Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)
Vol. 10, No. 3, September 2024, pp. 544-555
ISSN: 2338-3070, DOI: 10.26555/jiteki.v10i3.29450 

# From Text to Truth: Leveraging IndoBERT and Machine Learning Models for Hoax Detection in Indonesian News

Muhammad Yusuf Ridho, Evi Yulianti
Faculty of Computer Sciences, University of Indonesia, Depok 16424, Indonesia

**ARTICLE INFO**

**Article history:**

Received July 17, 2024
Revised August 31, 2024
Published September 10, 2024

**Keywords:**

IndoBERT;
Fake news detection;
Indonesian News Dataset;
Machine Learning;
Natural Language Processing;
Oversampling-SMOTE;
Text Classification;
Deep Learning;
Comparative model

**ABSTRACT**

In the era of technology and information exchange online content being deceitful poses a serious threat to public trust and social harmony on a global scale. Detective mechanisms to identify content are essential for safeguard the populace effectively. This study is dedicated to creating a machine learning system that can automatically spot deceptive content in Indonesian language by utilizing IndoBERT. A model specifically tailored for the intricacies of the Indonesian language. IndoBERT was selected due to its capacity to grasp the linguistic nuances present, in Indonesian text which are often challenging for other models built upon the BERT framework. The key focus of this study lies in conducting an assessment of the IndoBERT model in relation to other approaches used in past research for identifying fake news like CNN LSTM and various classification models such as Logistic Regression and Naïve Bayes among others. To address the issue of imbalanced data between valid labels in fake news detection tasks we employed the SMOTE oversampling technique, for data augmentation and balancing purposes. The dataset employed consists of Indonesian language news articles publicly available and categorized as either hoax or valid following assessment by three judges voting system. IndoBERT Large demonstrated performance by achieving an accuracy rate of 98% outperform the original datasets 92% when tested on the oversampled dataset. Utilizing the SMOTE oversampling technique aided in data balance and enhancing the models performance. These outcomes highlight IndoBERTs capabilities in detecting fake news and pave the way for its potential integration, into real world scenarios.



**Corresponding Author**:

Evi Yulianti, Faculty of Computer Sciences, University of Indonesia, Depok 16424, Indonesia
Email: evi.y@cs.ui.ac.id

## 1. INTRODUCTION

The role of the internet and technology in society, especially with the increasing amount of news circulating and easily obtained in seconds. In addition, technology has made the dissemination of news more effective by providing various channels such as social media and blogs and sites where people can read the latest news whenever they want. These news sites can now be easily distributed through chat groups and social media. This allows them to stay up to date with the latest news and provides an opportunity to discuss hot topics quickly [1].

However, despite having several advantages such as speed and ease of access, it also brings its own challenges, especially in terms of verifying information. Unverified information can spread widely in a short time, causing disinformation that has an impact on public opinion and social stability [2]. People often accept and spread information without verifying it, which can cause social unrest and conflict. With more and more information being disseminated through the internet and social media, the need for accurate and targeted

information is increasing. Alternatively, a reliable automatic detection system is becoming increasingly urgent. Detecting fake news and misinformation is a key solution to combating the spread of misinformation [3].

Hoax detection systems are designed to identify inaccurate or misleading information and flag it before it spreads further. They use a variety of methods, including natural language analysis, fact-checking, and algorithmic modeling, to assess the credibility of information [4]. The importance of this detection system lies in its ability to protect the public from the negative impacts of fake news, including a decline in public trust, negative impacts on mental health, and disruption to social stability. An effective detection system allows the public to more easily identify and avoid fake news, ensuring that the integrity of the information is maintained and decisions are made based on correct information. This also helps build a healthier and more sustainable information ecosystem [5].

Various efforts have been made to tackle the spread of fake news. The government, including the President, has emphasized how important it is to stop the spread of fake news, especially on social media, and reject untrue information. On the technical side, researchers have developed various tools and models to automatically detect inaccurate news [6].

Research on hoax detection has been conducted using the Long Short Term Memory (LSTM) model and the CheckThat!2021 task3a dataset. LSTM was chosen because it can handle continuous text and long-term relationships in messages. This step includes data processing (word abbreviations, stop words, spelling, punctuation) and text cleaning and word embedding using Word2Vec. The model was trained using TensorFlow 2.2.0 with hyperparameters such as LSTM size 128, attrition rate 0.25, batch size 32, and 30 epochs. As a result, the model achieved 98% accuracy on the training data and 55% accuracy on the validation data [1].

The next research proposes the use of the Naive Bayes model to detect fake news from a dataset consisting of 250 hoax and valid news articles in Indonesia. The data was divided into three ratios: 70:30, 80:20, and 60:40. At the 80:20 ratio, the accuracy reached 76%. Dataset development included web crawling, item identification, HTML removal, case folding, tokenization, stopword removal, and mutual information. The best results were achieved with a 70:30 ratio and an accuracy of 78.6%. The hoax precision rate was 67.1%, validity was 91.6%, hoax recall rate was 89.4%, and validity was 71.4%. This dataset is available on Mendeley OpenDataset under the name “Indonesian Hoax News Detection Dataset” [7].

The next study that will be discussed aims to overcome the shortcomings of previous models that use English datasets by adopting the Indonesian dataset, namely the "Indonesian Hoax News Detection Dataset" by Faisal *et al*. The models tested included LSTM, Bi-LSTM, GRU, Bi-GRU, and 1D-CNN, with 1D-CNN having the best accuracy of 97.9%. Key findings include the effectiveness of 1D-CNN with batch normalization as a feature extractor for NLP. Dropout techniques are useful for unidirectional recurrent neural networks and bidirectional neural networks, but are less suitable for GRUs because they can introduce NaNs. One-dimensional GRUs show better performance and efficiency than one-dimensional LSTMs, and neural networks outperform traditional classifiers in natural language processing [8].

Further research is to apply multilayer perceptron (MLP) to binary text classification in detecting fake news by comparing two feature extraction methods: TF-IDF and Bag of Words. In addition, an N-gram model is applied to improve accuracy. As a result, the model achieves a macro F1 score of 0.82. The study found that preprocessing steps such as stemming and stopword removal had minimal, and in some cases, negligible, impact on model performance. Using a combination of bag of words, character bigrams, and max_iter = 300, the model achieved a precision of 0.84 and a recall of 0.73 for fake news. Even with a max_iter of only 3, the model was able to achieve a macro F1 of 0.8, speeding up the process without significantly reducing accuracy [9]. The study further proposes a new approach called FakeBERT. It leverages BERT as a deep learning model due to its ability to capture semantic information and long-range relationships in text bidirectionally. FakeBERT combines BERT with a single-layer parallel CNN block with different kernel and filter sizes to improve feature extraction. The model uses parallel 1D CNN, pooling layers to reduce dimensionality, and a combination of dense and dropout layers to prevent overfitting. As a result, FakeBERT achieved 98.90% accuracy, outperforming previous models using one-way word embedding and 1D-CNN [10].

