Building of Informatics, Technology and Science (BITS)
Volume 4, No 4, Maret 2023 Page: 1815−1820
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i4.3084

Hoax Detection Using Long Short-Term Memory (LSTM) and Gate Recurrent Unit (GRU) on Social Media

Dion Pratama Putra*, Erwin Budi Setiawan

School of Computng, Informatics Engineering, Telkom University, Bandung, Indonesia
Email: 1,*dionpputraaa@student.telkomuniversity.ac.id, erwinbudisetiawan@telkomuniversity.ac.id
Correspondence Author Email: dionpputraaa@student.telkomuniversity.ac.id
Submitted: **02/02/2023**; Accepted: **27/03/2023**; Published**: 31/03/2023**

Abstract−There are negative effects of the ease of obtaining information in today's society, one of which is the rise of hoaxes on the internet. As much as 92.40% of social media platforms such as Twitter are used to spread hoaxes. Launched on July 13, 2006, Twitter is a microblogging service where users can spread information at no cost to themselves or others. In this study, the authors will conduct hoax news detection on Twitter social media using the Long Short - Term Memory (LSTM) method and Gate Recurent Unit (GRU) and gloVe feature expansion. with a dataset of 25,234 data which produces accuracy results in TF-IDF on each model, namely 97.33% in LSTM and 96.75% in GRU, and an increase in accuracy of 0.22% in the tweet corpus on LSTM and an increase in accuracy of 0.15 in the BeritaTweet corpus on GRU.

Keywords: Hoax; Twitter; Long Short Term Memory (LSTM); Gate Recurent Unit (GRU); GloVe

1. INTRODUCTION

Information has become increasingly accessible over time thanks to several media, for example social media. Social media is a technology where users can share, communicate and get information from social media. However, this has led to undesirable consequences, such as the rapid spread of information and fake news through social media [1]One of the most widely used social media today is Twitter, where users can share tweets consisting of 280 characters. because it is very easy to access and free, many twitter users abuse existing features such as hoax tweets and also hate speech. A hoax is news or information that contains dubious or completely false claims. It is not easy to find credible sources of information and news in the modern world. According to a survey conducted by Mastel in 2017, 44.3 percent of the 1,146 respondents said that they often encounter false information, with 17.2 percent saying that they often encounter it every day. And 92.4 percent of people use social media such as Twitter to spread hoaxes [2]. Founded on July 13, 2006 [3], Twitter is a social networking service built on short updates known as microblogging. In this study, 25,324 data were collected from Twitter and evaluated to see if they were real or fake. The material is collected according to the problem studied in this research. The research limitation in this Final Project is a dataset of 25,324 Indonesian tweets with the topic of Kanjuruhan and Ferdy Sambo, the labeling of the dataset is done manually with 2 labels, namely '0' for non hoax and '1' for hoax.

The purpose of this study is to measure the performance value, and analyze the results of the hoax detection system built using the TF-IDF feature extraction method and GloVe feature expansion on Indonesian-language tweet data in a dataset that has been previously crawled. Several studies have been conducted previously on hoax detection using deep learning. The review classification process will make it easier for users to categorize an opinion that is hoax or non hoax more precisely [4]. This study uses Gate Recurrent Unit and Long Short Term Memory approaches, where the Long Short Term Memory method, and using TF-IDF feature extraction and feature expansion with gloVe. In this study, the authors built 3 corpus using gloVe to get an increase in the accuracy of each model used. the corpus built is a tweet corpus, news corpus and a combined corpus between tweets and news, namely the newsatweet corpus.

In 2019. Researchers Faisal Rahutomo, Inggrid Yanuar Risca Pratiwi, and Diana Mayangsari Ramadhani found an average accuracy of 82.6% when utilizing the Nave Bayes approach to detect hoaxes using static testing on 600 news reports. And dynamic testing is done by entering news content into the system. Of the 60 news items tested, 41 news items resulted in the same news classification as the manual mark and 19 news items resulted in a different classification from the manual mark. The percentage of results that are suitable is 68.33% and not suitable is 31.67%. [5]and in 2019 also, Shu, K., Cui, L., Wang, S., Lee, D., & Liu H also conducted research related to hoax detection on social media using the GRU and Bi-LSTM models and obtained 71.2% accuracy on GRU and 84.6% on Bi-LSTM[6].

