FEATURE EXTRACTION USING GLOVE AND FASTTEXT FOR SENTIMENT ANALYSIS ON AUGMENTED TEXT DATA

MOHAMMAD HASBI ABIYADI[1] , AMALIA ZAHRA[2]

[1]Student, Master Program in Computer Science, BINUS Graduate Program, BINUS University, Indonesia [2]Lecturer, Master Program in Computer Science, BINUS Graduate Program, BINUS University, Indonesia E-mail: [1]mohammad.abiyadi@binus.ac.id, [2]amalia.zahra@binus.ac.id

ABSTRACT

In this modern era, technological development and globalization have completely changed human habits towards the digital world. leading to a significant increase in user activity on social media platforms such as Instagram, Facebook, and Twitter, which in turn generates large volumes of data known as big data.This research aims to improve the accuracy of sentiment analysis on Indonesian text data using the GloVe and FastText methods as effective feature extraction techniques in text processing. The research framework systematically describes the research process, from problem identification to research objectives. The identified problem is the low accuracy of sentiment analysis, especially for rare words. The literature review explores various sentiment analysis methods, word embeddings, and previous related research. The research hypotheses are formulated based on the literature review, stating that GloVe and FastText methods produce different accuracy in sentiment analysis on Indonesian text data and that FastText yields higher accuracy compared to GloVe. Data collection involves gathering Indonesian text data with sentiment labels from sources such as social media, online forums, and product reviews. The research provides a clear and detailed guide on the effective use of GloVe and FastText methods in sentiment analysis, assisting researchers and practitioners in decision-making when choosing methods

Keywords: *Sentiment Analysis, Feature Extraction, GloVe, FastText, Data Augmentation, Word Embedding, Indonesian Text, IndoBERT, Machine Learning.*

1. INTRODUCTION

A In this modern era, technological development and globalization have completely changed human habits towards the digital world. This change is especially pronounced in the social media sector, such as Instagram, Facebook, Twitter, and the like, which offer various features and applications. Social media not only facilitates business and work, but also serves as a means to deliver news, conduct promotions, and even as a channel to express emotions. The impact of this development is an increase in the number of users and activities on social media, which generates a large amount of data known as "big data" (Putra,2023).

Big data is a collection of data that has a very large and complex volume. Data derived from the use of social media is also included in the big data category, with three main characteristics: large volume, diverse variations, and speed of data change. The existence of big data is very beneficial for businesses, business developers, large companies, and governments that have a wide scope. In his State of the Nation Address on August 16, 2019 at the Presidential Palace, President Joko Widodo stated that "data is a new type of wealth for our country, and data is now more valuable than oil." This large amount of data not only contains important information but also includes information that is less relevant when viewed directly. However, what matters most is not the content or importance of the data, but the processing that can make it valuable. One of the techniques that can be used to process and utilize big data is text processing techniques (Dewi ,2020).

According to the results of research conducted by (Permatasari, et al., 2021), the application of classification techniques is faced with social media data, which mostly consists of text in the form of tweets or posts from users. Sentiment analysis is

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

carried out to gain understanding or decisions that can be useful for various related parties.

is faced with social media data, which mostly consists of text in the form of tweets or posts from users. Sentiment analysis is carried out to gain understanding or decisions that can be useful for various related parties. It is intended that these data also provide broad benefits.

Sentiment analysis is a process to identify and understand the opinions or feelings contained in a text. Sentiment information can be used for various purposes, such as understanding public opinion on a product or service, monitoring brand reputation, and analyzing market trends.

Feature extraction is a crucial step in sentiment analysis, transforming raw text into a numerical representation that can be processed by machine learning algorithms. Two popular feature extraction methods are GloVe (Global Vectors for Word Representation) and FastText. Both are word embedding techniques aimed at capturing the semantic meaning of words in a corpus. These word embeddings can then be used as input for various classification models, such as Support Vector Machine (SVM), Artificial Neural Network (ANN), XGBoost, and Random Forest. (Putra, 2023)

GloVe and FastText have different approaches to building word embeddings. GloVe trains embeddings based on a global word co-occurrence matrix, which records how often words appear together in a corpus. FastText, on the other hand, uses a subword-based approach, breaking words down into smaller units (n-grams) and training embeddings for each n-gram. This subword approach allows FastText to handle out-of-vocabulary words more effectively and capture morphological information from words.

Recent studies have demonstrated that FastText frequently outperforms GloVe in sentiment analysis tasks, particularly when applied to complex and heterogeneous textual data. As reported by Putra (2023), FastText achieved an accuracy of 85% and a precision of 87%, whereas GloVe achieved an accuracy of 83% and a precision of 85%. Nonetheless, the performance of both embedding methods may fluctuate depending on the inherent characteristics of the dataset and the tuning of model parameters.

According to the results of research conducted by (Permatasari, et al., 2021), the application of classification techniques

After word embedding is generated by GloVe or FastText, the next step is to select the appropriate classification algorithm. Some commonly used algorithms in sentiment analysis.

Support Vector Machine (SVM), an algorithm that is effective in high-dimensional spaces. SVM aims to find the optimal hyperplane that separates data classes.

Artificial Neural Network (ANN) a model inspired by the structure of the human nervous system. ANNs can learn complex patterns in data and are highly flexible in handling various types of classification problems.

XGBoost a powerful and efficient gradient boosting algorithm. XGBoost combines multiple weak models to produce an accurate prediction model.

Random Forest an ensemble learning algorithm that builds multiple decision trees and combines their predictions. Random Forest tends to be more stable and less prone to overfitting.

The selection of the optimal classification algorithm depends on the characteristics of the data and the purpose of the analysis.

One of the challenges in sentiment analysis is the lack of sufficient labeled data to train the model. Data augmentation is a technique to increase the size and variety of a dataset by creating new variations of existing data. In the context of sentiment analysis, data augmentation can be done in various ways, which will be discussed in the next chapter.

Based on the above background, both methods have been proven effective in generating meaningful word vector representations. However, there has been no comprehensive study comparing these two methods in the context of sentiment analysis on Indonesian text data. To date, there has been no study explicitly comparing the performance of the two word embedding methods, GloVe and FastText, in the context of Indonesian text data that has been augmented. Therefore,

This study aims to evaluate and compare the performance of GloVe and FastText word embedding methods in Indonesian sentiment analysis using augmented review datasets with imbalanced sentiment distribution. The research further investigates the effectiveness of several machine learning classifiers, namely Support Vector Machine (SVM), Artificial Neural Network (ANN), Random Forest, and XGBoost, in classifying positive, neutral, and negative sentiments. In addition, this study aims to analyze the impact of text augmentation techniques on improving classification performance through evaluation metrics including accuracy, precision, recall, and F1-score.

The novelty of this study lies in the integration of text augmentation techniques with comparative word embedding methods, namely GloVe and FastText, for Indonesian multiclass sentiment analysis on imbalanced review datasets. This research systematically evaluates the performance of multiple machine learning classifiers, including Support Vector Machine (SVM), Artificial Neural Network (ANN), Random Forest, and XGBoost within the same experimental framework. In addition, this study investigates the impact of data augmentation on improving sentiment classification performance for Indonesian-language datasets with imbalanced sentiment distribution, thereby providing a more comprehensive evaluation of embedding-based sentiment analysis approaches.

2. RELATED WORKS

Several Machine learning methods have been used to detect hate speech on social media sites, but the majority of methods focus on extracting features from text. Some research has used lexical features including dictionaries and bag-of-words to detect aggressive posts containing images and text.

Many algorithms for modeling terms have been suggested, and many have shown good results, including Naive Bayes (NB), Logistic Regression (LR) and Support Vector Machines (SVM) . Deep learning-based approaches have become increasingly popular for this task, and many models have blossomed in the context of natural language processing (NLP).

