Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)
Vol 5, No.1, March 2022, pp. 44 – 53
p-ISSN: 2614-3372 | e-ISSN: 2614-6150

Sentiment Analysis of Expedition Customer Satisfaction using BiGRU and BiLSTM

[1]Salsabila Zahirah Pranida, [2]Arrie Kurniawardhani
[1,2]Department of Informatics, Universitas Islam Indonesia
Email: [1]irasalsabila@gmail.com, [2]arrie.kurniawardhani@uii.ac.id

Article Info

Article history:
Received Jun 09th, 2022
Revised Jun 30th, 2022
Accepted Jul 15th, 2022

Keyword:
BiGRU
BiLSTM
Expedition Customer
Sentiment Analysis

ABSTRACT

The occurrence of a pandemic caused behavioral changes that occurred in Indonesian society, especially in increasing interest in online purchases. The increased purchases of goods increased the volume of four expeditions, namely: JNE, JNT Express, Sicepat, and Anteraja. To find out the customer satisfaction of the users of the four expeditions automatically, sentiment analysis was conducted based on the thousand tweet data from the opinions of expedition users in three-class categories, which are positive, negative, and neutral. Two deep learning methods were used to analyze the sentiment of expedition customer satisfaction: BiGRU and BiLSTM. The activities conducted during the sentiment analysis were crawling, preprocessing, data labeling, modeling, and evaluation. The performance evaluation results of the two methods use an accuracy matrix over 1,217 test data. The BiGRU method produces an accuracy performance of 71.5% and the BiLSTM method produces an accuracy performance of 66.5%.

Corresponding Author:
Arrie Kurniawardhani,
Department of Informatics,
Universitas Islam Indonesia,
Email: arrie.kurniawardhani@uii.ac.id
Jl. Kaliurang KM. 14,5 Sleman, Yogyakarta 55584.

DOI: http://dx.doi.org/10.24014/ijaidm.v5i1.17361

1. INTRODUCTION

In 2020, Indonesia was shocked by the corona virus pandemic which has not yet been completed. Enforcement of Restrictions on Community Activities which in Indonesia is called Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) promoted by the central government is also one of the programs that is able to suppress the spread of the corona virus [1]. The existence of a pandemic that has not ended and the PPKM program that continues, has made changes in behavior that occur in Indonesian society. One of the behavioral changes that have occurred is the increasing public interest in using e-commerce to buy goods online. According to a survey conducted by GWI, more than 87% of the people who have access to e-commerce, buy things online [2].

Purchasing goods online has also resulted in an increase in the volume of package delivery by several expeditions. Several expeditions provided by e-commerce include PT. Jalur Nugraha Ekakurir (JNE), Jet and Tony Express (JNT Express), Sicepat, and Anteraja. The increase in volume was felt by all four expeditions. 80% of JNE shipments are dominated by online shopping customers [3]. JNT Express noted that there were at least a total of 25 million packages delivered by the end of the year [4]. Sicepat informs that there are at least more than 2.8 million package shipments per day during 2021 [5]. Meanwhile, Anteraja experienced a significant increase of 550 thousand packages towards the end of 2021 [6].

However, the number of package deliveries made by the expedition does not necessarily indicate that the expedition is recommended for use. Indonesian Twitter users often upload tweets regarding complaints, appreciations, and questions to the customer service of the expedition they use. Some complaints such as break-ins, and goods that do not arrive, delays, often appear on several expeditions. Appreciation is also often given by users if the expedition meets their expectations, which is to arrive quickly and the goods are safe.

Sentiment analysis, commonly referred as opinion mining, is a method that focuses on opinions, then classifies the polarity of the text in sentences to determine a category, including such positive, negative, or neutral sentiment [7]. In sentiment analysis, crawling is done to get tweets that will be used to analyze customer satisfaction. Crawling tweets is a step that aims to collect data from the Twitter server database, in the form of users and all of their attributes [8].

There have been many studies conducted on sentiment analysis regarding JNT and JNE Express expeditions. Research on JNE and JNT Express expedition services conducted by [9] uses the Naive Bayes method. The study showed that the Naive Bayes method and the application of k-fold cross validation resulted in a score of 76% for the JNE expedition and 75% for the JNT Express expedition. The results show that the Naive Bayes algorithm works quite well in performing classification. And a similar study was also conducted by [10], who used Sicepat's data expedition to perform sentiment analysis using the SVM (Support Vector Machine) model. The results obtained on the performance of the accuracy matrix are quite good with an average value of 84%, the performance of the precision matrix of 84.28%, and the performance of the recall matrix of 83.57%.