Further research also proposes an automated model for detecting fake news in resource-limited languages, such as Bengali, using a combination of CNN and LSTM with pre-trained GloVe word embeddings, Ta. CNN is used for feature extraction and LSTM processes long data sequences. The model is equipped with dropout layers and batch normalization to prevent overfitting. The results show an accuracy of 98.94% on the BanFakeNews and English Fake News datasets, surpassing the accuracy of the previous model combining CNN and GRU, which was 98.71% [11].

Further research aims to improve the ability to detect fake news by comparing the detection accuracy based on text features using the Logistic Regression (LR) and Support Vector Machine (SVM) algorithms. In

this study, both the LR and SVM algorithms were performed on fake news samples consisting of 311 datasets using a G-power of 80 and an alpha value of 0.05. The LR algorithm was used to evaluate its ability to detect fake news accurately, and the SVM algorithm was used for comparison. Based on the analysis results, the accuracy of detecting fake news with the LR algorithm reached 95.12%, and the accuracy of the SVM algorithm reached 91.68%. A statistically significant difference was found between the two sample groups, with a significance value of 0.079 for precision and 0.125 for precision. These results indicate that the LR algorithm is effective in identifying fake news compared to the SVM algorithm [12].

Another study proposed an innovative point-based fake news detection framework that automates the process of identifying fake news from multiple news sources. This framework aims to address the challenges of ensuring the authenticity of messages published on social media platforms. The first method uses the TF-IDF (Term Frequency – Inverted Document Frequency) technique to extract text-based features from news articles that are considered authentic and fake. Next, the credibility score of the news source is calculated based on the site_url and top-level domain (TLD) features. This framework can estimate the credibility level of news by combining text-based features and credibility scores of multiple sources. The proposed framework is applied to various machine learning (ML) classifiers to test its performance in detecting fake news. Experimental results show that the framework using the gradient boosting algorithm achieves the highest efficiency of about 99.5% [13].

This future research proposes an innovative method to classify proactive personality data in predicting fake and real news on social media, using a dataset of 25,000 entities with five attributes. Decision Tree (DT) and Random Forest (RF) algorithms were applied, with DT achieving 96% accuracy and RF 93%. With a significance value of ($P<0.05$), DT proved superior in accuracy and precision to RF. The dataset was treated with entropy method, feature addition, punctuation and stop words removal, and Z-score standardization. The evaluation results show that DT is more effective in identifying fake and real news than RF [14].

Further research investigates the challenges of managing imbalanced datasets in real-world applications, such as noise, overlapping classes, and small data fractions that affect classification accuracy. This study provides a comparative analysis of SMOTE variants with the aim of improving data complexity handling. In this study, by conducting experiments on 24 imbalanced datasets, we observe and observe the changes in complexity measures produced by different SMOTE variants, both in terms of F1 score and data complexity metrics. The experimental results yield two main conclusions. SMOTE can improve the classification performance of machine learning models on imbalanced datasets and reduce the complexity of the dataset. Second, the negative correlation indicates that reducing data complexity is associated with improving classification performance as measured by F1 score. In this case, reducing the data complexity metric N1 has a positive impact on classification performance [15].

Data balancing techniques such as class weighting, random sampling, smote, smotenn are based on fake news detection models such as Xgboost, Random Forest, CNN, Bigru, Bilstm, CNN-LSTM, CNN-Bigru. Evaluation is based on precision, AUC, precision, recall, and F1 scores in a balanced and unbalanced dataset. The results show that Smoteenn significantly improves the performance of the model in terms of F1 scores, precision, and acquisition. This study highlights the need for more advanced data balance strategies and the development of fake news detection models that are more accurate using oversampling techniques such as Smote, especially for Indonesia and data from a trusted news site [16].

Data imbalance is a common problem in text classification, especially when the number of one class (in this case fake news) is much smaller than other classes (real news). This imbalance can make the machine learning model biased to the majority class, causing poor detection performance for minority classes. To overcome this problem, use the oversampling technique to balance class distribution by adding synthetic copies or variants of a small number of data [15].

Further's research is contributing to running the Indobert model compared to the previous research modeling approach that can detect fake news, such as CNN-LSTM, Logistics Regression, Classification of Gradient Increased, Decision Trees, Random Forests, and Naive Bayes Evaluation. And technical applications propose oversampling to overcome data treatment. Smote technique (synthetic minority oversampling technique) was chosen because of its ability to produce synthetic examples of minority class, which helps improve the performance of the model in detecting fake news. Smote makes a new sample based on the interpolation between the existing minority class samples [17]. This technique not only increases the amount of minority data, but also avoids the overfitting problem that often occurs when simply copying existing examples [18]. By adding diversity in the minority data, SMOTE helps the model to better capture relevant patterns and improve fake news detection capabilities [19]. The dataset used in this study is taken from the Indonesian Hoax News Detection Dataset source published on Mendeley by Faisal Rahutomo *et al*. Thus, this

research not only contributes in terms of fake news detection methods but also offers a practical solution to overcome the challenge of data imbalance in the context of Indonesian language.

## 2. METHODS

This research methodology includes several things including dataset collection and data preparation, including cleaning and text vectorization. Then we use several deep learning models namely IndoBERT Large, IndoBERT Base, and CNN-LSTM, as well as machine learning methods such as Linear Regression, Decision Tree, Naive Bayes, Random Forest, and Gradient Boosting Classifier. After training these models, we analyzed their performance with various metrics. This research also analyzes the results using datasets with SMOTE oversampling technique and without oversampling. Full details about the methodology and dataset can be seen in Fig. 1.

