p-ISSN: 2252-3006
e-ISSN: 2685-2411

User X's Sentiment on Cryptocurrency: Comparison of SVM and LSTM Methods

I Gede Yoga Asta Pramana[a1], Putu Veda Andreyana[a2], Putu Widiadnyana[b3], Dean Agnia[a4], Kadek Boy Brahmanda Satrya Montana[a5], Krisna Juniartha[a6]
[a]Information Technology Program, Faculty of Engineering, Udayana University
Bukit Jimbaran, Bali, Indonesia-80361
[b]STIKOM Bali Institute of Technology and Business
e-mail: [1] yoga.asta.pra@gmail.com, [2]putuveda@unud.ac.id, [3]widiadnyana@stikom-bali.ac.id, [4]deanagnia7@gmail.com, [5]boybrahmanda03@gmail.com, [6]gedekrisna22@gmail.com

Abstrak

Penelitian ini menganalisis sentimen pengguna X terhadap cryptocurrency menggunakan metode Support Vector Machine (SVM) dan Long Short-Term Memory (LSTM). Penelitian ini bertujuan untuk menentukan efektivitas metode tersebut dalam mengklasifikasikan sentimen sebagai positif, negatif, atau netral. Data dikumpulkan dari kumpulan data tweet terkait crypto. Langkah-langkah praproses meliputi tokenisasi, penghapusan stop-word, dan normalisasi. Model SVM dan LSTM kemudian dilatih dan diuji pada dataset ini. Hasil penelitian menunjukkan bahwa kedua metode tersebut efektif, dengan model LSTM menunjukkan kinerja yang sedikit lebih baik dalam menangani data berurutan dibandingkan dengan model SVM. Model LSTM memiliki akurasi pelatihan yang lebih tinggi dibandingkan SVM dengan persentase akurasi pelatihan sebesar 94,45% pada model LSTM berbanding 88,66% pada model SVM. Akurasi pengujian model LSTM menghasilkan persentase sebesar 77,00% sedangkan model SVM menghasilkan persentase sebesar 78,65%.

Kata kunci: *Analisis Media Sosial, Analisis Sentimen, Cryptocurrency, Long Short-Term Memory, Support Vector Machine*

Abstract

This research analyzes the sentiment of X users towards cryptocurrency using Support Vector Machine (SVM) and Long Short-Term Memory (LSTM) methods. This research aimed to determine the effectiveness of these methods in classifying sentiment as positive, negative, or neutral. Data was collected from a dataset of crypto-related tweets. Preprocessing steps include tokenization, stop-word removal, and normalization. SVM and LSTM models were then trained and tested on this dataset. The results show that both methods are effective, with the LSTM model showing slightly better performance in handling sequential data compared to the SVM model. LSTM model has a higher training accuracy compared to SVM with a training accuracy percentage of 94.45% in the LSTM model compared to 88.66% in the SVM model. Testing accuracy of the LSTM model produces a percentage of 77.00% while the SVM model produces a percentage of 78.65%.

Keywords: *Sentiment Analysis, Cryptocurrency, Support Vector Machine, Long Short-Term Memory, Social Media Analysis*

1. Introduction

Sentiment analysis on X has become an important tool in understanding public opinion, especially in the cryptocurrency field to make price predictions [1]. X can be a data source used in sentiment analysis techniques so that it can extract large volumes and variety of information [2]. Cryptocurrencies are a hot topic of conversation among X users. The volatile nature of cryptocurrencies requires an accurate and real-time understanding of market sentiment to predict trends and make the right investment decisions [3].

p-ISSN: 2252-3006
e-ISSN: 2685-2411

Among various machine learning algorithms, Support Vector Machine (SVM) and Long Short Term Memory (LSTM) networks stand out for their effectiveness in classification tasks. Support vector machine (SVM) is a powerful machine learning technique that relies on the principle of structural risk minimization [4]. SVM is a powerful tool for data classification [5]. This algorithm works by finding the hyperplane that best separates data into different classes [6]. On the other hand, The proposed distributed framework with long short term memory (LSTM) framework is an alternative to conventional sentiment analysis approaches in analyzing large volumes of data in a potential stream [7]. The LSTM results are comprehensive and identify that LSTM has better performance than the basic model [7]. Model evaluation shows that the combination of LSTM models performs very well on news articles with a large number of relevant sentiment words [8].

