Sentiment Analysis Using Recurrent Neural Network (Rnn) Method With Long Short Term Memory (Lstm) On Traveloka Application Comment Review

Husen Yarbo[1*], Rais Rais[2], Iman Setiawan[3]

[1,2,3] Statistic Study Program, Tadulako University, Palu, Indonesia

husenyarbo56@gmail.com

Abstract. In the industrial era 5.0, everything can be done online. The same applies to travel, whether booking public transportation tickets for a vacation, or booking a hotel room at the desired destination. One example of such an application is Traveloka on the Google Play Store for android users. Long Short Term Memory (LSTM) is one of the popular forms of recurrent neural networks (RNN) specifically designed to solve long-term dependency problems and is particularly suitable for time series processing and prediction. the highest frequency of words in this study is in the word 'disappointed' as many as 663, then the frequency of the word 'easy' with the number 529, the word 'buy' with the number of ferquencies "462", on the word 'fast' with the number "320", and on the number of words 'good' with the number "240" words. Sentiment analysis on Traveloka application comments using the Long Short Term Memory method on the 80:20 training and testing division has an accuracy of 83% correctly.

Keywords: Sentiment, LSTM, Traveloka, Classification, Word Cloud

1 Introduction

Today, in the era of industry 5.0, everything has moved online. Whether it's shopping, studying, making payments, or ordering food, everything can be done online. All of these activities can be accessed from home easily using a smartphone and an internet connection. One example of such an application is traveloka which can be downloaded through the Google Play Store for android users. This application is used to order public transportation tickets such as train tickets or plane tickets [1].

The traveloka application has provided various kinds of convenience and comfort to its users, but not all users are fully satisfied with its services. To accommodate this, the app provides a comment section on the Google Play Store so that users can provide reviews and ratings of their experience with the app. However, there are currently some users who give the highest rating but express negative reviews. Therefore, to make it easier for related parties to get information about the strengths and weaknesses of their

Y. Yuyun et al. (eds.), *Proceedings of the 5th International Seminar on Science and Technology (ISST 2023)*, Advances in Physics Research 10,
https://doi.org/10.2991/978-94-6463-520-1_12

application, sentiment analysis is carried out to classify review data from Traveloka App users based on the content of the reviews [2]. Recurrent Neural Network Method with Long Short Term Memory (LSTM) This is one of the Deep Learning models that can be used to perform sentiment classification. This technique has the ability to process data sequentially, including text, voice, and video [3]

Several studies have been conducted on sentiment analysis using Recurrent Neural Network with Long Short Term Memory (LSTM). One of them was conducted by [3], who used RNN with LSTM to analyze the sentiment of Instagram data. The results of this study produced a sentiment classification system with a test accuracy rate of 65% and an application accuracy rate of 79.46%.

Another study by [4] also used the LSTM method to analyze sentiment on reviews on TripAdvisor. The purpose of this study was to classify visitor reviews about the influence of COVID-19 on tourist attractions in Bali from TripAdvisor. The test results show an accuracy value of 71.67%.

In research by [5], sentiment analysis was carried out on Movie Review using Word2vec and the LSTM Deep Learning method. The dataset used contains 25,000 review documents with an average length per review of around 233 words. The dimensional size test of the word vector was carried out in variations of 50, 60, 100, 150, 200, and 500 dimensions to determine its effect on accuracy. The results show that the best accuracy is obtained at a word vector dimension of 100 with a value of 88.17%, and the lowest accuracy is obtained at a word vector dimension of 500 with a value of 85.86%.

2 Material and Methods

2.1 Traveloka Application

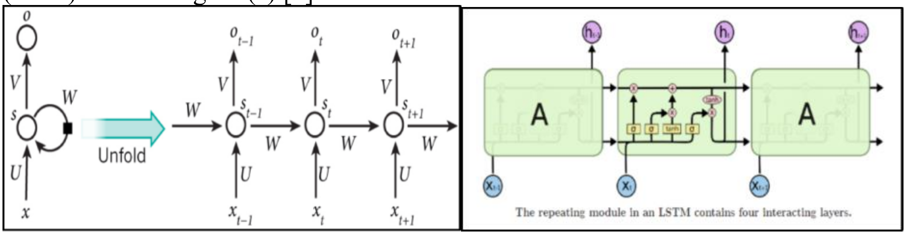
The company Traveloka focuses on online booking services for flight, train, bus, travel, and hotel reservations. The Traveloka app was founded by Ferry Unardi, Derianto Kusuma, and Albert Zhang as an initial concept to compare prices. The app was then launched on July 31, 2014 with the main goal of saving time in the ticket booking process [6]

2.2 Artificial Neural Network

Artificial Neural Network, also known as Neural Network, is a system that mimics biological neural networks by using computational methods to process information. The main principle is to mimic the workings of the human brain which has the ability to process in parallel, is able to process a large number of elements, and has a tolerance for errors [7].