Early sentiment analysis relied on lexical and statistical approaches, such as bag-of-words (BoW), n-grams, and TF-IDF weighting, combined with classifiers like Naïve Bayes (NB), Logistic Regression (LR), and Support Vector Machines (SVM). While effective in small-scale tasks, these methods struggled with semantic representation and out-of-vocabulary (OOV) words, especially in morphologically rich languages like Indonesian (Mandal & Sen, 2014).

The development of word embeddings advanced this field significantly. Word2Vec (Mikolov, Chen, Corrado, & Dean, 2013) and GloVe (Pennington, Socher, & Manning, 2014) mapped words into dense vector spaces that preserved semantic similarity. GloVe, trained on global co-occurrence statistics, proved efficient for capturing corpus-wide relationships. In contrast, FastText (Bojanowski, Grave, Joulin, & Mikolov, 2017) introduced subword representations that improved performance on rare words and morphologically complex terms, offering advantages for Indonesian sentiment tasks.

Deep learning models such as Convolutional Neural Networks (CNN) for sentence classification (Kim, 2014) and Recurrent Neural Networks (RNNs/LSTMs) further improved text classification, though they remained limited by static embeddings that did not adapt to context.

Recent years have seen a paradigm shift towards contextual embeddings and transfer learning. ELMo (Peters et al., 2018) introduced deep contextualized word representations, while ULMFiT (Howard & Ruder, 2018) demonstrated the power of inductive transfer learning for text classification. The introduction of BERT (Devlin, Chang, Lee, & Toutanova, 2019) and its derivatives, such as IndoBERT (Wilie et al., 2020), brought state-of-the-art performance across multiple NLP tasks, including sentiment analysis. These transformer-based models leverage bidirectional attention and large-scale pretraining, outperforming static embeddings like GloVe and FastText on many benchmarks (Wang et al., 2020).

Nevertheless, transformer models come with computational costs and data requirements, which may limit their adoption in resource-constrained scenarios. Furthermore, while studies have extensively benchmarked BERT and related models, fewer works have examined the relative effectiveness of classical embeddings (GloVe vs. FastText) in Indonesian sentiment analysis under data augmentation strategies. Augmentation techniques such as back translation, synonym replacement, and contextual paraphrasing have shown promise in addressing data imbalance (Wei, Zou, Chen, & Lu, 2019; Harywanto et al., 2022), but their impact on embedding performance remains less explored.

Sentiment analysis on mobile application reviews has emerged as a distinct and practically valuable research area, given that app store reviews exhibit unique linguistic characteristics compared to general social media text. Unlike tweets or forum posts, app reviews tend to be short, domain-specific, and heavily influenced by user experience with particular product features (Rodríguez-Ibáñez et al., 2023). In the Indonesian fintech sector specifically, mobile banking applications such as those offered by Bank Mandiri and Bank BRI have seen rapid user growth, generating large volumes of user-generated review data on platforms such as Google Play Store and the App Store. Analyzing the sentiment of such reviews provides direct insights into customer satisfaction, service quality, and feature preferences (Permatasari et al., 2021). However, studies that specifically apply word embedding-based feature extraction to Indonesian mobile banking app reviews remain highlighting a domain-specific gap that this research aims to address.

A recurring challenge in real-world sentiment analysis datasets, particularly those derived from app store reviews, is the problem of class imbalance. Users of mobile applications tend to submit reviews predominantly when they have strong positive or negative experiences, which results in highly skewed sentiment distributions. In Indonesian app review datasets, the positive class frequently dominates, leaving the neutral and negative classes severely underrepresented (Rahma, 2023). This imbalance directly affects model training, causing classifiers to develop a bias toward the majority class and producing inflated overall accuracy while failing to correctly identify minority sentiments. Data augmentation techniques have been proposed as an effective strategy to address this imbalance by synthetically generating new samples for underrepresented classes (Wei et al., 2019). In particular, contextual augmentation using masked language models such as IndoBERT has shown promise in generating semantically coherent augmented samples for low-resource Indonesian text scenarios (Harywanto et al., 2022), making it a suitable augmentation strategy for the imbalanced app review data used in this study.

While transformer-based models such as IndoBERT have demonstrated state-of-the-art performance across Indonesian NLP benchmarks, their practical adoption in resource-constrained or production-level classification tasks remains limited due to high computational costs and large memory requirements (Devlin et al., 2019). In contrast, classical word embedding methods such as GloVe and FastText offer a computationally efficient alternative that remains competitive for domain-specific short-text classification tasks such as app reviews (Nurdin et al., 2020; Putra, 2023). FastText's subword representation is particularly relevant for morphologically rich languages like Indonesian, where word variations and affixations are common in informal review writing (Bojanowski et al., 2017). GloVe, trained on global co-occurrence statistics, captures corpus-wide semantic relationships that are beneficial when the review vocabulary is relatively consistent within a specific domain such as banking services (Pennington et al., 2014). The comparative evaluation of these two methods on domain-specific Indonesian mobile banking review data therefore provides practical guidance for researchers and practitioners operating under real-world computational constraints.

Thus, this research contributes to compare between GloVe and FastText embeddings with augmented Indonesian text data across multiple classifiers (SVM, ANN, Random Forest, XGBoost). The study provides insights into whether classical embeddings can still be competitive in augmented low-resource contexts, offering a complementary perspective alongside transformer-based approaches.

3. PROPOSED METHODOLOGY

This research uses a comparative analytical approach to compare the performance of the GloVe and FastText methods in improving the accuracy of sentiment analysis on Indonesian text data. The data used will be analysed descriptively through accuracy, precision, recall, and F1-score metrics. This research model clearly describes the steps to be taken in data collection, data analysis, and interpretation of results to achieve the research objectives.

3.1 Study Design

This research adopts a quantitative experimental comparative design. The study systematically compares the performance of two independent word embedding methods — GloVe and FastText — as feature extraction techniques for Indonesian sentiment analysis, under two data conditions: original and augmented. The independent variables are: (1) the word embedding method (GloVe vs.

Journal of Theoretical and Applied Information Technology
30st June 2026. Vol.104. No.12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

FastText), and (2) the data type (original vs. IndoBERT-augmented). The dependent variables are the classification performance metrics: accuracy, precision, recall, and F1-score. Controlled variables include the classification algorithms (SVM, ANN, Random Forest, XGBoost), the dataset source (Brimo and Livin mobile banking reviews), the preprocessing pipeline, and the train-test split ratio (80:20).

The research hypotheses are as follows. H1: GloVe and FastText produce significantly different accuracy scores in sentiment analysis on Indonesian text data. H2: Data augmentation using IndoBERT masked language modeling significantly improves model performance when combined with GloVe feature extraction. H3: Data augmentation does not consistently improve model performance when combined with FastText feature extraction, due to FastText's sensitivity to subword-level token substitutions. These hypotheses are evaluated using accuracy metrics, classification reports, and error-based metrics (MSE, RMSE, MAE) across all classifier and embedding combinations.

3.2 System Design

The system design in this research involves mapping the sentiment analysis business process on Indonesian text data, identifying weaknesses or problems in the process, and preparing solutions offered using the GloVe and FastText methods. The framework used in this research is the SDLC framework, which includes the stages of planning, analysis, design, implementation, testing, and maintenance. System design in this research is the main key in ensuring that sentiment analysis on Indonesian text data can be done effectively and efficiently.