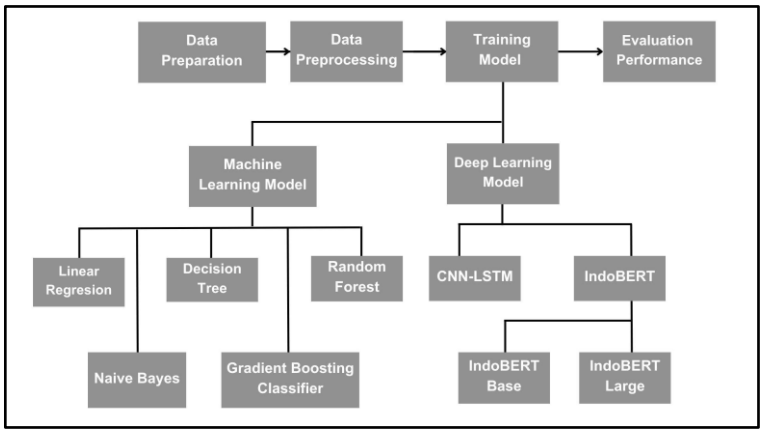


**Fig. 1.** Percentage of datasets

### 2.1. Dataset

In this study we conducted research using a dataset from previous researchers [7] which can be accessed at Mendeley Opendataset with a total of 600 data. The data consists of 372 validly labeled data and 228 invalidly labeled data with a total of 10 news topics and 12 news keywords. The dataset will be divided into train : test : val with a ratio of 80 : 10 : 10. The news titles include “Catfish contains cancer cells”, “Needle-pricked finger helps stroke patient”, “Iphone 6 bends easily”, “Reog Ponorogo burned in the Philippines”, “212 protesters can't enter Istiqlal mosque”, “Toothbrush made of pig hair”, and many more. The dataset comes from several news sites in Indonesia whose authenticity has been verified by experts and experts to certify that the news is valid, these news sites include teknoliputan6.com, tribunnews.com, madiunpos.com and tekno.kompas.com. The number of datasets can be seen in Fig. 2.

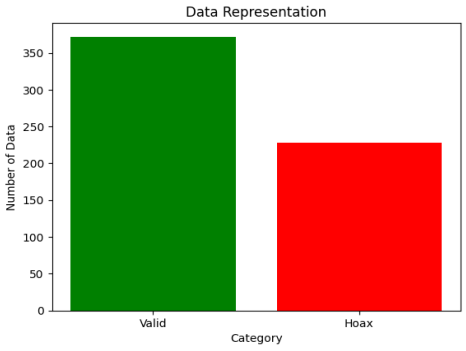


**Fig. 2.** Percentage of datasets

On the dataset used there is uniqueness, namely unbalanced data. Therefore, special techniques are needed to balance it. For this reason, we apply oversampling techniques using Synthetic Minority Oversampling (Smote) techniques. Previous research shows that this technique helps the in -depth learning model achieve higher accuracy in unbalanced data sets [20]. SMOTE works by producing new synthetic samples from minority classes, not only by duplicating existing samples, but also by interpolating between samples in the

minority class. In this way, Smote expands the minority class features space, allowing learning models from class patterns that are more diverse and rich [21]. By increasing the number of samples in the minority class, Smote ensures that the model has more data to be studied, so as to produce better prediction accuracy, especially for minority classes. In addition, Smote also helps prevent overfitting due to data reflection by giving more variations in minority classes and making the model not too dependent on the dominant pattern of the majority class [22].

The number of datasets can be seen in the following diagram. After applying the oversampling technique using SMOTE, the amount of data on each label becomes balanced. Previously, the dataset consisted of 372 data labeled as valid and 228 data labeled as hoax. After the SMOTE process, the amount of data for both labels becomes the same, which is 372 data for each label. An illustration of the number of datasets after the SMOTE process can be seen in Fig. 3.

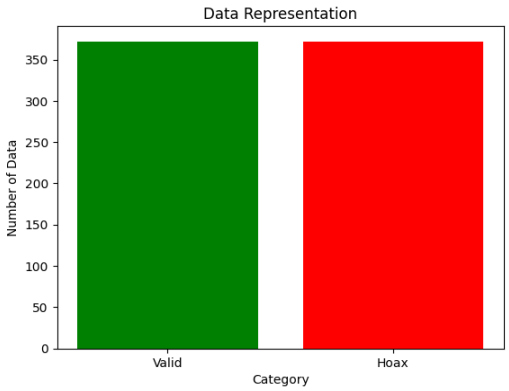


**Fig. 3.** The amount of data that has been oversampled

### 2.2. Preprocessing

In this study, the data used will be processed using several major techniques to ensure that the data is clean and ready to use. First, punctuation is removed to eliminate foreign elements that can interfere with analysis. This step is important to ensure that the model is not influenced by foreign symbols [23]. Second, the text is converted into lowercase letters to reduce capitalization differences that are not contextual, and the same words with different capitalization can be considered equal in the model [24]**.** The stemming and lemmatization technique is then applied to restore the word to its basic form, reducing unnecessary word variations and allows the model to more easily recognize patterns in data [25]**.** The removal of stop word is also carried out to reduce noise in the data by eliminating words that often arise but have no significance in the context of analysis [26]**.** The text normalization process is applied to maintain the consistency of the format, allowing the model to better understand and analyze data. In addition, improvisation of Indonesian grammar and morphology is also carried out, including Tokenisasi according to existing rules [27]. These comprehensive steps ensure that the data used to train and evaluate the model is clean, consistent, and representative, thereby improving the performance of the model in understanding and processing the community.

### 2.3. Text Vectorization

In this research, text vectorization is performed using various techniques to convert text data into a format that can be used by machine learning and deep learning models. One of the techniques used is TF-IDF (Term Frequency-Inverse Document Frequency), which is applied in traditional machine learning models. TF-IDF works by calculating the frequency of occurrence of each word in a given document (TF) and multiplying that value by the inverse logarithm of the frequency of documents containing that word in the entire data set (IDF) [28]. This means that a word that frequently appears in a particular document but rarely appears in the entire data set will have a higher weight, making it more significant in the analysis. In addition, the One Hot Encoding technique is used for the CNN-LSTM model. This technique converts text data into a binary vector where each word is represented by a value of 0 or 1. For example, if there are 10 unique words in the text, then each word will be represented by a vector of length 10 where one position is 1 and the rest are 0, depending on the position of the word in the pre-built dictionary [29]. One Hot Encoding allows deep learning models, such as CNN-LSTM, to process text data in a format that is easier to process and understand during the training process [30]. Meanwhile, for transformer-based models such as IndoBERT, text data is directly used without the need for vectorization processes such as TF-IDF or One Hot Encoding [31]. This is because IndoBERT has been pre-

trained with the Indonesian language corpus, allowing this model to better handle text representations using already optimized embeddings. The IndoBERT Base and IndoBERT Large models, used in this research, utilize the BERT architecture to capture the relationship between words in a sentence, thus being able to understand the context more deeply [32]. The various text vectorization techniques used, such as TF-IDF and One Hot Encoding, have proven to be very useful in training fake news detection models. The use of proper vectorization can improve the model's performance in identifying patterns in text, which in turn has a positive impact on prediction accuracy. By combining these approaches, this research seeks to obtain optimal results in detecting hoax news in Indonesian online media [33], [34].