The best accuracy of this study was achieved by a 2021 paper written by Ema Utami, Ahmad Fikri Iskandar, Wahyu Hidayat, Agung Budi Prasetyo, and Anggit Dwi Hartanto who used the KNN classification technique with Jaccard Space with Nazief & Adriani stemming. With K = 5, the KNN model in Jaccard Space, based on the work of Nazief and Adriani, achieved an accuracy of 75.89%. [7]

In 2022, Candra Surya Sriyano and Erwin Budi Setiawan showed that Naive Bayes Multinomial and TF-IDF weighting method can be used in detecting hoax news on twitter. The accuracy result obtained from the experiment conducted by the author with TF-IDF is 72.06%, then the accuracy result of the largest N-Gram, Unigram with 80% train data is 72.06%, with the weighting feature with TF-IDF, N-Gram has an effect because it makes the accuracy result obtained increase 0.46% greater than without using N-Gram[8].

Building of Informatics, Technology and Science (BITS)
Volume 4, No 4, Maret 2023 Page: 1815−1820
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i4.3084

2. RESEARCH METHODOLOGY

2.1 System Design

The system that has been created can be seen in Figure 1, This research begins with crawling data, labeling data, preprocessing data, feature extraction, split data, feature expansion, and modeling. Here's an explanation for each stage.

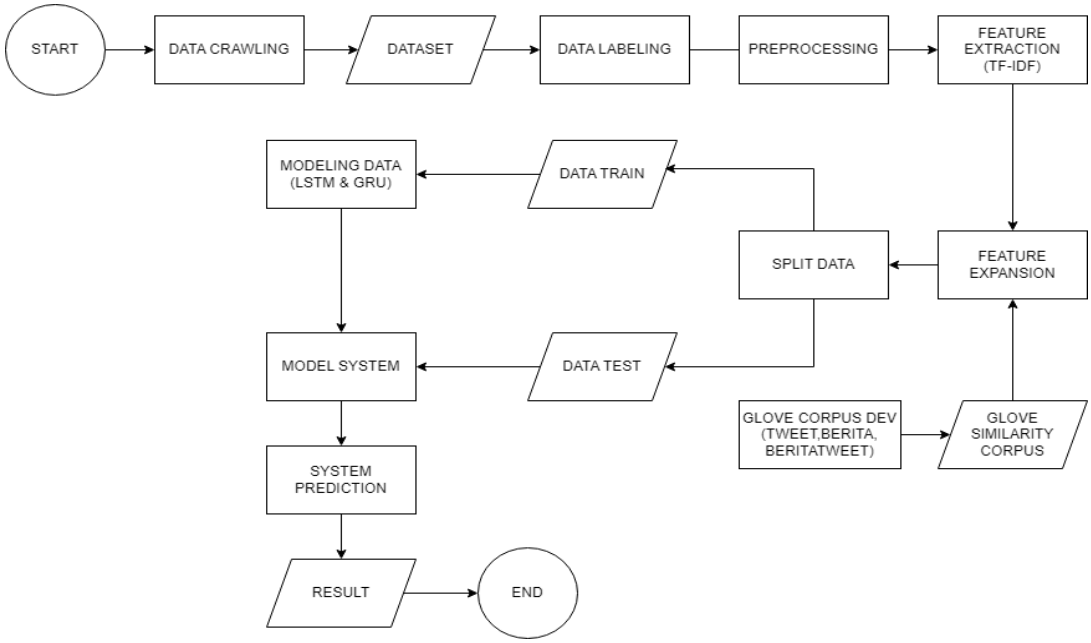

Figure 1. Flowchart System

2.2 Crawling Data

Datasets are obtained from tweets on Twitter using the Twitter API by crawling data according to the topic we want to identify the truth. Data retrieval is done with the API (Application Program Interface) provided by Twitter developers as a liaison from Twitter so that the data can be retrieved and processed [9].

After successfully collecting the dataset, the next step is to store the data in the database and the data can be reused during data pre-processing. data searched through tweets containing the words Kanjuruhan and Ferdy Sambo. The resulting data amounted to 25324 data to detect hoaxes. after the data is successfully crawled, the data is stored in a csv file and then the data is labeled, namely 0 for non hoax and 1 for hoax. The authors perform the labeling process manually by searching for news on the internet, can be seen in table 1 examples of data that has been labeled.

