JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 7, Nomor 3, Juli 2023, Page 1297-1305
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v7i3.6438

Misogyny Text Detection on Tiktok Social Media in Indonesian Using the Pre-trained Language Model IndoBERTweet

Perwira Hanif Zakaria, Dade Nurjannah, Hani Nurrahmi*

School of Computing, Informatics Study Program, Telkom University, Bandung, Indonesia
Email: [1]perwirahanifz@student.telkomuniversity.ac.id, [2]dadenurjannah@telkomuniversity.ac.id,
[3,*]haninurrahmi@telkumuniversity.ac.id
Correspondence Author Email: haninurrahmi@telkumuniversity.ac.id

Abstract−Social media is a popular communication and information platform due to its ease and speed of access. By using social media, one can express himself freely. This triggers irresponsible individuals to utter hate speech with the aim of bringing down a person or group of people. Misogyny is a form of hate speech directed at women. The problem of misogyny should not be underestimated because misogyny can be one of the main reasons women feel miserable. In this study, a model will be built to detect misogyny text on the Indonesian language TikTok social media using the IndoBERTweet pre-trained model. IndoBERTweet is a pre-trained model based on the BERT model, which has been trained using Indonesian language datasets taken from the previous Twitter social media, resulting in a good performance for detecting misogynous texts on social media by classifying them. The dataset used is in the form of text data taken from misogyny comments by focusing on forms of misogyny in the form of stereotypes, dominance, sexual harassment, and discredit in short video content on women's TikTok social media accounts. The performance of built model performs hyperparameter settings which include batch size 16, epochs 10, and learning rate 7e-5 and is evaluated using a confusion matrix with the best accuracy results of 76.89%.

Keywords: Misogyny; BERT; Pre-trained Model; IndoBERTweet

1. INTRODUCTION

Social media is one of the most popular communication and information platforms today. The popularity of social media cannot be separated from the visuals, convenience, and speed of access, which are its main attractions. By using social media, a person can express himself freely without any restrictions. This freedom of expression triggers many irresponsible individuals to commit crimes in the form of hate speech online that aim to bully or bring down a person or group of people [1].

One of the social media that is becoming a trend in Indonesia is TikTok. According to Statista, a website that collects statistical data on social media in the world, recorded until July 2022, Indonesia ranks second in terms of the number of TikTok users, with a total of 99 million active users.

Hate speech is carried out online through social media regardless of gender. According to OXIS, an Oxford survey website, both a man and a woman can still be the target of online hate speech by irresponsible persons, but women are still more likely to receive hate speech when compared to men.

Misogyny is one form of negativity that is often shared or given on social media. Misogyny is a form of hate speech against women, whether it is directed at individuals or groups. Furthermore, misogyny is a problem that cannot be underestimated. This is because misogyny is the main reason women around the world feel miserable [1], [2]. Misogyny can be categorised into several forms of behaviour, such as Stereotype, Dominance, Sexual Harassment, and Discredit [3]. Misogyny behaviour is increasingly common along with the development of social media as a forum for expressing free and anonymous opinions, especially during the COVID-19 pandemic [4]. The continued increase in misogyny behaviour every year is a concern that should not be underestimated, and a solution must be sought to resolve it.

Misogyny can be categorized into several forms of behaviour such as humiliating, sexual harassment, domination, and discrediting. Shaming is a form of misogyny that restricts or belittles women because of some physical characteristic. Sexual harassment is a form of misogyny in the form of requests or statements to take actions that are sexually directed, such as sexual comments, crude jokes, and constant invitations to have extramarital affairs that can make you uncomfortable. Discredit and domination are forms of misogyny in the form of harsh expressions and/or statements that men are superior to women [1], [5], [6].

Misogyny text detection is a task that is done to detect whether a text or sentence is an expression of misogyny or not. In terms of text detection or recognition, the task of misogyny text detection has some comfort with sentiment analysis. In sentiment analysis, we detect whether a text has a positive or negative meaning towards a target label, while in misogyny text detection, we search whether a text has misogyny or not [3].

Research related to misogyny text detection has been done before. As in research [1], [3], [5], analysis of the problem of misogyny text detection was carried out using datasets from social media Twitter using 2 main tasks, where the first task aims to detect misogyny behaviour using 2 labels namely misogyny and non-misogyny, while the second task focuses more on detecting misogyny behaviour more specifically, such as incriminating, thwarting, discrediting, domination, sexual harassment, stereotyping & objectification, threats of violence, and so on. In the first study [1], the detection of misogyny texts was carried out using an Arabic language dataset taken from social media Twitter. The model used in this research is BERT using the pre-trained MARBERT model with the best performance results from the model being 91.74% accuracy on task 1 and 80.81% accuracy on task 2.

