Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 3, No. 4, August 2022, Page 356−362
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v3i4.2146

Radicalism Speech Detection in Indonesia on Twitter Using Backpropagation Neural Network Method

Muhammad Rajih Abiyyu Musa*, Yuliant Sibaroni

Informatics Study Program, Telkom University, Bandung, Indonesia
Email: 1,*muhammadrajih19@gmail.com, 2yuliant@telkomuniversity.ac.id
Submitted: **19/08/2022**; Accepted: **29/08/2022**; Published: **30/08/2022**

Abstract−In this modern era, many people use social media easily and freely. One of the social media used is Twitter. The reason people use Twitter is that they can express their opinion freely. However, this freedom does not always have a positive impact on other Twitter users. One of the negative impacts for users is that they can spread radical content. Therefore, this research aims to detect whether a tweet contains radical elements or not using the backpropagation neural network method. The process is carried out by taking data on Twitter, after which the preprocessing process is carried out. Then the data is processed using imbalanced handling, where the data is divided into oversampling and undersampling data. After the data is divided, the next process is to do stopword and then look for accuracy by comparing different epoch values, namely 100, 150, 200, and 250. The best epoch value obtained is 200, with a final accuracy result of 86%.

Keywords: Radicalism; Twitter; Backpropagation Neural Network

1. INTRODUCTION

In this digitalization era, many people use the internet network in their daily activities, from parents, to teenagers, to children. There are many positive things gained from utilizing the internet, such as learning new things as a source of insight and can also be used as entertainment media. However, there are also negative sides to using the internet, one of which is the spread of radicalism. With the ease of accessing the internet, many people are irresponsible in spreading radicalism.

Radical has a very diverse meaning in Indonesia. Radical content is more identical and often associated with ethnic and religious issues (SARA) [1]. According to the National Counterterrorism Agency, radicalization can encourage or provoke a group or individual to commit acts of violence in the name of religion [1]. The spread of radicals is often found on social media. Twitter is one of the social media that is often used in Indonesia. Twitter is a text-based microblogging social media where users can write a status called a tweet [2]. Twitter networks can be formed by users who talk about the same topic [3].

Since Twitter has a character limit for each tweet, there are many abbreviations, a figure of speech, and writing errors. Radicalization has diverse meanings in different countries. Radicalization is widely associated with current Islamist groups that have been circulating for several centuries [4]. The so-called Islamic State is an organization on social media for sharing propaganda, raising funds, and for radicalization [5].

Many radical detectors use datasets from Twitter. Some studies propose and also predict the influence of radicalization. The methods used also vary, namely Naïve Bayes, Support Vector Machine, Logistic Regression, and KNN. In this research [1], the author proposes to select features that are categorized as radical or not radical using Human Brain and DF-Threshold. In this study, the results of the KNN method with the value of k=7 are the most stable compared to algorithms that achieve precision, recall, f-measure, and accuracy up to 66,37%. Research [6], analyzed more than 114 thousand tweets containing the term radicalization. By using the Naïve Bayes and Support Vector Machine Method, the author uses semantics because it can help strengthen radical detection. This research proposes an approach for semantic context representation of radical rhetoric terms.

In research [7], the author used machine learning to classify radical text. The result obtained in this study reached 92,3% using random forest. The lack of methods as a comparison in conducting research is one of the weaknesses of this study. The author provides other solutions for readers to combine or try other methods in order to improve the results to be more optimal. In research [8], the author uses the backpropagation method, which aims to get parameter values with the best accuracy. This study uses 100 training data and 20 test data that will be used to find the smallest iteration value with the greatest accuracy. Next, find the smallest iteration value with the greatest accuracy with the parameter used is the learning rate value obtained previously, which is 0,5. After getting the smallest iteration value, the author tests and analyzes the effect of iteration on accuracy results. Based on the testing process, the accuracy value reaches 90.

