

Automatic Dewey Decimal Classification of Indonesian Book Metadata Using IndoBERT with Weighted Loss and Context Enhancement

Joko Purwanto^{1*}, Fajar Mahardika², Adlan Nugroho³

^{1,3}Teknologi Rekayasa Perangkat Lunak, Politeknik Negeri Cilacap, Indonesia

²Teknik Informatika, Politeknik Negeri Cilacap, Indonesia

^{1*}jokopurwanto@pnc.ac.id, ²fajarmahardika@pnc.ac.id, ³adlannugroho@pnc.ac.id

Abstract: This study proposes an automatic Dewey Decimal Classification (DDC) classification framework for Indonesian book metadata by integrating the IndoBERT model strengthened through weighted loss and context enhancement mechanisms. The current escalation of digital book collections poses significant challenges in classification efficiency and information retrieval, while the manual DDC classification process still relies on librarian expertise and is relatively time-consuming. The dataset used includes 2,516 book metadata obtained through the Google Books API and mapped into 14 DDC categories. The context enhancement strategy is implemented by integrating book titles and descriptions into a single text representation, while weighted cross-entropy loss, random oversampling, and simple data augmentation techniques are applied to address class imbalance issues. Model performance is evaluated based on accuracy, precision, recall, and F1-score metrics. Experimental results show that the proposed approach achieves an accuracy of 90.14% and a weighted F1-score of 90.15%, outperforming the baseline IndoBERT model, which only achieved an accuracy of 47.82% and a weighted F1-score of 47.06%. These findings indicate that the combination of weighted loss and contextual text representation can improve the semantic understanding of book metadata while reducing bias towards the majority class in Transformer-based DDC classification.

Keywords: context enhancement; Dewey Decimal Classification; IndoBERT; Automatic Classification; Weighted Loss

1. INTRODUCTION

The development of digital libraries and online book repositories has significantly increased the number of digital book collections in recent years. Google Books is one of the largest digital book repositories, providing access to millions of digital book collections and is widely used in digital library services and information research [1]. This growth in digital collections poses challenges in the process of organizing, classifying, and efficiently retrieving information in modern library systems [2], [3]. Dewey Decimal Classification (DDC) is one of the most widely used library classification systems for systematically grouping knowledge sources through a subject-based hierarchical structure [4]. This system plays a crucial role in supporting collection management and information retrieval

in digital library environments [2], [4]. However, the manual DDC classification process still requires time, high consistency, and in-depth librarian expertise. This condition is considered less effective in handling the ever-growing scale of digital data, thus driving the need for an automatic classification system that can group book metadata quickly and accurately.

In the field of Natural Language Processing (NLP), various text classification approaches have been developed to support document automation. Conventional machine learning methods such as Support Vector Machine (SVM), Naïve Bayes, and TF-IDF weighting have been widely implemented in various previous studies. However, these methods still have limitations in capturing semantic relations between words, especially in Indonesian language book metadata which generally has short text contexts and diverse linguistic structures [4], [5]. As a result, classification performance tends to decrease when faced with data with limited semantic representation and imbalanced class distribution.

The evolution of the Transformer model through Bidirectional Encoder Representations from Transformers (BERT) has contributed significantly to improving performance in text classification tasks through its capability to generate contextual representations that surpass traditional methods [6]. In the domain of natural language processing (NLP) for Indonesian, IndoBERT has become one of the most reliable pre-trained models due to its ability to construct a deep understanding of local linguistic characteristics [7]. Numerous empirical studies have validated IndoBERT's superiority in various applications, ranging from multi-class classification and sentiment analysis to document clustering [8], [9]. However, the implementation of IndoBERT on class-imbalanced datasets still faces crucial challenges, where the model shows a tendency to bias towards the majority class, which results in a degradation of classification performance for the minority class [10], [11].

