

Sentiment Classification of AI Photo Studio: Pro Editor User Reviews on Google Play Store Using Multinomial Naïve Bayes

Ranti Maharani Putri^A, Yulison H. Chrisnanto^B, Gunawan Abdillah^C

Abstract

This study analyzes the sentiment of Indonesian-language user reviews of the AI Photo Studio: Pro Editor application collected from the Google Play Store using Multinomial Naïve Bayes. A total of 2,028 unique reviews were obtained after duplicate removal, and 1,976 reviews remained after text preprocessing. The reviews were categorized into three sentiment classes—negative, positive, and neutral—using a lexicon-based labeling approach. The preprocessed reviews were represented using Term Frequency–Inverse Document Frequency (TF-IDF), divided into training and testing sets using an 80:20 stratified split, and balanced using the Synthetic Minority Over-sampling Technique (SMOTE) applied only to the training data. The Multinomial Naïve Bayes classifier was then trained and evaluated using accuracy, precision, recall, F1-score, and a confusion matrix. The final model achieved an accuracy of 75.00%. The negative class achieved a precision of 0.80, recall of 0.88, and F1-score of 0.84; the neutral class achieved a precision of 0.65, recall of 0.35, and F1-score of 0.46; while the positive class achieved a precision of 0.66, recall of 0.74, and F1-score of 0.70. The results indicate that the model performed best in classifying negative reviews, whereas the neutral class remained the most challenging category. These findings provide application-specific evidence of textual sentiment in AI-assisted photo-editing reviews and highlight the importance of interpreting written feedback alongside numerical ratings.

Keywords: Sentiment Analysis, Multinomial Naïve Bayes, Application Review, AI Photo Studio: Pro Editor, Google Play Store, Text Mining.

INTRODUCTION

The rapid development of digital technologies has transformed the way users interact with information, digital services, and creative applications. Within the mobile application ecosystem, user-generated reviews have become an important source of evidence for evaluating perceived usability, functionality, and satisfaction. Google Play Store facilitates this process by allowing users to provide both numerical ratings and written evaluations of applications (Bigne et al., 2021a; Nilashi et al., 2023). While star ratings offer a general indication of user satisfaction, textual reviews provide richer information because users can describe particular features they appreciate, difficulties they

^AUniversitas Jenderal Achmad Yani, Cimahi, Indonesia, Email: yhc@if.unjani.ac.id

^BUniversitas Jenderal Achmad Yani, Cimahi, Indonesia

^CUniversitas Jenderal Achmad Yani, Cimahi, Indonesia

encounter, and reasons underlying their evaluations. The analysis of textual feedback is therefore particularly relevant for applications whose performance depends on multiple functional and experiential attributes (Cooper, 2024; Egger et al., 2019).

AI Photo Studio: Pro Editor represents an increasingly relevant case because it is an artificial intelligence (AI)-based image editing application designed to assist users in creating, adding, removing, and modifying image elements. According to the research document, the application was released on February 18, 2023, and had exceeded 100,000 downloads with a reported rating of 4.5 during the period considered in the study. Its Google Play Store page provides a corpus of user-generated reviews that reflects different experiences with the application. The available reviews include favorable assessments of the application's editing capabilities as well as complaints concerning photo accessibility, feature complexity, and paid functionality. Such variation indicates that the overall rating may not adequately capture the specific dimensions of user experience expressed in the textual feedback.

Sentiment analysis provides a means of converting these unstructured reviews into interpretable categories (Kisilevich et al., 2010; Liang et al., 2024; Zhu et al., 2024). By identifying positive, negative, and neutral orientations, the method enables a large volume of user feedback to be analyzed systematically rather than relying on individual manual inspection. For application developers, this classification can facilitate the identification of prevailing user responses and reveal whether a favorable overall rating is accompanied by substantial negative or neutral expressions in the underlying text (Bigne et al., 2021b; Hu et al., 2020; Patel et al., 2020). The present study therefore treats textual sentiment as a distinct analytical object rather than assuming that star ratings can serve as a direct substitute for the sentiment expressed in written reviews.

Previous studies have demonstrated the applicability of Naïve Bayes for sentiment classification of mobile application reviews. However, the findings across application contexts vary according to the characteristics of the reviewed applications and their users. More importantly, application ratings and written reviews do not necessarily convey the same information. A high numerical rating may coexist with negative or neutral expressions in the accompanying text, while a single rating value cannot explain the specific experiences or concerns expressed by users. Therefore, textual sentiment should be examined as a distinct analytical dimension rather than treated as a direct substitute for star ratings.

The research gap is more specific than the absence of a study on one particular application. First, the existing studies identified in the source document evaluate different application domains, meaning their classification performance and sentiment patterns are derived from review corpora shaped by different types of user interaction. None of the cited studies directly examines AI Photo Studio: Pro Editor or an equivalent AI-assisted photo-editing application. This leaves an empirical

gap concerning whether the sentiment characteristics observed in other application categories also occur in user feedback for AI-assisted image editing. The distinction is important because users of AI photo-editing applications evaluate outputs and functions that differ from those encountered in e-commerce, social media, or transportation-related applications. Evidence obtained from other application domains cannot be directly interpreted as evidence about AI Photo Studio: Pro Editor.

Second, the study focuses on the discrepancy between numerical ratings and textual sentiment. After duplicate removal, the dataset contained 2,028 unique reviews, of which 1,682 received the maximum rating of five, indicating a highly favorable rating distribution. In contrast, after preprocessing and lexicon-based labeling, the 1,976 reviews were classified as 1,175 negative, 430 positive, and 371 neutral. This divergence indicates that aggregate ratings alone may not fully capture the sentiment expressed in users' written feedback. From an analytical perspective, examining textual sentiment can reveal dissatisfaction, mixed experiences, or neutral evaluations that may be obscured by a single numerical score. From a managerial perspective, this information can help application developers identify issues in user experience, prioritize feature improvements, and distinguish between generally favorable ratings and specific concerns expressed within individual reviews. This condition raises a substantive question that cannot be answered from star ratings alone: what sentiment patterns emerge when users' written expressions are analyzed independently of their numerical scores?

Third, the lexicon-based labeling produced an imbalanced three-class dataset, with negative reviews substantially outnumbering positive and neutral reviews. This imbalance creates a methodological challenge because a classifier may favor the majority class. Therefore, the study incorporates TF-IDF for textual representation and SMOTE to balance the training data before applying Multinomial Naïve Bayes. This design enables the classification performance to be examined under a three-class sentiment structure with substantially unequal class frequencies.

Based on these gaps, this study examines Indonesian-language reviews of AI Photo Studio: Pro Editor by combining lexicon-based sentiment labeling, TF-IDF representation, SMOTE-based class balancing, and Multinomial Naïve Bayes classification. The analysis focuses on the distribution of positive, negative, and neutral sentiment and evaluates classification performance using accuracy, precision, recall, F1-score, and a confusion matrix. By analyzing written reviews independently from numerical ratings, the study provides application-specific evidence of user sentiment and highlights information that may not be apparent from aggregate ratings alone.

