



Sentiment Classification on Indodax Using Term Frequency, FastText, and Neural Attention Models

Dedy Hartama¹, Ginanti Riski²

¹Department of Information Systems, ²Department of Informatics Engineering, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

¹dedyhartama@amiktunasbangsa.ac.id, ²ginantiriski@gmail.com

Abstract

The rapid growth of mobile-based investment platforms such as Indodax has triggered a surge in user-generated reviews that reflect public perception and sentiment. This study aimed to develop and evaluate sentiment classification models that can accurately classify Indonesian user reviews on the Indodax app into negative, neutral, and positive sentiments. A dataset of 11,000 reviews was collected via web scraping from the Google Play Store. Reviews were preprocessed, labeled using a lexicon-based unsupervised method, and balanced using oversampling. Two models were built: a Bidirectional LSTM (BiLSTM) with attention mechanism using FastText embeddings, and a Feedforward Neural Network (FFNN) using a hybrid feature vector combining TF-IDF and FastText. The evaluation was performed using accuracy, classification report, confusion matrix, and PCA visualization. The FFNN model outperformed the BiLSTM-Attention model with an accuracy of 97.07% compared to 96.00%. Both models demonstrated strong performance in classifying three sentiment classes, though the FFNN showed better separation in PCA space and higher macro-average metrics. This study demonstrates the effectiveness of combining statistical and semantic feature representations for sentiment classification in Indonesian text. The proposed approach is particularly valuable for low-resource languages and informal user-generated content.

Keywords: fasttext embedding; hybrid feature representation; lexicon-based labeling; sentiment analysis; BiLSTM attention

How to Cite: D. Hartama and G. Riski, "Sentiment Classification on Indodax Using Term Frequency, FastText, and Neural Attention Models", *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 9, no. 5, pp. 1173 - 1182, Oct. 2025.
Permalink/DOI: <https://doi.org/10.29207/resti.v9i5.6871>

Received: June 30, 2025

Accepted: August 21, 2025

Available Online: October 25, 2025

*This is an open-access article under the CC BY 4.0 License
Published by Ikatan Ahli Informatika Indonesia*

1. Introduction

Amid the rapid growth of the digital economy In Indonesia, cryptocurrency trading through platforms such as Indodax has become increasingly popular, especially among the younger generation and retail investors. This phenomenon has triggered a surge of public opinion and sentiment widely disseminated in cyberspace, particularly through user reviews and social media. According to Bappebti data as of January 2024, the number of crypto investors in Indonesia has reached 18.83 million people [1], while the national transaction value from January to November 2024 soared to IDR 556.63 trillion—an increase of approximately 356% compared to the previous year [2]. This significant growth indicates that crypto is no longer a fleeting trend but has become part of the financial lifestyle of Indonesian society. In this context, sentiment analysis becomes highly relevant for understanding consumer perceptions and emotions toward crypto platform

services. This research stands out by combining linguistic feature representations and artificial intelligence techniques to examine local data an area still rarely explored academically while also addressing the challenges of extracting meaning from Indonesian texts that are often complex and context-sensitive.

Previous studies have explored sentiment analysis in financial and cryptocurrency applications using both classical and deep learning approaches. Classical machine learning techniques such as Naive Bayes and SVM have demonstrated strong performance in sentiment classification tasks. For instance, studies on Pluang and DANA app reviews achieved up to 99.50% and 87% accuracy, respectively. Similar performance was observed in domains such as cryptocurrency tweets and gaming reviews, with accuracies ranging from 80% to 94%[3-7].

In terms of feature representation for Indonesian text, FastText embeddings have shown strong suitability. A

study on Telkomsel product reviews achieved an F1-score above 95%, highlighting its effectiveness for local language processing [8]. This embedding technique has also been adopted in hybrid models, such as CNN-BiLSTM, for detecting cyberbullying and hate speech, yielding accuracies around 80% [9].

As data complexity increases, deep learning models are gaining traction. For example, a BiLSTM combined with Word2Vec achieved up to 87% accuracy in public sentiment analysis, underscoring the value of capturing sequential context [10]. However, the addition of attention mechanisms to BiLSTM did not consistently improve performance across different evaluation settings [11]. Meanwhile, hybrid deep learning architectures—such as CNN-LSTM—have shown stable performance on Twitter sentiment classification [12]. Combining LSTM, VADER, and TF-IDF also outperformed standalone models, suggesting a synergistic benefit from integrating lexical and neural-based methods [13].

Beyond binary sentiment classification, recent studies have extended these methods to broader NLP tasks, including multi-label classification. In Indonesian news categorization, for instance, SVM achieved 94.24% accuracy while LSTM reached 95%, underscoring their robustness and generalizability across text classification problems [14].

These gaps include: (1) a lack of focus on crypto-specific platforms like Indodax, (2) limited evaluation of hybrid methods combining statistical and semantic features, and (3) a scarcity of studies comparing deep learning architectures with hybrid models within a consistent experimental framework. This leaves a significant gap in the context of Indodax, the largest crypto trading platform in Indonesia, which has unique characteristics and a distinct user base. Additionally, many previous studies evaluate only one type of approach—either classical methods such as TF-IDF and Naive Bayes or deep learning like BiLSTM—without directly comparing their strengths and weaknesses within a consistent study. To the best of the authors' knowledge, this study is the first to explore and directly evaluate a hybrid feature representation combining TF-IDF and FastText specifically for sentiment classification in the Indonesian cryptocurrency context. Moreover, despite the growing popularity of BiLSTM with attention, its comparative performance against hybrid approaches has rarely been evaluated in a systematic manner that combines classical and modern features.