To address the problem of data imbalance, previous research has applied techniques such as weighted loss, focal loss, and data augmentation [12], [13], [14], [15], [16], [17]. The weighted loss approach works by giving greater weight to the minority class so that the model can learn a fairer data distribution. On the other hand, context enhancement is also starting to be widely used to enrich semantic representations by adding contextual information to text data [18], [19]. This approach has been proven to improve the quality of short text representations and help models understand semantic relationships more deeply, especially in transformer-based multi-class classification [20].

However, most previous studies have focused on general text classification and have not investigated the combined use of weighted loss and context enhancement in DDC-based Indonesian book metadata classification. This combination is important because DDC metadata classification commonly involves short and semantically sparse text alongside imbalanced category distributions. Context enhancement is expected to improve semantic representation by integrating complementary textual information, while weighted loss can reduce model bias toward majority classes during training. Nevertheless, previous studies mainly emphasized model architecture optimization or fine-tuning strategies without addressing both challenges simultaneously. Consequently, classification performance on minority categories and model generalization capability remain limited [11], [20].

Based on these issues, this study proposes a context-enriched IndoBERT framework integrated with weighted loss for automatic classification of book metadata based on DDC. The proposed approach combines book titles and descriptions as richer textual representations, then applies weighted cross-entropy loss to reduce bias towards the majority class during the training process. This study aims to improve model robustness and achieve more balanced classification performance across DDC categories, especially for minority classes. Furthermore, this research is expected to contribute to the development of a Transformer-based automatic classification system to support more effective and efficient digital library management.

2. RESEARCH METHODOLOGY

This study applies an experimental design to develop a framework for automatic classification of book metadata based on Dewey Decimal Classification (DDC). The proposed model integrates the IndoBERT architecture with weighted loss and context enhancement strategies. The use of transformer-based text classification was chosen due to its superior performance track record in handling Natural Language Processing (NLP) tasks, especially in multi-class classification scenarios with imbalanced datasets [21], [22]. This framework is specifically designed to mitigate the limitations of semantic context and the imbalance of class distribution in Indonesian book metadata. This research architecture includes systematic stages starting from data collection to performance evaluation, as illustrated in Figure 1

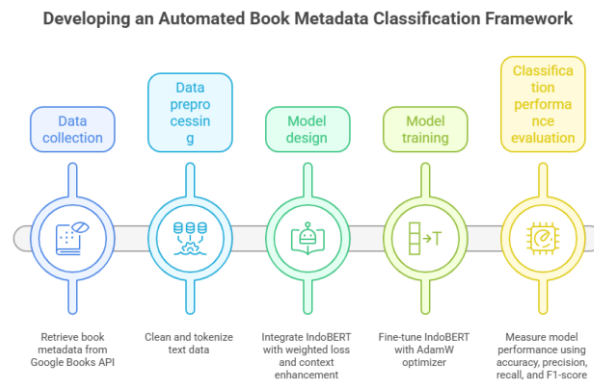


Figure 1. Research Workflow for DDC Classification Using IndoBERT

The dataset used consisted of 2,516 Indonesian-language book metadata entries extracted through the Google Books API. Metadata components included book titles, descriptions, and categories, which were then mapped to the DDC classification scheme. After data acquisition through the API request mechanism, data cleansing was performed to eliminate duplication and entries with incomplete category or description information.

A crucial preprocessing step is performed to ensure data consistency before the training phase. This procedure includes case folding, text normalization, non-alphabetic character removal, punctuation cleaning, and text noise reduction [11], [23]. Next, the data is processed using IndoBERT's built-in tokenizer to ensure compatibility with the transformer model's input. To maintain computational efficiency without sacrificing semantic representation, the maximum input length is set at 256 tokens.

In a context enhancement effort, book titles and descriptions are contextualized into a single unified text representation. This information integration aims to enrich semantic representations and contextual embedding, which has been proven effective in improving accuracy in transformer-based text classification [24], [20]. This strategy allows the model to capture relationships between topics more comprehensively, especially in metadata with very short descriptions.

