

Comparative Sentiment Analysis of Digital Banking Reviews Using Knowledge Discovery IndoBERT

Musdalifa Putri Casanova ¹, Allsela Meiriza ^{2*}, Ken Ditha Tania ³, Nabila Rizky Oktadini ⁴, Dedy Kurniawan ⁵
Sistem Informasi, Universitas Sriwijaya, Indonesia
09031282227040@student.unsri.ac.id ¹, allsela@unsri.ac.id ², kenya.tania@gmail.com ³, nabilarizky@unsri.ac.id ⁴,
dedykurniawan@unsri.ac.id ⁵

Article Info

Article history:

Received 2026-02-06

Revised 2026-02-25

Accepted 2026-04-08

Keyword:

*Sentiment Analysis,
Knowledge Discovery,
Text Mining,
IndoBERT,
Blu by BCA.*

ABSTRACT

The ever-evolving digital banking services require a comprehensive understanding of user perceptions. This study applies the Knowledge Discovery in Databases (KDD) approach, a structured process involving data selection, preprocessing, data mining, and interpretation, to analyze sentiment in user reviews of digital banking services. Data was collected by gathering reviews from the Google Play Store platform, resulting in 22,923 valid reviews after filtering and preprocessing. Sentiment labels were automatically assigned based on rating values, where ratings of 4–5 were categorized as positive and ratings of 1–3 as negative. After cleaning, the dataset consisted of 5,806 positive reviews and 4,193 negative reviews. Three classification algorithms—Support Vector Machine, Convolutional Neural Network, and IndoBERT—were compared. Model performance was evaluated using accuracy, precision, recall, F1-score, and AUC-ROC metrics. The results show that IndoBERT achieved the most stable and superior performance, with accuracy, precision, recall, and F1-score values of around 0.87, and the highest AUC-ROC value of 0.89, outperforming CNN and SVM. Additionally, WordCloud analysis was used as part of the knowledge discovery process to identify dominant words in each sentiment category, revealing aspects of the service that users frequently appreciate, as well as issues that often arise in reviews. Overall, this study produced knowledge discovery in the form of sentiment patterns and characteristics of user perceptions of the Blu by BCA application based on public review data.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. INTRODUCTION

The rapid advancement of information and communication technology has transformed the way people interact economically and financially [1]. Digital banking applications enable users to conduct financial transactions conveniently, quickly, and flexibly via mobile devices[2]. In Indonesia, the high adoption rate of digital technology has contributed to this phenomenon, with around 79.5% of the total population using the internet in their daily activities. This has encouraged the banking industry to continue innovating through the digitization of services in order to improve competitiveness and service quality for customers.

One of the innovations in Indonesia's digital banking sector is the Blu by BCA app, which was launched in July 2021 as a digital banking service with the concept of direct digital

banking and Bank as a Service (BaaS). The large volume of user reviews generated on application platforms constitutes a valuable source of data for understanding user perceptions and experiences.

As the use of mobile banking applications increases, users are becoming more active in expressing their opinions through online reviews. The reviews contain various experiences, ranging from satisfaction to complaints about the quality of the application's service. However, the increasing number of reviews, along with the use of informal language, abbreviations, and code-mixing, makes manual analysis inefficient. Previous research shows that reviews of mobile banking applications in Indonesia tend to be dominated by negative sentiments related to technical issues, such as login, verification, and transaction problems[3]. Therefore,

automated sentiment analysis is required to process review data systematically and objectively.

Sentiment analysis in natural language processing is commonly employed to categorize user opinions into sentiment classes. Various machine learning and deep learning methods have been applied in this domain. In the context of digital applications, sentiment analysis has an important role in describing users' perceptions of service quality.

A number of studies show that the use of machine learning and deep learning models can improve sentiment classification performance on text data. Transformer-based models such as IndoBERT have proven to have higher accuracy than conventional methods due to their ability to understand non-standard Indonesian language contexts and code-mixing phenomena[4]. Sentiment analysis results can generate new insights that represent user opinion trends[5]. To manage large amounts of heterogeneous user review data, a Knowledge Discovery in Databases (KDD) approach is needed as a structured framework[6]. KDD includes the stages of data collection, text cleaning and preprocessing, feature transformation, application of data mining techniques, and evaluation and interpretation of results. This approach enables the processing of unstructured text data so that patterns and meaningful information can be extracted systematically[7].

In sentiment analysis research, the KDD stages generally involve language normalization, tokenization, feature extraction, and model training and evaluation using performance metrics such as accuracy, precision, recall, and F1-score[8]. There are various algorithms that have been applied in sentiment analysis research, including Support Vector Machine (SVM), Convolutional Neural Network (CNN), and IndoBERT. Several studies report that SVM is capable of producing high accuracy in text classification, such as in Shopee app reviews with an accuracy of 86.906%, and Motion Bank app reviews with an accuracy of 93.7% [9][10]. Meanwhile, CNN also performed well in sentiment classification, with an accuracy rate of 94% in social media data analysis[11]. On the other hand, IndoBERT consistently demonstrates more stable and superior performance compared to other models in various Indonesian text classification tasks, including review analysis and aspect detection[12].

However, most previous studies still focus on comparing algorithm performance without integrating the analysis results into a comprehensive knowledge discovery framework. The Knowledge Discovery in Databases (KDD) approach emphasizes not only the process of sentiment classification but also the systematic discovery of patterns and interpretation of results to generate meaningful information from data. This is reflected in studies that adopt KDD, wherein stages such as data selection, preprocessing, transformation, mining, and interpretation collectively facilitate the extraction of actionable knowledge[13]. Nevertheless, many studies on Indonesian mobile application reviews emphasize technical performance metrics and

examine models separately, with limited efforts to combine classical machine learning, deep learning, and transformer-based approaches within a unified framework. As a result, sentiment analysis findings are often reported only as accuracy measures, with less interpretation of sentiment patterns as knowledge representing user experience trends and dominant service issues. Therefore, this research applies the KDD framework to analyze user sentiment in Blu by BCA reviews by comparing SVM, CNN, and IndoBERT.