This study is the first to directly compare a hybrid feature-based method and a BiLSTM-Attention model for analyzing Indonesian-language reviews on the Indodax platform. Thus, the objective of this research is to systematically compare the classification performance of a hybrid TF-IDF + FastText + FFNN model and a BiLSTM-Attention model on Indonesian-language sentiment data from Indodax. Specifically, it aims to address the following research question: *How*

does the performance of the hybrid approach compare to that of BiLSTM-Attention in classifying user opinions on Indodax? These two approaches will be systematically evaluated using performance metrics such as accuracy, precision, recall, and F1-score to determine which method performs better in classifying Indonesian-language opinions in the crypto sector. Through this comprehensive analysis, the study seeks not only to enrich the literature on natural language processing (NLP) in the local context but also to provide practical contributions in the form of methodological recommendations that can be applied in developing user opinion analytics systems for Indonesian digital financial platforms. The proposed hybrid approach is expected to better capture the semantic nuances of the Indonesian language and be more relevant to the needs of the evolving financial technology industry.

By addressing these gaps, this study introduces a novel methodology that combines a hybrid feature representation integrating TF-IDF and FastText embeddings with a lexicon-based sentiment labeling approach specifically adapted for the Indonesian language. Unlike previous research that typically employs either classical statistical methods or deep learning models separately, this combined feature and labeling strategy aims to improve classification accuracy and better handle the unique challenges of processing Indonesian text within the cryptocurrency domain. Thus, this research not only applies existing models but also develops and evaluates methods tailored to the local context and the specific characteristics of user-generated data on the Indodax platform, contributing to more accurate and context-sensitive sentiment classification systems for the Indonesian cryptocurrency ecosystem.

2. Methods

2.1 Research Dataset

The dataset in this study was obtained through a web scraping process on the Indodax application page in the Google Play Store [15]. From this page, the application ID “*id.co.bitcoin*” was used as the main parameter to access the user review data. The data collection was conducted using the *google_play_scraper* library with language and country settings set to Indonesian. The *reviews()* function was executed with the parameter *count=11000*, indicating that the system attempted to retrieve up to 11,000 of the most relevant user reviews.

A total of 11,000 valid user reviews were collected and used in this research. The features extracted from each review include the username (*userName*), rating score (*score*), timestamp of the review (*at*), and the review text content (*content*). These four features serve as the basis for the sentiment classification process in the subsequent stages of the study. Each of these attributes plays a critical role: *the content* provides the textual data for determining sentiment polarity, *the score* acts as the ground truth label for supervised learning, *the*

timestamp enables temporal trend analysis, and *the username* can assist in detecting repeated or potentially spammy reviews in future research.

To ensure the representativeness of the dataset, the reviews were sorted by their timestamp, allowing the inclusion of both recent and older reviews to capture temporal diversity. Moreover, no filtering was applied to specific score ratings, ensuring that user reviews across all sentiment levels (from 1 to 5 stars) were included. This approach helps to maintain the diversity of user opinions and supports a balanced sentiment analysis. In addition, no filtering was applied to specific users, allowing the dataset to reflect various types of users with different backgrounds and preferences.

All user review data collected in this study are publicly available on the Google Play Store. No personally identifiable information was retrieved, and the data were used solely for academic research purposes in accordance with ethical research standards. Therefore, explicit user consent was not required as the data were obtained from a public domain.

2.2 Research Stages

This study was designed with a structured sequence of stages to ensure that each process is conducted systematically and contributes to improving model performance. The step-by-step workflow is illustrated in Figure 1.

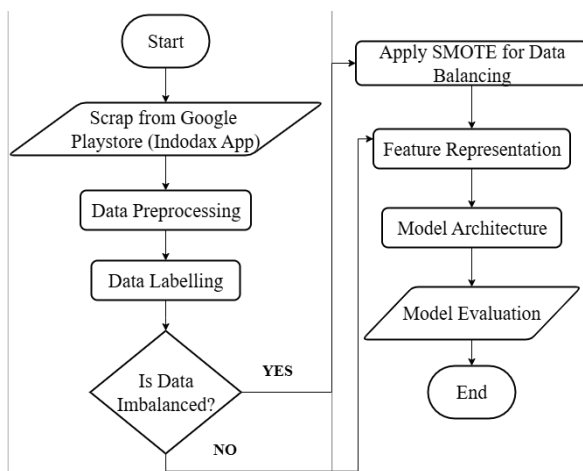


Figure 1. Research Stages

2.3 Data Preprocessing

The preprocessing stage was carried out to clean and prepare the review data before the classification process. The first step was cleaning, which included removing URLs, HTML tags, emojis, non-alphanumeric symbols, numbers, and converting text to lowercase. This was performed using sequential regular expression functions to reduce noise in the data [16], [17].

Next, normalization was applied by replacing informal or non-standard words with their standard equivalents using a predefined dictionary. This step was important to handle the informal language often found in

Indonesian user reviews. The process continued with tokenization, where the text was split into individual words for further analysis.

Then, stopwords removal was implemented to eliminate common words that do not significantly contribute to classification, based on the Indonesian stopwords list from the NLTK library. Finally, stemming was performed using the Sastrawi library to convert words to their root forms, ensuring a more consistent representation in the model.

2.4 Data Labeling

In this study, a lexicon-based sentiment labeling technique was employed to categorize the review data into three sentiment classes: positive, negative, and neutral. This approach employs a frequency-based heuristic, aligned with prior studies [18] - [20], in which high-frequency terms are considered indicative of sentiment relevance within a domain-specific corpus.

First, all tokens from the preprocessed text were aggregated, and the top 150 most frequent words were extracted. These were then manually divided into three sets: the top 50 most frequent words were treated as indicators of positive sentiment, the middle 50 as negative sentiment indicators, and the remaining as neutral.

A labeling function was then applied to count how many tokens from each review matched the positive or negative lexicon. If the positive matches exceeded the negative, the review was labeled as positive (2); if negative matches were higher, it was labeled as negative (0); otherwise, it was classified as neutral (1). This unsupervised technique enabled efficient and consistent sentiment annotation for the entire dataset without manual intervention.