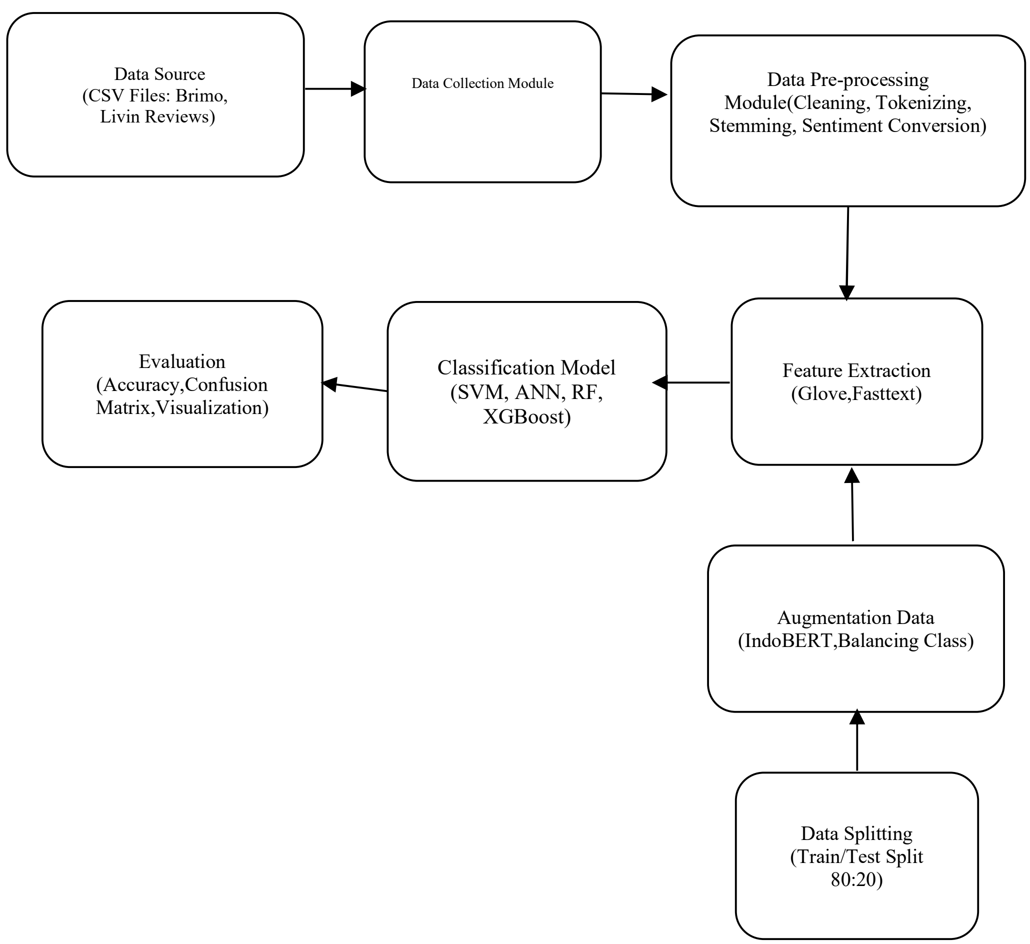

Figure 1: Research Flowcart

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

After the system is implemented, the testing stage will be conducted to ensure that the system can run properly and produce accurate results. In the figure 1. Research flowchart shows the technical steps of the research that divided into 5 main steps. The first main step is the data collection including Indonesian data text collection, breaking it down into training and testing data, and data preprocessing. Before data is being testing, the data should be balanced by the data augmentation technique that is what the research to conducted.

The second is the data analysis that includes feature extraction using two types of them that is Glove and Fasttext, model training, and then model testing.

The third is the discussion main step regarding to the model result such as model evaluation to see the result of the two feature extraction.

The fourth is evaluation that aims to evaluate the model using confusion matrix on the two feature extraction.

The last steps is the conclusion to conclude the result of the research regarding to data augmentation technique. All of the steps are explained detailly on it's own chapter.

3.3 Data Analysis

The analysis method to be applied in this research will include a comprehensive and detailed set of steps to evaluate the performance of the GloVe and FastText methods in sentiment analysis on Indonesian text data. These steps will include the use of standard metrics such as accuracy, precision, recall, and F1-score, as well as hypothesis testing using descriptive tests, relevant statistical tests, such as independent t-test.

Descriptive Statistical Analysis used to describe basic statistics of Indonesian text data, such as word count, sentence length, and sentiment distribution.

Data Visualisation: Used to visualise sentiment distribution, performance of GloVe and FastText methods, and patterns in the data.

Evaluation will used to measure how robust is the model to identify each class of sentiment using the evaluation metrics score.

Accuracy score is the proportion of correct predictions out of total predictions.

Precision score is proportion of correct positive predictions out of total positive predictions.

Recall is the proportion of correct positive predictions out of the total number of true positives.

Moreover, to assure the model are able to identify class sentiment and predicting accurately will require a further analysis.

Error Analysis is the most frequent types of errors made by the GloVe and FastText methods.

Performance analysis on data bubgroups is to analyse the performance of GloVe and FastText methods on different data subgroups, such as text data with different sentence lengths or text data with different topics.

To measure the accuracy of sentiment analysis, an accuracy metric will be used. Accuracy will measure the extent to which the GloVe and FastText methods are able to classify sentiment correctly. Furthermore, to measure the accuracy in identifying positive sentiments, precision will be used. Precision will provide information on how often the method classifies the sentiment as positive, when it is actually negative or neutral. Additionally, to measure how good the method is at finding all the true positive sentiments, recall will be used. Recall will provide information on how many positive sentiments the method can find out of the total positive sentiments that actually exist. Finally, to measure the balance between precision and recall, F1-score will be used.

4. EXPERIMENTS AND RESULTS

4.1 Experimental Setup

This study was obtained from the Kaggle website, which is a type of public data that can be accessed freely. This dataset consists of a collection of reviews of Bank Mandiri and Bank BRI mobile banking applications in Indonesian. This dataset was chosen because it is directly relevant to the purpose of the study, namely sentiment analysis using Indonesian text data.

The dataset consists of three types 1 review of

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Bank Mandiri's mobile banking app and 2 reviews of Bank BRI's mobile banking app. The Bank Mandiri dataset is in CSV format with 155,139 rows and 5 columns. The Bank BRI dataset has two types of reviews: Bank BRI mobile banking reviews on the Android app in CSV format with 17,219 rows and 8 columns, and Bank BRI mobile banking reviews on the iOS app in Excel format with 467 rows and 6 columns. The dataset includes variables related to the research topic, such as: Indonesian-language reviews, app ratings from 1 to 5, and the number of thumbs-up.

Table 1: Dataset Sample

Dataset	Size	Class
Livin mandiri	1551889	3
Brimo playstore	17220	3
Brimo appstore	449	3

Table 2: Livin Mandiri Dataset

Attribute	Data Type	Description
Date	Datetime	Date Review
Review	Text	User’s Review
Rating	Integer	User application score.
Thumbs_up	Integer	Number of likes based on the thumbs-up icon.
Version	Integer	Application Version
Attribute	Data Type	Description

Table 3: Brimo Playstore Dataset Description

Attribute	Data Type	Description
Reviewid	Integer	Reviewid
Content	Text	Review detail
Score	Integer	User’s Application Score
Thumbs_up	Integer	Number of likes based on the thumbs-up icon.
Review Created Version	Integer	Application version
At	Datetime	Date Review
Reply Content	Text	Review Reply
Reply At	Datetime	Review Reply
Review Created Version	Integer	Application version

Table 4: Brimo Appstore Dataset Description

Attribute	Data Type	Description
Date	Datetime	Date Review
Developer Respons	Json	Developer Response To review
Review	Text	Customer’s Review
Rating	Integer	Score
Is Edited	Boolean	Edited Review

4.2 Data Distribution

Data distribution is measured based on dataset labels, namely positive, negative, and neutral, which are measured from ratings of 1-5. Rating below 3 is classified as negative, rating equal to 3 as neutral, and rating above 3 as positive. The Livin Mandiri distribution shows a strong dominance of positive reviews (378 samples), followed by negative (119 samples), and a very small neutral class (16 samples). The Brimo dataset shows a similar positive-dominant distribution. This class imbalance directly motivated the use of data augmentation in this study.

