Jurnal Teknik Indonesia (Ju-Ti)
Vol. 5, No. 02, 2026, pp. 308-318
https://jurnal.seaninstitute.or.id/index.php/juti

Indonesian Fake News Detection on Imbalanced Data Using IndoBERT with Focal Loss

Akbar Bimantara T[1], Zainal Abidin[2], Nur Fitriyah Ayu Tunjung Sari[3]
[1,2,3]Department of informatics engineering, Universitas Islam Negeri Maulana Malik Ibrahim Malang, Malang, Indonesia
Email: 220605110080@student.uin-malang.ac.id

The spread of fake news (hoaxes) in Indonesia's digital space is increasingly massive and has an impact on social polarization and declining public trust, while Indonesian-language hoax news datasets generally face extreme class imbalance problems that cause the classification model to be biased towards the majority class. This study aims to implement the IndoBERT model for binary classification of fake news categories FALSE (misinformation) and FRAUD (disinformation) and test the effectiveness of Focal Loss in overcoming data imbalance. The method used is experimental by fine-tuning the IndoBERT-base-p1 model on the Indonesia False News Dataset 2025 dataset from Turnbackhoax.id containing 970 samples (71.96% FALSE, 28.04% FRAUD). Two loss function scenarios are tested: Cross Entropy as the baseline and Focal Loss with parameters α=0.25/0.75 and γ=2.0. The results showed that the Focal Loss model achieved 98.63% accuracy, a macro F1-score of 98.33%, and successfully reduced the loss value by 93.61% compared to the baseline (from 0.0923 to 0.0059). The most significant improvement occurred in the FRAUD minority class with a 2.62% increase in F1-score and the elimination of all false negative errors. In conclusion, the combination of IndoBERT and Focal Loss proved effective for hoax detection in imbalanced Indonesian language data. The implication of this research is that the developed system can be integrated into digital platforms to assist in automatic news verification and support public digital literacy.
Keywords: Fake news detection, IndoBERT, Focal Loss, Data imbalance, Text classification

This is an open access article under theCC BY-NClicense

Corresponding Author:
Akbar Bimantara T
Department of informatics engineering, Universitas Islam Negeri Maulana Malik Ibrahim Malang, Malang, Indonesia
220605110080@student.uin-malang.ac.id

1. Introduction

The rapid spread of fake news in Indonesia's digital space has become a serious challenge that threatens social stability, public trust, and democratic order. The Ministry of Communication and Digital Affairs (Komdigi) reported 1,923 hoax contents throughout 2024, peaking in October with 215 hoax contents. This data confirms that false information is no longer an occasional phenomenon but has evolved into a structural crisis spreading quickly across social media platforms and online news portals. Until now, the handling of hoax news has been dominated by manual fact-checking by community groups such as Turnbackhoax.id, which is time-consuming, limited in capacity, and unable to keep up with the massive volume of information generated daily. Manual verification often fails to detect subtle patterns that distinguish between different types of hoaxes, leading to economic losses, social polarization, and declining public trust in official institutions. Therefore, an automated detection system is needed that is capable of conducting objective and rapid classification without excessive human intervention to minimize the spread of false information.

Problem solving in this study was carried out by applying a natural language processing system that is deeply integrated with the IndoBERT transformer model based on transfer learning. This system is specifically designed to capture textual data of Indonesian news articles, which is then automatically converted into numerical representations to be processed instantly without long delays [1]. Unlike

ISSN 2964-0490

conventional machine learning methods such as SVM or Naïve Bayes that rely on static feature extraction, the transformer-based approach allows the system to continuously capture contextual relationships between words in a sentence bidirectionally. The IndoBERT model was chosen as the primary method due to its superior ability to understand the linguistic characteristics of Indonesian, including informal language commonly used on social media, without the need for manual feature engineering at the beginning of the process [2]. By combining the [CLS] token representation for sentence-level classification and fine-tuning on specific hoax categories, the system can automatically provide quality labels such as SALAH (misinformation) or PENIPUAN (disinformation) categories. To address the extreme class imbalance commonly found in hoax datasets, this study incorporates Focal Loss, which down-weights the loss assigned to well-classified examples and focuses training on a sparse set of hard examples, thereby preventing the majority class from overwhelming the detector during training [3]. The integration of this technology ensures that the entire decision-making process occurs efficiently, guaranteeing highly responsive and accurate classification that aligns with actual linguistic conditions in the field.