In addition to context enhancement, this study also applies simple data augmentation to several minority categories to expand the variety of text representations. The augmentation process involves randomly swapping the positions of words within the text metadata so that the model can learn different variations of sentence structure without changing the primary meaning of the text [19]. This approach is applied to a number of DDC categories with a relatively small amount of data to help improve model generalization during the training process.

The processed dataset is then separated into training data and test data using a stratified division technique with a ratio of 80:20 to maintain the proportion of class distribution in each data subset. Class distribution analysis was performed to identify data

imbalance conditions in the DDC categories used in this research. The distribution of the dataset by DDC category is shown in Table 1.

Table 1. Distribution of The Dataset by DDC Category

DDC Code Category	Category	Number of Data Points
005	Computer Programming, Programs & Data	231
150	Psychology	232
300	Social Sciences	243
370	Education	239
371	School Management & Special Education	212
372	Elementary Education	148
378	Higher Education	229
410	Linguistics	127
510	Mathematics	222
610	Medicine & Health	91
650	Management & Public Relations	162
700	Arts & Recreation	135
808	Writing & Rhetoric	146
820	English Literature	99
Total		2,516

Referring to Table 1, the data distribution in each DDC category shows class imbalance, where some categories have fewer data points than others. This situation has the potential to bias the model toward the majority class, necessitating a strategy to equalize the data distribution. To address this class imbalance, the study employed random oversampling of the minority category through a resampling technique, thus making the data distribution in each category more proportional [24].

The core model used in this research is IndoBERT, a transformer-based language model pre-trained for Indonesian Natural Language Processing (NLP) purposes [7]. IndoBERT was selected based on its ability to produce contextual embeddings capable of recognizing the linguistic characteristics of Indonesian and its proven superior performance on various text classification tasks [8], [25], [26]. In the fine-tuning phase, a classification layer was embedded at the end of the model to generate classification probabilities for each DDC category. Book metadata that has gone through the pre-processing and tokenization stages is then processed through the IndoBERT bidirectional transformer layer to form contextual embeddings used in the classification procedure. The model architecture schematic applied in this study is presented in Figure 2.

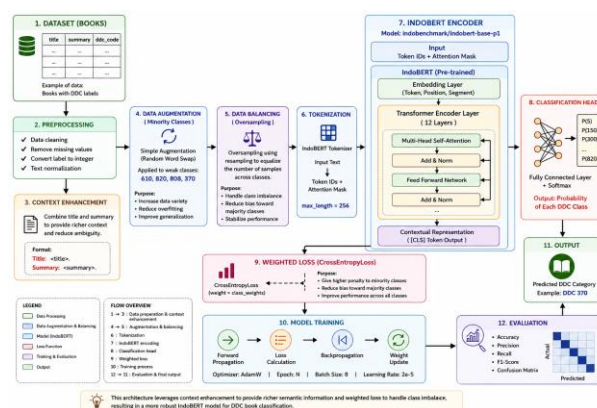


Figure 2. IndoBERT Architecture with Weighted Loss and Context Enhancement

To overcome the problem of disproportionate class distribution, this research uses weighted cross-entropy loss in the model learning stage. This loss weighting strategy is implemented by giving a larger value weight to the minority class so that the model is able to deepen the feature representation more balanced across all DDC categories [10], [12]. The calculation of class weights is executed by utilizing the following formulation:

$$w_i = \frac{N}{C \times n_i}$$

where w_i represents the weight of class i , N is the total number of data points, C is the total number of classes, and n_i is the number of data points in class i .

The model training process was carried out using the PyTorch framework and the HuggingFace Transformers library. The IndoBERT model was optimized through a fine-tuning stage using AdamW optimization [27] with a learning rate of rate of 2×10^{-5} , a batch size of 8, a token length limit of 256 tokens, and a training duration of 5 epochs with GPU acceleration support to improve computational efficiency. The 256-token length limit was chosen because most book metadata after the title and description merging process falls below that length, thus preserving important contextual information while maintaining efficient memory usage during the training process. The transformer fine-tuning approach that utilizes contextual embeddings and transfer learning has become a key technique in various modern text classification studies [21], [22], [28]. The hyperparameter configurations applied during the model training phase are shown in Table 2.