In addition to using machine learning methods, there are several studies related to sentiment analysis using deep learning methods, such as BiLSTM and BiGRU. There have been many studies on sentiment analysis regarding the methods used in sentiment analysis, such as [11], [12], and [13]. Research on the Bidirectional GRU (BiGRU) method has been carried out by [11] which discusses aspect-level sentiment analysis on SIoT to increase IoT functionality by using opinions or behavior of social media users. The BiGRU architecture layer (forward GRU and backward GRU) is used to obtain vectors from the input, by making the best use of the information obtained in each word. The performance of the resulting model increased significantly.

A similar study using the BiGRU method was also conducted by [12] which was used on the SentiDrugs dataset to determine aspect level sentiment analysis in drug use reviews. The use of bidirectional neural networks such as BiGRU and BiLSTM is expected to be able to give better results. The use of the bidirectional method plays an important role in the retrieval of two-way semantic information and the improvement of classification performance. The results of the BiGRU model from experiments carried out on the accuracy matrix and macro-f1 showed that it was 71.35% and 69.88%, respectively.

Moreover, the Bidirectional LSTM (BiLSTM) algorithm was presented and was used to analyze the sentiments of opinions on a dataset of hotel reviews conducted by [13]. The input towards the BiLSTM model involves word representation, which allow the algorithm to analyze the information included. Because it considers information in forward and backward sequences, BiLSTM is known to be able to extract semantic features from a context more effectively and efficiently. This is evidenced by the results of the precision matrix of 91.54%, recall matrix of 92.82% and F1-score matrix of 92.18%.

This research was then aimed to determine the classification of customer satisfaction of expedition users using the deep learning method in conducting sentiment analysis based on tweets about JNE, JNT Express, Sicepat, and Anteraja expeditions. The importance of doing sentiment analysis on expedition customer satisfaction is to find out the sentiments of the expedition's customer opinions automatically from the thousands of opinions uploaded on Twitter. There are several processes carried out namely crawling, preprocessing, labeling, modeling, and evaluation. Crawling or data collection using the Twitter API. Then, preprocessing the data which consists of cleaning, case folding, stop words removal, and normalization. The labeling of tweets data is done manually using three classes, namely positive, negative, and neutral. The modeling is done using BiGRU and BiLSTM methods, the bidirectional function was chosen because it has the advantage of doing forward and backward direction which is useful in storing information in sequence.

2. RESEARCH METHOD

The research begins with data collection or crawling activities to create an expedition dataset that contains data containing the keywords 'jne', 'jnt', 'sicepat', and 'anteraja'. After the crawling activity is done, then preprocessing is carried out which has four stages, including case folding, filtering, normalization, and stop words removal. After the preprocessing stage has been carried out, manual labeling is carried out, which classifies the expedition data into three classes, positive, negative, and neutral. Furthermore, modeling is done using Bidirectional GRU and Bidirectional LSTM. After the modeling has been completed, the existing models are evaluated in the form of an accuracy matrix, a precision matrix, and a recall matrix. Figure 1. shows a flow chart regarding the research methodology to be conducted.

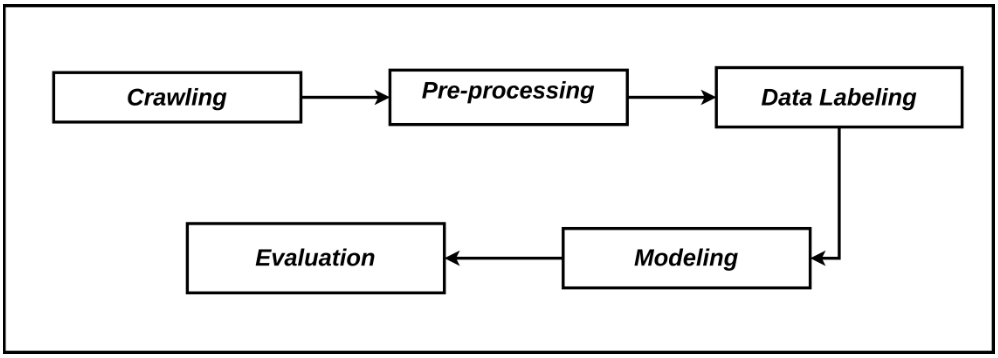

Figure 1. Research Method

2.1. Crawling

Crawling is a process or stage to obtain and collect data or information from the Twitter server database which contains tweets, users, and all their attributes. Crawling is carried out using the Twitter API which requires four keys to make API calls, including consumer_key, consumer_secret, access_token, and access_secret. In this research, only tweets from users that match the keywords will be used for labeling, while other attributes will be stored.