In this study, the author used the backpropagation neural network method, which is an artificial neural network model with guided training so that for each input pattern, there is an output pair. The contribution of this research is to find the effect of balanced labels on unbalanced label data, then look at the effect on modified literary stopwords, and finally, by using repetition of epoch values to determine the level of accuracy produced. Basically, the purpose of artificial neural networks to conduct training is to get accuracy between the network's ability to respond to input patterns during training and provide an expected value by providing other similar input patterns [9].

Detecting the spread of radicalism on social media must be done and important to do, where a system is needed that can detect tweets as radical or not automatically. Therefore, in this study, the author used the

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 3, No. 4, August 2022, Page 356−362
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v3i4.2146

backpropagation neural network classification method. The advantages of this method are that it can easily formulate the experience and knowledge of forecasters and are very flexible in changing forecaster rules [9].

The purpose of this research is to determine the optimal parameters for detecting radical speech, to determine the level of accuracy produced and to find out whether the backpropagation neural network can detect radical speech accurately using the backpropagation neural network algorithm by testing imbalanced handling, comparing *sastrawi* stopword with modified *sastrawi* stopwords, and testing with epoch values.

2. RESEARCH METHOD

This research aims to detect whether a tweet contains radical words or not. The following is the design of the system built in this study which can be seen in Figure 1.

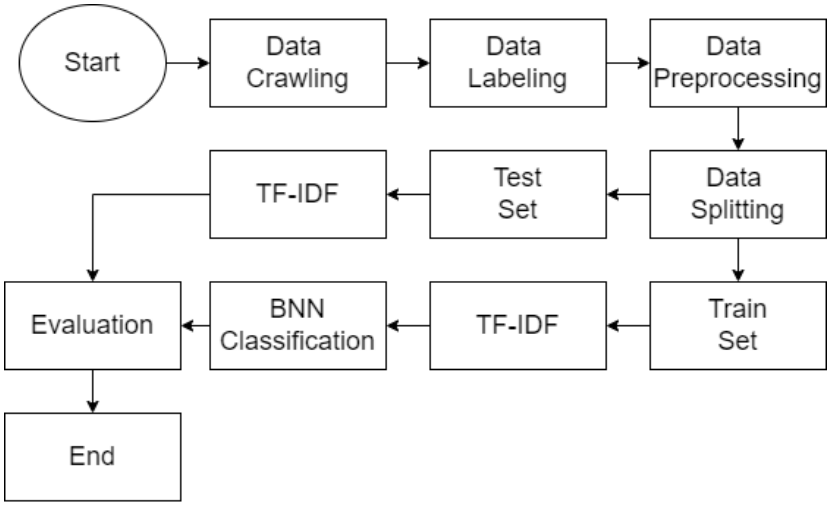

Figure 1. BNN Classification Model Process

2.1 Data Crawling

In this research, the datasets are obtained from the data crawling process on Twitter social media, where the crawling process uses the Twitter API with data retrieval time from January 21, 2016, to November 5, 2017. The dataset used uses the Indonesian language. The data that has been obtained is then labelled so that the data can be distinguished between data containing radical words or not, as shown in Table 1.

Table 1. Example of Dataset Used

tweet
@yys11_yeremia Indonesia ini berpegang pd pancasila atau agama islam Prasaan pancasila deh.
Terlalu banyak orang Islam yang bodoh di Indonesia. Baca Al Quran mu secara menyeluruh, jangan hanya sepotong2 ayat #IndonesiaBerkabung
Intinya Indonesia itu negara yang berlandaskan Bhinneka Tunggal Ika, negara yang mayoritas beragama Islam namun... https://t.co/LWrYc7ZeK3
@jokoanwar Umat islam di indonesia di adu domba giliran dia kena pasal malah kabur gimana ini ?
@wartapolitik @Handpro79 ayo kami dukung Panglima TNI, ambil alih tugas polisi, karena polisi tidak bisa dipercaya lagi rakyat indonesia yg mayoritas islam.
Vonis penjara Ahok adalah kemenangan Islam? Kemenangan atas apa? Perang apa? Image Islam Indonesia yg toleran hancur sudah di mata dunia.
HTI Akui Serukan Indonesia Jadi Negara Islam https://t.co/w4SfEwTlaE
@ustadtengkuzul @utama911 GIMN ISLAM NGK JDI MOMOK !!! KLO UMATNYA AJA SELALU MENDUKUNG KEJAHATAN.... INDONESIA RUSAK OLEH ORANG2 YG SOK BERAGAMA....
Mereka baca umat Islam Indonesia dgn referensi tahun 1970-an & 1980-an. Mereka tak ikuti lahirnya generasi baru yg umumnya penghafal Quran.

