

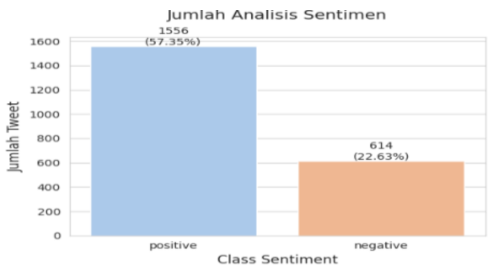
SMOTE was applied because the sentiment data from Kompas.id was imbalanced, with 528 positive samples (67.87%) and only 94 negative samples (12.08%). This imbalance could cause the model to be biased toward the majority class. After applying SMOTE, both classes were balanced to 528 samples each (50%), allowing the model to learn evenly and improving prediction accuracy.



Source : (Research result, 2025)
Figure 5. SMOTE Result CNN Indonesia

Vol. 22, No. 2 September 2025 | DOI https://doi.org/10.33480/techno.v22i2.7356

The sentiment data from CNN Indonesia exhibited a significant class imbalance, with 1,147 positive samples (70.72%) and only 150 negative samples (9.25%). This imbalance could bias the model toward predicting the majority class. To address this, SMOTE was applied, resulting in a balanced distribution of 1,147 samples (50%) for both positive and negative classes, enabling the model to learn fairly and improving prediction performance.



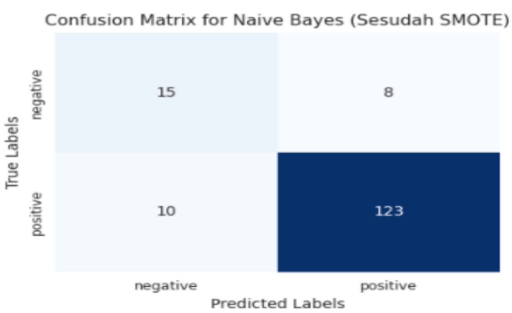
Source : (Research result, 2025)
Figure 6. Non-SMOTE Result Detikcom

The sentiment data from Detikcom showed 1,556 positive samples (57.35%) and 614 negative samples (22.63%), indicating some class imbalance, though not substantial. Therefore, SMOTE was not applied, as the proportional difference was still tolerable, allowing the model to learn effectively without synthetic data balancing.

Testing Model classification Naïve Bayes

The Naïve Bayes classification model was tested on user review data from three news applications: Kompas.id, Detikcom, and CNN Indonesia. The purpose of this testing was to evaluate how accurately the model could classify the sentiment of reviews based on the text representations processed in the previous stage.

1. Confusion Matrix Kompas.id



Source : (Research result, 2025)
Figure 7. Confusion Matrix Result Kompas.id

$$\text{Accuracy} = \frac{123+15}{156} = 0{,}88$$

$$\text{Precision (Positif)} = \frac{123}{123+8} = 0{,}94$$

$$\text{Precision (Negatif)} = \frac{15}{15+10} = 0{,}60$$

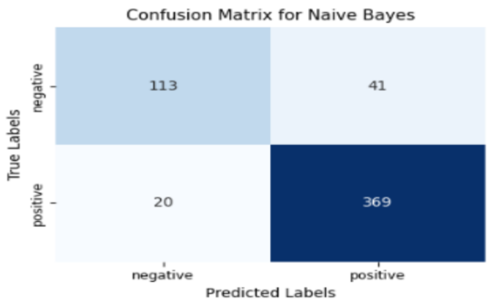
$$\text{Recall (Positif)} = \frac{123}{123+10} = 0{,}92$$

$$\text{Recall (Negatif)} = \frac{15}{15+8} = 0{,}65$$

$$\text{F1 – Score (Positif)} = 2 \text{ x} \frac{0{,}94 \text{ x } 0{,}92}{0{,}94+0{,}92} = 0{,}93$$

$$\text{F1 – Score (Negatif)} = 2 \text{ x} \frac{0{,}60 \text{ x } 0{,}65}{0{,}60+0{,}65} = 0{,}62$$