2.2. Preprocessing

Since the crawling results are usually untidy, the preprocessing phase is performed. One of the aspects that can improve the model's accuracy and processing time efficiency is the preprocessing procedure used [14]. At this stage, preprocessing is carried out, which include case folding, filtering, normalization, and stop words removal processes.

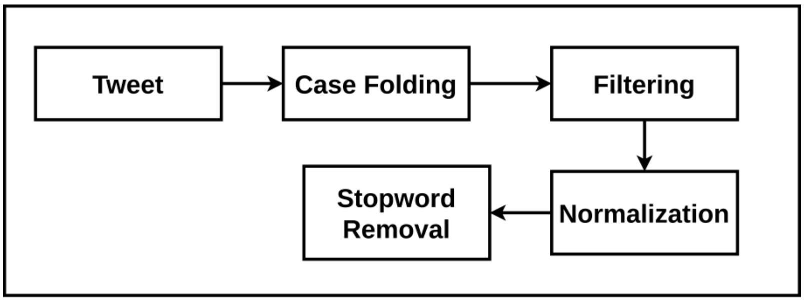

Figure 2. Flow of Preprocessing Stage

Figure 2 describes the flow that is performed in the preprocessing process. The preprocessing process begins with the expedition data as the input for the case folding method. Then, the output of the expedition data that has gone through case folding is used as an input to the next method, filtering. The expedition data which is the output of filtering then becomes the input for the next stage, namely normalization. In the normalization stage, the author uses a dictionary created by [15] and the author enriches the dictionary manually. The final stage after normalization is completed, the expedition data becomes the input for the stop words removal stage. Some words are removed in the stop words removal stage, such as *'ini'*, *'yang'*, *'adalah'*.

a) Case Folding
   Case Folding is a step taken to change all letters taken from the crawling process into lowercase or lowercase letters.
b) Filtering
   Filtering is a stage in preprocessing that is conducted to remove punctuation characters, which 14 such as period, question mark, exclamation point, comma, semicolon, colon, dash, hyphen, parentheses, brackets, braces, apostrophe, quotation marks, and ellipsis. Filtering is also done to remove usernames, URLs, and re-tweet (RT) in the expedition data.
c) Normalization
   Normalization is the stage to change non-standard or slang words into standard words. Normalization is done using an existing dictionary, by changing the words in the original column to

the replacement column. Table 1 shows examples of normalization dictionaries. For example, the slang word *'mengsedih'* is changed to *'sedih'* which is a standard word according to the KBBI.

Table 1. Normalization Dictionaries

Original	Replacement
ngejar	mengejar
krn	karena
mengsedih	sedih
pengennya	inginnya
ngga	tidak

d) Stop words Removal
Stop words removal is the step to remove familiar words such as *"ini", "ada", "adalah"*, and others, which are numerous and are considered meaningless. The list of words for stop words comes from Indonesian.

2.3. Data Labeling

Data labeling is a process of identifying a sentence or text into a predefined category or class. Model identification aims to provide context of information so that it can be studied by the model. Labeling is also frequently used in fields other than NLP, such as computer vision and speech recognition. After the data is cleaned and preprocessed, then the data is entered into labeling. Labeling is a step to assign one class label to the tweets data. There are three class labels, which are positive, negative, and neutral. Data labeling is done manually. The data that has been labeled will then be given a token for each word before being processed into a model.

2.4. Modeling

At the modeling stage, two methods were used, namely BiGRU (Bidirectional GRU) and BiLSTM (Bidirectional LSTM). In this stage, the distribution of training data and testing data is carried out with a ratio of 85:15 (training: testing). There is also data validation which has a ratio of 10:75 compared to the training data. Data validation is taken from the total amount of training data, which is then used to validate the data when the training is carried out.

a. *Bidirectional LSTM*

Long Short-Term Memory (LSTM) was created to complement traditional artificial neural networks which have the disadvantage of not being able to capture semantic information contained in sentences. LSTM is also intended to solve the vanishing gradient problem. LSTM is formed with the default behavior of learning information dependencies in the long term [16]. This causes it to be equipped with an input gate, output gate, forget gate and memory cell [13]. The basic components and calculations used in the LSTM are as follows:

$$f_t = \sigma_g(W_f X_t + U_f h_{t-1} + V_f C_{t-1} + b_f) \tag{1}$$

Forget gate or symbolized by f_t in equation (1) is a gate that determines the information that must be forgotten by the memory cell. There are two inputs of the forget gate, namely X_t and h_{t-1}. W_f is the connecting weight of X_t and forget gate. U_f is the connecting weight between h_{t-1} and the forget gate. C_{t-1} is the state of the last memory cell. V_f is the connecting weight of C_{t-1} and forget gate. b_f is a bias. σ_g is a sigmoid activation function.