This research aims not only to evaluate classification performance but also to generate knowledge in the form of sentiment patterns and dominant user issues, thereby contributing both methodologically and practically to sentiment analysis research on Indonesian textual data.

II. METHODS

This research uses a quantitative approach by applying sentiment analysis to identify user opinion trends regarding the Blu by BCA application. The analysis was conducted using machine learning and deep learning methods within the framework of Knowledge Discovery in Databases (KDD), which includes the stages of data collection, text preprocessing, sentiment classification, model evaluation, and interpretation of results. This approach is used to extract sentiment patterns and user opinion characteristics based on publicly available review data.

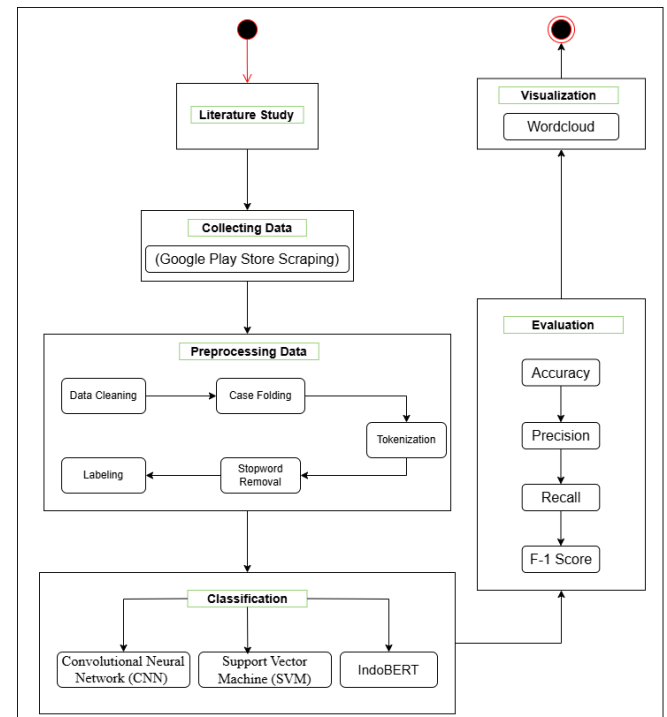


Figure 1. Flowchart Process

Figure 1 illustrates the overall research workflow based on the Knowledge Discovery in Databases (KDD) framework. The process begins with a literature review to establish a theoretical foundation and identify relevant methods in sentiment analysis and digital banking research.

The next step involves collecting data by gathering user reviews from the Google Play Store platform. The collected data is then processed through a pre-processing procedure, including data cleaning, case adjustment, tokenization, stop word removal, and sentiment tagging. Data cleaning removes irrelevant elements such as URLs, emojis, symbols, and numeric characters. Case adjustment standardizes text to lowercase format, while tokenization breaks text into individual word units. Stop word removal eliminates frequently occurring words with limited semantic contribution. The tagging process classifies sentiment based on rating values.

After preprocessing, the cleaned and labeled dataset was used in the classification stage. Three classification models were implemented, namely Convolutional Neural Network (CNN), Support Vector Machine (SVM), and IndoBERT. Each model was trained and evaluated to compare their performance in sentiment prediction.

The evaluation stage uses several performance metrics, including accuracy, precision, recall, and F1-score, to assess the effectiveness of classification. Finally, the visualization stage uses WordCloud to interpret the dominant terms in each sentiment category, facilitating the generation of thematic insights as part of the knowledge discovery process.

A. Literature Study

A literature study was conducted to obtain a conceptual basis related to sentiment analysis, text mining, and the implementation of the Knowledge Discovery in Databases (KDD) framework on user review text data. The study focuses on research that applies machine learning and deep learning methods, specifically SVM, CNN, and IndoBERT, in the classification of Indonesian text sentiment. In addition, previous literature on digital application review analysis was examined to identify methodological approaches, evaluation metrics, and research gaps that underlie the design of the methodology in this research.

B. Collecting Data

The research object is user reviews of the Blu by BCA application obtained from the Google Play Store platform. Data were collected using web scraping techniques through the Python programming language in the Google Colaboratory environment, utilizing the google-play-scraper library. The dataset comprised 25,000 reviews and several attributes, namely review ID, username, rating, review text, and date. All of the collected data was then stored in CSV format to facilitate further analysis.

The data collection process complies with applicable research ethics standards and platform policies. All reviews are obtained exclusively from the publicly accessible section of the Google Play Store without bypassing any security mechanisms. Only publicly available review text, rating scores, and timestamps are used in the analysis. No personal information is extracted, stored, or analyzed, and data is processed solely for academic research purposes to analyze

aggregate sentiment patterns rather than individual user behavior.