Table 2. Hyperparameter Configuration the Model Training

Parameter	Value
Model	IndoBERT
Optimizer	AdamW
Learning Rate	2×10^{-5}
Batch Size	8
Maximum Token Length	256
Epoch	5
Data Split Ratio	80:20

Before the training process, the book metadata data goes through a preprocessing stage which includes combining the book title and description into a single text representation, cleaning special characters, normalizing letters to lowercase, and tokenizing using the IndoBERT tokenizer. An example of preprocessing is shown in Table 3.

Table 3. Example of Book Metadata Preprocessing Stages

Stage	Results
Original Title	Basic Python Programming
Original Description	This book covers the basic concepts of Python programming for beginners.
Text Merge	Basic Python Programming This book covers the basic concepts of Python programming for beginners.
Lower case	basic python programming this book discusses the basic concepts of python programming for beginners.
Character Cleanup	basic python programming this book discusses the basic concepts of python programming for beginners

Model performance evaluation was performed using accuracy, precision, recall, and F1-score parameters to calculate classification effectiveness in each DDC category. The

weighted-average and macro-average F1-score metrics were used because the dataset has an uneven class distribution, which makes these two measures more reflective of the model performance evaluation results in the majority and minority class groups [10], [29]. In addition to global testing, a review of the performance of each class and a confusion matrix were also conducted to map the classification error patterns in the DDC categories. The proposed model was then compared with the baseline IndoBERT model without involving weighted loss and context enrichment to measure the impact of the implemented optimization scheme.

3. RESULT AND DISCUSSIONS

A series of tests were conducted to assess the effectiveness of the IndoBERT-based Dewey Decimal Classification (DDC) automatic classification model in two schemes: the baseline IndoBERT model without optimization techniques, and a model combined with weighted loss and context enrichment. The evaluation used accuracy, precision, recall, and F1-score parameters to measure classification results for each DDC category. Additionally, a confusion matrix was used to analyze patterns of misclassification among these categories.

In the initial phase, experiments were conducted using the baseline IndoBERT model without any additional optimization strategies. This model was trained using text metadata obtained through standard preprocessing procedures and tokenization using the IndoBERT tokenizer. The evaluation results of the baseline model are presented in Table 3.

Table 3. IndoBERT Baseline Evaluation Results

Metric	Value
Accuracy	0.4782
Precision	0.4870
Recall	0.4782
F1-Score	0.4706
Macro Average F1-Score	0.4698

Referring to the test results, the baseline model achieved an accuracy rate of 47.82% with a weighted precision of 48.70%, a weighted recall of 47.82%, and a weighted F1-score of 47.06%. The macro-average F1-score of 46.98% indicates that the model's effectiveness remains unstable across all DDC categories.

This finding suggests that the baseline model still struggles with consistent classification, especially for data sets with small quantities. The imbalance in the number of classes results in the model tending to focus more on patterns in the majority group than in the minority. Several categories showed relatively low results, for example, the DDC 372 category with an F1-score of 34.15%, the DDC 510 category at 37.33%, and the DDC 808 category at 29.63%. On the other hand, the DDC 410 category recorded superior performance with an F1-score of 65.31%.

The confusion matrix analysis of the baseline model presented in Figure 3 also reveals frequent misclassification between semantically related categories. Misclassification commonly occurred between DDC 300 (Social Sciences) and DDC 370 (Education), where many book descriptions contain similar educational and social terminology. For example, metadata such as "Pendidikan karakter dalam masyarakat modern" was incorrectly predicted as DDC 300 instead of DDC 370. Similar errors were also observed between DDC 510 (Mathematics) and DDC 004 (Computer Science-related materials) due to overlapping computational terminology. Furthermore, the prediction distribution in the confusion matrix is not yet concentrated along the main diagonal, indicating that the baseline model still has difficulty separating closely related DDC categories.