However, this lexicon-based unsupervised labeling method has several potential limitations. Since the sentiment labeling relies solely on the frequency of individual tokens, it may fail to capture nuanced or complex sentiments, such as sarcasm, irony, or context-dependent expressions. It also assumes that word frequency directly reflects sentiment polarity, which may not always hold true across diverse linguistic patterns. As a result, some reviews may be misclassified, particularly those with mixed or ambiguous sentiment. These limitations highlight the need for caution in interpreting the labeling results and suggest that future work may benefit from incorporating more sophisticated or hybrid labeling techniques.

2.5 Oversampling With SMOTE

To address the issue of class imbalance in the sentiment distribution, this study employed an oversampling strategy using the SMOTE (Synthetic Minority Over-sampling Technique) approach. The dataset initially exhibited a disproportionate number of reviews across the three sentiment classes: positive, neutral, and negative. To ensure balanced training data, synthetic samples were generated for the minority classes (neutral

and negative) until each class contained 9,170 instances.

SMOTE was chosen over random undersampling to avoid the loss of valuable information from the majority class, and over simple bootstrapping to introduce greater diversity in the synthetic samples. Unlike basic resampling, SMOTE generates new examples by interpolating between existing instances, which can lead to more robust decision boundaries during model training. Previous studies have also shown that synthetic oversampling methods like SMOTE can significantly improve classifier performance on imbalanced datasets [21], [22].

The resulting balanced dataset was then used for feature representation and model training, aiming to reduce bias toward the dominant class and enhance the model's generalization capabilities.

2.6 Feature Representation

Feature representation was carried out using two methods: TF-IDF and word embeddings. The TF-IDF approach extracted statistical features from stemmed tokens, resulting in 300-dimensional vectors [23]. In parallel, FastText was employed to generate 300-dimensional semantic embeddings based on subword information. The embedding vectors for each review were computed by averaging the word embeddings of its tokens.

These two representations were then concatenated into a 600-dimensional feature vector, which was used as input for the classification model. To further capture sequential patterns and highlight important words in the text, a Bidirectional LSTM architecture combined with an attention mechanism was utilized [24].

2.7 Model Architecture

The classification model was developed using two different approaches: a feedforward neural network architecture and a Bidirectional LSTM architecture with an attention mechanism. Both models received a 600-dimensional feature vector as input, formed by concatenating TF-IDF features (300 dimensions) and averaged FastText embeddings (300 dimensions).

The choice of 300-dimensional vectors for both the FastText and TF-IDF components was based on standard practices in prior research and supported by empirical evidence. Pretrained FastText models typically employ 300-dimensional embeddings—such as the *cc.id.300.vec* model developed by Facebook AI Research—because this dimensionality provides an optimal balance between semantic richness and computational efficiency. Previous studies have also shown that increasing the embedding size beyond 300 yields diminishing returns while incurring higher computational costs. Conversely, using lower-dimensional vectors risks a loss in representational fidelity [25].

The first model consisted of three dense layers: 256 units with ReLU activation followed by a 50% dropout, 128 units with ReLU activation followed by a 30% dropout, and a softmax output layer with 3 units for sentiment classification. The second model utilized pretrained FastText embeddings, processed through a Bidirectional LSTM layer with 64 units, followed by an attention layer, global average pooling, and a dense layer with 64 ReLU units before reaching the softmax output.

Both models were compiled using the categorical cross-entropy loss function and the Adam optimizer. To prevent overfitting and enhance model generalization, dropout and an ensemble approach were adopted as recommended in previous research [26].

2.8 Model Evaluation

The performance of the classification models was evaluated using quantitative metrics such as accuracy, precision, recall, and F1-score, along with a qualitative analysis via Principal Component Analysis (PCA) for feature visualization. The evaluation was conducted across three sentiment classes: negative, neutral, and positive.

The classification metrics were computed based on the values of true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN) derived from the confusion matrix. The formulas used are as shown in Formulas 1 to 4 [27].

Accuracy measures the proportion of correct predictions to the total number of predictions:

$$Accuracy = \frac{(a+d)}{Total\ Sampel} \times 100\% \quad (1)$$

Precision is the ratio of true positive predictions to the total predicted positive cases:

$$Precision = \frac{TP}{(TP+FP)} \quad (2)$$

Recall (or sensitivity) is the ratio of true positive predictions to the actual number of positive cases:

$$Recall = \frac{TP}{(TP+FN)} \quad (3)$$

F1-score is the harmonic mean of precision and recall, balancing both metrics:

$$1 - Score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \quad (4)$$

These evaluation metrics provided a comprehensive view of how well each model performed in distinguishing between sentiment classes.

3. Results and Discussions

3.1 Data Collection

A total of 11,000 user reviews were successfully collected from the Indodax application via web scraping and then imported into Google Colaboratory for further processing and analysis. The dataset includes four key

features: userName, score, at, and content, which serve as the foundation for the sentiment classification task.

Figure 2 displays the initial view of the dataset in Google Colab, showing the first few rows along with their respective columns and values.

	userName	score	at	content
0	Pengguna Google	5	2025-04-12 09:07:31	Aplikasi yg bagus dan mudah digunakan. sambil ...
1	Pengguna Google	2	2025-04-12 01:26:13	setelah. memakai aplikasi ini ternyata banyak ...
2	Pengguna Google	1	2025-04-11 16:02:03	Sistem apaan order book, saya aja dri exc plua...
3	Pengguna Google	5	2025-04-11 14:27:39	mantab bgt, indodax apk sebuah opsi btw milenial...
4	Pengguna Google	1	2025-04-11 11:49:43	aplikasi jelek, deposit lewat bank mandiri ga ...

Figure 2. Importing Data to Google Colaboratory

3.2 Data Preprocessing

After the dataset was loaded into Google Colab, duplicate entries were removed, resulting in 10,987 unique reviews. This step minimized potential bias caused by repeated data. The text was then preprocessed through a series of steps to transform raw, unstructured reviews into clean and consistent input for feature extraction and modeling.

Table 1 presents samples of raw review texts before preprocessing. The reviews still contain URLs, emojis, punctuation, and inconsistent capitalization.

