JOIN (Jurnal Online Informatika)
p-ISSN: 2528-1682, e-ISSN: 2527-9165
Volume 10 Number 2 | December 2025: 260-269
DOI: 10.15575/join.v10i2.1499

Comparative Analysis of IndoBERT and LSTM for Multi-Label Text Classification of Indonesian Motivation Letter

Yosep Setiawan[1], Lili Ayu Wulandhari[2]
[1,2]Department of Computer Science, Binus University Jakarta, Indonesia

Article Info

Article history:

Received November 18, 2024
Revised February 06, 2025
Accepted February 06, 2025
Published August 17, 2025

Keywords:

Comparative analysis
IndoBERT
LSTM
Motivation letter
Multi-label text classification

ABSTRACT

The evaluation of motivation letters is a crucial step in the student admission process for one of vocational institutions in Indonesia. However, the current manual assessment method is prone to subjectivity and inconsistency, making it less reliable for fair student selection. This research presents a comparative analysis of two deep learning models, IndoBERT and Long Short-Term Memory (LSTM), for multi-label text classification of motivation letters written in Indonesian. Using a dataset of 676 motivation letters labeled with nine predefined categories, we evaluate the models based on their classification performance. The results indicate that IndoBERT outperforms LSTM, achieving an F1-score of 81%, compared to 76% for LSTM. This research provides insights into the effectiveness of IndoBERT for multi-label classification tasks in the Indonesian language and serves as a benchmark for future research in automating motivation letter evaluations.

Corresponding Author:

Yosep Setiawan,
Department of Computer Science, Binus University Jakarta, Indonesia
Jl. Kebon Jeruk Raya No. 27, Kebon Jeruk Jakarta Barat 11530, Indonesia
Email: yosep.setiawan001@binus.ac.id

1. INTRODUCTION

The evaluation of motivation letters is a crucial step in the student selection process at vocational institutions in Indonesia. These letters reflect applicants' aspirations, motivations, and suitability for their chosen research programs, but manual evaluation is prone to subjectivity and inconsistency, raising concerns about fairness and reliability.

To improve objectivity and efficiency, this research proposes a multi-label text classification approach capable of assessing multiple aspects of motivation letters simultaneously, such as motivation, achievements, and instructional suitability. Instead of assigning only one label per document, multiple labels capture the different evaluation dimensions present in each letter.

Advances in deep learning have introduced models such as BERT and LSTM, which excel in handling complex text classification tasks. Prior studies have shown BERT's superior performance in document classification compared to CNN, LSTM, and SVM models [1]–[4]. In the Indonesian NLP context, IndoBERT, which is BERT pretrained on large-scale Indonesian text has demonstrated strong results in various multi-label classification tasks, including customer review analysis, toxic comment detection, and sentiment classification [5]–[8]. These works indicate IndoBERT's capacity to capture rich contextual information across domains.

LSTM-based architectures, on the other hand, are valued for their ability to capture sequential dependencies in text. They have been successfully applied in multi-label classification across languages and domains [9]–[11], and continue to be competitive, especially when enhanced with mechanisms such as attention layers.

However, no prior research has directly compared IndoBERT and LSTM for multi-label text classification of Indonesian motivation letters. Addressing this gap, our research provides a direct performance comparison and establishes a benchmark for future work in automated motivation letter evaluation, contributing to both educational technology and the broader field of Indonesian NLP.

2. METHOD

In this section, we outline the research methodology of the proposed approach, offering a detailed discussion of each step. We introduce a deep learning model specifically designed for the multi-label text classification of motivation letters written in Indonesian.

2.1. Flow Chart

This research aims to predict motivation letter scores by integrating text mining and deep learning techniques. The primary focus is to objectively identify motivation letters that meet institutional criteria. The initial step involves conducting a brief literature review to provide foundational information for the development of the methodology. We then adopt a multi-label text classification approach using the IndoBERT and LSTM models, as illustrated in Figure 1. The performance of both models will be compared to evaluate their effectiveness in predicting scores for motivation letters written in Indonesian.