TABLE I
SAMPLE OF DATA SCRAPING RESULTS

Review ID	Username	Rating	Review Text	Date
22754def-50c6-4439-8ce9-92e4c2c53ac3	Pengguna Google	1	gangguan trus..susah msuk banking	2025-05-19 12:59:49
4730f8dc-e25f-4861-b629-d9df8cee174b	Pengguna Google	3	tolong diperbaiki, penerimaan uang via qris dri bank atau ewallet selalu gagal	2025-05-19 12:26:03
55dec945-7231-4ec5-9000-a4d8887e8aa4	Pengguna Google	5	transaksi jadi lebih mudah	2025-05-19 06:06:26
...
9952f346-7e0b-48f9-bff1-2ef8025ce68e	Pengguna Google	5	bagussssssssss sssss	2025-05-19 01:48:24

C. Preprocessing Data

Preprocessing is performed to improve the quality of text data before it is used in the sentiment classification process. This is an important step in sentiment analysis because the level of data cleanliness and consistency significantly affects model performance. The accuracy of text classification models can be improved if preprocessing is carried out systematically, such as through lowercasing, noise removal, stopword removal, and tokenization[14].

In this research, preprocessing consisted of several stages, namely data cleaning, case folding, tokenization, stopword removal, text normalization, and labeling. Data cleaning was performed by removing non-informative elements such as URLs, emojis, symbols, and numerical characters[15]. Reviews containing only rating values without textual comments were also removed because they do not provide linguistic information for sentiment analysis[16]. Furthermore, case folding was applied to convert all text characters into lowercase format to maintain consistency in writing[17]. The tokenization process was conducted to segment text into word units, while stopword removal aimed to eliminate common words with limited semantic contribution[18].

In addition to basic preprocessing, text normalization was applied to convert informal or non-standard Indonesian words into their formal equivalents using a manually curated slang

dictionary (e.g., “gak” → “tidak”, “bgt” → “banget”). Repeated characters (e.g., “bagussss”) were reduced to their base form, and informal abbreviations were standardized. These normalization procedures enhance data consistency, minimize vocabulary sparsity, and improve the model’s ability to capture representative language patterns.

After the text cleaning and normalization stages were completed, sentiment labels were assigned to each review. Reviews with ratings of 4–5 were classified as positive sentiment, while reviews with ratings of 1–3 were classified as negative sentiment. This rating-based labeling approach assumes that user-provided ratings reflect the overall evaluative orientation expressed in the textual review. The labeling process produces a structured dataset that enables machine learning models to learn language patterns representing sentiment polarity in a systematic manner[19]. This automatic labeling strategy was selected to ensure objective ground truth generation and scalability for large-scale review data.

After preprocessing and labeling, the final dataset consisted of 9,999 reviews, comprising 5,806 positive reviews (58.06%) and 4,193 negative reviews (41.94%). This distribution indicates a moderate dominance of positive sentiment; however, the class proportions remain sufficiently balanced for classification modeling. Stratified sampling was applied during the train–test split to preserve the original class distribution in both subsets

TABEL II
PREPROCESSING RESULTS

Preprocessing Stage	Results
Before Preprocessing	Aplkasi blu simple...mudah diakses dan proses pembukaan rekeningnya cepet bgt. 📌
Data Cleaning	Aplkasi blu simple mudah diakses dan proses pembukaan rekeningnya cepet bgt
Case Folding	aplkasi blu simple mudah diakses dan proses pembukaan rekeningnya cepet bgt
Tokenization	['aplkasi', 'blu', 'simple', 'mudah', 'diakses', 'dan', 'proses', 'pembukaan', 'rekeningnya', 'cepat', 'bgt']
Stopword Removal	['aplkasi', 'blu', 'simple', 'mudah', 'diakses', 'proses', 'pembukaan', 'rekeningnya', 'cepat', 'bgt']

TABEL III
LABELLING RESULTS

Review Text	Rating	Sentiment Label
aplikasi blu simple mudah diakses proses pembukaan rekeningnya cepet bgt	5	Positive

Table 3 shows examples of sentiment labeling results for review data that has undergone preprocessing. Each review text is assigned a sentiment class based on user rating values, resulting in a labeled dataset that is used as training data in the classification process. This labeling enables machine learning

models to learn language patterns that represent sentiment polarity in a structured manner.

D. Classification

Prior to the training process, the labeled data were split into separate sets for training and testing through the hold-out approach. A total of 80% of the data was used as training data to build the classification models, while the remaining 20% was used as testing data to evaluate model performance. This data split allows the models to learn patterns from a substantial portion of the dataset while ensuring that performance is assessed on unseen data to measure generalization ability.

The sentiment classification stage was carried out by applying three algorithms, namely Support Vector Machine (SVM), Convolutional Neural Network (CNN), and IndoBERT. The three models were selected to represent conventional machine learning, deep learning, and transformer-based models in Indonesian text sentiment analysis.

Support Vector Machine (SVM) is a supervised margin-based learning method that determines the optimal separating boundary between two data categories. SVM is known to be effective in handling high-dimensional and nonlinear data through the use of kernel functions, so it is often used in text sentiment analysis[20].

Convolutional Neural Networks (CNN) were employed for their effectiveness in identifying local features within word vector representations through convolution and pooling layers. This model is capable of capturing combinations of words or phrases that effectively represent sentiment polarity. Previous research shows that CNN performs well in classifying Indonesian text sentiment, even with limited datasets[21][22].

In this study, the CNN model was implemented using the Keras Sequential architecture. Text data were first converted into numerical sequences using a tokenizer with a maximum vocabulary size of 10,000 words and an out-of-vocabulary (OOV) token. Each sequence was padded to a fixed length of 100 tokens using post-padding and post-truncation.

The model architecture consisted of an embedding layer with an input dimension of 10,000 and an embedding dimension of 128. This was followed by a one-dimensional convolutional layer with 128 filters and a kernel size of 5 using ReLU activation. A GlobalMaxPooling1D layer was applied to extract the most salient features from the convolution output. The pooled features were then passed to a fully connected dense layer with 64 units and ReLU activation. To reduce overfitting, a dropout layer with a rate of 0.5 was applied. The final output layer consisted of a single neuron with sigmoid activation for binary sentiment classification.