Submitted: **23/06/2023**; Accepted: **17/07/2023**; Published: **23/07/2023**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 7, Nomor 3, Juli 2023, Page 1297-1305
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v7i3.6438

Then in the second study [3], the detection on misogyny was carried out using English and Spanish datasets taken by social media Twitter using the Support Vector Machine or SVM model. The best performance results obtained in this study were 81.47% accuracy using the English dataset for task 1 and 54.22% F-target using the Spanish language dataset for task 2. In the third study [5], misogyny dedication was carried out using the dataset is in the form of a meme image embedded using GloVe to retrieve the text contained in the meme image. The main model used is BERT with the best performance results of 66.24% score for task 1 and 66.76% score for task 2.

In research [7], the problem of detecting misogyny was analysed using an Indonesian-language dataset taken from social media Twitter using only 2 labels, namely misogyny and non-misogyny. In this study, an experiment was conducted by comparing the effect of BERT Embedding on LR, CNN, and LSTM. The best performance results in this study were 86.15% accuracy and 81.37% F1-score obtained using the LSTM model.

The most visible shortcoming in previous studies is that most of the data used were in English and it was rare to find research using data sets in Indonesian. In addition, research on misogyny detection to classify using 2 labels, has an average performance of around 81%. However, this performance is better when compared to the performance of classification using more than 2 labels, which is equal to 67%. This is because the composition of certain labels in the dataset used indicates an imbalance condition [1], [3].

In previous studies, the Natural Language Processing or NLP approach and machine learning have been used to detect misogyny text on various types of social media [3], [8], [9]. The best performance result achieved in detecting misogyny text is 90%, which is achieved in detecting English content containing misogyny on social media Twitter using the BERTweet pre-trained method [9]. The BERTweet pre-trained model has been trained using datasets from social media Twitter with the limitation of only being trained using English datasets [8]. Similar to BERTweet, IndoBERTweet is a pre-trained model based on the BERT model that has been trained using datasets from Twitter in Indonesian [10]. Thus, using pre-trained IndoBERTweet can overcome the limitations of the pre-trained BERTweet model, which is only trained using English datasets.

The drawback of the previous misogyny text detection research is that most of the datasets used come from social media with English text [11] and Spanish [3]. In addition, research on the detection of misogyny on social media is mostly carried out using datasets from social media Twitter and Instagram.

This research will focus on detecting text misogyny on TikTok's Indonesian social media using the NLP approach and the BERT model. The BERT approach method is transfer learning by using the IndoBERTweet pre-trained model.

2. RESEARCH METHODOLOGY

2.1 General System

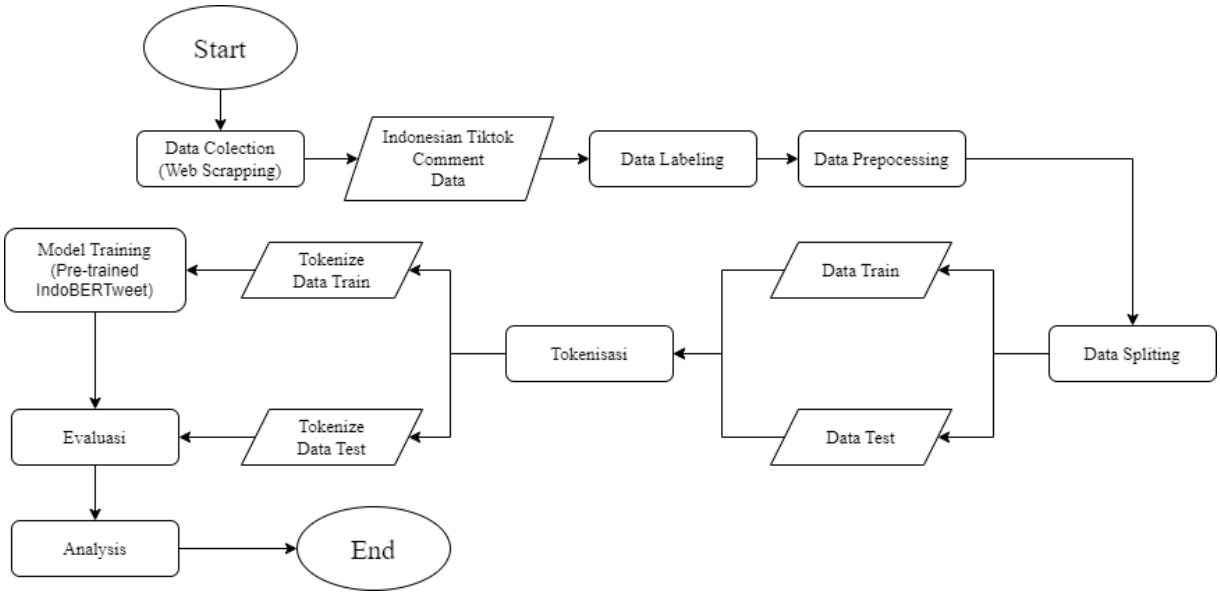

Figure 1. General System Design

In Figure 1, you can see the stages of developing a misogyny detection system starting from collecting misogyny data which focuses on forms of misogyny in the form of stereotypes, dominance, sexual harassment, and discredit on the Indonesian social media TikTok using the Web scraper method. The misogyny data collected is in the form of comments with the intention of insulting and/or dropping short video content on women's TikTok social media accounts. After collecting data, the process continues with data labelling, pre-processing, and data splitting. Data splitting is done by dividing the dataset into two parts, namely training data and test data. Before conducting training on the model, the tokenization process will be carried out first on the train data and test data,