2. Confusion Matrix Detikcom



Source : (Research result, 2025)
Figure 8. Confusion Matrix Detikcom Result

$$\text{Accuracy} = \frac{369+113}{543} = 0{,}89$$

$$\text{Precision (Positif)} = \frac{369}{369+41} = 0{,}90$$

$$\text{Precision (Negatif)} = \frac{113}{113+20} = 0{,}85$$

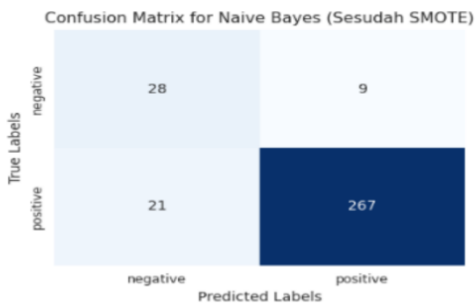
$$\text{Recall (Positif)} = \frac{369}{369+20} = 0{,}95$$

$$\text{Recall (Negatif)} = \frac{113}{113+41} = 0{,}73$$

$$\text{F1 – Score (Positif)} = 2 \text{ x} \frac{0{,}90 \text{ x } 0{,}95}{0{,}90+0{,}95} = 0{,}92$$

$$\text{F1 – Score (Negatif)} = 2 \text{ x} \frac{0{,}85 \text{ x } 0{,}73}{0{,}85+0{,}73} = 0{,}79$$

3. Confusion Matrix Aplikasi CNN Indonesia



Source : (Research result, 2025)
Figure 9. Confusion Matrix CNN Indonesia Result

$$\text{Accuracy} = \frac{267 + 28}{325} = 0{,}91$$

$$\text{Precision (Positif)} = \frac{267}{267 + 9} = 0{,}97$$

$$\text{Precision (Negatif)} = \frac{28}{28 + 21} = 0{,}57$$

$$\text{Recall (Positif)} = \frac{267}{267 + 21} = 0{,}93$$

$$\text{Recall (Negatif)} = \frac{28}{28 + 9} = 0{,}76$$

$$\text{F1 – Score (Positif)} = 2 \text{ x} \frac{0{,}97 \text{ x } 0{,}93}{0{,}97 + 0{,}93} = 0{,}92$$

$$\text{F1 – Score (Negatif)} = 2 \text{ x} \frac{0{,}57 \text{ x } 0{,}76}{0{,}57 + 0{,}76} = 0{,}65$$

Results Evaluation of Perception using Generative AI

1. Visualization Classification Sentiment

```
Classification report before SMOTE

Classifier: Naive Bayes
Accuracy: 0.8525641025641025
Classification Report:
              precision    recall  f1-score   support

    negative       0.00      0.00      0.00        23
    positive       0.85      1.00      0.92       133

    accuracy                           0.85       156
   macro avg       0.43      0.50      0.46       156
weighted avg       0.73      0.85      0.78       156
```

```
Classification report After SMOTE

Classifier: Naive Bayes (After SMOTE)
Accuracy: 0.8846153846153846
Classification Report:
              precision    recall  f1-score   support

    negative       0.60      0.65      0.62        23
    positive       0.94      0.92      0.93       133

    accuracy                           0.88       156
   macro avg       0.77      0.79      0.78       156
weighted avg       0.89      0.88      0.89       156
```

Source : (Research result, 2025)
Figure 10. Comparison Before & After SMOTE Kompas.id

Before applying SMOTE, the Naive Bayes model on Kompas.id reviews was biased toward the majority class (positive) with an accuracy of 85.3%, where the negative class was not detected at all. After applying SMOTE, the data distribution became balanced, enabling the model to recognize both classes more effectively. Accuracy increased to 88.5%, with improved performance on the negative class (precision 0.60, recall 0.65, f1-score 0.62), while the positive class remained strong (precision 0.94, recall 0.92, f1-score 0.93). The improvement in macro average and weighted average indicates more balanced classification, showing that SMOTE not only increased accuracy but also enhanced the model's ability to handle data imbalance.

```
Classifier: Naive Bayes
Accuracy: 0.8876611418047882
Classification Report:
              precision    recall  f1-score   support

    negative       0.85      0.73      0.79       154
    positive       0.90      0.95      0.92       389

    accuracy                           0.89       543
   macro avg       0.87      0.84      0.86       543
weighted avg       0.89      0.89      0.89       543
```