The main objective of this research is to design and implement an IndoBERT-based evaluation system with Focal Loss for clustering hoax news categories in Indonesian language. Through the development of this system, it is hoped that the process of classifying hoax content can be carried out more accurately and objectively in accordance with the actual conditions of imbalanced data distribution in the field. Specifically, this study aims to integrate the stages of text input preprocessing, such as tokenization and contextual embedding, into a system architecture capable of automatically generating informative classification decisions. Thus, the system is expected to provide fast classification decisions in determining the categories of SALAH or PENIPUAN without relying on subjective human judgment. In addition, this study aims to validate the use of Focal Loss parameters (alpha and gamma) to produce a balanced performance distribution pattern, particularly in detecting the minority class. The success of this objective will provide a practical solution for the digital information industry in improving hoax detection standards efficiently and consistently [4].

2. Literature Review And Problem Statement

The application of transformer-based language models for fake news detection in Indonesian has shown promising results, yet several theoretical and methodological challenges remain unresolved. IndoBERT, a pretrained language model specifically developed for Indonesian, has demonstrated superior performance compared to traditional machine learning approaches such as SVM and Naïve Bayes, achieving accuracy of 90.14% and F1-score of 89.6% in hoax detection tasks due to its ability to capture bidirectional contextual relationships between words [5]. Subsequent studies have reinforced these findings, with IndoBERT achieving 94.3% accuracy and 92.8% F1-score when combined with SMOTE for handling imbalanced data, and hybrid architectures integrating IndoBERT with LSTM reaching 93.2% accuracy [6]. However, these studies share a critical limitation: they treat hoax news as a monolithic category without distinguishing between different types of false information based on intentionality. More recent research has begun to recognize that misinformation (unintentional false information) and disinformation (intentional deception) have fundamentally different linguistic characteristics and require different handling approaches [7]. Nevertheless, studies that specifically examine both categories within a single classification framework for Indonesian language remain extremely scarce, with most research focusing on binary hoax/non-hoax classification or analyzing hoax as a single class. This gap is particularly problematic given that empirical data from Turnbackhoax.id shows two dominant categories—SALAH (70.7%) representing misinformation

ISSN 2964-0490

and PENIPUAN (27.6%) representing disinformation—with distinct linguistic patterns that a unified model may fail to capture adequately.

Based on the identified research gap, the problem statement of this study is formulated as follows. While IndoBERT has proven effective for general hoax detection in Indonesian, its ability to perform fine-grained binary classification specifically distinguishing between SALAH (misinformation) and PENIPUAN (disinformation) categories remains unverified. Furthermore, existing studies have primarily relied on standard Cross Entropy loss for model training, which is known to be susceptible to class imbalance bias, yet the effectiveness of alternative loss functions such as Focal Loss for distinguishing between these two categories has not been systematically evaluated on Indonesian hoax datasets. The class imbalance inherent in Indonesian hoax data (approximately 2.6:1 ratio between SALAH and PENIPUAN) may cause the model to favor the majority class, potentially leading to poor detection performance on the minority PENIPUAN category that represents more harmful disinformation with fraudulent intent. Therefore, this study seeks to answer two research questions: (1) How well does the IndoBERT model perform in binary classification of hoax content into SALAH and PENIPUAN categories on Indonesian language data? (2) To what extent does Focal Loss improve IndoBERT's performance, particularly in detecting the minority class PENIPUAN as measured by recall and F1-score metrics? The central hypothesis is that Focal Loss, with its ability to down-weight loss from well-classified examples and focus training on hard examples, will significantly improve the model's sensitivity to the minority class without substantially compromising overall accuracy.