$$i_t = \sigma_g(W_i X_t + U_i h_{t-1} + V_i C_{t-1} + b_i) \tag{2}$$

The input gate or symbolized by i_t in equation (2) is a gate that determines the information that must be updated by the memory cell. W_i is the connecting weight of X_t and the input gate. U_i is the connecting weight between h_{t-1} and the gate input. V_i is the connecting weight of C_{t-1} and the input gate. b_i represents bias.

$$C'_t = \sigma_c(W_c X_t + U_c h_{t-1} + V_c C_{t-1} + b_c) \tag{3}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot C'_t \tag{4}$$

The memory cell candidate gate or symbolized by C_t' in equation (3) is a gate containing candidate memory cell information. W_c is the connecting weight of X_t and the memory cell candidate gate. U_c is the weight of the link between h_{t-1} and the memory cell candidate gate. V_c is the connecting weight of C_{t-1} and input gate. f_t and it refers to the weights of C_{t-1} and C_t'. b_c is the bias. σ_c is an activation function of tanh.

$$o_t = \sigma_g(W_o X_t + U_o h_{t-1} + V_o C_{t-1} + b_o) \tag{5}$$

$$h_t = o_t \cdot \sigma_c(C_t) \tag{6}$$

The output gate or symbolized by o_t in equation (5) is a gate that issues a value from the LSTM. W_o is the connecting weight of X_t and the output gate. U_o is the connecting weight between h_{t-1} and the output gate. V_o is the connecting weight of C_{t-1} and the output gate. b_o represents bias. σ_g represents the sigmoid activation function. h_t represents the hidden state in equation (6).

Unfortunately, the LSTM architectural model only wins information in one direction, that is, forward. Therefore, in this study, it is proposed to use bidirectional so that the LSTM model can capture information in two directions, namely forward and backward. Bidirectional LSTM (BiLSTM) is also referred to as a stack LSTM, an improvisation of LSTM. The forward layer contained in the LSTM architecture is used to store context information afterwards. Meanwhile, the backward layer is used to store the previous context information [17]. An architectural illustration of BiLSTM can be seen in Figure 3. Figure 3 shows that each hidden layer component above and below is combined to form an architecture that is longer than the LSTM. BiLSTM which has a longer architectural framework also causes the processed information to be studied in more detail.

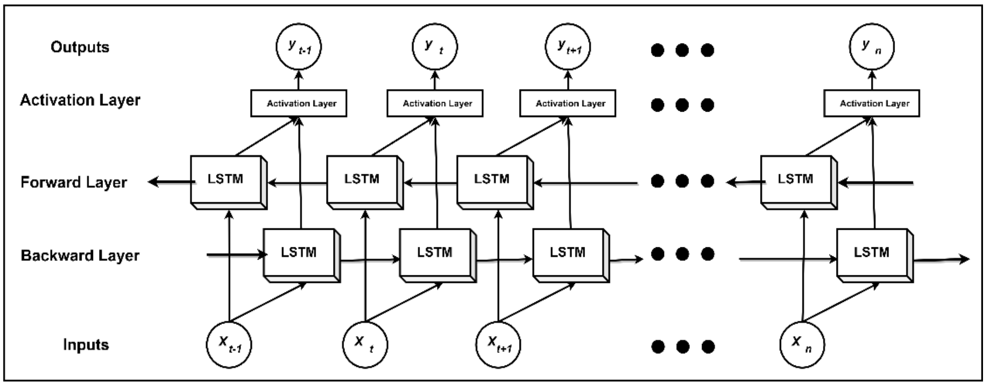

Figure 3. Bidirectional LSTM Architecture

b. *Bidirectional GRU*

Gated recurrent unit (GRU) is a method that is used for a method that is also used to handle short-term memory with two gates; reset gate and update gate [16]. The two gates in the GRU are used to control the vanishing gradient problem. GRU does not maintain cell state like LSTM, but retains hidden cells, which in turn makes the GRU model easier and faster to train small data sets. The basic components and calculations used in the GRU are as follows:

$$Z_t = \sigma(W_z \cdot [h_{t-1}, X_t]) \tag{7}$$

The update gate is denoted by Z_t in equation (7), which aims to eliminate the vanishing gradient problem. The problem can be eliminated because the model learns how much past information is passed on to the future. In the update gate, the variables X_t as inputs are multiplied by the weight of W_z. Similarly, to X_t, the variable h_{t-1} that stores information from the previous units is multiplied by the weight on W_z. Then, σ a sigmoid activation function is added to determine the output in the form of a value between 0 and 1.

