

TEXT CLASSIFICATION USING INDOBERT FINE-TUNING MODELING WITH CONVOLUTIONAL NEURAL NETWORK AND BI-LSTM

Alda Zevana Putri Widodo^{*1}, Dwiza Riana²

^{1,2}Computer Science, Department of Information Technology, Nusa Mandiri University, Indonesia
Email: ¹14210192@nusamandiri.ac.id, ²dwiza@nusamandiri.ac.id

(Article received: December 28, 2023; Revision: Januari 06, 2024; published: January 15, 2024)

Abstract

The technological advancements in goods delivery facilities have been increasing year by year in tandem with the growing online trade, which necessitates delivery services to fulfill the transactional process between sellers and buyers. Since 2000, top brand awards have often conducted official survey analyses to provide comparisons of goods or services, one of which includes delivery services. However, the survey rankings based on public opinion are less accurate due to users of delivery services and service companies being unaware of the specific success factors and weaknesses in their services. The aim of this research is to analyze the comparison of text mining using the Indonesian language transformation method, IndoBert. The algorithm utilized to demonstrate analysis performance employs Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory (Bi-LSTM). This method is utilized to determine the impact of opinion data from Twitter on the J&T Express expedition delivery service, incorporating both text preprocessing and data without text preprocessing. The IndoBert parameters vary in the learning rate section based on four factors: price, time, returns, and others. The research data consisted of 2525 comments from Twitter users regarding the delivery service spanning from January 1, 2021, to March 31, 2023. The testing showed that Bi-LSTM with text preprocessing performed 2% higher, achieving 79% at a learning rate of 1×10^{-6} , compared to without text preprocessing at the same learning rate, which reached 77%. Additionally, CNN outperformed by 3% with a rate of 83%, compared to 80% without text preprocessing at a learning rate of 1×10^{-5} . The highest accuracy, reaching 83%, was obtained by CNN with parameters set at 1×10^{-5} , and the preprocessing technique was considered superior to Bi-LSTM.

Keywords: Bi-LSTM, Convolutional Neural Network, Delivery, IndoBert Fine-Tuning, Twitter.

1. INTRODUCTION

Customer ratings on social media represent opinions derived from individuals' experiences as consumers of services or products [1]. These assessments serve as valuable reviews for other consumers and service or product sellers in enhancing their offerings. Twitter, a prevalent social media platform among Indonesians, facilitates internet-based communication, allowing users to post various photos and videos on their accounts [2]. J&T Express is a recently established company specializing in delivering goods, including documents and packages [3]. In 2023, J&T Express achieved the Top Brand Award in the Best Courier Service category, securing the highest score of 20 in a comparison of seven startups categorized as unicorns. Additionally, it claimed the highest delivery volume among five renowned shipping companies in Indonesia, namely JNE, SiCepat, Ninja Express, and SAP [4].

Text mining involves analyzing substantial amounts of unstructured data to generate new insights from previously unclear information [5]. Text Mining is used to extract useful information from data sources by identifying and finding interesting pattern

relationships. Text Mining also leads to the research field of Data Mining [6].

The learning rate, a crucial hyperparameter, governs the degree of model adjustment in response to the error estimate during weight updates [7]. Selecting an appropriate learning rate poses challenges since a small value can prolong the training process or lead to stagnation, while a large value may result in unstable training or suboptimal weight learning [8].

Adam optimization is a stochastic gradient descent method based on adaptive estimation of first- and second-order moments [9]. The name "Adam" comes from "Adaptive Moment Estimation" because Adam uses first and second moment gradient estimates to adjust the learning rate for each weight in the neural network. The Adam optimizer combines the best properties of the AdaGrad and RMSProp algorithms to provide a more optimal algorithm which can handle diffuse and noisy gradients [10]. Adam is relatively easy to configure with default configuration parameters working well for most problems generated during training.

