JINAV: Journal of Information and Visualization *ISSN: 2746-1440 (online)*
Vol. 4 No. 1 (2023)
https://doi.org/10.35877/454RI.jinav1525

Fake News (Hoaxes) Detection on Twitter Social Media Content Through Convolutional Neural Network (CNN) Method

Fauzaan Rakan Tama* & Yuliant Sibaroni

Bachelor of Informatics Study Program, Faculty of Informatics, Telkom University, Bandung, Indonesia,

Abstract

The use of social media is very influential for the community. Users can easily post various activities in the form of text, photos, and videos in social media. Information on social media contains fake news and hoaxes that will have an impact on society. One of the most social media used is Twitter. This study aims to detect fake news found on the Tweets using the Convolutional Neural Network (CNN) method by comparing the weighting features used of the Term Frequency Inverse Document Frequency (TF-IDF) and the Term Frequency-Relevance Frequency (TF-RF). The highest accuracy was obtained in the Term Frequency-Relevance Frequency (TF-RF) weighting feature with an accuracy of 84.11%, while in the Term Frequency Inverse Document Frequency (TF-IDF) weighting feature with an accuracy of 80.29%.

Keywords: Social Media; Hoax; Twitter; Convolutional Neural Network; TF-IDF; TF-RF.

1. Introduction

1.1. Background

The development of information and communication technology certainly has a great influence in this current era, not only in the real world but in the form of cyberspace, especially on social media. It can easily and quickly disseminate information connected to the internet network (Togay et al., 2022). Social media is a software that aims to enable users to indoctrinate, communicate, argue, share, network and other activities (Sriyano & Setiawan, 2021). For example, on the social media of Facebook, Instagram, and Twitter, users can easily and quickly spread information that changes the topic of being discussed, such as discussing the community environment, politics up to the entertainment industry. As for (Sriyano & Setiawan, 2021) stated that the hoax news is currently developing in social media circles (Dirjen *et al.*, 2017).

Fake news can provide the wrong information for the users so that those who read the news can believe that this news is misleading and can scare the users who receive it (Batoebara *et al.*, 2020a) This news seeks to deceive and trap the public and adversely affect users who easily believe (Batoebara *et al.*, 2020b). Social media is an online medium, whose users can easily participate, share, argue, and create content in the form of blogs, social networks, wikis, forums, and virtual worlds (Istiani & Islamy, 2020). Twitter is one of the social media that is widely used by Indonesian people (Dirjen *et al.*, 2017).

Twitter is a *micro-blogging* social media founded by Jack Dorsey in March 2006 and started to be used on July 2006. Twitter has the uniqueness of only being able to post a text with a maximum of 140 characters namely a *tweet.* This media often also uses more than one language or so-called bilingual language (use two or more languages) [3][11]. On Twitter, users can disseminate information through tweets (Tineges et al., 2020). The more *tweets*, the wider the information obtained. However, some fake news on *tweets* often occurs. This will have a very bad impact on recipients who easily believe its information, for example on the impact of COVID-19 which stated the number of hoax news scaring the public and can attack people's psychology (Yunanto *et al.*, 2021).

* Corresponding author.
E-mail address: fauzaanrt@student.telkomuniversity.ac.id

Hoax news is compiled not based on facts (Parewe *et al.*, 2021a). In 2019, Ananthi *et al.* detected hoax news and real news using deep learning (Parewe *et al.*, 2021b). Moreover, in their research, Ananthi *et al.* compared machine learning and deep learning techniques. The dataset taken comes from Kaggle.com website of hoax news and real news datasets. The methods used *K-Nearest Neighbor, Decision Tree, Naïve Bayes, Random Forest,* CNN (*Convolutional Neural Network*), and LSTM (*Long Short Term Memory)* using *Term* Frequency *Inverse Document Frequency* (TF-IDF) weighting. The results of deep learning techniques through the CNN and LSTM methods have a better accuracy value than using mechine learning techniques (Kurniawan & Mustikasari, 2021).