Recurrent Neural Network (RNN) is a type of artificial neural network that has the ability to identify hidden correlations in data. Its main applications are in speech recognition, natural language processing, and time series prediction. RNN are very effective in modeling data sequences as they are able to operate on input information as well as trace information from previous iterations through recurrent connections [8].

The following figure is the working architecture of the Recurrent Neural Network (RNN) method Figure (a) [9] :



(a)

(b)

Figure 1. Recurrent Neural Network Architecture (a) and Long Short Term Memory Network (b).

Long Short Term Memory (LSTM) is one of the popular forms of recurrent neural networks (RNN) specifically designed to solve long-term dependency problems and is particularly suitable for time series processing and prediction. The LSTM model consists of a series of single memory cells that replace the neurons in the hidden layer of an RNN, and the main component of the LSTM is the state of those memory cells [10].

2.3 Sentiment Analysis

Sentiment analysis or opinion mining is a part of data mining that is specifically used to analyze text data in the form of opinions that contain polarity, with the aim of producing information that has a positive or negative value [11]. The main task in sentiment analysis is to categorize the text in a particular sentence or document and then determine whether the opinions expressed in the text are positive or negative. In practice, sentiment analysis can be used to search for opinions about products, brands, or people, and determine whether those opinions tend to be positive or negative in nature [12]. Text Preprocessing is a process that aims to convert unstructured text data into structured text data [13]. Preprocessing has several stages

1. Remove punctuation: This stage aims to remove all non-alphabetic characters such as symbols, spaces, etc. from the text.
2. Case folding: Case folding is the process of converting all letters in a document into the same form, for example converting the entire document into lowercase letters or vice versa. In addition, non-letter characters are also removed.
3. Stemming: This process serves to return the words that have been generated from the filtering stage to their basic form. Initial affixes and final affixes are removed so that the base word is obtained.
4. Tokenizing: This process separates the text into pieces of sentences and words called tokens. The purpose of this stage is to get chunks of words that will become entities that have values in the matrix of text documents to be analyzed.
5. Bigrams and Trigrams: Bigrams and trigrams are part of the N-Gram model. N-Grams are representations of items in the form of letters, characters, or consecutive words that serve to predict the next word in a given word sequence.

Spelling Normalization aims to avoid the expansion of the number of word dimensions that occurs if words with incorrect spelling or abbreviations are not corrected, as this will cause them to be counted as different word entities in the matrix construction process.
A word cloud is a visual display of a set of words arranged in such a way that each word is presented in varying sizes depending on how much it appears in a given text or document. The more often a word appears in the text or document, the larger the font size of the word in the word cloud. Word clouds are used to visualize key words or topics that appear most frequently in a text or document, and can help facilitate understanding of the themes or topics discussed in the document [14]

2.4 Confusion Matrix

Confusion matrix is a table used to evaluate the performance of a classification system or prediction model. Confusion matrix shows the amount of data that is classified correctly or incorrectly in various categories or classes. Confusion matrix generally has four main cells or sections, namely true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). In a binary classification system (with only two classes), the confusion matrix will consist of two rows and two columns [19]. Confusion matrix can be used to calculate various performance evaluation metrics such as accuracy, precision, recall (sensitivity), F1 Score, and others. Through the confusion matrix, we can understand more about the performance of a classification system or prediction model, and improve performance more effectively and efficiently

3 Result and Discussion

3.1 Web Scraping

This research uses data obtained from the Google Play Store which is taken from comments given by users in the Traveloka application on the Google Play Store. To get comments on the comments column in the Traveloka application on the Google Play Store, the author scrapes using Python where the data taken starts from March 2, 2020 until the revocation of the Covid-19 emergency status by the World Health Organization on May 5, 2023. These comments will be the dataset used to analyze sentiment towards the Traveloka application. Existing comments were taken using the google play scraper library with the help of google colab. A total of 3,308 comments were successfully taken from the application review column which will then be preprocessed so that the amount of user data decreases to 3,238 comments because unnecessary sentences have been removed through the preprocessing stage.
Labeling is done manually by subjectively determining whether the comment belongs to a positive or negative class, if the comment given has a good impression such as praising or asking questions and providing input, it will be classified into the positive class. Conversely, if the sentiment given has a bad impression such as berating and sarcasm, it will be classified into the negative class. Ratings given by users also make it easier to determine classes with ratings 1, 2, and 3 will be categorized as negative

and ratings 4 and 5 will be given as positive ratings. The following is the amount of positive and negative data displayed in the bar chart image in the figure

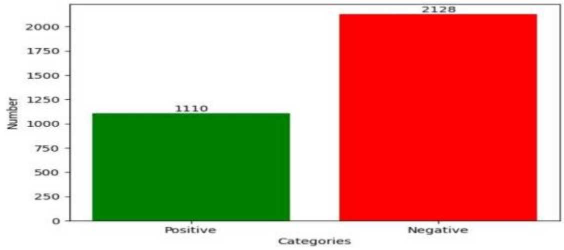


Fig. 2: Number of Comments for each Class

The figure above shows that there are 1.110 comments that fall into the positive class and 2.128 comments that fall into the negative class with a total of 3.238 data. The comparison of the number of positive and negative comments shows that there are more negative class comments than positive class comments. So it can be concluded that users of the traveloka application tend to give negative comments compared to positive comments.