### 2.4. Model Aprroach

In addition, we also adopt the latest BERT-based model, IndoBERT, in two variants: IndoBERT Base and IndoBERT Large, which have been pre-trained using Indonesian language corpus. IndoBERT is a transformer architecture-based language model designed to efficiently handle various natural language processing (NLP) tasks [35]**.** BERT (Bidirectional Encoder Representations from Transformers) is a language model designed to understand the context of a word by considering both directions, both from left to right and vice versa. The model is built on the transformer architecture [36]**,** which is a framework that enables parallel and efficient processing for NLP tasks [37]. The transformer works by using a self-attention mechanism, which allows the model to dynamically weigh the importance of each word in the context of the sentence. By using BERT, the model can capture the meaning of words in a broader context, which is crucial in tasks such as text classification, named entity recognition, and more [36]. Illustration of Bert architecture shown in Fig. 4.

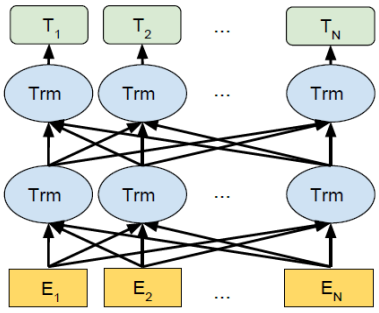


**Fig. 4.** Illustration of Bert architecture

IndoBERT was trained monolingually, meaning that this model was only trained using data from the Indonesian language. This training is done using the Huggingface framework, with settings similar to the original version of BERT designed for English. The model comes in two variants, namely IndoBERT Base and IndoBERT Large, which differ in size and complexity [35]. IndoBERT has been evaluated using the INDOLEM dataset, which is the largest and most comprehensive dataset for natural language processing tasks in Indonesian. The evaluation covers various important tasks such as part-of-speech tagging (POS tagging), named entity recognition (NER), sentiment analysis, text summarization, and more [38]. The evaluation results show that IndoBERT provides excellent performance and often outperforms other existing model [39]. Using IndoBERT, we were able to test how good the model is at detecting hoax news in Indonesian online media. This approach allows us to conduct an in-depth analysis and compare IndoBERT's performance with other models to determine which one is most effective in this task [40], [41].

### 2.5. Hyperparameter

In this study, we conducted experiments on Google Colab, utilizing the available GPU to speed up the model training process. The GPU used has a capacity of about 16GB, which is the standard GPU on the premium version of Google Colab. In addition, we also use 12GB of RAM, which is sufficient enough to run various machine learning and deep learning models on our dataset [42].For the IndoBERT-based model, we ran several experiments with different number of epochs. Given the complexity and size of the IndoBERT model, we chose to train for 5 to 10 epochs. This number of epochs was chosen based on the balance between the available computation time and the need to achieve model convergence. IndoBERT Base and IndoBERT Large each require more memory and computation time, so determining this number of epochs is important to avoid overfitting while still ensuring the model has learned the patterns from the data optimally [43]. In each epoch, the model updates the weights based on the trained data. By using available GPUs, the computation time for each epoch can be accelerated, allowing these experiments to be completed in a reasonable amount of

time despite using a model with high complexity such as IndoBERT [44]. In addition, we also tune other hyperparameters such as learning rate, batch size, and optimizer used [45]. For the IndoBERT model, we use the AdamW optimizer with a learning rate that is set incrementally using a learning rate scheduler. The use of this learning rate scheduler helps the model to adjust the learning rate during the training process, so that it can achieve its best performance [46]. In training the CNN-LSTM model and the traditional machine learning model, we also perform tuning on hyperparameters such as the number of neuron units, layers, and the type of activation function used. Each experiment conducted at Google Colab with this configuration has been customized to make optimal use of available resources and produce models that can perform hoax news detection with high accuracy [47].

### 2.6. Evaluation Models

In this study, we use several evaluation metrics to measure the performance of the proposed model, namely recall, precision, f1-score, accuracy, and confusion matrix. These metrics help us understand how well the model classifies the data, especially in detecting hoax news. Recall measures how well the model detects all true positive cases from all positive data. Simply put, recall shows how sensitive the model is to positive data [48]. Precision measures how accurate the model is in predicting positive data. It shows how many of the positive predictions are actually positive according to the ground truth [49]. F1-Score is a combined metric that calculates the harmonic mean of recall and precision. This metric is very useful when we want to balance between precision and recall, especially in situations where both are equally important [50]. Accuracy is the most commonly used metric to measure how often a model makes correct predictions (both positive and negative) out of all predictions made. It provides an overview of the model's performance [51]. Confusion Matrix is a table that provides a detailed overview of the model's performance. It shows the number of correct and incorrect predictions made by the model, divided into four categories: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) [52]. The formulas used to calculate these metrics are as follows:

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

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

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

$$F1\ Score = \frac{2x\ (Precision\ x\ Recal)}{Precision\ /\ Recall}$$

These metrics were chosen because each provides a different perspective on the model's performance, so by looking at all of them, we can get a more comprehensive picture of how well the model can predict hoax news. By understanding the strengths and weaknesses of the model through these metrics, we can be more effective in improving the model's performance in the future.

## 3. RESULTS AND DISCUSSION

### 3.1. Model Training

At this stage, various modeling approaches have been applied to improve the accuracy of fake news classification, with IndoBERT used as the main baseline. IndoBERT was chosen for its ability to deeply understand the context of Indonesian, especially in handling complex linguistic nuances. Trained using Masked Language Modeling (MLM) techniques on an Indonesian corpus, IndoBERT is able to capture rich and specific semantic context, making it highly effective for fake news detection in this language. Although it requires high computational resources, its advantage in improving classification accuracy on Indonesian datasets is significant [53]. CNN-LSTM architecture is applied to combine convolutional and sequential models, where CNN extracts high-level features from text and LSTM understands long-term relationships between words.