Based on the description above, the impact caused by hoax news is very detrimental to many parties. Therefore related research on the detection of hoax news is mostly carried out using the *Decision Tree*, *Support Vector* (SVM), *Self-Organizing Map*, and *Convolutional Neural Network* methods. (CNN). In this study, in order to detect hoax news using the CNN method, it is actually more effective and efficient using the CNN method to classify documents, but the author tries to use CNN in text form to find out whether CNN can work well in managing text or not. This study used the weightings of Term Frequency-Inverse Document Frequency (TF-IDF) and Term Frequency-Relevance Frequency (TF-RF) weighting (Farid *et al.*, 2020a).

1.2. Topics and The Limitations

The topic discussed in this study is the fake news (*hoaxes*) detection of *tweets* on Twitter social media using the *Convolutional Neural Network* (CNN) method by comparing the use of weighting features *of Term Frequency Inverse Document* Frequency (TF-IDF) and *Term Frequency-Relevance Frequency* (TF-RF). The limitations on the study used datasets based on tweets on Twitter social media using the hashtag "covid-19 vaccine" and the dataset retrieved only using Indonesian language.

1.3. Objective

The purpose of this study is to acquire the optimal CNN architecture in identifying hoax news of tweets on Twitter social media and the comparison of the success rate results of the CNN classification system with the TF-IDF weighting feature for CNN classification system with TF-RF weighting in identifying hoax news on Twitter social media.

1.4. Writing Organization

The research section is structured as the section in chapter 1 which contains an introduction, the section in Chapter 2 contains an explanation of the related studies, the section in Chapter 3 contains an explanation of the methods used and their application, the section in Chapter 4 contains the results of its testing and analysis, and the section in Chapter 5 contains the conclusions of this study.

2. Literature Review

There are several research journals that have been carried out related to the conducting of this research. This study applies the implementation of deep learning to determine real and hoax news in Indonesian using the Convolition Neural Networks (CNN) and Long Short Term Memory (LSTM) methods. The data collected amounted to 1786 news with the number of hoax news of 984 cases and real news of 802 cases. Data collection was taken from TurnbackHoax.id site.

The results of the study using the CNN method had a level of accuracy, precision, and recall of 0.88, while the use of LSTM method had a level of accuracy, precision, and recall of 0.83 (Kurniawan & Mustikasari, 2021).

Furthermore, the study detected hoax news on Twitter used the Feed-forward and Back-propagation Neural Networks method. This study used vectorization methods of TF-IDF and Word2vek. The dataset was derived from crawling using the Twitter API based on tweets according to keywords and hashtags. The accuracy of Neural Networks obtained was highest at 78.76% (Dirjen *et al.*, 2017).

In 2017, a study was conducted to detect hoax news on twitter social media with the *Naïve Bayes Multinomial* method using the TF-IDF weighting feature with the results of *Naïve bayes* accuracy and the TF-IDF weighting method obtained the highest result of 72.06%. Data collection was obtained as a result of *crawling* using the twitter API. The total data collected was 51,421 data with the number of hoax data of 25,329 and 26,029 of non-hoax data (Sriyano & Setiawan, 2021).

3. Methods

At the design stage, this system classifies fake news from selected tweets based on the hashtag "covid-19 vaccine" by comparing the use of the Term Frequency Inverse Document Frequency (TF-IDF) and the Term Frequency-Relevance Frequency (TF-RF) weighting features with the CNN method. The built system design can be seen in Figure 1 as follows:

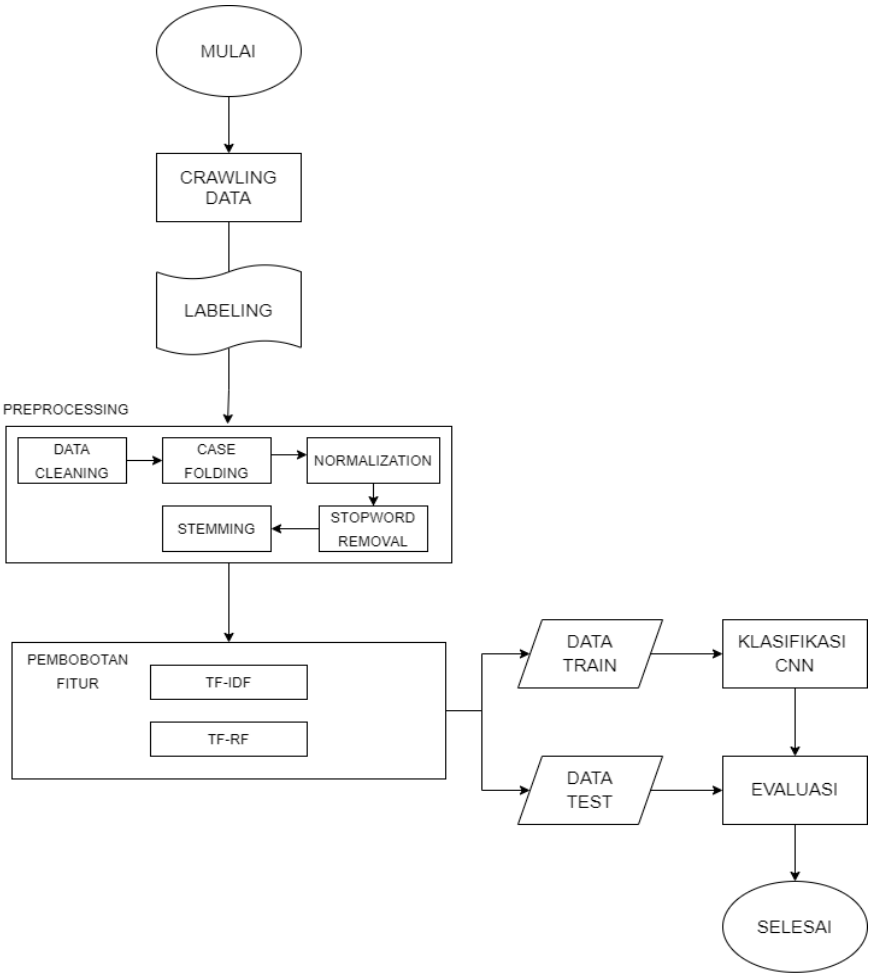

Figure 1. Flowchat System Design

3.1. Crawling Data

Crawling data is a stage of the data collection process. In this study, the researcher quoted tweets in the 2020-2022 period with the hashtag "covid-19 vaccine" as the process of retrieving data using the Application Programming Interface (API) provided by Twitter developers with a total of 16.000 tweets.

3.2. Data Labeling

Labeling is the process of manually labeling which is divided into 2, namely the index label "0" of real news and the index label "1" of hoax news. In this labeling process, there are three references that must be considered, namely the use of username that must be checked whether there are numbers or symbols and also using real names or pseudonyms, then the tweets that must be checked whether they have elements of hatred, panic, anxiety, provocation, or cornering other parties, the use of location with URLs in each tweet is checked whether it is appropriate to be posted or not, and last, then the number of followers of the account has been verified or not.

Table 1. Data Labeling Examples

Label	Tweets	Information
Hoax	Masuk Poin Ini, Dilarang Vaksinasi COVID-19!	Provocation, and the covid vaccine is mandatory for everyone
Fact	Booster sinovac bisa tingkatkan antibodi.	Fact because Sinovac booster can increase antibodies

Table 2. Number of Data Labeling

Label	Sum	Information
Hoax	871	Total hoax data 871
Fact	3853	Total real data 3853

3.3. Pre-processing Data

This has several stages, namely the stages of data cleaning, case folding, normalization, stop word removal, and steming. The initial stage is data cleaning as the part of the stage of cleaning words on data that contain noise in the data, such as emoticons and punctuation marks. Next case folding which there is this part of the stage to convert characters that use capital letters to lower-case letters. Furthermore, normalization is the stage to justify abbreviation words into actual words. Furthermore, the stop word removal stage is to remove words that are not required. Then, the last stage of stemming is to reverse the affix words into the real words.

