

Sentiment Analysis Of Energy Independence Tweets Using Simple Recurrent Neural Network

Kurnia Muludi¹, Mohammad Surya Akbar^{*2}, Dewi Asiah Shofiana³, Admi Syarif⁴

^{1,2,3,4}Department of Computer Science, FMIPA Universitas Lampung, Lampung, Indonesia
e-mail: ¹kmuludi@yahoo.com, ^{*2}suryaakbar675@gmail.com, ³dewi.asiah@fmipa.unila.ac.id,
⁴admi.syarif@fimpa.unila.ac.id

Abstrak

Analisis sentimen adalah bagian dari penelitian komputasi yang mengekstrak data tekstual untuk mendapatkan nilai positif, atau negatif, terkait suatu topik. Dalam penelitian terbaru, data biasanya diperoleh dari media sosial, termasuk Twitter, di mana pengguna sering memberikan pendapat pribadi mereka tentang subjek tertentu. Kemandirian energi pernah menjadi trending topic di Indonesia, karena pendapatnya yang beragam, pro dan kontra, menarik untuk dianalisis. Pembelajaran mendalam adalah cabang pembelajaran mesin yang terdiri dari lapisan tersembunyi jaringan saraf dengan menerapkan transformasi non-linier dan abstraksi model tingkat tinggi dalam database besar. Jaringan saraf berulang (RNN) adalah metode pembelajaran mendalam yang memproses data berulang kali, terutama cocok untuk tulisan tangan, data multi-kata, atau pengenalan suara. Penelitian ini membandingkan tiga algoritma: Simple Neural Network, Bernoulli Naive Bayes, dan Long Short-Term Memory (LSTM) dalam analisis sentimen menggunakan data kemandirian energi dari Twitter. Berdasarkan hasil penelitian, Simple Recurrent Neural Network menunjukkan kinerja terbaik dengan nilai akurasi 78% dibandingkan dengan nilai Bernoulli Naive Bayes 67% dan LSTM dengan nilai akurasi 75%.

Kata kunci—Analisis Sentimen, Simple RNN, LSTM, Bernoulli Naive Bayes, Kemandirian Energi

Abstract

Sentiment analysis is part of computational research that extracts textual data to obtain positive, or negative values related to a topic. In recent research, data are commonly acquired from social media, including Twitter, where users often provide their personal opinion about a particular subject. Energy independence was once a trending topic discussed in Indonesia, as the opinions are diverse, pros and cons, making it interesting to be analyzed. Deep learning is a branch of machine learning consisting of hidden layers of neural networks by applying non-linear transformations and high-level model abstractions in large databases. The recurrent neural network (RNN) is a deep learning method that processes data repeatedly, primarily suitable for handwriting, multi-word data, or voice recognition. This study compares three algorithms: Simple Neural Network, Bernoulli Naive Bayes, and Long Short-Term Memory (LSTM) in sentiment analysis using the energy independence data from Twitter. Based on the results, the Simple Recurrent Neural Network shows the best performance with an accuracy value of 78% compared to Bernoulli Naive Bayes value of 67% and LSTM with an accuracy value of 75%.

Keywords— Sentiment Analysis, Simple RNN, LSTM, Bernoulli Naive Bayes, Energy Independence

1. INTRODUCTION

The internet is currently one of the mandatory electronic needs of today's society. Most people use it as a source of information, communication, and social media to connect with others. The Indonesian Internet Service Providers Association (APJII) revealed that internet users in Indonesia reached 196.71 million people from 2019 to June 2020. It is common for people to express their opinions on social media, and one of the most popular public discussions is about energy independence. Energy independence is one of the sustainable development plans of the government of the Republic of Indonesia for energy distribution in areas lacking in good energy sources. [1]

One of the social media that is used as a place to express opinions is Twitter. Twitter is a social media from the United States with unique characteristics and formats with symbols or specific format because Twitter users can only send and read messages with a maximum limit of 140 characters, which is known as a tweet [2]. Tweets can be in the form of posts with text or images, sometimes some people express opinions, suggestions, or criticisms on a specific topic that vary from Twitter users.