Submitted: **23/06/2023**; Accepted: **17/07/2023**; Published: **23/07/2023**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 7, Nomor 3, Juli 2023, Page 1297-1305
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v7i3.6438

to adjust the shape of the data in such a way that it can be accepted by the BERT model. After training the model, the process continues with evaluation and analysis to draw conclusions from this research.

2.2 Data Collection (Web Scrapping)

The dataset used in this study is in the form of text data taken from comments with the intention of insulting or dropping on accounts on the TikTok social media belonging to women, which are limited to Indonesian as many as 1576 data. TikTok itself is a social media that is becoming a trend in Indonesia. Judging from the number of TikTok users in the world, Indonesia is ranked second in the world, with approximately 99 million active users.

Table 1. The Example of Data Collection Using Export Comments

Unique ID	Name	Likes	Comment
User_1	user1	7	pacarankah sma Thoriq?🙄kayaknya di TT Thoriq sma Fuji🙄🙏
User_2	user2	0	@khilma249 kpn"buat pake baju putih abu"😁
User_3	user3	164	orang2 kok sibuk bet sama pribadi orang!!😅
User_4	user4	6	ya allah masih sma kok kayak mama muda 😁
User_5	user5	26	semangat Ka Cikaaaaaa😈

The method used to collect data in this study is web scraping, using a tool in the form of the web named Export Comments, which can collect data from TikTok social media. Table 1 shows the results of the data collection using the web scrapper method, with the help of the Export Comments tool. Apart from using the web scraping method, data was collected by manually inputting it.

2.3 Data Labelling

The NLP approach can be used to detect text misogyny on social media by classifying text [12]. To solve text classification problems, the supervised learning method is used by giving targets or labels to the used dataset. In this study, data labelling was done by labelling "0" for text data that did not contain misogyny and "1" for text data containing misogyny. Data labelling was done by five people and then validated by a psychiatrist in Jakarta, Indonesia. The Example of data labelling can be seen in Table 2.

Table 2. Data Labelling Example

User	Comments	Label
@user1	pacarankah sma Thoriq?🙄kayaknya di TT Thoriq sma Fuji🙄🙏	1
@user2	@khilma249 kpn"buat pake baju putih abu"😁	1
@user3	orang2 kok sibuk bet sama pribadi orang!!😅	0
@user4	ya allah masih sma kok kayak mama muda 😁	1
@user5	semangat Ka Cikaaaaaa😈	0

2.4 Pre-Processing Data

The data obtained using the web scraper method may be inconsistent, incomplete, and/or unstructured. To overcome this problem, data pre-processing is carried out. Pre-processing data is a data mining technique whose purpose is to convert raw data into ready-to-use data [13]. In this study, the techniques used in the data pre-processing process include data cleaning, case folding, tokenisation, and normalization.

2.4.1 Data Cleaning

Data Cleaning is a data pre-processing technique, the purpose of which is to clean data by removing unnecessary characters or symbols, such as hashtags, emoticons, punctuation, usernames, and so on. An example of a comparison of sentences before and after using data-cleaning techniques can be seen in Table 3.

Table 3. Data Cleaning Example

Before Cleaned	After Cleaned
3emang kalau jadi cewe harus gitu ya?	emang kalau jadi cewe harus gitu ya
@pandabetina cape deh klo ga bisa tobat🤣🤣🤣	cape deh klo ga bisa tobat
Udah gila ni cwe #foryou	Udah gila ni cwe foryou

2.4.2 Case Folding

Case folding is a data pre-processing technique, which aims to change all text characters in a sentence in the dataset to lowercase [14], [15]. An example of a comparison of sentences before and after using the case folding technique can be seen in Table 4.

Submitted: **23/06/2023**; Accepted: **17/07/2023**; Published: **23/07/2023**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 7, Nomor 3, Juli 2023, Page 1297-1305
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v7i3.6438

Table 4. Case Folding Example

Before Case Folding	After Case Folding
Semangat terus Kakak	Semangat terus kakak
HARUSNYA SIH MALU YAA	Harusnya sih malu yaa
Udah gila ni cwe	udah gila ni cwe foryou

2.4.3 Tokenization

Tokenization is one of the data pre-processing techniques, which aims to cut or separate consecutive words in a sentence into several parts of words [14], [15]. An example of a comparison of sentences before and after using the tokenization technique can be seen in Table 5.