2.2. Data Labeling

After getting data from Twitter, the next step is to label the data into two labels, namely positive and negative. Positive data is determined when finding a sentence or word that has an extreme understanding or tarnishes the ideology of Pancasila, which is categorized as radical, and the rest is labelled negative.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 3, No. 4, August 2022, Page 356−362
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v3i4.2146

Table 2. Example of Labeling Data

Tweet	Label
@JWOLFFH @CIA Maksud saya adalah Indonesia sedang mengalami pergeseran demografis. Beberapa menginginkan reformasi dan sekularisme yang lain memahami bahwa Islam tidak dimaksudkan untuk direformasi	Negative
@DailyCaller Islam menyebalkan, penyalahguna, kultus yang tidak adil sebagai malaikat jahatnya Allah membutuhkan manusia untuk Menyesatkan & lindungi @Indonesia @kompascom @Metro_TV #Indonesia	Positive

2.3. Preprocessing

At this stage, the data that has been taken previously will be preprocessed before classification. The preprocessing stages are as follows:

a) Data Cleaning
Data cleaning is done to separate sentences into words or terms separated by punctuation marks such as spaces, periods, commas, question marks, and exclamation marks. The text of the previous sentences is separated into consecutive marks [2].

b) Case Folding
Case folding is the stage of converting all words in a sentence into lowercase letters. This process can improve accuracy in distinguishing similar words [2].

c) Tokenization
Tokenization is the process of cutting strings or words in a sentence into several fractions. All punctuation marks and hyphens will be removed [1][10].

d) Stemming
Stemming is a process of getting the base word from the derived word [11]. Stemming techniques remove pluralization suffixes from words or more complicated approaches that try to preserve meaning and include dictionaries [12].

e) Stopword Removal
Stopword is a process to remove words that are not used. Stopword is a word with low discrimination power, and it is only used for connecting high discriminating power words while building sentences [13].

Table 3. Preprocessing Process and Results

Process	Result
Sentences	“_19 _ary _irena "Pd Negara {Kemenag} sj.. Kemenag kan wadah bg smua agama.. Utk ursn intern Islam,sy hrp smua ulama Indonesia ambil sikap mllui MUI... ."
Data Cleaning	“ary irena Pd Negara Kemenag sj Kemenag kan wadah bg smua agama Utk ursn intern Islam sy hrp smua ulama Indonesia ambil sikap mllui MUI”
Case Folding	“ary irena pd negara kemenag sj kemenag kan wadah bg smua agama utk ursn intern islam sy hrp smua ulama indonesia ambil sikap mllui mui”
Tokenization	"['ary', 'irena', 'pd', 'negara', 'kemenag', 'sj', 'kemenag', 'kan', 'wadah', 'bg', 'smua', 'agama', 'utk', 'ursn', 'intern', 'islam', 'sy', 'hrp', 'smua', 'ulama', 'indonesia', 'ambil', 'sikap', 'mllui', 'mui']"
Stemming	"['ary', 'irena', 'pd', 'negara', 'kemenag', 'sj', 'kemenag', 'kan', 'wadah', 'bg', 'smua', 'agama', 'utk', 'ursn', 'intern', 'islam', 'sy', 'hrp', 'smua', 'ulama', 'indonesia', 'ambil', 'sikap', 'mllui', 'mui']"
Stopword	"['ary', 'irena', 'pd', 'negara', 'kemenag', 'sj', 'kemenag', 'wadah', 'bg', 'smua', 'agama', 'utk', 'ursn', 'intern', 'islam', 'sy', 'hrp', 'smua', 'ulama', 'indonesia', 'ambil', 'sikap', 'mllui', 'mui']"