$$R_t = \sigma(W_r \cdot [h_{t-1}, X_t]) \tag{8}$$

The reset gate, denoted by R_t in equation (8), has the opposite function from the update gate, because it determines how much past information should be forgotten. The reset gate formula is very similar to the update gate, which differs only in weight and function.

$$\hat{h}_t = \tanh(W \cdot [r_t * h_{t-1}, X_t]) \quad (9)$$

The memory cell is denoted by $\hat{h}_t$ in equation (9). The memory cell has an input of X_t, which is then multiplied by the weight of W. The application of element-wise multiplication to the reset gate and the previous h_{t-1} outputs allow for passing relevant past information. Both were added along with the implementation of the tanh activation function.

$$h_t = (1 - Z_t) * h_{t-1} + Z_t * \hat{h}_t \quad (10)$$

The final memory cell is denoted as h_t in equation (10). The h_t vector stores a lot of information on the unit and will be forwarded further down the network. Implementation of element-wise multiplication into update gate and memory cell. The element-wise multiplication implementation is also applied to $(1 - Z_t)$. The results of the implementation of the two are then added together.

GRU is considered suitable for sequential data, especially with large amounts of data. GRU can also capture data of the long sequence required for NLP learning [18]. The use of the bidirectional method, whose architecture is shown in Figure 4, plays an important role in the retrieval of semantic information that occurs in two directions in improving classification performance. Bidirectional method captures data sequentially with two directions, namely forward and backward. The forward layer in BiGRU captures sequential data that contains information in the next sequence. Meanwhile, the backward layer in BiGRU captures the sequential data in the previous sequence.

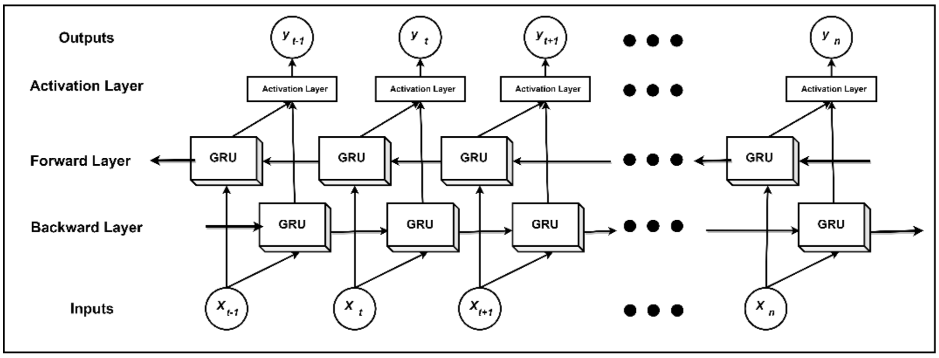

Figure 4. Bidirectional GRU Architecture

2.5. Evaluation

Several matrices were used to evaluate the model's performance, such as the accuracy matrix, recall matrix, and precision matrix. Evaluation generally uses a confusion matrix which consists of four variables, including True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). The accuracy matrix is defined in equation (11) as the ratio of the number of correct predictions compared to the total predictive data.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (11)$$

The precision matrix is defined in equation (12) as the ratio of the number of correct predictions that are positive compared to the total number of positive predicted prediction data.

$$Precision = \frac{TP}{TP+FP} \quad (12)$$

The recall or sensitivity matrix is defined in equation (13) as the ratio of the number of correctly positive predictions compared to the total number of correctly positive predictive data.

$$Recall = \frac{TP}{TP+FN} \quad (13)$$

3. RESULTS AND ANALYSIS

3.1. Crawling Result

The expedition dataset collected was 19.973 tweets data, which was taken via the Twitter API, based on four keywords including 'jne', 'anteraja', 'sicepat', and 'jnt'. Crawling was carried out on December 23, 2021 using the Python programming language. The results of the crawl contain information about Twitter's user_id, user profile photos, tweets content, geo location and so on. Several API calls were made, each of which a maximum of 100 tweets were taken containing the desired keywords. The author only takes the contents of tweets containing keywords and then stored in a Python data frame.

3.2. Preprocessing Result

The dataset obtained as many as 19.973 from the results of the crawling carried out still contains many things to be cleaned, such as punctuation, stop words, slang and non-standard words, up to hundreds of duplicated data. So, preprocessing is conducted which aims to clean the data so that the data can be easily processed at the next stage. Several methods used in preprocessing include case folding, filtering, normalization, and stop words removal. In addition to the four processes used, the author also cleans duplicated data. A total of 8.261 data from preprocessing are ready for labeling.