Sentiment analysis is a branch of text mining that identifies text and then extracts information from the identified text into subjective information [3]. In addition to extracting information, sentiment analysis also focuses on opinion managers that contain polarity, namely the value of positive or negative sentiments from a text [4]. Problems in sentiment analysis are usually difficult to define, including the concept of problems, subproblems, and objectives that serve as a performance framework in various studies [5].

In data collection, sentiment analysis uses various algorithms originating from the branch of Artificial Intelligence, such as the Machine Learning algorithm consisting of Naive Bayes, Support Vector Machine (SVM), Decision Trees, and regression, and the Deep Learning algorithm which consists of a Recurrent Neural Network (RNN), Deep Neural Network (DNN), Deep Belief Network (DBN), and Convolutional Neural Network (CNN). There are many studies related to sentiment analysis using the Naive Bayes algorithm. The advantage of the Naive Bayes algorithm is that it does not require a large amount of data, minimal process of training data since it is based on the probability value, and its efficiency in the calculation process [6].

Recurrent Neural Network (RNN) is an algorithm within the Deep Learning family. RNN has made rapid progress and revolutionized fields such as natural language processing, time-series data processing, and so on [7]. Deep learning is suitable for sentiment analysis as it is part of Natural Language Processing (NLP) [8]. The Recurrent Neural Network algorithm has various developing methods such as the Simple Recurrent Neural Network and Long Short-Term Memory (LSTM).

Several previous studies regarding sentiment analysis by applying deep learning has been conducted. Zulfa and Winarko [9] conducted an analysis of the sentiment of Indonesian-language tweets. Research conducted by Zulfa and Winarko used an algorithm that is included in the Deep Learning category, namely the Deep Belief Network (DBN). DBN uses a layer or stack method of several algorithms with feature extraction that makes the most of all resources. This study indicated a high accuracy value for sentiment analysis, reaching 93.31%.

Based on this explanation, this research will apply deep learning to analyze and classify tweet data related to the topic of energy independence in Indonesia. The classification results will be used to compare the classification values (accuracy, precision, recall, and f1-score) between three algorithms, namely Simple Recurrent Neural Network and Long Short Term

Memory - Naive Bayes algorithm, namely Bernoulli Naive Bayes.

2. METHODS

Figure 1 shows the research flow starting from Twitter data collection, data preprocessing, sentiment classification, and evaluation of the model. The sentiment classification will apply three algorithms: Simple Recurrent Neural Network, Long Short Term Memory, and Bernoulli Naive Bayes.