Table 5. Tokenization Example

Before Tokenization	After Tokenization
emang kalau jadi cewe harus gitu ya	[emang, kalau, jadi, cewe, harus, gitu, ya]
cape deh klo ga bisa tobat	[cape, deh, klo, ga, bisa, tobat]
harusnya si malu ya	[harusnya, si, malu, ya]

2.5 Data Splitting

After going through the data preprocessing process, data splitting is carried out by dividing the dataset into several parts and ratios [16]. In this study, data splitting was carried out by dividing the dataset into two parts, namely the train data and test data with a ratio of 8:2, which means 80% of the dataset is for train data and 20% of the dataset is for test data. The training data is used to provide training on the model being built, while the test data is used to evaluate the model being built.

2.6 BERT

BERT is the first model developed and introduced by the AI team at Google. The BERT model is designed to conduct training on unlabeled text in a dataset to give weight to the model, then fine-tuned the previously labelled text, according to research needs. This BERT model is bidirectional or commonly called bidirectional. Thus, this model can understand the correlation of a word with the surrounding words by reading the entire word order at once, which is useful for helping to understand the sentiment of a word in a sentence [5], [17], [18]. In previous research, a comparison of the performance of the BERT model with several traditional machine learning models such as KNN, Logistic Regression, Multilayer Perceptron, and Random Forest was carried out to solve the problem of misogyny identification [5].

Table 6. Score from Task A [5]

Model	Training	Validation	Score
KNN	78.05%	64.27%	57.54%
LR	73.95%	71.03%	58.69%
MLP	89.97%	71.32%	58.05%
RF	98.56%	61.97%	58.81%
BERT	**87.26%**	**82.10%**	**66.24%**

From this research, it was found that the BERT model has better performance when compared to the traditional machine learning model tested [5].

2.7 Fine Tuning

Fine Tuning is the ability of a deep learning model to transfer knowledge or experience from training results to solve different problems or tasks by adding new training to match the requirements of the task [19].

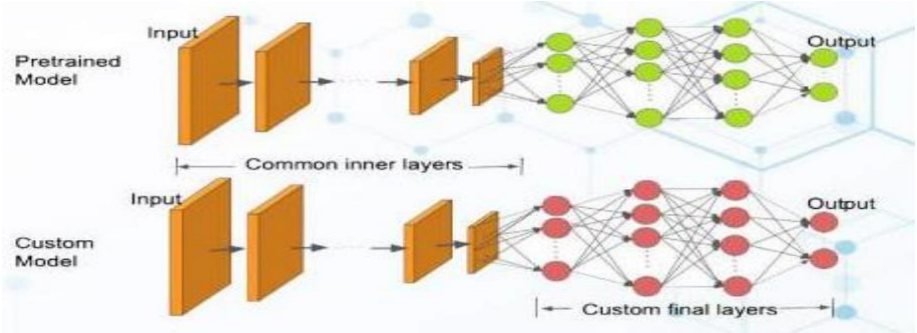

Figure 2. Fine Tuning Process [19]

Submitted: **23/06/2023**; Accepted: **17/07/2023**; Published: **23/07/2023**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 7, Nomor 3, Juli 2023, Page 1297-1305
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v7i3.6438

In Figure 2, you can see an overview of how fine-tuning works, where the pre-trained model will distribute layers if needed to the custom model. After receiving the layers from the pre-trained model, the custom model will train again by adding layers to get the desired output.

2.8 Pre-trained Model IndoBERTweet

Pre-trained models are models that have been trained, developed, and provided beforehand to solve new problems. In general, pre-trained models are used to solve deep learning-based problems and are trained using large datasets.

IndoBERTweet is a pre-trained model based on the BERT model that has been trained by IndoLEM by following the same procedure as the IndoBERT model. IndoBERTweet was trained using a dataset of 26 million Twitter comment data with 409millionword tokens, which means that the training data used to train the IndoBERTweet model is 2 times larger when compared to the training data used on IndoBERT [10], [20].

A study on public opinion on the COVID-19 vaccine in Indonesia uses the CNN-LSTM, IndoBERT and IndoBERTweet models. In this study, the IndoBERTweet pre-trained model obtained the best performance results compared to IndoBERT and CNN-LSTM [20].

IndoBERTweet uses a fine-tuning approach by calling the "AutoModelForSequenceClassification" function from the pre-trained model IndoBERTweet-based-uncased, one of the pre-trained models based on the BERT model trained and provided by IndoLEM [10], [21].

2.8.1 Import Library

To implement classification with IndoBERTweet, first import "AutoTokenizer" and "AutoModel" from the Transformers library provided by Huggingface. AutoTokenizer is used as a tool for tokenisation, while AutoModel is used as a tool for loading IndoBERTweet pre-trained models [10].