This architecture is ideal for sequential data as it is able to capture local patterns and temporal relationships, however its complexity can lead to overfitting on small datasets and requires longer training time. To overcome the problem of data imbalance in fake news classification, this research applies the oversampling technique SMOTE (Synthetic Minority Over-sampling Technique). This research uses SMOTE

(Synthetic Minority Over-sampling Technique) to overcome data imbalance in fake news classification. Unlike undersampling, which reduces the amount of data from the majority class to balance the dataset, SMOTE keeps the original dataset size intact while increasing the amount of minority class data. Undersampling can lead to the loss of valuable information from the majority class, which can result in decreased model performance on larger and more complex data [54]. Other oversampling techniques such as ADASYN (Adaptive Synthetic Sampling) and Borderline-SMOTE were also considered, but SMOTE was chosen due to its simplicity and proven effectiveness in various studies, as well as the ability to maintain a stable balance without adding additional complexity to the sampling process [55]. For comparison, several classic machine learning algorithms such as Logistic Regression, Naïve Bayes, Random Forest, Decision Tree, and Gradient Boosting Classifier are used. These models were selected based on their best performance in previous studies. Logistic Regression is known for its simplicity and probabilistic interpretation, although it is less effective on non-linear data. Naïve Bayes is efficient for large datasets, but its assumptions often do not hold. Random Forest and Decision Tree excel in handling complex data, but Random Forest requires longer training time and Decision Tree is prone to overfitting. Gradient Boosting Classifier is used for its ability to improve accuracy, although it requires longer training time and risks overfitting. These comparator models were selected based on recommendations from previous studies that have proven their effectiveness in fake news detection. By using proven models, this study provides a comprehensive evaluation of IndoBERT's performance in a broader context, ensuring high-standard comparisons. The implementation of IndoBERT is expected to make a significant contribution to Indonesian fake news detection, as the model is trained with an Indonesian language corpus, thus capturing the linguistic context more accurately than other models that are not adapted for this language.

### 3.2. Result

The discussed models will be run and evaluated using various metrics to understand the performance differences, analyze the results, and uncover the strengths and weaknesses of each model. This step is important to ensure that the best model is selected based on the most optimal performance in detecting fake news, while providing more accurate and relevant recommendations for real-world applications. Model using the original dataset shown in Table 1. Confusion Matrix the original dataset shown in Fig. 5.

**Table 1.** Model using the original dataset

| Machine Learning Model Name | Acc | P | R | F1 | Deep Learning Model Name | Acc | P | R | F1 |
|---|---|---|---|---|---|---|---|---|---|
| CNN-LSTM | 73% | 73% | 73% | 73% | Naive Bayes | 75% | 74% | 75% | 74% |
| IndoBert Base | 88% | 89% | 88% | 89% | Logistic Regression | 65% | 45% | 65% | 53% |
| IndoBert Large | 93% | 90% | 90% | 91% | Gradient Boosting Classifier | 55% | 43% | 55% | 48% |
| | | | | | Decision Tree | 61% | 57% | 61% | 57% |
| | | | | | Random Forest | 60% | 59% | 60% | 59% |

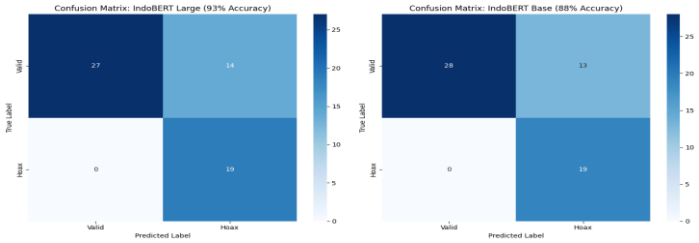


**Fig. 5.** Confusion Matrix the original dataset

The model developed in this study will also be compared with results from previous studies that used the SMOTE over-sampling technique (Table 2). This comparison aims to evaluate the effectiveness of the approach used, as well as ensuring that the results obtained can be compared equally with other methods. Confusion Matrix the Oversampling dataset shown in Fig. 6.

**Table 2**. Model using the Oversampling dataset

| Machine Learning Model Name | Acc | P | R | F1 | Deep Learning Model Name | Acc | P | R | F1 |
|---|---|---|---|---|---|---|---|---|---|
| CNN-LSTM | 44% | 44% | 43% | 44% | Naive Bayes | 78% | 75% | 79% | 75% |
| IndoBert Base | 97% | 97% | 98% | 96% | Logistic Regression | 76% | 77% | 77% | 76% |
| IndoBert Large | 98%% | 98% | 97% | 98% | Gradient Boosting Classifier | 70% | 67% | 69% | 68% |
| | | | | | Decision Tree | 67% | 70% | 68% | 66% |
| | | | | | Random Forest | 72% | 70% | 73% | 75% |

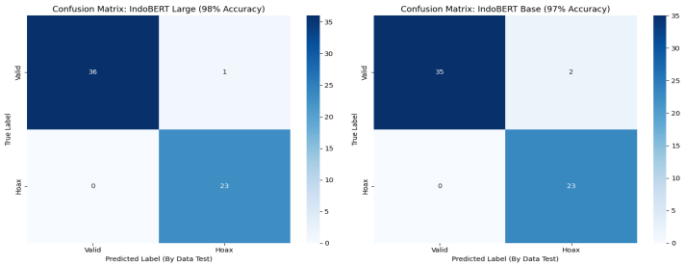


**Fig. 6.** Confusion Matrix the Oversampling dataset

### 3.3. Analysis

The experimental results showed significant performance differences between tested models, both in the original dataset and after oversampling. The large and simple version of Indobert always performs well. Indobert large reached 93%accuracy with 90%precision, 90%recall, and 91%F1 score. Indobert Base recorded 88%precision, 89%precision, 88%recall, and 89%F1 score. This advantage is thanks to the transformer architecture that is able to capture the context and relations between words in Indonesian speaking sources. Conversely, traditional models such as Naive Bayes and logistics regression have poor performance. Naïve Bayes recorded precision 75%, precision 74%, recall 75%, and 74%F1 score due to the assumption of feature independence, which is not suitable for complex text. Logistics regression is even worse, with 65%precision, 45%precision, 65%recall, and 53%F1 score due to the difficulty of capturing relationships between complex words. Decision -based models such as increased gradients, decision trees, and random forests also show less optimal results with accuracy between 55 and 61%. Despite oversampling, Indobert's performance continues to improve. Indobert large reached 98%accuracy with 98%precision, recall 97%, and F1 scores 98%, while Indobert Base reached 97%precision with 97%precision, recall 98%, and F1 scores 96%. This increase is caused by the ability of the transformer model in utilizing additional data effectively. However, the CNN-LSTM model has decreased performance after oversampling, so it only results in precision 44%, precision 44%, recall 43%, and F1 scores 44%due to overfitting in additional data. The traditional model also increases performance after oversampling, but not as much as Indobert. Naïve Bayes increased to 78% precision, 75% precision, 79% recall, and 75% F1 scores, while logistics regression reached 76% precision, 77% precision, 77% recall, and 76% score. Overall, Indobert is the most reliable system in detecting fake news in Indonesia, both from the original data and after oversampling, thanks to its sophisticated transformation architecture and the ability to adapt to additional additional data.