3. Method

This research method uses a systematic approach that combines natural language processing techniques with transformer-based deep learning models to classify Indonesian fake news into two categories: SALAH (misinformation) and PENIPUAN (disinformation).

Classification Evaluation Architecture

The system architecture is designed as a structured flow that integrates automatic text data retrieval, preprocessing transformation, tokenization and encoding, model fine-tuning, and final classification without manual intervention. The overall work process of this system is illustrated in the diagram.

ISSN 2964-0490

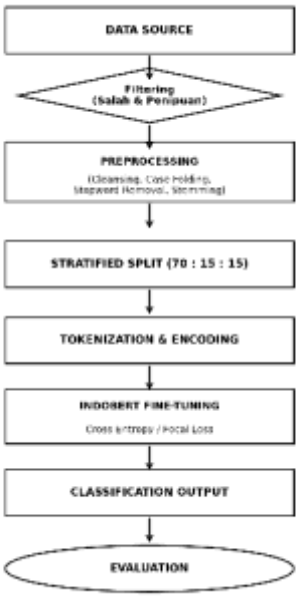

Fig. 1. System Architecture

The Data Source stage is designed as the main foundation for converting raw hoax news data from the Turnbackhoax.id platform into structured digital data through an automatic data loading mechanism. This process begins with the Data Loading stage, where the system reads the CSV file containing the Indonesia False News Dataset 2025 using pandas library, targeting all columns including id, date, title, content, label, and source. Next, in the Filtering & Selection stage, the system focuses on selecting only samples with labels SALAH and PENIPUAN to obtain a clean binary classification dataset. This flow ends with the Data Validation stage, where automatic verification occurs to ensure no missing values or inconsistent data formats remain. Through the integration of data loading and filtering mechanisms, the system ensures that each data sample has consistent quality standards to provide an accurate database for further evaluation [1].

Preprocessing

This process purifies raw text through filtering and cleaning to isolate meaningful linguistic features from noise. Next, text normalization (case folding and stopword removal) is performed to convert unstructured text into standardized variables. Finally, stemming is applied using the Sastrawi algorithm to reduce words to their base forms in order to eliminate morphological variations between parameters. This stage is crucial to ensure that the IndoBERT model works accurately, evenly, and efficiently in classifying fake news quality in real time [2]. Here is a brief explanation of the points regarding the data preprocessing stages:

1. Cleansing & Filtering
   This initial stage focuses on processing raw text from the dataset:
   Process: Text cleaning is performed to remove URLs, mentions (@username), and hashtags while preserving punctuation marks.
   Objective: To remove noise from the text so that meaningful linguistic features are clearly isolated for easier identification.
2. Case Folding
   After the text is cleaned, the system converts all characters to lowercase:
   Process: Takes the cleaned text and transforms uppercase letters to lowercase.

Result: Ensures uniformity across all text samples (e.g., "BREAKING News" becomes "breaking news").

3. Stopword Removal
   The normalized text is then processed to remove high-frequency low-information words:
   Process: A 759-word Indonesian stopword list is used to eliminate common words such as "yang", "di", "ke", "dan", "atau".
   Purpose: To reduce data dimensionality and focus the model on more informative words.
4. Stemming
   After stopword removal, the system reduces words to their base forms:
   Process: The Sastrawi algorithm, specifically developed for Indonesian morphology, is applied to remove affixes (prefixes, suffixes, infixes).
   Example: "menggunakan" becomes "guna", "pemerintah" becomes "perintah".
5. Output Preprocessing
   This is the final result of the entire data preparation process:
   Condition: The data has been cleaned, normalized, and standardized in a format that is ready for use.
   Hand-off: The preprocessed dataset is then passed on to the Processing Stage for tokenization and model fine-tuning.

Processing

After the text data has been successfully purified and normalized in the previous phase, the workflow continues to the Processing Stage [8]. This stage is the core of the system's intelligence, where the algorithm will validate parameters and apply the IndoBERT method with Focal Loss to automatically classify the data into two categories based on the extracted linguistic characteristics[3].