2.8.2 Dataset Preparation

Dataset preparation is carried out to adjust the form of the dataset to the form of input that can be accepted by the model that has been built. To adjust the shape of the dataset prior to IndoBERTweet modelling, tokenisation is performed. Tokenization is done by separating consecutive words in a sentence into several parts of words [14], [15]. After a complete sentence is separated into several word parts, the signification process is continued by giving tokens to each part of the word that was previously separated. Tokenization is carried out using the help of "AutoTokenizer", one of the tools for tokenizing specifically for the Indonesian BERT model [21].

2.8.3 Training Model

By using the fine-tuning approach, the model training process is not carried out from scratch, but instead carries out the process of transferring knowledge or experience from the model that has been previously trained to the model to be built. The model built will later be added with a little training so that it can detect misogyny text on TikTok social media with optimal performance. When carrying out the fine-tuning approach, it is necessary to adjust several hyperparameters, which include batch size, learning rate, and epoch. This is done to optimize the performance of the model being built [18].

2.9 Evaluation System

The confusion matrix is the most common method for visualizing the performance of an algorithm [22]. Each row in the confusion matrix represents the predicted results of the algorithm, while the columns represent the facts in the dataset used. The confusion matrix contains four parts, namely true positive (TP), true negative (TN), false positive (FP), and false negative (FN) [23].

Table 7. Confusion Matrix

		Fact	
		Positive	**Negative**
Prediction	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

In Table 7, you can see an example image of a confusion matrix diagram, with TP which means the prediction results and facts in the dataset used are positive, FP which means the prediction results are positive but the facts in the dataset used are negative, FN which means the prediction results are worth negative but the facts in the dataset used are positive, and TN which means the prediction results and the facts in the dataset used are negative [23].

To see the evaluation of the system more easily and clearly, calculations are made on the values of accuracy, precision, recall, and f1-score, which is a method for summarizing the results of the confusion matrix [24]. Accuracy is the result of calculating the total predicted value that is correct or in accordance with the facts divided by the total value of all data samples in the confusion matrix [24], as in the formula (1), accuracy can be

Submitted: **23/06/2023**; Accepted: **17/07/2023**; Published: **23/07/2023**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 7, Nomor 3, Juli 2023, Page 1297-1305
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v7i3.6438

obtained by adding up the values of True Positive and True Negative and then dividing by the result of the number of True Positive, True Negative, False Positive, and False negative.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

Precision is the result of calculating the total predicted value and positive facts divided by the total positive predicted value. When written mathematically, as in the formula (2), precision can be obtained by dividing the True Positive value by the result of True Positive plus False Positive.

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

Recall is the result of calculating the total positive value of predictions and facts divided by the total value of positive facts. When written mathematically, as in the formula (3), recall can be obtained by dividing the True Positive value by the result of True Positive plus False Negative.

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

F1-score is the result of calculating the mean value of the precision and recall values. When written mathematically, as in the formula (4).

$$F1\ score = 2\left(\frac{Precision \times Recall}{Precision+Recall}\right) \quad (4)$$

3. RESULT AND DISCUSSION

In this study, the detection of misogyny text on TikTok social media was carried out using pre-trained IndoBERTweet as a model for classifying misogyny text data. Misogyny text data was taken from 1.576 comments aimed at women's TikTok social media accounts and continued by labelling the dataset with the symbol "1" for text data that contains misogyny and "0" for the text that does not contain misogyny. There are a total of 1.141 data labelled as non-misogyny and 435 data labelled as misogyny, which indicates an imbalance in the dataset.

3.1 Pre-processing

Before the data is processed by the model, it is necessary to explore and prepare data or commonly known as data preprocessing. The purpose of this data preprocessing is to turn raw data into clean data and ready to be used or processed for processing. The dataset used in this study is text data taken from comments on social media taken through web scraping. Data obtained from comments on social media is usually unstructured data because it contains emoticons, HTML tags, links, and so on which makes it difficult for models to learn. Therefore, it is necessary to preprocess the data by processing the data so that it is more structured to make it easier for the model to understand the dataset in the learning process.