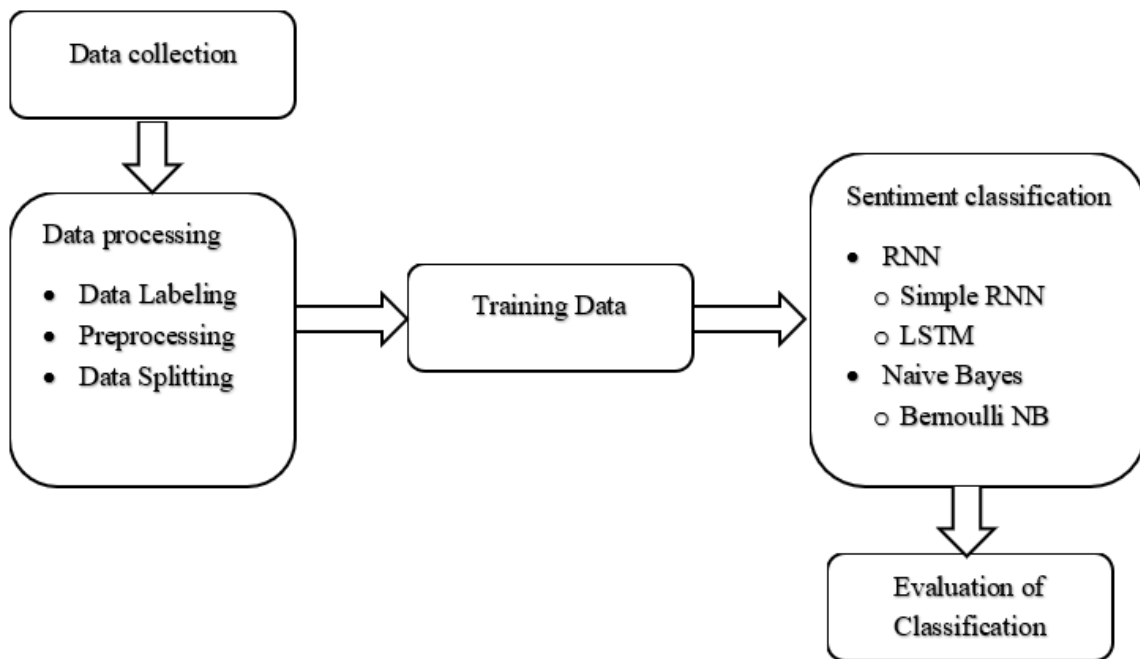


Figure 1. Research Stages

2.1. Data Collection

Data collection or data crawling is carried out using the Twitter API, which is an application created by Twitter to make it easier for developers to access Twitter web information. The data collected from Twitter in this study is related to energy independence in Indonesia. Data is collected using the Python Tweepy library and Twitter API in the form of text data from tweets data. Before collecting data, it is necessary to check the consumer key and access token so that the system can retrieve data from Twitter. The data are then stored in a CSV file automatically according to the search.

2.2. Data Processing

The data obtained from Twitter will be labelled as positive or negative manually, depends whether the text data contains pros (positive) or cons (negative). In addition, data preprocessing is also carried out to remove URLs, punctuation marks, special characters, and symbols of the alphabet, which is called the tokenization process.

2.2.1. Data Labelling

Data labelling is a process to assign a value or class to data. The data that has been previously collected do not have positive or negative classes because Twitter does not provide such classes.

2.2.2 Preprocessing

Data preprocessing is needed to obtain clean data and increase the value of tweet classification. The steps taken when preprocessing are text normalization, text tokenization, and stopword removal. The functions of the preprocessing stage are as follows:

2.2.2.1. Text Normalization

The aim of text normalization is to improve the structure of words and vocabulary in sentences. In this study, normalization will be carried out on all data collected. The text normalization process includes deleting URLs, punctuation marks, special characters, and symbols in the form of alphabet letters so that there is no ambiguity in the text data.

2.2.2.2. Text Tokenization

Text tokenization is a process of cutting strings or input sentences based on each word that makes up the sentence in the text data. Tokenization broadly breaks down a set of characters in a text into word units, distinguishing certain characters that can be treated as word separators.

2.2.2.3. Stopword Removal

Stopwords are a collection of words that often appear but will not change the sentence's meaning if it is deleted, or in other words, deleting a meaningless or irrelevant word in the sentence.

2.2.3. Data Splitting

After data preprocessing is conducted, the data is randomly divided into two for classification: 80% training data and 20% test data.

2.3. Training Data

The data training process was carried out using Word2vec, followed by classification using three algorithms: Simple Recurrent Neural Network (Simple RNN), Long Short-Term Memory (LSTM), and Bernoulli Naive Bayes.

2.3.1. Word2vec

Word2vec will train the dataset and obtain the vector value of the trained word. It will also acquire the number of total words and the vocabularies that appears from the dataset training process. CBOW (Continuous Bag of Words) model is using for word2vec model training because this model uses continuous distributed representation of the context [10]. For architecture CBOW model is shown at Figure 2.