LITERATURE REVIEW

Previous Studies

Research on sentiment analysis of user reviews has developed alongside the rapid growth of digital platforms and mobile applications.

A number of studies have examined user-generated reviews on Google Play Store using text classification techniques, particularly the Naïve Bayes algorithm. These studies provide relevant methodological references for identifying sentiment polarity, evaluating classifier performance, and interpreting user feedback. The following studies are particularly relevant to the present research.

Table 1. Previous Studies

No.	Authors	Title	Research Findings	Research Distinction
1	(Praneswara & Cahyono, 2023)	<i>Sentiment Analysis of TikTok Shop Seller Center Reviews on Google Play Store Using the Naïve Bayes Algorithm</i>	The study analyzed 5,000 user reviews of the TikTok Shop Seller Center and found that negative sentiment was more prevalent than positive sentiment. The Multinomial Naïve Bayes classifier achieved its strongest performance using a 90:10 training-to-testing data ratio, with an accuracy of 87.6%, precision of 85.8%, recall of 95.5%, and F1-score of 90.3%.	The research focused specifically on the TikTok Shop Seller Center application, whereas the present study examines a different application and its associated user reviews.
2	(Apriani & Gustian, 2019)	<i>Sentiment Analysis Using Naïve Bayes on Tokopedia Application Reviews</i>	From 1,500 reviews, 63.53% were classified as negative and 36.47% as positive. The findings indicated a relatively high proportion of negative comments despite the application's favorable user ratings. The Naïve Bayes model obtained an accuracy of 97.13%, precision of 1.00, recall of 95.49%, and AUC of 0.980. [6]	The object of analysis was Tokopedia, with the study concentrating on sentiment patterns in e-commerce application reviews. The present research applies the same classification approach to a different application context.
3	(Hasan, 2024)	<i>Sentiment Analysis of CapCut Application Users Based on Play Store Reviews Using the Naïve Bayes Method</i>	The study examined 880 reviews, consisting of 528 training records and 352 testing records. Negative sentiment slightly exceeded positive sentiment, with 455 negative and 425 positive reviews. Complaints were frequently associated with recent application updates that affected user satisfaction. The classifier produced an	The study investigated CapCut user reviews and emphasized dissatisfaction related to application updates. The present research addresses another application and develops a distinct analytical focus based on its review characteristics.

			accuracy of 84.09%, precision of 91.91%, and recall of 73.53%. [7]	
--	--	--	---	--

The reviewed studies demonstrate the suitability of Naïve Bayes for classifying sentiment in application reviews collected from Google Play Store. Across the three studies, differences in sentiment distribution and classifier performance were observed despite the use of comparable methodological approaches. Such variation suggests that the characteristics of an application, the vocabulary used by its users, the size of the dataset, and preprocessing procedures may influence classification outcomes. The present study extends this line of research by applying sentiment analysis to a different application and examining its user reviews through a comparable text-mining framework.

Google Play Store

Google Play Store is a digital distribution platform that enables Android users to obtain mobile applications and other digital products. In addition to providing access to applications, the platform facilitates user-generated evaluations through rating and review features. These reviews contain direct expressions of user experiences, perceptions, satisfaction, dissatisfaction, and expectations regarding the applications they use (Martanto & Istiono, 2024).

User reviews constitute an important source of qualitative information because prospective users frequently consider ratings and written feedback when deciding whether to install an application. The large volume of textual feedback available on Google Play Store makes the platform suitable for computational analysis, particularly for identifying recurring opinions and sentiment tendencies. Applications are organized into categories and include both free and paid products, allowing users to locate applications according to their particular needs (Agistina & Budiyantera, 2026).

From a research perspective, Google Play Store reviews can be treated as unstructured textual data that reflect real-world user experiences. Their continuous generation also provides a dynamic source of information for assessing public responses to application features, updates, performance, and service quality.

Naïve Bayes

Naïve Bayes is a probabilistic classification algorithm based on Bayes' theorem. The method estimates the probability that an observation belongs to a particular class by considering the probability of its observed features and the prior probability of each class. Its defining assumption is that the features contribute independently to the classification outcome, even though such independence may not fully exist in practical datasets (Karim et al., 2020).

The algorithm has been widely applied in text classification tasks, including sentiment analysis, spam filtering, and document categorization. Its computational simplicity, relatively low processing

requirements, and effectiveness in handling high-dimensional textual data make it suitable for applications involving large collections of reviews (Putra & Febriawan, 2024). In sentiment classification, Naïve Bayes determines the probability that a review belongs to a predefined sentiment category based on the occurrence of particular terms within the text.

Sentiment Analysis

Sentiment analysis is a computational approach for identifying and interpreting opinions expressed in textual data. In the context of application reviews, the technique is used to determine whether users express positive, negative, or neutral attitudes toward an application (Laksono & Suryono, 2025). The analytical process generally involves text processing, feature representation, classification, and interpretation of the resulting sentiment categories.

The meaning of sentiment cannot be determined solely from individual words because polarity is influenced by linguistic context and the way expressions are combined within a sentence. User reviews may contain explicit opinions, arguments, evaluations, complaints, or expressions of satisfaction, all of which contribute to the overall sentiment conveyed by the text (Goulart et al., 2024; Rizaldi et al., 2023). Sentiment analysis provides a systematic means of transforming these textual expressions into structured information that can be examined quantitatively.

Data Scraping

Data scraping refers to the automated extraction of information from websites or other digitally accessible sources using software, scripts, or specialized algorithms. The process is intended to collect data in a structured form so that it can be stored, processed, and analyzed for research purposes (Grewal et al., 2021).

In studies involving application reviews, scraping is particularly useful because manually collecting thousands of reviews is inefficient and susceptible to inconsistency. Automated extraction allows researchers to gather larger datasets within a defined period and establish a reproducible data collection procedure. The resulting dataset can then be subjected to cleaning, preprocessing, feature extraction, and classification.

Data Mining

Data mining refers to the process of discovering meaningful patterns, relationships, and previously unrecognized information from datasets through methods derived from statistics, mathematics, artificial intelligence, and machine learning (Peres et al., 2024). It involves selecting relevant data and applying analytical techniques to identify structures or patterns that may support interpretation and decision-making.

In textual research, data mining, as discussed by Paymode and Malode (2022), is useful for converting large volumes of raw data into interpretable knowledge. Classification, clustering, association analysis, and prediction are among the techniques commonly employed within data-mining applications. For sentiment analysis, classification is particularly relevant because textual records can be assigned to predefined sentiment categories based on their linguistic features.

Text Mining

Text mining is a branch of data mining concerned with extracting meaningful information and knowledge from textual documents. The process identifies relevant terms, linguistic patterns, and relationships that can represent the content of a collection of documents (Subasi et al., 2024). Because much of the information available in textual form is unstructured, text mining employs computational procedures to transform raw language into structured representations that can be analyzed systematically (Andreas et al., 2023; Das & Singh, 2023; He & Pan, 2026).