Table 1. Labeling dataset

No	Text	Label
1	Negara ini benar-benar memiliki drama, astaga, saya telah diberitahu bahwa semua bom telah beres	0
2	Video yang diarsipkan, terkadang diputar ulang ribuan kali, mengumpulkan doa untuk para korban monumen kehidupan pahlawan. Satu keinginan, kekuatan untuk membagi sumber	0
3	Putri Candrawathi , istri mantan Kabag Propam Irjen Pol Ferdy Sambo, dicurigai membunuh Brigadir J	0
4	brigadir rr alias ricky rizal mengancam hukuman mati jika ada bukti yang terlibat dalam rencana pembunuhan itu bagaimana Irjen Sambo	1
5	hukum kematian sambo wow	1
6	putri candrawathi saya lebih baik mati daripada melaporkan hal ini sayang puteri candrawathi	1

2.3 Data Preprocessing

When the dataset has been collected, the data is still in an unstructured state, therefore several processes are needed to convert unstructured data into structured data which is commonly called preprocessing [10]. preprocessing has several stages, namely Data Cleaning, Case Folding, Tokenizing, Stopwords Removal, and Stemming [11].

The following is an explanation of some of the preprocessing stages:

a. Data Cleaning, removing noise in the text such as usernames, hashtags, URL links, numbers, and punctuation.
b. Case Folding, converts uppercase letters to lowercase letters,
c. Tokenizing, breaking the text into tokens or per word contained in the text.
d. word contained in the text.

Building of Informatics, Technology and Science (BITS)
Volume 4, No 4, Maret 2023 Page: 1815−1820
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i4.3084

e. Stopword Removal, removing words in tweets that contain words that are considered considered not to have an important influence in determining classification such as conjunctions.
f. Stemming, making it only as a base word, eliminating suffixes and prefixes.

2.4 Data Split

After preprocessing, the data is continued to the next stage, namely splitting data. splitting data is dividing data into test data and train data, in this study the data used is 70:30 with details of 70% for training data and 30% for test data.

2.5 Term Frequency-Inverse Document Frequency (TF-IDF)

The TF-IDF method is used to isolate functions. Each word is calculated individually using this method, resulting in a weighted value for the word [12]. Equation (1) is used to calculate the mass, which is then used in the fraud detection procedure. Using TF-IDF, we can see how different scales have evolved over time. TF is multiplied by IDF to get the TF-IDF value.

$$idfj = log\ log\ (\ D\ dfj\ ) \quad (1)$$

Then, TF-IDF calculation can be done to get the result. Equation (2) is the TF-IDF calculation formula.

$$wij = tfij\ x\ idfj \quad (2)$$

Information:
tfij : Number of occurrences of the term in the document
wij : Weight of term in document
D : Total number of documents
idfj : Distribution of terms in document
dfj : Number of documents containing the term

2.6 GloVe

gloVe is a word embeddings that performs unsupervised train method for word representation, and gloVe is superior effective for learning vector representation of words [13] When compared with word2vec, the corpus, vocabulary, window size, and training time are the same [14].

Tabel 2. Similarity word

Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Rank 6	Rank 7	Rank 8	Rank 9
tragedi	sdawet	rusuh	stadion	disaster	malang	peristiwa	gbla	rssa

2.7 Long Short Term Memory (LSTM)

As one type of deep learning model, Long Short-Term Memory (LSTM) deserves a closer look. The Long Short-Term Memory Model (LSTM) is an evolution of the RNN technique. To maintain its relevance level, LSTM can use loops by inserting gates In [15] , Long Short-Term Memory (LSTM) is used to process sorted data by controlling the transmission of information sequentially with vector gates at each place. The fact that each neuron in LSTM has its own memory cell and gate unit contributes to this benefit [16]

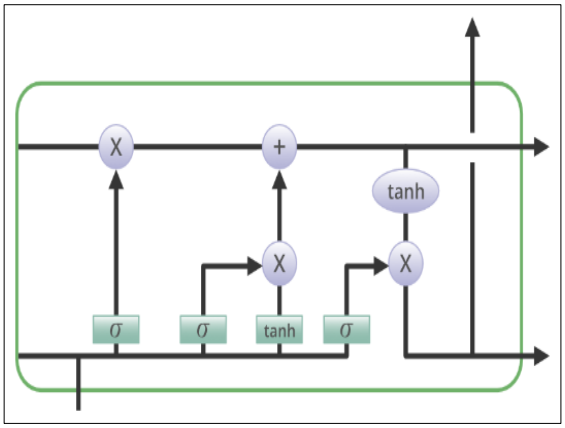

Figure 2. Long Short-Term Memory model architecture

2.8 Gate Recurent Unit (GRU)

Although GRU is computationally simpler than LSTM, it is very successful in dealing with Vanishing Gradient problems because each recurrent unit can be adapted to capture relationships over multiple time scales adiptively [17] . Here is a picture of the architecture of the Gate Recurent Unit model [18]