Source : (Research result, 2025)
Figure 11. Result Classification Report Non-SMOTE Detikcom

For the Naive Bayes classification model trained on data from the Detikcom application, with an overall accuracy of 88.77%. This report details the performance for the 'negative' and 'positive' classes using precision, recall, and f1-score metrics. Although the data shows an imbalance (389 'positive' samples versus 154 'negative' ones), oversampling techniques were likely not used for several reasons. First, the imbalance is not considered extreme, and the model's performance on the majority class ('positive') is already very good with a recall of 0.95. Second, developers may have wanted to avoid the risk of overfitting that can occur when creating synthetic data. Finally, if the current model performance already meets the target, adding oversampling would only increase complexity without providing significant benefits to the Detikcom application.

```
Classification report before SMOTE

Classifier: Naive Bayes
Accuracy: 0.9076923076923077
Classification Report:
              precision    recall  f1-score   support

    negative       1.00      0.19      0.32        37
    positive       0.91      1.00      0.95       288

    accuracy                           0.91       325
   macro avg       0.95      0.59      0.63       325
weighted avg       0.92      0.91      0.88       325
```

```
Classification report After SMOTE

Classifier: Naive Bayes (After SMOTE)
Accuracy: 0.9076923076923077
Classification Report:
              precision    recall  f1-score   support

    negative       0.57      0.76      0.65        37
    positive       0.97      0.93      0.95       288

    accuracy                           0.91       325
   macro avg       0.77      0.84      0.80       325
weighted avg       0.92      0.91      0.91       325
```

Source : (Research result, 2025)
Figure 12. Comparison Before & After SMOTE CNN Indonesia

On CNN Indonesia reviews, the Naive Bayes model reached 90.7% accuracy but was biased toward positives, with the negative class showing very low recall (0.19). After SMOTE, accuracy remained 90.7%, but negative class performance improved significantly (precision 0.57, recall 0.76, f1-score 0.65), while the positive class stayed strong. This shows SMOTE balanced the model, maintaining accuracy while enhancing minority class detection. maintained accuracy but also enhanced the model's ability to detect the minority class effectively.

2. Generative AI

The results generated using the Gemini API for three news applications, namely Kompas.id, Detikcom, and CNN Indonesia, present an evaluation of user perceptions covering the level of satisfaction, strengths, and weaknesses of each application as follows :

```
======================== HASIL EVALUASI PERSEPSI PENGGUNA ========================

1.  Secara keseluruhan, persepsi dan kepuasan pengguna terhadap aplikasi berita ini menunjukkan gambaran yang bervariasi
    namun cenderung positif, didominasi oleh apresiasi terhadap kualitas konten dan kredibilitas berita yang disajikan.
    Banyak pengguna sangat menghargai berita yang aktual, akurat, mutakhir, informatif, tajam, independen, dan
    terpercaya, seringkali menyebut Kompas sebagai sumber informasi utama untuk menambah wawasan dan pemahaman. Aplikasi
    ini juga banyak dipuji karena antarmuka yang bagus, mudah digunakan, ringan, serta fitur-fitur seperti teks suara
    dan mode malam yang meningkatkan pengalaman membaca, menjadikannya alternatif yang nyaman dan efektif dibandingkan
    koran fisik. Namun, terdapat sejumlah besar pengguna yang menyatakan ketidakpuasan, terutama terkait sistem
    langganan yang dianggap mahal, proses pendaftaran atau aktivasi langganan yang rumit dan sering bermasalah,
    kesulitan akses setelah berlangganan, serta adanya iklan yang mengganggu meskipun sudah berlangganan. Beberapa
    keluhan teknis juga muncul, seperti aplikasi yang lambat, gagal pasang, atau masalah tampilan teks yang tertutup,
    yang mengurangi kenyamanan penggunaan. Selain itu, beberapa ulasan negatif juga menyoroti isu netralitas berita atau
    tuduhan penggiringan opini publik yang berujung pada rasa kecewa. Jadi, meskipun inti produk berupa kualitas berita
    sangat dihargai, masalah pada aspek monetisasi (langganan) dan kinerja teknis aplikasi menjadi penyebab utama
    ketidakpuasan pengguna.

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

Source : (Research result, 2025)
Figure 13. Generative AI Kompas.id Result

It shows that, in general, the Kompas news application is considered positive because it is able to present news that is up-to-date, accurate, informative, independent, and reliable. The diverse content is seen as helpful for users in obtaining relevant information while also broadening their insights, supported by an application interface that is easy to use, lightweight, and equipped with a text-to-speech feature that facilitates access to news without the need for direct reading. However, users also highlight weaknesses such as technical issues including slow performance, frequent crashes, obscured text, prolonged loading times, as well as excessive advertisements that reduce comfort. Thus, although the quality of the news content receives appreciation, technical aspects and monetization remain the main causes of user dissatisfaction