Batch size, representing the number of samples processed before updating the model, must range from one to the total samples in the training dataset

[11] Given that deep neural networks cannot process the entire dataset at once, batch size divides the dataset into manageable groups or segments.

An epoch refers to one round of training a neural network with all available training data [12]. The number of epochs dictates the number of times the deep learning algorithm runs. However, excessive epochs can lead to underfitting, where the model fails to discern meaningful relationships between input and output data [13].

The dropout layer within neural networks prevents overfitting by excluding individual nodes during training processes using probabilities [14]. In this process, individual nodes are excluded in various training processes using probabilities, as if they were not part of the network architecture at all [15]. In general, 0.5 is used, which means removing 50% of the unit. In the dropout layer, the dropout unit is selected randomly, then the unit is not used. Meanwhile, other units will continue learning.

The research carried out this time analyzed the comparison of text mining with the Indonesian language transformation method of Bidirectional Encoder Representations form Transformsis (BERT), namely IndoBERT. IndoBERT is a language modeling extension of BERT which is used in various types of linguistic tasks, such as sentiment analysis, question answering, text prediction, text generation, and text summarization [16].

The algorithm used to display analysis performance uses Convolutional Neural Network (CNN) and Bidirectional Long Short Term Memory (Bi-LSTM), this method is used to determine the influence of opinion data from Twitter on the J&T Express expedition delivery service which has been carried out text preprocessing and data text preprocessing was not carried out with IndoBERT parameters which varied in the learning rate [17].

This study analyzes the comparison of text mining using the IndoBERT transformation method, which is an extension of BERT specifically designed for the Indonesian language. It employs Convolutional Neural Network (CNN) and Bidirectional Long Short Term Memory (Bi-LSTM) algorithms to evaluate the impact of Twitter opinions on J&T Express delivery services. By varying IndoBERT parameters concerning learning rates, the study aims to determine the optimal algorithm performance. Previous studies have classified tweet reviews into positive, neutral, and negative sentiments, achieving 75% accuracy using a learning rate parameter of 3×10^{-4} with CNNs [18].

Additionally, research utilizing IndoBERT Fine-tuning with LSTM, BiLSTM, CNN, and GRU algorithms achieved the highest accuracy of 84.47% with CNN [19].

Bi-LSTM, a neural network with two LSTM layers, processes information bidirectionally to capture both past and future contexts [20]. This

architecture amalgamates two opposing hidden layers into one output [21].

CNN, an integral deep learning architecture, obviates the need for manual feature extraction by integrating convolution into artificial neural networks [22]. It aims to recognize new objects or images based on detected features.

This research compares the fine-tuning of the IndoBERT base model on preprocessed and non-preprocessed datasets, aiming to ascertain the influence of the J&T Express delivery service dataset on the IndoBERT base model using Bert Embedding. Two deep neural network algorithm architectures—Bidirectional LSTM and CNN—are assessed using training and test data, evaluating performance metrics such as accuracy, F1 score, precision, and recall. The goal is to determine the best performing model after fine-tuning the IndoBERT base model and to optimize the model by varying learning rate optimization values to gauge their impact on model performance.

2. RESEARCH METHODS

The research methods in this study can be seen in Figure 1, and Figure 2 illustrates the framework used in this research.

