Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15260
e-ISSN : 2541-2019
p-ISSN : 2541-044X

# Fairer Public Complaint Classification on LaporGub: Integrating XLM-RoBERTa with Focal Loss for Imbalance Data

**Azzula Cerliana Zahro[1)], Farrikh Alzami[2)], Ramadhan Rakhmat Sani[3)], Amiq Fahmi[4)], Rama Aria Megantara[5)*] Muhammad Naufal[6)], Harun Al Azies[7)]. Iswahyudi[8)]**
[1,2,3,4)]Sistem Informasi, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Semarang, Indonesia
[5,6,7)]Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Semarang, Indonesia
[8)]Dinas Komunikasi dan Informatika Provinsi Jawa Tengah
[1)]azzulacerliana@gmail.com, [2)]alzami@dsn.dinus.ac.id, [3)]ramadhan_rs@dsn.dinus.ac.id,
[4)]amiq_fahmi@dosen.dinus.ac.id, [5)]aria@dsn.dinus.ac.id, [6)]m.naufal@dsn.dinus.ac.id,
[7)]harun.alazies@dsn.dinus.ac.id, [8)]iswahyudi@jatengprov.go.id

**Submitted** :Aug 22, 2025 | **Accepted** : Sep 18, 2025 | **Published** : Oct 2, 2025

**Abstract:** The advancement of digital technology has provided opportunities for governments to improve the quality of public services through citizen complaint channels. One example of this implementation in Indonesia is *Lapor Gub*, managed by the Dinas Komunikasi dan Informasi Provinsi Jawa Tengah (Communication and Information Agency of Central Java Province). This platform receives thousands of complaints daily, ranging from infrastructure, social issues, to illegal levies. However, the large volume of data and the imbalanced distribution of categories pose significant challenges for both manual and automated processing. This study aims to classify citizen complaint texts using XLM-RoBERTa combined with Focal Loss as an approach to handle data imbalance. The dataset consists of 53,774 complaints after data cleaning and text preprocessing. The training process applied a stratified split (78% training, 18% validation, 10% testing) and fine-tuning for 10 epochs. Model performance was evaluated using accuracy, precision, recall, and macro F1-score. The results show that the model without Focal Loss achieved 78.1% accuracy with a macro F1-score of 0.606, while the model with Focal Loss improved the macro F1-score to 0.625 with 78.5% accuracy. Unlike previous studies that relied on resampling or data augmentation, this research demonstrates how Focal Loss improves fairness without altering data distribution, offering a more robust solution for imbalanced text classification. These findings highlight the novelty of integrating Focal Loss with XLM-RoBERTa in the Indonesian public complaint context. In practice, this approach enables governments to process citizen complaints more fairly and transparently, ensuring that minority categories such as “illegal levies” and “tourism” are not overlooked while maintaining strong performance on majority classes.

**Keywords:** RoBERTa, Focal Loss, Text Classification

## INTRODUCTION

In today's digital era, the government's ability to interact with and listen to the public is key to effective and responsive governance. Public complaints are one way to encourage public participation in supporting government efforts to improve the quality of public services (Kristina et al., 2023) . Various digital platforms and online complaint channels have emerged as important bridges for citizens to convey their aspirations, criticisms, and complaints regarding public services. The existence of an efficient and transparent complaint system not only increases government accountability but also directly contributes to improving service quality and public trust.

To facilitate communication between the public and the government, Dinas Komunikasi dan Informasi Provinsi Jawa Tengah (Communication and Information Agency of Central Java Province) has established a public complaints channel called **"Lapor Gub"**. This channel receives various types of complaints from the public, such as potholes**,** inadequate public facilities, illegal levies, environmental issues, and so on. However, this ease of access also presents significant challenges. The volume of complaints received daily is enormous, often exceeding

*name of corresponding author

This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

manual processing capacity. This makes it quite difficult for administrators to categorize public complaints (Afida et al., 2021) . Furthermore, it also poses significant obstacles to the process of disposing of complaints quickly and accurately to relevant agencies or services. Therefore, an effective automation method for categorizing complaints that can manage the massive volume of data is needed so that complaints can be followed up effectively.

Text mining is the process of converting unstructured text into a structured format to discover patterns or useful information (Sani et al., 2022) . This process is part of Natural Language Processing (NLP) which focuses on identifying and extracting meaning in text (Mufidah et al., 2022) . To support this process, an appropriate text classification method is needed to obtain accurate and optimal results. Various methods have been applied, such as the Naïve Bayes Classifier, K-Nearest Neighbor (K-NN), Support Vector Machine (SVM), and Decision Tree. In the realm of Natural Language Processing (NLP), the task of **text classification** has undergone rapid evolution. However, these traditional approaches show significant limitations in handling the semantic complexity, informal language, and unstructured nature of complaint texts in Indonesia.

RoBERTa, as a transformer-based language model, has demonstrated outstanding capability in capturing contextual meanings within sentences. Unlike traditional approaches that require complex preprocessing such as stemming, stopword removal, and punctuation elimination, RoBERTa is able to process raw text while retaining its structural and semantic richness. This strength makes it highly relevant for handling public complaints, which often contain varied writing styles, informal expressions, and local dialects that may not conform to standard linguistic structures. However, despite the advantages offered by RoBERTa, the issue of imbalanced data remains a critical challenge. In the *Lapor Gub* dataset, certain categories such as “Infrastructure” dominate the majority of complaints, while others, like “Tourism and Culture” or “Illegal Levies,” appear far less frequently. This imbalance can cause the model to perform poorly on minority classes, as it tends to prioritize learning patterns from the majority categories, leading to biased and inaccurate classification results (Rahma & Suadaa, 2023).

To address this issue, this study integrates Focal Loss into the fine-tuning process of the RoBERTa model. Focal Loss is designed to give greater weight to misclassified or minority class samples, ensuring that the model focuses more attention on the harder-to-learn instances. By adjusting the learning emphasis, Focal Loss mitigates the model’s bias toward majority categories without the need for oversampling or undersampling techniques, which often distort the original data distribution. The use of Focal Loss has been widely recognized in handling imbalanced datasets, particularly in computer vision tasks. Its application in Natural Language Processing, however, has gained growing attention in recent years as researchers increasingly encounter textual datasets with skewed distributions. While Focal Loss has been widely used in computer vision tasks, its application in Natural Language Processing especially in Indonesia remains limited. Research on public complaint text classification in Indonesia has rarely addressed fairness for minority categories such as illegal levies and tourism. This study aims to fill this gap by evaluating the effectiveness of combining RoBERTa with Focal Loss for fairer complaint classification (Nemoto et al., 2021).