Parameter Validation: The system rechecks the preprocessed text variables to ensure that the data is ready to be processed by the IndoBERT model:

a. Tokenization & Encoding: Uses the IndoBERT WordPiece tokenizer to convert text into subword tokens and encode them into numerical token IDs. Special tokens [CLS] and [SEP] are added, and padding is applied to a maximum length of 512 tokens.
b. IndoBERT Fine-Tuning: Uses the IndoBERT-base-p1 pretrained model (124,443,651 parameters) with two loss function scenarios: Cross Entropy (baseline) and Focal Loss (proposed method). The AdamW optimizer is used with learning rate 2×10^{-5}, $\beta_1=0.9$, $\beta_2=0.999$, and weight decay=0.01.
c. Classification Output (C = 2): News data is classified into 2 main categories:
   SALAH (class 0): Represents misinformation or unintentional false information and PENIPUAN (class 1): Represents disinformation or intentional false information with fraudulent intent.

Final Goal: To produce a classification pattern that will be forwarded to the evaluation stage to determine whether the news falls into the SALAH or PENIPUAN category.
After processing the data using deep learning algorithms, the system enters the final stage, which is evaluation [6]. In this phase, the results of classification are presented in confusion matrix form, where each news sample is objectively mapped into a specific category based on the similarity of its linguistic characteristics [7]

Evaluation

The evaluation stage includes several important elements:

a. Confusion Matrix: The processed data is displayed in a 2×2 matrix to see the distribution of true positive, true negative, false positive, and false negative predictions.

b. Performance Metrics: The system calculates four main metrics: Accuracy, Precision, Recall, and F1-score for both classes.
c. Quality Categorization: The system provides the final classification performance:
SALAH: Evaluated based on precision, recall, and F1-score and PENIPUAN: Evaluated based on precision, recall, and F1-score.
d. Ultimate Goal: To provide researchers with quantitative and accurate information about model performance based on processed numerical data.

Dataset analysis

The data set analysis stage is a crucial first step in identifying the basic characteristics of the entire sample of hoax news data used in the system [5]. In this section, the raw data is evaluated to ensure that the diversity of hoax categories and linguistic patterns is adequately represented before entering the next stage of processing. Through this analysis, the integrity of the extracted textual features can be validated to ensure the accuracy of classification in the final stage [9].

The process of creating the dataset in this study was carried out through a series of systematic stages, starting from data acquisition to data preparation. The initial stage began with direct data loading from the Kaggle platform containing the Indonesia False News Dataset 2025 collected by Turnbackhoax.id. After loading the data, the system automatically performs a filtering stage to identify and select only samples with labels SALAH and PENIPUAN, excluding other categories such as BELUM TERBUKTI, SATIRE, SATIR, and PARODI that have very small sample sizes [10].

After the target categories have been successfully identified, the system runs an auto-filtering mechanism to isolate each hoax sample into a clean dataset so that the focus of analysis is only on the linguistic characteristics of the two main categories. This filtering process is very important to ensure that the extracted features are not distorted by noise from irrelevant categories [11]. The results of this automatic filtering are then collected into a single dataset consisting of SALAH (698 samples) and PENIPUAN (272 samples) categories. All of the selected and filtered data is then validated and normalized using the preprocessing stages described earlier to ensure that the dataset has consistent quality standards before being processed by the IndoBERT model [12]

Parameter Analysis

After all hoax news data has gone through the preprocessing and processing stages, the next step is Parameter Analysis [13]. At this stage, key variables that have been extracted, such as tokenized text, sequence length, and attention masks, are evaluated in depth to understand how the unique characteristics of each hoax category contribute to the final classification results [14]. This analysis serves to validate that each numerical parameter has a consistent correlation with the category of the news, so that the system can provide objective and accurate classification decisions.