The model was compiled using the Adam optimizer with a learning rate of 0.001 and binary cross-entropy as the loss function. Training was conducted for 6 epochs with a batch size of 128 and a validation split of 0.2. The dataset was

divided into 80% training data and 20% testing data using stratified sampling to preserve class distribution.

IndoBERT is a pretrained language model based on the Transformer architecture that was developed specifically for the Indonesian language. This model is capable of understanding language context in both directions and is effective in handling non-standard language variations and code-mixing phenomena. Research by Fatihaturrahmah and Tania (2025) shows that IndoBERT provides stable performance in classifying the sentiment of digital application reviews containing language variations, so this model was chosen as one of the main approaches in this research[23].

In this study, the pretrained model "indobenchmark/indobert-base-p1" was used as the base architecture for sentiment classification. The tokenizer associated with this model was applied to convert text into tokenized inputs with automatic padding and truncation. The maximum sequence length was set to 128 tokens.

The dataset was divided into 80% training data and 20% testing data using stratified sampling to preserve class distribution. The encoded data were converted into TensorFlow datasets and trained with a batch size of 16. Fine-tuning was performed for 2 epochs.

The model was optimized using the AdamW optimizer with an initial learning rate of $3e-5$ and no warm-up steps. The loss function used was Sparse Categorical Cross-Entropy with logits, and accuracy was employed as the evaluation metric during training. The classification head consisted of a fully connected layer with two output labels corresponding to positive and negative sentiment classes.

E. Evaluation

Model performance evaluation is conducted to measure the ability of each algorithm to classify user review sentiment. The evaluation metrics used include accuracy, precision, recall, and F1-score, which are calculated based on the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values represented in the confusion matrix. These evaluation measures are widely applied in text classification studies to offer a thorough assessment of model effectiveness. The formulation of each metric is presented below.

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

Accuracy indicates the proportion of correct predictions relative to the entire data set.

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

Precision measures how accurately the model predicts positive classes.

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

Recall shows the model's ability to detect all positive data.

$$F1\text{-Score} = \frac{2}{\frac{1}{\text{Recall}} + \frac{1}{\text{Precision}}}$$

The F1-score represents the balance between precision and recall.

F. Visualization

The final stage of the research is the visualization of classification results and interpretation of findings within the framework of Knowledge Discovery in Databases (KDD). A bar chart is used to visualize the comparison of model performance to show the algorithm with the best performance. In addition, WordCloud is used to display dominant words in positive and negative sentiments, thereby providing a deeper understanding of language patterns in user reviews.

The use of WordCloud as a tool for exploring dominant words was also applied in sentiment analysis research on transportation service applications and proved to be capable of providing additional insights into word trends in each sentiment category[24]. Visualization of sentiment analysis results also plays an important role in supporting the process of discovering new patterns and knowledge, as shown in research on mobile-based government applications[25].

III. RESULTS AND DISCUSSION

A. Support Vector Machine (SVM) Classification Results

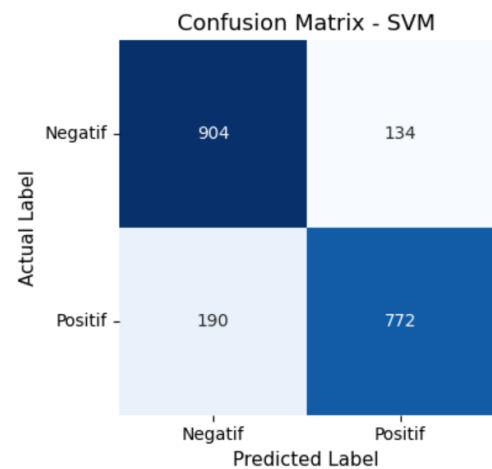


Figure 2. Confusion Matrix of the SVM Model

The SVM confusion matrix model shows that 904 negative data points and 772 positive data points were classified correctly, while 134 negative data points and 190 positive data points were misclassified. These results indicate that the SVM algorithm performs well because the number of correct predictions is greater than the number of errors. Furthermore, the classification report based on the SVM confusion matrix is presented below.



Figure 3. SVM Classification Report Model

The SVM model showed good classification performance with an accuracy of 83.8%. Evaluation by class showed that the model was more sensitive in identifying negative sentiment (precision 0.82; recall 0.87; f1-score 0.85) than positive sentiment (precision 0.85; recall 0.80; f1-score 0.83). The macro average value of 0.84 and weighted average of 0.87 indicate consistent model performance despite the possibility of data distribution imbalance. These results show that SVM is able to capture sentiment patterns in Blu by BCA user reviews quite consistently.

B. Convolutional Neural Network (CNN) Classification Results

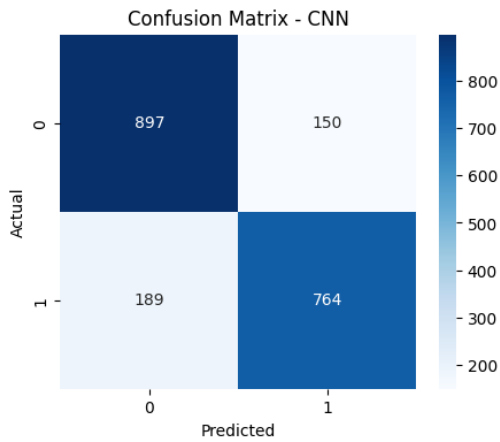


Figure 4. Confusion Matrix of the CNN Model

Based on the confusion matrix, the CNN model shows good classification performance with 897 True Negatives and 764 True Positives, although there are still 150 False Positives and 189 False Negatives. The higher number of correct predictions compared to errors indicates that the CNN model has adequate capabilities in classifying text sentiment.

