IJCCS (Indonesian Journal of Computing and Cybernetics Systems)
Vol.15, No.2, April 2021, pp. 121~130
ISSN (print): 1978-1520, ISSN (online): 2460-7258
DOI: 10.22146/ijccs.65176

Implementation of K-Nearest Neighbor (K-NN) Algorithm For Public Sentiment Analysis of Online Learning

Auliya Rahman Isnain*[1], Jepi Supriyanto[2], Muhammad Pajar Kharisma[3]
[1,2,3]Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia, Lampung, Indonesia
e-mail: ***[1]auliyarahman@teknokrat.ac.id**, [2]japi.supriyanto@gmail.com,
[3]pajarkharisma@teknokrat.ac.id

Abstrak

Penelitian ini dilakukan untuk menerapkan algoritma KNN (K-Nearest Neighbor) dalam melakukan sentimen analisis pengguna Twitter tentang isu terkait kebijakan pemerintah tentang Pembelajaran Daring. penelitian menggunakan data Tweet sebanyak 1825 data tweet Bahasa Indonesia data dikumpulkan sejak tanggal 1 Februari 2020 sampai dengan 30 September 2020. Menggunakan library python yaitu Tweepy .pembobotan kata menggunakan TF-IDF, akan dilakukan klasifikasi nilai sentimen ke dalam dua kelas yaitu positif dan negatif. Setelah dilakukan pengujian dengan K sebanyak 20 didapatkan hasil akurasi tertinggi terdapat Pada saat K = 10 dengan nilai akurasi 84,65% dengan presisi mencapai 87%, recall 86% f measure 87% serta error rate mencapai 0,12% dan di dapatkan pula kecenderungan opini publik terhadap Pembelajaran Daring Cenderung Positif.

Kata kunci—*Sentimen analisis, pembelajaran daring, K Nearest Neighbor,TF-IDF,Confusion Matrix.*

Abstract

This research was conducted to apply the KNN (K-Nearest Neighbor) algorithm in conducting sentiment analysis of Twitter users on issues related to government policies regarding Online Learning. Research using Tweet data as much as 1825 Indonesian tweet data data were collected from February 1, 2020 to September 30, 2020. Using the python library, Tweepy. word weighting using TF-IDF, will be classified into two classes of sentiment values, positive and negative. After testing with K of 20, the highest accuracy results were obtained when K = 10 with an accuracy value of 84.65% with a precision of 87%, a recall of 86% f measure 87% and an error rate of 0.12% and a tendency was also obtained. public opinion on online learning tends to be positive.

Keywords— *Sentiment analysis, online learning, K Nearest Neighbor, TF-IDF, Confusion Matrix.*

1. INTRODUCTION

Twitter is a microblogging site that allows its users to write about various topics and discuss issues that are hotly discussed. Twitter provides a service for its users to send or read tweets that have been shared with characters that are limited to a maximum of 140. With this service, people prefer to give their opinion through social media rather than convey it directly. Conversations between users on Twitter often contain random information, but can be exploited by 3rd parties [1][2]. Twitter users until January 2013 had more than 500 million internet users registered on Twitter. The increase in Twitter users usually occurs when there are popular events such as the current topic about Covid-19.

Received April 8th,2021; Revised April 27th, 2021; Accepted April 29th, 2021

Since the announcement of positive cases of Corona virus (Covid-19) in Indonesia on March 2, 2020, until July 31, 2020, a total of 108,376 confirmed cases were recorded, with details of patients in care reaching 37,338 people spread across all hospitals in Indonesia, 65,907 The Ministry of Health of the Republic of Indonesia, (2020) The global pandemic that occurred in Indonesia made many parties try to participate in overcoming it. General practitioners and specialists spoke together to give a brief explanation to the community and appeal to maintain personal and environmental hygiene while not leaving the house much [3].

One of the impacts of the Corona virus pandemic is on education around the world, which has led to the closure of schools, madrassas, universities and Islamic boarding schools. In connection with these developments, the Ministry of Education and Culture (Kemendikbud) has also taken a policy as a guide in dealing with this disease at the level education unit by holding distance learning from the Ministry of Education and Culture (2020) responding to the policy of the Ministry of Education and Culture, many of them reap opinions among the public that they pour through social media, especially on Twitter.To interpret and understand these opinions, algorithms and programs are needed to process information and opinion data, as well as analyze the opinions of social media users. which is also called the sentiment analysis of [4].