```
======================== HASIL EVALUASI PERSEPSI PENGGUNA ========================

1. Evaluasi persepsi dan kepuasan pengguna secara keseluruhan berdasarkan ulasan menunjukkan
   gambaran yang terbagi. Di satu sisi, banyak pengguna sangat menghargai Detikcom sebagai sumber
   berita yang cepat, akurat, informatif, dan terpercaya. Mereka memuji kelengkapan dan relevansi
   berita, serta kemudahan aplikasi dalam menyajikan informasi terkini yang penting untuk menambah
   wawasan. Pengguna yang puas sering menyebutkan tampilan yang rapi, kemudahan akses, dan kinerja
   aplikasi yang bagus secara keseluruhan, merasa terbantu dan menganggapnya sebagai aplikasi
   berita yang mantap dan oke. Namun, di sisi lain, terdapat tingkat ketidakpuasan yang sangat
   tinggi dan meluas, terutama dipicu oleh masalah iklan yang berlebihan dan mengganggu. Banyak
   ulasan negatif secara spesifik mengeluhkan iklan popup yang menutupi layar, iklan yang secara
   otomatis mengarahkan ke toko daring atau situs judi, serta video iklan yang memutar otomatis dan
   menguras kuota data. Selain itu, versi aplikasi yang baru atau perubahan tampilan seringkali
   dikritik karena dianggap rumit, membingungkan, lambat, dan rentan galat atau crash. Masalah
   terkait akses komentar yang sulit, keharusan untuk login berulang kali, serta hilangnya beberapa
   fitur yang disukai (seperti jadwal olahraga), juga berkontribusi pada pengalaman pengguna yang
   negatif. Meskipun konten berita seringkali dipandang positif, masalah teknis dan gangguan iklan
   ini secara signifikan mengurangi kenyamanan dan kepuasan pengguna, bahkan mendorong banyak dari
   mereka untuk menghapus aplikasi.

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

Source : (Research result, 2025)