Table 1. Raw Review Text Before Preprocessing

Index	Content
1	mantab bgt, indodax apk sebuah opsi btw milenial dan kaum Gen z, platform investasi simple praktis utk mmbngun Sebuah aset di era globalisasi 🍷 ID : Ody Mazz
2	Sangat cocok untuk pumula yang jngin memulai deposit yang rendah, saya sangat suka dengan ada nya indodax info dan berita sehingga bisa membantu mengetahui update-update terbaru 🍷 @prassss #AsetMasaDepan

The first stage, cleaning, removed irrelevant elements such as special characters, numbers, emojis, HTML tags, and converted all text to lowercase. As shown in Table 2, the review texts became simpler and more uniform.

Table 2. Review Text After Cleaning

Index	Content
1	mantab bgt indodax apk sebuah opsi btw milenial dan kaum gen z platform investasi simple praktis utk mmbngun sebuah aset di era globalisasi id ody mazz
2	sangat cocok untuk pumula yang jngin memulai deposit yang rendah saya sangat suka dengan ada nya indodax info dan berita sehingga bisa membantu mengetahui updateupdate terbaruprassss asetmasadepan

Next, normalization replaced informal or misspelled words with their formal equivalents, improving consistency especially in user slang. An example is shown in Table 3.

The tokenization stage split each review into individual words, as shown in Table 4.

Table 3. Review Text After Normalization

Index	Content
1	mantab banget indodax aplikasi sebuah opsi btw milenial dan kaum gen saja platform investasi simple praktis untuk membangun sebuah aset di era globalisasi id ody mazz
2	sangat cocok untuk pemula yang jngin memulai deposit yang rendah saya sangat suka dengan ada ya indodax info dan berita sehingga bisa membantu mengetahui update terbaru prassss aset masa depan

Table 4. Tokenized Review Text

Index	Content
1	mantab, banget, indodax, aplikasi, sebuah, opsi, btw, milenial, dan, kaum, gen, saja, platform, investasi, simple, praktis, untuk, membangun, sebuah, aset, di, era, globalisasi, id, ody mazz
2	sangat, cocok, untuk, pemula, yang, jngin, memulai, deposit, yang, rendah, saya, sangat, suka, dengan, ada, ya, indodax, info, dan, berita, sehingga, bisa, membantu, mengetahui, update, terbaru, prassss, aset, masa, depan

After the tokenization process, stopword removal was performed to eliminate common words (e.g., 'yang', 'dan', 'di') that do not carry significant meaning for sentiment analysis. The results of this process are presented in Table 5.

Table 5. Tokens After Stopword Removal

Index	Content
1	['mantab', 'banget', 'indodax', 'aplikasi', 'opsi', 'btw', 'milenial', 'kaum', 'gen', 'saja', 'platform', 'investasi', 'simple', 'praktis', 'membangun', 'aset', 'di', 'era', 'globalisasi', 'id', 'odymazz']
2	['cocok', 'pemula', 'jngin', 'deposit', 'rendah', 'saya', 'suka', 'ya', 'indodax', 'info', 'berita', 'membantu', 'mengetahui', 'update', 'terbaru', 'prassss', 'aset', 'masa', 'depan']

Finally, stemming was applied using the Sastrawi stemmer to reduce each word to its root form. This minimized vocabulary redundancy and improved consistency. The stemmed tokens are shown in Table 6.

Table 6. Final Tokens After Stemming

Index	Content
1	mantab banget indodax aplikasi opsi btw milenial kaum gen platform investasi simple praktis membangun aset di era globalisasi id ody mazz
2	cocok pemula jngin deposit rendah saya suka ya indodax info berita membantu mengetahui update terbaru prassss aset masa depan

These preprocessing steps produced standardized and cleaner review data, which were then ready for feature representation and model training in the next phase.

3.3 Data Labeling

To automatically assign sentiment categories, this study employed a lexicon-based method grounded in token frequency. The 150 most frequent words—after cleaning and stopword removal—were selected and divided into three categories: the top 50 were assumed to reflect positive sentiment, the middle 50 negative sentiment, and the bottom 50 neutral.

Each review was then labeled by comparing the number of matching words in the positive and negative sets. A

review was labeled as positive (2) if it contained more positive tokens, negative (0) if it contained more negative tokens, and neutral (1) if the counts were equal or none matched.

Table 7 presents the sentiment distribution of the 10,987 reviews labeled using this approach.

Table 7. Sentiment Label Distribution

Sentiment	Label	Count
Positive	2	9,170
Neutral	1	1,025
Negative	0	792

For illustration, Table 8 presents two sample reviews that were preprocessed and labeled as positive sentiment.

Table 8. Sample Reviews After Labeling

Content	Sentiment
mantab banget indodax aplikasi opsi btw milenial kaum gen platform investasi simple praktis membangun asset di era globalisasi id odyamaz cocok pemula ingin deposit rendah saya suka ya indodax info berita membantu mengetahui update terbaru prassss asset masa depan	2

3.4 Oversampling With SMOTE

The initial sentiment distribution was highly imbalanced, with the majority of reviews labeled as positive (2), while neutral (1) and negative (0) labels were underrepresented. Figure 3 illustrates this imbalance.

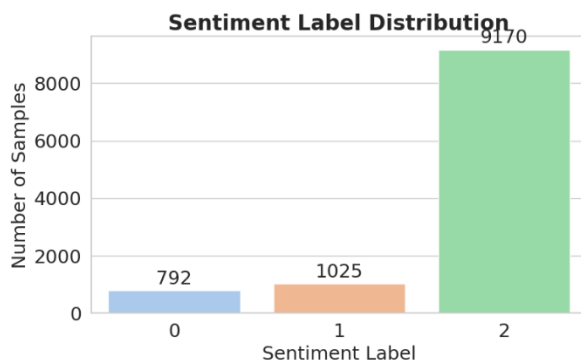


Figure 3. Sentiment Label Distribution Before SMOTE

To address this issue, the Synthetic Minority Oversampling Technique (SMOTE) was applied. SMOTE generates synthetic samples for the minority classes based on feature similarity. After applying SMOTE, the dataset was balanced, with each sentiment class containing 9,170 entries.