Sentiment analysis or Opinion Minning, which is a field of textual data management that conducts studies based on a person's opinion, sentiment, evaluation, behavior and emotions that can be used as evaluation material [5][6]. The goal of this sentiment analysis is to find opinions / opinions in the writing of the text, get the sentiments that are in the opinions found, and in the end get the polarity or classification of whether positive / supportive or unsupportive, this sentiment analysis is divided into three types consisting of : document level sentiment analysis, sentences level sentiment analysis, and aspect level sentiment analysis [7].

In a study conducted by [8], with the title "sentiment analysis of television shows based on public opinion on social media Twitter using the K – Nearest Neighbor method and weighting the number of retweets". The calculation of accuracy uses various types of weighting, namely textual weighting, non-textual weighting and combining the two weights. Textual weighting reaches 82.50%, non-textual weighting reaches 60%, using the combination of the two reaches 83.33 Based on the accuracy test, the results of the classification accuracy review from the sentiment analysis application using KNN are 96.61.

Research by [9] conducted a sentiment analysis study of public opinion on 2019 Homecoming Facilities and Transportation on Twitter Using a comparison between the Naïve Bayes Algorithm, Neural Network, KNN, and Svm, the test results of this study show that K-NN produces accuracy. The highest is (90.76%), SVM (89.03%), Naïve Bayes (78.16%), while the Neural Network algorithm produces accuracy (52.73%) In contrast to research conducted by [10] in word weighting using the TF feature -IDF and training data using consent similarity get classification results using the K-Nearest Neighbor method. The results are obtained with an accuracy rate of (94.23%). while research [11] conducted a sentiment analysis on Twitter users using the K-Nearest Neighbor algorithm with a total of 2000 tweet data with the highest accuracy (67.2%). In contrast to the research conducted by [12] with the title Sentiment Analysis of Automotive Products from Twitter Using a Combination of the Knn Algorithm and the Lexicon Approach by testing using the Confusion Metrix model, the results, with a level of accuracy from the combination of KNN with the Lexicon approach are greater than previous research conducted by [11] resulted in an accuracy value of (83%).

While the research conducted by [13]. Implementation of the K - Nearest Neighbor Algorithm Against Restaurant Review Sentiment Analysis with Indonesian Text This test uses a confusion matrix and the accuracy results obtained reach 96.61%, the value of k = 1, and the error rate value of 3.39%. From the results of this study, it shows that the K Nearest Neighbor algorithm can produce a good percentage and is suitable for use in sentiment analysis, while [14] conducted research on sentiment analysis using a combination of the K - Nearest Neighbor

Algorithm with the Levenshtein distance algorithm. The highest accuracy obtained from the combination of these two algorithms has a value of 65.625%, this result is lower when compared to previous research conducted by [13] which only uses the KNN algorithm.

The purpose of this research is to analyze Twitter users' sentiment towards online learning. With input in the form of tweet data in Indonesian, classification will be carried out using the KNN (K-Nearest Neighbor) algorithm to determine whether the tweet is positive or negative.

2. METHODS

The sentiment analysis system design will be built in accordance with the results of the analysis. The system design that will be proposed in this study uses the K Nearest Neighbor (KNN) method which can be seen in Figure 1.

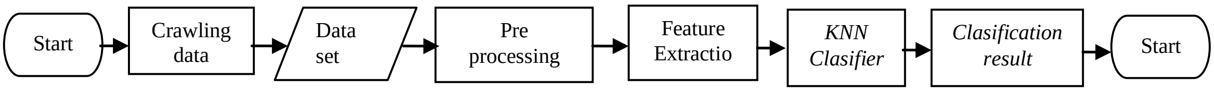

Figure 1. Public Analysis Sentiment Stages Of Online Learning

Figure 1 is an overview of the process that will be carried out to conduct public analysis sentiment on online learning.