Another advantage of combining RoBERTa with Focal Loss is the ability to preserve the richness of the original dataset while enhancing model performance. Since no artificial balancing methods are introduced, the model learns directly from the authentic complaint data. This is particularly valuable in government contexts, where accuracy and fairness in representing public concerns are critical to maintaining public trust and ensuring that less frequently raised issues are not ignored. In addition to improving fairness, the proposed model contributes to efficiency in processing large-scale textual data. With more than 50,000 complaint records in the dataset, manual classification is impractical. An automated system that combines the linguistic strength of RoBERTa with the balancing mechanism of Focal Loss has the potential to accelerate the categorization process, enabling government agencies to respond more quickly and appropriately to citizens’ concerns (Abdel-salam, 2022).

The significance of this research lies not only in its technical contribution to NLP but also in its social and administrative impact. By providing a more accurate and balanced classification of complaints, the model supports better allocation of government resources, reduces response time, and ensures that minority issues receive due attention. This aligns with the broader goals of e-government, which emphasizes inclusivity, responsiveness, and accountability in public service delivery. Furthermore, the implementation of advanced machine learning techniques in public complaint management reflects the growing role of Artificial Intelligence (AI) in governance. It demonstrates how governments can harness AI-driven solutions to enhance transparency and responsiveness, thereby strengthening democratic practices through participatory channels. When citizens see their complaints categorized and addressed fairly, their trust in public institutions is reinforced (Song et al., 2021).

This study also highlights the importance of evaluating multiple performance metrics in imbalanced classification tasks. While overall accuracy provides a general indication of performance, it can be misleading in cases of imbalance, as the model may perform well on majority classes but fail to recognize minority categories. Metrics such as macro F1-score, precision, and recall offer a more balanced and comprehensive assessment of the model’s true capabilities. In conclusion, this study emphasizes the dual gap: the limitations of traditional models in complaint classification (Mufidah et al., 2022; Sani et al., 2022) and the lack of Focal Loss application in NLP Indonesia (Kunaefi et al., 2025; Rahma & Suadaa, 2023). By addressing these issues, the research not only

*name of corresponding author

This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

e-ISSN : 2541-2019
p-ISSN : 2541-044X

advances text classification methodology but also offers practical implications for improving governance and public service delivery (Azadi et al., 2024, 2024).

## LITERATURE REVIEW

RoBERTa, as a *Transformer- based language mode*, has a good ability to capture the contextual meaning of words in sentences. This advantage means that RoBERTa does not require complex *preprocessing*, such as punctuation removal, because its architecture is designed to understand the complete structure of language, including punctuation, stopwords, stemming, etc. Furthermore, RoBERTa has demonstrated superior performance in various natural language processing (NLP) tasks and has high flexibility in handling various types of text. Previous research has shown that the use of pre-trained language models, particularly XLM-RoBERTa, has brought significant improvements in sentiment analysis for Indonesian (Wiciaputra et al., 2021) .

While RoBERTa offers great potential, the "Lapor Gub" complaint data likely has an imbalanced category distribution. Some complaint categories, such as Saber Pungli (Extortion), Pariwisata (Tourism), and others, may have significantly fewer *instances* than others. This imbalanced data condition causes classification models to tend to ignore minority classes and focus on the majority class during training. If left unaddressed, it can impact classification model performance, leading to bias and inaccuracy in predicting classes with fewer instances (Rahma & Suadaa, 2023).

To address this *imbalanced data issue*, this study will integrate **Focal Loss** into the RoBERTa model. **Focal Loss** is a technique used to increase the model's sensitivity to minority classes by focusing the learning process on difficult samples. Focal Loss's advantage lies in its ability to achieve this without requiring data modifications, such as oversampling or undersampling (Kunaefi et al., 2025) .

Focal Loss offers a more efficient solution for handling data imbalance because it doesn't alter the original data distribution. This allows the model to learn from the full range of available data. Therefore, combining XLM-RoBERTa with Focal Loss is a powerful approach for improving classification performance, especially for underrepresented or minority categories.

Several studies in the field of Natural Language Processing (NLP) have emphasized the importance of transformer-based models in advancing text classification. The introduction of BERT and its variants, including RoBERTa, has marked a significant shift from traditional methods such as Naïve Bayes and Support Vector Machines. These older approaches often rely heavily on handcrafted features and bag-of-words representations, which fail to capture semantic relationships within text. In contrast, transformer-based models learn contextual embeddings that allow for deeper understanding of linguistic nuances, making them more effective for complex tasks such as public complaint classification. RoBERTa, which stands for *A Robustly Optimized BERT Pretraining Approach*, refines BERT's pretraining methodology by removing the Next Sentence Prediction (NSP) objective, increasing training data size, and extending training time. Liu et al. (2019) demonstrated that RoBERTa significantly outperformed BERT across a wide range of NLP benchmarks. This robustness and scalability have made RoBERTa a preferred choice for many classification tasks, including multilingual applications where language variations pose additional challenges (Zhao et al., 2024).

In the Indonesian context, the application of transformer-based models has shown promising results. Wiciaputra et al. (2021) demonstrated that XLM-RoBERTa performed exceptionally well in bilingual text classification tasks, outperforming conventional machine learning algorithms. Their study underlined the adaptability of RoBERTa to Indonesian text, which is often characterized by informal language, abbreviations, and mixed vocabulary. This finding supports the potential of RoBERTa in handling user-generated complaint texts from *Lapor Gub*. However, the issue of class imbalance remains a recurring challenge in text classification research. Rahma & Suadaa (2023) highlighted that imbalanced datasets can cause models to misclassify minority classes while achieving deceptively high overall accuracy. This issue is particularly problematic in public complaint datasets, where less frequent categories, such as "Illegal Levies" or "Tourism," may still hold critical importance for policy decisions. Addressing this imbalance is therefore essential to ensure fairness and inclusivity in automated complaint management (Vasyl et al., 2024).

Various strategies have been proposed to overcome data imbalance, including resampling techniques like oversampling minority classes or undersampling majority classes. While effective to some extent, these approaches often distort the natural distribution of the dataset, leading to overfitting or information loss. Cost-sensitive learning, which adjusts the weight of loss functions, has also been explored. Among these, Focal Loss has emerged as a particularly effective method due to its ability to emphasize harder-to-classify samples without altering the data distribution. Originally introduced in computer vision for object detection tasks, Focal Loss modifies the standard cross-entropy loss by adding a modulation term that reduces the relative loss for well-classified examples. Lin et al. (2017) showed that this mechanism forces the model to focus more on difficult or misclassified samples. In NLP, the adaptation of Focal Loss has shown potential in handling imbalanced datasets, as it increases model sensitivity toward minority categories while maintaining performance in majority classes (Nemoto et al., 2021).