3.2 Long Short Term Memory Classification

This research uses the Long Short Term Memory (LSTM) algorithm as a classification algorithm for sentiment analysis of traveloka application comments on the play store. The application of the LSTM algorithm uses a hard module module with Sequential in python. The parameters used in determining the model to be formed are as follows:

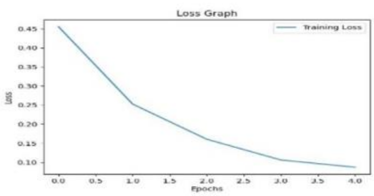
1. Hidden Layer : 4
2. Neuron Hidden : 50
3. Batch Size : 32
4. Drop Out : 0,2
5. Epoch : 5
6. Optimizer : Adam
7. Activation Function : Tanh and sigmoid

The parameters refer to previous research namely [20] and the development of the author by changing the number of batch sizes and the number of epochs. Determination of parameters there are no definite rules regarding the model formed, so it can adjust the needs of the research. The model formed is influenced by the amount of data input and the parameters formed. The parameterized sequetial model can be seen in figure (a) below:

Model : "sequential"

Layer (type)	*Output* Shape	Param
embedding (Embedding)	(None, 36, 3238)	16190000
spatial_dropout1d (SpatialDropout1D)	(None 36, 3238)	0
lstm (LSTM)	(None, 50)	657800
dense (Dense)	(None, 2)	102

Total patams: 16,847,902
Trainable: 16,847,902
Non-trainable params: 0



(a) (b)

Fig. 3: Word cloud of traveloka app comments (a) and Bar chart of 5 most frequent words (b)

In fig (a) above, we can see that the model consists of 4 hidden layers. In the first layer, the input in the form of 2 dimensions is formed from samples (many rows of data) time step and features. In the SpatialDropout1D layer, there are no parameters because there is no weight value in the layer. Then, in the third layer, there are 50 neurons that produce a total of 657,800 parameters. Finally, in the dense layer, there are 2 output values, namely the target value (y) for positive results and the value (y) for negative results. The total number of parameters in this model is 16,847,902 weights and biases. Figure (b) graph of the loss model formed in this study. The LSTM architecture used can be seen in Figure below

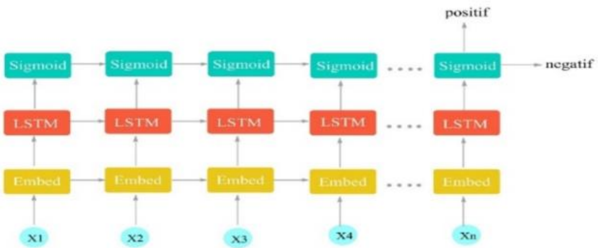


Fig. 4: Arsiterkur Long Short-Term Memory

In the fig above, it can be seen that the input layer is variable X where the number of input vectors is up to Xn or with a total amount of data, namely 3.238 data. While the second hidden layer is embeding which functions to display data in 2-dimensional form. Then in the third hidden layer is a neuron model of LSTM of 50 neurons which means the model can predict 50 long-term patterns from the given data sequence. In the fourth hidden layer there is a sigmoid activation function that functions to predict the value of the cell state generated in the LSTM model with a sigmoid activation function so that it can determine the resulting value will enter the negative class if $0 \leq \sigma \leq 0.2$ and enter the positive class if $0.2 \leq \sigma \leq 1$.

3.3 Evaluation of Results

The result evaluation stage is carried out by looking at the confusion matrix from the results of sharing the best testing and training data based on the highest accuracy is with a ratio of 80: 20 by using the model that has been generated. This model was formed with 2,580 training data and tested with 648 testing data. The confusion matrix results can be seen in the table below

TABLE 10. Confusion matrix result

		Actual Class	
		Positive	Negative
Actual Class	Positive	398	41
	Negative	70	139

observations that are correctly predicted as positive classes and 139 negative observations that are correctly predicted as Negative classes. In addition, classification errors occur where there are 70 observations that belong to the negative class but are predicted to enter the Positive class and 41 observations that belong to the Positive class but are predicted to enter the Negative class.