This research aims to utilize these advanced machine learning techniques to analyze user sentiment towards cryptocurrencies. By leveraging big data analytics and sentiment analysis, this research aims to provide a deeper understanding of market dynamics, thereby offering valuable insights for stakeholders in the cryptocurrency field.

2. Research Method

The research methodology explains the research stages that underlie the data collection process, as well as comparative testing of the SVM algorithm with LSTM.

2.1 Process Stages

The stages carried out in this include data retrieval, data labeling, and after that data labeling, data preprocessing is divided into 5 stages. *handling duplicate, cleaning, handling missing value, tokenize, normalization, stemming,* divide the data into test data and train data, then create an SVM & LSTM model, carry out model evaluation, model testing, compare the two algorithm models used, make comparison results visually, and after that get *preprocessing result*.

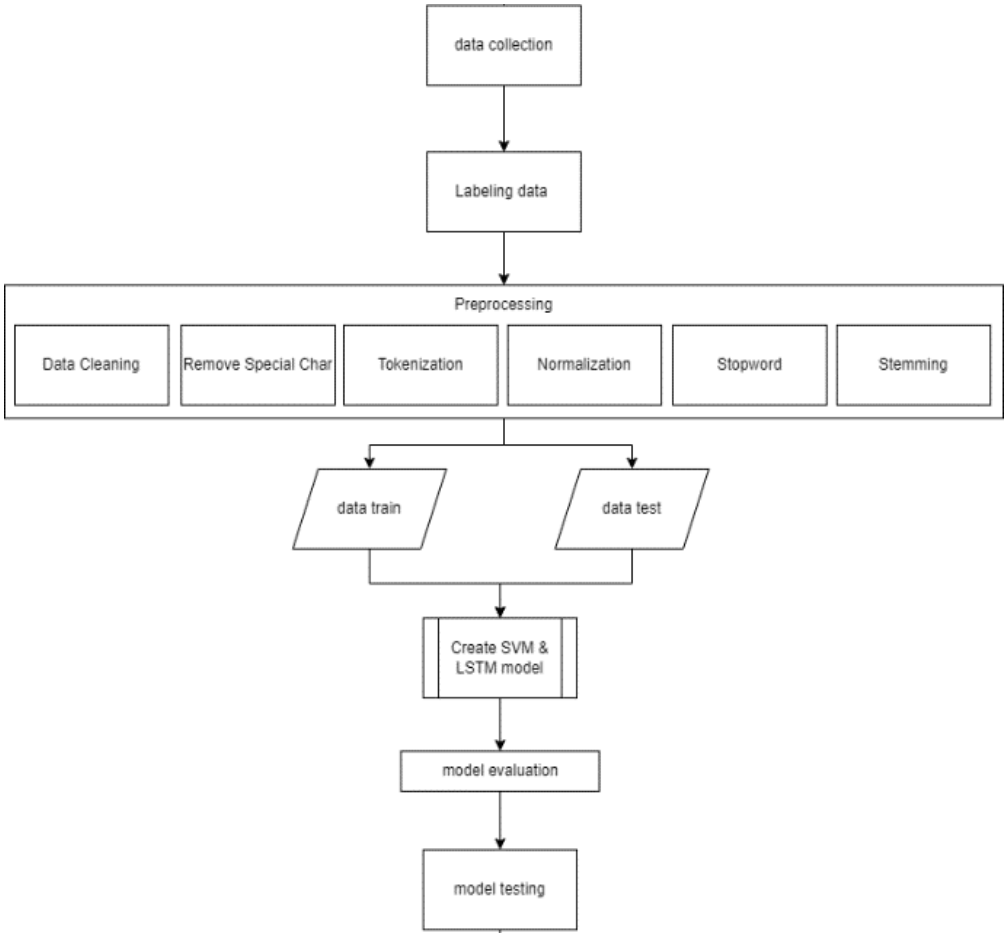

Figure 1 Research Stage

p-ISSN: 2252-3006
e-ISSN: 2685-2411

Figure 1 depicts the research process for developing the SVM model (*Support Vector Machine*) and LSTM (*Long Short-Term Memory*) for text analysis, from the initial stages to the final results. The first stage is data collection, where relevant data is collected for training and testing purposes. The second stage is data labeling to create a supervised learning dataset. The third stage is data preprocessing, which includes several steps: data cleaning to remove unnecessary elements, removing special characters from the text, tokenization to break the text into tokens or words, normalization to ensure data consistency, stopword removal to remove common words that do not have important meaning, and stemming to reduce words to their basic forms. After preprocessing is complete, the data is divided into training (training data) and testing (test data) sets. In the fourth stage, the model is created using the SVM and LSTM algorithms (creating an SVM & LSTM model). The fifth stage is evaluating the performance of these models (model evaluation) followed by testing them with test data (model testing). The sixth stage is to compare the results of the SVM and LSTM models to determine which model has better performance (comparison). The final stage is to visualize the data and model performance results (data visualization), and conclude the process (completion). This workflow outlines the detailed steps from data collection to visualization of model results.

3. Literature Study

The literature study discusses the general theoretical basis which is used as the basis for the research process of Comparison of Support Vector Machine and LSTM for analyzing X user sentiment regarding cryptocurrency.

3.1. Sentiment Analysis

Sentiment Analysis is a natural language processing that is able to detect a person's emotions and opinions on a particular topic [9]. These emotions and opinions can be obtained from a text which is categorized into three, including positive, negative or neutral categories[10].