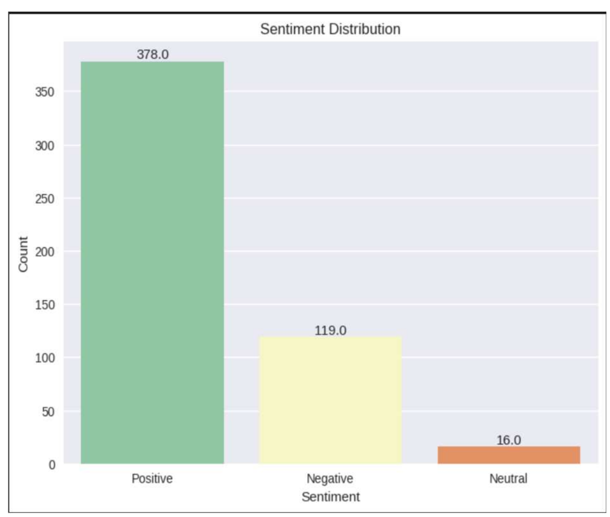

Figure 2: Livin Mandiri Data Distribution

Journal of Theoretical and Applied Information Technology
30st June 2026. Vol.104. No.12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

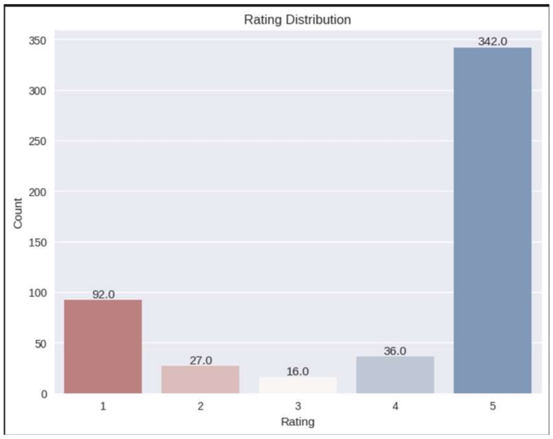

Figure 3: Brimo Data Distribution

4.3 Baseline Model

The baseline model is used as an initial benchmark to assess the performance of more complex advanced models. In this research, the baseline model helps compare the effectiveness of feature extraction methods (GloVe and FastText) on sentiment analysis of augmented Indonesian text data.

1. Data

The source is Indonesian text data that has been labeled with sentiment (positive, negative, neutral) and has been augmented there are, Peprocessing such as, tokenization, lowercasing, stopword removal, and Stemming/Lemmatization

2. Feature Extraction

Pre-trained word embeddings such as GloVe and FastText represent words as dense, real-valued vectors that capture semantic meaning and contextual similarity. word embeddings are learned from large corpora and encode linguistic relationships between words.

GloVe and FastText are serve as strong baselines because they provide richer word representations and enable direct comparison with more complex, contextual models. FastText has the additional advantage of handling out-of-vocabulary words through subword representations, which improves robustness.

3. Data Augmentation

In this study the researcher used model augmentation data indobenchmark and synonym technique replacement. For the model augmentation Indobenchmark/IndoBERT-base-p2 masked language modeling to handle imbalance class. Random words in sentences are replaced with [MASK] tokens, and IndoBERT predicts the most probable replacement (synonym replacement), generating augmented textual samples to enrich minority classes.

For the technique in this study the researcher used the synonyms to obtained from a website that can be accessed using a module that can be imported via Python. The word replacement process in the synonym replacement technique is based on previous research that has been modified for the Indonesian language (Jungiewicz & Smywinski-Pohl, 2019). Additionally, not all words in a sentence are replaced with their synonyms. There are two conditions for a word to be replaced in the synonym replacement flow mechanism. The first condition is that the word tag must have a counterpart on the thesaurus website. If condition 1 is met, then a further check is carried out for condition 2, namely that the combination of the tag and the word is found to have a synonym on the thesaurus website. (Rahma, 2023)

4. Classification Algorithm

In this study the model that will be train such as Logistic Regression that is Random Forest and XGBoost classifier, Support Vector Machine (SVM), ANN.

The model as baseline classification were chosen because, they are classic algorithms commonly used in NLP and they are relatively lightweight, quick to train, and capable of providing stable baseline performance.

Hyperparameter tuning is performed using GridSearchCV to optimize each model's performance on training data.

4.4 Evaluation

$$\boldsymbol{Accuracy} = \frac{\boldsymbol{TP + TN}}{\boldsymbol{TP + TN + FP + FN}} \qquad (1)$$

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

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

$$\boldsymbol{F1\ Score} = \mathbf{2} \times \frac{\boldsymbol{Recall} \times \boldsymbol{Precision}}{\boldsymbol{Recall} + \boldsymbol{Precision}} \qquad (\mathbf{4})$$

In this section will calculate the whole score of the model using matrix evaluation above, where TP is True Positive, TN is True Negative, FP is False Positive, dan FN is False Negative. In the matrix evaluation it has 4 type of score such as eq (1). accuracy score is the correct proportion predictions out of total predictions, eq (2) precision score is the proportion of correct positive predictions, eq (3) recall score is the proportion of correct positive predictions out of total true positive predictions, eq (4) f1 score is the average harmonized precision and recall score.

4.5 Error Evaluation

In this section will calculate the error metrics score of MSE,RMSE, and MAE to measure how far is the prediction value approaching the actual value. The less value of the error metrics score, means the closest prediction value to the actual label is. Conversely, the highest score, means the prediction score deviating the actual labels and far from the accurate prediction.

4.6 Experimental Design

The system design in this study involves mapping the sentiment analysis business process on Indonesian text data, identifying weaknesses or problems in the process, and developing solutions using the GloVe and FastText methods.

This sentiment analysis system is designed using the Python programming language with a machine learning-based text processing and classification approach. It performs sentiment analysis on application reviews (Brimo, Livin) taken from the App Store and Google Play Store. The purpose of this sentiment analysis system is to classify reviews into positive, negative, and neutral sentiments based on the ratings in the dataset. Then, it builds a model with 4 classification models (SVM, ANN, Random Forest, XGBoost) using two types of word embedding feature extraction (GloVe and FastText) on the original data and the augmented data. Finally, it improves the performance of the sentiment classification model by using a contextual-based data augmentation technique (IndoBERT).

This system uses several main libraries such as Sastrawi for Indonesian stemming, Gensim for loading word embeddings (GloVe and FastText), and Scikit-learn for machine learning modeling. In addition, the NLTK library is used for text tokenization.

Before data can be processed, pre-processing must first be carried out as described in the baseline model chapter, such as data cleaning, tokenization, and stemming, so that the dataset can be processed and converted according to the needs of the machine learning model.

Then, the processed dataset is divided into two parts, namely training data and testing data with a ratio of 80:20, which will then be entered into the data augmentation process.

Data augmentation methods have also been used for Natural Language Processing (NLP) (Feng et al., 2021; Li, Hou, and Che, 2022) such as text classification (Bayer et al., 2022). Data augmentation methods have also been used for time series (Wen et al., 2020) and tabular datasets for various applications. The following are the steps in the data augmentation process.

Identify Text Length serves to filter reviews that have more than 2 words for augmentation.

IndoBERT Model (Contextual Embedding): Uses the indobenchmark/IndoBERT-base-p2 model from Hugging Face as an unmasker to generate context-based synonyms.

Augmentation Generation: Randomly select words in a sentence, replace them with [MASK], and use IndoBERT to predict the most likely replacement words, generating new augmented sentences.

Identify underrepresented sentiment classes (e.g., Negative and Neutral if Positive is the largest target class) and generate augmentations until the number of samples in those classes approximates the number of samples in the target class so that each class has a balanced data distribution.

The system utilizes pretrained word embeddings models such as GloVe and FastText to represent words in the form of numerical vectors that can be understood by machine learning algorithms.

Both GloVe and FastText feature dimensions have 300 dimensions. The sequence length in

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

GloVe and FastText usually depends on the number of words in the analyzed text. The sequence length is determined by the number of words in the text input. For example, if the text consists of 50 words, the sequence length is 50.