In application-review research, text mining enables large numbers of user comments to be processed according to consistent criteria. The approach can reveal frequently occurring terms, recurring themes, and sentiment patterns that may not be readily identified through manual examination alone. Text mining therefore forms an essential analytical foundation for sentiment classification based on user-generated reviews.

Text Preprocessing

Text preprocessing is a critical stage in text classification because raw textual data generally contain inconsistencies, irrelevant elements, and variations in word forms that may reduce classification effectiveness (Agrawal, 2023; Putri, 2025; Sepúlveda-Oviedo et al., 2023). The preprocessing stage prepares collected documents so that they can be represented in a more standardized and analyzable form.

The procedure commonly includes data cleaning and other transformations intended to reduce noise and improve the quality of textual representations. Individual sentences are segmented into smaller linguistic units, allowing words to be analyzed according to their contribution to the document (Bigne et al., 2021a; Jin & Ye, 2022; Patel et al., 2020). Appropriate preprocessing is particularly important in sentiment analysis because symbols, duplicate expressions, unnecessary characters, and inconsistent word forms can influence the representation of sentiment-bearing terms.

Term Frequency-Inverse Document Frequency (TF-IDF)

Term Frequency-Inverse Document Frequency (TF-IDF) is a feature-weighting technique used to represent the relative importance of terms within a collection of documents. The method combines two measures, namely Term Frequency (TF) and Inverse Document Frequency (IDF). TF measures how frequently a term appears within a

particular document, while IDF reflects the extent to which that term is distributed across the entire document collection (Tsao & Nogues, 2024).

A term receives a relatively high TF-IDF value when it occurs frequently in a particular document but appears in relatively few documents across the corpus. This characteristic makes TF-IDF useful for distinguishing informative terms from words that occur broadly throughout the dataset.

Term Frequency (TF)

The Term Frequency component can be expressed as:

$$[TF(t,d) = \frac{f_{\{t,d\}}}{\sum_k f_{\{k,d\}}}]$$

where:

- $(TF(t,d))$ = Term Frequency of term (t) in document (d)
- $(f_{\{t,d\}})$ = number of occurrences of term (t) in document (d)
- $(\sum_k f_{\{k,d\}})$ = total number of terms in document (d)

This formulation normalizes the frequency of a term according to the total number of terms contained in the document.

Inverse Document Frequency (IDF)

The Inverse Document Frequency component is defined as:

$$[IDF(t) = \log\left(\frac{N}{df(t)}\right)]$$

where:

- (N) = total number of documents in the corpus
- $(df(t))$ = number of documents containing term (t)

A term appearing in many documents receives a lower IDF value, whereas a term occurring in relatively few documents receives a higher value.

TF-IDF

The combined TF-IDF weight is calculated as:

$$[TF \text{ } IDF(t,d) = TF(t,d) \times IDF(t)]$$

where:

- $(TF(t,d))$ = Term Frequency value of term (t) in document (d)
- $(IDF(t))$ = Inverse Document Frequency value of term (t)

The resulting value represents the relative importance of a term within a document while accounting for its distribution throughout the corpus.

Model Evaluation

Model evaluation is performed to assess the classification performance of the Naïve Bayes model. A confusion matrix is commonly used as the primary evaluation framework for classification models. It summarizes the correspondence between actual class labels and predicted class labels through four fundamental components: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

True Positive represents observations that are correctly classified as belonging to a target class, while True Negative represents

observations that are correctly classified as not belonging to that class. False Positive refers to observations that actually belong to another class but are incorrectly assigned to the target class. False Negative describes observations that belong to the target class but are incorrectly classified as another class.

The confusion matrix provides the basis for several performance measures, including accuracy, precision, recall, and F1-score. Examining these metrics together provides a more comprehensive assessment of classifier performance than relying on a single measure.

Accuracy

Accuracy measures the proportion of correctly classified observations among all observations:

$$[\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}]$$

where:

- (TP) = number of correctly predicted positive observations
- (TN) = number of correctly predicted negative observations
- (FP) = number of negative observations incorrectly predicted as positive
- (FN) = number of positive observations incorrectly predicted as negative

A higher accuracy value indicates a greater proportion of correct predictions in the overall dataset.

Precision

Precision indicates the proportion of observations predicted as positive that are actually positive:

$$[\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}]$$

where:

- (TP) = number of correctly predicted positive observations
- (FP) = number of negative observations incorrectly predicted as positive

Precision is particularly informative when the reliability of positive predictions is an important consideration.

Recall

Recall measures the proportion of actual positive observations that are correctly identified by the classifier:

$$[\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}]$$

where:

- (TP) = number of correctly predicted positive observations
- (FN) = number of positive observations incorrectly predicted as negative

A high recall value indicates that the classifier is effective at identifying observations belonging to the target class.

F1-Score

F1-score provides a harmonic mean of precision and recall:

$$F1\text{-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

The metric is useful when a balance between precision and recall is required, particularly in datasets where the distribution of classes is not completely balanced. Taken together, accuracy, precision, recall, and F1-score provide complementary perspectives on the predictive capability of the Naïve Bayes classifier.

METHOD

Research Approach

This study employed a quantitative approach to classify user sentiment toward the AI Photo Studio: Pro Editor application based on reviews collected from the Google Play Store. Data processing was conducted using Python through a sequence of data collection, text preprocessing, sentiment labeling, feature extraction, classification, and model evaluation.

The study adopted the Knowledge Discovery in Databases (KDD) framework. The research process comprised data selection, preprocessing, transformation, data mining, and evaluation. Reviews were collected using the *google-play-scraper* library and processed through text cleaning, tokenization, stopword removal, short-word removal, and stemming using Sastrawi. Sentiment labels were generated using a lexicon-based approach, while TF-IDF was applied to transform the textual data into numerical features. SMOTE was used to address class imbalance in the training data before classification with Naïve Bayes.

Model performance was assessed using accuracy, precision, recall, F1-score, and a confusion matrix. The resulting model was subsequently deployed through a Gradio interface to enable sentiment prediction for new reviews.

Research Procedure

The research consisted of several sequential stages, beginning with data collection and continuing through preprocessing, sentiment labeling, feature extraction, classification, evaluation, and model deployment. The workflow is presented in Figure 1.