2.1 Data

The data in this study were obtained from Twitter about public comments on online learning. The data was taken from February 1, 2020 to September 30, 2020. The data was retrieved by crawling the data used to retrieve tweets by utilizing the features of the twitter API (Application Programming Interface) and tweepy library in the Python programming language and stored in the database. The data taken is only tweets in Indonesian. In this study, the tweet data will be grouped manually into two polarity sentiments, namely positive and negative sentiments. The following is an example of crawling comment data from Twitter using the Tweepy library, Example data can be seen in Table 1:

Table 1 The results of data crawling

No	Tweet
1	@rahmadhan_set17Sekolah daring memang dirasa tepat karena upaya melindungi anak-anaknya agar terhindar dari covid-19.#belajardarirumah#covid19
2	@zTnaime_ Shitt!!!Alangkah tidak efektifnya Pembelajaran daring ini #covid19#pembelajarandaring
3	@nyimas_ayu Biarkan Pemerintah mencari solusi atas hambatan pembelajaran (daring), sembari menunggu daerahnya aman covid-19 #covid19#pembelajarandaring=).@Merdeka_belajar \nhttps://t.co/ig2svIektv
4	@janji_jiw"gw ga paham kenapa siswa malah jdi ngeluh dan menyalahkan guru dengan adanya pembelajaran daring. siswa pada ngeluh
5	"RT @nu_online: Ayo Dukung Program Pembelajaran Daring untuk Siswa Dhuafa
6	@satrio28"masa pandemi,masa normal baru,belajar dari rumah,pembelajaran Daring,Luring,kombinasi dll itulah yang saat ini kita lakukkan slama pandemi

After the data is obtained from the results of crawling, the data will then enter the preprocessing stage.

2.2. Preporcessing

The preprocessing stage is managing tweet data to make it easier to carry out the next process. In preprocessing, there are two steps that must be done, namely[2]: 1) Case Folding and Cleansing, At this stage the entire text is changed to lowercase (lowercase), then deleting all tweet data that contains numbers (0-9), links (http: //), username (@), (#) hash mark, comma (,) and period (.) as well as other punctuation marks. 2) Stopword Removal, After going through the case folding and cleansing process, entering the Stopword removal process is a process of removing words that are deemed unnecessary, for example the words 'yang', 'to', 'and', 'di', 'at', 'it' and all words. in the stopword dictionary that has been created. The purpose of this process is to minimize the number of words stored in the token list which will be carried out in the next process. 3) Tokenizing, In this section, the process of trimming a document or sentence into words or also known as tokens will be carried out. Table 3 illustrates the tokenizing process. 4.) Stemming, The last part of this preprocessing process is carried out to change all the words in the document to be converted into a root word by eliminating all affixes consisting of a prefix, a suffix, an infix, and a combination of prefix suffixes (confix).

2. 3 Feature Selection with TF-IDF

After going through the stemming stage, the weight is calculated using the TF-IDF TF-IDF method to improve the performance of the classifier which is useful for increasing accuracy and reducing computation time. In this method, the weight calculation is done by multiplying the value of the Term Frequency with the Inverse Document Frequency [15].

$$IDF = log\left(\frac{d}{df}\right), W_{d.t} = tf_{d.t}.IDF.t \quad (1)$$

where d is the d document, t is the t-word of the keyword, W is the weight of the d-document against the t-word. D is the total document, df is the number of documents that contain the word searched while tf is the number of words searched for in a document.

2. 4 Classification

Classification is a method for grouping data system according to predetermined rules and regulations. Classification can also be interpreted as grouping new data or objects based on observed variables to predict an object whose class or category is still unknown [16]. Classification is a data mining technique that looks at the attributes of a predefined data group. So that it can classify new data by manipulating the classified data and using the results to provide some rules. These rules are applied to new data for further classification. The purpose of classification is so that records that are not known in the previous category can be grouped accurately by [17].

The use of K-Nearest Neighbor aims to classify new objects based on learning data that is closest to the new object. The K-Nearest Neighbor algorithm technique is easy to implement. In this case, the amount of data or commonly referred to as the closest neighbor is determined by the user which is stated by k. As for the steps in the K Nearest Neighbor method: 1) Calculates Euclidean distances. Euclidean distance formula:

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^{n}(a_r(x_i) - a_r(x_j))^2}$$

Where d (x_i, x_j) is euclidean distance, x_i is record ti i, x_j record to j and a_r data to r. 2) Sort by Euclidean distance value, 3) Determine the k nearest classification record, and 4) The target output is the majority class.