3.2. Cryptocurrency

Cryptocurrency is a technology based *blockchain* which is widely used as digital currency. Ownership of *cryptocurrency* not controlled by the government or banks [11]. User *cryptocurrency* can be found in almost all parts of the world as market interest increases worldwide and explicitly, there is no doubt that cryptocurrency users will continue to increase as the market becomes more attractive [12].

3.3. Support Vector Machine

Support Vector Machine (SVM) is a method *machine learning* which is useful to find *hyperlane* best that divides the two parts of the class in the input space. SVM is usually used for linear and non-linear data classification. SVM uses kernel tricks to train training samples into a high-dimensional feature space and identify *hyperlane* optimal separator [13]. Over the past few years, SVM has become a superior method for classification patterns and has a high success rate when applied in various aspects [14].

3.4. Long Short-Term Memory

LSTM is an evolution of the RNN method through the addition of cells (*cell*) LSTM on RNN architecture [15]. LSTM uses a structure called *gate* to help store the necessary information, so that it can maintain long-term information. *Cell Gates* serves to regulate the information that will be released to *cell state* or next unit [16]. *Cell State* is the path at the top to send information to the next unit [17]. In addition, there is a formula to calculate the output from the layer *Forget Gate* which means deleting information and determining important information. The LSTM algorithm is commonly used for prediction and classification. Apart from that, the LSTM algorithm is also suitable for use on data that has sequences such as text because each data is connected to each other [18].

p-ISSN: 2252-3006
e-ISSN: 2685-2411

3.5. Evaluation Matrix

The Evaluation Matrix is a matrix used to carry out evaluations which has four categories, including True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) [19]. Then precision, recall, F1-Score and macro F1-Score will also be seen with the following calculations.

1. Precision

Precision is the True Positive (TP) of the total positive predicted value.

$$Precision = \frac{City}{TP + FP}$$

2. Recall

Recall is the proportion of True Positive values from the total actual positive values.

$$Recall = \frac{City}{TP + FN}$$

3. F1-score

F1-score is the harmonic average of precision and recall.

$$F1 - Score = \frac{2 * Precision * Recall}{Precision + Recall}$$

4. Macro F1-score

Macro average is the average of performance within a class model. The F1-score macro is used to see the average F1-score performance of each class, so that these classes can be treated the same.

$$Macro\ F1 - Score = \frac{\sum_{i=1}^{n} \quad F1 - Score_i}{n}$$

4. Result and Discussion

In this section, we discuss and analyze the results of User X's Sentiment on Cryptocurrency: Comparison of SVM and LSTM Method research. There are several processes that will be discussed in this section, the following are the processes discussed in this research.