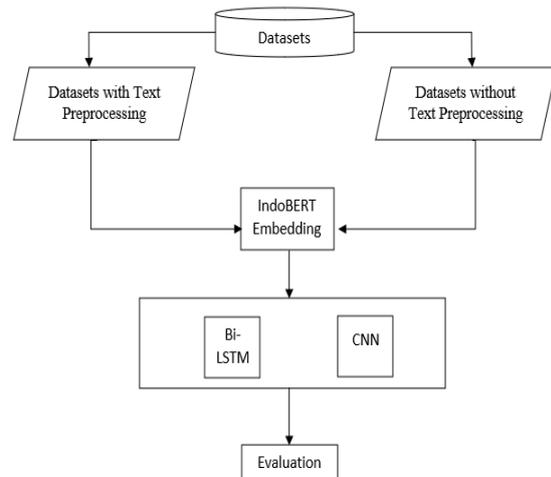


Figure 1. Research Flow

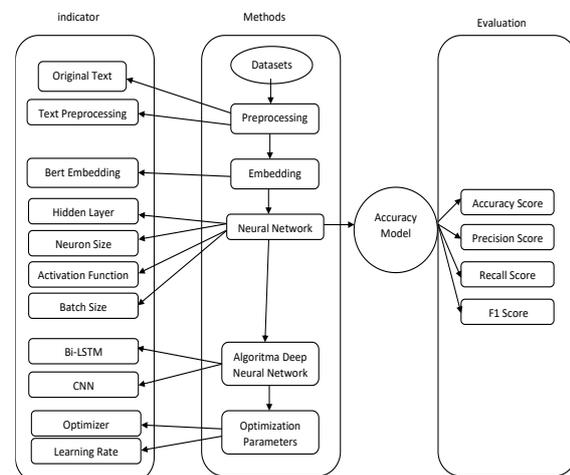


Figure 2. Framework

This research utilizes Python programming language version 3.10.0 and the Visual Studio Code application. The library packages employed include TensorFlow, Pandas, NumPy, Keras for modeling using IndoBert with the Bi-LSTM and CNN algorithms, and for visualizing image plots, Pylot and Graphiz were utilized.

The dataset for this research comprises textual data in the form of tweets sourced from the official J&T Express Twitter account spanning from January 20, 2021, to March 31, 2023, totaling 2525 data entries. Experts from J&T Express categorized the dataset into four categories: price, time, returns, and others. The dataset is structured with three columns: date, tweet, and category. Data collection was executed using the Python programming language, specifically leveraging Tweepy, and the collected data is stored in CSV format. The details of the dataset are outlined in Table 1, providing attributes, descriptions, and data types for the research information.

Table 1. Research Dataset Information

Attribute	Description	Type
<i>Date</i>	The date and time of the opinion tweet	<i>Date</i>
<i>Tweet</i>	Data to be classified (Contents)	<i>String</i>
<i>Category</i>	Tweet data categories	<i>Char</i>

This research involves four different categories, namely:

1. Price: 112 tweets.
2. Time: 2052 tweets
3. Returns: 224 tweets
4. Others: 137 tweets

To enhance the data analysis process, the label encoder is applied to convert the category from a character data type to an integer type. Table 2 is the encoder label output regarding data labeling based on categories.

Table 2. Encoder Label Information

Category	Label
Price	1
Time	2
Returns	3
Other	4

Data splitting involves dividing the dataset for a machine learning model. The training data is utilized to construct the model, while the validation data is employed to test each algorithm model created during training, ensuring its validation. Subsequently, testing data is used to assess the constructed model.

Text preprocessing of the data involves cleaning it to eliminate noise or interference, resulting in refined, clean data. This process relies on seven key indicators for text preprocessing:

1. Case Folding (Processing text data into all lowercase letters)

2. Remove Tweet Special (Remove tabs, new lines, emoticons, hashtags, links, hashtags)
3. Eliminate punctuation
4. Remove whitespace
5. Eliminate single characters
6. Use Indonesian stop words
7. Using literary stemming

The purpose behind employing these 7 indicators is to rectify abbreviated vocabulary, repeated letters, and incorrectly written words. The objective is to transform them into complete and appropriate words without altering their intended meanings, adhering to the KBBI Online guidelines. Additionally, the removal of mentions aims to eliminate sentence ambiguities that could potentially impact the sentence pattern relationship within the model.

For IndoBert implementation, the IndoBert-lite-base-p2 indobenchmark is utilized for both tokenization and modeling. This modeling process encompasses datasets both with and without preprocessing. The primary goal is to assess how input data affects the IndoBERT base model. Optimization of the model involved the use of Adam optimization, employing various learning rate optimization values as outlined in the subsequent table. The research incorporates 2 learning rate hyperparameters, detailed in table 3.