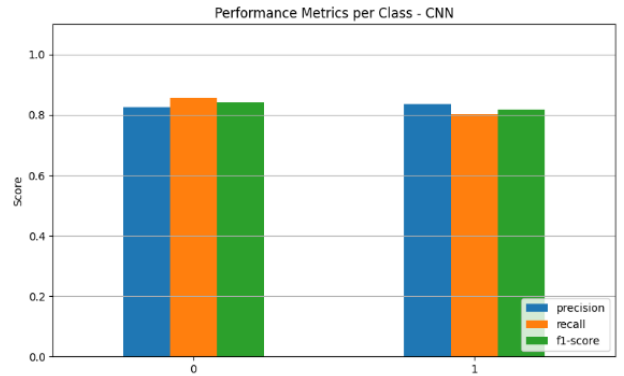


Figure 5. CNN Classification Report Model

The CNN model showed good classification performance with an accuracy of 83.05%. In the negative class, precision was 0.83, recall was 0.86, and the F1 score was 0.84, while in the positive class, they were 0.84, 0.80, and 0.82, respectively. Macro and weighted average values in the range of 0.83 indicate that the model's performance is relatively balanced across classes, suggesting that CNN is capable of consistently capturing sentiment patterns in review data

C. IndoBERT Classification Results

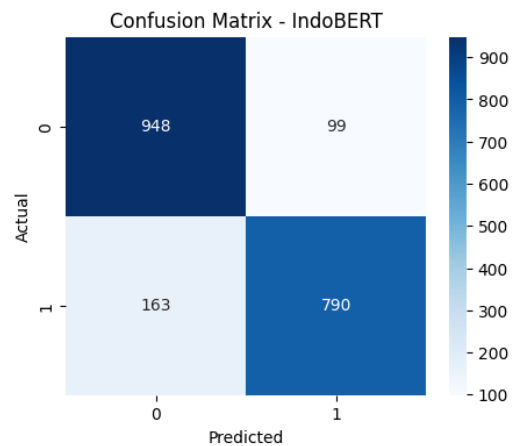


Figure 6. Confusion Matrix of the IndoBERT Model

Based on the confusion matrix, the IndoBERT model shows excellent classification performance with 790 True Positives and 948 True Negatives. The prediction error is relatively low, consisting of 163 false negatives and 99 false positives. Compared to other models, IndoBERT provides higher accuracy and consistency in predictions, especially in recognizing negative classes, demonstrating its superior ability to represent Indonesian language for sentiment analysis.



Figure 7. IndoBERT Classification Report Model

The IndoBERT model shows stable sentiment classification performance in both classes. For negative reviews, the model achieved precision of 0.85, recall of 0.90, and an F1 score of 0.88, while for positive reviews it achieved precision of 0.89, recall of 0.83, and an F1 score of 0.86. These results show that IndoBERT is capable of understanding the context of user reviews well and producing balanced classification performance, thereby effectively supporting the knowledge discovery process through the revelation of sentiment patterns in review data.

D. Comparison of Model Performance

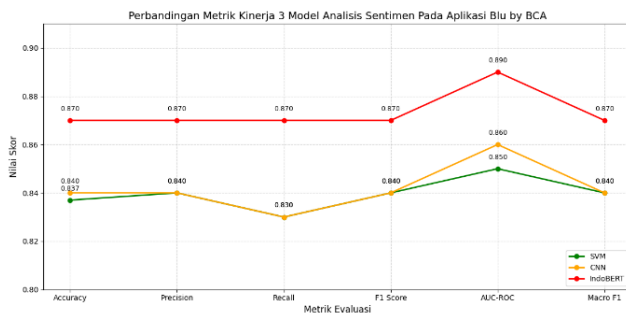


Figure 8. Comparison of Model Performance

Based on six evaluation metrics, the IndoBERT model showed the most consistent and superior performance compared to CNN and SVM, with accuracy, precision, recall, F1-score, and macro F1 values of 0.87, as well as the highest AUC-ROC of 0.89. This value exceeds that of CNN (0.83–0.86) and SVM (0.83–0.85), demonstrating IndoBERT's superior ability to capture linguistic context and semantic relations in Indonesian text. These findings provide empirical evidence that the transformer-based approach is more effective than conventional machine learning methods and other deep learning methods in sentiment analysis tasks. Within the framework of knowledge discovery, these results show differences in model performance characteristics in extracting sentiment patterns from textual data.

Although IndoBERT consistently recorded the best performance across all evaluation metrics, this study did not include formal statistical significance tests to examine whether the observed performance differences between

classifications were statistically significant. Future research could include statistical comparison methods, such as the McNemar test, to strengthen the empirical validation of the model's superiority.

E. WordCloud Visualization

As part of the Knowledge Discovery process, WordCloud visualization is not only used to explore the frequency of dominant words but also to identify broader thematic patterns in user sentiment. In addition to descriptive frequency analysis, dominant terms are interpreted by grouping them into service-related aspects such as transaction efficiency, ease of use, feature completeness, system reliability, and operational stability. This thematic interpretation enhances the practical contribution of the study by linking sentiment patterns to specific areas for improvement in digital banking services.

User sentiment reviews of the Blu by BCA app are categorized into two classes: positive and negative. Based on the labeling results using user ratings, 5,806 reviews were classified as positive sentiment and 4,193 reviews were classified as negative sentiment. This distribution shows the dominance of positive sentiment in the review data analyzed. WordCloud visualizations for each category are presented in Figure 9 and Figure 10.