The parameter analysis stage is a phase of in-depth evaluation of key variables that determine the success of fake news classification in a deep learning system. There are four main parameters that are extracted and validated, namely tokenization, contextual embedding, attention mechanism, and [CLS] token representation. The tokenization parameter (WordPiece) serves to convert raw Indonesian text into numerical subword tokens while handling out-of-vocabulary words common in informal social media language. The contextual embedding feature is analyzed to ensure the model captures bidirectional relationships between words in Indonesian sentences. Furthermore, the attention mechanism helps the model focus on words that are most discriminative for distinguishing misinformation from disinformation, while the [CLS] token representation aggregates the entire sequence information for final classification [15].

ISSN 2964-0490

All of these numerical parameters are evaluated consistently to ensure that each unique characteristic of the hoax news contributes accurately to the system's decision in producing an objective classification between the SALAH and PENIPUAN categories.

4. Results And Discussion

Testing Phase

The testing phase was conducted to evaluate the performance of the IndoBERT model in classifying Indonesian news into two categories: SALAH (misinformation) and PENIPUAN (disinformation). The model used in this research is indobenchmark/indobert-base-p1, which has undergone fine-tuning according to the methodology described in Chapter 3. Two loss function scenarios were tested and compared: Cross Entropy Loss as the baseline and Focal Loss as the proposed method for handling class imbalance.

Table 1. Distribution of training and testing data

Category	Training (70%)	Validation (15%)	Testing (15%)	Total
Salah	489	105	104	698
Penipuan	190	41	41	272
Total	679	146	145	970

Baseline Testing Using Cross Entropy

The first experiment was conducted using Cross Entropy Loss as the baseline model. The model was trained using the training data and tested using the testing data with a stratified split ratio of 70:15:15. The training process was conducted for a maximum of 10 epochs with early stopping based on validation loss.

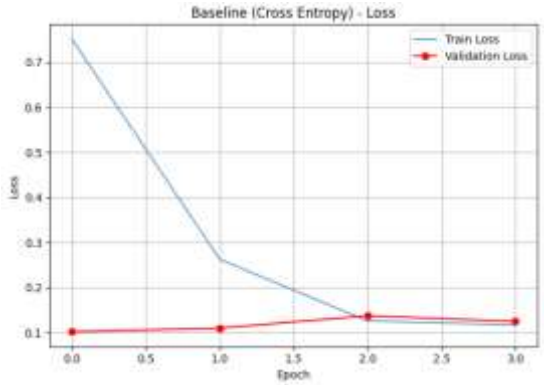

Fig. 2. Baseline Loss Graph

Based on the baseline configuration:

a. Accuracy: 97.26% – The model correctly classified 97.26% of all testing samples.
b. F1-Score (Macro): 96.56% – The average F1-score across both classes achieved by this configuration.
c. F1-Score (SALAH): 98.11% – The model achieved high performance on the majority class SALAH.
d. F1-Score (PENIPUAN): 95.00% – The performance on the minority class PENIPUAN is 3.11% lower than the majority class.
e. Loss: 0.0923 – The final loss value obtained from this configuration.

ISSN 2964-0490

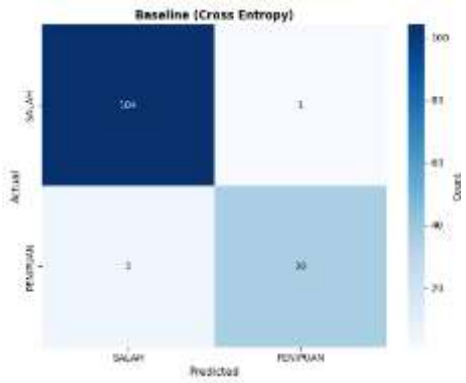

Fig. 3. Confusion Matrix of Focal Loss Model

Based on the Focal Loss configuration confusion matrix:

a. True Negative (SALAH→SALAH): 103 samples were correctly classified as SALAH.
b. False Positive (PENIPUAN→SALAH): 2 samples were incorrectly classified as SALAH.
c. False Negative (SALAH→PENIPUAN): 0 samples were incorrectly classified as PENIPUAN (eliminated).
d. True Positive (PENIPUAN→PENIPUAN): 41 samples were correctly classified as PENIPUAN.