4.1. Data Collection Results

The initial process carried out in this research was data collection. Data collection is done by doing *data collection* by retrieving data on the Kaggle platform. When collecting data for this research, not all columns were used for this research, the researcher only used the tweet_row and labels columns. There are three labels in the data taken, these labels are *negative, neutral,* And *negative.* The following are the results of data collection carried out by researchers.

	Content	Labels
0	hong kong to kick ripple xrp off the top crypt...	neutral
1	i rather love how many commenters have not rea...	positive
2	coinbase says 43 increase in subscription and ...	positive
3	texas bitcoin miners buy electricity from the ...	negative
4	yall fake ass be changing up like this weather	neutral
...	...	...
4593	i really never been to kalahari or chic fil a	neutral
4594	todays meeting has been hawkish very hawkish ...	negative
4595	this claim is true reformparty_uk has candida...	neutral
4596	dogecoin high time preference dont shtcoin i...	neutral
4597	1 duhhh thats why dot is your daddy built usin...	negative

4598 rows × 2 columns

Figure 2 Total Data

p-ISSN: 2252-3006
e-ISSN: 2685-2411

The figure above is the result of data collection carried out by researchers. The amount of data obtained in this research amounted to 4597 data with detailed data *negative* of 1504, data *neutral* 1594, and data *positive* 1500.

4.2. Data Preprocessing

Next, the researcher carried out data preprocessing. There were several stages of data preprocessing carried out in this research. The data preprocessing stages carried out in this research are data cleaning, remove special char, tokenization, normalization, stopword, and stemming.

	Content	Labels
0	all tethers are liars and nothing more than useful idiots to our enemies	neutral
1	this research defines 6 main categories of depin projects compute ie filecoin akashnet_ livepeer wireless ie helium wicrypt energy ie toucanprotocol ai ie fetch_ai servic...	neutral
2	qqq running premarket news article out discordggws6egbv2vn sol avax sei bonk trb inj tia near btc eth sui data rune ace icp shib dia lunc woo ete ctk auction shib stx iotx min...	positive
3	they should have called it arg instead of xrp	negative
4	viral x post claims chickfila charges more in drivethru than inside restaurant abc6onyoursidecomnewsoffb	positive
...	...	...
4597	livestreaming development of spelljammer combat exploration supplement publishing dms_guild illustrating paraelemental planets pipedvideowatchv74bekzdl wizards_dnd ttrpgs f...	neutral
4598	todays meeting has been hawkish very hawkish fed funds future market briefly priced in a terminal rate of 26 aug 2023 contract fedguy12s call is correct once again	negative
4599	this guy agrees	neutral
4600	invested in the future of cryptocurrency hodling dogecoin to the moon cryptocraze tothemoonandbeyond	neutral
4601	1 duhhh thats why dot is your daddy built using the dot sdk 2 eth was once 721 in its infancy look below at the pic dot will be more than 1000x than eth currently is more th...	negative

4602 rows × 2 columns

Figure 3 Preprocessed Data

The figure above is the result of data taken from the Kaggle platform which has carried out all the data preprocessing processes as mentioned previously. This preprocessing data has a big influence on the model accuracy results during model training later.

4.3. Model Build

The next process in this research is to build a model using a method or algorithm *Long Short Term Memory Networks* (LSTM) and *Support Vector Machine* (SVM). This research uses this algorithm because this algorithm is an algorithm that is often used in text classification, especially in making sentiment analysis.