Table 8. Data Preprocessing Example

User	TikTok Comments	Pre-processing result
@user1	pacarankah sma Thoriq?🙄kayaknya di TT Thoriq sma Fuji🙄🙏	[pacarankah, sma, thoriq, kayaknya, di, tt, thoriq, sma, fuji]
@user2	@khilma249 kpn"buat pake baju putih abu"😁	[khilma, kpn, buat, pake, baju, putih, abu]
@user3	berat ya?semangat yaaa!!	[berat, ya, semangat, yaaa]
@user4	ya allah masih sma kok kayak mama muda 😁	[ya, allah, masih, sma, kok, kayak, mama, muda]
@user5	anak SMA jaman sekarang sudah mulai aktif yaa ...	[anak, sma, jaman, sakarang, sudah, mulai, aktif, yaa]

In this study, preprocessing was carried out using data cleaning, case folding, and tokenization techniques to clean the dataset from links, HTML tags, and symbols other than letters, such as punctuation marks, emoticons, usernames, and numbers which can be seen in Table 8. After the dataset is clean, case folding is carried out by changing all text characters to lowercase to generalize the form of data comments on the dataset to make it more consistent. The sentences in the comments on the dataset that have been consistent will then be broken down into several words to make it easier for the model to do the analysis.

3.2 Hyperparameter

To improve model performance, hyperparameter settings are made which include bach size, epochs, and learning rate. The batch size used in this study was 16, while the epochs and learning rate hyperparameters were determined by conducting experiments with the aim of getting the model performance in the form of the best accuracy.

Submitted: **23/06/2023**; Accepted: **17/07/2023**; Published: **23/07/2023**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 7, Nomor 3, Juli 2023, Page 1297-1305
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v7i3.6438

Experiments were carried out by setting the epochs values of 10 and 20, while the learning rate values were 2e-5, 7e-5, and 9e-4.

Table 9. Hyperparameter Test

		Accuracy	
		Epocs : 10	**Epocs : 20**
Learning Rate	2e-5	76.58	76.26
	7e-5	**76.89**	76.26
	9e-4	74.05	74.05

In Table 9, shows the details of the experimental results on the hyperparameters that have been carried out. In the first experiment, the epochs were set with a value of 10 and learning rate values were 2e-5 and produced an accuracy of 76.58. Experiments were continued by changing the learning rate value to 7e-5 and 9e-4 with the same epochs value setting and producing accuracy values of 76.89 and 74.05 respectively. The experiment was continued by changing the epochs setting to 20 with learning rate values of 2e-5, 7e-5, and 9e-4 and producing accuracy values of 76.26, 76.26, 74.05, respectively. The best results in this experiment were achieved with an epoch value of 10 and a learning rate of 7e-5.

3.3 Confusion Matrix and Error Analysis

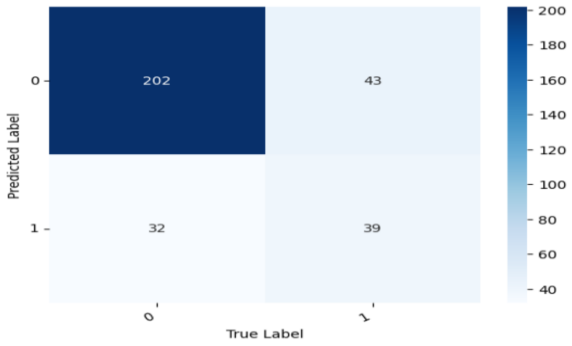

Figure 3. IndoBERTweet Confusion Matrix Diagram

Model evaluation was carried out using the confusion matrix method using hyperparameter settings according to previous experiments. The best result of the built model is an accuracy of 76.89%, with a true positive (TP) value of 202, a true negative (TN) of 39, a false positive (FP) of 43, and a false negative (FN) of 32, which can be seen in Figure 3. From the results obtained, it can be seen that the model can make predictions about texts that do not contain misogyny meaning better with details of 202 correct texts and 43 wrong texts out of a total of 245 texts tested, when compared to predictions of texts containing misogyny meanings which only succeed make correct predictions of 39 texts and incorrect predictions of 32 texts out of a total of 71 texts tested.

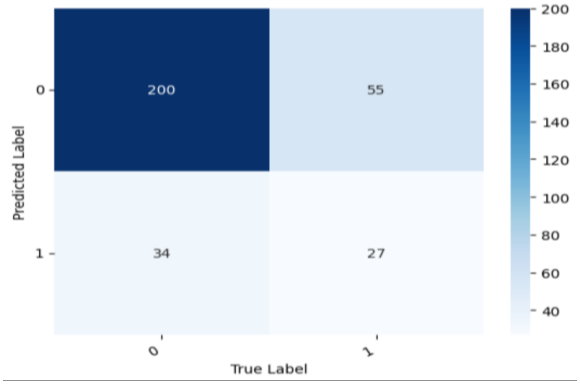

Figure 4. IndoBERT Confusion Matrix Diagram

In addition, researchers also used the IndoBERT model with the same hyperparameter settings to be compared with the tested IndoBERTweet model. The evaluation results for the IndoBERT model yielded true positives (TP) of 200, true negatives (TN) of 27, false positives (FP) of 55, and false negatives (FN) of 34, with a model accuracy of 71.8%, which can be seen in Figure 4.