The post-SMOTE label distribution is shown in Table 9, indicating that all three sentiment classes now contain an equal number of samples.

Table 9. Data Distribution After SMOTE

Sentiment	Label	Count
Positive	2	9,170
Neutral	1	9,170
Negative	0	9,170

This balanced distribution ensures that the model can learn fairly across all sentiment classes during training.

3.5 Feature Representation

In the feature representation stage, the stemmed review texts were transformed into numerical vectors to be used as input for sentiment classification models. The first approach applied was TF-IDF (Term Frequency – Inverse Document Frequency), which produced 300-dimensional vectors based on the most significant words in the corpus. This method emphasizes words that frequently appear in individual reviews but are rare across others, helping the model distinguish between different contexts.

Next, pretrained FastText word embeddings in the Indonesian language were employed. Each word was represented as a 300-dimensional vector, and sentence-level representations were obtained by averaging the vectors of all words within a review. This approach enabled the model to capture the semantic meaning of entire sentences more effectively than purely statistical methods.

To enrich the representation, the outputs from both TF-IDF and FastText were concatenated into a single 600-dimensional vector. This hybrid vector was then used as input to a dense-layer neural network, with dropout regularization applied to prevent overfitting.

Alternatively, the review texts were also represented as token sequences and processed using a Bidirectional LSTM model. This model utilized the FastText embedding layer and was equipped with an attention mechanism, allowing it to focus on key words within each sentence. The output of the attention layer was aggregated using global average pooling, followed by dense layers for sentiment prediction. This approach accounts not only for word frequency but also the contextual relationships between words in a sequence.

3.6 Model Architecture

In this study, two model architectures were implemented to classify sentiment in user reviews from the Indodax platform. The first model utilizes a *Bidirectional Long Short-Term Memory* (BiLSTM) network enhanced with an attention mechanism. This model captures the sequential and contextual meaning of words by processing tokenized and padded inputs through a fixed embedding layer using pre-trained 300-dimensional Indonesian FastText vectors. The BiLSTM layer reads the input in both forward and backward directions, and the attention mechanism assigns greater weight to important words. The resulting context vector is passed through global average pooling, dropout, and dense layers, and finally through a softmax layer to classify reviews into three sentiment categories: negative, neutral, or positive.

The second model employs a *Feedforward Neural Network* (FFNN) based on fixed feature representations. It combines 300-dimensional TF-IDF vectors with 300-dimensional averaged FastText

vectors, resulting in a 600-dimensional input. This input is processed by two dense layers (with 256 and 128 neurons, respectively), each followed by dropout for regularization. The output layer is a softmax classifier that predicts one of the three sentiment classes.

By applying these two distinct approaches, the study enables a performance comparison between context-aware sequence modeling and fixed feature-based classification. The full architectures of both models are illustrated in Figure 4.

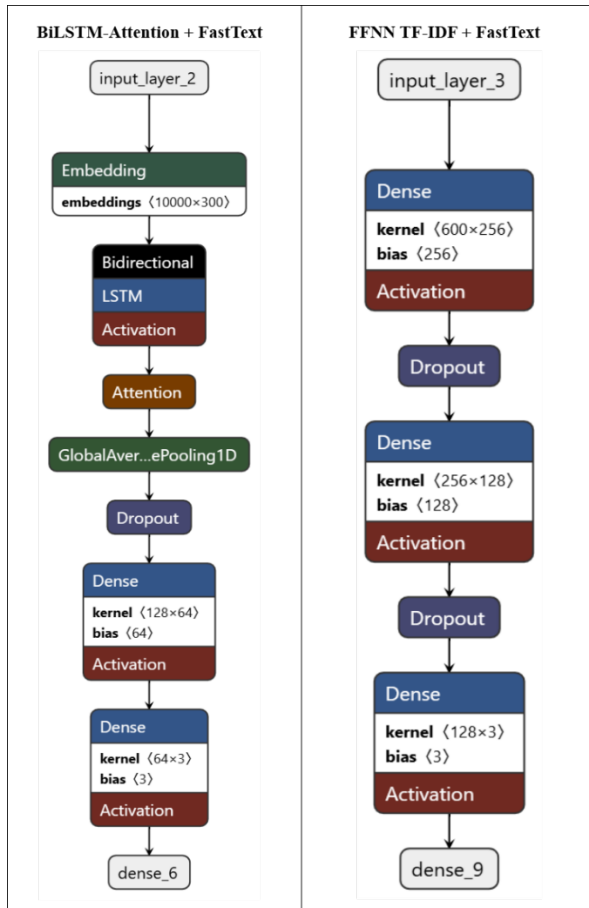


Figure 4. Model architectures: BiLSTM-Attention with FastText and FFNN with TF-IDF + FastText

3.7 Model Evaluation

The performance of both models was evaluated using accuracy, confusion matrix, classification report, and PCA visualization. The classification task includes three sentiment categories: negative, neutral, and positive. The results are summarized as follows.

The BiLSTM-Attention model utilizing FastText embeddings achieved an accuracy of 96%. Its classification performance is illustrated in Figure 5, which presents the classification report containing the precision, recall, and F1-score for each sentiment class.

As shown in Figure 5, the model demonstrated balanced performance across all classes. The negative class achieved a precision of 0.96, a recall of 0.99, and an F1-score of 0.97. The neutral class attained a precision of 0.93, a recall of 0.97, and an F1-score of 0.95.

Meanwhile, the positive class recorded a precision of 0.99, a recall of 0.92, and an F1-score of 0.95. These metrics indicate the model's ability to consistently identify sentiments across different categories, with a macro average of 0.96 for all evaluation metrics.