4.3.1. Long Short Term Memory Networks (LSTM)

Long Short Term Memory Networks (LSTM) is an algorithm model that is improved from Recurrent Neural Network (RNN). Researchers build a model using the LSTM algorithm, then training will be carried out on the model that has been built previously, and in this research a model evaluation will also be carried out. The following are the results of model building and model training results in the form of accuracy and accuracy validation.

```
Model: "sequential_3"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding_3 (Embedding)      (None, 620, 400)          4000000
lstm_3 (LSTM)                (None, 400)               1281600
dense_3 (Dense)              (None, 3)                 1203
=================================================================
Total params: 5282803 (20.15 MB)
Trainable params: 5282803 (20.15 MB)
Non-trainable params: 0 (0.00 Byte)
_________________________________________________________________
None
Epoch 1/10 87/87 [==============================] - 644s 7s/step - loss: 0.9290 - accuracy: 0.5551 - val_loss: 0.6103 - val_accuracy: 0.7098
Epoch 2/10 87/87 [==============================] - 683s 8s/step - loss: 0.4072 - accuracy: 0.8332 - val_loss: 0.4918 - val_accuracy: 0.7592
Epoch 3/10 87/87 [==============================] - ETA: 0s - loss: 0.1614 - accuracy: 0.9445
Accuracy achieved, training stopped...
87/87 [==============================] - 680s 8s/step - loss: 0.1614 - accuracy: 0.9445 - val_loss: 0.6364 - val_accuracy: 0.7625
```

Figure 4 LSTM Result

The figure above is the result of model training carried out in this research. The results of this model training obtained an accuracy of 0.9445 or around 94% with a validation accuracy of

p-ISSN: 2252-3006
e-ISSN: 2685-2411

0.7625 or around 76%. The training was carried out for ten epochs and stopped at the 3rd epoch because the callbacks target had been reached.

4.3.2. Support Vector Machine (SVM)

Support Vector Machine (SVM) is an algorithm used to carry out classification. This algorithm is an algorithm that handles classification problems from various types of data. This research carries out a model development and training using this algorithm. The following are the results of model training using SVM in the form of accuracy.

Table 1 SVM Validation Result

	Precision	Recall	F1-score	support
Negative	0.70	0.62	0.66	298
Neutral	0.67	0.71	0.69	320
Positive	0.93	0.98	0.95	302

The evaluation table shows the performance of the SVM model dividing the data into three categories: Negative, Neutral, and Positive. For the Negative category, the model achieved precision of 0.70 and recall of 0.62, resulting in an F1-score of 0.66 from 298 samples. In the Neutral category, precision and recall are 0.67 and 0.71 respectively, with an F1-score of 0.69 from 320 samples. The Positive category shows the best performance with precision 0.93 and recall 0.98, resulting in an F1-score of 0.95 from 302 samples.

Table 2 SVM Training Result

	Precision	Recall	F1-score	support
Negative	0.85	0.77	0.81	1274
Neutral	0.78	0.84	0.81	1274
Positive	0.97	1.00	0.98	1274

4.4. Comparasion

In this research, model training for each algorithm used has previously been carried out. After training the models for each algorithm, the researcher will then compare the model accuracy from the results of the model training that has been carried out. Below we will show a comparison of the results of the model training that has been carried out in this research.

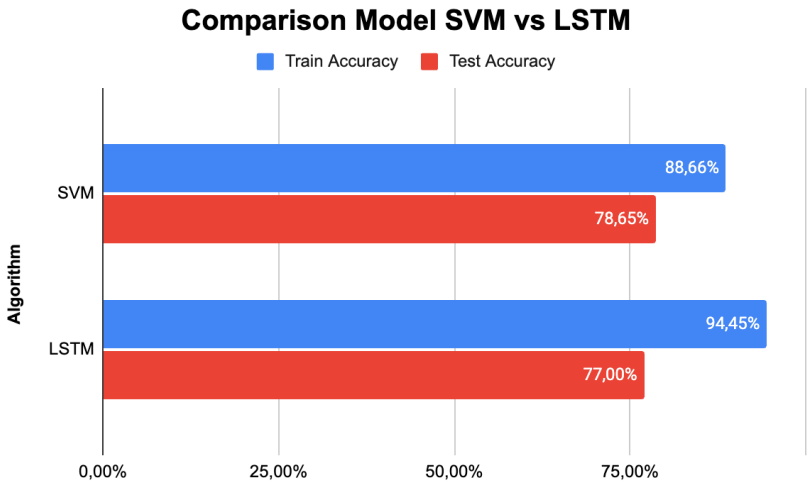

Figure 5 Comparison Model

The figure above shows a performance comparison between two machine learning models, namely Support Vector Machine (SVM) and Long Short-Term Memory (LSTM). There