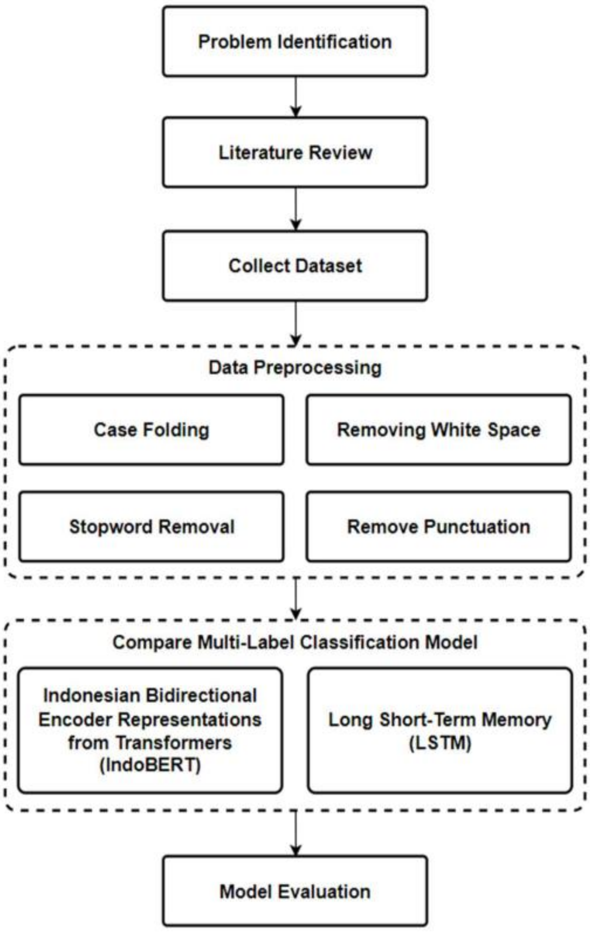


Figure 1. Flow Chart

p-ISSN: 2528-1682
e-ISSN: 2527-9165

2.2. Data Preparation

This research utilized 676 motivation letter files submitted by prospective students in 2022 and 2023. Initially in Portable Document Format (PDF), the dataset was converted to Comma Separated Values (CSV) format using the Python library Pdfplumber [12]. Figure 2 illustrates the data processing workflow, highlighting the conversion from PDF to CSV.

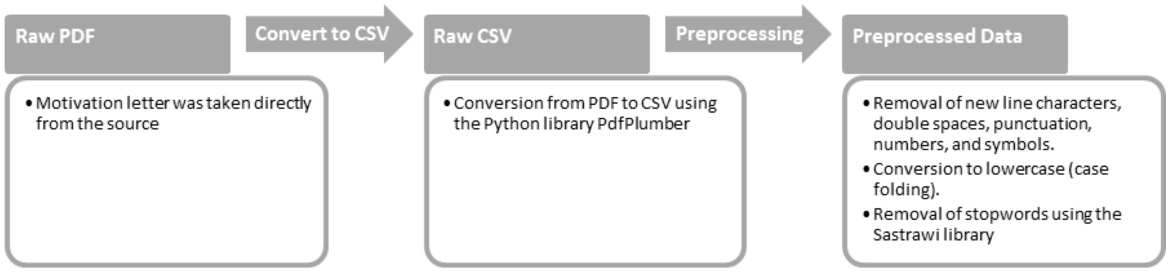
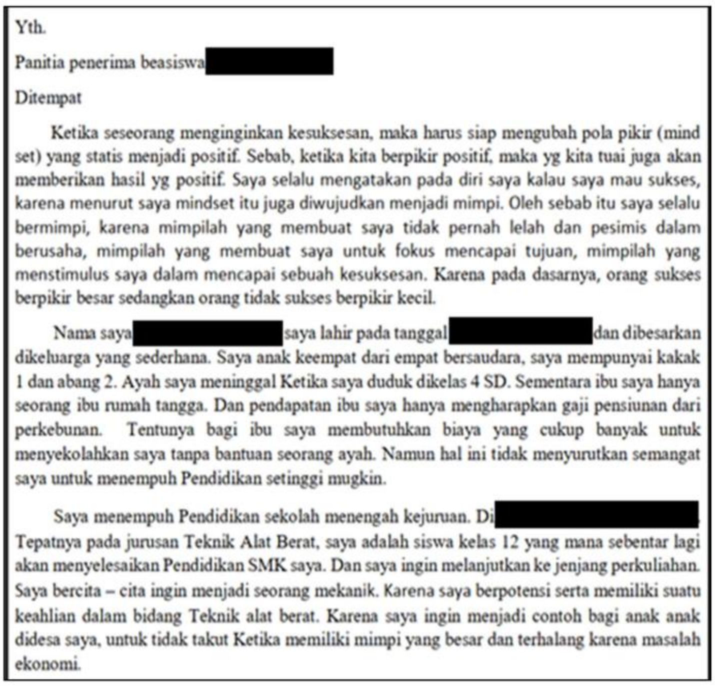


Figure 2. Data Processing Workflow

This dataset, after conversion to CSV, remained raw and required preprocessing. We performed data preprocessing by removing newline characters, double spaces, punctuation, numbers, and symbols, as well as applying case folding to lowercase and removing stopwords using the Sastrawi library [13].