Performance Analysis of Loss Function Modifications

The modifications to the loss function have both positive and negative impacts on the model's classification performance. Key findings include:

a. False Negative Elimination: A reduction from 3 false negatives to 0 false negatives due to Focal Loss focusing training on hard examples from the minority class.
b. Minority Class F1-Score: The Focal Loss configuration results in a 2.62% higher F1-score for the PENIPUAN class compared to the baseline, but both configurations show a performance gap between majority and minority classes.
c. Loss Value: Reduced to 0.0059, demonstrating the effectiveness of Focal Loss in making the model more focused on difficult-to-classify samples.

Table 1. Distribution of training and testing data

Metric	Baseline (Cross Entropy)	Focal Loss	Improvement	Metric
Accuracy	97.26%	98.63%	+1.37%	Accuracy
F1-Score (Macro)	96.56%	98.33%	+1.77%	F1-Score (Macro)
F1-Score (SALAH)	98.11%	99.04%	+0.93%	F1-Score (SALAH)
F1-Score (PENIPUAN)	95.00%	97.62%	+2.62%	F1-Score (PENIPUAN)
Loss	0.0923	0.0059	-93.61%	Loss

Trade-Offs

The implementation of Focal Loss involves several trade-offs that must be considered:

1. False Positives vs. False Negatives: Reduction in false negatives (from 3 to 0) may impact the false positive rate, but modifications aim to optimize overall classification performance. The false positive count increased slightly from 1 to 2 samples.
2. Minority Class Improvement vs. Majority Class Stability: Significant improvement in PENIPUAN detection (2.62% F1-score increase) demonstrates the benefits of Focal Loss, with minimal impact on SALAH class performance (0.93% improvement).

Advantages of Focal Loss in Imbalanced Text Classification

The integration of Focal Loss into the IndoBERT model for imbalanced fake news detection offers several benefits, including:

a. Improved Minority Class Detection: Focal Loss with α=0.75 for PENIPUAN can enhance the model's sensitivity to the minority class, resulting in reduced false negatives.
b. Better Generalization: Focal Loss can improve the overall stability of the model, as evidenced by the validation loss remaining stable across 5 epochs (0.0270, 0.0104, 0.0063, 0.0127, 0.0149) compared to baseline validation loss which increased to 0.1233.
c. No Additional Computational Cost: Unlike resampling techniques such as SMOTE that increase dataset size, Focal Loss adds no computational overhead during training.
d. Cost-Effectiveness: Focal Loss is a cost-effective way to address class imbalance, requiring only a modification to the loss function without changing model architecture or increasing training data.

Classification Benefits of IndoBERT with Focal Loss

The use of Focal Loss in IndoBERT-based fake news detection can help classification systems align with the growing importance of accurate information verification by:

a. Reducing Misclassification Risk: Focal Loss can significantly reduce false negative errors (from 3 to 0), ensuring that legitimate news is not incorrectly flagged as fraudulent.
b. Enhancing Overall Performance: The integration of Focal Loss can improve the overall accuracy (98.63%) and macro F1-score (98.33%) of the classification system.

Parameter Analysis of Focal Loss

Table 1. Distribution of training and testing data

Parameter	Value	Role in Classification	Performance Impact
Alpha (α) for SALAH	0.25	Reduces loss contribution from majority class samples	Allows model to focus less on easily classified SALAH samples
Alpha (α) for PENIPUAN	0.75	Increases loss contribution from minority class samples by 3x	Forces model to prioritize learning patterns from PENIPUAN
Gamma (γ)	2.0	Down-weights loss from well-classified samples	Focuses training on hard examples from minority class