## 4. CONCLUSION

The conclusion of this experiment shows that the Indobert model is better in detecting fake news in Indonesia than other models, both in large and basic scale versions. In the original data collection, Indobert Large and Base has accuracy, precision, recall, and F1 scores that are much higher than traditional models such as Naive Bayes, Logistics Regression, and Decision Tree Based Models such as Decision Forest and Random Forest. The application of oversampling techniques significantly increases accuracy, precision, recall, and F1 scores, thus providing significant benefits for Indobert. Indobert reached 98% accuracy after oversampling, which shows its superior ability to utilize additional data to understand the complex linguistic context. Conversely, the CNN-LSTM model shows a decrease in performance after oversampling, which shows overfitting and the inability of the model to generalize. Although traditional models also increase performance

after oversampling, the increase is not comparable to Indobert. Overall, Indobert has proven to be the most effective model in detecting fake news in Indonesia. The benefits of highlighting the ability of transformer architecture to understand the nuances and context of complex language. Future research: For future research, several approaches can be applied to more complex attention mechanisms, such as: B. Exploring variations in the mechanism of attention such as attention and cross attention can increase models that capture informational inflation. In addition, transformer architecture is integrated with techniques such as graphic nerve networks (GNN) to create a hybrid model to understand complex relationships between entities in text. Finally, Indobert can be combined with other models to apply ensemble techniques to optimize the performance of fake news detection.

## Acknowledgments

I would like to express my deepest gratitude to Universitas Indonesia for their support and resources. Special thanks to Mrs. Evi Yulianti, S.Kom., M.Kom., Ph.D., my lecturer in the Information Retrieval course, for her invaluable guidance and encouragement. I also extend my heartfelt thanks to the LPDP scholarship program for providing the funding for my Master's studies in Computer Science at Universitas Indonesia. This research was funded by the Directorate of Research and Development, Universitas Indonesia, under Hibah PUTI Pascasarjana 2024 (Grant No. NKB-28/UN2.RST/HKP.05.00/2024).

## REFERENCES

[1] B. Majumdar, M. RafiuzzamanBhuiyan, M. A. Hasan, M. S. Islam, and S. R. H. Noori, "Multi Class Fake News Detection using LSTM Approach," i*n 2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART)*, pp. 75–79, 2021, https://doi.org/10.1109/SMART52563.2021.9676333.

[2] R. D. Abdiansyah, D. Mutiara, S. P. Sumedha, and N. Hanafiah, "Effective Methods for Fake News Detection: A Systematic Literature Review," *in 2021 1st International Conference on Computer Science and Artificial Intelligence (ICCSAI)*, pp. 278–283, 2021, https://doi.org/10.1109/ICCSAI53272.2021.9609777.

[3] M. A. Rahmat, Indrabayu, and I. S. Areni, "Hoax Web Detection For News in Bahasa Using Support Vector Machine," *in 2019 International Conference on Information and Communications Technology (ICOIACT)*, pp. 332–336, 2019, https://doi.org/10.1109/ICOIACT46704.2019.8938425.

[4] J. T. H. Kong, W. K. Wong, F. H. Juwono, and C. Apriono, "Generating Fake News Detection Model Using A Two-Stage Evolutionary Approach," *IEEE Access*, vol. 11, pp. 85067–85085, 2023, https://doi.org/10.1109/ACCESS.2023.3303321.

[5] S. S. Syam, B. Irawan, and C. Setianingsih, "Hate Speech Detection on Twitter Using Long Short-Term Memory (LSTM) Method," *in 2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, pp. 305–310, 2019, https://doi.org/10.1109/ICITISEE48480.2019.9003992.

[6] L. M. R. Rizky and S. Suyanto, "Improving Stance-based Fake News Detection using BERT Model with Synonym Replacement and Random Swap Data Augmentation Technique," *in 2021 IEEE 7th Information Technology International Seminar (ITIS)*, pp. 1–6, 2021, https://doi.org/10.1109/ITIS53497.2021.9791600.

[7] I. Y. R. Pratiwi, R. A. Asmara, and F. Rahutomo, "Study of hoax news detection using naïve bayes classifier in Indonesian language," *in 2017 11th International Conference on Information & Communication Technology and System (ICTS),* pp. 73–78, 2017, https://doi.org/10.1109/ICTS.2017.8265649.

[8] B. P. Nayoga, R. Adipradana, R. Suryadi, and D. Suhartono, "Hoax Analyzer for Indonesian News Using Deep Learning Models," *Procedia Comput.* Sci., vol. 179, pp. 704–712, 2021, https://doi.org/10.1016/j.procs.2021.01.059.

[9] A. Rusli, J. C. Young, and N. M. S. Iswari, "Identifying Fake News in Indonesian via Supervised Binary Text Classification," *in 2020 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*, pp. 86–90, 2020, https://doi.org/10.1109/IAICT50021.2020.9172020.

[10] R. K. Kaliyar, A. Goswami, and P. Narang, "FakeBERT: Fake news detection in social media with a BERT-based deep learning approach," *Multimed. Tools Appl.*, vol. 80, no. 8, pp. 11765–11788, 2021, https://doi.org/10.1007/s11042-020-10183-2.

[11] A. J. Keya, S. Afridi, A. S. Maria, S. S. Pinki, J. Ghosh, and M. F. Mridha, "Fake News Detection Based on Deep Learning," *in 2021 International Conference on Science & Contemporary Technologies (ICSCT),* pp. 1–6, 2021, https://doi.org/10.1109/ICSCT53883.2021.9642565.

[12] S. S. Birunda and R. K. Devi, "A Novel Score-Based Multi-Source Fake News Detection using Gradient Boosting Algorithm," *in 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS),* pp. 406–414, 2021, https://doi.org/10.1109/ICAIS50930.2021.9395896.

[13] B. Ganesh and Dr. K. Anitha, "Implementation of Personality Detection and Accuracy Prediction for identification of fake and true news using Decision Tree and Random Forest Algorithms*," in 2022 International Conference on Business Analytics for Technology and Security (ICBATS)*, pp. 1–5, 2022, https://doi.org/10.1109/ICBATS54253.2022.9759039.