Both generate 300-dimensional vectors for each word, which can be combined into a single representation vector for the entire text. This is so that the vectors from GloVe and FastText can be input into the classification model. Features from GloVe and FastText are input into the classification model as a single vector per sentence, even if the sentence consists of many words. This process is done by averaging the word vectors.

Several classification models were tested, including Support Vector Machine (SVM), Artificial Neural Network (ANN), Random Forest (RF), and XGBoost (XGB). Hyperparameter tuning was performed using GridSearchCV from the sklearn.model_selection library. GridSearchCV systematically tried all parameter combinations specified in param_grid and performed cross-validation to find the best combination.

The system evaluates model performance using accuracy metrics, with a training scheme that uses augmented data and testing using the original data to measure how effectively the model learns. It also uses classification reports to assess sentiment prediction results. The evaluation metric scores are a collection of test results from all combinations of models, feature extraction methods, and data types (original vs. augmented).

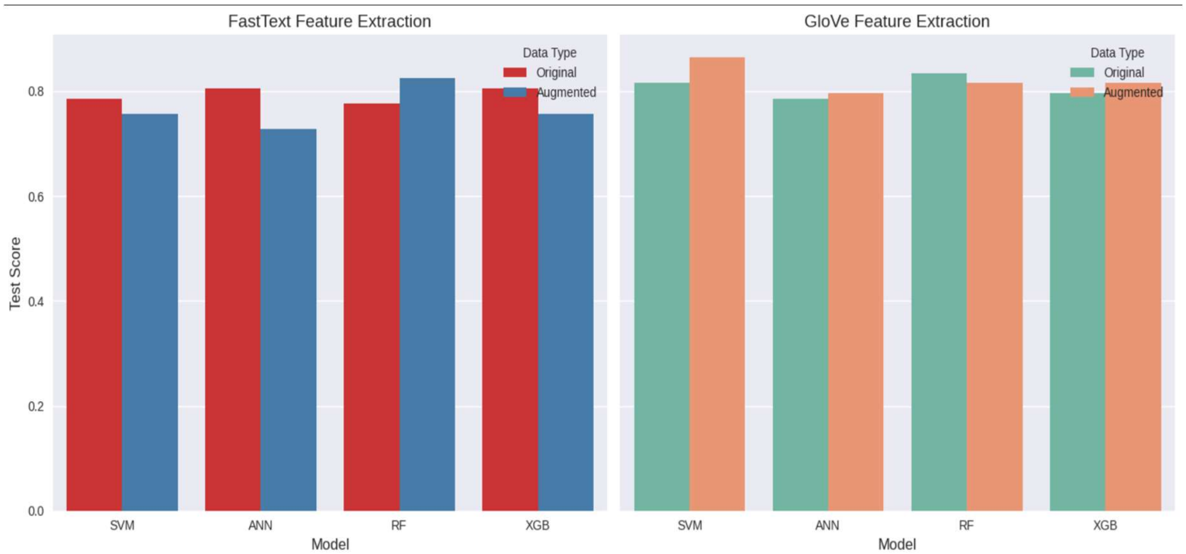

Figure 4 : All Model Result

4.7 Result

This research compares two feature extraction methods, namely GloVe and FastText, for sentiment analysis on augmented Indonesian text data. The evaluation was conducted using several machine learning classification models, namely Artificial Neural Network (ANN), Random Forest, and Support Vector Machine (SVM), with the following main results as shown on the figure 4.

1. Artificial Neural Network (ANN)

Table 5: ANN Accuracy Score

Word Embedding	Data Type	Test Score
GloVe	Original	0.786408
GloVe	Augmented	0.796117
FastText	Original	0.805825
FastText	Augmented	0.728155

Table 6: ANN Error Metrics Score

Word Embedding	Data Type	Error Metrics Type	Score
GloVe	Original	MSE	0.7670
GloVe		RMSE	0.8758
GloVe		MAE	0.3981
GloVe	Augmented	MSE	0.7282
GloVe		RMSE	0.8533
GloVe		MAE	0.3786
Fastext	Original	MSE	0.7184
Fastext		RMSE	0.8476
Fastext		MAE	0.3689
Fastext	Augmented	MSE	0.8544
Fastext		RMSE	0.9243
Fastext		MAE	0.4660

ANN can essentially capture non-linear data patterns, making it suitable for GloVe features that contain semantic information for each word. By using GloVe vectors as features, ANN obtains a dense, meaningful, and contextual representation of each word. Table 6 shows the test results before and after augmentation of the Glove and FastText word embeddings. Glove experienced an increase after augmentation, but not significantly, while FastText decreased. Figure 7 shows the results of the model prediction error calculations, which also show alignment with the results of the previous table. Glove experienced a decrease in error, while FastText experienced an increase.

The ANN model trained with GloVe features from the original data shows a test accuracy of around 78.64%. The relatively moderate MSE, RMSE, and MAE values indicate that the model has an acceptable error rate.

This finding aligns with prior studies showing that GloVe's global co-occurrence representation produces stable, corpus-wide semantic vectors that benefit from increased data variety through augmentation (Pennington et al., 2014). Data augmentation in ANN provides a slight performance improvement because the data variation can help ANN learn more robustly against various linear and non-linear patterns. The increase in scores also shows that ANN is quite good at mapping the meaning of input word representations to sentiment labels.

With the FastText feature, ANN obtains representations that can capture word morphology and variations, making it stronger in handling word variations due to FastText's subword feature. As previously, the accuracy of the ANN model with the FastText feature decreased drastically after data augmentation, to around 72.82%. The increase in MSE, RMSE, and MAE values was also very significant. This occurred because the augmented data generated by IndoBERT introduced ambiguity or semantic mismatches that degraded the FastText representation. Moreover due to increased noise or variations that are not well represented in the FastText features used, causing the ANN model to struggle with generalizing on the augmented data.

The degradation of FastText performance after IndoBERT-based augmentation is consistent with the known sensitivity of subword-based embeddings to token-level substitutions. As noted by (Bojanowski et al. 2017), FastText constructs word representations from character n-grams, meaning that even minor lexical substitutions can alter the subword composition and shift the resulting vector representation. When augmented tokens generated by a masked language model do not preserve the original morphological structure of Indonesian words, the resulting n-gram sequences can introduce noise into the embedding space, a problem also observed by (Harywanto et al. 2022) in augmented Indonesian text classification tasks.

2. Random Forest

Table 7: Random Forest Accuracy Score

Word Embedding	Data Type	Test Score
GloVe	Original	0.834951
GloVe	Augmented	0.815534
FastText	Original	0.776699
FastText	Augmented	0.825243

Table 8: Random Forest Error Metrics Score

Word Embedding	Data Type	Error Metrics Type	Score

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

GloVe	Original	MSE	0.6019
GloVe		RMSE	0.7758
GloVe		MAE	0.3107
GloVe	Augmented	MSE	0.6796
GloVe		RMSE	0.8244
GloVe		MAE	0.3495
Fastext	Original	MSE	0.8350
Fastext		RMSE	0.9138
Fastext		MAE	0.4272
Fastext	Augmented	MSE	0.6117
Fastext		RMSE	0.7821
Fastext		MAE	0.3204

Random Forest is an ensemble learning method that combines multiple decision trees, much like a real tree with branches. By using bootstrap sampling and proportional random feature selection, the model is able to capture complex patterns and reduce overfitting. Table 7 shows the results of the Random Forest test which experienced an increase in accuracy score after augmentation on the FastText word embedding while GloVe decreased its test accuracy score, but the figure 5 shows a decrease in the error metric score on GloVe and an increase on FastText. This indicates the FastText test score is not valid due to the error on the prediction value.