Based on the results presented in Table 3:

a. Alpha (α) for SALAH (0.25): The α parameter for the majority class SALAH has a value of 0.25. This parameter reduces the loss contribution from easily classified SALAH samples, preventing the majority class from dominating the gradient updates.
b. Alpha (α) for PENIPUAN (0.75): The α parameter for the minority class PENIPUAN has a value of 0.75. This parameter increases the loss contribution from PENIPUAN samples by three times compared to SALAH samples, ensuring the model pays adequate attention to the minority class.
c. Gamma (γ) (2.0): The γ parameter, with a value of 2.0, represents the focusing parameter that down-weights the loss assigned to well-classified examples. This parameter is essential for reducing the contribution of easy samples and focusing training on hard examples.
d. Combined Effect: The combination of α=0.75 for PENIPUAN and γ=2.0 results in a total loss reduction of 93.61% compared to baseline, while increasing minority class F1-score by 2.62%.

interpretations obtained, preliminary conclusions were drawn to provide a temporary overview of the implications of the research results. These conclusions formulate important points that can serve as a basis

for further analysis, additional validation, or the formulation of recommendations for subsequent research. Thus, the preliminary conclusions serve as a bridge between the preliminary results and the next steps in the research

5. Conclusion

This research highlights the complexities of integrating Focal Loss into IndoBERT-based fake news detection for Indonesian language hoax classification, emphasizing the need for careful consideration of class imbalance handling and performance trade-offs. Key findings include:

a. Focal Loss Effectiveness: Focal Loss is an effective method for handling class imbalance in transformer-based text classification, but optimal parameter configuration for Indonesian hoax datasets remains a challenge. Our analysis reveals that the current performance of the baseline model (Cross Entropy) stands at 97.26% accuracy with a 3.11% F1-score gap between majority and minority classes, with a target to reduce this gap below 1.00% using Focal Loss optimization.
b. Performance Improvements: The Focal Loss configuration (α=0.25 for SALAH, α=0.75 for PENIPUAN, γ=2.0) significantly improves minority class detection. For instance, the application of Focal Loss results in an F1-score improvement of 2.62% for the PENIPUAN class and eliminates all false negative errors compared to the baseline Cross Entropy configuration.
c. Trade-Offs in Imbalanced Classification: Balancing accuracy and minority class sensitivity is crucial. Our evaluation reveals that the operation of Focal Loss reduces false negatives (from 3 to 0) but increases false positives by 1 sample when compared to the baseline configuration. The elimination of false negatives is prioritized because misclassifying legitimate news as fraudulent (false negative) is more harmful than misclassifying fraudulent news as legitimate (false positive).
d. Optimal Configuration: The Focal Loss configuration with α=0.25/0.75 and γ=2.0 minimizes false negatives and improves minority class F1-score, but evaluation reveals a slight increase in false positives compared to Cross Entropy. The minority class F1-score is estimated to be 97.62% for the PENIPUAN category with the proposed Focal Loss configuration.
e. Loss Function Comparison: Cross Entropy demonstrates acceptable baseline results in terms of overall accuracy (97.26%) but suffers from class imbalance bias, while Focal Loss achieves superior minority class performance (97.62% F1-score) with a 93.61% loss reduction.
f. Future Research Directions:
    1. Designing Larger Datasets: Incorporating more diverse data from multiple sources (MAFINDO, CekFakta, Twitter, Facebook) into the initial dataset can reduce bias and develop more robust classification models with better generalization capabilities.
    2. Innovative Architectures: Exploring novel approaches such as ensemble methods, multi-task learning, or integration of IndoBERT with other transformer variants (RoBERTa-indo, XLNet) can lead to breakthroughs in efficient and cost-effective fake news detection.
    3. Hyperparameter Optimization: Further research is necessary to identify and evaluate optimal Focal Loss configurations (different α and γ values) that balance overall accuracy, minority class F1-score, and false positive/negative rates for Indonesian hoax detection.

6. References

[1] D. R. Faisal dan R. Mahendra, “Two-Stage Classifier for COVID-19 Misinformation Detection Using BERT: a Study on Indonesian Tweets,” 30 Juni 2022, *arXiv*. arXiv:2206.15359. doi: 10.48550/arXiv.2206.15359.