Table 3. Deep Neural Network Model Optimization Parameters

Optimization	Learning Rate
Adam	1×10^{-5}
	1×10^{-6}

This research employs the CNN algorithm, incorporating various forms of hyperparameter tuning, including learning rate, batch size, epoch, activation function, pooling, Conv1D, dense layers, and dropout.

Additionally, this research utilizes the Bi-LSTM algorithm, employing diverse hyperparameter tuning methods such as learning rate, batch size, epoch, activation function, pooling, bidirectional layers, dense layers, and dropout.

The evaluation encompasses four performance metrics: accuracy, precision, recall, and F1 score. It compares the performance between the Bi-LSTM algorithm and CNN after implementing IndoBert modeling, utilizing two different learning rates.

3. RESULT AND DISCUSSION

The data utilized in this research, whether subjected to text preprocessing or not, consists of four categories with different data types. From the collected data, 90% is allocated as training data, while the remaining 10% is designated as test data. This distribution is maintained for both the preprocessed and non-preprocessed data sets. Table 4 presents examples of text data subjected to text preprocessing and data that underwent no text preprocessing.

Table 4. Text Data Differences

Data Without Pre-Processing	Data With Pre-Processing
@J&T_express sortir di gudang sunternya manual kah? Sampe ngestuck seminggu? https://t.co/0n4bmk3brT	sortir di gudang sunternya manual kah sampe ngestuck seminggu
Halo @J&T_express tlong cek resi sya 10006930707622 in dr kebunem k jkrt emank lewat bantu ya https://t.co/50j2g5LouH	halo tlong cek resi sya in dr kebunem jkrt emank lewat bantu ya
@J&T_express Mending urusin paket gue yg muter2 di jaktim ga jelas sumpah,hampir 1 minggu muter2 aja kayak g tau jalan atau ga punya bensin	mending urusin paket gue yg muter di jaktim ga jelas sumpahhampir minggu muter aja kayak tau jalan atau ga punya bensin

ReLU activation function is applied in the hidden layer, while the softmax function is utilized in the output layer. The loss function employs sparse categorical cross-entropy due to the classification targets numbering more than two categories. A dropout layer of 0.2 is utilized within the hidden layers to regulate the model, thereby reducing overfitting. Each model was constructed with a batch value of 128 and an epoch value of 4. The IndoBERT base model, the Hugging Face indobenchmark, and the developed IndoNLU parameters are detailed in table 5.

Table 5. Hyperparameter Tuning

Parameter	Numbers
Batch Size	128
Learning Rate	1×10^{-5} dan 1×10^{-6}
Epoch	4
Loss	Sparse Categorical Cross-entropy
Optimizer	Adam
Activation	Softmax
Pool	Global Max Pooling
Bi-LSTM	32
Dropout	(0,2)
Dense	4
ConvID (CNN)	(32,5)

The deep neural network algorithm model is built with 2 layers which are explained in table 6.

Table 6. Research Testing

Data	Model Architecture	Learning Rate	Batch
• Data with Preprocessing	Bi-LSTM	1×10^{-5}	128
		1×10^{-6}	128
• Data without Preprocessing	CNN	1×10^{-5}	128
		1×10^{-6}	128

Table 7 and table 8 are the results of BI-LSTM research with data using text preprocessing and data without text preprocessing.

Table 7. Bi-LSTM Without Text Preprocessing

Learning Rate	Accuracy	Precision	Recall	F1
1×10^{-5}	79%	77%	79%	79%
1×10^{-6}	77%	76%	79%	79%

Table 8. Bi-LSTM With Text Preprocessing

Learning Rate	Accuracy	Precision	Recall	F1
1×10^{-5}	80%	80%	79%	79%
1×10^{-6}	79%	79%	79%	79%

From a comparison of data processed using the Bi-LSTM algorithm without or with text preprocessing, it was found that higher accuracy results were obtained from data processed first using text preprocessing with a learning rate of 1×10^{-5} .