[14] N. A. Azhar, M. S. Mohd Pozi, A. M. Din, and A. Jatowt, "An Investigation of SMOTE Based Methods for Imbalanced Datasets with Data Complexity Analysis (Extended Abstract)," *in IEEE 40th International Conference on Data Engineering (ICDE),* pp. 5735–5736, 2024, https://doi.org/10.1109/ICDE60146.2024.00499.

[15] E. Aljohani, "Enhancing Arabic Fake News Detection: Evaluating Data Balancing Techniques Across Multiple Machine Learning Models," *Eng. Technol. Appl. Sci. Res*., vol. 14, no. 4, pp. 15947–15956, 2024, https://doi.org/10.48084/etasr.8019.
[16] M. Das, "A Comparative Study on TF-IDF feature Weighting Method and its Analysis using Unstructured Dataset," *arXiv preprint arXiv:2308.04037,* 2017, https://doi.org/10.48550/arXiv.2308.04037.
[17] T. Pan and J. Yang, "An Oversampling Method Based on KL-Divergence for Imbalanced Datasets and Credit Risk Assessment," *in 2023 International Conference on New Trends in Computational Intelligence (NTCI)*, pp. 78–82, 2023, https://doi.org/10.1109/NTCI60157.2023.10403689.
[18] R. Das, S. Kr. Biswas, D. Devi, and B. Sarma, "An Oversampling Technique by Integrating Reverse Nearest Neighbor in SMOTE: Reverse-SMOTE," *in 2020 International Conference on Smart Electronics and Communication (ICOSEC),* pp. 1239–1244, 2020, https://doi.org/10.1109/ICOSEC49089.2020.9215387.
[19] S. U. Sabha, A. Assad, N. M. U. Din, and M. R. Bhat, "Comparative Analysis of Oversampling Techniques on Small and Imbalanced Datasets Using Deep Learning," *in 2023 3rd International conference on Artificial Intelligence and Signal Processing (AISP)*, pp. 1–5, 2023, https://doi.org/10.1109/AISP57993.2023.10134981.
[20] S. Veerla, A. V. Devadasan, M. Masum, M. Chowdhury, and H. Shahriar, "E-SMOTE: Entropy Based Minority Oversampling for Heart Failure and AIDS Clinical Trails Analysis," in *2024 IEEE 48th Annual Computers, Software, and Applications Conference (COMPSAC),* pp. 1841–1846, 2024, https://doi.org/10.1109/COMPSAC61105.2024.00291.
[21] A. Patil, A. Framewala, and F. Kazi, "Explainability of SMOTE Based Oversampling for Imbalanced Dataset Problems*," in 2020 3rd International Conference on Information and Computer Technologies (ICICT),* pp. 41–45, 2020, https://doi.org/10.1109/ICICT50521.2020.00015.
[22] K. Sharifani, M. Amini, Y. Akbari, and J. A. Godarzi, "Operating Machine Learning across Natural Language Processing Techniques for Improvement of Fabricated News Model," *International Journal of Science and Information System Research*, vol. 12, no. 9, pp. 20-44, 2022, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4251017.
[23] T. Felber, "Constraint 2021: Machine Learning Models for COVID-19 Fake News Detection Shared Task," arXiv: *arXiv:2101.03717*, 2024, http://arxiv.org/abs/2101.03717.
[24] L. de F. Santos and M. V. da Silva, "The effect of stemming and lemmatization on Portuguese fake news text classification," *arXiv: arXiv:2310.11344*, 2024, http://arxiv.org/abs/2310.11344.
[25] T. Khan, A. Michalas, and A. Akhunzada, "Fake news outbreak 2021: Can we stop the viral spread?," *J. Netw. Comput. Appl*., vol. 190, p. 103112, 2021, https://doi.org/10.1016/j.jnca.2021.103112.
[26] S. Pandey, S. Prabhakaran, N. V. Subba Reddy, and D. Acharya, "Fake News Detection from Online media using Machine learning Classifiers," *J. Phys. Conf. Ser.*, vol. 2161, no. 1, p. 012027, 2022, https://doi.org/10.1088/1742-6596/2161/1/012027.
[27] R. Kumar, "Fake News Detection using Passive Aggressive and TF-IDF Vectorizer," *International Research Journal of Engineering and Technology (IRJET)*, vol. 7, no. 12, 2020, https://www.irjet.net/archives/V7/i12/IRJET-V7I12158.pdf.
[28] J. Huang, "Detecting Fake News With Machine Learning," *J. Phys. Conf. Ser*., vol. 1693, no. 1, p. 012158, 2020, https://doi.org/10.1088/1742-6596/1693/1/012158.
[29] O. Oriola, "Exploring N-gram, Word Embedding and Topic Models for Content-based Fake News Detection in FakeNewsNet Evaluation," *Int. J. Comput. Appl*., vol. 176, no. 39, pp. 24–29, 2020, https://doi.org/10.5120/ijca2020920503.
[30] M. F. Mubaraq and W. Maharani, "Sentiment Analysis on Twitter Social Media towards Climate Change on Indonesia Using IndoBERT Model," *J. MEDIA Inform. BUDIDARMA*, vol. 6, no. 4, p. 2426, 2022, https://doi.org/10.30865/mib.v6i4.4570.
[31] D. I. Putri, A. N. Alfian, M. Y. Putra, and P. D. Mulyo, "IndoBERT Model Analysis: Twitter Sentiments on Indonesia's 2024 Presidential Election," *J. Appl. Inform. Comput*., vol. 8, no. 1, pp. 7–12, 2024, https://doi.org/10.30871/jaic.v8i1.7440.
[32] K. Li, "HAHA at FakeDeS 2021: A Fake News Detection Method Based on TF-IDF and Ensemble Machine Learning," *In IberLEF@ SEPLN*, pp. 630-638, 2021, https://ceur-ws.org/Vol-2943/fakedes_paper5.pdf.
[33] S. Coelho and A. Hegde, "MUCS@DravidianLangTech2023: Malayalam Fake News Detection Using Machine Learning Approach," *In Proceedings of the Third Workshop on Speech and Language Technologies for Dravidian Languages*, pp. 288-292, 2023, https://aclanthology.org/2023.dravidianlangtech-1.42.
[34] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," *arXiv: arXiv:2011.00677,* 2020, http://arxiv.org/abs/2011.00677.
[35] A. Vaswani *et al*., "Attention Is All You Need," *arXiv: arXiv:1706.03762*, 2023, http://arxiv.org/abs/1706.03762.
[36] L. M. Riza Rizky and S. Suyanto, "Improving Stance-based Fake News Detection using BERT Model with Synonym Replacement and Random Swap Data Augmentation Technique," *in 2021 IEEE 7th Information Technology International Seminar (ITIS),* pp. 1–6, 2021, https://doi.org/10.1109/ITIS53497.2021.9791600.
[37] S. M. Isa, G. Nico, and M. Permana, "IndoBERT for Indonesian Fake News Detection," *ICIC Express Letters*, vol. 16, no. 3, pp. 289-297, https://doi.org/10.24507/icicel.16.03.289.
[38] I. R. Hidayat and W. Maharani, "General Depression Detection Analysis Using IndoBERT Method," *Int. J. Inf. Commun. Technol. IJoICT*, vol. 8, no. 1, pp. 41–51, 2022, https://doi.org/10.21108/ijoict.v8i1.634.