2.5. Data classification with K Nearest Neighbor (KNN)

The classification in this study uses the K nearest Neighbor method with feature selection. Feature selection in classification is expected to be more efficient by reducing the amount of data analyzed by identifying features which will then be processed based on the classifier model that has been generated from the training process. In this study, the data is divided into two parts, namely training data and testing data, in Table 2 the training data is the sequence from D1 to D4 where the data already has Class while the D5 data (testing data) is unknown.

Table 2 Data Tweet

No	Tweet	Class
1	partai Golkar dan Demokrat akan Bertanding pada saat kampanye 2009	Negative
2	Pertandingan pertama antara Persema vs Persebaya malang	Positive
3	Sangat besar hapan wasit saay tanding sepakbola dpt berlaku adil	Positive
4	partai demokrat menang 2019 karena ada figur sby	Negative
5	Pertandingan sepakbola persebaya pada kampanye pemilu 2009 akan ditunda	?

There are 4 training documents (D1, D2, D3, D4) and 1 testing document (D5). The text documents in the training data are classified into 2 categories, namely positive class and negative class, first-class Positive while the second class is negative, determine the class of D5 following the steps that will be carried out. The first step is to give weight to each document using the TF IDF technique with the assumption that the data in Table 2 above has passed the preprocessing stage. The following results are shown in Table 3:

Table 3 TF-IDF calculation results

	tf						idf	wdt=tf.idf				
term	D1	D2	D3	D4	D5	DF	log(n/dff)	D1	D2	D3	D4	D5
partai	1			1		2	0,39794	0,39794	0	0	0,39794	0
golkar	1					1	0,69897	0,69897	0	0	0	0
demokrat	1			1		2	0,39794	0,39794	0	0	0,39794	0
tanding	1	1	1		1	4	0,09691	0,09691	0,09691	0,09691	0	0,09691
kampanye	1				1	2	0,39794	0,39794	0	0	0	0,39794
2009	1			1	1	3	0,22185	0,221849	0	0	0,221849	0,221849
pertama		1				1	0,69897	0	0,69897	0	0	0
persema		1				1	0,69897	0	0,69897	0	0	0
persebaya		1			1	2	0,39794	0	0,39794	0	0	0,39794
malang		1				1	0,69897	0	0,69897	0	0	0
besar			1			1	0,69897	0	0	0,69897	0	0
wasit			1			1	0,69897	0	0	0,69897	0	0
sepakbola			1		1	2	0,39794	0	0	0,39794	0	0,39794
adil			1			1	0,69897	0	0	0,69897	0	0
menang				1		1	0,69897	0	0	0	0,69897	0
pemilu				1	1	2	0,39794	0	0	0	0,39794	0,39794
figur				1		1	0,69897	0	0	0	0,69897	0
sby				1		1	0,69897	0	0	0	0,69897	0
tunda					1	1	0,69897	0	0	0	0	0,69897

Next is to calculate the similarity between documents using the cosine similarity calculation because what will be sought is the sentiment value of the last tweet data, then the tweet data (D5) is calculated for its similarity with all the data that has the results as follows.

Table 4. TFIDF calculation results

Wd5*wdi			
D1	D2	D3	D4
0,21696	0,16775	0,16775	0,20757

The next step is to calculate the length of each document including D5 by quadrading the weight of each term in each document, adding up the quadrad value and then rooting the results of the quadratic document as in the example in the Table 5 below.