*name of corresponding author
This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15260
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Recent research in Indonesia has also applied Focal Loss in the context of text classification. Kunaefi et al. (2025) explored its application in hoax news classification using IndoBERT and reported significant improvements in macro F1-score compared to models trained with standard cross-entropy loss. Their study highlights that Focal Loss can enhance the robustness of transformer-based models when facing skewed data distributions. This evidence supports its integration with RoBERTa for classifying *Lapor Gub* complaints. Beyond technical improvements, the use of Focal Loss also carries implications for governance. By enabling models to more accurately classify underrepresented categories, minority issues that might otherwise be overlooked can receive proper attention. For instance, complaints about illegal levies or tourism may not occur as frequently as infrastructure-related issues, but they remain vital for improving public services. Thus, applying Focal Loss ensures that all categories of complaints are represented fairly in the classification process (Younes & Mathiak, 2021).

Moreover, evaluating classification performance in imbalanced contexts requires more than accuracy as a single metric. Researchers often recommend macro-averaged precision, recall, and F1-score to provide a fairer assessment across categories (Sani et al., 2022). These metrics treat each category equally, regardless of its frequency, making them particularly suitable for evaluating models trained with Focal Loss. By adopting these metrics, researchers can more transparently assess whether the model genuinely improves performance for minority classes.

**Table 1.** Summary of Related Works

| Author(S) & Year | Method | Dataset | Finding/Result | Limitations |
|---|---|---|---|---|
| (Wiciaputra et al., 2021) | XLM-Roberta (transfer learning) | Bilingual test (Indonesian-English) | Achieved significant accuracy improvements compare to traditional method | Did not address fairness and imbalance issues |
| Rahma & Suadaa (2023) | Text augmentation for text classification | Indonesian text dataset | Augmentation helped reduce imbalance | Did not apply advanced loss functions such as Focal loss |
| Kunefi et al. (2025) | IndoBERT + Focal Loss | Indonesian hoax news dataset | Macro F1-score improved significantly with Focal Loss | Limited to hoax news, not public complaint classification |
| (Nemoto et al., 2021) | Imbalance learning for NER | Multilingual NER dataset | Focal Loss effective in handling minority classes | Not focused on Indonesian public text |
| Younes & Mathiak (2021) | Pre-trained LM for imbalance detection | Dataset mention detection | Improved sensitivity to minority classes | Did not address e-government context |
| Jurn & Kim (2025) | Data augmentation + NER metadata | Call center conversation dataset | Echanced classification performance in imbalance scenarious | More focussed on augmentation than fairness directly |

Table 1 summarizes previous related works in text classification using traditional and transformer-based models. As seen, while transformer-based approaches such as XLM-RoBERTa and IndoBERT have improved performance, most studies still struggle with imbalanced data issues and rarely emphasize fairness in minority categories. Recent works (Jurn & Kim, 2025; Vasyl et al., 2024; Zhao et al., 2024) also highlight the importance

*name of corresponding author
This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15260
e-ISSN : 2541-2019
p-ISSN : 2541-044X

of fairness-aware NLP, especially in public service or governance contexts. However, there remains a lack of studies focusing on Indonesian public complaint classification using Focal Loss as a balancing mechanism. This gap motivates our study to integrate XLM-RoBERTa with Focal Loss to ensure fairer and more inclusive complaint management in the e-government setting.

This study focuses on Focal Loss integration with XLM-RoBERTa rather than alternative imbalanced learning approaches for several domain-specific reasons. Data augmentation techniques (SMOTE, back-translation) alter the original distribution of public complaints, potentially compromising authentic representation of citizen concerns—critical in governance applications where natural complaint frequencies reflect real public priorities. Cost-sensitive learning requires manual weight tuning and may become suboptimal as complaint distributions evolve in dynamic government systems. Focal Loss offers key advantages: (1) preserves original data authenticity while dynamically focusing on difficult samples, (2) requires minimal hyperparameter tuning, (3) has proven effectiveness for Indonesian text classification (Kunaefi et al., 2025), and (4) introduces negligible computational overhead suitable for real-time processing. This methodological choice prioritizes both classification fairness and practical deployment considerations essential for sustainable government complaint systems.

## METHOD

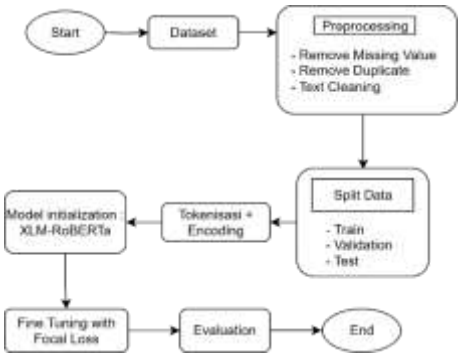


Figure 1. Research Method Flow

The research method flow can be seen in Figure 1. This research begins with the data pre-processing stage, which includes text cleaning, handling *missing values*, and removing duplicate data. After this process is completed, the data is divided into three sets, namely train data (72%), validation data (18%), and test data (10%). The next stage is the application of the RoBERTa model, which begins with the tokenization and encoding process, then continues with the fine-tuning process on the prepared data. To overcome the problem of class imbalance (data imbalance) in the dataset, fine-tuning is carried out using Focal Loss as a loss function that is more adaptive to minority classes.

### Data Sources

This study uses a dataset of Indonesian-language public complaints collected from the Lapor Gub channel between 2022 and March 2025. The available data is 54,212 with 10 features, one of which is the category feature, which contains 18 types of labels such as infrastructure, social, and others. However, the data distribution between categories is uneven, making this imbalance an important consideration in selecting and implementing an appropriate model. Examples of complaint data and their category labels can be seen in Table 1, which illustrates that text units from the Lapor Gub channel can be classified into the identified categories.

**Table 2.** Sample Dataset

| Text | Label |
|---|---|
| The station road to the east is damaged, there are many small and large holes. | Infrastructure |
| Last month, January 17, 2025, I paid taxes via the New Sakpole application, and the payment was successful, but in the application I checked the payment status and it was not paid in full. | Finance and Assets |
| Is the Central Java Prosperous Card no longer valid? | Social Community |

*name of corresponding author

This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15260
e-ISSN : 2541-2019
p-ISSN : 2541-044X

**Preprocessing**

*Pre-processing* stage involves three main steps. First, data containing blank values is removed to prevent disruption in the analysis process. Next, duplicate data is also removed to eliminate redundancy. After that, *text cleaning is performed* , which includes several important processes, such as converting entities and removing HTML tags, converting all text to lowercase , removing emojis, normalizing words, and removing excess whitespace. All of these steps aim to ensure the data is clean and ready for use in the next analysis stage.