In a study [20], research using the IndoBERTweet model has been carried out. In this study, the results of the accuracy of the IndoBERTweet model were 73%, as shown in Table 10. In this study, the accuracy results obtained using the IndoBERTweet model were slightly better at 76.89%.

Submitted: **23/06/2023**; Accepted: **17/07/2023**; Published: **23/07/2023**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 7, Nomor 3, Juli 2023, Page 1297-1305
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v7i3.6438

Table 10. Performance Model [20]

Model	Accuracy	F1-Score
IndoBERT	64%	68%
IndoBERTweet	73%	73%
CNN-LSTM	66%	61%

From the results obtained, the built model of difficulty in identifying misogyny texts. This is due to the large amount of misogyny text data which does not contain negative words but intends to satirize, insult, or harass which causes the model to be unable to learn more optimally. In the example sentences in Table 11, the words "keluar", "masuk", “asli”, “link” and "gesek" literally does not contain the meaning of misogyny, but these words mean misogyny with the intention of committing sexual harassment.

Table 11. Example of Ambiguous Text

Comments	Fact	Prediction
makasih kak aku dah keluar hehe	1	0
gisel di bayar brp om itu klau kali masuk	1	0
asli kah	1	0
info link	1	0
yg dibawah buat gesek kartu atm bukan si	1	0

4. CONCLUSION

This study aims to analyze the detection of text misogyny on TikTok social media using the NLP approach and the IndoBERTweet pre-training model with 1,576 data taken from text comments on women's TikTok social media accounts using a tool in the form of the Export Comments web. The IndoBERTweet model that was built can detect misogyny text by setting hyperparameters including batch size, epoch, and learning rate to improve model performance. The hyperparameter is determined by testing the epoch value and learning rate, while the batch size value has been determined at number 16. From the test results, the best accuracy value is obtained with the hyperparameter value which includes bach size 16, epochs 10, and learning rate 7e-5, with an accuracy of 76.89%. With these performance results, the IndoBERTweet model can detect misogyny text better when compared to the IndoBERT model which only gets an accuracy of 71.8%. Imbalance and the amount of ambiguous data cause the performance of the model still needs to be improved. The number of non-misogynous words that have misogynous meanings in a sentence causes the sentence to become ambiguous which in turn causes errors in the process of detecting misogynous texts. Suggestions from researchers for further research are to add data by paying more attention to the diversity or uniqueness of the data so that the dataset can be more balanced. It is hoped that this will cover the deficiencies in this research by optimizing and improving the performance of the built model.

REFERENCES

[1] A. El Mahdaouy, A. El Mekki, A. Oumar, H. Mousannif, and I. Berrada, “Deep Multi-Task Models for Misogyny Identification and Categorization on Arabic Social Media,” Jun. 2022, [Online]. Available: http://arxiv.org/abs/2206.08407
[2] M. E. Moloney and T. P. Love, “Assessing online misogyny: Perspectives from sociology and feminist media studies,” *Sociol Compass*, vol. 12, no. 5, May 2018, doi: 10.1111/soc4.12577.
[3] J. S. Canós, “Misogyny Identification Through SVM at IberEval 2018,” In *IberEval@SEPLN*, Sep. 2018, pp. 229-233.
[4] N. Dehingia, R. Lundgren, A. K. Dey, and A. Raj, “BIG DATA AND GENDER IN THE AGE OF COVID-19: A BRIEF SERIES FROM UC SAN DIEGO,” 2020. [Online]. Available: https://www.aljazeera.com/news/2020/5/5/bois-locker-room-indian-teen-probed-over-instagram-rape-chats
[5] G. Sharma, G. S. Gitte, S. Goyal, and R. Sharma, “IITR CodeBusters at SemEval-2022 Task 5: Misogyny Identification using Transformers,” In *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*, 2022, pp. 728-732, doi: 10.18653/v1/2022.semeval-1.100
[6] E. Fersini, P. Rosso, and M. Anzovino, “Overview of the Task on Automatic Misogyny Identification at IberEval 2018,” In *Proceedings of the third workshop on evaluation of human language technologies for Iberian Languages (IberEval 2018)*, 2018, pp. 1-15. CEUR.org
[7] R. S. Angeline, D. Nurjanah, and H. Nurrahmi, “Misogyny Speech Detection Using Long Short-Term Memory and BERT Embeddings,” in *2022 5th International Conference on Information and Communications Technology (ICOIACT)*, IEEE, Aug. 2022, pp. 155–159, doi: 10.1109/ICOIACT55506.2022.9972171.
[8] C. Graney-Ward, B. Issac, L. Ketsbaia, and S. M. Jacob, “Detection of Cyberbullying Through BERT and Weighted Ensemble of Classifiers,” 2021, doi: 10.36227/techrxiv.17705009.v1
[9] Y. Guo, X. Dong, M. A. Al-Garadi, A. Sarker, C. Paris, and D. Mollá-Aliod, “Benchmarking of Transformer-Based Pre-Trained Models on Social Media Text Classification Datasets,” 2020. [Online]. Available: https://nlp.stanford.edu/projects/glove/
[10] F. Koto, J. H. Lau, and T. Baldwin, “IndoBERTweet: A Pretrained Language Model for Indonesian Twitter with Effective Domain-Specific Vocabulary Initialization,” Sep. 2021, [Online]. Available: http://arxiv.org/abs/2109.04607