Building of Informatics, Technology and Science (BITS)
Volume 4, No 4, Maret 2023 Page: 1815−1820
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i4.3084

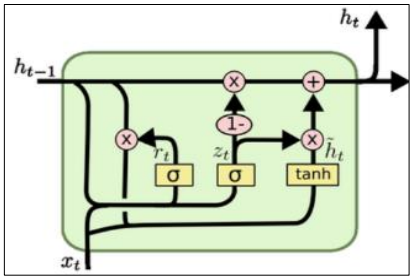

Figure 3. Gate Recurent Unit model architecture

2.9 Performance Measurement

Performance measurement and evaluation of the created algorithm is essential to assess and evaluate the performance of the resulting algorithm. The performance of the system uses binary classification metrics from various pre-existing binary classifications to determine the accuracy score [19]. Whether a given value is optimal for an existing model depends on its precision. The Confusion Matrix is the framework for this investigation. For this evaluation, we use a four-category scale of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) [20].

$$\text{Precision} = \frac{TP}{TP+FP} \quad (6)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (7)$$

$$\text{F1 Score} = \frac{precision\ x\ recall}{precision\ recall} \quad (8)$$

3. RESULT AND DISCUSSION

3.1 Data Distribution

Tweet data that has been obtained from the The crawling process amounted to 25,324 Indonesian tweets containing tweets with the keywords ferdy sambo and kanjuruhan. There are 2 labels, namely 0 for non hoax and 1 for hoax. Then, researchers conducted corpus building for complementary data in this study to get the best accuracy. There are 3 that were built, namely the Tweet corpus, the News corpus, and the combined corpus of the two previous corpuses, namely the BeritaTweet corpus.

3.2 Test Result

This stage includes system testing and analysis of test results. Test and analysis in accordance with the purpose of writing previously described in the introduction. The test system has 3 scenarios, the first scenario uses unigram to find the data ratio that produces the greatest accuracy to be used as a baseline in this study. For the second scenario the author tests using N-Gram to find the highest accuracy of each model using Unigram, Bigram, and Trigram. And in the last scenario the author conducted tests using the corpus that had been built, namely the tweet corpus, Berita corpus, and BeritaTweet corpus.

3.2.1 Scenario 1 (Baseline + TF-IDF)

In this scenario, model testing is carried out by determining the same data ratio in each model, namely 80:20, 70:30 and 90:10. This data ratio division is done to find the data ratio that produces the greatest accuracy to be used as a baseline, and in table 3 the data ratio that has the greatest accuracy is 70:30. Model testing is carried out on the available data. Testing includes 2 models, namely LSTM and GRU.

Tabel 3. Ratio testing data using Unigram

MODEL	70:30	80:20	90:10
LSTM	**87.33**	87.12	87.09
GRU	**87.33**	87.26	87.09

Table 4 shows that the accuracy results with a data ratio of 70:30 using TF-IDF and the LSTM model have an accuracy value of 97.33% and the GRU model has an accuracy value of 96.75%. This shows that the LSTM model is the most accurate model compared to other models.

Tabel 4. Testing using TF-IDF

Model	Ratio Data	Accuracy	F1 Score
LSTM	70:30	**97.33%(+10.00%)**	96.96%
GRU		**96.75%(+9.42%)**	96.79%

Building of Informatics, Technology and Science (BITS)
Volume 4, No 4, Maret 2023 Page: 1815−1820
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i4.3084

3.2.2 Scenario 2 (N-gram Testing)

After finding the data ratio that has the best accuracy results in scenario 1, in scenario 2 researchers search for accuracy using N-grams in each model used.

Tabel 5. Testing Using N-Gram in LSTM Model

Model	N-Gram	Accuracy
LSTM	Unigram	87.12
	Bigram	**87.33**
	Trigram	87.33

Tabel 6. Testing Using N-Gram in GRU Model

Model	N-Gram	Accuracy
GRU	Unigram	87.26
	Bigram	87.33
	Trigram	**87.37**

From the N-Gram test results which can be seen in table 5 and table 6, the GRU model has the best average value of each N-Gram with an Unigram value of 87.26%, Bigram 87.26%, Trigram 87.37%.