2.3 Split Data

In the data splitting stage, the dataset is divided into three parts: training data, validation data, and test data. First, 10% of the total data is allocated as test data to evaluate the model's performance after training. The remaining 90% is used for training and validation purposes. Of this 90%, a further split is performed, taking 20% as validation data and the remaining 80% as training data. This splitting process is carried out stratified , namely maintaining the proportion of the label distribution in each subset to be representative of the entire data. The distribution of the split data can be seen in Table 2.

**Table 3.** Distribution of split data

| | |
|---|---|
| Train | 72% |
| Validation | 18% |
| Test | 10% |

**Tokenization**

Tokenization aims to break down text into smaller units called tokens. In the RoBERTa model, this process is carried out using a *subword tokenization approach* that uses *the Byte-Pair Encoding (BPE) algorithm* . Unlike some other models, RoBERTa maintains the original form of uppercase and lowercase letters (case-sensitive). In the *BPE stage* , text is parsed into subword pieces based on the frequency of occurrence of character pairs or bytes in the entire data (Pusung & Dewi, 2024) . This approach allows RoBERTa to process unknown (out-of-vocabulary) words by breaking them down into subwords already contained in the model's dictionary, so that they can still be understood and processed by the model effectively. After the tokenization process, the next stage is encoding, which is converting each token into a numeric representation in the form of a token ID according to the dictionary ( *vocabulary* ) owned by the RoBERTa model. In addition to the token ID, the encoding process also produces *an attention mask* that indicates the position of important tokens in the input. The encoding results are then used as input to the model for the training or prediction process.

**Model XLM-RoBERTa**

XLM-RoBERTa (Cross-lingual RoBERTa) is a transformer-based language model developed by Facebook AI, designed to handle multiple languages effectively. It is built upon the RoBERTa architecture, which is an improved version of BERT, trained on a much larger dataset with optimized hyperparameters. Unlike monolingual models, XLM-RoBERTa is trained on 2.5TB of filtered CommonCrawl data covering 100 different languages, making it highly effective for cross-lingual tasks such as text classification, sentiment analysis, and machine translation. Its ability to understand multilingual contexts allows it to capture semantic relationships across different languages without requiring explicit alignment between them. In the context of the attached diagram, XLM-RoBERTa is used as the base model initialization before fine-tuning. The dataset undergoes preprocessing, tokenization, and encoding steps to transform raw text into suitable input for the model. By leveraging XLM-RoBERTa, the system can generalize better across diverse and possibly noisy datasets, ensuring robust performance even in multilingual complaint text classification. Fine-tuning with Focal Loss further improves its ability to handle imbalanced data, enhancing accuracy in predicting underrepresented categories while maintaining strong overall performance (Pusung & Dewi, 2024).

**Fine Tuning**

In the model training stage, a *fine-tuning process was carried out* on the RoBERTa model by adjusting the model weights to the previously prepared public complaint data. Given that the label distribution in the data is unbalanced, the model has the potential to be biased towards the majority class (Kunaefi et al., 2025) . Therefore, *focal loss* was used as the loss function to address the *class imbalance problem* . Focal loss is a modification of *cross-entropy loss* designed to place greater emphasis on samples that are difficult to classify, while reducing the contribution of samples that are easily recognized by the model. Thus, the model will focus more on learning from data included in the minority category. The use of focal loss is expected to improve the overall model performance, especially in recognizing categories with much less data than the majority category. The *focal loss formula* is

*name of corresponding author

This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15260
e-ISSN : 2541-2019
p-ISSN : 2541-044X

written in formula (1). The *fine-tuning process* was carried out with 10 epochs, a batch size of 16, and a learning rate of 2e-5.

*FL (Pt) = - at(1 - pt)yloh(pt)*

Where:

$pt$ : predicted probability for the correct class
$\gamma$ : focus parameter
$at$ : balancing factor between classes

**Evaluation**

To assess the performance of each approach, a quantitative evaluation was conducted using *a confusion matrix* and a *classification report*. This evaluation included four main metrics in classification: *accuracy, precision* , *recall*, and, *Macro F1-score*. These four metrics are able to provide a comprehensive overview of model performance, especially in the context of data with uneven label distribution. Through this approach, the effectiveness of each loss function can be measured objectively, based on the model's ability to distinguish between the majority and minority classes. The evaluation results of each model are presented in Table.

## RESULT

In the initial stage of the research, the data obtained was 54,212 rows with 9 columns/features. However, after checking the data quality, empty data (missing values) and duplicate data were found. To ensure good and accurate data quality, data cleaning was carried out by removing 2 empty rows of data that had no values in several important columns, and removing 372 rows of duplicate data that contained the same information and could have reduced the quality of the analysis. After data cleaning, text preprocessing was carried out on the column containing public complaints. This process included removing emojis, normalization, lower case, and other cleaning processes. The cleaned text was then stored in a new column named *cleaned text.* With the addition of the *cleaned text column,* the total number of columns in the dataset increased to 10 columns, while the number of data after cleaning became 58,774 rows.

Table 4. Summary of Initial and Final Research Datasets

| Dataset Condition | Number of rows | Number of columns | Information |
|---|---|---|---|
| Before cleaning | 54,212 | 9 | There are missing values and duplicates |
| After cleaning | 53,774 | 10 | Clean data and added cleaned text column |

Figure 2 below shows the class distribution graph in the dataset. It can be seen that the data is not evenly distributed. The infrastructure category dominates, while other categories, such as tourism and culture and saber pungli, have very few complaints. This imbalance indicates **a class imbalance** that requires attention.