are two performance metrics that are compared, namely: Train Accuracy shows the accuracy of the model when trained with training data. For SVM, the training accuracy is 88.66%, while for LSTM it is 94.45%. Then Test Accuracy shows the accuracy of the model when tested with independent test data. For SVM, the test accuracy is 78.65%, while for LSTM it is 77.00%.

5. Conclusion

From the comparison of the Support Vector Machine (SVM) and Long Short-Term Memory (LSTM) method tests to analyze user sentiment X towards cryptocurrency, it can be seen that the LSTM model has a higher training accuracy than SVM with a training accuracy percentage of 94.45% in the LSTM model compared to 88.66% in the SVM model. While the LSTM testing accuracy is lower than SVM with a testing accuracy percentage of 78.65% in the SVM model and 77.00% in the LSTM model. Therefore, referring to the training accuracy results, it can be concluded that LSTM is the best method that can be applied to sentiment analysis in discussions related to cryptocurrencies on social media with a higher level of accuracy.

References

[1] X.-F. Jiang *et al.*, “Analyst sentiment and earning forecast bias in financial markets,” *Physica A: Statistical Mechanics and its Applications*, flight. 589, p. 126601, Mar. 2022, doi: 10.1016/j.physa.2021.126601.

[2] F. Wunderlich and D. Memmert, “A big data analysis of Twitter data during premier league matches: do tweets contain information valuable for in-play forecasting of goals in football?,” *Soc Netw Anal Min*, vol. 12, no. 1, p. 23, Dec. 2022, doi: 10.1007/s13278-021-00842-z.

[3] Z. Chen, C. Li, and W. Sun, “Bitcoin price prediction using machine learning: An approach to sample dimension engineering,” *J Comput Appl Math*, vol. 365, p. 112395, Feb. 2020, doi: 10.1016/j.cam.2019.112395.

[4] A. Roy and S. Chakraborty, “Support vector machine in structural reliability analysis: A review,” *Reliab Eng Syst Saf*, vol. 233, p. 109126, May 2023, doi: 10.1016/j.ress.2023.109126.

[5] R. G. Mayopu, V. Nalluri, and L.-S. Chen, “Classification ChatGPT Generated News and True News Using Support Vector Machines,” in *2023 12th International Conference on Awareness Science and Technology (iCAST)*, IEEE, Nov. 2023, pp. 228–232. doi: 10.1109/iCAST57874.2023.10359313.

[6] G. Lantzanakis, Z. Mitraka, and N. Chrysoulakis, “X-SVM: An Extension of C-SVM Algorithm for Classification of High-Resolution Satellite Imagery,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 5, pp. 3805–3815, May 2021, doi: 10.1109/TGRS.2020.3017937.

[7] D. C. Edara, L. P. Vanukuri, V. Sistla, and V. K. K. Kolli, “Sentiment analysis and text categorization of cancer medical records with LSTM,” *J Ambient Intl Humaniz Comput*, vol. 14, no. 5, pp. 5309–5325, May 2023, doi: 10.1007/s12652-019-01399-8.

[8] S. Usmani and J. A. Shamsi, “LSTM based stock prediction using weighted and categorized financial news,” *PLoS One*, vol. 18, no. 3, p. e0282234, Mar. 2023, doi: 10.1371/journal.pone.0282234.

[9] D. Oktavia and Y. R. Ramadahan, "Sentiment Analysis of the Implementation of the E-Ticket System on Twitter Social Media Using the Support Vector Machine (SVM) Algorithm," *CLICK: Scientific Study of Informatics and Computers*, vol. 4, no. 1, pp. 407–417, 2023, doi: 10.30865/klik.v4i1.1040.