Such preprocessing steps are crucial for Indonesian NLP, as inconsistent orthography, affixation, and common stopwords can reduce the performance of downstream models if not addressed. Tokenization and stopword removal in Indonesian differ significantly from English, making the use of libraries like Sastrawi highly relevant.

An example of the raw data can be seen in Figure 3.



Yth.

Panitia penerima beasiswa [redacted]

Ditempat

Ketika seseorang menginginkan kesuksesan, maka harus siap mengubah pola pikir (mind set) yang statis menjadi positif. Sebab, ketika kita berpikir positif, maka yg kita tuai juga akan memberikan hasil yg positif. Saya selalu mengatakan pada diri saya kalau saya mau sukses, karena menurut saya mindset itu juga diwujudkan menjadi mimpi. Oleh sebab itu saya selalu bermimpi, karena mimpilah yang membuat saya tidak pernah lelah dan pesimis dalam berusaha, mimpilah yang membuat saya untuk fokus mencapai tujuan, mimpilah yang menstimulus saya dalam mencapai sebuah kesuksesan. Karena pada dasarnya, orang sukses berpikir besar sedangkan orang tidak sukses berpikir kecil.

Nama saya [redacted] saya lahir pada tanggal [redacted] dan dibesarkan dikeluarga yang sederhana. Saya anak keempat dari empat bersaudara, saya mempunyai kakak 1 dan abang 2. Ayah saya meninggal Ketika saya duduk dikelas 4 SD. Sementara ibu saya hanya seorang ibu rumah tangga. Dan pendapatan ibu saya hanya mengharapkan gaji pensiunan dari perkebunan. Tentunya bagi ibu saya membutuhkan biaya yang cukup banyak untuk menyekolahkan saya tanpa bantuan seorang ayah. Namun hal ini tidak menyurutkan semangat saya untuk menempuh Pendidikan setinggi mugkin.

Saya menempuh Pendidikan sekolah menengah kejuruan. Di [redacted]. Tepatnya pada jurusan Teknik Alat Berat, saya adalah siswa kelas 12 yang mana sebentar lagi akan menyelesaikan Pendidikan SMK saya. Dan saya ingin melanjutkan ke jenjang perkuliahan. Saya bercita – cita ingin menjadi seorang mekanik. Karena saya berpotensi serta memiliki suatu keahlian dalam bidang Teknik alat berat. Karena saya ingin menjadi contoh bagi anak anak didesa saya, untuk tidak takut Ketika memiliki mimpi yang besar dan terhalang karena masalah ekonomi.

Figure 3. Example of Raw Motivation Letter Data

After text preprocessing, this dataset has an average of 3,427 characters and 468 words per document. In the statistical analysis, the shortest document in the dataset contains 565 characters and

20 words, while the longest document reaches 7,849 characters and 1,092 words. The results of the text cleaning and stopword removal process can be seen in Figure 4.