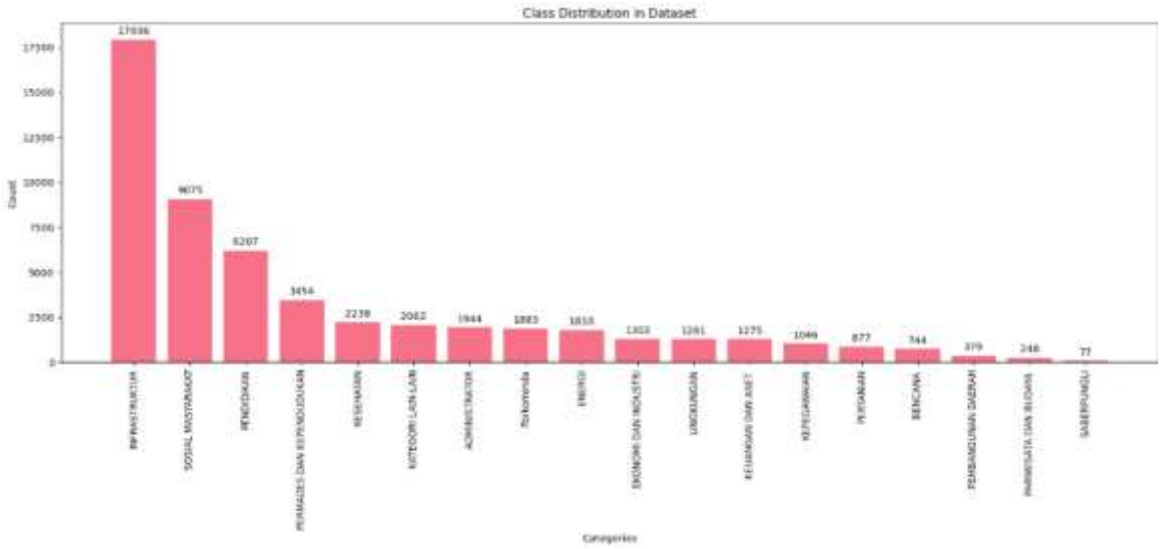


**Figure 2.** Category Distribution

*name of corresponding author

This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15260
e-ISSN : 2541-2019
p-ISSN : 2541-044X

After analyzing the data distribution, the next step is to divide the dataset using the **stratified split method** to maintain a balanced proportion of each category. This division is carried out according to predetermined proportions. The following is the resulting data distribution after the split process.

Table 5. Distribution of split data

| Description | Results |
|---|---|
| Train | 38,716 |
| Validation | 9,680 |
| Test | 5,378 |

The model was trained for 10 *epochs* using a *batch size* of 16 and *a learning rate* of 2e-5. The training process used the *Focal Loss function*, specifically designed to address the problem of imbalanced data distribution. Focal Loss works by assigning a greater penalty weight to prediction errors in the minority class, thereby helping the model learn more optimally in recognizing less representative categories. For comparison, this study also conducted model training without using Focal Loss.

The first model was trained without **Focal Loss**, and the training process was automatically stopped at **the 9th epoch** using an **early stopping mechanism**. This technique was implemented to prevent the model from overfitting, a condition where the model overfits to the training data but is unable to generalize well to the test data. Based on the evaluation results, the model achieved an accuracy of 0.781, a precision of 0.771, and a recall of 0.781. However, the macro F1-score value was only 0.606, indicating that the model tends to be better at recognizing the majority class than the minority class. This indicates that there is still an imbalance in performance between categories, and other approaches such as the use of Focal Loss are needed to improve it. The results can be seen in the following table.

- Roberta without focal loss
-

**Table 6.** Distribution of results without focal loss

| Epoch | Training Loss | Validation Loss | Accuracy | Precision | Recall | F1 | Macro Precision | Macro Recall | Macro F1 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.650.600 | 1.328.082 | 0.697658 | 0.749183 | 0.697658 | 0.711426 | 0.512432 | 0.630136 | 0.548391 |
| 2 | 1.200.300 | 1.208.959 | 0.714581 | 0.768004 | 0.714581 | 0.724314 | 0.529471 | 0.660070 | 0.561121 |
| 3 | 0.957700 | 1.155.531 | 0.747291 | 0.780134 | 0.747291 | 0.755794 | 0.572985 | 0.675390 | 0.610161 |
| 4 | 0.775000 | 1.080.297 | 0.726654 | 0.783376 | 0.726654 | 0.742821 | 0.573934 | 0.692443 | 0.616977 |
| 5 | 0.667600 | 1.169.164 | 0.749974 | 0.785723 | 0.749974 | 0.758391 | 0.582362 | 0.682549 | 0.615511 |
| 6 | 0.538700 | 1.194.422 | 0.747807 | 0.791300 | 0.747807 | 0.761652 | 0.583998 | 0.684782 | 0.621218 |
| 7 | 0.437500 | 1.275.450 | 0.764008 | 0.787678 | 0.764008 | 0.771425 | 0.595207 | 0.683993 | 0.629817 |
| 8 | 0.353800 | 1.322.027 | 0.759983 | 0.786823 | 0.759983 | 0.769656 | 0.592768 | 0.675027 | 0.625745 |
| 9 | 0.312500 | 1.427.969 | 0.767929 | 0.786966 | 0.767929 | 0.774504 | 0.606004 | 0.670432 | 0.632144 |
| 10 | 0.257100 | 1.450.161 | 0.763079 | 0.784068 | 0.763079 | 0.770729 | 0.597114 | 0.666858 | 0.626258 |

After testing the model without Focal Loss, further experiments were conducted **using Focal Loss** to address class imbalance. The training process was automatically terminated at **the 8th epoch** through early stopping, which serves to prevent overfitting. The evaluation results showed an improvement in model performance, with an accuracy of 0.785, a precision of 0.778, a recall of 0.785, and a macro F1-score of 0.625.

Compared to the previous model, the use of Focal Loss provides improvements across all evaluation metrics, particularly the macro F1-score, which rose from 0.606 to 0.625. This indicates that although the overall gain in macro F1 appears moderate (an increase of only 0.019), it reflects a fairer distribution of performance across all categories, particularly minority classes such as Saber Pungli, Tourism and Culture, and Finance and Assets.

Compared with previous studies, our findings are consistent with those reported by (Kunaefi et al., 2025), who applied IndoBERT with Focal Loss for hoax news classification and observed improvements in macro F1 ranging between 0.02–0.03. Similarly, (Rahma & Suadaa, 2023) showed that imbalance-handling techniques such as text augmentation improved minority-class recognition but often yielded only modest increases in macro-level

*name of corresponding author
This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15260
e-ISSN : 2541-2019
p-ISSN : 2541-044X

performance. These results suggest that improvements in imbalanced classification tasks are typically incremental rather than drastic, as models must trade off performance between majority and minority classes.

The moderate increase in macro F1 in this study can be explained by two factors. First, the data imbalance in the Lapor Gub dataset is extreme: categories such as Infrastructure dominate, while others like Saber Pungli and Tourism are severely underrepresented, making it difficult for the model to learn robust patterns for minority classes. Second, although Focal Loss directs more attention to minority samples, it does not introduce new information; hence its effect is limited by the scarcity of training data in those categories. This limitation highlights the need for future work exploring complementary techniques such as data augmentation, cost-sensitive learning, or hybrid approaches that may further enhance fairness without compromising majority-class performance.