After data augmentation, the test accuracy of the Random Forest model with GloVe features decreased to 81.55%. This is likely due to the additional data containing new patterns or noise that were not fully addressed by Random Forest, although the model remained fairly stable as indicated by the decrease in MSE, RMSE, and MAE values, which also show that data augmentation benefits this model, helping it to generalize better.

Interestingly, when the data was trained with data augmented using FastText, the test accuracy increased significantly to around 82.52%. This increase shows that data augmentation has an effect on FastText, possibly because FastText is more sensitive to data variation and can utilize the additional diversity generated by augmentation to learn stronger representations.

The contrasting behavior of Random Forest under GloVe versus FastText augmentation reflects the well-documented sensitivity of ensemble tree methods to feature space consistency. (Jinyong et al. 2024) noted that Random Forest performance is heavily influenced by the stability of input feature distributions across training and testing data. GloVe's averaged vector representation tends to produce a more consistent feature space after augmentation, whereas FastText's subword-derived vectors are more susceptible to distributional shifts introduced by synonym-based augmentation, which can alter n-gram composition even for semantically similar replacements.

3. Support Vector Machine (SVM)

Table 9: SVM Accuracy Score

Word Embedding	Data Type	Test Score
GloVe	Original	0.815534
GloVe	Augmented	0.864078
FastText	Original	0.786408
FastText	Augmented	0.757282

Table 10: SVM Error Metrics Score

Word Embedding	Data Type	Error Metrics Type	Score
GloVe	Original	MSE	0.6796
GloVe		RMSE	0.8244
GloVe		MAE	0.3495
GloVe	Augmented	MSE	0.4854
GloVe		RMSE	0.6967
GloVe		MAE	0.2524
Fastext	Original	MSE	0.7961
Fastext		RMSE	0.8923
Fastext		MAE	0.4078
Fastext	Augmented	MSE	0.8544
Fastext		RMSE	0.9243
Fastext		MAE	0.4466

GloVe is a word embedding method based on a matrix of word co-occurrence in a large corpus that captures the frequency of word occurrence. Table 8 shows the test results before and after augmentation of the Glove and FastText word embeddings. Glove experienced an increase after augmentation, whileFastText decreased. Figure 5 shows the results of the model prediction error calculations. Glove experienced a decrease in error, while FastText experienced an increase.

GloVe generates word representations in a vector space that captures the semantic and contextual relationships of words. With relatively low MSE, RMSE, and MAE values, it indicates that the model's predictions are quite close to the actual sentiment values after conversion to numerical values. This indicates that GloVe's word representations are capable of capturing the semantic meaning relevant to sentiment classification with the original data.

This indicates that data augmentation using the IndoBERT unmasker model successfully enriched the training dataset, enabling the SVM model to learn more robust patterns and generalize better on the test data.

FastText is a word embedding model that relies on subwords in the form of n-gram characters by breaking sentences into letter segments to generate word representations even for words not present in the training corpus or OOV. FastText is effective for languages with word variations or new words, as its subwords can capture morphological information.

However, The superior performance of SVM with GloVe features after augmentation achieving an accuracy of 0.864 is consistent with prior findings in Indonesian sentiment analysis. Daniati and Utama (2023) demonstrated that SVM combined with word embedding features consistently outperforms other classical classifiers on Indonesian text data, attributing this to SVM's ability to find optimal decision boundaries in high-dimensional embedding spaces. Furthermore, the persistent class imbalance observed in the neutral class (only 2 test samples) echoes findings by Rahma (2023), who highlighted that token substitution-based augmentation strategies tend to underperform for severely underrepresented classes in Indonesian datasets, as the generated samples may not adequately represent the full semantic range of neutral sentiment expressions in domain-specific mobile banking language.

4. XGBoost

Table 11:XGBoost Accuracy Score

Word Embedding	Data Type	Test Score
GloVe	Original	0.796117
GloVe	Augmented	0.815534
FastText	Original	0.805825
FastText	Augmented	0.757282

Table 12: XGBoost Error Metrics Score

Word Embedding	Data Type	Error Metrics Type	Score
GloVe	Original	MSE	0.6990
GloVe		RMSE	0.8361
GloVe		MAE	0.3689
GloVe	Augmented	MSE	0.6214
GloVe		RMSE	0.7883
GloVe		MAE	0.3301
Fastext	Original	MSE	0.6893
Fastext		RMSE	0.8303
Fastext		MAE	0.3592
Fastext	Augmented	MSE	0.7670
Fastext		RMSE	0.8758
Fastext		MAE	0.4175

XGBoost is known for its ability to handle complex tabular data and efficiently capture non-linear patterns. By using the GloVe feature, the XGBoost model leverages the global semantic knowledge contained in word embeddings. Table 11 shows the results of testing the XGBoost model before and after augmentation using Glove and FastText. Glove's word embedding performance increased, while FastText's performance decreased. Figure 7 shows the model's

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Label	precision	recall	f1-score	support
Negative	0.70	0.64	0.67	25
Neutral	0.33	0.50	0.40	2
Positive	0.91	0.92	0.92	76
Macro avg	0.65	0.69	0.66	103
Weighted avg	0.85	0.84	0.84	103

prediction error calculation, which, consistent with the table 12, indicates a decrease in error for GloVe and an increase for FastText. Means the GloVe is closer to right prediction value.

After data augmentation, the test accuracy of the XGBoost model with the GloVe feature increased to 81.55%. The improvement in GloVe accuracy after augmentation indicates that data augmentation successfully enhances the quality of GloVe representations, possibly by filling gaps in the vocabulary or providing more context for less common words. The XGBoost model with FastText on the original data performs slightly better in terms of error rates compared to GloVe, as seen in the MSE, RMSE, and MAE values.

After data augmentation, the accuracy of FastText actually decreased. This could indicate several things: augmentation may have introduced noise or irrelevant variations that confused the model. FastText, which already has the ability to handle OOV through n-gram characters, may benefit less from masking-based augmentation compared to GloVe.

The overall pattern across all four classifiers where GloVe consistently benefits from augmentation while FastText generally degrades parallels observations in comparable multilingual NLP studies. Wei et al. (2019) found that the effectiveness of data augmentation is highly dependent on the nature of the feature extraction method used, with context-insensitive embeddings benefiting more from augmentation diversity than subword-sensitive models. This suggests that for Indonesian mobile banking review data, where informal language and morphological variation are prevalent, masked language model augmentation strategies are better suited to global co-occurrence embeddings like GloVe than to subword-based methods like FastText.

Best Model Evaluation (SVM)

The following table shows the scores from the metric evaluation and images from the confusion matrix. The scores and images represent the model with the highest score, which is the SVM model.

The table 13 above shows the results of the sentiment classification model evaluation with three categories: Negative, Neutral, and Positive. The data imbalance is very apparent, with the positive class dominating with 76 samples, followed by Negative with 25 samples, and Neutral with only 2 samples. The Positive class achieved a precision of 0.91, recall of 0.92, and an F1-score of 0.92 from the 76 samples. For the Negative class, the model achieved a precision of 0.70, recall of 0.64, and an F1-score of 0.67 from the 25 samples. However, the model faced significant challenges in classifying the neutral class, achieving a precision of only 0.33, recall of 0.50, and an F1-score of 0.40 from just 2 samples. This resulted in a striking difference between the macro average (0.65 precision, 0.69 recall, 0.66 F1-score) and the weighted average (0.85 precision, 0.84 recall, 0.84 F1-score). The higher weighted average indicates that the overall performance of the model is influenced by the dominance of the positive class, which has the best performance.

In the confusion matrix figure 6 below, you can see that the test score above has a value of 0.845, which means that this value is considered good.

