JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

VOL. 10. NO. 1 AUGUST 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i1.5280

COMPARISON OF DEEP LEARNING METHODS ON SENTIMENT ANALYSIS USING WORD EMBEDDING

Rizal Gibran Aldrin Pratama[1]; Nuri Cahyono[2*]

Informatics[1,2]
Universitas Amikom Yogyakarta, Indonesia[1,2]
https://home.amikom.ac.id[1,2]
aldrin@students.amikom.ac.id[1], nuricahyono@amikom.ac.id[2*]

(*) Corresponding Author
(Responsible for the Quality of Paper Content)

Abstract— *According to ICW, corruption cases in Indonesia in the last 5 years have increased and the amount of losses suffered by the state from 2012-2022 reached Rp138.39 trillion. According to Transparency International, Indonesia's CPI ranking decreased in 2023 to 115 compared to 2022 at 110 out of 180 countries. These results show that the response to corruption is still slow and continues to deteriorate due to a lack of support from stakeholders. The purpose of this study is to test and compare the performance of deep learning model algorithms (RNN/LSTM/GRU/Bi-GRU/Bi-LSTM) on sentiment classification using word embedding, and obtain a model architecture that can determine the polarity of a sentence about public sentiment related to corruption in Indonesia, which can help governments, researchers, and practitioners in designing more effective anti-corruption strategies. The dataset used amounted to 1793 derived from crawling Twitter with 3 classes namely positive, negative and neutral. This research starts from data collection, preprocessing, word embedding, splitting the dataset which is divided into 80% training data and 20% test data, deep learning model testing, model evaluation and result representation. Word embedding uses word2vec with a dimension of 300. Based on the experimental results obtained, Bi-GRU has better performance than other architectural models with an accuracy value of 88%, precision 88.07%, recall 86.97% and f1-score 87.51%. The data used in this research is relatively small, it is recommended that future research can overcome it.*

Keywords: *deep learning, GRU, LSTM, sentiment analysis, word embedding*

Intisari— *Menurut ICW, kasus korupsi di Indonesia dalam 5 tahun terakhir mengalami peningkatan dan semakin besarnya kerugian yang dialami oleh negara sejak tahun 2012-2022 mencapai Rp138,39 triliun. Menurut Transparency International, peringkat CPI Indonesia menurun pada tahun 2023 menjadi 115 dibandingkan tahun 2022 di peringkat 110 dari 180 negara. Hasil tersebut menunjukkan bahwa respon terhadap korupsi masih lambat dan terus memburuk akibat dari kurangnya dukungan dari pemangku kepentingan. Tujuan dari penelitian ini menguji dan membandingkan kinerja algoritma model deep learning (RNN/LSTM/GRU/Bi-GRU/Bi-LSTM) terhadap klasifikasi sentimen menggunakan word embedding, serta mendapatkan arsitektur model yang dapat mengetahui polaritas pada suatu kalimat tentang sentimen publik terkait korupsi di Indonesia, yang dapat membantu pemerintah, peneliti, dan praktisi dalam merancang strategi anti-korupsi yang lebih efektif. Dataset yang digunakan berjumlah 1793 berasal dari crawling Twitter dengan 3 kelas yakni positif, negatif dan netral. Penelitian ini dimulai dari pengumpulan data, preprocessing, word embedding, splitting dataset yang dibagi menjadi data latih 80% dan data uji 20%., pengujian model deep learning, evaluasi model dan representasi hasil. Word embedding menggunakan word2vec dengan dimensi 300. Berdasarkan hasil percobaan diperoleh bahwa Bi-GRU memiliki kinerja yang lebih baik daripada model arsitektur lainnya dengan nilai akurasi 88%, precision 88,07%, recall 86,97% dan f1-score 87,51%. Data yang digunakan pada penelitian ini relatif kecil, disarankan penelitian mendatang dapat mengatasinya.*

Kata Kunci: *deep learning, GRU, LSTM, analisis sentimen, word embedding,*

INTRODUCTION

Corruption in Indonesia is old and widespread, and has disrupted the rule of law, economic growth, good governance, social welfare, and national and global economic competitiveness. According to Indonesia Corruption Watch (ICW), corruption cases in Indonesia in the last 5 years have increased and the amount of losses suffered by the state from 2012-2022 reached Rp138.39 trillion. According to Transparency International, Indonesia's Corruption Perceptions Index (CPI) in 2022 was 34/100 and ranked 110th out of 180 countries surveyed [1] and in 2023 with the same score ranked 115th. This score decreased by 4 points from 2021 and is the worst decline in the history of reform. Indonesia's CPI also ranked 5th as the most corrupt country in Southeast Asia in 2022 [2].