- Roberta + Focal Loss
-

**Table 7**. Distribution of results with focal loss

| Epoch | Training Loss | Validation Loss | Accuracy | Precision | Recall | F1 | Macro Precision | Macro Recall | Macro F1 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.807900 | 0.595432 | 0.757301 | 0.726649 | 0.757301 | 0.725309 | 0.551324 | 0.509768 | 0.501023 |
| 2 | 0.539300 | 0.510344 | 0.778145 | 0.749491 | 0.778145 | 0.756100 | 0.605431 | 0.575205 | 0.579769 |
| 3 | 0.413800 | 0.466228 | 0.795790 | 0.777095 | 0.795790 | 0.779211 | 0.667851 | 0.607622 | 0.615030 |
| 4 | 0.344500 | 0.457812 | 0.787225 | 0.781308 | 0.787225 | 0.782120 | 0.652415 | 0.636403 | 0.631361 |
| 5 | 0.286300 | 0.470455 | 0.792797 | 0.778209 | 0.792797 | 0.781686 | 0.641262 | 0.632469 | 0.629544 |
| 6 | 0.225000 | 0.486685 | 0.789392 | 0.779598 | 0.789392 | 0.781897 | 0.631556 | 0.647344 | 0.633627 |
| 7 | 0.179900 | 0.509475 | 0.791766 | 0.780104 | 0.791766 | 0.783982 | 0.634011 | 0.639941 | 0.633561 |
| 8 | 0.141500 | 0.525847 | 0.785368 | 0.778192 | 0.785368 | 0.778831 | 0.624367 | 0.635137 | 0.625089 |

After applying **Focal Loss**, the model's performance showed significant improvements in several categories, especially those with smaller data sets. For example, for the **Forkominda label**, the previous model was only able to correctly recognize **67% of the time**, but after using Focal Loss, this increased to **73%**. The **Administrator category** also saw an increase in correct predictions, from **25%** to **29%**. This suggests that Focal Loss helps the model focus more on **minority classes** that were previously often overlooked.

On the other hand, prediction accuracy in majority categories such as **Infrastructure** and **Education remained high, at 0.90** and **0.86,** respectively, indicating that the use of Focal Loss did not compromise performance in the majority class. However, the most obvious improvement was seen in **minority categories** such as saberpungli (illegal levies), from 62% to 75%, tourism and culture (62% to 68%), and finance and assets (65% to 72%). This average increase was also in line with **the increase in the macro F1-score value** from **0.606** to **0.625**, indicating that the model became fairer and more balanced in predicting all classes.

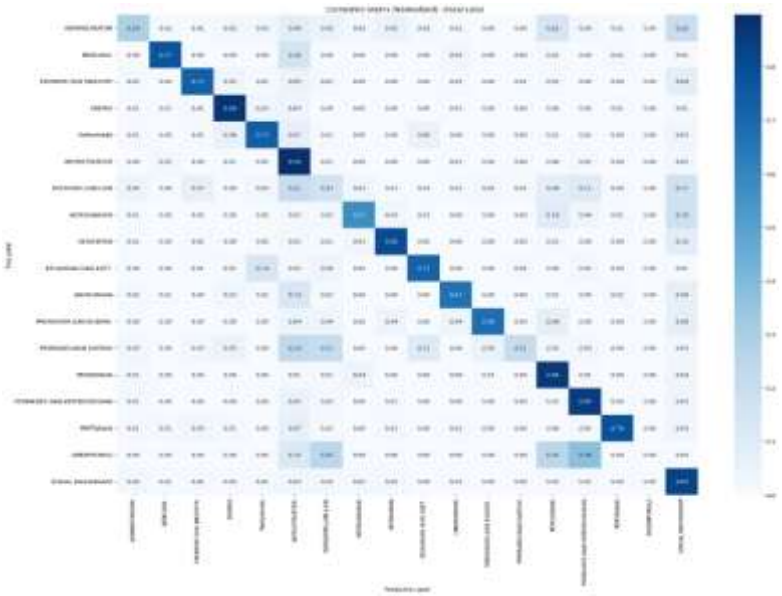

**Figure 3.** Correlation Matrix

*name of corresponding author

This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15260
e-ISSN : 2541-2019
p-ISSN : 2541-044X

The image shows a normalized correlation matrix for a multi-class classification model trained with Focal Loss, representing how well the model predicts different categories of public complaints. Each row corresponds to the true label, while each column indicates the predicted label. The diagonal values represent correct predictions, with darker blue shades indicating higher accuracy for that category, such as *INFRASTRUKTUR* (0.90), *KESEHATAN* (0.80), and *PENDIDIKAN* (0.86). Off-diagonal values represent misclassifications, where the model confused one category with another, for example, *ADMINISTRATOR* being misclassified as *SOSIAL MASYARAKAT*. Overall, the matrix highlights the strengths and weaknesses of the model across different categories, showing that while some classes achieve relatively high accuracy, others still suffer from misclassification.

## DISCUSSION

Based on the test results, it appears that the RoBERTa model applied to classify public complaint texts from *the Lapor Gub channel* performed quite well. The model trained without Focal Loss achieved an accuracy of 78.1% and a Macro F1-score of 0.606. While the overall accuracy is high, the lower Macro F1-score indicates performance disparities between categories, particularly in minority categories, which the model tends to overlook. *Class imbalance* is a major factor affecting model performance in recognizing categories with fewer data sets. This aligns with the findings of (Rahma & Suadaa, 2023), who stated that classification models tend to be biased toward the majority class if no specific data imbalance is addressed. This imbalance is reflected in the data distribution, which is heavily dominated by the infrastructure category, while several categories, such as *Saber Pungli* and *Tourism and Culture,* have only limited data representation (Arham et al., 2025).

The application of Focal Loss in the RoBERTa fine-tuning process showed significant improvements in several evaluation metrics, particularly the Macro F1-score, which increased to 0.625. This improvement indicates that the model is more balanced in predicting both the majority and minority classes. Furthermore, the model's accuracy also slightly increased to 78.5%, indicating that the use of Focal Loss not only benefits the minority class but also maintains performance in the majority class. The most notable performance improvements were seen in several minority categories, such as *Saber Pungli*, *Forkominda*, and *Tourism and Culture*. The increase in correct prediction rates in these categories indicates that Focal Loss is effective in helping the model focus more on cases that were previously difficult to learn. This is in line with research by Kunaefi et al. (2025) which showed that Focal Loss can increase the model's sensitivity to minority classes in cases of imbalanced data (Abdel-salam, 2022).