From the heatmap visualization above, you can see that the SVM model with GloVe correctly predicted 70 classes as genuine positive reviews, then the model predicted 6 classes of false negative reviews and 0 classes of false neutral reviews. Next, for genuine negative classes, the model was able to predict 16 negative classes, 7 false positive classes, and 2 false neutral classes. Based on the model's previous prediction distribution, it can be seen that the model is very good at predicting positive classes and only 6 classes were predicted as false negatives.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

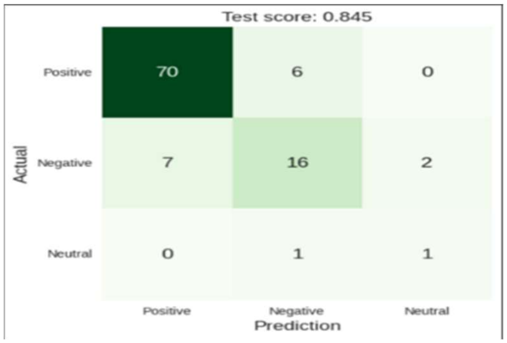

Figure 4: Confusion Matrix

Then, for the negative class, the model is quite capable of predicting the original negative class, but there are still errors in predicting the wrong class, namely 7 false positive classes and 2 false neutral classes. This shows that the model is still wrong in predicting the positive class that is actually negative and also the neutral class that is actually negative. For the neutral class, since the data distribution is still small, the model's predictions are not strong because the model has not been enriched with data augmentation for the neutral class. This means there is still an imbalance in the neutral class.

Nevertheless, overall, the model already has good performance based on the previous confusion matrix calculations and analysis.

The experimental results indicate that GloVe achieved better performance compared to FastText in several classification scenarios on the Indonesian sentiment dataset especially *SVM*. Although previous studies reported that FastText performs effectively for morphologically rich languages due to its subword representation capability, this study indicates that GloVe provided more stable semantic representations for the augmented text used in this research. This result suggests that global co-occurrence information captured by GloVe may be more effective for representing sentiment patterns. In addition, the application of text augmentation techniques contributed to improved classification performance, particularly for minority sentiment classes in the imbalanced dataset. These findings support earlier studies that reported augmentation methods can increase data variability and improve model generalization in sentiment analysis tasks. Furthermore, the comparative evaluation using SVM, ANN, Random Forest, and XGBoost demonstrates that classifier performance is strongly influenced by the quality of feature representation and dataset balance. Therefore, the combination of augmentation techniques and embedding-based feature extraction can provide a more robust framework for Indonesian sentiment classification

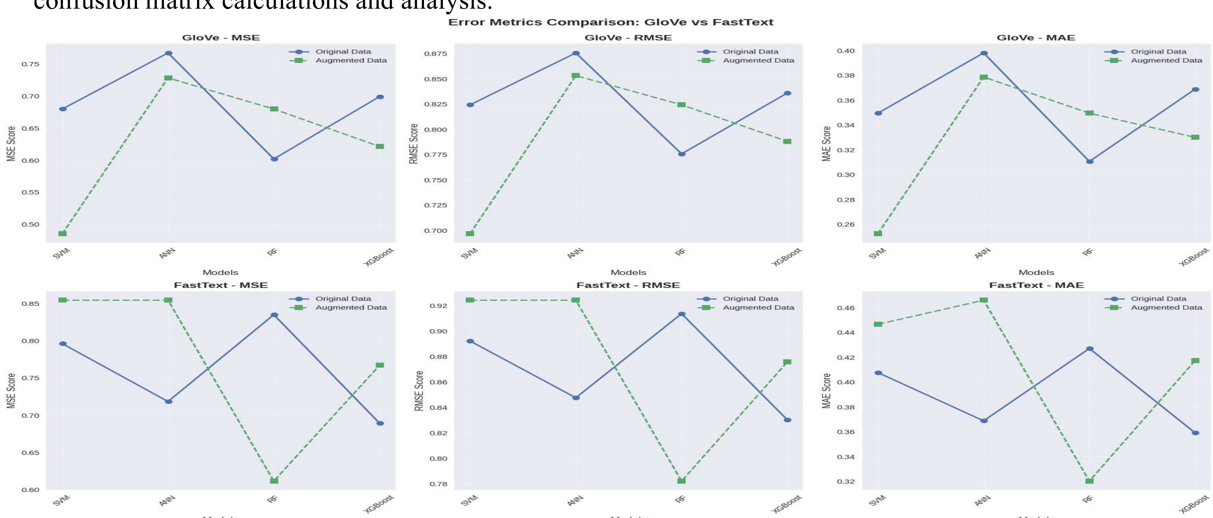

Figure 5 : All Model Error Metrics

5. CONCLUSION

This study successfully evaluated and compared the performance of GloVe and FastText word embedding methods for Indonesian multiclass sentiment analysis using augmented datasets with imbalanced sentiment distribution of class. The experimental results demonstrated that the implementation of text augmentation techniques contributed to improving classification performance by increasing dataset variability and reducing class imbalance effects. Among the evaluated embedding methods, GloVe generally achieved better performance in several experimental scenarios due to the GloVe provided more effective semantic representations for the dataset used in this study. Furthermore, the comparative evaluation using SVM, ANN, Random Forest, and XGBoost classifiers showed that classification performance is influenced not only by the classifier architecture but also by the quality of feature representation and dataset balance.

The findings of this study contribute to the development of Indonesian-language sentiment analysis by providing a comprehensive experimental comparison between embedding methods, augmentation strategies, and machine learning classifiers within a unified framework. This research also demonstrates that combining augmentation techniques with embedding-based feature extraction can improve the robustness of sentiment classification models for imbalanced datasets. Future research may explore the use of transformer-based embeddings such as BERT or IndoBERT, larger datasets, and hybrid deep learning architectures to further improve sentiment classification performance.

Data augmentation has been shown to increase recall in the minority class, namely negative, resulting in a more balanced SVM decision boundary. On the other hand, if sentiment words with strong meanings to represent the semantic meaning of features are replaced by data augmentation, there is a risk of decreasing the precision score. This demonstrates that GloVe's global co-occurrence-based representation is capable of providing word vectors that are quite robust against noise after data augmentation.

On average, FastText performance decreased for most classification models after augmentation. This can be explained because FastText forms subword-based embeddings, making them more sensitive to single-word substitutions resulting from augmentation. Token substitutions that are not fully consistent with sentiment polarity have the potential to produce noisy labels and degrade representation quality. Judging from the error analysis matrix scores in FastText, this is consistent with the majority of models experiencing an increase in error scores despite augmentation. Therefore, the effect of data augmentation depends on the embedding method used and the model's tuning parameters. For embeddings that are sensitive to noise due to subwords in FastText, a more careful augmentation strategy is required.

In future research, it is recommended to use larger and more diverse datasets such as other social media, forums, or product reviews. This will improve the generalization and validity of sentiment analysis results in Indonesian texts.

In addition, the limitations of this study also contribute to its shortcomings. The augmentation technique in this study was limited to using only one token substitution with IndoBERT (fill-mask). This process does not always preserve the meaning or sentiment polarity, thereby risking label noise. Furthermore, class imbalance is not fully optimized. The balancing process only refers to the positive class without using other classes as a target reference, resulting in less optimal predictions for labels other than positive. Therefore, improvements are needed to address the limitations of this study.

Future research is recommended to not only rely on token substitution methods based on masked language models but also combine them with other augmentation techniques such as back-translation, synonym replacement, random insertion or deletion, or contextual paraphrasing. Future research should also adjust to the actual majority class or use an adaptive oversampling approach, such as SMOTE for text, to ensure that the class distribution is truly balanced without losing semantic representation.