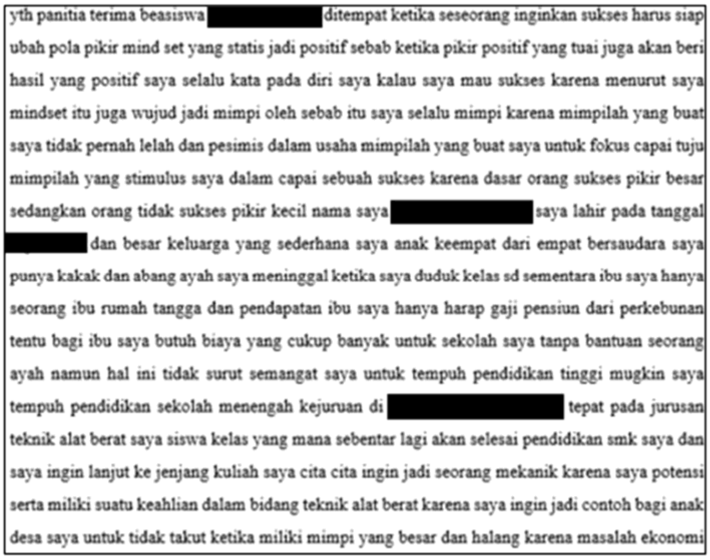

yth panitia terima beasiswa [redacted] ditempat ketika seseorang inginkan sukses harus siap ubah pola pikir mind set yang statis jadi positif sebab ketika pikir positif yang tuai juga akan beri hasil yang positif saya selalu kata pada diri saya kalau saya mau sukses karena menurut saya mindset itu juga wujud jadi mimpi oleh sebab itu saya selalu mimpi karena mimpilah yang buat saya tidak pernah lelah dan pesimis dalam usaha mimpilah yang buat saya untuk fokus capai tuju mimpilah yang stimulus saya dalam capai sebuah sukses karena dasar orang sukses pikir besar sedangkan orang tidak sukses pikir kecil nama saya [redacted] saya lahir pada tanggal [redacted] dan besar keluarga yang sederhana saya anak keempat dari empat bersaudara saya punya kakak dan abang ayah saya meninggal ketika saya duduk kelas sd sementara ibu saya hanya seorang ibu rumah tangga dan pendapatan ibu saya hanya harap gaji pensiun dari perkebunan tentu bagi ibu saya butuh biaya yang cukup banyak untuk sekolah saya tanpa bantuan seorang ayah namun hal ini tidak surut semangat saya untuk tempuh pendidikan tinggi mugkin saya tempuh pendidikan sekolah menengah kejuruan di [redacted] tepat pada jurusan teknik alat berat saya siswa kelas yang mana sebentar lagi akan selesai pendidikan smk saya dan saya ingin lanjut ke jenjang kuliah saya cita cita ingin jadi seorang mekanik karena saya potensi serta miliki suatu keahlian dalam bidang teknik alat berat karena saya ingin jadi contoh bagi anak desa saya untuk tidak takut ketika miliki mimpi yang besar dan halang karena masalah ekonomi

Figure 4. Example of Preprocessed Motivation Letter Data

In addition to the motivation letter data, labeled data for the motivation letters is available in a Microsoft Excel file provided by the Department of Student and Alumni Affairs. This dataset includes the registration number, name, study program, and nine labels. Each label is represented in binary form (1 or 0), indicating the presence or absence of a specific context within the motivation letter. This binary labeling scheme aligns with the conventional representation of multi-label tasks in NLP research [14], allowing models to predict each label independently at the output layer.

It is important to note that these label values, assigned as 1 or 0, are based on manual assessments carried out by human evaluators. For instance, a label is assigned a value of 1 if a particular context is present in the letter, and 0 if it is absent. The data is then manually converted into CSV format for further processing. Table 1 provides an example of the motivation letter label data.

Table 1. Example of Motivation Letter Label Data

Registration Number	Name	Study Program	Label 1	...	Label 9
222300020	Mr. A	MI	1	...	0
222300025	Mr. B	P4	1	...	1
222300048	Mr. C	P4	0	...	1
222300050	Mr. D	MO	1	...	1
222300087	Mr. E	MI	1	...	0

These nine labels have been assigned, with each label representing a distinct context. Table 2 provides an explanation of the meaning of each of these labels.

Table 2. Label Data Context

Label #	Context
1	Willingness to Learn
2	Readiness for Further Studies
3	Firmness of Choice
4	Motivation
5	Self-Advantages
6	Suitability of Instructions
7	Achievement
8	Language
9	Experience

In our analysis, we focused particularly on instances where the label had a value of 1. We carefully examined the distribution of these instances within the dataset to assess the prevalence of positive labels. Table 3 reveals a notable data imbalance, especially with label 7 and label 9, which shows a significantly lower frequency of occurrences with a value of 1 compared to the other labels.