Table 5 Vector Length document

Vector Length				
D1	D2	D3	D4	D5
0,158356	0	0	0,158356	0
0,488559	0	0	0	0
0,158356	0	0	0,158356	0
0,009392	0,009392	0,009392	0	0,009392
0,158356	0	0	0	0,158356
0,049217	0	0	0,049217	0,049217
0	0,488559	0	0	0
0	0,488559	0	0	0
0	0,158356	0	0	0,158356
0	0,488559	0	0	0
0	0	0,488559	0	0
0	0	0,488559	0	0
0	0	0,158356	0	0,158356
0	0	0,488559	0	0
0	0	0	0,488559	0
0	0	0	0,158356	0,158356
0	0	0	0,488559	0
0	0	0	0,488559	0
0	0	0	0	0,488559
1,022236	1,633425	1,633425	1,989963	1,180592
1,01106	1,27806	1,27806	1,41066	1,08655

Then the next step is to look for similarities between documents, calculate the similarities between D1 and D5, D2 and D5, D3 and D5 and D4 and D5 as follows:

Table 6 calculate the similarity of documents

Document	Formula	Result
COS(D5,D1)	0,26196/(1,08655*1,01106)	0,1975
COS(D5,D2)	0,16775/(1,08655*1,27806)	0,1208
COS(D5,D3)	0,16775/(1,08655*1,27806)	0,1208
COS(D5,D4)	0,20757/(1,08655*1,41066)	0,13542

The next step is to determine the class from D5 by taking from K as many as K (K = 3) which has the highest level of similarity to D1, then the results are as follows:

Table 7 Determining the Class of D1

D1	D4	D2
Negative	Negative	Positive

Choose the class that appears the most. For K = 3: Positive class, represented by 2 documents, namely D1 and D4. Negative class, represented only by D2. Conclusion D5 is clarified to Negative Class.

3. RESULTS AND DISCUSSION

The data used in this study were taken from Twitter, with tweets in Indonesian with the keyword "Online Learning" which were obtained from February 1, 2020, to September 30, 2020, with 1825 Tweets. The data was divided into two classes, positive and negative, which means agree or disagree with the Online Learning policy carried out by the Indonesian government, especially the Ministry of Education and Culture. This testing phase was carried out to find out how well the accuracy of the method used in this study was the K Nearest Neighbor (KNN) method with TF-IDF feature extraction in the policy.

3.1. The accuracy results from the K Nearest Neighbor Algorithm

Table 8 K Nearest Neighbor Algorithm Accuracy

No	Value K	accuracy	No	Value K	Accuracy
1	K=1	75%	11	K=11	81,64%
2	K=2	74,79%	12	K=12	83,56%
3	K=3	77,80%	13	K=13	82,73%
4	K=4	81,91%	14	K=14	82,46%
5	K=5	82,73%	15	K=15	82,19%
6	K=6	84,38%	16	K=16	82,19%
7	K=7	82,19%	17	K=17	82,46%
8	K=8	82,73%	18	K=18	82,19%
9	K=9	83,28%	19	K=19	80,82%
10	**K=10**	**84,93%**	20	K=20	82,46%

In the testing phase of the accuracy of the K Nearest Neighbor algorithm to find the best accuracy, testing is carried out with the values of K = 1, K = 2, K = 3, K = 4, K = 5, K = 6, K = 7, K = 8, K = 9, K = 10, K = 11, K = 12, K = 13, K = 14, K = 15, K = 16, K = 17, K = 18, K = 19, K = 20.

Based on the test results, the optimal accuracy of the K value is at the value of K = 10 with the accumulation obtained is 84.93%, from the results of the accuracy value ranging from K = 1 to K = 10 the more the K value, the higher the accuracy. , but at K = 11 to K = 20 it is not always high, even in vulnerable K = 13 to K = 20 the accuracy tends to be flat, therefore the author only includes Table K only up to 10. This happens because the more documents the more also the features or number of words of visualization of the data.