Figure 14. Generative AI Detikcom Result

It shows that the Detikcom news application is generally considered positive as it is able to provide fast, accurate, informative, and reliable news, with a neat interface, user-friendly design, and relevant content. Nevertheless, users also highlight weaknesses such as disturbing pop-up ads, slow or crashing performance, and difficulties in accessing certain content when the network is unstable. Some features, such as the sports schedule, are also deemed less optimal. Thus, although the quality of the content is appreciated, technical aspects and excessive advertising remain the main factors of user dissatisfaction.

```
======================== HASIL EVALUASI PERSEPSI PENGGUNA ========================

1. Evaluasi persepsi dan kepuasan pengguna secara keseluruhan menunjukkan bahwa sebagian besar
   pengguna merasa sangat puas dengan aplikasi berita ini, menganggapnya bagus dan mantap karena
   menyajikan berita yang cepat, akurat, aktual, dan mudah dibaca. Banyak pengguna menghargai
   aplikasi ini sebagai sumber informasi yang terpercaya, lengkap, dan informatif, serta merasa
   terbantu untuk mendapatkan wawasan terkini baik nasional maupun internasional. Namun, keluhan
   yang paling menonjol dan berulang kali muncul adalah mengenai banyaknya iklan yang mengganggu
   pengalaman membaca, mulai dari iklan yang otomatis muncul, iklan layar penuh, hingga iklan judi
   daring yang dianggap tidak pantas, yang pada akhirnya mengurangi kenyamanan dan bahkan membuat
   sebagian pengguna mempertimbangkan untuk menghapus aplikasi. Selain masalah iklan, ada juga
   beberapa masukan minor terkait performa aplikasi yang terkadang lambat, pemberitaan yang dirasa
   kurang netral, adanya hoaks sesekali, atau judul berita yang kurang sesuai dengan isinya, tetapi
   secara keseluruhan, kepuasan terhadap kualitas berita dan kemudahan akses cukup tinggi, terlepas
   dari hambatan iklan tersebut.

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

Source : (Research result, 2025)

Figure 15. Generative AI CNN Indonesia Result

It shows that the CNN Indonesia application is considered positive for delivering fast, accurate, easy-to-read news and being trusted as a comprehensive source of information. Users find it helpful in gaining both national and international insights. However, the main complaint is the excessive presence of disruptive ads, ranging from automatic pop-ups and full-screen ads to online gambling ads that are deemed inappropriate. In addition, there is feedback regarding the app's occasional slow performance, news perceived as less neutral, and headlines that do not match the content. Overall, the quality of the content and ease of access are appreciated, although excessive advertising remains the main factor of user dissatisfaction

The comparison of user perceptions across the three news applications shows clear differences. CNN Indonesia is considered the most satisfying because it delivers fast, accurate, and easily accessible news, although advertisements remain a major complaint. Detikcom is also highly appreciated for its clean interface, quick news access, and stable performance, even though excessive ads continue to disturb users. Meanwhile, Kompas.id stands out for the quality and credibility of its news but faces more criticism due to its expensive and complicated subscription system as well as technical issues within the app. This comparison highlights that beyond content quality, technical performance, ease of access, and user comfort play a more dominant role in shaping positive user perceptions.

3. Research Comparison

This study analyzes sentiment in news applications by comparing Naive Bayes as a baseline with IndoBERT. While previous works mostly relied on Naive Bayes, results highlight IndoBERT's superior ability to capture Indonesian linguistic nuances. Furthermore, the main innovation of this study lies in the integration of Generative AI for two critical purposes: automatic text normalization during the preprocessing stage and summarizing research findings into an easily understandable narrative. The use of generative AI not only improves the quality of input data but also provides richer and more in-depth interpretations compared to merely presenting statistical metrics. Thus, this study not only proves IndoBERT's superiority but also presents a comprehensive framework that integrates various AI technologies to produce more accurate and insightful sentiment analysis.

CONCLUSION

This study evaluated user perceptions of the digital news applications Detikcom, Kompas.id, and CNN Indonesia through sentiment analysis using the Naive Bayes method labeled with IndoBERT.