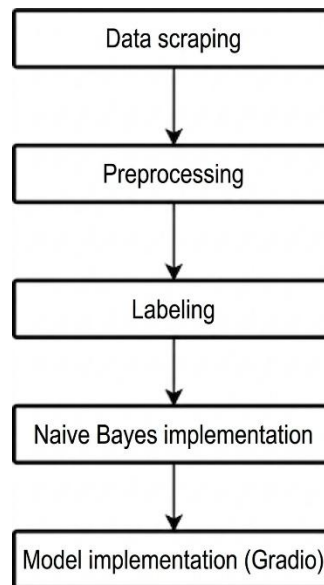


Figure 1. Research Procedure

Data Collection

User reviews of the AI Photo Studio: Pro Editor application were collected from the Google Play Store using Python in Google Colab and the google-play-scraper library. The collection was configured for Indonesian-language reviews using the language code id and country code id. Reviews available through the scraping process were collected without restricting the analysis to a specific star-rating category. Duplicate reviews and records with missing values were subsequently removed. The scraping process initially yielded 3,964 reviews, which were reduced to 2,028 unique reviews after duplicate removal.

Data Preprocessing

The collected reviews were subjected to preprocessing to reduce textual noise and standardize the data. The process included case folding, cleaning, tokenization, stopwords removal, short-word removal, and stemming. The preprocessing stages are summarized in Table 2.

Table 2. Text Preprocessing Procedures

Stage	Description
Case folding	Converts all text into lowercase characters.
Cleaning	Removes punctuation, numbers, emojis, URLs, and other irrelevant characters.
Tokenization	Divides the review into individual words or tokens.
Stopword removal	Removes common words with limited contribution to sentiment classification.
Short-word removal	Removes tokens containing fewer than two characters.
Stemming	Converts Indonesian words into their root forms using the Sastrawi library.

For example, the review “*Aplikasinya KEREN!!! 👍*” is transformed into “*aplikasinya keren*” after case folding and cleaning. The resulting text is then tokenized and normalized before feature extraction.

Sentiment Labeling

Sentiment labeling was performed using a lexicon-based approach with two sentiment dictionaries, `positive.tsv` and `negative.tsv`. Each preprocessed token was matched against the corresponding sentiment lexicon and assigned a sentiment weight. Positive terms contributed positive scores, whereas negative terms contributed negative scores. The cumulative score of each review was used to determine its sentiment category: positive when the score was greater than zero, negative when the score was less than zero, and neutral when the score was equal to zero. The lexicon used in this study contained 3,609 positive words and 6,607 negative words.

Because the reviews contain informal Indonesian expressions, slang, abbreviations, repeated characters, and context-dependent expressions, the lexicon-based approach may not capture all semantic variations accurately. The preprocessing stage was therefore applied before lexicon matching to normalize the review text. However, contextual meanings, negation, sarcasm, and words whose sentiment polarity depends on context may still produce labeling errors. This limitation is considered when interpreting the classification results.

```
[
Label =
\begin{cases}
Positive, & \text{Score} > 0 \\
Negative, & \text{Score} < 0 \\
Neutral, & \text{Score} = 0
\end{cases}
]
```

The resulting sentiment labels were used as the target classes in the classification stage.

An important limitation concerns the source of the sentiment labels used as the target variable. The labels were generated automatically using a lexicon-based approach rather than manually annotated by human reviewers. Therefore, the reported classification performance primarily reflects the ability of the Multinomial Naïve Bayes model to reproduce the automatically generated lexicon labels and should not be interpreted as a direct measure of agreement with human sentiment judgments. In addition, lexicon-based labeling may produce errors when words have context-dependent meanings, contain informal expressions, or involve negation and sarcasm. These limitations should be considered when interpreting the reported accuracy and class-level performance.

Feature Extraction and Class Balancing

The labeled dataset showed a substantial class imbalance, with 1,175 negative reviews, 430 positive reviews, and 371 neutral reviews. This imbalance may cause the classifier to favor the majority class and reduce its ability to recognize minority sentiment categories. Therefore, SMOTE was applied only to the training data to generate synthetic

observations for the minority classes. After oversampling, the training data contained 939 observations for each sentiment category. The resulting balanced distribution was used only for model training and does not represent the actual sentiment distribution of users. The original sentiment distribution of 1,175 negative, 430 positive, and 371 neutral reviews was retained for describing the dataset.

Naïve Bayes Classification

The Multinomial Naïve Bayes algorithm was used to classify reviews into positive, negative, and neutral sentiment categories. TF-IDF feature vectors served as the model inputs, while the sentiment labels generated through the lexicon-based method served as the target variable. The classifier was trained using the training dataset and then applied to the testing dataset. Multinomial Naïve Bayes was selected because of its relatively simple implementation, computational efficiency, and suitability for text-classification problems. TF-IDF feature vectors served as the model inputs, while the sentiment labels generated through the lexicon-based method served as the target variable.

The classifier was trained using the training dataset and then applied to the testing dataset. Naïve Bayes was selected because of its relatively simple implementation, computational efficiency, and suitability for text-classification problems.

Model Evaluation

Model performance was evaluated using a confusion matrix, accuracy, precision, recall, and F1-score. These measures were used to assess the ability of the classifier to correctly identify each sentiment category.

Accuracy indicates the proportion of correctly classified observations. Precision measures the proportion of predicted observations that are correctly assigned to a class, whereas recall measures the proportion of actual observations that are successfully identified. F1-score combines precision and recall into a single performance measure.

For the three-class classification task, the evaluation results were reported for the positive, negative, and neutral categories. The confusion matrix was also examined to identify patterns of correct and incorrect predictions across the three classes.

Model Deployment

The trained Multinomial Naïve Bayes model was deployed through a Gradio-based web interface to enable interactive sentiment classification. The trained classifier was stored as `model_nb.pkl`, while the TF-IDF vectorizer was stored as `tfidf.pkl` using Joblib. A new review entered through the interface was processed using the same preprocessing and feature-extraction procedures before being classified by the trained Multinomial Naïve Bayes model.

RESULT AND DISCUSSION

Data Scraping

Data scraping constitutes the initial stage of this research, involving the extraction of reviews for the "AI Photo Studio: Pro Editor" application using the Python programming language. The process targeted the 'id' language and country codes, ensuring the collected data originated from Indonesia and was written in Indonesian. The scraping process yielded 3,964 user reviews from the Google Play Store. Following data collection, a cleaning process was performed to remove duplicate reviews. This process resulted in 2,028 unique reviews, which served as the research dataset. Duplicate data was removed to prevent repetitive entries from skewing the sentiment analysis results, thereby ensuring the dataset was more representative and of higher quality.

RINGKASAN DATA HASIL SCRAPING

```
=====
Jumlah Review      : 2028
Rating Maksimum    : 5
Rating Minimum     : 1
Rata-rata Rating   : 4.58
```

Figure 2. Summary of Scraped Data

Based on the summary of the scraped data, the maximum rating is 5 and the minimum is 1, with an average rating of 4.58. This average indicates that users generally rate the "AI Photo Studio: Pro Editor" application highly. The resulting dataset comprises two primary attributes: "content," containing the user review text, and "score," representing the rating given by the user on a scale of 1 to 5. These attributes serve as input data for the subsequent stage of the process.

The rating distribution reveals that the review data is not evenly distributed across the rating categories. Out of the 2,028 reviews analyzed, there are 133 reviews with a rating of 1, 23 with a rating of 2, 63 with a rating of 3, 127 with a rating of 4, and 1,682 with a rating of 5. These results demonstrate that the majority of users awarded the highest rating of 5, indicating a predominantly positive response to the application under study.