Table 2 shows the difference between tweets that have not been preprocessed and those that have been preprocessed.

Table 2. Case Folding

Process	Before	After
Case Folding	@sun_ny_____ @regularonce @jntexpressid pernah gini juga, ngga inget waktu itu pake jnt atau bukan, tulisan paket sudah diterima padahal belum terima sama sekali, tp besoknya langsung sampe rumah itu pakenya, dianter seperti biasa. pernah baca katanya sengaja digituin krn ngejar target(?) atau gimana gitu.	@sun_ny_____ @regularonce @jntexpressid pernah gini juga, ngga inget waktu itu pake jnt atau bukan, tulisan paket sudah diterima padahal belum terima sama sekali, tp besoknya langsung sampe rumah itu pakenya, dianter seperti biasa. pernah baca katanya sengaja digituin krn ngejar target(?) atau gimana gitu.
Filtering	@sun_ny_____ @regularonce @jntexpressid pernah gini juga, ngga inget waktu itu pake jnt atau bukan, tulisan paket sudah diterima padahal belum terima sama sekali, tp besoknya langsung sampe rumah itu pakenya, dianter seperti biasa. pernah baca katanya sengaja digituin krn ngejar target(?) atau gimana gitu.	pernah gini juga ngga inget waktu itu pake jnt atau bukan tulisan paket sudah diterima padahal belum terima sama sekali tp besoknya langsung sampe rumah itu pakenya dianter seperti biasa pernah baca katanya sengaja digituin krn ngejar target atau gimana gitu
Normalization	pernah gini juga ngga inget waktu itu pake jnt atau bukan tulisan paket sudah diterima padahal belum terima sama sekali tp besoknya langsung sampe rumah itu pakenya dianter seperti biasa pernah baca katanya sengaja digituin krn ngejar target atau gimana gitu	pernah begini juga tidak ingat waktu itu pakai jnt atau bukan tulisan paket sudah diterima padahal belum terima sama sekali tetapi besoknya langsung sampai rumah itu pakenya diantar seperti biasa pernah baca katanya sengaja dibegitukan karena mengejar target atau bagaimana begitu
Stopword Removal	pernah begini juga tidak ingat waktu itu pakai jnt atau bukan tulisan paket sudah diterima padahal belum terima sama sekali tetapi besoknya langsung sampai rumah itu pakenya diantar seperti biasa pernah baca katanya sengaja dibegitukan karena mengejar target atau bagaimana begitu	waktu pakai jnt tulisan paket diterima besoknya rumah pakenya diantar baca sengaja dibegitukan mengejar target
Pre-processing Final	@sun_ny_____ @regularonce @jntexpressid pernah gini juga, ngga inget waktu itu pake jnt atau bukan, tulisan paket sudah diterima padahal belum terima sama sekali, tp besoknya langsung sampe rumah itu pakenya, dianter seperti biasa. pernah baca katanya sengaja digituin krn ngejar target(?) atau gimana gitu.	waktu pakai jnt tulisan paket diterima besoknya rumah pakenya diantar baca sengaja dibegitukan mengejar target

3.2. Data Labeling Result

The expedition dataset obtained from the crawling results will be assigned a class label manually, which is positive, negative, or neutral. The data to be labeled is clean because it has been preprocessed. The labeling stage is carried out to identify or categorize the tweets data so that the model can study it. The distribution results of tweets data after labeling are as follows:

Table 3. Result of Data Labeling

Keywords	Positive	Neutral	Negative
JNE	340	385	175
JNT	475	1067	2911
SiCepat	772	594	561
Anteraja	596	156	228
Total	2183	2202	3875

A total of 8,261 tweets data that have been cleaned after preprocessing. Based on the distribution of the labeling data in Table 3, the JNT keywords have the most negative and neutral data, while the Sicepat keywords have the most positive data. Negative data is mostly filled with consumer dissatisfaction or expedition users regarding delivery delays, packages sitting in place, and insecurity of packages. While the positive data are mostly filled with friendly expedition couriers, fast delivery, and packages that arrive safely. Then the neutral data is mostly filled with questions with customer service and sentences that just stand alone or without context.

3.3. Evaluation Result

Table 4 shows the results of the evaluation carried out using three matrices, including the accuracy matrix, the precision matrix, and the recall matrix.