3.2.3 Scenario 3 (Feature Expansion *GloVe*)

In this scenario, feature expansion with GloVe is carried out on the test results in scenario 2 using the corpus that has been built before, the data is trained to get the accuracy value of each top-n feature. accuracy value of each top-n feature, in this study 3 categories are used, namely top-1 features, top-5 features, and top-10 features. It can be seen in tables 7 and 8 that the best results in this third scenario are in the expansion feature using the top 1 tweet corpus and the LSTM model, with an accuracy value of 97.55% which can be seen in table 7 and in the expansion feature using the top 1 BeritaTweet corpus and the GRU model, with an accuracy value of 96.85% in table 8

Tabel 7. Performance of the GloVe corpus on the LSTM model

RANK	Akurasi (%)				
	Baseline	TF-IDF	TF-IDF + Tweet	TF-IDF + Berita	TF-IDF + BeritaTweet
TOP 1	87.33%	97.33%	97.55%(+**0.22**)	97.30%(-0.03)	97.35%(+0.02)
TOP 5	87.33%	97.33%	97.21%(-0.12)	97.12%(-0.21)	97.28%(-0.05)
TOP 10	87.33%	97.33%	97.16%(-0.17)	96.54%(-0.79)	97.27%(-0.16)

Tabel 8. Performance of the GloVe corpus on the GRU model

RANK	Akurasi (%)				
	Baseline	TF-IDF	TF-IDF + Tweet	TF-IDF + Berita	TF-IDF + BeritaTweet
TOP 1	87.33%	96.75%	96.70%(-0.05)	96.73%(-0.02)	96.85%(+**0.15**)
TOP 5	87.33%	96.75%	96.66%(-0.09)	96.69%(-0.06)	96.78%(+0.03)
TOP 10	87.33%	96.75%	96.79%(+0.04)	96.53%(-0.22)	96.80% (+0.05)

4. CONCLUSION

In this research, hoax detection has been built using GloVe feature expansion with Long Short Term Memory (LSTM) and Gate Recurent Unit (GRU) methods. GloVe itself is a feature expansion used to reduce vocabulary mismatches in sentences in the dataset. Feature expansion is done by building 3 GloVe corpus (Tweet. News and NewsTweet) and using Top 1, Top 5 and Top 10 from each corpus to find the best model, and the result using the top 1 tweet corpus and the LSTM model, with an accuracy value of 97.55% and in the expansion feature using the top 1 BeritaTweet corpus and the GRU model, with an accuracy value of 96.85%. Researchers also conducted model testing by determining the same data ratio in each model, namely 80:20, 70:30 and 90:10. This data ratio division is done to find the data ratio that produces the greatest accuracy, and the data ratio that has the greatest accuracy is 70:30. Then, the data ratio is used to find the accuracy of each model using TF-IDF, and it is found that the Long Short Term Memory model has greater accuracy than the Gate Recurent Unit model, which is 97.33%, while the Gate Recurent unit model has an accuracy of 96.75%. Suggestions for further research are to try using other feature extraction and feature expansion methods in order to produce higher accuracy.

REFERENCES

[1] M. P. Utami, O. D. Nurhayati, and B. Warsito, “Hoax Information Detection System Using Apriori Algorithm and Random Forest Algorithm in Twitter,” in 6th International Conference on Interactive Digital Media, ICIDM 2020, Dec. 2020. doi: 10.1109/ICIDM51048.2020.9339648.

Building of Informatics, Technology and Science (BITS)
Volume 4, No 4, Maret 2023 Page: 1815−1820
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i4.3084