Table 3. Positive Label Count

Label #	Positive Label Count	Variance (%)
1	595	88.02
2	507	75.00
3	469	69.38
4	449	66.42
5	424	62.72
6	604	89.35
7	345	51.04
8	472	69.82
9	216	31.95

Since the motivation letter dataset is in text form and involves multi-label classification, we acknowledge that no additional steps can be taken to balance the data. The next step was to integrate the motivation letter data (in CSV format) with the corresponding label data (also in CSV format). The registration number served as the key element for linking the two datasets. This integration process was carefully monitored to ensure that no data was mismatched, guaranteeing that each motivation letter was paired with its correct label. As a result, this integration produced a cohesive and structured dataset that will be used in the subsequent stages of the research.

2.3. Model Design

This research compares two models for performing multi-label text classification on motivation letters: BERT for Indonesian text (IndoBERT) and LSTM. Each model uses different tokenization and vectorization processes to transform the data into numerical vector representations. An illustration of this process can be seen in Figure 5.

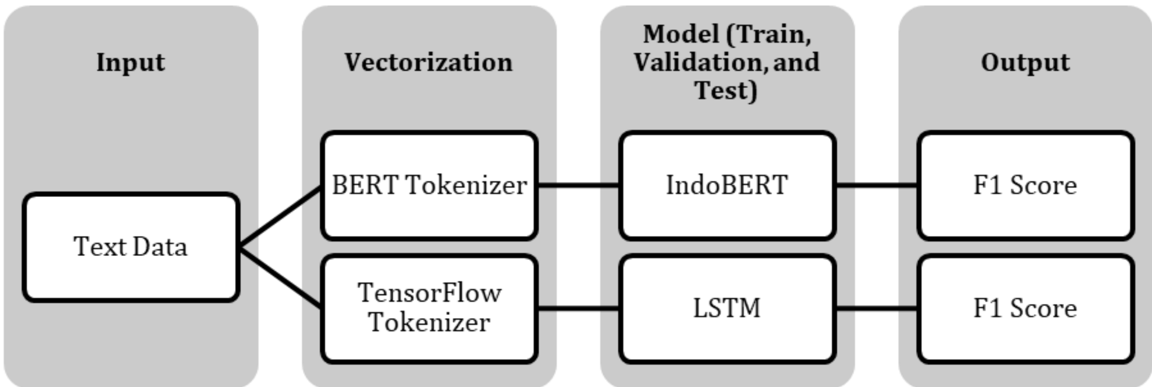


Figure 5. Multi-Label Text Classification Model Process

In the IndoBERT model, vectorization is performed using the pretrained BERTTokenizer, while LSTM uses TensorFlow's tokenizer. After obtaining the vector representations, multi-label text classification is carried out, with example of final output represented as a binary, as shown in Table 4.

Table 4. Example of Multi-Label Text Classification Results for Motivation Letter

Label #	Context	Example of Classification Result
1	Willingness to Learn	0
2	Readiness for Further Studies	1

Label #	Context	Example of Classification Result
3	Firmness of Choice	0
4	Motivation	0
5	Self-Advantages	0
6	Suitability of Instructions	1
7	Achievement	0
8	Language	0
9	Experience	0

2.4. IndoBERT Model

Bidirectional Encoder Representations from Transformers (BERT), developed by Google using the Transformer architecture [15], revolutionizes natural language processing (NLP) with its self-attention mechanism [16]. BERT enables deep contextual understanding of text, excelling in tasks like text classification and named entity recognition through pretraining and fine-tuning [15].

While BERT is widely used for English text, IndoBERT, pre-trained on over 220 million Indonesian words, serves as a strong alternative for Bahasa Indonesia [17]. IndoBERT achieves over 70% accuracy across various NLP tasks and shows great potential for multi-label text classification [18].

This IndoBERT model used in this research has undergone pre-training, allowing the researchers to conduct experiments by adjusting hyperparameters such as the number of epochs, learning rate, batch size, and max length, as shown in Table 5.

Table 5. IndoBERT Model Parameter Configuration

Experiment #	Epoch	Learning Rate	Batch Size	Max Length
1	20	0.00001	32	64
2	20	0.00001	32	128
3	50	0.00001	32	256
4	50	0.00001	32	384
5	50	0.00001	16	384
6	50	0.00001	16	512

2.5. LSTM Model

Long Short-Term Memory (LSTM) networks [19] retain context in text through memory cells that manage information flow using forget, input, and output gates [20], effectively handling long-term dependencies in text. While transformer-based models have gained prominence, LSTM remains competitive for smaller datasets and is less resource-intensive, making it a relevant baseline in educational NLP applications [20]. In our research, we implemented a BiLSTM architecture with dropout regularization and experimented with varying batch sizes, as shown in Table 6 and Table 7.