The stagnation of scores and the downward ranking of Indonesia's CPI in 2023 show that the response to corruption is still slow and continues to deteriorate due to a lack of support from stakeholders. Under the Rome Statute, corruption is a problem that is aligned with extraordinary crimes, such as terrorism, drug abuse, or serious environmental destruction. This is because it causes great damage to various sectors, leading to large state losses. Corruption is carried out in a systemic, complex and planned manner by state officials, starting from the lowest level to the highest positions in the government. It has taken away the basic rights of the people to obtain a decent livelihood or public service. It also threatens world order and has a negative impact on humanity. In other words, corruption creates an environment that allows extraordinary crimes to flourish and survive.

With the development of technology as it is today, the dissemination of various information can be more easily and quickly disseminated and accessed through online media. Currently, people use Twitter as an alternative platform to respond and express their opinions on a problem that occurs in the surrounding environment. The increasing number of Twitter users and more than 500 million tweets per day are sent [3]. Therefore, this study uses data collected from user uploads on twitter social media to find out the responses given by users regarding corruption that has occurred in Indonesia in recent years. One of them is sentiment analysis which is currently used by researchers for research.

Text sentiment analysis is an automated process for determining whether a segment of text contains objective or opinion content, and can then determine the polarity of the sentiment. Sentiment analysis is very useful, as it can collect and classify public opinion by analyzing large social data.

In research [4] using Naive Bayes and LSTM methods in sentiment analysis of Permendikbud No.30. Dataset as many as 2765 tweets, with negative and positive labels. After preprocessing the data becomes 471 data, with the division of training data and test data by 80% and 20%. Next, the weighting process is carried out using the TF-IDF method and continued with the calculation of the method. The results of this study obtained an accuracy value of 77% for LSTM and 76% for Naïve Bayess. In other research [5] regarding sentiment analysis of covid-19 tweets using word embedding with the LSTM, RNN and Naïve Bayes methods. The dataset amounted to 1364 data from twitter. After data collection, labeling is carried out with 2 positive and negative categories and then preprocessing is carried out. Researchers used a dimension of 300 for word embedding with word2vec. Classification is done with epoch 100, batch 30 and learning rate 0.0001. This research model has 81% accuracy for LSTM, 71% for RNN and 74% for Naïve bayes.

In research [6] analyzed movie review sentiment using Word2Vec and the LSTM method. Researchers use the CBOW and Skip-gram methods with the dimensions of the vector ranging from 50 to 500. Researchers divided the data into 80% training data and 20% testing data. The performance of LSTM obtained the best accuracy at a word vector dimension size of 100 with Skip-gram of 88.17% and obtained an accuracy of 87.68% at a word vector dimension size of 200 with CBOW. In research [7] with office product review datasets with five product rating classes from amazon, using the LSTM, CNN and GRU algorithms. Researchers do preprocessing by cleaning from empty data, then dividing the data into 80% training data and 20% testing data. The data is divided into "summary" and "overall" features for word embedding processing. Researchers using epoch 50 obtained an accuracy value of 77.96% for LSTM, 77.99% for CNN and 77.97% for GRU.

In research [8] using Bi-LSTM for sentiment analysis of Bali tourist destination reviews. Review data comes from the tripadvisor website. After preprocessing, the researchers conducted pretraining using word2vec with CBOW and hierarchical softmax evaluation method with 200 dimensions, then divided the data 80% training data and 20% testing data. Classification is done by researchers with a learning rate of 0.0001 and a dropout of 0.5. This research resulted in an accuracy value of 96.86%, precision 96.53%, recall 96.31%, f1 measure 96.41%. In research [9] LSTM for