[10] F. M. J. M. Shamrat *et al.*, “Sentiment analysis on twitter tweets about COVID-19 vaccines using NLP and supervised KNN classification algorithm,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 23, no. 1, pp. 463–470, Jul. 2021, doi: 10.11591/ijeecs.v23.i1.pp463-470.

[11] S. Rouhani and E. Abedin, “Crypto-currencies narrated on tweets: a sentiment analysis approach,” *International Journal of Ethics and Systems*, vol. 36, no. 1, pp. 58–72, Jan. 2020, doi: 10.1108/IJOES-12-2018-0185.

[12] R. Azhar, A. Surahman, and C. Juliane, "Sentiment Analysis of Cryptocurrency Based on Python TextBlob Using the Naïve Bayes Algorithm," *Journal of Computer Science & Informatics (J-SAKTI*, vol. 6, no. 1, pp. 267–281, 2022.

[13] G. Angga and N. Piarsa, “Comparison of Support Vector Machine and K-Nearest Neighbor for Baby Foot Identification based on Image Geometric Characteristics,” *MERPATI SCIENTIFIC JOURNAL (Information Technology Academic Research Tower)*, vol. 9, no. 1, 2021.

[14] E. Suryati, A. Ari Aldino, N. Corresponding Author, and E. Suryati Submitted, "Online Transportation Sentiment Analysis Using Word2vec Text Embedding Model Feature Extraction and Support Vector Machine (SVM) Algorithm," vol. 4, no. 1, pp. 96–106, 2023, doi: 10.33365/jtsi.v4i1.2445.

[15] D. R. Alghifari, M. Edi, and L. Firmansyah, "Implementation of Bidirectional LSTM for Sentiment Analysis of Grab Indonesia Services," *Journal of Information Management (JAMIKA)*, vol. 12, no. 2, pp. 89–99, Sept. 2022, doi: 10.34010/jamika.v12i2.7764.

[16] K. Xia, J. Huang, and H. Wang, “LSTM-CNN Architecture for Human Activity Recognition,” *IEEE Access*, vol. 8, pp. 56855–56866, 2020, doi: 10.1109/ACCESS.2020.2982225.

[17] R. Dwi, W. Santosa, M. Arif Bijaksana, and A. Romadhony, "Implementation of the Long Short-Term Memory (LSTM) Algorithm to Detect the Use of Abusive Sentences in Indonesian Texts," *Final Project Journal of the Faculty of Informatics*, vol. 8, no. 1, 2021.

[18] Y. Widhiyasana, T. Semiawan, I. Gibran, A. Mudzakir, and M. R. Noor, "Application of Convolutional Long Short-Term Memory for Classification of Indonesian News Texts (Convolutional Long Short-Term Memory Implementation for Indonesian News Classification)," 2021.

[19] Anak Agung Mayra Candra Dyanggi, I Wayan Agus Surya Darma, and Ni Nyoman Ayu J. Sastaparamitha, "FastText and Bi-LSTM for Sentiment Analysis of Tinder Application Reviews," Vol 12 No 1 (2024): Vol. 12, No. 1, April 2024.

[20] Tita Lattifia, Putu Wira Buana, and Ni Kadek Dwi Rusjayanthi, "Model Prediksi Cuaca Menggunakan Metode LSTM" 2022.

[21] Ni Wayan Sumartini Saraswati, Dewa Ayu Putu Rasmika Dewi, and Poria Pirozmand, "Comparative Analysis of SVM and CNN for Pneumonia Detection in Chest X-Ray," Vol 15 No 1 (2024): Vol. 15, No. 1 April 2024.

[22] K. Adrian Manurung, “Sentiment Analysis of Tourist Attraction Review from TripAdvisor Using CNN and LSTM,” *International Journal on Information and Communication Technology (IJoICT)*, vol. 9, no. 1, pp. 73–85, Jul. 2023, doi: 10.21108/ijoict.v9i1.756.