Table 6. LSTM Model Parameter Configuration

Experiment #	Epoch	Learning Rate	Batch Size
1	30	0.001	16
2	30	0.001	32

Table 7. LSTM Neural Network Architecture

Layer #	Layer Types	Input	Output	Activation Function
1	Embedding	Sequence	(Sequence, 512)	-
2	Bidirectional LSTM	(Sequence, 512)	(1024,)	-
3	Bidirectional LSTM	(1024,)	(1024,)	-
4	Dense	(1024,)	(256,)	ReLU
5	Dropout (0.5)	(256,)	(256,)	-
6	Dense (Output)	(256,)	(9,)	Sigmoid

2.6. Model Evaluation

In each model, data was split into 80% for training, 10% for validation, and 10% for testing. The evaluation was based on metrics such as accuracy, precision, recall, and F1 score, with the F1 score selected for its ability to balance precision and recall. In the results and discussion chapter, only the F1 score was reported, as it provided a more comprehensive measure of classification performance.

For F1 score calculation, the classification_report function from scikit-learn was used with the micro-average method. The micro-average F1 is chosen due to class imbalance in the dataset, as it aggregates contributions of all labels, providing a more reliable global measure than macro-average F1

p-ISSN: 2528-1682
e-ISSN: 2527-9165

in imbalanced scenarios [21]. This method gives equal weight to each label and calculates the F1 score based on the total number of true positives (TP), false negatives (FN), and false positives (FP) across all labels, as shown in the equation below.

$$F1_{micro} = \frac{(2 * \sum_{i=1}^{9} TP_i)}{(2 * \sum_{i=1}^{9} TP_i) + \sum_{i=1}^{9} FP_i + \sum_{i=1}^{9} FN_i} \quad (1)$$

Explanation:
TP_i = True positives for label i
FN_i = False negatives for label i
FP_i = False positives for label i

3. RESULT AND DISCUSSION

3.1. *Results*

In this research, we explored several hyperparameter combinations for the training and validation of the IndoBERT and LSTM models. The results, including train accuracy, train loss, validation accuracy, and validation loss, are presented in the form of line graphs.

For the IndoBERT model, we conducted experiments with various hyperparameter combinations, as listed in Table 5. The experimental results are organized according to the numbers in the table and are shown in Table 8.

Table 8. Summary of IndoBERT Training and Validation Performance

Experiment #	Epoch	Batch Size	Max Length	Train Loss	Train Accuracy	Validation Loss	Validation Accuracy
1	20	32	64	0.3961	0.8321	0.5932	0.6699
2	20	32	128	0.3695	0.8496	0.6506	0.6797
3	50	32	256	0.1185	0.9872	0.7493	0.6471
4	50	32	384	0.0960	0.9928	0.6997	0.6846
5	50	16	384	0.0508	1.0000	1.0147	0.6569
6	**50**	**16**	**512**	**0.0441**	**1.0000**	**0.9369**	**0.7010**

Based on the train loss, train accuracy, validation loss, and validation accuracy values in Table 8, IndoBERT demonstrates good performance. We conclude that as the Max Length parameter increases, the IndoBERT model becomes more accurate in classification. However, it is important to note that as the Batch Size decreases, the validation loss tends to increase, indicating that the model may lose its ability to generalize to new, unseen data.

In the LSTM model, we conducted experiments with various hyperparameter combinations listed in Table 6. The experimental results will be organized according to the numbers in the table, and the results of the experiments can be seen in Table 9.

Table 9. Summary of LSTM Training and Validation Performance

Experiment #	Epoch	Batch Size	Train Loss	Train Accuracy	Validation Loss	Validation Accuracy
1	30	16	0.0721	0.9731	1.6010	0.6127
2	**30**	**32**	**0.0678**	**0.9750**	**1.5526**	**0.6356**

Based on the train loss, train accuracy, validation loss, and validation accuracy values in Table 9, the LSTM model shows decent progress, though not as good as the IndoBERT model. The performance gap reflects IndoBERT's ability to capture bidirectional context and token semantics better than sequential models like LSTM, especially in morphologically rich languages such as Indonesian.