[39] R. Anggrainingsih, G. M. Hassan, and A. Datta, "CE-BERT: Concise and Efficient BERT-Based Model for Detecting Rumors on Twitter," *IEEE Access*, vol. 11, pp. 80207–80217, 2023, https://doi.org/10.1109/ACCESS.2023.3299858.
[40] H. Jayadianti, W. Kaswidjanti, A. T. Utomo, S. Saifullah, F. A. Dwiyanto, and R. Drezewski, "Sentiment analysis of Indonesian reviews using fine-tuning IndoBERT and R-CNN," Ilk. J. Ilm., vol. 14, no. 3, pp. 348–354, 2022, https://doi.org/10.33096/ilkom.v14i3.1505.348-354.
[41] S. Vats, B. B. Sagar, K. Singh, A. Ahmadian, and B. A. Pansera, "Performance Evaluation of an Independent Time Optimized Infrastructure for Big Data Analytics that Maintains Symmetry," Symmetry, vol. 12, no. 8, p. 1274, 2020, https://doi.org/10.3390/sym12081274.
[42] B. Richardson and A. Wicaksana, "Comparison of IndoBERT-lite and RoBERTa in Text Mining for Indonesian Language Question Answering Application," *Int. J. Innov. Comput. Inf. Control*, vol. 18, no. 6, pp. 1719-1734, 2022, https://doi.org/10.24507/ijicic.18.06.1719.
[43] A. Jazuli, Widowati, and R. Kusumaningrum, "Aspect-based sentiment analysis on student reviews using the Indo-Bert base model," *E3S Web Conf.*, vol. 448, p. 02004, 2023, https://doi.org/10.1051/e3sconf/202344802004.
[44] C. F. Shaw, "A Transformer Based Architecture for Indonesian Sentiment Analysis - Exploring IndoBERT Variations, Training Size, and Self-Supervised Model Training," *Training Size, and Self-Supervised Model Training*, 2023, https://scholar.afit.edu/etd/7013/.
[45] P. F. Supriyadi and Y. Sibaroni, "Xiaomi Smartphone Sentiment Analysis on Twitter Social Media Using IndoBERT," *JURIKOM (Jurnal Riset Komputer)*, vol. 10, no. 1, pp. 19-30, 2023, https://garuda.kemdikbud.go.id/documents/detail/3324067.
[46] A. Zevana and D. Riana, "Text Classification Using Indobert Fine-Tuning Modeling with Convolutional Neural Network And Bi-Lstm," *J. Tek. Inform. Jutif*, vol. 4, no. 6, pp. 1605–1610, 2024, https://doi.org/10.52436/1.jutif.2023.4.6.1650.
[47] S. Chen, Y. Hou, Y. Cui, W. Che, T. Liu, and X. Yu, "Recall and Learn: Fine-tuning Deep Pretrained Language Models with Less Forgetting," *arXiv: arXiv:2004.12651*, 2020, http://arxiv.org/abs/2004.12651.
[48] S. R. Nandakumar *et al.,* "Mixed-Precision Deep Learning Based on Computational Memory," *Front. Neurosci.*, vol. 14, p. 406, 2020, https://doi.org/10.3389/fnins.2020.00406.
[49] R. Yacouby and D. Axman, "Probabilistic Extension of Precision, Recall, and F1 Score for More Thorough Evaluation of Classification Models," *in Proceedings of the First Workshop on Evaluation and Comparison of NLP Systems, Online: Association for Computational Linguistics*, pp. 79–91, 2020, https://doi.org/10.18653/v1/2020.eval4nlp-1.9.
[50] A. E. Maxwell, T. A. Warner, and L. A. Guillén, "Accuracy Assessment in Convolutional Neural Network-Based Deep Learning Remote Sensing Studies—Part 1: Literature Review," *Remote Sens.*, vol. 13, no. 13, p. 2450, 2021, https://doi.org/10.3390/rs13132450.
[51] M. Vakili, M. Ghamsari, and M. Rezaei, "Performance Analysis and Comparison of Machine and Deep Learning Algorithms for IoT Data Classification," *arXiv preprint arXiv:2001.09636*, 2020, https://doi.org/10.48550/arXiv.2001.09636.
[52] R. Merdiansah, S. Siska, and A. Ali Ridha, "Analisis Sentimen Pengguna X Indonesia Terkait Kendaraan Listrik Menggunakan IndoBERT," *J. Ilmu Komput. Dan Sist. Inf. JIKOMSI*, vol. 7, no. 1, pp. 221–228, 2024, https://doi.org/10.55338/jikomsi.v7i1.2895.
[53] F. R. -Torres, J. F. M. -Trinidad, and J. A. C. -Ochoa, "An Oversampling Method for Class Imbalance Problems on Large Datasets," *Appl. Sci.*, vol. 12, no. 7, p. 3424, 2022, https://doi.org/10.3390/app12073424.
[54] M. Muntasir Nishat *et al.,* "A Comprehensive Investigation of the Performances of Different Machine Learning Classifiers with SMOTE-ENN Oversampling Technique and Hyperparameter Optimization for Imbalanced Heart Failure Dataset," *Sci. Program.*, vol. 2022, pp. 1–17, 2022, https://doi.org/10.1155/2022/3649406.

## BIOGRAPHY OF AUTHORS

**Muhammad Yusuf Ridho** is currently pursuing a Master's degree in Computer Science at Universitas Indonesia. His research interests include computer vision, natural language processing, and artificial intelligence. He can be contacted at muhammad.yusuf25@ui.ac.id.

**Evi Yulianti** is a lecturer and researcher at the Faculty of Computer Science, Universitas Indonesia (Fasilkom UI). Her research interests include Information Retrieval, Summarization, and Machine Translation. She can be contacted at evi.y@cs.ui.ac.id.