2.4. Term Frequency – Inverse Document Frequency

Before using the backpropagation neural network, each word in the pre-processed dataset is weighted with TF-IDF. TF-IDF or Term Frequency – Inverse Document Frequency is one of the most effective ways to calculate term weight. Words with high TF values have importance in the document. On the other hand, DF implies the number of times a particular word appears in the document collection [14].

$$w_{ij} = tf_{ij} \times idf \quad (1)$$

$$idf = log \frac{N}{df_j} \quad (2)$$

w_{ij} = weight of word i in document j.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 3, No. 4, August 2022, Page 356−362
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v3i4.2146

N = number of document.
tf_{ij} = number of words i in document j.
df_i = number of document j containing word i.

2.5. Backpropagation Neural Network

The backpropagation neural network algorithm includes supervised learning designed for multi-layer perceptron operations [2]. In this algorithm, there are three layers, namely, the input layer, hidden layer, and output layer [15], as shown in Figure 2.

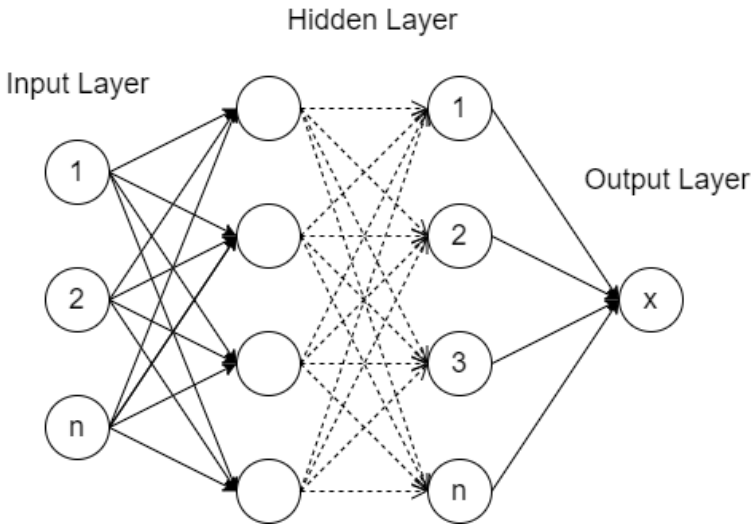

Figure 2. Backpropagation Neural Network Architecture

Backpropagation neural network uses error to change the value of its weight in the backward direction. In this phase, each output unit receives a target pattern to calculate the error value [16]. To get the error value, the forward propagation phase must be done first [2]. While the weight modification phase aims to reduce the error that occurs [16].

2.6. Confusion Matrix

The confusion matrix is a 2x2 matrix-shaped measurement tool used for the accuracy of the algorithm used. The confusion matrix can be seen in Table 4.

Table 4. Confusion Matrix Table

		Actual Class	
		True	**False**
Prediction Class	True	True Positive (TP)	False Positive (FP)
	False	False Negative (FN)	True Negative (TN)

Description:
True Positive (TP) = if the predicted data is positive and matches the actual value
False Positive (FP) = if the predicted data does not match the actual value
False Negative (FN) = if the predicted data is negative and the actual is positive
True Negative (TN) = if the predicted data is negative and the actual is negative

It is common to evaluate classification performance using f-measure, recall, and precision. Precision and recall are used to measure the performance of binary classification tests. Recall is the percentage of correctly recognized positive labels. Here are the equations of f-measure, recall, precision, and accuracy [1]:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

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

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

$$\text{F-Measure} = \frac{2*precision*recall}{Precision+recall} \quad (6)$$

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 3, No. 4, August 2022, Page 356−362
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v3i4.2146

3. RESULT AND DISCUSSION

In this research, several stages are carried out. The first stage is crawling, after which labelling is done in the dataset to be used. Then preprocessing is done, and data splitting is done. Then the data is classified using a backpropagation neural network. The following is the scenario used in this research:
a) Scenario 1: Testing the effect of imbalanced handling
b) Scenario 2: Comparison between *sastrawi* stopword and modified *sastrawi*
c) Scenario 3: Testing the effect of epoch values.