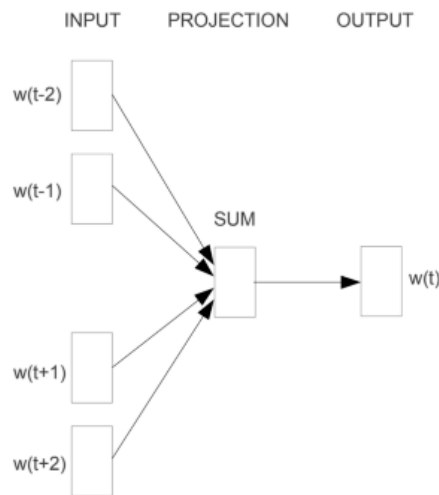


Figure 2. CBOW Model

2.4. Sentiment Classification

The sentiment classification process was carried out using divided test data. The classification will be carried out using three algorithmic models: Simple Recurrent Neural Network (Simple RNN), Long Short-Term Memory (LSTM), and Bernoulli Naive Bayes. The results of the classification are in the form of confusion matrix values which then get the accuracy, precision, recall and f1-score values.

2.4.1. Simple Recurrent Neural Network

The Simple Recurrent Neural Network method performs data training repeatedly, as seen in Figure 3. This Deep Learning algorithm has an input layer, a context layer with a hidden layer, and an output layer [11].

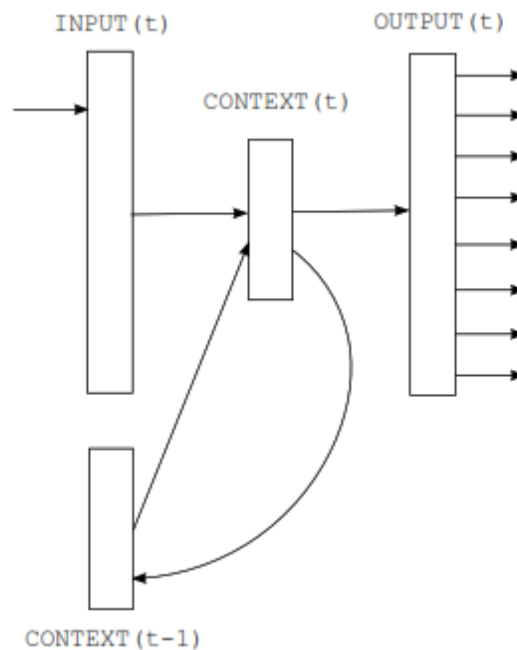


Figure 3. Simple Recurrent Neural Network

In Figure 3, we can see the flow of the Simple Recurrent Neural Network. The number of times this method conducts training depends on the number of epochs. Epoch is a parameter that determines how many times an algorithm or method runs, specifically for algorithms related to Deep Learning or Neural Networks. The Simple Recurrent Neural Network in this study uses 12 epochs for training data before classifying them.

2.4.2. Long Short Term Memory

In addition to Simple Recurrent Neural Network, Long Short-Term Memory (LSTM) also includes training data before classification. For training, the LSTM training is carried out through several layers (done repeatedly) with the provisions of the epoch value. The LSTM training also uses 12 epochs. LSTM is an artificial Recurrent Neural Network (RNN) architecture, part of Deep Learning. LSTM is well-suited for classifying, processing, and making predictions based on time series data [12].

2.4.3. Bernoulli Naïve Bayes

Classifying data in Bernoulli Naive Bayes applies the probability concept, in which the data acts as samples [13]. In the Bernoulli Bayes Model, features are independent Booleans (binary variables) describing the inputs. Like the Multinomial Naive Bayes model, this model is notable for document classification tasks where binary term occurrence features are used rather than term frequencies.

2.4.4. Confusion Matrix Evaluation

The results of each classification algorithm are evaluated using the confusion matrix. The accuracy, precision, recall, and f1-score values are derived from the confusion matrix. Confusion matrix table is shown in Table 1.

Table 1. Confusion matrix

True Label	Prediction Label	
	Negative	Positive
Negative	TN (True Negative)	FP (False Positive)
Positive	FN (False Negative)	TP (True Positive)

Confusion matrix in classification problem with two classes is shown in Table 1. It can be concluded from the figure that there are possible four different results predictions. Really positive and really negative outcomes are correct classification, while the false positive and false negative outcomes are two possible types of errors [14]. True Negative (TN) value is the number of negative data detected correctly, while False Negative (FN) is positive data detected negative. True Positive (TP) values are positive data that is detected correctly, while False Positive (FP) is negative data but detected positive. The confusion matrix table is used to measure the performance of a classification method by calculating the value of accuracy, precision, recall, and f1-score.