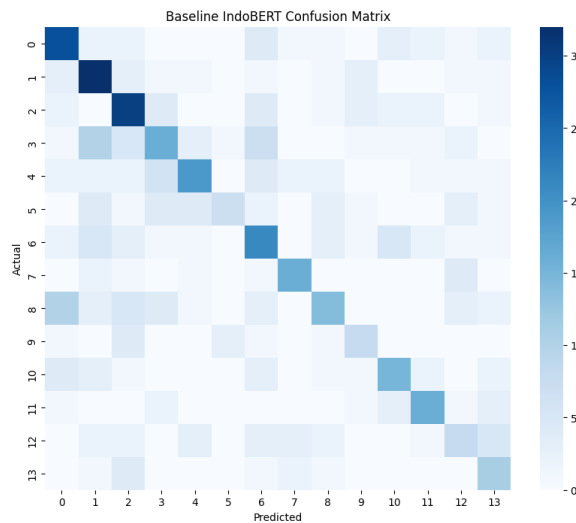


Figure 3. Baseline IndoBERT Confusion Matrix

To further analyze the contribution of each proposed component, an ablation study was conducted by separately applying weighted loss and context enhancement before combining both approaches into the proposed model. The evaluation results are shown in Table 4.

Table 4. Ablation Study Results

Metric	Accuracy	Precision	Recall	F1-Score
Baseline IndoBERT	0.4782	0.4870	0.4782	0.4706
IndoBERT + Weighted Loss	0.5600	0.5700	0.5600	0.5600
IndoBERT + Context Enhancement	0.5853	0.6037	0.5853	0.5874
Proposed Method	0.9014	0.9027	0.9014	0.9015

In the next experiment, weighted loss and context enrichment were combined into the IndoBERT model to address class inequality and improve the quality of semantic representation in book metadata. The context enhancement strategy was implemented by combining book title and synopsis data into a single text representation, while the weighted loss was implemented using weighted cross-entropy loss during the model learning phase. The evaluation findings of the proposed model are shown in **Table 5**.

Table 5. Evaluation Results of the Proposed Method

Metric	Value
Accuracy	0.9014
Precision	0.9027
Recall	0.9014
F1-Score	0.9015
Macro Average F1-Score	0.9015

Referring to the test results, the proposed model achieved an accuracy rate of 90.14% with a weighted precision of 90.27%, a weighted recall of 90.14%, and a weighted F1-score of 90.15%. The macro-average F1-score of 90.15% indicates that the model's effectiveness is significantly more stable across all DDC categories compared to the baseline model. This performance improvement demonstrates that combining weighted

loss and context enrichment improves the model's ability to capture contextual relationships in book metadata while minimizing bias toward the majority class.

Performance improvements are also evident in the assessment of each class. The DDC 610 category achieved an F1-score of 92.55%, the DDC 808 category 92.55%, and the DDC 820 category 98.46%. Furthermore, the DDC 410 category achieved the highest F1-score with an F1-score of 95.34%. This finding indicates that the weighted loss scheme can boost classification results in categories with initially small amounts of data, while context enrichment helps the model more thoroughly explore the semantic background of book metadata.

The confusion matrix for the proposed method, presented in Figure 4, shows significant improvements in classification results compared to the baseline model. The distribution of predictions is more concentrated in the main diagonal area of the confusion matrix, indicating that most of the data has been successfully mapped to the correct DDC categories. Furthermore, the frequency of misclassifications between categories was also significantly reduced.