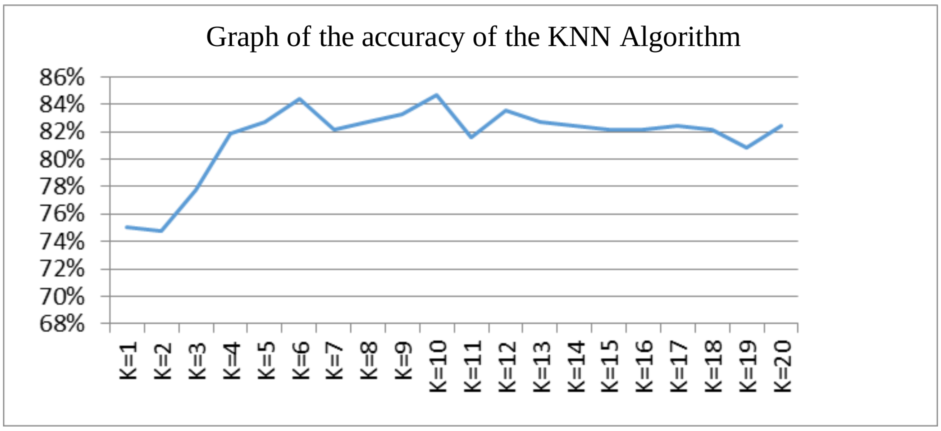

Figure 1 Graph of the accuracy of the KNN Algorithm

3.2. Testing with the Confusion Matrix

In this study, the confusion matrix method is used to test the performance of the K Nearest Neighbor algorithm in analyzing sentiment about public comments on Online Learning, testing the accuracy refers to TP (positive results detected correctly), TF (negative results detected correctly), FP (Positive results. detected incorrectly), and FN (negative results read incorrectly) the results of the test can be seen in Table 9 and Figure 2 reports from the Confusion Matrix.

Table 9 Confusion Matrix Result

Testing	Result
Accuration	85%
Precision	87%
Recall	86%
F measure	87%
Error Rate	0.12%

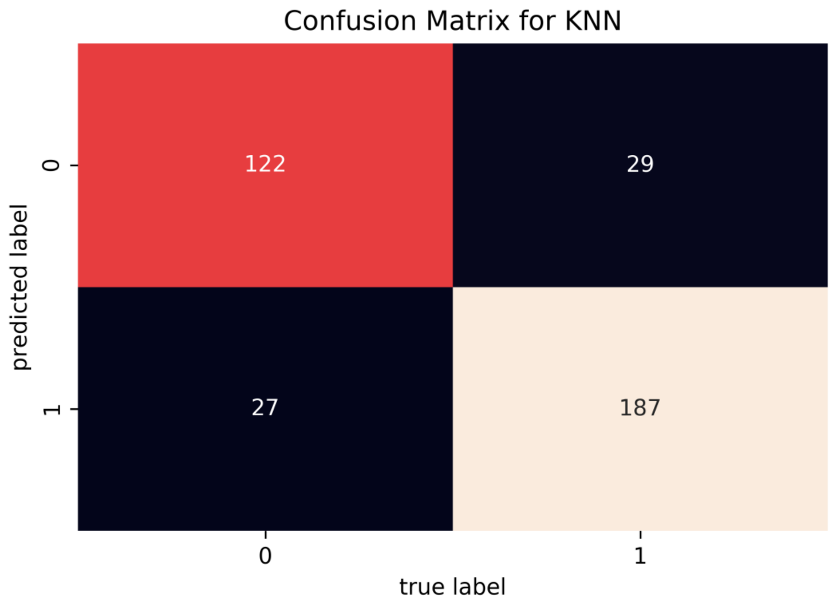

Figure 2 Confusion Matrix Test Results

4. CONCLUSIONS

Based on the results and discussion presented in the previous chapter, it can be seen that as many as 56.24% of tweets are positive while negative tweets are 43.76% with a total data of 1825. K Nearest Neighbor algorithm weighting the word TF-IDF, with a dataset of 1825 data . The dataset is divided into two data, namely 80% training data and 20% testing data, resulting in an accuracy rate of 84.93% on the K = 10 test. While the precision reaches 87%, recall 87%, f measure 87% and has an error rate of 0.12%.

The author realizes that the sentiment analysis made by researchers still has many shortcomings. Therefore, several things can be considered to be able to develop this research so that it is better in the future, namely: 1) Increasing the amount of training data used in the classification process. 2) Implementation of a larger data model. 3) Future research is expected to use two or more methods to improve the accuracy of the system.

REFERENCES

[1] R. R. Setiawan *et al.*, "Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, 2017, doi: 10.1074/jbc.M209498200.

[2] A. R. Isnain, N. S. Marga, and D. Alita, "Sentiment Analysis Of Government Policy On Corona Case Using Naive Bayes Algorithm," vol. 15, no. 1, pp. 55–64, 2021.