Although the improvement in Macro F1-score may appear modest (0.606 → 0.625), this pattern aligns with findings from previous research. Kunaefi et al. (2025) reported similar incremental gains when applying Focal Loss to hoax news classification with IndoBERT, while Rahma & Suadaa (2023) highlighted that imbalance-handling strategies rarely produce dramatic increases but remain critical for ensuring fairer recognition of minority classes. This suggests that even small improvements are meaningful in highly imbalanced datasets, as they demonstrate that minority categories are better represented without sacrificing majority-class performance.

However, it's worth noting that the performance improvements in minority classes didn't completely eliminate the existing imbalance. Some categories, such as *Administrator,* still showed low accuracy despite the improvements. This suggests that extreme imbalance remains a challenge that deserves more attention in future research. Additional strategies, such as data augmentation or cost-sensitive learning, could be considered to further improve performance. From a technical perspective, using RoBERTa offers advantages because this model is able to capture the context and meaning of public complaint texts without requiring complex preprocessing. This approach is highly relevant considering that complaint texts often contain language variations, non-standard spellings, and the use of local terms. As stated by Wiciaputra et al. (2021), Transformer-based models like RoBERTa are able to handle the complexity of natural language better than traditional methods (Song et al., 2021).

Overall, the results of this study indicate that the combination of RoBERTa with Focal Loss is an effective approach for text-based public complaint classification in addressing data imbalance issues. The application of this technology can help the government manage and respond to public complaints more efficiently, accurately, and fairly, thus supporting a more responsive and transparent government. Another important implication of this study is the role of balanced classification in ensuring fairness in governance. When minority complaint categories such as “Illegal Levies” or “Tourism” are better recognized, it signals that all citizens’ voices are considered equally, regardless of frequency. This inclusivity strengthens public trust in digital platforms like *Lapor Gub* and encourages broader civic participation in reporting issues (Azadi et al., 2024).

Furthermore, the slight increase in accuracy observed when using Focal Loss demonstrates that addressing class imbalance does not necessarily sacrifice performance in dominant categories. On the contrary, the approach ensures that both frequent and infrequent complaint types are well represented. This balance is crucial in government applications, where overlooking minority issues could result in social dissatisfaction or policy gaps. The evaluation metrics also highlight the importance of moving beyond accuracy as the sole indicator of model performance. While accuracy remained high in both experiments, the improvement in macro F1-score provides

*name of corresponding author
This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

stronger evidence that Focal Loss contributed to fairer predictions. This finding supports the argument made by Sani et al. (2022), who emphasized the use of precision, recall, and F1-score in imbalanced classification tasks.

Despite these improvements, the persistence of lower performance in certain categories, such as "Administrator," underlines the limitations of Focal Loss alone. Extreme data imbalance, where some categories contain only a handful of examples, may require additional interventions. Data augmentation techniques, such as back translation or synonym replacement, could help increase representation in minority categories without compromising authenticity. Another potential solution lies in cost-sensitive learning, where misclassifications in minority classes are penalized more heavily. This method, when combined with Focal Loss, may offer even greater sensitivity to underrepresented categories. Future research could explore hybrid strategies that integrate multiple balancing approaches to further enhance classification performance (Zhao et al., 2024).

From a governance perspective, the findings of this study have practical implications. An improved classification system can accelerate the process of directing complaints to the relevant government departments, reducing delays in response and resolution. Faster and more accurate routing of complaints helps ensure that issues are addressed promptly, ultimately improving citizen satisfaction with public services. The adaptability of RoBERTa also ensures that the model remains effective in handling the diversity of language used in public complaints. Informal expressions, abbreviations, and local dialects—often seen as challenges in text classification were better understood due to the model's contextual learning capabilities. This reinforces the suitability of transformer-based models for real-world government applications (Vasyl et al., 2024).

From the perspective of local governments, this study provides three important practical implications. First, improved classification enables faster responses by ensuring that complaints are routed more efficiently to the appropriate agencies. Second, better recognition of minority categories supports more equitable service delivery, so that less frequent but still critical issues are not overlooked. Third, by treating all categories consistently, the system reinforces accountability and transparency in public service management, thereby strengthening public trust in local governance.

Additionally, the success of this study demonstrates the broader potential of AI in the public sector. Beyond complaint management, similar approaches could be applied to analyze citizen feedback in other contexts, such as surveys, social media engagement, or policy consultations. Such applications would contribute to data-driven governance and more informed decision-making. It is also important to highlight the scalability of this approach. As digital platforms continue to grow and the volume of citizen complaints increases, traditional manual categorization will become even less feasible. Automated systems based on RoBERTa and Focal Loss can scale effectively to handle such growth, ensuring that governments remain responsive in the face of rising citizen engagement. In conclusion, the discussions reinforce that while challenges remain, particularly with extreme data imbalance, the integration of RoBERTa and Focal Loss represents a significant step toward fairer and more efficient complaint classification. The improvements observed in minority categories underscore the value of adopting advanced NLP methods in governance. This study thus contributes not only to academic research in text classification but also to practical solutions that support more transparent, inclusive, and accountable public service delivery (Younes & Mathiak, 2021).

It worth mentioning, this study presents point estimates for model performance improvements but does not include statistical significance testing (e.g., McNemar tests, paired t-tests) due to computational constraints and timeline limitations. Training multiple XLM-RoBERTa model variants on our large-scale dataset (53,774 complaints) requires substantial computational resources and time, making comprehensive statistical validation beyond the scope of this initial investigation. Future research should implement rigorous statistical testing frameworks to validate the observed improvements and establish confidence intervals for performance metrics.

## CONCLUSION

This study concludes that integrating XLM-RoBERTa with Focal Loss provides an effective solution for addressing the challenges of class imbalance in public complaint classification. The experimental results show that while the baseline RoBERTa model achieved strong overall accuracy, its performance in minority categories was limited, as reflected by the relatively low macro F1-score. The incorporation of Focal Loss improved the model's sensitivity to underrepresented categories, resulting in a more balanced classification without compromising accuracy in majority classes. These findings highlight the potential of advanced transformer-based models combined with adaptive loss functions to enhance the fairness, efficiency, and transparency of digital governance systems such as Lapor Gub, ultimately supporting more responsive public service delivery.

However, this study is not without limitations. The dataset remains heavily dominated by certain categories, particularly Infrastructure, while others such as Tourism or Illegal Levies are underrepresented. This extreme imbalance restricts the model's ability to learn robust patterns for minority classes, even when using Focal Loss.