Next, we conducted testing using 10% of the entire dataset, which consists of 68 data. The values reported include the F1 scores for each label as well as the micro-average for all labels. A comparison of the performance between the IndoBERT and LSTM classification models can be seen in Table 10. For

IndoBERT, the reported F1 score is the highest when using a Max Length hyperparameter of 512. For LSTM, the highest F1 score is reported when using a Batch Size hyperparameter of 32.

Table 10. Classification Performance of IndoBERT and LSTM Models

Model	**F1 Score**									
	Label 1	**Label 2**	**Label 3**	**Label 4**	**Label 5**	**Label 6**	**Label 7**	**Label 8**	**Label 9**	**Micro-Average**
IndoBERT	0.95	0.85	0.87	0.84	0.68	0.97	0.60	0.80	0.15	**0.81**
LSTM	0.91	0.83	0.73	0.68	0.68	0.96	0.53	0.79	0.24	**0.76**

From the series of tests conducted, it can be concluded that the use of two algorithms, IndoBERT and LSTM, for multi-label text classification on the Indonesian motivation letter dataset resulted in varying F1 scores. The results show that IndoBERT slightly outperforms LSTM, both for individual labels and the micro-average. However, it should be noted that for label 9 (Experience), both models performed poorly in classification. This was due to the inadequacy of the dataset for training, validation, or testing on this particular label.

Overall, the test results indicate that IndoBERT outperforms the LSTM model, thanks to its better capability in multi-label classification, with a micro-average F1 score of 0.81, despite the limited number of labels and dataset size.

In general, the results suggest a strong relationship between the frequency of positive label occurrences (as shown in Table 3) and the resulting F1 scores. Labels with a higher frequency of positive occurrences tend to have higher F1 scores, while those with lower frequencies tend to have lower F1 scores.

3.2. Discussions

The results of this research provides a comparative analysis of IndoBERT and LSTM for multi-label text classification of motivation letters. Both models perform well on the Indonesian motivation letter dataset. However, the findings indicate that the proposed approach can mitigate subjectivity inherent in manual assessments. IndoBERT, with its deep contextualized embeddings, captures nuanced information from the text, making it more reliable than LSTM, which primarily relies on sequential patterns. This enhanced performance suggests that IndoBERT can improve the fairness and consistency of motivation letter evaluations.

A key contribution of this research is establishing IndoBERT as a benchmark model for multi-label classification of motivation letters. The evaluation results show that IndoBERT achieves an F1-score of 81%, positioning it as a promising candidate for automated motivation letter assessment in Indonesian. This benchmark provides a valuable reference point for future research in this area.

Nonetheless, a significant challenge observed in this research is the uneven distribution of labels, particularly for Label 7 (Achievement) and Label 9 (Experience), which negatively impacted model performance. To improve classification accuracy for these underrepresented labels, future work should focus on acquiring a more balanced dataset or employing data augmentation techniques. However, given the difficulty in generating realistic examples for such specific labels, prioritizing the collection of additional real-world data is recommended.

Several refinements can further enhance the models' effectiveness. For instance, fine-tuning IndoBERT on domain-specific data related to student applications could improve its accuracy. Additionally, exploring hybrid models that combine IndoBERT's contextual richness with the sequential modeling capabilities of LSTM may yield better results. Integrating attention mechanisms into the LSTM architecture could also enhance its ability to capture long-term dependencies in text classification.

4. CONCLUSION

This research addressed the challenge of achieving multi-label text classification of Indonesian texts by comparing two deep learning models, IndoBERT and LSTM. Our experiments demonstrate that IndoBERT, with its deep contextual representations, achieved a micro-average F1 score of 81% compared to 76% for LSTM, clearly establishing its advantage and setting a benchmark for Indonesian text classification.

While our analysis reveals data imbalance issues, particularly for underrepresented labels, it also highlights the need for future work to expand and balance datasets, incorporate advanced techniques like N-grams and attention mechanisms, and explore alternative transformer-based architectures.

The findings contribute to both academic research and practical applications, particularly in automating the assessment of motivation letters to improve fairness and reduce subjectivity in student admissions. This aligns with broader trends in applying AI for educational evaluation tasks

Overall, our research contributes to the development of more objective and reliable multi-label classification methods for Indonesian texts, with broad implications for various real-world applications. Future work should also consider hybrid architectures that integrate IndoBERT embeddings with lightweight sequential models for deployment in resource-constrained educational environments.