Table 3. Data Scraping Results

Content	Score
<i>Bagus.</i>	5
<i>Lumayan bagus, bisa langsung jernih</i>	5
<i>Bener2 bagus.. gx bo'ong</i>	5
<i>sangat bagusAA</i>	5
<i>Bagus kali ya</i>	5

Although the dataset is dominated by 5-star reviews, The presence of reviews across ratings 1 to 4 provides additional textual information that may reflect a range of user experiences and evaluations. The entire scraped dataset is utilized as the basis for sentiment analysis. However, the imbalanced rating distribution requires attention during the classification stage, as it can affect the model's performance in identifying

classes with fewer data points. Thus, this data distribution serves as an initial overview of the dataset's characteristics prior to proceeding to the next stages.

Preprocessing

Data preprocessing involves cleaning and normalizing scraped review data before it is used in the sentiment labeling and classification stages. The primary objective of preprocessing is to transform raw text data into cleaner, structured, and higher-quality data, thereby improving the quality of the subsequent sentiment analysis and classification process.

```
=====
DATA BERHASIL DIMUAT
=====
Jumlah data awal : 2028

Memulai preprocessing...
Preprocessing selesai.

File berhasil disimpan:
preprocessed_data.csv

=====
RINGKASAN HASIL PREPROCESSING
=====
Jumlah data akhir : 1976
Jumlah kolom      : 3
Nama kolom       : ['content', 'score', 'cleaned']
```

Figure 3. Summary of Preprocessing Results

Based on the preprocessing stage, the initial dataset of 2,028 reviews was successfully processed into 1,976 reviews ready for use in the subsequent research phase. During preprocessing, 52 entries were removed because they resulted in empty text after undergoing data cleaning steps such as the removal of special characters, numbers, stopwords, and meaningless words leaving no analyzable information. The resulting dataset was saved in a file named `preprocessed\_data.csv`, comprising three attributes: `content`, `score`, and `cleaned`. The `content` column contains the original review text, `score` holds the user-assigned rating, and `cleaned` represents the final output of the preprocessing stage, intended for use as input in the next process.

Table 4. Table of Preprocessing Results Data

Content	Cleaned	Score
<i>Mantap untuk saat ini</i>	<i>mantap</i>	5
<i>lumayan bagus, bisa langsung jernih</i>	<i>lumayan bagus langsung jernih</i>	5
<i>sangat bagusAA</i>	<i>sangat bagus</i>	5
<i>bagus kali ya</i>	<i>bagus kali</i>	5
<i>Bagus</i>	<i>bagus</i>	5

The preprocessing results indicate that each review underwent a series of text processing stages: case folding, cleaning, tokenizing, stopword removal, removal of short words, and stemming. For instance, the review "*lumayan bagus, bisa langsung jernih*" (pretty good, can clear up immediately) became "*lumayan bagus langsung jernih*" after the word

"bisa" was removed as a stopword and punctuation marks were eliminated. Another example is the review "Mantap untuk saat ini" (Great for now), which became simply "mantap" (great). These changes demonstrate that the preprocessing process successfully removed irrelevant elements while retaining words containing information crucial for sentiment analysis.

The rating distribution also underwent slight changes following the removal of data that did not meet the criteria. Out of a total of 1,976 reviews, there were 132 reviews with a rating of 1, 23 with a rating of 2, 62 with a rating of 3, 121 with a rating of 4, and 1,638 with a rating of 5. Although the data count in each category decreased slightly compared to the pre-processing stage, the rating distribution pattern shows that the majority of users gave a rating of 5; thus, the dataset remains dominated by positive ratings. This indicates that the preprocessing process did not alter the dataset's primary characteristics but rather improved data quality by eliminating irrelevant information.

The preprocessing stage concluded with a check for missing values in each column. The results showed that the content, score, and cleaned columns each contained zero missing values, leading to the conclusion that the dataset is clean and complete, the preprocessed data meets the requirements for use in the subsequent stage.

Labeling

Labeling is the process of assigning sentiment labels (e.g., positive, negative, and neutral) to text intended for training a classification model[23]. While data labeling can be performed manually, the process is extremely time-consuming and costly, the labeling process in this study employs a lexicon-based approach.

```
=====
DATA BERHASIL DIMUAT
=====
Jumlah data : 1976

Informasi Lexicon
-----
Jumlah kata positif : 3609
Jumlah kata negatif : 6607
Total kata lexicon : 9074

Memulai pelabelan menggunakan Lexicon...

File berhasil disimpan : labeled_data.csv
```

Figure 4. Summary of Labeling Results Data

Based on the sentiment labeling process using the lexicon-based method, 1,976 preprocessed review entries were successfully labeled for sentiment automatically. The labeling process utilized a sentiment dictionary comprising 3,609 positive words and 6,607 negative words, resulting in a total of 9,074 words serving as the reference for determining the sentiment polarity of each review. Each word in the dictionary was assigned a specific weight; these weights were summed to produce a lexicon score. This score was then used to determine the

sentiment category: positive if the score was greater than zero, negative if less than zero, and neutral if equal to zero. The resulting labeled dataset was saved to the file `labeled\_data.csv`. The resulting labeled dataset was saved to the file labeled_data.csv for use in the subsequent Multinomial Naïve Bayes classification stage. The labeling process resulted in changes to the data distribution across the respective labels, as shown in Table 5 below.

Table 5. Number of Data Points After the Labeling Process

Label	Number
Negative	1,175
Positive	430
Neutral	371

The labeling results show that, out of 1,976 review records, 1,175 were classified as negative sentiment, 430 as positive sentiment, and 371 as neutral sentiment. This distribution indicates that the dataset was dominated by negative sentiment labels. This finding differs from the application's rating distribution, which was previously dominated by five-star ratings. The discrepancy may arise because the Lexicon-Based method does not determine sentiment from the numerical rating assigned by users, but rather from the cumulative weights of words contained in the review text. Accordingly, a review with a high rating may still be classified as negative when it contains a greater number of words associated with negative sentiment in the sentiment lexicon.

The distribution of lexicon scores shows that sentiment scores ranged from -63 to 13. Negative scores indicate that a review contained more words associated with negative sentiment, whereas positive scores indicate a greater presence of positively weighted words. In addition, 371 reviews obtained a score of 0 and were therefore classified as neutral sentiment. The variation in sentiment scores indicates that the polarity of each review differs according to the number and weighting of sentiment-bearing words identified by the lexicon.