Table 3. Pre-processing Data

Preprocessing	Data	Information
Tweets	Jangan ragukan utk lagi kehalalan dan kemanan vaksin covid-19 #acehviral #aceh #VaksinCovid19 #covid19indonesia	Original data that has not been preprocessed
Data Cleaning	Jangan ragukan utk lagi kehalalan dan kemanan vaksin covid acehviral aceh VaksinCovid covid indonesia	Remove hashtag and numbers in previous data
Case Folding	jangan ragukan utk lagi kehalalan dan kemanan vaksin covid acehviral aceh vaksincovid covid indonesia	Change capital letters to the lowercase
Normalization	jangan ragukan untuk lagi kehalalan dan kemanan vaksin covid acehviral aceh vaksincovid covid indonesia	Change the abbreviation word to the actual word. Example data besides "*utk*" becomes "*untuk*"
Stop word removal	jangan ragukan untuk kehalalan kemanan vaksin covid acehviral aceh vaksincovid covid indonesia	Remove unimportant words. For example, removing the word "*dan*"
Stemming	jangan ragu untuk halal keman vaksin covid acehviral aceh vaksincovid covid indonesia	Return the word affix. Example, "*ragukan*" to "*ragu*"

3.4. TF-IDF Weighting Features

TF-IDF is an algorithmic method whose function is to calculate the weight of words that are commonly used in a document. This method includes easy, efficient, and accurate methods. The TF-IDF calculates the value of each word in the document through an inverse comparison of the words frequency with the percentage of the document in which the word appears[2].

$$W_{(t,d)} = tf_{(t,d)}\, x\, idf_{(t)} = tf_{(t,d)} = (log\ N\ df_{(t)}) \quad (1)$$

Information:

W(t,d) : the value of the term (t) weight in the document(d)
tf(t,d) : the number of occurrences of term (t) in the document(d)
idf(t): number of inverses of document frequency per word
df(t) : the number of frequencies of each word
N : total number of documents

This process also uses the n-gram weighting feature which is divided into 3, namely unigram, bigram, and trigram (Sriyano & Setiawan, 2021). N-Gram is the N-character chunk of a longer string (MacLaughlin & Greenwood, 2010).

Table 4. N-gram example

Data	Jokowi membuat gedung baru
Unigram	| Jokowi| |membuat| |gedung| |baru|
Bi-gram	| Jokowi membuat| gedung baru|
Tri-gram	| Jokowi membuat gedung| membuat gedung baru|

3.5. TF-RF Weighting Features

Relevance Frequency is a relatively new method created in an effort to improve previous TF methods. This method assesses the relevance of documents seen from the frequency of occurrence in related categories (Lan *et al.*, 2006). This process also used the n-gram weighting feature which was divided into 3, namely unigrams, bigrams, and trigrams (Sriyano & Setiawan, 2021)

$$tf_{(t,d)}\, rf_{(t)} = tf_{(t,d)} * (log\ (2 + \frac{b}{max1,c}) \quad (2)$$

Description:

tf*rf : Weighting of documents into vector model space
tf(t,d): The number of occurrences of the word T in a document
b : number of documents containing word t documents
c : number of documents that do not contain the word t

3.6. Convolutional Neural Network Classification (CNN)

Convolition Neural Networks (CNN) is an algorithm of deep learning that can accept input in the form of images. Besides, it can determine what aspects of the object in the form of images I used "learning" machines that can distinguish one from another (Shafirra & Irhamah, 2020). CNN has several layers namely the input layer, the hidden layer, and the output layer. Hidden layers involve Convolutional layers, pooling layers, activation layers (generally Relu), fully connected layers and loss layers (Farid *et al.*, 2020b; Yoviananda & Fahrudin, 2022).