Table 4. Result of Evaluation

Methods	Matrix		
	Accuracy	Precision	Recall
BiGRU	0.715	0.832	0.698
BiLSTM	0.665	0.704	0.543

Based on the evaluation results in Table 4, the BiGRU method is capable of classifying with an accuracy value of 0.715, a precision value of 0.832, and a recall value of 0.698. Meanwhile, the BiLSTM method results in an accuracy value of 0.665, a precision value of 0.704, and a recall value of 0.543. The accuracy is not good enough because the training data used is not balanced and the labeling is given manually, so the data is subjective, and errors can occur which then have an impact on misclassification when testing the model.

The word *'gila'* which is usually labeled as negative, is often used by the Twitter community as an affix word. So, in manual data labeling, the word is also used as an affix for excited expressions as shown in sentence column both in Table 5 and Table 6.

Table 5. Result of BiGRU Experiments

Sentences	Actual	Prediction
Gilakk tumben dah sampenya lama banget	negative	negative
Gilakk!!! Tumben dah nyampe	positive	positive
Ongkir surabya jogja brp min	neutral	neutral
Duh!! Ekspedisinya ruwet banget gak sampe-sampe	negative	positive
Alhamdulillah kurir di tempatku Amanah sih	positive	negative
Halo min, tolong dong cek dm kuuu	neutral	neutral

In addition to using several evaluation matrices in Table 4, the author also tried the BiGRU and BiLSTM models with repetitive and ambiguous sentences. There are five sentences used in the experiment. In Table 5 shows that the BiGRU model can predict at least 4 out of 6 sentences correctly. Meanwhile, Table 6 shows that the BiLSTM model can predict at least 5 out of 6 correctly.

Table 6. Result of BiLSTM Experiments

Sentences	Actual	Prediction
Gilakk tumben dah sampenya lama banget	negative	negative
Gilakk!!! Tumben dah nyampe	positive	negative
Ongkir surabya jogja brp min	neutral	neutral
Duh!! Ekspedisinya ruwet banget gak sampe-sampe	negative	negative
Alhamdulillah kurir di tempatku Amanah sih	positive	positive
Halo min, tolong dong cek dm kuuu	neutral	neutral

4. CONCLUSION

The expedition dataset used in this study was 19,973 tweets data, which was taken from the results of crawling using the Twitter API with four keywords, namely 'jne', 'jnt', 'sicepat', 'anteraja'. A series of preprocessing was carried out which then resulted in a clean data of 8.261 tweets. The net data were then manually labeled, with 2.183 tweets data with positive sentiment, 2.203 tweets data with neutral sentiment, and 3.875 tweets data with negative sentiment. The research was carried out to achieve the goal of knowing the classification of customer satisfaction of expedition users in sentiment analysis on JNE, JNT Express, Sicepat, and Anteraja expeditions. The performance evaluation results obtained from the use of two deep learning methods, namely Bidirectional GRU and Bidirectional LSTM, in classifying sentiment analysis in the evaluation are the performance accuracy matrix from the use of the BiGRU method is 71.5% and the accuracy matrix performance from the use of the BilSTM method is 66.5%. The accuracy performance is quite low due to the influence of unbalanced data and subjective data labeling because the labeling is done manually, which can lead to errors and misclassification when labeling.

REFERENCES

[1] A. D. Darmawan, “PPKM Darurat Mampu Tekan Penyebaran Covid-19 di Jakarta,” Jul. 21, 2021. https://databoks.katadata.co.id/datapublish/2021/07/21/ppkm-darurat-mampu-tekan-penyebaran-covid-19-di-jakarta (accessed Apr. 14, 2022).

[2] S. Kemp, “Digital 2021: the latest insights into the ‘state of digital’ - We Are Social UK,” Jan. 27, 2021. https://wearesocial.com/uk/blog/2021/01/digital-2021-the-latest-insights-into-the-state-of-digital/ (accessed Apr. 14, 2022).

[3] R. Yati, “80 Persen Pengiriman JNE Didominasi Pelanggan e-Commerce,” Jun. 12, 2021. https://ekonomi.bisnis.com/read/20210613/98/1404786/80-persen-pengiriman-jne-didominasi-pelanggan-e-commerce (accessed Apr. 14, 2022).

[4] Tribunnews, “J&T Express Catat Pengiriman Tertinggi Pada Harbolnas 12.12, Total Hingga 25 Juta Paket,” Dec. 16, 2021. https://www.tribunnews.com/bisnis/2021/12/16/jt-express-catat-pengiriman-tertinggi-pada-harbolnas-1212-total-hingga-25-juta-paket (accessed Apr. 14, 2022).