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

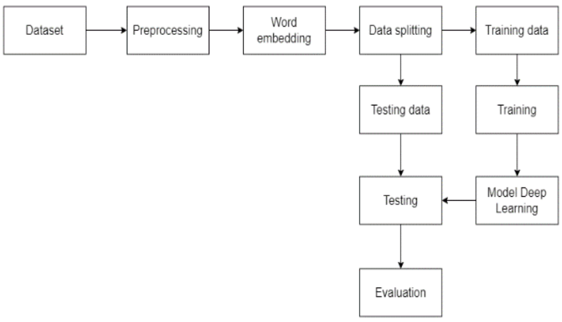
VOL. 10. NO. 1 AUGUST 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i1.5280

sentiment Covid-19 vaccine classification. The dataset is divided into 3 classes, positive, neutral and negative with a data distribution of 2564 data in each class. Researchers used the word2vec model with a dimension of 100. This research model uses batch 64 and 50 epochs for LSTM and Naïve Bayes. The test results of this study, LSTM has an accuracy value of 66%, precision 75%, recall 53%, and f1-score 54%, Naïve Bayes accuracy value of 61%, precision 58%, recall 60%, and f1-score 61%.

Based on the above problems, this research uses RNN, LSTM, GRU, Bi-LSTM and Bi-GRU to classify the sentiment of tweets about corruption in Indonesia. This research will produce a deep learning model architecture that can determine the polarity of a sentence with the aim of providing deeper insight into public sentiment regarding corruption in Indonesia, which can help the government, researchers, and practitioners in designing more effective anti-corruption strategies. This research also aims to test and compare the performance of deep learning model algorithms for sentiment classification using word embedding. The contribution of this research is to test the deep learning model in order to obtain the best model and the results of the model performance comparison in this study can be used as a reference for further research in developing Indonesian-based sentiment classification models.

MATERIALS AND METHODS

This research, focusing on sentiment classification related to corruption issues in Indonesia, tests and compares the performance of the RNN, LSTM, GRU, Bi-LSTM and Bi-GRU algorithms using word embedding. This research has several stages which can be seen in Figure 1. This research starts from data collection, preprocessing, word embedding using word2vec, splitting datasets, model testing using deep learning algorithms, model evaluation and result representation.



Source: (Research Results, 2024)
Figure 1. Research Procedure

A. Dataset

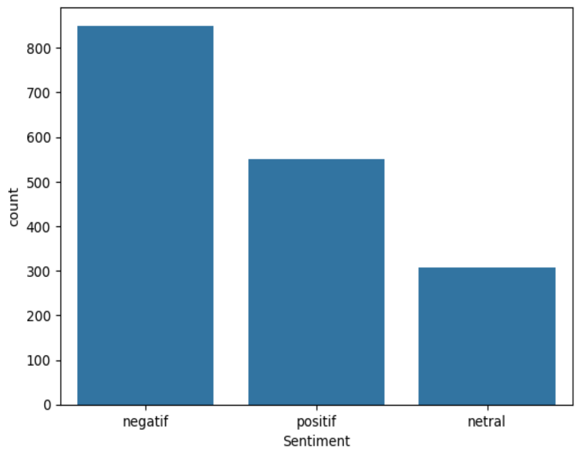
The dataset is taken from a collection of tweets derived from crawling on Twitter tweet-harvest, with the keywords "corruption" and #corruption taken specifically in September 2023. Data retrieval from Twitter is done several days, due to the limitation of data retrieval in a day, which has been set by Twitter, depending on the type of account owned. This causes duplication of tweets in the data. This problem can be overcome at the data preprocessing stage.

Table 1. Crawling Result

Create_at	Tweet	Username
Mon Dec 07 05:19:16 +0000 2020	In Zimbabwe, let alone what is halal, even what is haram is hard to come by and fiercely contested. #Corruption	huzprince90
Sun Mar 21 06:21:44 +0000 2021	This country ðŸ‡²ðŸ‡¨ is no longer respected on the international stage, perhaps due to too much news about injustice and #Corruption #Corruptors.	Anom1945
Sat Jan 30 12:06:51 +0000 2021	Oh God! Keep me away from the misleading devil and the cursed corruptor! #SaturdayNight #Corruption #Corruptors #Devil	satiristikid

Source: (Research Results, 2024)