- a. Accuracy is the ratio of true predictions (positive and negative) to the overall data.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (1)$$

- b. Precision is the ratio of true positive predictions to overall positive predictions.

$$precision = \frac{TP}{TP + FP} \times 100\% \quad (2)$$

- c. Recall is the ratio of true positive predictions to the overall true positive data.

$$recall = \frac{TP}{TP + FN} \times 100\% \quad (3)$$

- d. F1-Score is a weighted average comparison of precision and recall.

$$f1\ score = 2 \times \frac{(recall \times precision)}{(recall + precision)} \quad (4)$$

3. RESULTS AND DISCUSSION

This section will explain the results of the classification of sentiment analysis from three methods, namely the Simple Recurrent Neural Network, Long Short-Term Memory, and Bernoulli Naïve Bayes. In addition to the discussion of classification results, it will also discuss the results of the previously mentioned training data. There are 3000 data in total for this study, divided into training data and test data. Details of the data split provided in Table 2.

Table 2. Details of data splitting

Class	Training Data	Test Data
Positive	1548	405
Negative	852	195
Total	2400	600

Before training the data, all tweets (data) are manually labelled based on the type of sentiment, categorized as positive or negative. For example, a tweet "it is impossible for energy independence since resources is not evenly distributed" will be labelled with a number 0 or negative class because it has a negative sentiment value that states disbelief about energy independence.

The Word2vec process resulted in a total of 3392 words from the training data of 2400 texts. In addition, 608 vocabularies will act as training data are obtained, along with the detection of 758 stopwords are then removed.

After the sentiment analysis classification is carried out, the models are then evaluated using the confusion matrix. We can determine the accuracy, precision, recall, and f1-score values for each sentiment analysis classification algorithm from this matrix. The confusion matrix value is the value of the number of tweets that matches the prediction of the calculation results from each classification algorithm with the class label. The more tweet data matches the prediction of classification with data labelling, the higher the accuracy, precision, recall, and f1-score values are obtained. The confusion matrix from each sentiment classifier is given in Table 3, 4, and 5.

Table 3. Confusion matrix Simple Recurrent Neural Network

True label	Prediction label	
	Negative	Positive
Negative	96	99
Positive	28	377

Table 4. Confusion matrix Long Short-Term Memory

True label	Prediction label	
	Negative	Positive
Negative	80	115
Positive	36	369

Table 5. Confusion matrix Bernoulli Naive Bayes

True label	Prediction label	
	Negative	Positive
Negative	8	187
Positive	12	393

Bernoulli Naive Bayes classification does not perform training such as Simple Recurrent Neural Networks and Long Short-Term Memory. However, it uses Word2vec training and probability using the data from the previous Word2vec training. The classification results of each algorithm are provided in Table 6.

Table 6. Comparison of Classification Results

Method	Accuracy	Precision	Recall	F1-score
<i>Simple Recurrent Neural Network</i>	78%	78%	78%	77%
<i>Long Short-Term Memory</i>	75%	74%	75%	73%
<i>Bernoulli Naive Bayes</i>	67%	58%	67%	56%

Comparing the classification results of each algorithm in Table 6 shows that the Simple Recurrent Neural Network has an advantage in classification with an accuracy value of 78%. On the other hand, the Long Short-Term Memory (LSTM) classification has an accuracy value of 75% and Bernoulli Naive Bayes' 67%. These three algorithms use the same data, 2400 training data and 600 testing data. The only difference is that Simple Recurrent Neural Network and LSTM carry out training using 12 epochs, whilst Bernoulli Naive Bayes relies on probability and training from Word2vec.