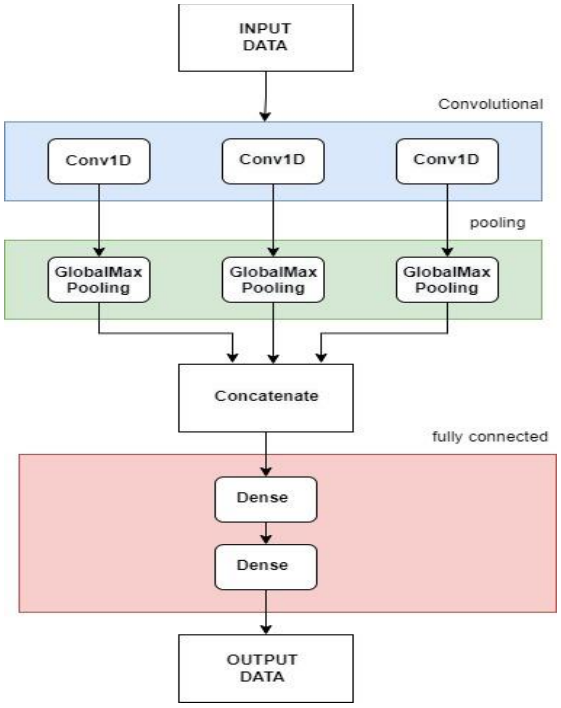

Figure 2. CNN architecture model

3.7. Evaluation Metrics

The last stage in this study is the evaluation of the performance related to the system built. This stage is the stage of performance measurement, collecting, analyzing, and evaluating the performance to be designed. This performance measurement uses the Confusion Matrix and accuracy, precision, recall, and F1 Score calculations.

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

$$Precision\ (P) = \frac{TP}{TP+FP} \quad (4)$$

$$Recall(R) = \frac{TP}{TP+FN} \quad (5)$$

4. Result and Discussion

This study aims to identify hoax news on social media of tweets on Twitter social media using the Convolutional Neural Network (CNN) method by comparing the use of Term Frequency-Inverse Document Frequency (TF-IDF) and Term Frequency-Relevance Frequency (TF-RF) weighting features.

4.1. Test Results

The first scenario used by TF-IDF aims as a word weighting feature by testing unigrams, bigrams, and trigrams and conducted classification with the Convolutional Neural Network (CNN) method. Accuracy results were obtained from the unigram, bigram, and trigram in Tables 5,6, and 7

Table 5. Unigram with TF-IDF feature Weighting

K- fold	Accuracy
1	78.43%
2	73.15%
3	75.47%
4	75.68%
5	79.87%
6	75.42%
7	79.02%
8	76.90%
9	78.81%
10	76.69%

Based on Table 5, unigram testing with TF-Idf feature weighting using cross validation with the k-fold number of 10, then the K-fold 5 showed the highest accuracy result of 79.87% and the lowest result on K-fold 2 of 73.15%.

Table 6. Bigram with TF-IDF Feature Weighting

K- fold	Accuracy
1	78.43%
2	74.20%
3	78.22%
4	78.43%
5	76.48%
6	77.33%
7	76.27%
8	75.21%
9	79.02%
10	**80.29%**

Based on Table 6, bigram testing with TF-Idf feature weighting using cross validation with the k-fold number of 10, then, the K-fold 10 showed the highest accuracy result of 80.29% and the lowest result on K-fold 2 of 74.20%.

Table 7. Trigram with TF-IDF feature weighting

K- fold	Accuracy
1	78.22%
2	79.69%
3	77.37%
4	75.68%
5	79.44%
6	78.38%
7	79.23%
8	76.69%
9	**80.29%**
10	78.38%

Based on table 7, unigram testing with TF-Idf feature weighting using cross validation with the k-folds number folds of 10, then, the K-fold 9 provides the highest accuracy result of 80.29% and the lowest result on K-fold 4 is 75.68%. The trigrams that tend to be higher in accuracy. Furthermore, the second scenario used TF-RF as a word weighting feature by testing unigrams, bigrams, and trigrams and performed classification by the *Convolutional Neural Network* (CNN) method. This second test aims to compare weighting features to be more accurately based on the accuracy of the unigram,bigram, and trigram. Accuracy results were obtained from the unigram, bigram, and trigram in Tables 8,9 and 10.