Table 6. Data Summary After Lexicon Processing

Content	Cleaned	Score	Lexicon_score	Label
<i>mantap</i>	<i>mantap</i>	5	5.0	Positive
<i>saya sangat puas dengan hasil yang ini behhh persisi</i>	<i>sangat puas hasil behhh persisi lah</i>	5	1.0	Positive
<i>lah pokok ha fidio saya foto saya</i>	<i>pokok fidio foto</i>			
<i>bagus</i>	<i>bagus</i>	5	-4.0	Negative
<i>mantap untuk saat ini</i>	<i>mantap</i>	5	5.0	Positive
<i>yeah</i>	<i>yeah</i>	5	0.0	Neutral

The review “bagus” received a lexicon score of -4 and was therefore automatically classified as negative. This result does not necessarily represent the actual semantic meaning of the review, because “bagus” generally expresses a positive evaluation. The finding indicates a limitation of the lexicon used in this study, particularly in handling

Indonesian vocabulary and contextual polarity. Therefore, the lexicon-generated labels should be interpreted as automated polarity estimates rather than definitive human-validated sentiment labels. This potential labeling error may also affect the resulting sentiment distribution and subsequent classification performance.

The labeling process successfully produced a dataset containing positive, negative, and neutral sentiment classes. The labeled dataset was first divided into training and testing subsets using an 80:20 stratified split. The TF-IDF vectorizer was fitted only on the training data and subsequently used to transform both the training and testing data. SMOTE was then applied only to the training feature matrix, while the testing data remained unchanged and was not used during feature fitting, oversampling, or model training.

Multinomial Naïve Bayes Implementation

The Naïve Bayes implementation was conducted to develop and evaluate a sentiment classification model using the Multinomial Naïve Bayes algorithm. The model was designed to classify application user reviews into three sentiment categories: positive, negative, and neutral, based on the labels generated in the preceding stage. In addition to Multinomial Naïve Bayes, the study employed the Synthetic Minority Over-sampling Technique (SMOTE) to address the imbalance in the number of observations across the sentiment classes.

```
=====
DATA BERHASIL DIMUAT
=====
Jumlah Data : 1976

Distribusi Label Sebelum SMOTE

  Label  Jumlah
0  negatif  1175
1  positif   430
2  netral    371

Jumlah Data Training : 1580
Jumlah Data Testing  : 396

Melakukan TF-IDF...
Jumlah Fitur : 1263

Menerapkan SMOTE...

Distribusi Label Setelah SMOTE

  Label  Jumlah
0  positif  939
1  netral   939
2  negatif  939

Melatih Model Multinomial Naive Bayes...

Melakukan Prediksi...
```

Figure 5. Summary of Naive Bayes Implementation Results

Based on the implementation of the Multinomial Naïve Bayes algorithm, the study used 1,976 review records that had undergone preprocessing and sentiment labeling. Before model training, the data distribution was imbalanced, consisting of 1,175 negative, 430 positive, and 371 neutral reviews. This imbalance could cause the model to favor

the majority class during the learning process. To address this issue, the *Synthetic Minority Over-sampling Technique (SMOTE)* was applied to balance the sentiment classes before model training.

The dataset was divided into 80% training data and 20% testing data using a train-test split with stratification to preserve the proportion of each sentiment class. This procedure resulted in 1,580 training records and 396 testing records. Feature extraction was then performed using *Term Frequency-Inverse Document Frequency (TF-IDF)*, producing 1,263 unique features derived from the preprocessed reviews. These features provided a numerical representation of the textual data required by the Multinomial Naïve Bayes classifier.

SMOTE was subsequently applied to the training data to address class imbalance. After oversampling, each sentiment class contained 939 training observations, resulting in an equal distribution across the negative, positive, and neutral categories. A balanced training dataset allows the classifier to learn the characteristics of each sentiment category more consistently and reduces the potential influence of the majority class.

Following model implementation, an evaluation stage was conducted to assess the performance of the Multinomial Naïve Bayes classifier in categorizing reviews of the AI Photo Studio: Pro Editor application. The model generated predictions using the testing dataset, and these predictions were compared with the actual sentiment labels using several evaluation metrics. The assessment focused on the model's ability to correctly distinguish among negative, positive, and neutral reviews.

```
=====
HASIL EVALUASI MODEL
=====
Accuracy : 0.7500

Classification Report

              precision    recall  f1-score   support

negatif      0.80      0.88      0.84        236
netral       0.65      0.35      0.46         74
positif      0.66      0.74      0.70         86

accuracy          0.75        396
macro avg      0.70      0.66      0.66        396
weighted avg   0.74      0.75      0.74        396
```

Figure 6. Summary of Evaluation Results Data

After the training process, the Multinomial Naïve Bayes model was evaluated using 396 testing samples that were not used during model training. The model achieved an accuracy of 75.00%. The classification performance varied across the three sentiment categories. The negative class achieved a precision of 0.80, recall of 0.88, and F1-score of 0.84. The positive class achieved a precision of 0.66, recall of 0.74, and F1-score of 0.70, while the neutral class achieved a precision of 0.65, recall of 0.35, and F1-score of 0.46. The results indicate that the model

performed best in identifying negative sentiment, whereas neutral sentiment remained the most difficult class to classify.

The final model achieved an accuracy of 75.00%. The negative class achieved a precision of 0.80, recall of 0.88, and F1-score of 0.84. The positive class achieved a precision of 0.66, recall of 0.74, and F1-score of 0.70, while the neutral class achieved a precision of 0.65, recall of 0.35, and F1-score of 0.46.

The sentiment labeling process resulted in three sentiment categories: negative, positive, and neutral. Based on the final labeled dataset, 1,175 reviews were classified as negative, 430 as positive, and 371 as neutral, yielding a total of 1,976 reviews. The distribution indicates that negative sentiment constituted the largest class, followed by positive and neutral sentiment. This distribution is presented in Figure 7.

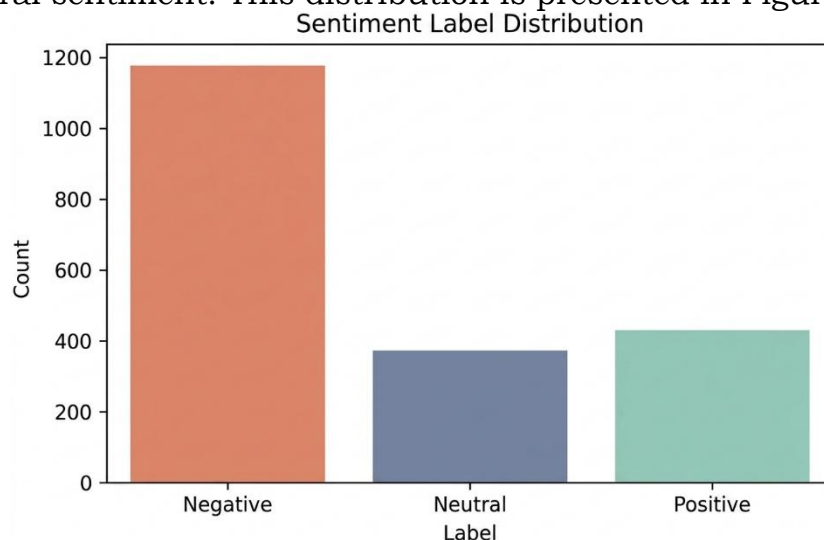


Figure 7. Sentiment Distribution of User Reviews

The sentiment label distribution chart reveals an imbalanced dataset resulting from the lexicon-based labeling method. The negative sentiment class contains the highest volume of data, comprising 1,175 reviews, whereas the positive and neutral sentiment classes consist of 430 and 371 reviews, respectively. This distribution confirms that negative sentiment was the largest class in the lexicon-based labeling results. This significant disparity indicates that the labeling process categorized a larger number of reviews as negative compared to positive or neutral. Such an imbalance can affect the classification algorithm's learning process, as models tend to more easily recognize patterns associated with classes that have a larger volume of data. Consequently, prior to model training, this study employs the Synthetic Minority Over-sampling Technique (SMOTE) to balance the data across classes, thereby ensuring a more optimal learning process that is not biased toward the majority class.