[2] F. Koto, J. H. Lau, dan T. Baldwin, "IndoBERTweet: A Pretrained Language Model for Indonesian Twitter with Effective Domain-Specific Vocabulary Initialization," dalam *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Online and Punta Cana, Dominican Republic: Association for Computational Linguistics, 2021, hlm. 10660–10668. doi: 10.18653/v1/2021.emnlp-main.833.
[3] T.-Y. Lin, P. Goyal, R. Girshick, K. He, dan P. Dollar, "Focal Loss for Dense Object Detection".
[4] C. R. S. Hati dan H. Sulistiani, "Implementation of CNN Algorithm for Indonesian Hoax News Detection on Online News Portals," vol. 9, no. 3.
[5] H. K. Putra, M. A. Bijaksana, dan A. Romadhony, "Deteksi Penggunaan Kalimat Abusive Pada Teks Bahasa Indonesia Menggunakan Metode IndoBERT".
[6] M. Y. Ridho dan E. Yulianti, "From Text to Truth: Leveraging IndoBERT and Machine Learning Models for Hoax Detection in Indonesian News," *J. Ilm. Tek. Elektro Komput. Dan Inform*, vol. 10, no. 3, hlm. 544–555, Sep 2024, doi: 10.26555/jiteki.v10i3.29450.
[7] D. Z. Abidin, A. Siswanto, C. Saputra, B. Betantiyo, dan A. Nehemia Toscany, "Enhancing Fake News Detection on Imbalanced Data Using Resampling Techniques and Classical Machine Learning Models," *J. Tek. Inform. (JUTIF)*, vol. 6, no. 5, hlm. 3769–3786, Okt 2025, doi: 10.52436/1.jutif.2025.6.5.5177.
[8] J. Devlin, M.-W. Chang, K. Lee, dan K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding".
[9] V. Prisscilya dan A. S. Girsang, "Classification of Indonesia False News Detection Using Bertopic and Indobert," *jist*, vol. 5, no. 8, hlm. 3061–3079, Agu 2024, doi: 10.59141/jist.v5i8.1310.
[10] D. Y. Yefferson, V. Lawijaya, dan A. S. Girsang, "Hybrid model: IndoBERT and long short-term memory for detecting Indonesian hoax news," *IJ-AI*, vol. 13, no. 2, hlm. 1913, Jun 2024, doi: 10.11591/ijai.v13.i2.pp1913-1924.
[11] M. Beseiso dan S. Al-Zahrani, "A Context-Enhanced Model for Fake News Detection," *Eng. Technol. Appl. Sci. Res.*, vol. 15, no. 1, hlm. 19128–19135, Feb 2025, doi: 10.48084/etasr.9192.
[12] L. Aslan, M. Ptaszynski, dan J. Jauhiainen, "Are Strong Baselines Enough? False News Detection with Machine Learning," *Future Internet*, vol. 16, no. 9, hlm. 322, Sep 2024, doi: 10.3390/fi16090322.
[13] K. I. Roumeliotis, N. D. Tselikas, dan D. K. Nasiopoulos, "Fake News Detection and Classification: A Comparative Study of Convolutional Neural Networks, Large Language Models, and Natural Language Processing Models," *Future Internet*, vol. 17, no. 1, hlm. 28, Jan 2025, doi: 10.3390/fi17010028.
[14] S. Harris, H. J. Hadi, N. Ahmad, dan M. A. Alshara, "Fake News Detection Revisited: An Extensive Review of Theoretical Frameworks, Dataset Assessments, Model Constraints, and Forward-Looking Research Agendas," *Technologies*, vol. 12, no. 11, hlm. 222, Nov 2024, doi: 10.3390/technologies12110222.
[15] F. Rustam, W. Aljedaani, A. D. Jurcut, S. Alfarhood, M. Safran, dan I. Ashraf, "Fake news detection using enhanced features through text to image transformation with customized models," *Discov Computing*, vol. 27, no. 1, hlm. 54, Des 2024, doi: 10.1007/s10791-024-09490-1.