*name of corresponding author

This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Future research should therefore explore complementary strategies, including data augmentation techniques (e.g., back translation, synonym replacement) to enrich minority categories, cross-dataset testing to validate generalizability beyond Lapor Gub, and integration of multi-platform complaint data (e.g., social media, national complaint portals) to build more comprehensive and inclusive classification models. Such directions would further strengthen the reliability and applicability of automated complaint classification in supporting fair and accountable governance.

## ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the Dinas Komunikasi dan Informasi Provinsi Jawa Tengah (Communication and Information Agency of Central Java Province) for providing access to the *Lapor Gub* complaint dataset, which made this research possible. Appreciation is also extended to colleagues, mentors, and academic peers who provided valuable insights and constructive feedback throughout the development of this study. Finally, the authors acknowledge the continuous support from their respective institutions, whose encouragement and resources contributed significantly to the successful completion of this work.

## REFERENCES

Afida, D., Udayanti, ED, & Kartikadarma, E. (2021). Text Mining Application for Clustering Public Complaints in Semarang City Using the K-means Algorithm. *Transformatika Journal* , *18* (2), 215–224. https://doi.org/10.26623/transformatika.v18i2.2362

Kristina, EC, Setyawati, E., & Wati, L. (2023). Android-Based Public Complaint Information System for Banyumas. *Electro Luceat* , *9* (2), 1–13. https://doi.org/10.32531/jelekn.v9i2.684

Kunaefi, A., Abidin, Z., & Kusumawati, R. (2025). CLASSIFICATION OF HOAX NEWS IN INDONESIAN LANGUAGE USING INDOBERT FINE-TUNING WITH A FOCAL LOSS APPROACH ON UNBALANCED DATA. *JIPI (Scientific Journal of Informatics Research and Learning)* , *10* (2), Article 2. https://doi.org/10.29100/jipi.v10i2.7811

Mufidah, FS, Winarno, S., Alzami, F., Udayanti, ED, & Sani, RR (2022). Analysis of Public Sentiment Towards Shopeefood Services Through Twitter Social Media Using the Naïve Bayes Classifier Algorithm. *JOINS (Journal of Information System)* , *7* (1), 14–25.

Pusung, EM, & Dewi, IN (2024). RoBERTa Optimization with Hyperparameter Tuning for Text-based Emotion Detection. *National Journal of Technology and Information Systems* , *10* (3), Article 3. https://doi.org/10.25077/TEKNOSI.v10i3.2024.240-248

Rahma, IA, & Suadaa, LH (2023). Application of Text Augmentation to Overcome Imbalanced Data in Indonesian Text Classification. *Journal of Information Technology and Computer Science* , *10* (6), 1329–1340. https://doi.org/10.25126/jtiik.2023107325

Sani, RR, Pratiwi, YA, Winarno, S., Udayanti, ED, & Alzami, F. (2022). Comparative Analysis of Naive Bayes Classifier and Support Vector Machine Algorithms for Hoax News Classification in Indonesian Online News. *Journal of Informatics Society* , *13* (2), 85–98.

Wiciaputra, Y., Young, J., & Rusli, A. (2021). Bilingual Text Classification in English and Indonesian via Transfer Learning using XLM-RoBERTa. *International Journal of Advances in Soft Computing and Its Applications* , *13* (3), 73–87. https://doi.org/10.15849/IJASCA.211128.06

Abdel-salam, R. (2022). reamtchka at SemEval-2022 Task 6: Investigating the effect of different loss functions for Sarcasm detection for unbalanced datasets. 896–906.

Arham, M., Mohan, R., & Kadiyala, R. (2025). 1-800-SHARED-TASKS @ NLU of Devanagari Script Languages: Detection of Language, Hate Speech, and Targets using LLMs. 1(1), 1–13.

Azadi, A., Ansari, B., Zamani, S., & Eetemadi, S. (2024). Bilingual Sexism Classification: Fine-Tuned XLM-RoBERTa. 1(1), 1–8.

Jurn, S., & Kim, W. (2025). Improving Text Classification of Imbalanced Call Center Conversations Through Data Cleansing, Augmentation, and NER Metadata. Electronics, 14(11), 2259. https://doi.org/10.3390/electronics14112259

Kunaefi, A., Abidin, Z., & Kusumawati, R. (2025). KLASIFIKASI BERITA HOAKS BAHASA INDONESIA MENGGUNAKAN INDOBERT FINE-TUNING DENGAN PENDEKA-TAN FOCAL LOSS PADA DATA TIDAK SEIMBANG. JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika), 10(2), Article 2. https://doi.org/10.29100/jipi.v10i2.7811

Mufidah, F. S., Winarno, S., Alzami, F., Udayanti, E. D., & Sani, R. R. (2022). Analisis Sentimen Masyarakat Terhadap Layanan Shopeefood Melalui Media Sosial Twitter Dengan Algoritma Naïve Bayes Classifier. JOINS (Journal of Information System), 7(1), 14–25.

Nemoto, S., Kitada, S., & Iyatomi, H. (2021). Majority or Minority: Data Imbalance Learning Method for Named Entity Recognition. 1(1).

*name of corresponding author

This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Rahma, I. A., & Suadaa, L. H. (2023). Penerapan Text Augmentation untuk Mengatasi Data yang Tidak Seimbang pada Klasifikasi Teks Berbahasa Indonesia. Jurnal Teknologi Informasi dan Ilmu Komputer, 10(6), 1329–1340. https://doi.org/10.25126/jtiik.2023107325

Sani, R. R., Pratiwi, Y. A., Winarno, S., Udayanti, E. D., & Alzami, F. (2022). Analisis Perbandingan Algoritma Naive Bayes Classifier dan Support Vector Machine untuk Klasifikasi Berita Hoax pada Berita Online Indonesia. Jurnal Masyarakat Informatika, 13(2), 85–98.

Song, G., Huang, D., & Xiao, Z. (2021). A Study of Multilingual Toxic Text Detection Approaches under Imbalanced Sample Distribution. 1–16.

Vasyl, D., Vitalii, B., Renat, A., & Mykola, B. (2024). EVALUATING CUSTOMER EXPERIENCE IN E-COMMERCE : MULTILINGUAL SENTIMENT ANALYSIS OF USER REVIEWS USING. 0.

Younes, Y., & Mathiak, B. (2021). Handling Class Imbalance when Detecting Dataset Mentions with Pre-trained Language Models.

Zhao, H., Chen, H., Ruggles, T. A., Feng, Y., Singh, D., & Yoon, H.-J. (2024). Improving Text Classification with Large Language Model-Based Data Augmentation. 11(2535), 1–14.

*name of corresponding author

This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.