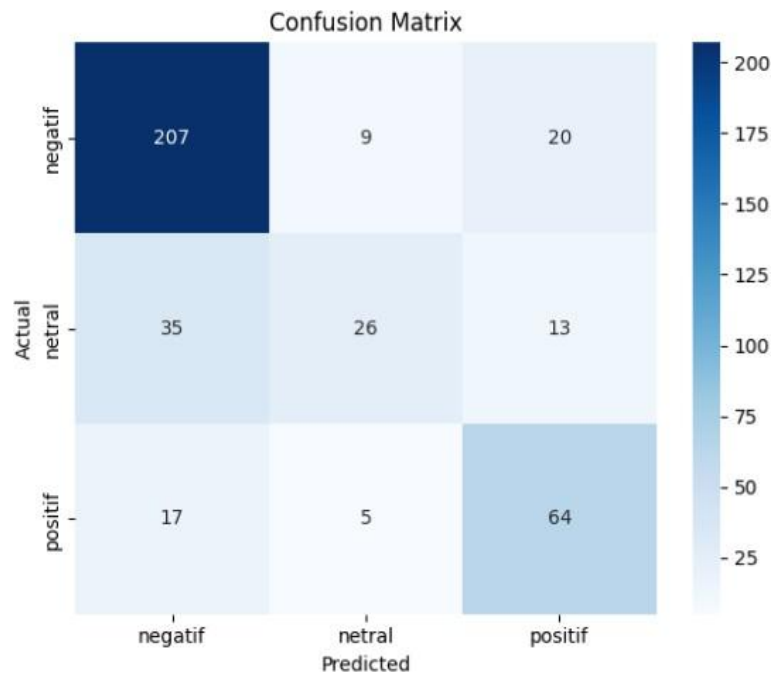


Figure 8. Confusion Matrix

The neutral class was the most difficult category to classify. Of the 74 actual neutral reviews in the testing set, only 26 were correctly classified as neutral, while 35 were classified as negative and 13 as positive. This result is consistent with the low recall of 0.35 for the neutral class. One possible explanation is the labeling rule used in the lexicon-based approach, in which a review is assigned a neutral label when its cumulative lexicon score is exactly zero. A zero score does not necessarily indicate that the review expresses genuinely neutral sentiment; it may also occur when positive and negative lexical weights cancel each other or when the review contains few sentiment-bearing words. Consequently, neutral reviews may have linguistic characteristics that overlap with both positive and negative categories, making them more difficult for the classifier to distinguish.

File berhasil disimpan:
hasil_klasifikasi_sentimen.xlsx

Contoh Hasil Prediksi

	cleaned	actual_label	predicted_label
0	ngasih gratis cma gmn guna mau puas	positif	positif
1	kerennnn	netral	negatif
2	ahhhh jujur puas makasih yaa buat	positif	positif
3	iklan nya tlg jangan banyak	negatif	negatif
4	sangat puas hasil	positif	positif
5	woww keren	netral	netral
6	mntp klo gratis lama aja	positif	positif
7	oke sekali	positif	positif
8	tarik	positif	negatif
9	wah sangat bagus cerah	negatif	negatif

Figure 9. Prediction Results

Based on the sample predictions, the majority of the data was correctly classified; for instance, the review "ngasih gratis cma gmn guna

mau puas" was predicted as positive, matching its actual label, while the review "iklan nya tlg jangan banyak" was successfully predicted as negative. However, there were still some classification errors, such as the review "kerennnn," which was labeled neutral but predicted as negative, and the review "tarik," which was labeled positive but predicted as negative. These prediction errors indicate that certain reviews possess word characteristics that are difficult for the model to distinguish, particularly within the neutral and positive classes.

Model Implementation (Gradio)

The model implementation utilizes Gradio to integrate the developed Naïve Bayes model into a web-based application, enabling users to directly perform sentiment classification on opinions or reviews. The application interface is designed using Gradio Blocks, featuring a simple and user-friendly layout. The main page comprises an "Enter Opinion" field, "Classify" and "Cancel" buttons, and a "Sentiment Result" field. The "Classify" button triggers the entire analysis process ranging from preprocessing to sentiment prediction using the Naïve Bayes model while the "Cancel" button clears the input field and the prediction result. The application is executed using the ``demo.launch(share=True)`` command, generating a temporary web address (Public URL) accessible via a web browser.

Figure 10. Web Page Appearance

Based on the web-based application, the system detects that the input opinion results in a negative sentiment. It can be concluded that the application is capable of detecting the sentiment of the input opinion.

CONCLUSION

This study demonstrates that the Multinomial Naïve Bayes algorithm can be applied to classify user reviews of the AI Photo Studio: Pro Editor application into positive, negative, and neutral sentiment categories. Using 1,976 preprocessed reviews, the final model achieved an accuracy of 75.00% on 396 testing samples. The negative class showed the strongest performance, with a precision of 0.80, recall of 0.88, and F1-score of 0.84. In contrast, the neutral class remained the most challenging category, with a recall of 0.35 and an F1-score of 0.46. These findings indicate that the model can distinguish negative and positive sentiment more effectively than neutral sentiment in the

evaluated dataset. However, the sentiment labels were generated automatically using a lexicon-based approach and were not validated through human annotation. The reported performance should be interpreted as the ability of the model to reproduce lexicon-derived labels rather than as direct evidence of agreement with human sentiment judgments. The findings provide application-specific evidence of textual sentiment in AI Photo Studio: Pro Editor reviews and demonstrate the value of analyzing written feedback alongside numerical ratings.