1) Positive Sentiment WordCloud

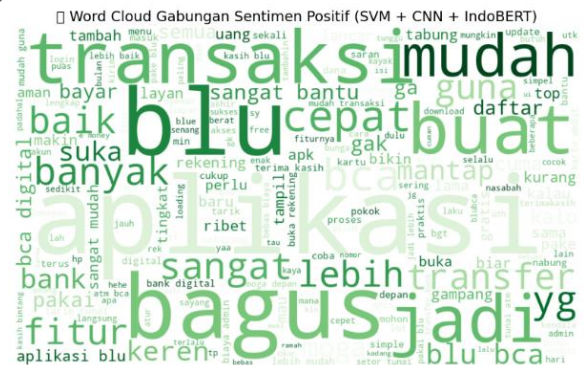


Figure 9. Positive Sentiment WordCloud

The visualization of positive sentiment WordCloud shows the dominance of words such as “transaksi”, “mudah”, “bagus”, “cepat”, and “fitur”, indicating that users consider the Blu app to provide ease in digital financial transactions and is supported by adequate features. The appearance of the words “aman”, “bantu”, and “suka” reflects a positive perception of the convenience and reliability of the service, while the words “blu”, “bca”, and “aplikasi” reinforce users' associations with the identity of the digital product. Some examples of reviews that represent these dominant words include:

- “Transaksi *cepat dan mudah* digunakan.”
- “Aplikasinya *bagus, fitur lengkap dan sangat membantu.*”
- “Blu bikin *transfer jadi lebih simpel dan efisien.*”

WordCloud visualization is part of the knowledge discovery process that presents textual data in visual form, making it easier to identify dominant words in positive sentiment reviews. The dominance of words related to ease of use, speed of service, and application features indicates a positive perception of the Blu by BCA application among users. These findings provide a descriptive overview of user perceptions based on sentiment analysis results.

2) Negative Sentiment WordCloud



Figure 10. Negative Sentiment WordCloud

The WordCloud of negative sentiment in user reviews of Blu by BCA is dominated by words that represent obstacles to use and dissatisfaction, such as “gak”, “masuk”, “transaksi”, “daftar”, “lama”, and “ribet”. These findings indicate that users' main complaints relate to login difficulties, inefficient registration processes, and recurring transaction disruptions, reflecting problems with the reliability of the application service. Some complaints relevant to these findings include:

- “*aplikasi tidak bisa masuk meskipun jaringan stabil,*”
- “*proses pendaftaran memakan waktu lama dan sering gagal,*”
- “*transaksi error dan harus mengulang dari awal.*”

This WordCloud interpretation is part of the knowledge discovery process that aims to systematically reveal problem patterns in text data. The visualization provides an overview of the issues most frequently expressed by users in the context of negative sentiment based on the available text data. Thus, the results of this negative sentiment WordCloud analysis serve as an empirical representation of users' perceptions of their experience using the Blu by BCA app, as recorded in public reviews.

Beyond frequency-based visualization, the dominant terms were further interpreted by grouping them into specific service aspects relevant to digital banking performance. Positive sentiment terms predominantly reflect transaction efficiency, usability, and feature effectiveness, whereas negative sentiment terms are strongly associated with system reliability issues, performance instability, and access-related constraints. This aspect-oriented interpretation enhances the practical relevance of the findings by linking sentiment patterns to specific service dimensions, thereby providing

structured insights for improving system robustness, optimizing transaction processes, and strengthening overall user experience consistency.

F. Discussion

This research proposes the application of the Knowledge Discovery in Databases (KDD) framework to analyze user sentiment reviews of the Blu by BCA application based on public text data. The analysis process was conducted through structured stages, including data collection, preprocessing, sentiment classification, model performance evaluation, and interpretation of results through visualization.

The main contribution of this research lies in comparing three sentiment classification approaches, namely Support Vector Machine (SVM), Convolutional Neural Network (CNN), and IndoBERT, on Indonesian review data. Evaluation using accuracy, precision, recall, and F1-score metrics provides a quantitative overview of the performance characteristics of each model in the context of text sentiment classification.

The SVM model demonstrates higher precision but relatively lower recall, indicating a trade-off in classification behavior. This suggests that the model tends to be conservative in predicting sentiment classes, producing fewer false positives but failing to identify some true instances. Such a pattern often occurs when the model prioritizes prediction certainty, particularly in datasets containing short, ambiguous, or linguistically varied reviews. As a result, certain sentiment expressions are not detected, leading to reduced recall despite maintaining high precision. In contrast, CNN and IndoBERT exhibit more balanced precision and recall values, suggesting better generalization in capturing sentiment instances across diverse review patterns.

This finding is consistent with previous sentiment analysis studies, in which traditional machine learning models such as SVM often achieve high precision but lower recall due to their dependence on explicit feature representation and limited contextual understanding. Deep learning and transformer-based approaches, on the other hand, typically provide more balanced performance by capturing broader semantic relationships within text data. The relatively stable and balanced performance observed in CNN and IndoBERT in this study aligns with prior research emphasizing the advantage of contextual modeling in handling informal language and linguistic variability in Indonesian text classification.

These results have practical implications for decision-makers in digital banking services. High precision indicates that the negative reviews identified by the model are highly reliable, which is useful for detecting clearly expressed user complaints. However, lower recall implies that some user dissatisfaction may remain undetected, potentially leading to an underestimation of service issues. Therefore, relying solely on conservative models may limit the effectiveness of automated feedback monitoring. The superior and more balanced performance of IndoBERT suggests that transformer-based approaches are more suitable for large-

scale monitoring of user feedback in Indonesian-language applications. This enables organizations to implement data-driven evaluation mechanisms, identify recurring issues more comprehensively, and prioritize service improvements based on dominant sentiment patterns.