Table 9 and table 10 are a comparison of the process with the CNN algorithm with or without text preprocessing first.

Table 9. CNN Without Text Preprocessing

Learning Rate	Akurasi	Presisi	Recall	F1
1×10^{-5}	80%	80%	80%	80%
1×10^{-6}	79%	79%	79%	79%

Table 10. CNN With Text Preprocessing

Learning Rate	Akurasi	Presisi	Recall	F1
1×10^{-5}	83%	83%	100%	91%
1×10^{-6}	82%	82%	100%	90%

From this second comparison, CNN with text preprocessing parameters 1×10^{-5} is 3% superior to the one without text preprocessing.

4. DISCUSSION

In the comparison of data processed using the CNN algorithm, the data subjected to text preprocessing showed higher accuracy compared to the data without preprocessing. The accuracy was observed at a learning rate of 1×10^{-5} , which outperformed the accuracy at a learning rate of 1×10^{-6} .

When comparing the two algorithms, IndoBert modeling with the CNN algorithm achieved better accuracy than the Bi-LSTM algorithm, with a 3% difference in accuracy. This research successfully utilized a language transition model integrated with a deep learning algorithm. In previous research, the analysis of delivery service opinions employed Bi-GRU, resulting in a 71% accuracy [23].

Bert Embedding, applied to this dataset, transforms complete input sentences into token embeddings. The first token of each sequence in the dataset is transformed into a special classification token. In the embeddings segment, when multiple sentences exist, they are combined into a single sequence and differentiated using special tokens. Additionally, positional encoding is utilized to preserve the word order for vectorizing the dataset.

Within the encoder transformer layer of Bert, the attentional head mechanism enables accurate connections between words in sentences within the dataset. This allows for an accurate understanding of the contextual semantic meaning of the same words in different contexts within a sentence.

5. CONCLUSION

The research has concluded that text preprocessing significantly influences the input dataset for Bert Embedding in the IndoBert Base model. Text preprocessing ensures that the dataset is

free from interference or noise, facilitating further data processing.

This study successfully tested two algorithmic model architectures, Bi-LSTM and CNN, achieving the highest accuracy of 83% with datasets subjected to text preprocessing using the CNN algorithm.

The tests revealed that Bi-LSTM with text preprocessing scored 79% accuracy at a learning rate of 1×10^{-6} , which is 2% higher than the score obtained without text preprocessing at the same learning rate, namely 77%. Additionally, CNN exhibited a 3% superiority with an accuracy of 83% compared to 80% without text preprocessing at a learning rate of 1×10^{-5} . The optimal results, obtained by CNN with parameters set at 1×10^{-5} , reached 83%, with the preprocessing technique considered superior to Bi-LSTM.