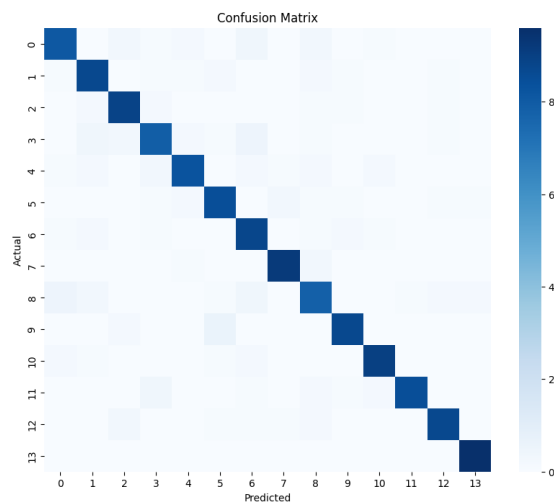


Figure 5. Confusion Matrix of the Proposed Method

To provide a more comprehensive evaluation, the proposed method is compared with several other classification approaches commonly used in text classification research, namely TF-IDF + SVM, baseline IndoBERT, IndoBERT + weighted loss, and IndoBERT + context enhancement. The comparison results are shown in Table 5.

Table 6. Model Performance Comparison

Model	Accuracy	Precision	Recall	F1-Score
Baseline IndoBERT	0.4782	0.4870	0.4782	0.4706
TF-IDF + SVM	0.5400	0.5400	0.5400	0.5400
IndoBERT + Weighted Loss	0.5600	0.5700	0.5600	0.5600
IndoBERT + Context Enhancement	0.5853	0.6037	0.5853	0.5874
Proposed Method	0.9014	0.9027	0.9014	0.9015

Experimental results show that the TF-IDF + SVM method achieved an accuracy and weighted F1-score of 54.00%. This performance is slightly better than the IndoBERT baseline, which only achieved a weighted F1-score of 47.06%. This indicates that the TF-IDF-based feature representation is still quite effective in capturing word patterns in short

book metadata, while standard IndoBERT fine-tuning without class imbalance handling and context enrichment is not yet able to produce optimal semantic representation.

Applying weighted loss to IndoBERT increased the weighted F1-score to 56.00%, while context enhancement resulted in a weighted F1-score of 58.74%. These results indicate that each approach positively contributes to improving classification performance. Weighted loss helps reduce model bias toward the majority category, while context enhancement enriches semantic representation by integrating book titles and descriptions.

The proposed method, which simultaneously combines weighted loss and context enhancement, achieved the best performance with an accuracy of 90.14% and a weighted F1-score of 90.15%. This significant performance improvement demonstrates that the combination of class imbalance management and semantic context enrichment can improve the model's generalization capability in DDC-based book metadata classification.

Overall, the experimental results demonstrate that the integration of weighted loss and context enhancement within the IndoBERT framework is an effective solution for automatic DDC classification of Indonesian-language digital book metadata. The proposed method not only improves overall classification performance but also enhances prediction stability for minority categories previously difficult to classify by baseline models.

4. CONCLUSION

This study proposes an automatic Dewey Decimal Classification (DDC) framework for Indonesian-language book metadata using IndoBERT integrated with weighted loss and context enhancement. The proposed approach is designed to address two key issues in DDC classification: limited semantic representation and imbalanced class distribution. Combining book titles and descriptions as richer context representations, along with the application of weighted cross-entropy loss, enables the model to learn semantic patterns more evenly across each DDC category.

Experimental results show that the proposed framework delivers superior performance compared to the baseline model and other benchmark methods. The integration of weighted loss and context enhancement significantly improves the model's ability to understand contextual relationships in book metadata, reduces bias toward the majority class, and produces more stable classification performance in minority categories. Confusion matrix analysis also shows a decrease in classification errors between semantically similar DDC categories.

This research contributes to the development of a transformer-based automatic classification system to support more effective and efficient digital library management. The research findings indicate that a combination of semantic enrichment strategies and data imbalance handling can improve the model's robustness in the multi-class classification task of Indonesian language book metadata.