Beyond methodological implications, the findings of this study also provide strategic insights for improving digital banking service quality. The predominance of positive sentiment related to transaction convenience and usability suggests that maintaining interface simplicity and transaction efficiency should remain a central priority in service development. Conversely, recurring negative sentiment associated with login failures, transaction errors, and registration complexity underscores the need to enhance system reliability, strengthen technical infrastructure stability, and optimize user onboarding processes. By systematically integrating sentiment analysis into service evaluation mechanisms, digital banking institutions can implement proactive and data-driven quality improvement strategies to identify emerging issues, prioritize system refinements, and improve overall customer experience consistency.

Furthermore, the integration of WordCloud visualization within the KDD framework complements quantitative model evaluation by enabling thematic interpretation of dominant sentiment expressions. This combined analytical approach strengthens the knowledge discovery process by linking statistical classification outcomes with interpretable linguistic patterns derived from user-generated content.

Overall, this research presents a measurable and replicable KDD-based sentiment analysis framework, contributing to the development of Indonesian text analysis studies in the domain of digital service applications.

IV. CONCLUSION

The findings demonstrate that sentiment analysis using machine learning and deep learning techniques can systematically reveal user perceptions of the Blu by BCA application while identifying challenges related to service quality. Through the stages of preprocessing, implementation of Support Vector Machine (SVM), Convolutional Neural Network (CNN), and IndoBERT models, and interpretation within the Knowledge Discovery in Databases (KDD) framework, it was found that IndoBERT achieved the highest performance across all evaluation metrics, including accuracy, precision, recall, F1-score, macro-F1, and AUC-ROC. CNN exhibited competitive performance, while SVM showed comparatively lower performance due to its limitations in capturing the contextual complexity of Indonesian review text.

The WordCloud analysis further indicates that positive sentiment is primarily associated with transaction convenience, usability, and feature effectiveness, whereas negative sentiment is related to login difficulties, registration complexity, and transaction disruptions. These findings confirm that the knowledge discovery approach applied to user review data can generate structured sentiment patterns and empirically represent user experience characteristics.

This study is subject to several limitations that should be acknowledged. First, sentiment labeling was performed solely based on user rating values, which may not consistently represent the actual polarity conveyed in the textual content of the reviews. In addition, user-generated reviews may be influenced by self-selection bias, as individuals with particularly positive or negative experiences are generally more motivated to provide feedback. This potential bias may affect the overall sentiment distribution observed in the dataset. Second, the dataset was limited to a specific observation period and focused on a single digital banking application, which may limit the generalizability of the findings. Public opinion expressed on digital platforms is inherently dynamic and may fluctuate over time due to system updates, feature enhancements, service disruptions, or broader technological changes. Furthermore, the analysis was restricted to binary sentiment classification and did not incorporate aspect-based or multi-class sentiment approaches, which could provide more comprehensive and fine-grained insights into user perceptions.

Future research may extend the dataset across multiple digital banking platforms and longer observation periods to capture dynamic sentiment trends. Additionally, the integration of aspect-based sentiment analysis, advanced transformer fine-tuning strategies, and explainable artificial intelligence (XAI) techniques may further enhance model interpretability and deepen the understanding of user perception patterns.

REFERENCES

- [1] T. Takdirmin, F. Ramadhani, I. Ramadhani, Nurfadilah, and V. Dwiastya, "Keuangan Digital dan Fintech: Inovasi, Dampak, dan Tantangan di Era Digital," *Journal of Authentic Research*, vol. 4, no. 1, pp. 338–350, 2025, doi: 10.36312/jar.v4i1.2828.
- [2] P. A. Lestari and M. I. Fasa, "Transformasi Digital Banking: Manfaat dan Risiko Transaksi Online Modern (Internet Banking dan Mobile Banking)," *Jurnal Media Akademik*, vol. 3, no. 4, 2025.
- [3] D. Munandar, M. Afdal, Zarnelly, and R. Novita, "Analisis Sentimen Ulasan Pengguna Aplikasi Mobile Banking Menggunakan Algoritma K-Nearest Neighbor," *Jurnal Teknologi Sistem Informasi dan Aplikasi*, vol. 7, no. 3, pp. 1309–1318, 2024, doi: 10.32493/jtsi.v7i3.41409.
- [4] K. S. Nugroho, A. Y. Sukmadewa, H. Wuswilahaken, F. A. Bachtiar, and N. Yudistira, "BERT Fine-Tuning for Sentiment Analysis on Indonesian Mobile Apps Reviews," in *Proceedings of the International Conference on Sustainable Information Engineering and Technology*, 2021.
- [5] A. Farah and E. R. Syahputra, "Analisis Sentimen terhadap Ulasan Cashless Menggunakan Metode Knowledge Discovery Databases Berbasis Web," *Jurnal Ilmiah Sistem Informasi dan Ilmu Komputer*, vol. 5, no. 3, pp. 198–214, 2025, doi: 10.55606/juisik.v5i3.1507.
- [6] R. Zulfiqri, B. N. Sari, and T. N. Padilah, "Analisis Sentimen Ulasan Pengguna Aplikasi Media Sosial Instagram pada Situs Google Play Store Menggunakan Naïve Bayes Classifier," *JITET (Jurnal Informatika dan Teknik Elektro Terapan)*, vol. 12, no. 3, pp. 2965–2973, 2024.
- [7] P. Ayuningtyas, K. D. Tania, and W. K. Sari, "Sentiment-Based Knowledge Discovery pada Aplikasi iPusnas Menggunakan Metode Machine Learning dan Deep Learning," *Journal of Applied Informatics and Computing*, vol. 9, no. 5, pp. 2486–2497, 2025.