[5] E. Catriana, “SiCepat Ekspres Catatkan Pengiriman Paket Lebih dari 2,8 Juta Per Hari Selama 2021,” Dec. 21, 2021. https://money.kompas.com/read/2021/12/21/153000126/sicepat-ekspres-catatkan-pengiriman-paket-lebih-dari-28-juta-per-hari-selama (accessed Apr. 14, 2022).

[6] K. Safitri, “Jelang Akhir Tahun, Pengiriman Harian Anteraja Naik Signifikan,” 2021. https://money.kompas.com/read/2020/12/30/161845026/jelang-akhir-tahun-pengiriman-harian-anteraja-naik-signifikan (accessed Mar. 10, 2022).

[7] S. Samsir, A. Ambiyar, U. Verawardina, F. Edi, and R. Watrianthos, “Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes,” *stmik-budidarma.ac.id*, vol. 5, no. 1, pp. 157–163, 2021, doi: 10.30865/mib.v5i1.2604.

[8] D. Ramadhan and E. B. Setiawan, “Analisis Sentimen Program Acara di SCTV pada Twitter Menggunakan Metode Naive Bayes dan Support Vector Machine,” *... .telkomuniversity.ac.id*, vol. 6, no. 2, pp. 9736–9743, 2019, [Online]. Available: https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/10708

[9] N. Ika, P. Kalingara, O. N. Pratiwi, and H. D. Anggana, “Analisis Sentimen Review Customer Terhadap Layanan Ekspedisi Jne Dan J & T Express Menggunakan Metode Naïve Bayes Sentiment Analysis Review Customer of Jne and J & T Express Expedition Services Using Naïve Bayes Method,” vol. 8, no. 5, pp. 9035–9048, 2021.

[10] F. Prasetiawan, S. Widiyanesti, and T. Widarmanti, “Analisis Sentimen Mengenai Kualitas Layanan Jasa Ekspedisi Barang Sicepat Di Media Sosial Twitter Sentiment Analysis Regarding Quality of Sicepat Expedition Services On Twitter Social Media,” vol. 9, no. 2, pp. 147–160, 2022.

[11] W. Ali, Y. Yang, X. Qiu, Y. Ke, and Y. Wang, “Aspect-Level Sentiment Analysis Based on Bidirectional-GRU in SIoT,” *IEEE Access*, vol. 9, pp. 69938–69950, 2021, doi: 10.1109/ACCESS.2021.3078114.

[12] Y. Han, M. Liu, and W. Jing, “Aspect-Level Drug Reviews Sentiment Analysis Based on Double BiGRU and Knowledge Transfer,” *IEEE Access*, vol. 8, pp. 21314–21325, 2020, doi: 10.1109/ACCESS.2020.2969473.

[13] G. Xu, Y. Meng, X. Qiu, Z. Yu, and X. Wu, “Sentiment analysis of comment texts based on BiLSTM,” *IEEE Access*, vol. 7, pp. 51522–51532, 2019, doi: 10.1109/ACCESS.2019.2909919.

[14] P. Chandrasekar and K. Qian, “The Impact of Data Preprocessing on the Performance of a Naïve Bayes Classifier,” *Proc. - Int. Comput. Softw. Appl. Conf.*, vol. 2, pp. 618–619, Aug. 2016, doi: 10.1109/COMPSAC.2016.205.

[15] Ramaprakoso, “Analisis Sentimen,” 2019. https://github.com/ramaprakoso/analisis-sentimen (accessed Mar. 25, 2022).

[16] M. R. Raza, W. Hussain, and J. M. Merigo, “Cloud Sentiment Accuracy Comparison using RNN, LSTM and GRU,” *Proc. - 2021 Innov. Intell. Syst. Appl. Conf. ASYU 2021*, 2021, doi: 10.1109/ASYU52992.2021.9599044.

[17] G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami, and C. Dyer, “Neural Architectures for Named Entity Recognition”, Accessed: Mar. 21, 2022. [Online]. Available: https://arxiv.org/pdf/1603.01360.pdf

[18] M. Zulqarnain, R. Ghazali, M. G. Ghouse, and M. F. Mushtaq, “Efficient processing of GRU based on word embedding for text classification,” *Int. J. Informatics Vis.*, vol. 3, no. 4, pp. 377–383, 2019, doi: 10.30630/JOIV.3.4.289.

BIBLIOGRAPHY OF AUTHORS

Salsabila Zahirah Pranida, a student from the Department Informatics S1 Program, Univeritas Islam Indonesia. Research interest in NLP and Sentiment Analysis

Arrie Kurniawardhani, a lecturer from Department Informatics, Universitas Islam Indonesia. Research interest in Computer Vision, Data Science, and Machine Learning.