However, this study still has several limitations. The dataset used only covers 14 DDC categories and is still limited to metadata in the form of book titles and descriptions. Future research can be conducted by expanding the scope of DDC categories, adding other bibliographic metadata, applying more complex semantic augmentation techniques, and exploring the latest transformer models such as XLM-RoBERTa, ModernBERT, and instruction-tuned Large Language Models (LLMs) to further optimize classification performance.

5. REFERENCES

- [1] M. Borovic, M. Ojstersek, and D. Strnad, "A Hybrid Approach to Recommending Universal Decimal Classification Codes for Cataloguing in Slovenian Digital Libraries," *IEEE Access*, vol. 10, no. July, pp. 85595–85605, 2022, doi: 10.1109/ACCESS.2022.3198706.

- [2] X. Yang and Z. Zhang, "Enhancing book genre classification with BERT and InceptionV3: a deep learning approach for libraries," *PeerJ Comput. Sci.*, vol. 11, pp. 1–20, 2025, doi: 10.7717/peerj-cs.2934.
- [3] M. P. Satija and A. Kyrios, *A Handbook of History, Theory and Practice of the Dewey Decimal Classification System*. Facet, 2023. doi: 10.29085/9781783306114.
- [4] S. Gao et al., "Limitations of Transformers on Clinical Text Classification.," *IEEE J. Biomed. Heal. Informatics*, vol. 25, no. 9, pp. 3596–3607, Sep. 2021, doi: 10.1109/JBHI.2021.3062322.
- [5] R. Su, S. Gao, K. Zhao, and J. Zhang, "Adaptive feature interaction enhancement network for text classification," *Sci. Rep.*, vol. 15, no. 1, pp. 1–14, 2025, doi: 10.1038/s41598-025-95492-y.
- [6] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.*, vol. 1, no. M1m, pp. 4171–4186, 2019.
- [7] B. Willie et al., "IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding," *Proc. 1st Conf. Asia-Pacific Chapter Assoc. Comput. Linguist. 10th Int. Jt. Conf. Nat. Lang. Process. AACL-IJCNLP 2020*, pp. 843–857, 2020, doi: 10.18653/v1/2020.aacl-main.85.
- [8] H. Ahmadian, T. F. Abidin, H. Riza, and K. Muchtar, "Hybrid Models for Emotion Classification and Sentiment Analysis in Indonesian Language," *Appl. Comput. Intell. Soft Comput.*, vol. 2024, 2024, doi: 10.1155/2024/2826773.
- [9] K. E. Saputra and Riccosan, "Indonesian news article authorship attribution multilabel multiclass classification using IndoBERT," *IAES Int. J. Artif. Intell.*, vol. 13, no. 4, pp. 4688–4694, 2024, doi: 10.11591/ijai.v13.i4.pp4688-4694.
- [10] M. Arslan and C. Cruz, "Business text classification with imbalanced data and moderately large label spaces for digital transformation," *Appl. Netw. Sci.*, vol. 9, no. 1, 2024, doi: 10.1007/s41109-024-00623-5.
- [11] D. Refai, S. Abu-Soud, and M. J. Abdel-Rahman, "Data Augmentation Using Transformers and Similarity Measures for Improving Arabic Text Classification," *IEEE Access*, vol. 11, no. October, pp. 132516–132531, 2023, doi: 10.1109/ACCESS.2023.3336311.
- [12] Y. Cui, M. Jia, T. Y. Lin, Y. Song, and S. Belongie, "Class-balanced loss based on effective number of samples," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2019-June, pp. 9260–9269, 2019, doi: 10.1109/CVPR.2019.00949.
- [13] N. Mahmoodi, H. Shirazi, M. Fakhredanesh, and K. DadashtabarAhmadi, "Automatically weighted focal loss for imbalance learning," *Neural Comput. Appl.*, vol. 37, no. 5, pp. 4035–4052, 2025, doi: 10.1007/s00521-024-10323-x.
- [14] R. Escobar Díaz Guerrero, L. Carvalho, T. Bocklitz, J. Popp, and J. L. Oliveira, "A Data Augmentation Methodology to Reduce the Class Imbalance in Histopathology Images," *J. Imaging Informatics Med.*, vol. 37, no. 4, pp. 1767–1782, 2024, doi: 10.1007/s10278-024-01018-9.
- [15] Z. Cai, Z. Li, Y. Liu, L. Guo, and Y. Song, "Ustnlp16 at SemEval-2025 Task 9: Improving Model Performance through Imbalance Handling and Focal Loss," *ArXiv*, vol. abs/2505.00021, 2025, doi: 10.48550/arxiv.2505.00021.
- [16] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *J. Big Data*, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0192-5.
- [17] A. A. Khan, O. Chaudhari, and R. Chandra, "A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation," *Expert Syst. Appl.*, vol. 244, no. May 2023, p. 122778, 2024, doi: 10.1016/j.eswa.2023.122778.
- [18] Q. Xie, Z. Dai, E. Hovy, M. T. Luong, and Q. V. Le, "Unsupervised data augmentation