Based on the results above, it can be seen that the classification managed to get the best Accuracy based on three tests that have been carried out of 0.829 or if rounded to 0.83, the classification ability using test data can predict the class correctly is 83%, in testing with training data of 80% and testing data of 20%. Precision is a measure of how well the model classifies Positive classes that are truly Positive classes, a Precision value of 0.90 means that the classification successfully predicts Positive classes that are truly Positive by 90%. While recall is a measure of how well the model finds all existing positive samples, a recall value of 0.850 or 85% high recall indicates the model's ability to predict existing positive samples. F1 Score is 0.874 or can make accurate model predictions of 87% in the calcification model. F1 Score combines Precision and recall into one value to see the sensitivity of the calcification model, and in this case, the model has good sensitivity between the two metrics.

4 Conclusion

Based on the results of the implementation and test results of the system, the conclusions of the sentiment analysis on Traveloka application comments using the Long Short Term Memory method produces an accuracy of 0.829 or if it is monthized by 0.83 in the training and testing division of 80: 20. In other words, the LSTM model in classifying Traveloka application comments can predict 83% correctly.

5 Acknowledgments

We would like to thank the Faculty of Mathematics and Natural Science, Tadulako University for funding this research.

References

1. F. Bei and S. Sudin, “Analisis Sentimen Aplikasi Tiket Online Di Play Store Menggunakan Metode Support Vector Machine (Svm),” Sismatik, vol. 01, no. 01, 2021.
2. B. Faradilla, “Analisis Sentimen Terhadap Aplikasi Traveloka Menggunakan Metode Learning Vector Quantization (Lvq) Berdasarkan Ulasan Di Google Play Store,” 2020. [Online]. Available: http://repository.uin-suska.ac.id/29632/%0Ahttp://repository.uin-suska.ac.id/29632/1/File lengkap sampai lampiran kecuali hasil penelitian.pdf
3. R. Cahyadi, A. Damayanti, and D. Aryadani, “Recurrent Neural Network (RNN) dengan Long Short Term Memory (LSTM) untuk Analisis Sentimen

data instagram," Jurnal Informatika dan Komputer, vol. 5, no. 1, 2020.
4. J. Nurvania, Jondri, and K. M. Lhaksamana, "Analisis Sentimen Pada Ulasan di TripAdvisor Menggunakan Metode Long Short-Term Memory (LSTM)," e-Proceeding of Engineering, vol. 8, no. 4, 2021.
5. W. Widayat, "Analisis Sentimen Movie Review menggunakan Word2Vec dan metode LSTM Deep Learning," JURNAL MEDIA INFORMATIKA BUDIDARMA, vol. 5, no. 3, 2021, doi: 10.30865/mib.v5i3.3111.
6. N. Fikria, "ANALISIS KLASIFIKASI SENTIMEN REVIEW APLIKASI," UNIVERSITAS ISLAM INDONESIA, YOGYAKARTA, 2018.
7. D. Setiabudi, "Sistem Informasi Peramalan Beban Listrik Jangka Panjang di Kabupaten Jember Menggunakan JST Backpropagation," SMARTICS Journal, vol. 1, no. 1, 2015.
8. E. Budianita and U. D. Arni, "Penerapan Learning Vector Quantization Penentuan Bidang Konsentrasi Tugas Akhir (Studi Kasus : Mahasiswa Teknik Informatika UIN Suska Riau)," CoreIT, vol. 1, no. 2, 2015.
9. C. Tian, J. Ma, C. Zhang, and P. Zhan, "A deep neural network model for short-term load forecast based on long short-term memory network and convolutional neural network," Energies (Basel), vol. 11, no. 12, 2018, doi: 10.3390/en11123493.
10. Q. Yin, R. Zhang, and X. Shao, "CNN and RNN mixed model for image classification," MATEC Web of Conferences, vol. 277, 2019, doi: 10.1051/matecconf/201927702001.
11. A. Singh, "Anomaly Detection for Temporal Data using Long Short-Term Memory (LSTM)," Trita-Ict-Ex Nv - 2017:124, vol. Independen, 2017.
12. Colah, "Understanding LSTM Networks -- colah's blog." Accessed: Nov. 06, 2022. [Online]. Available: https://colah.github.io/posts/2015-08-Understanding-LSTMs/
13. V. K. S. Que, A. Iriani, and H. D. Purnomo, "Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization," Jurnal Nasional Teknik Elektro dan Teknologi Informasi, vol. 9, no. 2, 2020, doi: 10.22146/jnteti.v9i2.102.
14. T. Kurniawan, "Implementasi Text Mining Pada Analisis Sentimen Pengguna Twitter Terhadap Media Mainstream Menggunakan Naïve Bayes Classifier Dan Support Vector Machine Media Mainstream Menggunakan Naïve Machine," IT Journal, vol. 23, 2017.