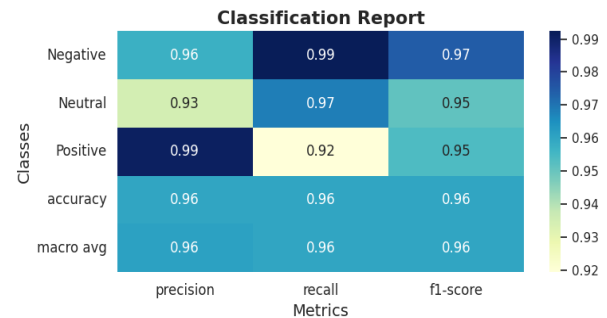


Figure 5. Classification Report BiLSTM-Attention with FastText

Further insights into the model's prediction distribution are provided in Figure 6, which displays the confusion matrix.

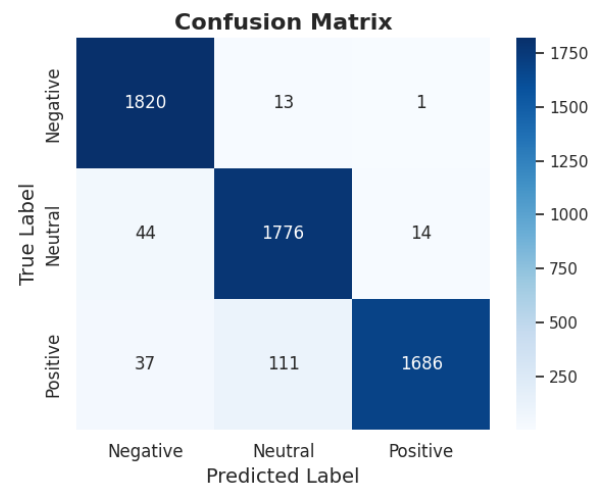


Figure 6. Confusion Matrix BiLSTM-Attention with FastText

The matrix reveals that the model accurately classified 1820 out of 1834 samples in the negative class. For the neutral class, 1776 predictions were correct, although some misclassifications occurred, with 44 instances labeled as negative and 14 as positive. The positive class also performed well, with 1686 correctly predicted samples, although 111 were misclassified as neutral and 37 as negative.

Meanwhile, The FFNN model utilizing combined TF-IDF and FastText features achieved a classification accuracy of 97,07%. Figure 7 displays the classification report, highlighting strong performance across all sentiment classes.

Specifically, the model reached a precision of 0.97 for the Negative class, with a recall of 0.99 and an F1-score of 0.98. The Neutral class showed a precision of 0.95, a recall of 0.99, and an F1-score of 0.97. Meanwhile, the Positive class achieved the highest precision of 1.00, a recall of 0.93, and an F1-score of 0.96. These scores indicate that the model maintains a balanced and high-

quality classification capability, with a macro average of 0.97 across all metrics.

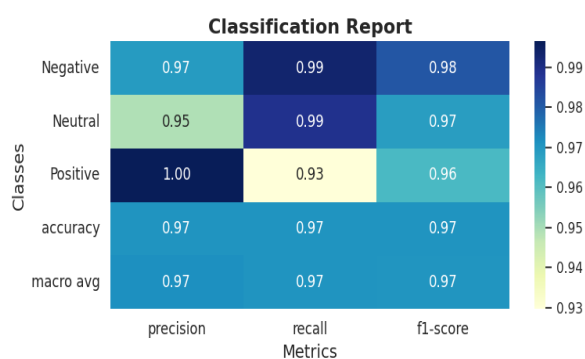


Figure 7. Classification Report FFNN with TF-IDF + FastText

Figure 8 presents the confusion matrix, which further illustrates the model's predictive distribution.

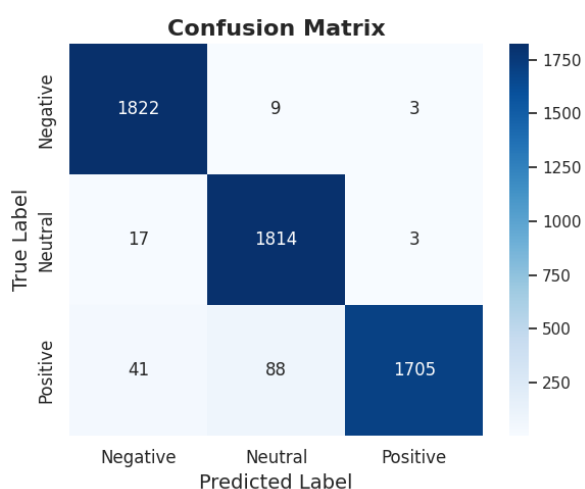


Figure 8. Confusion Matrix FFNN with TF-IDF + FastText

The Negative class was predicted almost perfectly, with only 12 misclassifications out of 1834 instances. For the Neutral class, 17 samples were incorrectly labeled as Negative, and 3 as Positive. The Positive class experienced 88 misclassifications as Neutral and 41 as Negative, which aligns with the slightly lower recall observed in the classification report.

Overall, the confusion matrix highlights that most classification errors occurred between the *neutral* and *positive* classes, which often share overlapping semantic characteristics. Despite this, both models—BiLSTM-Attention and FFNN—exhibited strong and consistent classification capabilities across all sentiment categories, demonstrating their effectiveness in handling sentiment classification tasks with high accuracy.

The frequent confusion between neutral and positive sentiment classes likely stems from lexical and contextual overlap in user-generated content. Reviews labeled as *neutral* often contain mildly positive expressions or vague praise without strong sentiment indicators, making them semantically close to *positive* reviews. Additionally, certain informal words or

expressions (e.g., "*oke*", "*lumayan bagus*", "*cukup membantu*") may ambiguously fall between neutral and positive, leading to misclassification.

To mitigate this issue, future improvements could include fine-tuning word embeddings with sentiment-specific corpora, incorporating syntactic or semantic rules to disambiguate borderline cases, or employing transformer-based models such as BERT, which are more capable of capturing nuanced sentiment within context. Furthermore, expanding the training dataset with more clearly defined neutral examples may enhance class separability and reduce ambiguity.