BIBLIOGRAPHY

- [1] I. R. M. Association, *Research Anthology on Social Media Advertising and Building Consumer Relationships*. IGI Global, 2022.
- [2] S. Smith, *Social Media Marketing for Business 2020*. Samuel Smith, 2021.
- [3] N. F. Rahmah, "The Influence of Service Quality on Consumer satisfaction Delivery Services at JNT Express in Garut," *bircu*, vol. 5, no. 3, p. 13, 2022, doi: <https://doi.org/10.33258/birci.v5i3.6628>.
- [4] F. Dwita, "THE EFFECT OF TOXIC LEADERSHIP AND JOB STRESS ON TURNOVER INTENTION IN LOGISTIC COURIER BEKASI CITY," *Airlangga J. Innov. Manag.*, vol. 3, no. 2, p. 10, 2022, doi: [10.20473/ajim.v3i1.39982](https://doi.org/10.20473/ajim.v3i1.39982).
- [5] A. George, *Python Text Mining*. BPB Publications, 2022.
- [6] J. Yang, "Brief introduction of medical database and data mining technology in big data era," *J. Evid. Based. Med.*, vol. 13, no. 1, pp. 57–69, 2020, doi: <https://doi.org/10.1111/jebm.12373>.
- [7] K. You, "HOW DOES LEARNING RATE DECAY HELP MODERN NEURAL NETWORKS?," *Mach. Learn.*, vol. 2, p. 15, 2019, doi: <https://doi.org/10.48550/arXiv.1908.01878>.
- [8] F. He, "Control Batch Size and Learning Rate to Generalize Well: Theoretical and Empirical Evidence," p. 4, 2019.
- [9] K. K. Chandriah, "RNN / LSTM with modified Adam optimizer in deep learning approach for automobile spare parts demand forecasting," *SpringerLink*, 2021.
- [10] A. Glassner, *Deep Learning: A Visual Approach*. Amerika Serikat: No Starch Press, 2021.
- [11] I. Kandel, "The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset," *ScienceDirect*, vol. 6, p. 4, 2020, doi: <https://doi.org/10.1016/j.icte.2020.04.010>.
- [12] N. Aneja, "Transfer Learning using CNN for Handwritten Devanagari Character Recognition," *IEEE*, 2019, doi: <https://doi.org/10.1109/ICAIT47043.2019.8987286>.
- [13] L. F. W. Anthony, "Carbontracker: Tracking and Predicting the Carbon Footprint of Training Deep Learning Models," *Comput. Soc.*, vol. 3, p. 11, 2020, doi: <https://doi.org/10.48550/arXiv.2007.03051>.
- [14] A. Labach, "Survey of Dropout Methods for Deep Neural Networks," *Neural Evol. Comput.*, vol. 2, p. 13, 2019, doi: <https://doi.org/10.48550/arXiv.1904.13310>.
- [15] B. A. Skourt, "Mixed-pooling-dropout for convolutional neural network regularization," *J. King Saud Univ. – Comput. Inf. Sci.*, vol. 34, no. 8, p. 6, 2022, doi: <https://doi.org/10.1016/j.jksuci.2021.05.001>.
- [16] S. Alaparathi, "Bidirectional Encoder Representations from Transformers (BERT): A sentiment analysis odyssey," *Comput. Lang.*, p. 15, 2020, doi: <https://doi.org/10.48550/arXiv.2007.01127>.
- [17] A. S. A. Ali, "CNN-Bi LSTM Neural Network for Simulating Groundwater Level," *Comput. Res. Prog. Appl. Sci. Eng.*, p. 7, 2022, doi: <https://doi.org/10.52547/crpase.8.1.2748>.
- [18] M. Pota, "An Effective BERT-Based Pipeline for Twitter Sentiment Analysis: A Case Study in Italian," *Sensors*, p. 21, 2020, doi: <https://doi.org/10.3390/s21010133>.
- [19] L. P. M. Aliyah Kurniasih, "On the Role of Text Preprocessing in BERT Embedding-based DNNs for Classifying Informal Texts," *Aliyah Kurniasih*, vol. 13, p. 8, 2022.
- [20] W. Lu, "A CNN-BiLSTM-AM method for stock price prediction," *Neural Comput. Appl.*, vol. 33, no. 4, pp. 4741–4753, 2020, doi: <https://doi.org/10.1007/s00521-020-05532-z>.
- [21] F. Theis, *Artificial Neural Networks and Machine Learning – ICANN 2019: Text and Time Series*. Springer International Publishing, 2019.
- [22] M. Tan, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," *Int. Conf. Mach. Learn. Proc. Mach. Learn. Res.*, vol. 97, pp. 6105–6114, 2019.
- [23] S. Z. Pranida, "Sentiment Analysis of

Expedition Customer Satisfaction using
BiGRU and BiLSTM,” *IJAIDM*, vol. 5, no. 1,
p. 10, 2022.