The findings indicate that positive reviews were more dominant, although each application carried distinct user perceptions. Detikcom received 89.3% positive reviews and was appreciated for its informative and up-to-date news content but was widely criticized for excessive advertisements that disrupted user convenience. Kompas.id obtained 88.5% positive reviews, praised for its credible brand image, independent and reliable news quality, and valuable subscription features, although some users considered the subscription costs expensive and the process complicated. CNN Indonesia achieved 90.8% positive reviews, recognized for its speed in delivering accurate and timely news, although several technical issues were noted by users.

The application of Generative AI proved effective in identifying sentiment patterns and key issues from thousands of reviews efficiently and objectively. Future research is recommended to expand the use of other classification models such as Random Forest, SVM, Transformer or Deep Learning and to enrich data sources beyond the Google Play Store, including the App Store. From a practical perspective, the findings provide valuable input for developers to reassess the frequency, type, and placement of advertisements, enhance user experience, and become more responsive to user comments and reviews on application platforms.

REFERENCE

Arlovin, T., Kusrini, & Kusnawi. (2024). Analisis Sentimen Review Pengguna Aplikasi Fizzo Novel Di Google Play Menggunakan Algoritma Naive Bayes. *Jurnal Informatika Teknologi Dan Sains (Jinteks)*, *6*(1), 65–70. https://doi.org/10.51401/jinteks.v6i1.3909

Br Sinulingga, J. E., & Sitorus, H. C. K. (2024). Analisis Sentimen Opini Masyarakat terhadap Film Horor Indonesia Menggunakan Metode SVM dan TF-IDF. *Jurnal Manajemen Informatika (JAMIKA)*, *14*(1), 42–53. https://doi.org/10.34010/jamika.v14i1.11946

Ernianti Hasibuan, & Elmo Allistair Heriyanto. (2022). Analisis Sentimen Pada Ulasan Aplikasi Amazon Shopping Di Google Play Store Menggunakan Naive Bayes Classifier. *Jurnal Teknik Dan Science*, *1*(3), 13–24. https://doi.org/10.56127/jts.v1i3.434

Fatkhudin, A., Artanto, F. A., & Safli, N. A. (2024). Decision Tree Berbasis SMOTE Dalam Analisis Sentimen Penggunaan Artificial Intelligence Untuk Skripsi. *REMIK: Riset Dan E ...*, *8*(April), 494–505. Retrieved from https://www.jurnal.polgan.ac.id/index.php/remik/article/view/13531%0Ahttps://www.jurnal.polgan.ac.id/index.php/remik/article/download/13531/2453

Haas, J., Yolland, W., & Rabus, B. (2022). *Inducing Early Neural Collapse in Deep Neural Networks for Improved Out-of-Distribution Detection*. 1–19. Retrieved from http://arxiv.org/abs/2209.08378

Hapsari, S. K., & Priliantini, A. (2025). *PROSES PRODUKSI BERITA PADA LAMAN*. *7*(1), 57–73.

Kusnia, U., & Kurniawan, F. (2022). Analisis Sentimen Review Aplikasi Media Berita Online Pada Google Play menggunakan Metode Algoritma Support Vector Machines (SVM) Dan Naive Bayes INFO ARTIKEL ABSTRAK. *Jurnal Keilmuan Dan Aplikasi Teknik Informatika*, *14(1)*(36), 24–25. Retrieved from https://doi.org/10.35891/explorit

Larasati, F. A., Ratnawati, D. E., & Hanggara, B. T. (2022). Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *6*(9), 4305–4313.

Nadira, A., Setiawan, N. Y., & Purnomo, W. (2023). Analisis Sentimen Pada Ulasan Aplikasi Mobile Banking Menggunakan Metode Naïve Bayes Dengan Kamus Inset. *Indexia*, *5*(01), 35. https://doi.org/10.30587/indexia.v5i01.5138

Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. *Jurnal Sains Komputer & Informatika (J-SAKTI*, *5*(2), 697–711.

Nurtikasari, Y., Syariful Alam, & Teguh Iman Hermanto. (2022). Analisis Sentimen Opini Masyarakat Terhadap Film Pada Platform Twitter Menggunakan Algoritma Naive Bayes. *INSOLOGI: Jurnal Sains Dan Teknologi*, *1*(4), 411–423. https://doi.org/10.55123/insologi.v1i4.770

Rina Noviana, & Isram Rasal. (2023). Penerapan Algoritma Naive Bayes Dan Svm Untuk Analisis Sentimen Boy Band Bts Pada Media Sosial Twitter. *Jurnal Teknik Dan Science*, *2*(2), 51–60. https://doi.org/10.56127/jts.v2i2.791

Samiaji, A., Hananto, B., & Kom, S. (2022). Analisis Sentimen Review Aplikasi Berita Online Pada Google Play Menggunakan Metode Naïve Bayes Studi Kasus: Tribunnews.Com. *Seminar Nasional Mahasiswa Ilmu Komputer Dan Aplikasinya (SENAMIKA)*, (2), 733–743.

Septiani, D., & Isabela, I. (2022). Analisis Term Frequency Inverse Document Frequency (Tf-Idf) Dalam Temu Kembali Informasi Pada Dokumen Teks. *SINTESIA: Jurnal Sistem Dan Teknologi Informasi Indonesia*, *1*(1), 81–88.

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. https://doi.org/10.35957/jatisi.v9i2.1835

Tarwoto, Nugroho, R., Azka, N., & Graha, W. S. R. (2025). Analisis Sentimen Ulasan Aplikasi Mobile JKN di Google PlayStore Menggunakan IndoBERT. *Jurnal JTIK (Jurnal Teknologi Informasi Dan Komunikasi)*, *9*(2), 495–505. https://doi.org/10.35870/jtik.v9i2.3340