3.8 Research Contribution

This research contributes to the growing body of literature on sentiment analysis in financial and cryptocurrency domains, particularly by applying and comparing modern neural architectures—namely BiLSTM with attention mechanism and feedforward neural networks—on Indonesian language user comments from Indodax. The study offers several key contributions:

First, it introduces a novel implementation of hybrid feature representations that combine term frequency and FastText embeddings, demonstrating their effectiveness in capturing both lexical and semantic information within noisy and informal financial text data. This combination enhances classification performance, particularly in differentiating semantically overlapping sentiment classes such as neutral and positive.

Second, the use of the attention mechanism within the BiLSTM architecture significantly improves the model's ability to focus on relevant contextual cues, which is especially valuable in analyzing short texts like social media comments or platform-based reviews. The results highlight the advantages of neural attention in sentiment classification tasks, showing superior precision and recall across all sentiment classes.

Third, this study contributes an empirical benchmark for sentiment classification on Indonesian cryptocurrency comments—an area that remains underexplored. The dataset derived from Indodax represents real-world financial discourse in Bahasa Indonesia, and the annotated dataset and model evaluation results can serve as a useful reference for future work in fintech-related natural language processing.

Finally, this research provides practical insights for financial technology platforms and investors. By enabling accurate sentiment classification, the models developed can potentially be integrated into sentiment-aware decision support systems, helping stakeholders to monitor public perception, detect sentiment shifts, and make informed investment decisions based on user feedback in real time.

4. Conclusions

This study aimed to evaluate and compare the performance of two neural network architectures—Bidirectional LSTM with attention mechanism and Feedforward Neural Network (FFNN)—for sentiment classification on Indonesian-language user reviews from the Indodax cryptocurrency platform. The hypothesis was that combining statistical and semantic features—specifically TF-IDF and FastText embeddings—would enhance classification performance.

The findings support this hypothesis. The BiLSTM-Attention model, utilizing FastText embeddings, achieved 96% accuracy and balanced macro-average precision, recall, and F1-score of 0.96. The FFNN model, which used a 600-dimensional concatenation of TF-IDF and FastText features, slightly outperformed BiLSTM with 97.07% accuracy and macro-average precision, recall, and F1-score of 0.97.

Interestingly, the FFNN outperformed the more complex BiLSTM-Attention model. A possible explanation lies in the nature of the input features and the labeling approach: FFNNs are generally effective when trained on structured, fixed-length features. Since the dataset was lexicon-labeled and carefully preprocessed, the FFNN may have benefited from the noise reduction and clearer signal in the engineered feature space. On the other hand, BiLSTM—while capable of capturing sequential dependencies—may be more vulnerable to subtle noise or mislabeling in weak supervision settings, such as lexicon-based annotation.

The FFNN + TF-IDF + FastText model has significant potential for real-world fintech applications in Indonesia. Its lightweight architecture and high accuracy make it suitable for direct deployment in sentiment monitoring dashboards, customer feedback analysis tools, or as an alert system for detecting negative sentiment spikes on trading platforms. Because of its relatively low computational cost, it can be integrated into existing analytic infrastructures without major architectural overhauls.

This study employed weakly supervised lexicon-based labeling, which enabled scalable and efficient annotation of sentiment data from cryptocurrency-related social media. This approach represents a key strength, as it significantly reduces annotation costs while supporting large-scale data utilization. However, it also exhibits limitations in identifying linguistic nuances such as sarcasm, slang, and complex sentiment shifts. Additionally, the dataset was confined to a single platform (Indodax) and a limited time frame, which may constrain the generalizability of the findings.

Future research is encouraged to incorporate transformer-based models such as IndoBERT to better capture contextual and nuanced sentiment. It is also advisable to expand the dataset across diverse platforms and broader time ranges to enhance generalization. Furthermore, employing manual validation or semi-

automated gold-standard labeling could improve the reliability of sentiment annotations. Lastly, conducting ablation studies on various feature combinations may provide deeper insights into the key drivers of model performance.