REFERENCES

- Agistina, S. D., & Budiyanantara, A. (2026). Klasifikasi Metode Naïve Bayes pada Ulasan Pengguna Aplikasi Dazzcam untuk Pengeditan Foto Vintage di App Store. *DIGINTEL-AI: DIGital INnovation and InTELLigence-AI*, 1(2).
- Agrawal, A. (2023). *No-Code Artificial Intelligence: The new way to build AI powered applications (English Edition)*. Bpb Publications.
- Andreas, A., Mavromoustakis, C. X., Song, H., & Batalla, J. M. (2023). Optimisation of CNN through transferable online knowledge for stress and sentiment classification. *IEEE Transactions on Consumer Electronics*, 70(1), 3088–3097.
- Apriani, R., & Gustian, D. (2019). Analisis sentimen dengan Naïve Bayes terhadap komentar aplikasi Tokopedia. *Jurnal Rekayasa Teknologi Nusa Putra*, 6(1), 54–62.
- Bigne, E., Ruiz, C., Cuenca, A., Perez, C., & Garcia, A. (2021a). What drives the helpfulness of online reviews? A deep learning study of sentiment analysis, pictorial content and reviewer expertise for mature destinations. *Journal of Destination Marketing & Management*, 20, 100570.
- Bigne, E., Ruiz, C., Cuenca, A., Perez, C., & Garcia, A. (2021b). What drives the helpfulness of online reviews? A deep learning study of sentiment analysis, pictorial content and reviewer expertise for mature destinations. *Journal of Destination Marketing & Management*, 20, 100570.
- Cooper, R. G. (2024). The AI transformation of product innovation. *Industrial Marketing Management*, 119, 62–74.
- Das, R., & Singh, T. D. (2023). Multimodal sentiment analysis: a survey of methods, trends, and challenges. *ACM Computing Surveys*, 55(13s), 1–38.
- Egger, M., Ley, M., & Hanke, S. (2019). Emotion recognition from physiological signal analysis: A review. *Electronic Notes in Theoretical Computer Science*, 343, 35–55.
- Goulart, L., Matte, M. L., Mendoza, A., Alvarado, L., & Veloso, I. (2024). AI or student writing? Analyzing the situational and linguistic characteristics of undergraduate student writing and AI-generated assignments. *Journal of Second Language Writing*, 66, 101160.
- Grewal, D., Guha, A., Satornino, C. B., & Schweiger, E. B. (2021). Artificial intelligence: The light and the darkness. *Journal of Business Research*, 136, 229–236.

- Hasan, F. N. (2024). Analisis Sentimen Pengguna Aplikasi CapCut Pada Ulasan di Play Store Menggunakan Metode Naïve Bayes. *Media Online*, 4(4).
- He, X., & Pan, X. (2026). What are the Chinese public's attitudes towards AI-generated portrait images? A text mining analysis based on social media. *Aslib Journal of Information Management*, 1–17.
- Hu, S., Kumar, A., Al-Turjman, F., Gupta, S., & Seth, S. (2020). Reviewer credibility and sentiment analysis based user profile modelling for online product recommendation. *Ieee Access*, 8, 26172–26189.
- Jin, K., & Ye, J. (2022). Artificial intelligence and deep learning in ophthalmology: Current status and future perspectives. *Advances in Ophthalmology Practice and Research*, 2(3), 100078.
- Karim, A., Azhari, A., Belhaouri, S. B., Qureshi, A. A., & Ahmad, M. (2020). Methodology for analyzing the traditional algorithms performance of user reviews using machine learning techniques. *Algorithms*, 13(8), 202.
- Kisilevich, S., Rohrdantz, C., & Keim, D. (2010). “Beautiful picture of an ugly place”. Exploring photo collections using opinion and sentiment analysis of user comments. *Proceedings of the International Multiconference on Computer Science and Information Technology*, 419–428.
- Laksono, U. H., & Suryono, R. R. (2025). Sentiment analysis of online dating apps using Support Vector Machine and Naïve Bayes algorithms. *Jurnal Teknik Informatika (Jutif)*, 6(1), 229–238.
- Liang, S., Wu, D., & Zhang, C. (2024). Enhancing image sentiment analysis: A user-centered approach through user emotions and visual features. *Information Processing & Management*, 61(4), 103749.
- Martanto, M. L. F., & Istiono, W. (2024). Sentiment analysis of M-Paspor app reviews using multinomial Naive Bayes. *Journal of Logistics, Informatics and Service Science*, 11(10), 311–326.
- Nilashi, M., Abumalloh, R. A., Ahmadi, H., Samad, S., Alrizq, M., Abosaq, H., & Alghamdi, A. (2023). The nexus between quality of customer relationship management systems and customers' satisfaction: Evidence from online customers' reviews. *Heliyon*, 9(11).
- Patel, K., Mehta, D., Mistry, C., Gupta, R., Tanwar, S., Kumar, N., & Alazab, M. (2020). Facial sentiment analysis using AI techniques: state-of-the-art, taxonomies, and challenges. *IEEE Access*, 8, 90495–90519.
- Paymode, A. S., & Malode, V. B. (2022). Transfer learning for multi-crop leaf disease image classification using convolutional neural network VGG. *Artificial Intelligence in Agriculture*, 6, 23–33.
- Peres, R., Schreier, M., Schweidel, D. A., & Sorescu, A. (2024). The creator economy: An introduction and a call for scholarly research. In *International Journal of Research in Marketing* (Vol. 41, Number 3, pp. 403–410). Elsevier.
- Praneswara, A. O., & Cahyono, N. (2023). Analisis Sentimen Ulasan Aplikasi TikTok Shop Seller Center di Google Playstore

- Menggunakan Algoritma Naive Bayes. *The Indonesian Journal of Computer Science*, 12(6).
- Putra, S. H. W., & Febriawan, D. (2024). Analisis Sentimen Ulasan Aplikasi Digital Korlantas POLRI Menggunakan Naïve Bayes pada Google Play Store. *Kaji. Ilm. Inform. Dan Komput*, 4(4).
- Putri, S. (2025). Sentiment Analysis of Shopee User Reviews on the Play Store Using the Naive Bayes Algorithm for Evaluating User Satisfaction. *Digital Informatics*, 1(2), 84–91.
- Rizaldi, S. A. R., Alam, S., & Kurniawan, I. (2023). Analisis Sentimen Pengguna Aplikasi JMO (Jamsostek Mobile) Pada Google Play Store Menggunakan Metode Naive Bayes. *STORAGE: Jurnal Ilmiah Teknik Dan Ilmu Komputer*, 2(3), 109–117.
- Sepúlveda-Oviedo, E. H., Travé-Massuyès, L., Subias, A., Pavlov, M., & Alonso, C. (2023). Fault diagnosis of photovoltaic systems using artificial intelligence: A bibliometric approach. *Heliyon*, 9(11).
- Subasi, A., Qaisar, S. M., & Nisar, H. (2024). *Artificial Intelligence and Multimodal Signal Processing in Human-Machine Interaction*. Elsevier.
- Tsao, J., & Nogues, C. (2024). Beyond the author: Artificial intelligence, creative writing and intellectual emancipation. *Poetics*, 102, 101865.
- Zhu, X., Guo, C., Feng, H., Huang, Y., Feng, Y., Wang, X., & Wang, R. (2024). A review of key technologies for emotion analysis using multimodal information. *Cognitive Computation*, 16(4), 1504–1530.