In addition to GloVe and FastText, research can be expanded by comparing other embedding methods such as Word2Vec, BERT, or the latest transformer models. This approach can identify the most optimal method for Indonesian language characteristics and augmented data, as well as test the performance of various classification algorithms, such as LightGBM, or deep learning models like CNN, to determine the optimal combination of feature extraction methods and classification algorithms in sentiment analysis.

a. Acknowledgement

The author would like to express my gratitude to the anonymous editors and reviewers for their time and expertise in reviewing this manuscript. Their scientific suggestions and comments have been invaluable in improving the quality of this work.

b. Funding Information

The author did not gain any additional funding to this research.

c. Author's Contributions

Mohammad Hasbi Abiyadi: Observation, Experimenting, Data Collecting, Methodology, Code Writing, Writing- Original Draft

Amalia Zahra: Formal Analysis, Supervisor, Writing- Review and Editing.

d. Ethics

The authors confirm that this article is original and contains unpublished material. The corresponding author also certifies that all co-authors have reviewed and approved the manuscript and that no ethical conflicts exist.

REFERENCES:

[1] Arliyanti Nurdin, B. A. (2020). Perbandingan Kinerja Word Embedding Word2vec, Glove, Dan Fasttext Pada Klasifikasi Teks. Teknokompak, Vol. 14, No. 2, 74-79.

[2] Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching Word Vectors with Subword Information. Facebook AI Research, 12.

[3] Daniati, E., & Utama, H. (2023). Analisis Sentimen Dengan Pendekatan Ensemble Learning Dan Word Embedding Pada Twitter . Journal Of Information System Management, 1-7.

[4] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Google AI Language, 1-16.

[5] Dewi, A. O. (2020). Big Data di Perpustakaan dengan Memanfaatkan Data Mining. ANUVA, 4(2), 223-230.

[6] Dqlab. (2022, November 29). Selami 7 Fase pada Pembuatan Machine Learning Model. Retrieved from Dqlab: https://dqlab.id/selami-7-fase-pada-pembuatan-machine-learning-model

[7] Goldberg, Y., & Levy, O. (2014). word2vec Explained: Deriving Mikolov et al.'s Negative-Sampling Word-Embedding Method. Computer Science Computational Language, 1-5.

[8] Harywanto, Gabriela & Veron, Juan & Suhartono, Derwin. (2022). A BERTweet-based design for monitoring behaviour change based on five doors theory on coral bleaching campaign. Journal of Big Data. 9. 10.1186/s40537-022-00615-1.

[9] Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., . . . Adam, H. (2019). SearchingforMobileNetV3. Computer Vision and Pattern Recognition, 1-11.

[10] Howard, J., & Ruder, S. (2018). Universal Language Model Fine-tuning for Text Classification. Computational Language, 1-12.

[11] Iftitah Athiyyah Rahma, L. H. (2023). Penerapan Text Augmentation Untuk Mengatasi Data Yang Tidak Seimbang Pada Klasifikasi Teks Berbahasa Indonesia Studi Kasus: Deteksi Judul Clickbait Dan Komentar Hate Speech Pada Berita Online. Jurnal Teknologi Informasi Dan Ilmu Komputer (Jtiik) , 1329-1340.

[12] Indah, Y. M., Aristawidya, R., Fitrianto, A., Erfiani, & Jumansyah, L. D. (2025). Comparison of Random Forest, XGBoost and LightGBM Methods on the Human Development Index Classification. Jambura, 14-18.

[13] Jacob Murel Ph.D., E. K. (2023, Desember 10). What are stemming and lemmatization? Retrieved from IBM: https://www.ibm.com/topics/stemming-lemmatization

[14] Jim Holdsworth, M. S. (2024, Juni 17). What is deep learning? Retrieved from IBM : https://www.ibm.com/topics/deep-learning#:~:text=Deep%20learning%20is%20a%20subset,applications%20in%20our%20lives%20today.

[15] Jinyong, X., Zhenhua, W., & Yunyun, D. (2024). A Review of Machine Learning Classification Research Based on the Random Forest Algorithm. Hans, 143-152.

[16] Jou, L., & Li, X. (2016). FastText: A fast

and numerical representation of text. In Proceedings of the 25th ACM International Conference on Information and Knowledge Management (CIKM) (https://arxiv.org/abs/1607.04606 ed.). New York: Cornel University.

[17] Mandal, A. K., & Sen, R. (2014). Supervised Learning Methods For Bangla Web Document Categorization. International Journal of Artificial Intelligence & Applications (IJAIA), 13.

[18] Margarita Rodríguez-Ibanez, A. C.-V.-M.-M.-J. (2023). A review on sentiment analysis from social media platforms. Expert Systems With Applications 223.

[19] Maryanto, B. (2017). Big Data Dan Pemanfaatannya Dalam Berbagai Sektor. Media Informatika, 16(2), 14-20.

[20] Mikolov, T., Checn, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vectors Space. Google Inc., Mountain View, CA, 12.

[21] Mojumder, P., Hasan, M., Hossain, M. F., & Hasan, K. A. (2020). A Study of fastText Word Embedding Effects in Document Classification in Bangla Language. Research Gate, 13.

[22] Naira, A. R., Singha, R. P., Guptaa, D., & Kumar, P. (2024). Evaluating the Impact of Text Data Augmentation on Text Classification Tasks using DistilBERT. Procedia: Computer Science, 102-111.

[23] Ovirianti, N. H., Zarlis, M., & Mawengkang, H. (2022). Support Vector Machine Using A Classification Algorithm. Sinkron, 1-5.

[24] Pennington, J., Socher , R., & Manning, C. (2014). Glove: Global vectors for word representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) (https://rcpedia.stanford.edu/topicGuides/textProcessingWord_Embeddings.html ed.). California: Stanford University.

[25] Permatasari, P., Linawati, & Lie Jasa. (2021). Survei Tentang Analisis Sentimen Pada Media Sosial . Majalah Ilmiah Teknologi Elektro, 20(1), 177-186 .

[26] Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. Allen Institute for Artificial Intelligence, 1-15.

[27] Putra, K. T. (2023). Analisis Feature Extraction Pada Text Processing Untuk Analisis Sentimen. Thesis (Tidak diterbitkan ed.). Malang: Universitas Islam Negri Maulana Malik Ibrahim.

[28] Qamar, R., & Zardari, B. A. (2023). Artificial Neural Networks: An Overview. Mesopotamian, 130-139.

[29] Qorina, E. S. (2020). Analisis Perbandingan Metode Fast Text Dan Word2vec Pada Query Kesamaan Semantik Sistem Temu Kembali Informasi Sirah Nabawiyah. Jakarta: Universitas Islam Negeri Syarif Hidayatullah.

[30] Rodríguez-Ibanez, M., Casanez-Ventura, A., Castejon-Mateos, F., & Cuenca-Jim´enez, P.-M. (2023). A review on sentiment analysis from social media platforms. Expert Systems With Applications, 14.

[31] Thalib, I. (2019, November 1). NLP Preprocessing : Teknik Tokenisasi Untuk Memecah Kalimat menjadi Kata-Kata Pada Python. Retrieved from Medium: https://medium.com/@irfandy.thalib/teknik-tokenisasi-untuk-memecah-kalimat-menjadi-kata-kata-pada-python-12f799b74d49

[32] Vaswani, A., Shaazer, N., Parmar, N., Uszkoreit, J., Jonez, L., Gomez, A. N., . . . Polosukhin, I. (2023). Attention Is All You Need. Neural Information Processing Systems, 1-15.

[33] Wamidh K. Mutlag, S. K. (2020). Feature Extraction Methods: A Review . Journal of Physics: Conference Series , 1742-6596.

[34] Wang, A., Singh, A., Pruksachatkun, Y., Michael, J., Hill, F., Bowman, S. R., . . . Levy, O. (2020). SuperGLUE: A Stickier Benchmark for. Computational Language, 1-29.

[35] Wei , J., Zou, K., Chen , Y., & Lu, J. (2019). EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks. . Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Join Conference on Natural Language Processing (EMNLP-IJCNLP), (pp. 6381–6387). Hong kong.

[36] Yoon, K. (2014). Convolutional Neural Networks for Sentence Classification. New York University, 6.

[37] Ziedhan Alifio Dieksona, M. R. (2022). Sentiment analysis for customer review: Case study of Traveloka. 682–690.