References

- [1] A. A. Artanti, "Jumlah Investor Kripto Terus Naik di Awal 2024," *metrotvnews.com*, 2024. [Online]. Available: <https://www.metrotvnews.com/read/NgxCV8qB-jumlah-investor-kripto-terus-naik-di-awal-2024>
- [2] A. Malik, "OJK Mulai Awasi Kripto – Investor Wajib Tahu Dampaknya," *bareksa.com*, 2025. [Online]. Available: https://www.bareksa.com/berita/reksa-dana/2025-02-06/ojk-mulai-awasi-kripto-investor-wajib-tahu-dampaknya?utm_source=chatgpt.com
- [3] Muhammad Afif Raihan and Erwin Budi Setiawan, "Aspect Based Sentiment Analysis with FastText Feature Expansion and Support Vector Machine Method on Twitter," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 6, no. 4, pp. 591–598, 2022, doi: 10.29207/resti.v6i4.4187.
- [4] R. Rahmadani, A. Rahim, and R. Rudiman, "Analisis Sentimen Ulasan 'Ojol the Game' Di Google Play Store Menggunakan Algoritma Naive Bayes Dan Model Ekstraksi Fitur TF-IDf Untuk Meningkatkan Kualitas Game," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, 2024, doi: 10.23960/jitet.v12i3.4988.
- [5] B. A. Maulana, M. J. Fahmi, A. M. Imran, and N. Hidayati, "Analisis Sentimen Terhadap Aplikasi Pluang Menggunakan Algoritma Naive Bayes dan Support Vector Machine (SVM)," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, pp. 375–384, 2024, doi: 10.57152/malcom.v4i2.1206.
- [6] D. A. Putri, V. E. Irawan, F. Alfaraby, M. A. Fathurrazaq, and G. P. Insany, "Klasifikasi Sentimen Analisis Terhadap Cryptocurrency Exchange Menggunakan Naive Bayes Untuk Mendukung Keputusan Investor Pemula," *Sentim. (Seminar Nas. Teknol. Informasi, Mekatronika dan Ilmu Komputer)*, 2024.
- [7] A. Syahrul, A. I. Purnamasari, and I. Ali, "Analisis Sentimen Twitter Terhadap Cryptocurrency Menggunakan Algoritma Naive Bayes dan Decision Tree," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 2, pp. 1–8, 2024.
- [8] Verra Budhi Lestari, Ema Utami, and Hanafi, "Combining Bi-LSTM And Word2vec Embedding For Sentiment Analysis Models Of Application User Reviews," *Indones. J. Comput. Sci.*, vol. 13, no. 1, pp. 312–326, 2024, doi: 10.33022/ijcs.v13i1.3647.
- [9] M. A. S. Nasution and E. B. Setiawan, "Enhancing Cyberbullying Detection on Indonesian Twitter: Leveraging FastText for Feature Expansion and Hybrid Approach Applying CNN and BiLSTM," *Rev. d'Intelligence Artif.*, vol. 37, no. 4, pp. 929–936, 2023, doi: 10.18280/ria.370413.
- [10] L. Saragih, M. Nababan, Y. Simatupang, and J. Amalia, "Analisis Self-Attention Pada Bi-Directional Lstm Dengan Fasttext Dalam Mendeteksi Emosi Berdasarkan Text," *Zo. J. Sist. Inf.*, vol. 4, no. 2, pp. 144–156, 2022, doi: 10.31849/zn.v4i2.10846.
- [11] C. Hayati, S. S. Hilabi, A. L. Hananto, S. Informasi, U. Buana, and P. Karawang, "Klasifikasi dan Prediksi Ulasan Aplikasi Dana pada Google PlayStore Menggunakan Algoritma Naive Bayes," *JINTEKS (Jurnal Inform. dan Sains)*, vol. 7, no. 2, pp. 596–605, 2025.
- [12] A. Fauzul, A. Kustiwa, M. Aminul Akbar, and A. Pinandito, "Studi Perbandingan pada Metode CNN-LSTM dan LSTM dalam Mendeteksi Emosi pada Data Teks Berbahasa Indonesia pada Media Sosial Twitter," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 1, pp. 2548–964, 2017, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [13] M. Chiny, M. Chihab, Y. Chihab, and O. Bencharef, "LSTM, VADER and TF-IDF based Hybrid Sentiment Analysis Model," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 7, pp. 265–275, 2021, doi: 10.14569/IJACSA.2021.0120730.
- [14] Alfando and R. Hayami, "Klasifikasi Teks Berita Berbahasa Indonesia Menggunakan Machine Learning dan Deep

- Learning: Studi Literatur,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 681–686, 2023.
- [15] google, “Indodax Kripto Mudah Aman Cuan,” play.google.com. Accessed: Jun. 29, 2025. [Online]. Available: <https://play.google.com/store/apps/details?id=id.co.bitcoin&hl=id>
- [16] N. Nurwanda, N. Suarna, and W. Prihartono, “Penerapan Nlp (Natural Language Processing) Dalam Analisis Sentimen Pengguna Telegram Di Playstore,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 1841–1846, 2024, doi: 10.36040/jati.v8i2.8469.
- [17] D. Normawati and S. A. Prayogi, “Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter,” *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 2, pp. 697–711, 2021.
- [18] K. Machová, M. Mikula, X. Gao, and M. Mach, “Lexicon-based sentiment analysis using particle swarm optimization,” *Electron.*, vol. 9, no. 8, pp. 1–22, 2020, doi: 10.3390/electronics9081317.
- [19] G. Li, B. Li, L. Huang, and S. Hou, “Automatic construction of a depression-domain lexicon based on microblogs: Text mining study,” *JMIR Med. Informatics*, vol. 8, no. 6, pp. 1–17, 2020, doi: 10.2196/17650.
- [20] H. Utama, E. Daniati, and A. Masruro, “Weak Supervision Dengan Pendekatan Labeling Function Untuk Analisis Sentimen Pada Twitter,” *Indones. J. Comput. Sci. Res.*, vol. 3, no. 1, pp. 49–57, 2024, doi: 10.59095/ijcsr.v3i1.93.
- [21] T. Wongvorachan, S. He, and O. Bulut, “A Comparison of Undersampling, Oversampling, and SMOTE Methods for Dealing with Imbalanced Classification in Educational Data Mining,” *Inf.*, vol. 14, no. 1, 2023, doi: 10.3390/info14010054.
- [22] S. Wang, Y. Dai, J. Shen, and J. Xuan, “Research on expansion and classification of imbalanced data based on SMOTE algorithm,” *Sci. Rep.*, vol. 11, no. 1, pp. 1–11, 2021, doi: 10.1038/s41598-021-03430-5.
- [23] K. Babić, S. Martinčić-Ipšić, and A. Meštrović, “Survey of neural text representation models,” *Inf.*, vol. 11, no. 11, pp. 1–32, 2020, doi: 10.3390/info11110511.
- [24] J. Philip Thekkekkara, S. Yongchareon, and V. Liesaputra, “An attention-based CNN-BiLSTM model for depression detection on social media text,” *Expert Syst. Appl.*, vol. 249, no. PC, p. 123834, 2024, doi: 10.1016/j.eswa.2024.123834.
- [25] A. Borah, M. P. Barman, and A. Awekar, *Are Word Embedding Methods Stable and Should We Care about It?*, vol. 1, no. 1. Association for Computing Machinery, 2021. doi: 10.1145/3465336.3475098.
- [26] A. Mohammadi and A. Shaverizade, “Ensemble deep learning for aspect-based sentiment analysis,” *Int. J. Nonlinear Anal. Appl.*, vol. 12, no. Special Issue, pp. 29–38, 2021, doi: 10.22075/IJNAA.2021.4769.
- [27] Ž. Vujović, “Classification Model Evaluation Metrics,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 6, pp. 599–606, 2021, doi: 10.14569/IJACSA.2021.0120670.