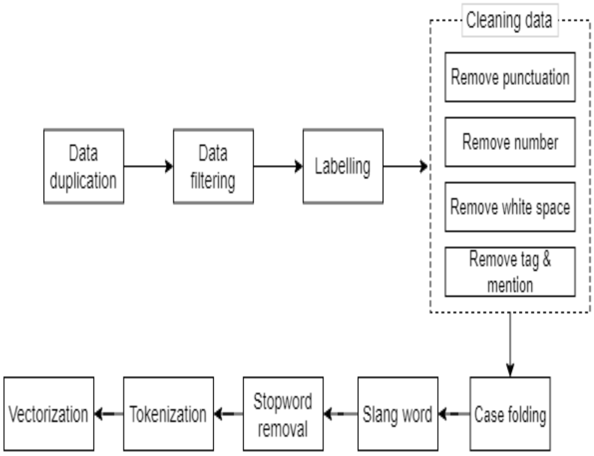
Based on the above references and the limitation on the number of tweets that can be taken by twitter, a dataset of 1793 tweets. Examples of tweets crawled from Twitter are shown in Table 1 and the overall result of the labeled dataset is illustrated in Figure 2. After the cleaning stage, the remaining data amounted to 1707 tweets with a composition of 849 negative tweets, 550 positive tweets, and 308 neutral tweets.



Source: (Research Results, 2024)
Figure 2. Sentiment Labeling Result

B. Preprocessing Data

Data preprocessing techniques can be performed to transform data into structured data considering that data from Twitter is usually unstructured data that will be difficult for algorithms to classify. To clean tweets from noise and spelling errors, a series of preprocessing steps are required. The purpose of performing this process is to clean the data and improve the quality of the data when used to build sentiment analysis models [10]. Figure 3 shows the stages of preprocessing. This stage ensures that the data conforms to the required standards. Table 2 shows the results of one of the sample datasets that underwent the preprocessing stage.



Source: (Research Results, 2024)
Figure 3. Preprocessing Stage

After crawling the dataset, the preprocessing stage starts from data duplication. Deleting duplicate tweet data is done by identifying and checking for duplicates in the dataset, then collecting data that contains duplicates, then deleting duplicates in the tweet column, because what is used later is only the tweet column in the crawling dataset results. Next, data filtering is done by deleting incomplete or empty data rows after data duplication. After the process, then manually labeling the dataset with positive, negative, and neutral classes.

The next stage is data cleaning, this process is done to clean things that are not needed such as numbers, symbols or punctuation marks, emoticons, url addresses, twitter mentions etc. Data cleansing is done so that the data is cleaner from symbols and words. Furthermore, the Case Folding process is used to change all capital letters into lower case letters. This is done without exception at the beginning of sentences, naming people, cities etc. Some Twitter users in Indonesia use non-standard language (slang) in making tweets [11]. To overcome this, each slang word must be restored to its original word form by using a previously published slang dictionary. Slang words reflect variations in everyday language that may not conform to formal or official language.

In the tokenization process, spaces are used as separators between words. In this research, the separating spaces in phrases are removed. Tokenization is the process of breaking down text into smaller units called "tokens." Tokens can be words, sub-words, phrases, or symbols that have meaning in the context of the language [12]. The aim is to transform the text into a more structured and compartmentalized form, thus facilitating further processing.

The last stage of preprocessing is vectorization. Vectorization converts these words into numerical representations for further analysis, such as model training or statistical calculations. For a model to process a sentence in natural language processing studies, a word must be converted into vector form.

Table 2. Preprocessing Stage Results

Process	Result
Cleaning	This country is no longer respected on the international stage, perhaps due to too much news about injustice and corruption.
Case folding	This country is no longer respected on the international stage, perhaps due to too much news about injustice and corruption.
Slang word	This country is no longer respected on the international stage, perhaps due to too much news about injustice and corruption.
Stopword removal	The country is respected on the international stage, but news of injustice, corruption, and corruptors persists.
Tokenizing	['country', 'respected', 'on the stage', 'internasional', 'news', 'injustice', 'fairness ', ' corruption ', ' corruptors ']
Stemming	The country price event on the international stage news unfairness corruption corruptors.