Submitted: **23/06/2023**; Accepted: **17/07/2023**; Published: **23/07/2023**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 7, Nomor 3, Juli 2023, Page 1297-1305
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v7i3.6438

[11] S. Frenda, B. Ghanem, M. Montes-Y-Gómez, and P. Rosso, "Online hate speech against women: Automatic identification of misogyny and sexism on twitter," *Journal of Intelligent and Fuzzy Systems*, vol. 36, no. 5, pp. 4743–4752, 2019, doi: 10.3233/JIFS-179023.
[12] Biere, Shanita, S. Bhulai, and M. B. Analytics. "Hate speech detection using natural language processing techniques." In *Master Business Analytics Department of Mathematics Faculty of Science*, Aug. 2018
[13] H. Kumar Sharma, K. Kshitiz, and Shailendra, "NLP and Machine Learning Techniques for Detecting Insulting Comments on Social Networking Platforms," in *2018 International Conference on Advances in Computing and Communication Engineering (ICACCE)*, 2018, pp. 265–272, doi: 10.1109/ICACCE.2018.8441728.
[14] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge University Press, 2008, doi: 10.1017/CBO9780511809071.
[15] U. Hasanah, T. Astuti, R. Wahyudi, Z. Rifai, and R. A. Pambudi, "An Experimental Study of Text Preprocessing Techniques for Automatic Short Answer Grading in Indonesian," in *2018 3rd International Conference on Information Technology, Information System and Electrical Engineering (ICITISEE)*, 2018, pp. 230–234, doi: 10.1109/ICITISEE.2018.8720957.
[16] A. Mahmood and J. L. Wang, "Machine learning for high performance organic solar cells: Current scenario and future prospects," *Energy and Environmental Science*, vol. 14, no. 1. Royal Society of Chemistry, pp. 90–105, Jan. 01, 2021, doi: 10.1039/d0ee02838j.
[17] S. González-Carvajal and E. C. Garrido-Merchán, "Comparing BERT against traditional machine learning text classification," May 2020, [Online]. Available: http://arxiv.org/abs/2005.13012
[18] J. Devlin, M.-W. Chang, K. Lee, K. T. Google, and A. I. Language, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," 2019, doi: 10.48550/arXiv:1810.04805
[19] A. Mohsin Abdulazeez, "Impact of Deep Learning on Transfer Learning : A Review," 2021, doi: 10.5281/zenodo.4559668.
[20] S. Saadah *et al.*, "Implementation of BERT, IndoBERT, and CNN-LSTM in Classifying Public Opinion about COVID-19 Vaccine in Indonesia," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 4, pp. 648–655, 2022, doi: 10.29207/resti.v6i4.4215
[21] M. Zidni Subarkah, M. Hildha, N. Tri Amanda, and E. Zukhronah, "Analisis Sentimen Review Tempat Wisata Pada Data Analisis Sentimen Review Tempat Wisata Pada Data Online Travel Agency Di Yogyakarta Menggunakan Model Neural Network IndoBERTweet Fine Tuning (Analysis of Sentiment Reviews of Tourist Attractions on Online Travel Agency Data in Yogyakarta Using the IndoBERTweet Fine Tuning Neural Network Model)."2022, doi: 10.34123/semnasoffstat.v2022i1.1246
[22] A. Luque, A. Carrasco, A. Martín, and A. de las Heras, "The impact of class imbalance in classification performance metrics based on the binary confusion matrix," *Pattern Recognit*, vol. 91, pp. 216–231, Jul. 2019, doi: 10.1016/j.patcog.2019.02.023.
[23] C. Das, A. K. Sahoo, and C. Pradhan, "Chapter 12 - Multicriteria recommender system using different approaches," in *Cognitive Big Data Intelligence with a Metaheuristic Approach*, S. Mishra, H. K. Tripathy, P. K. Mallick, A. K. Sangaiah, and G.-S. Chae, Eds., in Cognitive Data Science in Sustainable Computing. Academic Press, 2022, pp. 259–277, doi: https://doi.org/10.1016/B978-0-323-85117-6.00011-X.
[24] H. Yun, "Prediction model of algal blooms using logistic regression and confusion matrix," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 3, pp. 2407–2413, Jun. 2021, doi: 10.11591/ijece.v11i3.pp2407-2413.

Submitted: **23/06/2023**; Accepted: **17/07/2023**; Published: **23/07/2023**