3.1 Scenario 1: Testing the Effect of Imbalanced Handling

Since the dataset used in this study has a far comparison of labels, this scenario uses imbalanced handling to equalize the number of labels tested. This scenario tests to balance the train data in the backpropagation neural network model. The data is divided into oversampling and undersampling data. The test results can be seen in Table 5.

Table 5. Experiment Imbalanced Handling Test Results

Imbalance Handling	Accuracy	F1-Score
Oversampling	77%	43,63%
Undersampling	75%	74,75%

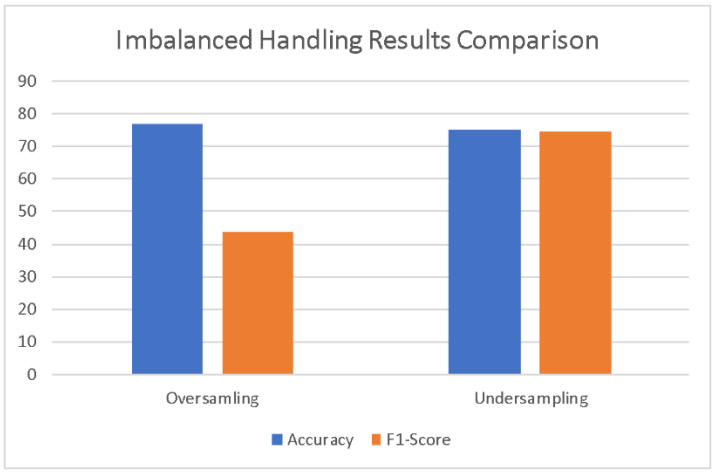

Figure 3. Imbalanced Handling Results Comparison

3.2 Scenario 2: Comparison Between *Sastrawi* Stopword and Modified *Sastrawi* Stopword

After getting the best performance on imbalanced handling, in this scenario using oversampling data, we will test the training data using *sastrawi* stopwords and modified *sastrawi* stopwords. This test is done to find out how much influence is obtained by using *sastrawi* stopwords with modified *sastrawi* stopwords by adding some of the remaining stopwords to the modified stopwords. The test results can be seen in Table 6.

Table 6. Stopword Removal Test Results

Stopword Removal	Accuracy	F1-Score
Sastrawi	83%	65,61%
Modified *sastrawi*	84%	68,45%

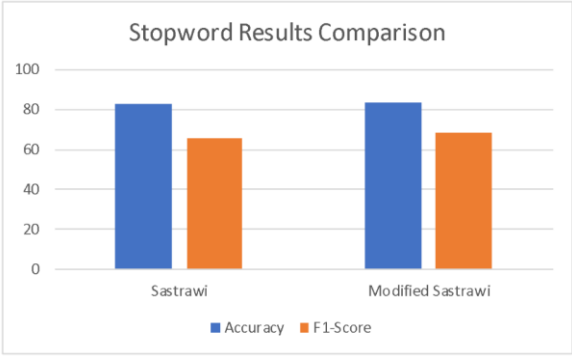

Figure 4. Stopword Removal Results Comparison

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 3, No. 4, August 2022, Page 356−362
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v3i4.2146

3.3 Scenario 3: Testing the Effect of Values

Because the accuracy results obtained using a different number of epochs, the results will be different too, so testing the number of epochs is needed. This scenario tests to find out how much accuracy is obtained by using a comparison of epoch values. The epoch values used are 100, 150, 200, and 250. The test results can be seen in Table 7.