Source: (Research Results, 2024)

C. Word Embedding

Word embedding is a distributed representation consisting of word properties in a vector of real numbers that retrieve syntactic features and semantic word relationships [13]. One of the models that can be used to generate word embedding with length N dimensions is word2Vec.
Word2Vec attempts to understand the meaning of words based on their context of use and provides a vector representation that reflects the semantic relationship between words. Word2Vec has the ability to generalize and handle words that are not

VOL. 10. NO. 1 AUGUST 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i1.5280

seen during training. This allows the model to provide a good representation for new words [14].

The use of Word2Vec also speeds up the training process, improves the accuracy of deep learning algorithms and has an influence on the performance of deep learning models in performing sentiment classification and the resulting word embedding has lower dimensions, but still retains meaning information [15]. Word2Vec developed by Mikolov et al. (2013) is one of the continuous learning tools to generate word embedding. The word2Vec algorithm is a combination of two techniques, namely the Continuous Bag Of Words (CBOW) and Skip-gram model. Both techniques are basic neural networks used to map words or words to target variables which can also be words or sentences [16].

At this stage, a word dictionary is created that is taken from a collection of articles on the Indonesian wikipedia dump page using the gensim library. Each article from Wikipedia is taken and written to a text file (id-wiki_dump_lower.txt). With a word vector dimension of 300, using the CBOW technique, epoch 10 and the maximum distance between the current and predicted words in a sentence (windows size) 5. The results of the .txt file are used as input for the word2vec model. Each word is given a unique index and used to convert words to numerical representations. The model generates a word dictionary from the dataset into a vector form with files in the form of word vectors, which are later used for the embedding layer of the deep learning model.

D. Classification

The training process in this study uses a device with core i5 9300H processor specifications with 8 cores CPU, Nvidia GeForce GTX 1650 and 8 GB RAM. The purpose of algorithm comparison in this research is due to the differences between RNN, LSTM and GRU. RNN consists of an array of interconnected, feed-forward neurons for sequence and data modeling whose output depends on the previous state. LSTMs include memory cells and have 3 gates that control the flow of information, which can capture long-term dependencies and learn complex patterns on long sequences. GRU has 2 gates that are simpler than LSTM, the reset gate and the update gate. Without sacrificing the ability to understand long contexts and more efficient in terms of the number of parameters.

The input sequence is a series of words that have been converted into numeric digits, then each digit is converted into a vector in the embedding layer to become input to the Recurrent Layers model (RNN/LSTM/GRU/Bi-LSTM/Bi-GRU). The word embedding used by the word2vec model uses the Continuous Bag Of Words (CBOW) technique [17] with a dimension value of 300. To test the performance of the word embeddings, the model was created for each test scenario. The dataset obtained was divided into 2 parts, training data and testing data with a distribution of 80:20. The number of epoch values used during the training process amounted to 5, with a combination of values 5, 10, 15, 20 and 25. The model was built using the TensorFlow framework and Keras. Furthermore, the output of the recurrent layers is used as input to the dense layer.

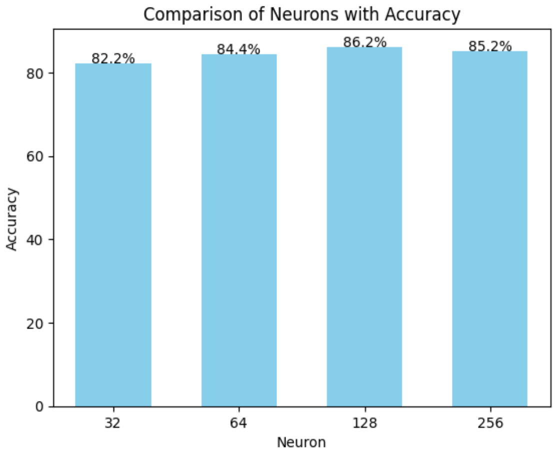
We then added a dropout layer, a technique commonly used for regularization purposes and to avoid overfitting. In fact, by using dropout, a number of neurons are randomly selected, to be ignored during training which has a great impact on model performance [18]. Then entering the last layer, a dense layer consisting of 3 neurons with Softmax activation function is used since the dataset has 3 classes. The Softmax function converts a vector of values into a probability distribution. In general, the way it claims this function will calculate the probability of each target class over all possible target classes. Later, the calculated probabilities are helpful for determining the target class for a given input. The biggest advantage of using Softmax is the range of output probabilities [19].