4. CONCLUSIONS

The Simple Recurrent Neural Network shows the best performance based on the research results, with accuracy reaching 78%, 78% precision, 78% recall, and 77% f1-score. The LSTM also performs quite well with an accuracy of 75%, while Bernoulli Naive Bayes with 67% accuracy in classification sentiment analysis. The Simple Recurrent Neural Network and LSTM methods are varied methods of the Recurrent Neural Network algorithm, whereas the Bernoulli Naive Bayes algorithm is derived from the Naive Bayes. From the comparison of the classification results, it can be concluded that the Recurrent Neural Network algorithm is superior in classifying sentiment analysis compared to the Naive Bayes algorithm.

REFERENCES

- [1] ESDM, *Indonesia Energy Out Look 2019*. Jakarta : Sekretariat Jenderal Dewan Energi Nasional, 2019.
- [2] L. Zhang, R. Ghosh., M. Dekhil, M. Hsu, and B. Liu, "Combining lexicon-based and learning-based methods for twitter sentiment analysis", *HP Laboratories Technical Report*, HPL-2011-89, 2011.
- [3] H. Soong, N. B. A. Jalil, R. K. Ayyasamy, and R. Akbar, "The essential of sentiment analysis and opinion mining in social media", *2019 IEEE 9th Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, pp. 272–277, 2019, <https://doi.org/10.1109/ISCAIE.2019.8743799>.
- [4] A. Novantirani, M. K. Sabariah, and V. Effendy, "Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine", *E-Proceeding of Engineering*, 2015.
- [5] B. Liu, "Sentiment Analysis and Subjectivity in: Handbook of Natural Language Processing", *Handbook of Natural Language Processing, Second Edition*, vol. 2, pp. 568, 2010.
- [6] K. Saranya, and S. Jayanthi, "Onto-based sentiment classification using machine learning techniques", *Proceedings of 2017 International Conference on Innovations in Information, Embedded and Communication Systems, ICIIECS 2017*, Vol. 2018-January, pp. 1–5. Institute of Electrical and Electronics Engineers Inc. 2018 <https://doi.org/10.1109/ICIIECS.2017.8276047>
- [7] A. Karpathy, "The Unreasonable Effectiveness of Recurrent Neural Networks", *Web Page*, pp. 1–28, 2015. Available <http://karpathy.github.io/2015/05/21/rnn-effectiveness/>
- [8] R. Socher, A. Perelygin, J. Y. Wu, J. Chuang, C. D. Manning, A. Y. Ng, and C. Potts, "Recursive deep models for semantic compositionality over a sentiment treebank". In *EMNLP 2013 - 2013 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*. pp. 1631-1642. Association for Computational Linguistics (ACL), 2013.
- [9] I. Zulfa, and E. Winarko, "Sentimen Analisis Tweet Berbahasa Indonesia dengan Deep Belief Network", *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 11(2), pp 187, 2017. <https://doi.org/10.22146/ijccs.24716>.

-
- [10] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient Estimation of Word Representations in Vector Space”, In *1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings*, International Conference on Learning Representations, ICLR. pp. 1–12, 2013.
- [11] T. Mikolov, M. Karafiát, L. Burget, J. Cernocký, and S. Khudanpur, “Recurrent neural network based language model”, *Proceedings of the 11th Annual Conference of the International Speech Communication Association. INTERSPEECH 2010*. (September), pp. 1045–1048, 2010.
- [12] J. Schmidhuber, “Deep learning in neural networks”: An overview. *Neural Networks*, vol. 61, pp. 85-117, 2015. Available doi: 10.1016/j.neunet.2014.09.003
- [13] G. Singh, B. Kumar, L. Gaur, and A. Tyagi, “Comparison between Multinomial and Bernoulli Naïve Bayes for Text Classification”, In *2019 International Conference on Automation, Computational and Technology Management, ICACTM 2019*. pp. 593–596. Institute of Electrical and Electronics Engineers Inc, 2019, Available <https://doi.org/10.1109/ICACTM.2019.8776800>
- [14] J. D. Novaković, A. Veljović, S. S. Ilić, Željko Papić, and T. Milica, “Evaluation of Classification Models in Machine Learning”, *Theory Appl. Math. Comput. Sci.*, vol. 7, no. 1, pp. Pages: 39 -, Apr. 2017.