ACKNOWLEDGEMENTS

This research was supported by Bina Nusantara University and Astra Polytechnic. We thank our colleagues from these institutions who provided insight and expertise that greatly assisted the research, although they may not agree with all of the interpretations/conclusions of this paper.

We also acknowledge the open-source NLP community, particularly contributors to the IndoBERT and HuggingFace Transformers projects, whose tools were instrumental in conducting this research.

REFERENCES

[1] A. Adhikari, A. Ram, R. Tang, and J. Lin, "DocBERT: BERT for Document Classification," arXiv:1904.08398, 2019.

[2] W.-C. Chang, H.-F. Yu, K. Zhong, Y. Yang, and I. Dhillon, "X-BERT: eXtreme Multi-label Text Classification with BERT," arXiv:1905.02331, 2019.

[3] L. Cai, Y. Song, T. Liu, and K. Zhang, "A Hybrid BERT Model That Incorporates Label Semantics via Adjustive Attention for Multi-label Text Classification," IEEE Access, 2020, doi:10.1109/ACCESS.2020.3017382.

[4] I. Chalkidis, M. Fergadiotis, P. Malakasiotis, and I. Androutsopoulos, "Large-Scale Multi-Label Text Classification on EU Legislation," arXiv:1906.02192, 2019.

[5] 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, Oct. 2023, doi:10.11591/ijece.v13i5.pp5641-5652.

[6] G. Z. Nabiilah, I. Nur, E. S. Purwanto, and M. F. Hidayat, "Indonesian multilabel classification using IndoBERT embedding and MBERT classification," Int. J. Electr. Comput. Eng., vol. 14, no. 1, pp. 1071–1078, Feb. 2024, doi:10.11591/ijece.v14i1.pp1071-1078.

[7] N. C. Mei, S. Tiun, and G. Sastria, "Multi-Label Aspect-Sentiment Classification on Indonesian Cosmetic Product Reviews with IndoBERT Model," Int. J. Adv. Comput. Sci. Appl., vol. 15, no. 11, 2024, doi:10.14569/IJACSA.2024.0151168.

[8] Y. Sagama and A. Alamsyah, "Multi-label Classification of Indonesian Online Toxicity Using BERT and RoBERTa," in Proc. IAICT, 2023, doi:10.1109/IAICT59002.2023.10205892.

[9] Y. Yan et al., "LSTM2: Multi-Label Ranking for Document Classification," Neural Process Lett 47, 117–138, 2018, doi:10.1007/s11063-017-9636-0.

[10] B. Alsukhni, "Multi-Label Arabic Text Classification Based On Deep Learning," 12th International Conference on Information and Communication Systems (ICICS), Valencia, Spain, 2021, pp. 475-477, doi: 10.1109/ICICS52457.2021.9464538.

[11] L. Enamoto et al., "Multi-label legal text classification with BiLSTM and attention," International Journal of Computer Applications in Technology, 2022, Vol. 68 No. 4, pp. 369-378, doi:10.1504/IJCAT.2022.125186

[12] S. J. Vine, "Pdfplumber: Plumb a PDF for Detailed Information About Each Char, Rectangle, and Line," 2025, [Online]. Available: https://pypi.org/project/pdfplumber/.

[13] H. A. Robbani, "Sastrawi: Library for Stemming Indonesian (Bahasa) Text," 2025, [Online]. Available: https://pypi.org/project/Sastrawi/.

[14] S. Huang, W. Hu, B. Lu, Q. Fan, X. Xu, X. Zhou, and H. Yan, "Application of Label Correlation in Multi-Label Classification: A Survey," Appl. Sci., vol. 14, no. 19, p. 9034, 2024, doi:10.3390/app14199034.

[15] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language

Understanding," arXiv:1810.04805v2, 2019.

[16] A. Vaswani et al., "Attention Is All You Need," arXiv:1706.03762v7, 2023.

[17] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," arXiv:2011.00677v1, 2020.

[18] Indolem, "IndoBERT: Indonesian Version of BERT Model," 2025, [Online]. Available: https://huggingface.co/indolem/indobert-base-uncased.

[19] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997, doi:10.1162/neco.1997.9.8.1735.

[20] S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu, and J. Gao, "Deep Learning Based Text Classification: A Comprehensive Review," arXiv:2004.03705v3, 2021.

[21] Z. C. Lipton, C. Elkan, and B. Narayanaswamy, "Thresholding Classifiers to Maximize F1 Score," arXiv:1402.1892v2, 2014.