For bidirectional, the first layer uses ReLu activation and added a dropout layer, using ReLu activation function which serves to reduce vanishing gradient, bring non-linearity to the model and the operation is simple [20]. And the last layer uses the same Softmax activation function as the algorithm without using bidirectional. The weights of each neuron are optimized during the training process using Adam, because it can make the model achieve optimal weight values in a short number of training iterations. Classification is performed using pre-trained embedding using 5 algorithms consisting of RNN, LSTM, GRU, Bi-LSTM, and Bi-GRU. Based on research [17] using a learning rate value of 0.005 is the most appropriate to improve the accuracy of deep learning models using pre-trained word embedding. Therefore, in this study we used a learning rate of 0.005 for each algorithm. Model performance during training is monitored using early stopping with a patience value of 3. Early Stopping monitors the validation loss during the training process. If early stopping finds that the validation loss increases for 3 iterations during training, then the training model will be stopped. The batch size value used is 32, based on research [21] the smaller the batch size or the amount of sample data propagated through the network in one iteration, the accuracy tends to increase.

RESULTS AND DISCUSSION

A. Neuron Testing

This test is to determine the neuron value in a certain range that is used to get the optimal value for the results of this deep learning method. By changing the neuron value using the LSTM algorithm, 5 trials were conducted for each neuron value. Starting with a neuron value of 32 with 4 trials with different neuron values. The neuron values used in this test are 32, 64, 128 and 256. Figure 4 shows the results of this test.



Source: (Research Results, 2024)
Figure 4. Neuron Testing

Figure 4 shows the highest accuracy obtained from this test is obtained from 128 neurons with an accuracy of 86.2%. The test was conducted with batch size parameter 32, epoch value 10 and learning rate 0.005. For more complete results can be seen in Table 3.

Table 3. Neuron Testing

Neuron	Loss	Acc (%)	Pres (%)	Recall (%)	F1 (%)
32	0,5305	82,22	82,33	82,39	82,17
64	0,5397	84,44	84,38	84,5	84,33
128	0,5048	86,2	86,04	86,1	85,91
256	0,5594	85,2	85,13	85,21	85,05

Source: (Research Results, 2024)

B. Epoch Value Testing

This stage of testing to determine the epoch value is used to get optimal results. By changing the epoch value using the LSTM algorithm, 5 trials were conducted for each epoch value. Starting with an epoch value of 5 with an interval of 5, as many as 5 trials with different epoch values. Table 5 shows the results of the test. Table 4 shows that the highest accuracy is obtained with an epoch value of 25. Based on the accuracy value obtained, the greater the epoch value used, the higher the accuracy obtained.

Table 4. Epoch Testing

Epoch	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
5	80,93	80,5	81,08	80,89
10	85,19	85,14	85,21	85,13
15	85,41	85,34	85,43	85,29
20	85,56	85,49	85,6	85,45
25	86,07	85,97	86,13	85,99

Source: (Research Results, 2024)

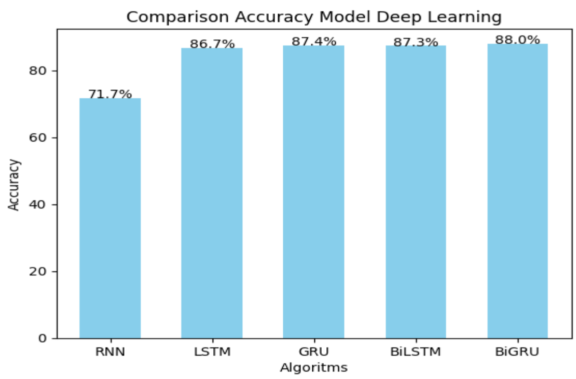
C. Model Comparison

Figure 5 shows that of the proposed deep learning models, the Bidirectional Gated Recurrent Unit (GRU) model gets the highest accuracy and is superior to the other four models which are also variants of Recurrent Neural Network (RNN). The accuracy obtained is 88% which is better than Bi-LSTM. Likewise, with the model without bidirectional, the GRU model gets superior accuracy compared to the LSTM model which is 87.44%. RNN itself obtained the lowest accuracy value compared to its variant models, because the GRU and LSTM models improved from the RNN model in order to get better results. It can be seen from the value of the confusion matrix in Table 5, Bi-GRU gets the highest value of the overall average value, it's just that the recall value is higher than the GRU model without bidirectional.