[3] E. M. Ferrazzi *et al.*, "COVID-19 Obstetrics Task Force, Lombardy, Italy: Executive management summary and short report of outcome," *Int. J. Gynecol. Obstet.*, 2020, doi: 10.1002/ijgo.13162.

[4] A. M. Ramadhani and H. S. Goo, "Twitter sentiment analysis using deep learning methods," 2017, doi: 10.1109/INAES.2017.8068556.

[5] D. Alita, Y. Fernando, and H. Sulistiani, "Sentiment Ana," *Tekno Kompak*, vol. 14, no. 2, pp. 86–91, 2020.

[6] A. R. Isnain, A. Sihabuddin, and Y. Suyanto, "Bidirectional Long Short Term Memory Method and Word2vec Extraction Approach for Hate Speech Detection," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 14, no. 2, p. 169, 2020, doi: 10.22146/ijccs.51743.

[7] A. Subhan, E. Sediyono, and O. Dwi, "Analisis Sentimen Berbasis Ontologi di Level Kalimat untuk Mengukur Persepsi Produk," vol. 02, pp. 84–97, 2015.

[8] W. E. Nurjanah, R. S. Perdana, and M. A. Fauzi, "Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat pada Media Sosial Twitter menggunakan Metode K-Nearest Neighbor dan Pembobotan Jumlah Retweet," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, 2017, doi: 10.1074/jbc.M209498200.

[9] M. W. Pertiwi, "Analisis Sentimen Opini Publik Mengenai Sarana Dan Transportasi Mudik Tahun 2019 Pada Twitter Menggunakan Algoritma Naïve Bayes, Neural Network, KNN dan SVM," *Inti Nusa Mandiri*, 2019.

[10] R. R. A. Siregar, Z. U. Siregar, and R. Arianto, "KLASIFIKASI SENTIMENT ANALYSIS PADA KOMENTAR PESERTA DIKLAT MENGGUNAKAN METODE K-NEAREST NEIGHBOR," *KILAT*, 2019, doi: 10.33322/kilat.v8i1.421.

[11] A. Deviyanto and M. D. R. Wahyudi, "PENERAPAN ANALISIS SENTIMEN PADA PENGGUNA TWITTER MENGGUNAKAN METODE K-NEAREST NEIGHBOR," *JISKA (Jurnal Inform. Sunan Kalijaga)*, 2018, doi: 10.14421/jiska.2018.31-01.

[12] P. Studi *et al.*, "OTOMOTIF DARI TWITTER MENGGUNAKAN KOMBINASI ALGORITMA K-NEAREST NEIGHBOR DAN PENDEKATAN LEXICON (STUDI KASUS : MOBIL TOYOTA) Skripsi," 2019.

[13] R. P. Fitrianti, A. Kurniawati, and D. Agusten, "Implementasi Algoritma K - Nearest Neighbor Terhadap Analisis Sentimen Review Restoran Dengan Teks Bahasa Indonesia," *Semin. Nas. Apl. Teknol. Inf. 2019*, pp. 27–32, 2019.

[14] M. Jabal Tursina, "Sentimen Analisis Sistem Zonasi Sekolah Pada Media Sosial Youtube Menggunakan Metode K-Nearest Neighbor Dengan Algoritma Levenshtein Distance," *Univ. Islam Negeri Syarif Hidayatullah Jakarta*, 2019.

[15] Y. Tu, "Machine learning," in *EEG Signal Processing and Feature Extraction*, 2019.

[16] T. H. Simanjuntak, W. F. Mahmudy, and Sutrisno Sutrisno, “Implementasi Modified K-Nearest Neighbor Dengan Otomatisasi Nilai K Pada Pengklasifikasian Penyakit Tanaman Kedelai,” *Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, No.2, no. 2, pp. 75–79, 2017, [Online]. Available: http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/15/21.

[17] S. Zainuddin, N. Hidayat, and A. A. Soebroto, “Penerapan Algoritma Modified K-Nearest Neighbour (MKNN) Pada pengklasifikasian Penyakit Tanaman Kedelai,” *Repos. J. Mhs. PTIIK UB*, vol. 3, p. 8, 2014.