- for consistency training," *Adv. Neural Inf. Process. Syst.*, vol. 2020-Decem, no. NeurIPS, pp. 1–13, 2020.
- [19] L. Xiao and J. S. Zhang, "A novel transformer-based semantic feature extraction method for multi-label text classification," *Sci. Rep.*, vol. 16, no. 1, pp. 1–12, 2026, doi: 10.1038/s41598-025-30925-2.
- [20] C. H. Lin and U. Nuha, "Sentiment analysis of Indonesian datasets based on a hybrid deep-learning strategy," *J. Big Data*, vol. 10, no. 1, 2023, doi: 10.1186/s40537-023-00782-9.
- [21] R. Qasim, W. H. Bangyal, M. A. Alqarni, and A. Ali Almazroi, "A Fine-Tuned BERT-Based Transfer Learning Approach for Text Classification," *J. Healthc. Eng.*, vol. 2022, pp. 1–17, Jan. 2022, doi: 10.1155/2022/3498123.
- [22] S. Jamshidi *et al.*, "Effective text classification using BERT, MTM LSTM, and DT," *Data Knowl. Eng.*, vol. 151, p. 102306, 2024, doi: <https://doi.org/10.1016/j.datak.2024.102306>.
- [23] Jacob Eisenstein, *Introduction*. The MIT Press, 2019.
- [24] G. Zhang and J. Hu, "Enhanced industrial text classification via hyper variational graph-guided global context integration," *PeerJ Comput. Sci.*, vol. 10, 2024, doi: 10.7717/peerj-cs.1788.
- [25] N. K. Nissa and E. Yulianti, "Multi-label text classification of Indonesian customer reviews using bidirectional encoder representations from transformers language model," *Int. J. Electr. Comput. Eng.*, vol. 13, no. 5, pp. 5641–5652, 2023, doi: 10.11591/ijece.v13i5.pp5641-5652.
- [26] F. Baharuddin and M. F. Naufal, "Fine-Tuning IndoBERT for Indonesian Exam Question Classification Based on Bloom's Taxonomy," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 9, no. 2, pp. 253–263, 2023, doi: 10.20473/jisebi.9.2.253-263.
- [27] Dr. Li Deng and Y. Liu, *Deep Learning in Natural Language Processing*. Singapore: Springer Singapore, 2018. doi: 10.1007/978-981-10-5209-5.
- [28] T. Y. Tandil, T. F. Abidin, and H. Riza, "Incorporation of IndoBERT and Machine Learning Features to Improve the Performance of Indonesian Textual Entailment Recognition," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 11, no. 2, pp. 173–186, 2025, doi: 10.20473/jisebi.11.2.173-186.
- [29] H. Lu, L. Ehwerhemuepha, and C. Rakovski, "A comparative study on deep learning models for text classification of unstructured medical notes with various levels of class imbalance," *BMC Med. Res. Methodol.*, vol. 22, no. 1, pp. 1–12, 2022, doi: 10.1186/s12874-022-01665-y.