- [8] P. Anggraini and Winarsih, "Komparasi Naïve Bayes, Support Vector Machine, dan Random Forest dalam Analisis Sentimen Aplikasi Shopee di Google Play Store," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 3, pp. 4451–4457, 2025.
- [9] N. Zelina and A. Afyati, "Analisis Sentimen Ulasan Pengguna Aplikasi M-Banking Menggunakan Algoritma Support Vector Machine dan Decision Tree," *Jurnal Linguistik Komputasional*, vol. 7, no. 1, 2024.
- [10] K. Muludi, M. N. Humam, D. A. Shofiana, and A. Syarif, "Perbandingan Kinerja CNN dan Naïve Bayes pada Analisis Sentimen Performa Manchester United di Twitter," *JIEET*, vol. 7, no. 2, pp. 83–91, 2023.
- [11] S. Imron, E. I. Setiawan, and J. Santoso, "Deteksi Aspek Review E-Commerce Menggunakan IndoBERT Embedding dan CNN," *INSYST*, vol. 5, no. 1, pp. 10–16, 2023, doi: 10.52985/insyst.v5i1.267.
- [12] M. B. M. Amin *et al.*, "Deteksi Spam Berbahasa Indonesia Berbasis Teks Menggunakan Model BERT," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 6, pp. 1291–1302, 2024, doi: 10.25126/jtiik.2024118121.
- [13] M. T. Nugraha, N. Sulistiyowati, and U. Enri, "Analisis Sentimen Ulasan Aplikasi Satu Sehat pada Google Play Store Menggunakan Naïve Bayes Classifier," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 5, pp. 3593–3601, 2023.
- [14] A. Hermawan, I. Jowensen, Junaedi, and Edy, "Implementasi Text-Mining untuk Analisis Sentimen pada Twitter dengan Algoritma Support Vector Machine," *Jurnal Sains dan Teknologi*, vol. 12, no. 1, pp. 129–137, 2023, doi: 10.23887/jstundiksha.v12i1.52358.
- [15] M. T. Nugraha, N. Sulistiyowati, and U. Enri, "Analisis Sentimen Ulasan Aplikasi Satu Sehat pada Google Play Store Menggunakan Naïve Bayes Classifier," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 5, pp. 3593–3601, 2023.
- [16] M. Y. Urrochman, H. Asy'ari, and A. Ro'uf, "Aspect-Based Sentiment Analysis of Tumpak Sewu Waterfall Tourist Reviews Using the Naive Bayes Classifier (NBC) Method," *Journal of Informatics Development*, vol. 4, no. 1, 2025, doi: 10.30741/jid.v4i1.1758.
- [17] D. Rifaldi, A. Fadlil, and Herman, "Teknik Preprocessing pada Text Mining Menggunakan Data Tweet 'Mental Health,'" *DECODE: Jurnal Pendidikan Teknologi Informasi*, vol. 3, no. 2, pp. 161–171, 2023, doi: 10.51454/decode.v3i2.131.
- [18] M. R. M. Akbar and I. Pratama, "Analisis Sentimen Ulasan Pengguna Aplikasi MyBCA dengan Naive Bayes, Random Forest, dan Decision Tree," *Sistemasi: Jurnal Sistem Informasi*, vol. 14, no. 5, pp. 2464–2478, 2025.
- [19] P. Arsi and R. Waluyo, "Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM)," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 8, no. 1, pp. 147–156, 2021, doi: 10.25126/jtiik.202183944.
- [20] K.-L. Du, B. Jiang, J. Lu, J. Hua, and M. N. S. Swamy, "Exploring Kernel Machines and Support Vector Machines: Principles, Techniques, and Future Directions," *Mathematics*, vol. 12, no. 24, p. 3935, 2024, doi: 10.3390/math12243935.
- [21] N. Hilmiaji, K. M. Lhaksana, and M. D. Purbolaksono, "Identifying Emotion on Indonesian Tweets using Convolutional Neural Networks," *RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 3, pp. 584–593, 2021, doi: 10.29207/resti.v5i3.3137.
- [22] M. R. Saputra, S. Agustian, Jasril, and Novriyanto, "Klasifikasi Sentimen Pada Dataset yang Terbatas Menggunakan Algoritma Convolutional Neural Network," *Bulletin of Computer Science Research*, vol. 5, no. 4, pp. 522–531, 2025, doi: 10.47065/bulletincsr.v5i4.613.
- [23] A. Fatihaturrachmah and K. D. Tania, "A Hybrid Approach of Aspect-Based Sentiment Analysis and Knowledge Extraction for Evaluating Security Perceptions in Digital Payment Applications," *Scientific Journal of Informatics*, vol. 12, no. 4, pp. 635–650, 2025, doi: 10.15294/sji.v12i4.31557.
- [24] J. A. Wibowo, V. C. Mawardi, and T. Sutrisno, "Visualisasi Word Cloud Hasil Analisis Sentimen Berbasis Fitur Layanan Aplikasi Gojek dengan Support Vector Machine," *Jurnal Serina Sains, Teknik dan Kedokteran*, vol. 2, no. 1, pp. 61–70, 2024, doi: 10.24912/jsstk.v2i1.32058.
- [25] R. B. Hamid, C. Andriansyah, D. I. Sensuse, S. Lusa, D. Elisabeth, and N. Safitri, "Sentiment Analysis and Topic Modeling for Discovering Knowledge in Indonesian Mobile Government Applications," *Jurnal Teknik Informatika (JUTIF)*, vol. 6, no. 5, pp. 3188–3203, 2025, doi: 10.52436/1.jutif.2025.6.5.4991.