Table 8. Unigram with TF-RF feature Weighting

K-fold	Accuracy
1	79.70%
2	74.63%
3	76.10%
4	77.80%
5	79.23%
6	78.38%
7	78.17%
8	76.90%
9	78.60%
10	79.44%

Based on Table 8, unigram testing with TF-RF feature weighting using cross validation with the number of k-folds of 10, then the K-fold 1 provided the highest accuracy result of 79.70% and the lowest result on K-fold 2 of 74.63%.

Table 9. Bigram with TF-RF feature Weighting

K-fold	Accuracy
1	79.70%
2	78.22%
3	82.24%
4	81.18%
5	80.93%
6	75.42%
7	80.29%
8	75.63%
9	82.62%
10	80.08%

Based on Table 9, through bigram testing with TF-RF feature weighting using cross validation with the number of k-folds 10, then, the K-fold 3 provided the highest accuracy result of 82.24% and the lowest result on K-fold 6 of 75.42%.

Table 10. Trigram with TF-RF feature weighting

K-fold	Accuracy
1	80.12%
2	78.64%
3	83.50%
4	80.12%
5	81.77%
6	76.27%
7	81.56%
8	79.87%
9	81.77%
10	**84.11%**

Based on Table 10, through bigram testing with TF-RF feature weighting using cross validation with the number of k-folds 10, the K-fold 10 provided the highest accuracy result of 84.11% and the lowest result on K-fold 6 of 76.27%. Based on the three n-grams data results above, the trigram has the highest accuracy.

4.2. Analysis of Test Results

Based on the test results from the two skenarios that have been carried out, it can be seen that using TF-IDF weighting feature with the CNN classification method tends to has smaller accuracy result than using the TF-RF weighting feature. Based on n-grams, trigrams have highest accuracy results than unigrams and bigrams. The accuracy result on the TF-RF feature is highest on the trigram with an accuracy of 84.11% and the highest results on the trigram and bigram TF-IDF weighting features with the same accuracy results, namely 80.29%. So, using the TF-RF weighting feature provides a great influence to determine the accuracy value in the process of the hoax news detection system on Twitter through the CNN classification method.

5. Conclusion

Based on this research, with the aim of detecting fake news (hoaxes) on Twitter social media content with the hashtag of "covid-19 vaccine" using deep learning classification method of the Convolutional Neural Network (CNN) is used to be able to minimize the spread of hoax news that has a negative impact on society. Through comparing the use of Term Frequency-Inverse Document Frequency (TF-IDF) and the Term Frequency-Relevance Frequency (TF-RF) weighting features, these greatly affects the level of accuracy in using the TF-RF weighting feature to get higher accuracy value in terms of any N-gram. The highest accuracy of the trigram is 84.11%. The accuracy value obtained from the test is not optimal because the influence of the preprocessing process carried out has the similar word. In addition, there is lost data when preprocessing so that the system cannot test optimally and resulting in a decrease in the accuracy value. The suggestion for further research is to use the TF-BinICF weighting feature using the CNN classification method to find out whether its weighting feature can improve the higher accuracy than other weighting features.

References

Batoebara, M. U., Suyani, E., & Nuraflah, C. A. (2020a). Literasi Media dalam Menanggulangi Berita Hoaks (Studi Pada Siswa SMKN 5 Medan). *Jurnal Warta Edisi 63*, *14*, 34–41. http://jurnal.dharmawangsa.ac.id/index.php/juwarta/article/download/541/530

Batoebara, M. U., Suyani, E., & Nuraflah, C. A. (2020b). Literasi Media dalam Menanggulangi Berita Hoaks (Studi Pada Siswa SMKN 5 Medan). *Jurnal Warta Edisi 63*, *14*(1), 34–41.

Dirjen, S. K., Riset, P., Pengembangan, D., Dikti, R., Wintang Kencana, C., Budi, E., #2, S., & Kurniawan, I. (2017). Terakreditasi SINTA Peringkat 2 Hoax Detection on Twitter using Feed-forward and Back-propagation Neural Networks Method. *Masa Berlaku Mulai*, *1*(3), 648–654.