[2] C. Juditha, "Interaksi Komunikasi Hoax di Media Sosial serta Antisipasinya Hoax Communication Interactivity in Social Media and Anticipation," 2018.
[3] A. Fauzi, E. B. Setiawan, and Z. K. A. Baizal, "Hoax News Detection on Twitter using Term Frequency Inverse Document Frequency and Support Vector Machine Method," in Journal of Physics: Conference Series, May 2019, vol. 1192, no. 1. doi: 10.1088/1742-6596/1192/1/012025.
[4] B. P. Nayoga, R. Adipradana, R. Suryadi, and D. Suhartono, "Hoax Analyzer for Indonesian News Using Deep Learning Models," in Procedia Computer Science, 2021, vol. 179, pp. 704–712. doi: 10.1016/j.procs.2021.01.059.
[5] F. Rahutomo, I. Y. R. Pratiwi, and D. M. Ramadhani, "Eksperimen Naïve Bayes Pada Deteksi Berita Hoax Berbahasa Indonesia," JURNAL PENELITIAN KOMUNIKASI DAN OPINI PUBLIK, vol. 23, no. 1, Jul. 2019, doi: 10.33299/jpkop.23.1.1805.
[6] K. Shu, L. Cui, S. Wang, D. Lee, and H. Liu, "Defend: Explainable fake news detection," in Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Jul. 2019, pp. 395–405. doi: 10.1145/3292500.3330935.
[7] E. Utami, A. F. Iskandar, W. Hidayat, A. B. Prasetyo, and A. D. Hartanto, "Covid-19 Hoax Detection Using KNN in Jaccard Space," IJCCS (Indonesian Journal of Computing and Cybernetics Systems), vol. 15, no. 3, p. 255, Jul. 2021, doi: 10.22146/ijccs.67392.
[8] C. S. Sriyano and E. B. Setiawan, "Pendeteksian Berita Hoax Menggunakan Naive Bayes Multinomial Pada Twitter dengan Fitur Pembobotan TF-IDF."
[9] J. Eka Sembodo, E. Budi Setiawan, and Z. Abdurahman Baizal, "Data Crawling Otomatis pada Twitter," Sep. 2016, pp. 11–16. doi: 10.21108/indosc.2016.111.
[10] Y. T. Zhang, L. Gong, and Y. C. Wang, "Improved TF-IDF approach for text classification," J Zhejiang Univ Sci, vol. 6 A, no. 1, pp. 49–55, Jan. 2005, doi: 10.1631/jzus.2005.A0049.
[11] A. Nurdin, B. Anggo, S. Aji, A. Bustamin, and Z. Abidin, "PERBANDINGAN KINERJA WORD EMBEDDING WORD2VEC, GLOVE, DAN FASTTEXT PADA KLASIFIKASI TEKS," Jurnal TEKNOKOMPAK, vol. 14, no. 2, p. 74, 2020.
[12] A. Aizawa, "An information-theoretic perspective of tf-idf measures q." [Online]. Available: www.elsevier.com/locate/infoproman
[13] J. Pennington, R. Socher, and C. D. Manning, "GloVe: Global Vectors for Word Representation." [Online]. Available: http://nlp.
[14] st Alfiyah Ramadian Jamaludin, "Deteksi Berita Hoax di Media Sosial Twitter dengan Ekspansi Fitur Menggunakan Glove."
[15] "Abstract Keywords-traffic flow prediction; LSTM; GRU; ARIMA A. Parameter Models."
[16] A. Khumaidi, R. Raafi, I. Permana Solihin, and J. Rs Fatmawati, "Pengujian Algoritma Long Short Term Memory untuk Prediksi Kualitas Udara dan Suhu Kota Bandung," Jurnal Telematika, vol. 15, no. 1.
[17] A. Hanifa, S. A. Fauzan, M. Hikal, and M. B. Ashfiya, "PERBANDINGAN METODE LSTM DAN GRU (RNN) UNTUK KLASIFIKASI BERITA PALSU BERBAHASA INDONESIA COMPARISON OF LSTM AND GRU (RNN) METHODS FOR FAKE NEWS CLASSIFICATION IN INDONESIAN." [Online]. Available: https://covid19.go.id/p/hoax-buster.
[18] B. P. Nayoga, R. Adipradana, R. Suryadi, and D. Suhartono, "Hoax Analyzer for Indonesian News Using Deep Learning Models," in Procedia Computer Science, 2021, vol. 179, pp. 704–712. doi: 10.1016/j.procs.2021.01.059.
[19] V. M. Patro and M. Ranjan Patra, "Augmenting Weighted Average with Confusion Matrix to Enhance Classification Accuracy," Transactions on Machine Learning and Artificial Intelligence, vol. 2, no. 4, Aug. 2014, doi: 10.14738/tmlai.24.328.
[20] X. Deng, Q. Liu, Y. Deng, and S. Mahadevan, "An improved method to construct basic probability assignment based on the confusion matrix for classification problem," Inf Sci (N Y), vol. 340–341, pp. 250–261, May 2016, doi: 10.1016/j.ins.2016.01.033.