Table 7. Epoch Test Results

Epoch Values	Accuracy	F1-Score
100	86%	79%
150	85%	78,59%
200	86%	79,38%
250	85%	77,69%

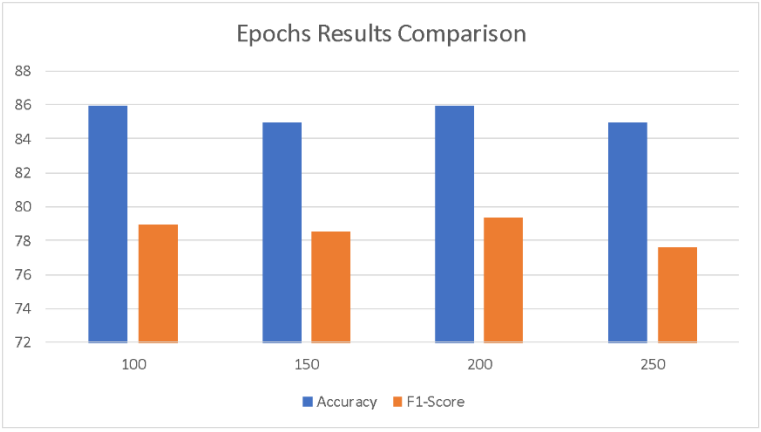

Figure 5. Epoch Results Comparison

After testing with the three scenarios using a learning rate of 0.01, the optimizer used is SGD, using three layers, namely the input layer, hidden layer, and output layer. It can be concluded that the stages performed in the scenario can affect the performance value of the backpropagation neural network model. In the first test, namely imbalanced handling, the accuracy value on oversampling data is greater with an accuracy value of 77% compared to undersampling data with an accuracy value of 75%. This is influenced because the train data on oversampling used is more than the train data on undersampling. In the second test, namely comparing *sastrawi* stopword. The accuracy obtained in *sastrawi* stopword is 83%, while the modified *sastrawi* stopword has a greater accuracy value of 84%. The addition of words to the modified *sastrawi* stopword gives better performance than the unmodified *sastrawi* stopword. By using *sastrawi* stopword data that has been modified previously, in this last scenario, namely testing on epoch values with the values used are 100, 150, 200, and 250. It was found that the accuracy value in this test was 86% with epoch values of 100 and 200, but the epoch with a value of 200 has a greater f1-score value of 79.38% than the epoch with a value of 100, which is 79%.

4. CONCLUSION

After conducting research on radicalism speech detection in Indonesia on Twitter social media using the backpropagation neural network method, it is concluded that imbalanced handling of train data using oversampling gets a greater accuracy value than using undersampling. The use of *sastrawi* stopword that has been modified also affects the accuracy value produced compared to unmodified *sastrawi*. In testing the value epochs, although the accuracy value produced by each value is not stable, the epoch value of 200 produces the best accuracy of 86%. Suggestion for further research is to increase the dataset to be used in research and have a balanced number of labels in order to produce a better accuracy value.

REFERENCES

[1] M. Subhan, A. Sudarsono, and A. Barakbah, "Preprocessing of radicalism dataset to predict radical content in Indonesia," *Proc. - Int. Electron. Symp. Knowl. Creat. Intell. Comput. IES-KCIC 2017*, vol. 2017-Janua, pp. 270–275, 2017, doi: 10.1109/KCIC.2017.8228598.

[2] N. A. Setyadi, M. Nasrun, and C. Setianingsih, "Text Analysis for Hate Speech Detection Using Backpropagation Neural Network," *Proc. - 2018 Int. Conf. Control. Electron. Renew. Energy Commun. ICCEREC 2018*, pp. 159–165, 2018, doi: 10.1109/ICCEREC.2018.8712109.

[3] E. Milani, E. Weitkamp, and P. Webb, "The visual vaccine debate on twitter: A social network analysis," *Media Commun.*, vol. 8, no. 2, pp. 364–375, 2020, doi: 10.17645/mac.v8i2.2847.

ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 3, No. 4, August 2022, Page 356−362
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v3i4.2146

[4] A. Kaya, "Islamist and nativist reactionary radicalisation in europe," *Polit. Gov.*, vol. 9, no. 3, pp. 204–214, 2021, doi: 10.17645/pag.v9i3.3877.
[5] M. Fernandez, M. Asif, and H. Alani, "Understanding the roots of radicalisation on twitter," *WebSci 2018 - Proc. 10th ACM Conf. Web Sci.*, pp. 1–10, 2018, doi: 10.1145/3201064.3201082.
[6] M. Fernandez and H. Alani, "Contextual semantics for radicalisation detection on Twitter," *CEUR Workshop Proc.*, vol. 2182, 2018.
[7] A. De Pablo, O. Araque, and C. A. Iglesias, "Radical text detection based on stylometry," *ICISSP 2020 - Proc. 6th Int. Conf. Inf. Syst. Secur. Priv.*, pp. 524–531, 2020, doi: 10.5220/0008971205240531.
[8] B. Andrianto and S. Adinugroho, "Analisis Sentimen Konten Radikal Melalui Dokumen Twitter Menggunakan Metode Backpropagation," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 12, pp. 7380–7385, 2018.
[9] M. D. Sapanta, "Backpropagation Pada Studi Peramalan Beban Menggunakan Metode Artificial Neural Network," *Skripsi Jur. Tek. Elektro Univ. Islam Indones.*, 2018.
[10] N. M. G. D. Purnamasari, M. A. Fauzi, Indriarti, and L. S. Dewi, "Identifikasi Tweet Cyberbullying pada Aplikasi Twitter menggunakan Metode Support Vector Machine (SVM) dan Information Gain (IG) sebagai Seleksi Fitur," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 11, pp. 5326–5332, 2018.
[11] R. Syafaat Amardita and M. Dwifebri Purbolaksono, "Analisis Sentimen terhadap Ulasan Paris Van Java Resort Lifestyle Place di Kota Bandung Menggunakan Algoritma KNN," *J. Ris. Komputer)*, vol. 9, no. 1, pp. 2407–389, 2022, doi: 10.30865/jurikom.v9i1.3793.
[12] D. Farrar and J. H. Hayes, "A comparison of stemming techniques in tracing," *Proc. - 2019 IEEE/ACM 10th Int. Work. Softw. Syst. Traceability, SST 2019*, pp. 37–44, 2019, doi: 10.1109/SST.2019.00017.
[13] D. J. Ladani and N. P. Desai, "Stopword Identification and Removal Techniques on TC and IR applications: A Survey," *2020 6th Int. Conf. Adv. Comput. Commun. Syst. ICACCS 2020*, pp. 466–472, 2020, doi: 10.1109/ICACCS48705.2020.9074166.
[14] S. W. Kim and J. M. Gil, "Research paper classification systems based on TF-IDF and LDA schemes," *Human-centric Comput. Inf. Sci.*, vol. 9, no. 1, 2019, doi: 10.1186/s13673-019-0192-7.
[15] Nur Ghaniaviyanto Ramadhan and Imelda Atastina, "Neural Network on Stock Prediction using the Stock Prices Feature and Indonesian Financial News Titles," *Int. J. Inf. Commun. Technol.*, vol. 7, no. 1, pp. 54–63, 2021, doi: 10.21108/ijoict.v7i1.544.
[16] F. A. Hizham, Y. Nurdiansyah, and D. M. Firmansyah, "Implementasi metode Backpropagation Neural Network (BNN) dalam sistem klasifikasi ketepatan waktu kelulusan mahasiswa," *Berk. Sainstek*, vol. 6, no. 2, pp. 97–105, 2018, [Online]. Available: https://www.researchgate.net/publication/330446472_Implementasi_Metode_Backpropagation_Neural_Network_BNN_dalam_Sistem_Klasifikasi_Ketepatan_Waktu_Kelulusan_Mahasiswa_Studi_Kasus_Program_Studi_Sistem_Informasi_Universitas_Jember