Eka Sembodo, J., Budi Setiawan, E., & Abdurahman Baizal, Z. (2016). *Data Crawling Otomatis pada Twitter*. *October 2018*, 11–16. https://doi.org/10.21108/indosc.2016.111

Farid, H. K., Setiawan, E. B., & Kurniawan, I. (2020a). Implementation Information Gain Feature Selection for Hoax News Detection on Twitter using Convolutional Neural Network (CNN). *Indonesia Journal on Computing (Indo-JC)*, *5*(3), 23–36. https://doi.org/10.34818/INDOJC.2020.5.3.506

Farid, H. K., Setiawan, E. B., & Kurniawan, I. (2020b). Implementation Information Gain Feature Selection for Hoax News Detection on Twitter using Convolutional Neural Network (CNN). *Indonesia Journal on Computing (Indo-JC)*, *5*(3), 23–36. https://doi.org/10.34818/INDOJC.2020.5.3.506

Istiani, N., & Islamy, A. (2020). Fikih Media Sosial Di Indonesia. *Asy Syar'Iyyah: Jurnal Ilmu Syari'Ah Dan Perbankan Islam*, *5*(2), 202–225. https://doi.org/10.32923/asy.v5i2.1586

Kurniawan, A. A., & Mustikasari, M. (2021). Implementasi Deep Learning Menggunakan Metode CNN dan LSTM untuk Menentukan Berita Palsu dalam Bahasa Indonesia. *Jurnal Informatika Universitas Pamulang*, *5*(4), 544. https://doi.org/10.32493/informatika.v5i4.6760

Lan, M., Tan, C. L., & Low, H. B. (2006). Proposing a new term weighting scheme for text categorization. *Proceedings of the National Conference on Artificial Intelligence*, *1*, 763–768.

MacLaughlin, H., & Greenwood, S. (2010). Weight management of obese patients on the renal ward. *Journal of Renal Nursing*, *2*(3), 116–121. https://doi.org/10.12968/jorn.2010.2.3.48079

Parewe, A. M. A. K., Aman, A., & Dewang, D. P. M. (2021a). Perbandingan Algoritma Winnowing dan Algoritma Manber dalam Mendeteksi Berita Hoax di Media Sosial. *Seminar Nasional Teknologi Informasi Dan Komputer*, 41–46.

Parewe, A. M. A. K., Aman, A., & Dewang, D. P. M. (2021b). Perbandingan Algoritma Winnowing dan Algoritma Manber dalam Mendeteksi Berita Hoax di Media Sosial. *Seminar Nasional Teknologi Informasi Dan Komputer*, 41–46.

Shafirra, N. A., & Irhamah, I. (2020). Klasifikasi Sentimen Ulasan Film Indonesia dengan Konversi Speech-to-Text (STT) Menggunakan Metode Convolutional Neural Network (CNN). *Jurnal Sains Dan Seni ITS*, *9*(1). https://doi.org/10.12962/j23373520.v9i1.51825

Sriyano, C. S., & Setiawan, E. B. (2021). Pendeteksian Berita Hoax Menggunakan Naive Bayes Multinomial Pada Twitter dengan Fitur Pembobotan TF-IDF. *E-Proceeding of Engineering : Vol.8, No.2*, *8*(2), 3396–3405.

Yoviananda, C., & Fahrudin, T. M. (2022). Implementation of Deep Learning to Detect Indonesian Hoax News with Convolutional Neural Network Method. *IJEEIT International Journal of Electrical Engineering and Information Technology*, *4*(2), 86–93. https://doi.org/10.29138/ijeeit.v4i2.1525

Yunanto, R., Purfini, A. P., & Prabuwisesa, A. (2021). Survei Literatur: Deteksi Berita Palsu Menggunakan Pendekatan Deep Learning. *Jurnal Manajemen Informatika (JAMIKA)*, *11*(2), 118–130. https://doi.org/10.34010/jamika.v11i2.5362