Table 5. The Evaluation Comparison

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
RNN	71,60	75,68	64,92	69,89
LSTM	86,71	86,93	86,26	86,60
GRU	87,44	87,90	87,15	87,43
BiLSTM	87,33	87,87	86,89	87,37
BiGRU	88,00	88,07	86,97	87,51

Source: (Research Results, 2024)



Source: (Research Results, 2024)
Figure 5. Evaluation Comparison Chart

VOL. 10. NO. 1 AUGUST 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i1.5280

CONCLUSION

From the research that has been done, the variant of RNN has good performance, able to exceed the RNN itself. The LSTM model has a difference of 15.11% from the RNN model. From these results LSTM overcomes the missing gradient problem associated with RNN, as well as GRU which offers an efficient alternative to LSTM. Judging from the results of the research that has been done, the performance of GRU can outperform the performance of LSTM both with bidirectional and without. The highest accuracy obtained by the Bi-GRU model is 88%, with precision 88.07%, recall 86.97% and f-measure 87.51%. Then GRU with an accuracy of 87.44%, precision 87.90%, recall 87.15% and f-measure 87.51%. Followed by the Bi-LSTM model which has better accuracy than LSTM by 0.62%, which has an accuracy of 87.33% while LSTM is 86.71%. Compared with the existing results, the proposed method has better feature extraction ability and variant model of RNN and better text sentiment classification effect. At the same time, the method of this paper also has some shortcomings, because the deep learning method requires a large amount of data, the data used for research is relatively small.

REFERENCE

[1] Transparency International, "Corruption Perception Index (CPI) Indonesia," Corruption Perception Index. Accessed: Feb. 01, 2024. [Online]. Available: https://www.transparency.org/en/countries/indonesia

[2] Databoks, "Indonesia Negara Terkorup ke-5 di Asia Tenggara pada 2022." [Online]. Available: https://databoks.katadata.co.id/datapublish/2023/02/02/indonesia-negara-terkorup-ke-5-di-asia-tenggara-pada-2022

[3] Ying Lin, "10 Twitter Statistics Every Marketer Should Know In 2023 [Infographic]," Oberlo. [Online]. Available: https://www.oberlo.com/blog/twitter-statistics

[4] Y. Romadhoni, K. Fahmi, and H. Holle, "Analisis Sentimen Terhadap PERMENDIKBUD No.30 pada Media Sosial Twitter Menggunakan Metode Naive Bayes dan LSTM," *Jurnal Informatika: Jurnal pengembangan IT (JPIT)*, vol. 7, no. 2, pp. 118-124, 2022, doi: 10.30591/jpit.v7i2.3191.

[5] M. Z. Rahman, Y. A. Sari, and N. Yudistira, "Analisis Sentimen Tweet COVID-19 menggunakan Word Embedding dan Metode Long Short-Term Memory (LSTM)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 11, pp. 5120–5127, 2021, [Online]. Available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/10188.

[6] W. Widayat, "Analisis Sentimen Movie Review menggunakan Word2Vec dan metode LSTM Deep Learning," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 3, p. 1018, Jul. 2021, doi: 10.30865/mib.v5i3.3111.

[7] M. Totox and H. F. Pardede, "Exploring the Effectiveness of Deep Learning in Analyzing Review Sentiment," *JIKO (Jurnal Informatika dan Komputer)*, vol. 6, no. 2, pp. 116-121, Aug. 2023, doi: 10.33387/jiko.v6i2.6372.

[8] D. I. Af'idah, D. Dairoh, S. F. Handayani, R. W. Pratiwi, and S. I. Sari, "Sentimen Ulasan Destinasi Wisata Pulau Bali Menggunakan Bidirectional Long Short Term Memory," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 21, no. 3, pp. 607–618, Jul. 2022, doi: 10.30812/matrik.v21i3.1402.

[9] M. Ihsan, Benny Sukma Negara, and Surya Agustian, "LSTM (Long Short Term Memory) for Sentiment COVID-19 Vaccine Classification on Twitter," *Digital Zone: Jurnal Teknologi Informasi dan Komunikasi*, vol. 13, no. 1, pp. 79–89, May 2022, doi: 10.31849/digitalzone.v13i1.9950.

[10] M. F. Karaca, "Effects of preprocessing on text classification in balanced and imbalanced datasets," *KSII Transactions on Internet and Information Systems*, vol. 18, no. 3, pp. 591–609, Mar. 2024, doi: 10.3837/tiis.2024.03.004.

[11] E. Utami, I. Oyong, S. Raharjo, A. Dwi Hartanto, and S. Adi, "Supervised learning and resampling techniques on DISC personality classification using Twitter information in Bahasa Indonesia," *Applied Computing and Informatics*, Sep. 2021, doi: 10.1108/ACI-03-2021-0054.

[12] K. K. Agustiningsih, E. Utami, and M. A. Alsyaibani, "Sentiment Analysis of COVID-19 Vaccines in Indonesia on Twitter Using Pre-Trained and Self-Training Word Embeddings," *Jurnal Ilmu Komputer dan Informasi*, vol. 15, no. 1, pp. 39–46, Feb. 2022, doi: 10.21609/jiki.v15i1.1044.

[13] D. Dessi, R. Helaoui, V. Kumar, D. R. Recupero, and D. Riboni, "TF-IDF vs Word Embeddings for Morbidity Identification in Clinical Notes: An Initial Study," *arXiv preprint arXiv:2105.09632*, May. 2021, doi: 10.5281/zenodo.4777594.
[14] H. Juwiantho *et al.*, "Sentiment Analysis Twitter Bahasa Indonesia Berbasis Word2Vec Menggunakan Deep Convolutional Neural Network," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 7, no. 1, pp. 181–188, Feb. 2020, doi: 10.25126/jtiik.202071758.
[15] D. Intan Af *et al.*, "Pengaruh Parameter Word2Vec terhadap Performa Deep Learning pada Klasifikasi Sentimen," *JPIT (Jurnal Informatika: Jurnal Pengembangan IT)*, vol. 6, no. 3, pp. 156-161, Oct. 2021, doi: 10.30591/jpit.v6i3.3016.
[16] H. Xia, "Continuous-bag-of-words and Skip-gram for word vector training and text classification," in *Journal of Physics: Conference Series*, Institute of Physics, vol. 2634, pp 012052, Nov. 2023. doi: 10.1088/1742-6596/2634/1/012052.
[17] M. Aydoğan and A. Karci, "Improving the accuracy using pre-trained word embeddings on deep neural networks for Turkish text classification," *Physica A: Statistical Mechanics and its Applications*, vol. 541, p.123288, Mar. 2020, doi: 10.1016/j.physa.2019.123288.
[18] M. U. Salur and I. Aydin, "A Novel Hybrid Deep Learning Model for Sentiment Classification," *IEEE Access*, vol. 8, pp. 58080–58093, 2020, doi: 10.1109/ACCESS.2020.2982538.
[19] G. S. N Murthy, S. Rao Allu, B. Andhavarapu, M. Bagadi, and M. Belusonti, "Text based Sentiment Analysis using LSTM," *International Journal Of Engineering Research & Technology (IJERT)*, vol. 09, no. 05, pp. 299-303, May 2020, [Online]. Available: www.ijert.org
[20] N. R. Bhowmik, M. Arifuzzaman, and M. R. H. Mondal, "Sentiment analysis on Bangla text using extended lexicon dictionary and deep learning algorithms," *Array*, vol. 13, p.100123, Mar. 2022, doi: 10.1016/j.array.2021.100123.
[21] A. Hikmah, S. Adi, and M. Sulistiyono, "The Best Parameter Tuning on RNN Layers for Indonesian Text Classification," in *2020 3rd International Seminar on Research of Information Technology and Intelligent Systems, ISRITI 2020*, Institute of Electrical and Electronics Engineers Inc., Dec. 2020, pp. 94–99. doi: 10.1109/ISRITI51436.2020.9315425.

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri