Journal of Science, Technology, and Engineering (JSTE)
Vol. 1, No. 2, October 2021
P-ISSN: 8888-888, E-ISSN: 8888-8888

Journal of Scinece, Technology, and Engineering *is licensed under*
a Creative Commons Attribution-NonCommercial 4.0 International License.

Analysis of the Sentiment of Social Media Users to the Teacher's Room Using the K-Nearest Neighbor (K-NN) Algorithm

Nuke L Chusna[1*)], Herwanto[2], Elsa Naracristy[3]
[1,2] Lecture of Universitas Krisnadwipayana, Jatiwaringin, Jakarta Timur, Indonesia
[3] Student of Universitas Krisnadwipayana, Jatiwaringin, Jakarta Timur, Indonesia

nukelchusna@unkris.ac.id[1,*)], herwanto_p@yahoo.com[2], elsanaraac@gmail.com[3]

ARTICLE INFO

Article history:
Received: Feb 20, 2021
Revised: March 24, 2021
Accepted: Oct 27, 2021

Keywords:
Sentiment analysis, KNN (K-Nearest Neighbor), Social Media, Ruangguru

ABSTRACT

This study was made to classify the KNN (K - Nearest Neighbor) algorithm in Twitter user sentiment analysis from Ruangguru in June during the pandemic 2020. Tweet data used were 700 Indonesian-language tweet data with the distribution of training data and test data using a combination of 80% - 20%. Using the KNN algorithm with TF-IDF word weighting, the sentiment values will be classified into two classes, positive and negative. From the test results it is known that the best accuracy value is 88.21% in the parameter value of k = 13, the highest precision is 70.98% in the parameter k = 15, the results of several tests show that the sentiment towards the Teacher's Roomin June tends to be positive.

1. Introduction

Indonesia continues to develop electronic learning both in terms of the process of implementing electronic learning and in terms of technology. Advances in science and technology have an impact, especially in improving the quality of education. Various factors affect the development of education in the future, including the rapid development of information technology, as well as increasingly fierce competition in obtaining employment. In this context, reform in the field of education and learning needs to be carried out continuously and must be a process that will not stop(Auntie, 201: p.275)

The positive or negative nature of the opinion will be used to measure how much support the opinion maker supports on an existing problem topic. For topic creators, this can be used to take the next step related to the topic of the problem(Subhan, Sediyono, & Dwi, 2015: p.85)

The solution to assist Ruang Guru in making decisions by utilizing existing tweet data is Text Mining research can be used as a solution. Approach method or algorithm in order to speed up the accurate classification process, researchers use the method for decision makers, namely K-Nearest Neighbor. The K-Nearest Neighbor (K-NN) method will be used to overcome the above case, namely by classifying twitter user activity in tweets in the API data which requires login access to get the required data.

The problem that exists in the object of research is to see how far the content of meaning and comments made by Twitter social media users on the effectiveness of Ruang Guru is positive and or negative and the effectiveness of the application of the K-Nearest Neighbors (K-NN) algorithm in classifying a review including into the class of negative or positive sentiment.

The objectives to be achieved in this study are to obtain the best accuracy value from the iteration of the K-Nearest Neighbor (K-NN) algorithm parameters and to find out the dominant results of sentiment analysis of twitter users on Ruangguru.

Dewi Onantya, Indriati, & Pandu Adikara (2019)Entitled "Sentiment Analysis on BCA Mobile Application Reviews Using BM25 and Improved K-Nearest Neighbor." The problem in this research is the existing mobile application, there is no sentiment analysis feature to classify or filter between positive and negative reviews. Until the classification of documents using Improved K-Nearest Neighbor. The results obtained based on the evaluation in the form of a 5-fold test got the best k-values of 10, with a precision value of 0.946, recall of 0.934, f-measure of 0.939, and accuracy of 0.942.

Mentari, Fauzi, & Muflikhah (2018)Titled "Analysis of 2013 Curriculum Sentiment on Twitter Social Media Using the K-Nearest Neighbor Method and the Feature Selection Query Expansion Ranking". The problem in this research is to analyze tweets about the 2013 Curriculum by classifying them as positive or negative opinions. Based on the tests that have been carried out, it is proven that feature selection increases the accuracy of the system. The best accuracy result of 96.36% is obtained at the value of k = 1 by using a 50% feature selection ratio.

Nurjanah, Perdana, & Fauzi (2017)Titled "Analysis of Sentiment to Television Impressions Based on Public Opinion on Twitter Social Media using the K-Nearest Neighbor Method and Weighting the Number of Retweets". Textual weighting results from the classification, the data used is in the form of public opinion on television shows on Twitter as many as 400. From the results of accuracy testing using textual weighting obtained 82.50%, using non-textual weighting 60%, and using a combination of both 83.33% with value k=3.

2. Method

Text Mining

Text Mining (text mining) is a computer-based mining or algorithmic approach to analyzing text, spanning various communities, including information retrieval, language processing and extracting meaningful information from a text. (Allahyari et al., 2017).

Sentiment Analysis

Sentiment analysis is a grouping of texts in the form of opinionated textual information. Therefore, the nature of sentiment analysis is subjective to one thing. What is meant by subjective is that it can be in the form of negative sentiments or positive sentiments. Textual information that is grouped into negative or positive will contain a value. This value will then be used as a parameter in determining a decision on a document(Dewi Onantya, Indriati, Putra, 2019: p. 2576)

Twitter APIs

*Twitter*is a platform of social media, communication is done online which forms a structure. Twitter users can post comments known as tweets. *Twitter* also provides Application Programming Interfaces (APIs) that provide access to tweet data from a specific time range, from a specific user, with specific keywords, or from a specific geographic area, but does not provide a feature to extract structure from tweets, and does not provide an overview of aggregated data. twitter on different topics (for example, frequency of tweets about a particular topic over time) (Karami, 2020)

Rapid Miner

Rapid Miner is a software platform or software on data science developed by the company of the same name that provides a unified environment for text mining, machine learning, deep learning, and predictive analytics.). (Nofitri & Irawati, 2019: p.201)

K-Nearest Neighbors (K-NN) Algorithm

The K-Nearest Neighbors Algorithm is a method for classifying data based on its surrounding neighbors as a predictive value for new data(Fauziah, Sulistyowati, Asra, 2019: p.24)This algorithm is often used to classify text on data. In carrying out the calculation process using the K-nearest neighbor algorithm, several processes are carried out, namely:

1. The initial stage is to determine the value of K, for example K = 1. Then the closest document 1 will be taken to be used as a determinant of the classification results

2. Calculates the distance between the new data in each data label and the distance of the training data. To calculate this distance using the Euclidean Distance equation. The calculation method is to use the following formula:

$$D\ (X,Y) = \sqrt{\sum_{k=1}^{n}(Xk - Yk)^2} \qquad (1)$$

Information:
D= distance variable value k
X= Test data
Y= Training data

Then sort the results of the euclidean distance based on the provisions of the K value that you want to use. If K = 3, then the 3 closest distances will be selected as the classification results.

3. Results And Discussion

The data used in this study are Indonesian-language tweets found on Twitter. The tweets used are tweets that display the text as you are looking for. Data search is done by using basic data, and data knowledge. To obtain this data, researchers took data from Twitter in June 2020 so that the total tweets used as data amounted to 700.

Research Stages

The stages of research carried out in this study consisted of several processes. The processes that will be carried out are: collecting opinions from tweet data, entering the processing process, the results of preprocessing are calculated using the tf-idf weight. is a method for calculating the weight of the words used. The flow of the initial stages of this research is shown in Figure 3.1

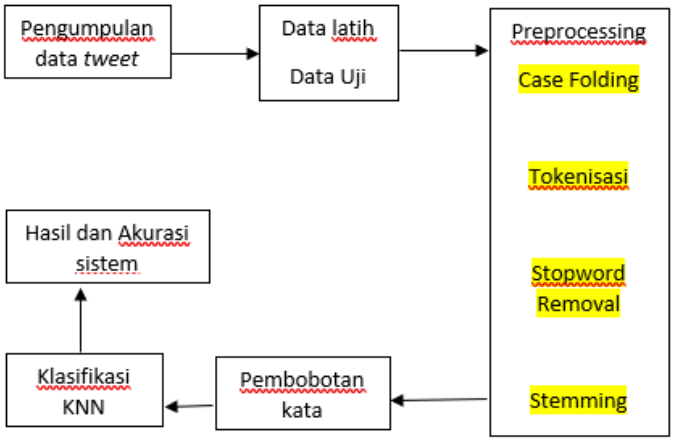

Fig 1. The flow of the research stage

Tweet Data Collection

Tweet data in this study was obtained by utilizing the API (Application Programming Interface) provided by twitter. By utilizing the API, an application was built to retrieve the tweet data from Twitter and then stored it in the database. The schematic of the tweet data retrieval process can be seen in Figure 3.2

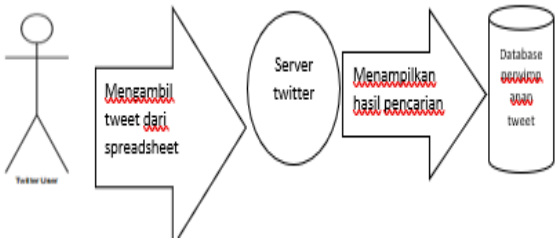

Fig 2. Schematic of the tweet retrieval process

The first step in retrieving tweets using the Twitter API is to have keys and an access token to connect to the Twitter API. The following captures the contents of the Twitter API key and access token. After connecting to the Twitter API, you need the keywords you want to search for as access rights to filter the required data. the following keywords are #RuangGuru , the teacher's room.

Preprocessing

Basically, the data obtained from this process has an irregular structure. Therefore, before the data is entered into the model, it must first go through the data processing stage. The processing stages include:

1. ***case folding***, namely the uniformity of the letter form, removing signs other than text. Case folding is done by changing words into lower case or lowercase letters.
2. **Tokenization**, namely the word splitting process. In sentences generally use a space character, so the space character is relied on in this tokenization process. At the same time, tokenization also removes certain characters which are considered as punctuation marks.
3. ***Stopword Removal***, namely checking every word in the comments, then eliminating words that are considered unimportant with a dictionary containing words such as 'and', 'or', 'at', 'to', 'this', 'that', etc.
4. ***Stemming***, namely the process of finding the basic word by eliminating all additional words or those attached to the word with the rule of stages is to remove the words 'lah','kah','-an','-i','kan','nya'(Adres , 2019):

Training data processing is done by using the training data input that has been made previously. At this stage, machine learning software, RapidMiner 9.7 is used and uses a data mining classification algorithm as a method to generate a model that can predict whether the tweet is in the positive or negative category. The formed model is stored for further testing on the testing data.

Word Weighting (TF-IDF)

Datasets that have passed the data preprocessing process are then weighted in order to obtain a value so that they can be classified, and is a profitable process for calculating the occurrence of words in a document, at this stage the weighting uses the term frequency-inverse document frequency or tf-idf method.

K-NN Klasifikasi Classification

After going through the Term Weighting stage, the data is ready to be processed using the K-Nearest Neighbor method. The purpose of this process is to group data into predetermined classes by looking at a number of "k" parameters as the closest distance value. With the data sample, it will be used to determine the existence of the word.

Results and Accuracy

The last stage is the data that has been processed, all of which have been divided into several classes of sentiment, both positive and negative sentiments. The results of this sentiment data process will be in the form of graphs and tables that will display the total value and presentation of Twitter user sentiment.

In this study, the data retrieval used was tweet data from the Twitter server, through the Spreadsheet the data was obtained by utilizing the API. Tweets obtained as many as 700 Tweets.

	A	B	C	D	E	F	G	H	
1	Date	Screen Name	Full Name	Tweet Text	Tweet ID	Link(s)	Media	Location	Retwee
12	21/06/2020 18:13	@meluruskanniat	lolos'21	Pertumbuhan dan perkembangan tumbuhan. Maap catatany	1274661635638956032		https://pbs.twimg.com/media/EbCB1S9\		
13	21/06/2020 18:12	@RuangrinduX1	Ruangguru | Bimbel Online N	Lah pacar gw ruang guru? #액스원데뷔_300일축하해 #300d	1274661412887949313		https://pbs.twimg.com/media/EbCBjLYU		
14	21/06/2020 18:10	@schfess	OPFOLL 19:00 WIB!! - SCHOC	sch! bimbel online SMA bagusan yang mana? nf, quipper, ru	1274661046012108801				
15	21/06/2020 18:09	@Ripkifdllh	SUKASUKA.COM	Ruang guru aja gratis, masa classroom bayar	1274660670047309825				
16	21/06/2020 17:18	@subschfess	ON!! - SUB SCHOOLFESS	Sch! Bagusan ruang guru / zenius?	1274647786248523777				
17	21/06/2020 16:27	@Senjaxoxo	Senja♡	Anak" pling g sk klo aku yg msuk ruang ujian mukany pd kec	1274634965255217152	https://twitter.com/aan__/status/1274523615250182144	https://twitter.cc		
18	21/06/2020 16:14	@seollie_	Alin. #IIBSNawasenaNight	Ralienka meninggalkan hall sejenak, melangkah menuju ru	1274631772148396032				
19	21/06/2020 16:06	@pp_igi	PP. Ikatan Guru Indonesia	Webinar Ruang Belajar Bersama; Perkenalan E-Learning C	1274629798602526720	https://www.igi.or.id/webinar-n	https://pbs.twimg.com/media/EbBlCseU		
20	21/06/2020 15:58	@jerabugi	enza	aku udah mikir bakal dijadiin brand ambassador ruang guru	1274627658014310400	https://twitter.com/jerabugi/sta	https://pbs.twimg.com/media/EbBhjofUE		
21	21/06/2020 14:16	@samyangstew	liyn	Yuk masukan kode referral ONLINEDIRUMAH dan dapatkan	1274602161129717761				
22	21/06/2020 14:00	@itsmee_ta17	dm	Lagi liat" lagu di youtube, jadi inget pas kls 3 sma karaokear	1274598059788218368				
23	21/06/2020 12:34	@RisyaPramana	Risya Pramana Situmorang	Ruang Kreatif Guru : Masyarakat Cyber 21: Belajar dari Kore	1274576456962551809	http://ruangkreatifguru.blogspot.com/2020/06/masyarakat-cyber-21-bel			
24	21/06/2020 12:32	@ppppajingan	bilun	kiamatnya diundur dlu, dajjalnya lagi rapat di ruang guru🙏	1274575970960203778				
25	21/06/2020 12:15	@masihkecilsaja	DiRumahAjaCariCuan	Selamat Siang Pacitan , Selamat weekend semesta! Terima	1274571544069410816				
26	19/06/2020 13:04	@lalo38381724	lalo	Prakerja plis insentif nya d proses. Banyak yg btuh duitnya ni	1273859223177138177				
27	15/06/2020 07:46	@perpusdikbud	perpustakaandikbud	How to Make Free Website:Membuat Website Pribadi (Guru)	1272329622361792512	http://repositori.kemdikbud.go	https://pbs.twimg.com/media/Eag5CxTU		
28	14/06/2020 21:06	@renywiradhana	Rey	Hal apa yang membuat kalian masuk ke ruang guru BP?? M	1272168608248946689				
29	14/06/2020 20:52	@tehirene_	7ØXвuиα Teh iren hadid	Ayok belajar bersama guru iren Di ruang buruk	1272165016913932288		https://pbs.twimg.com/media/EaejVxUU\		
30	14/06/2020 20:35	@brownsyugerrr	babo	belajar 3th,habis lulus jd gugelnya temen2 yg pd mo bikin k	1272160716460421120		https://pbs.twimg.com/media/EaefaP9U\		
31	14/06/2020 19:36	@Niaaoneit	Mujidin//Rest	Ruang guru notis aku dong hehe X1 REBOOT #끝난나자_액	1272145861321973760				
32	14/06/2020 19:20	@wnaaq	.	Yup, dulu pas SMP pernah debat sama guru matematika. Ka	1272141809024552961	https://twitter.com/ybrap/status	https://pbs.twimg.com/media/EadV_OgU		
33	14/06/2020 17:51	@dibungkusgula	txtdaripesertautbk	gue nanya ke guru mat dari pertama kali belajar bangun rua	1272119391631507457	https://twitter.com/ybrap/status	https://pbs.twimg.com/media/EadV_OgU		

Fig 3. Dataset

Data Input

From the data that already collected, several columns will be stored, namely text and screenname which are variables needed for a series of modeling processes, table 4.1 shows examples of training data that will be stored for the model formation process.

Table 1. Example of Training Data

Text	*Screenname*	*Sentiment*
Have any of your parents ever got a wa from Ruangguru like this or not, but you don't have a Ruangguru, I don't use a teacher's room even though. this is the chat like an ordinary number, not the Ruangguru number, what is it?	Pipelluv	*negative*
It's really confusing, even if you open the teacher's room and continue to work on the discussion, what's wrong with the brain...	Ditnocure	*Positive*
On the other twitter, there is Zenius on this twitter, this is the teacher's room =))	Potato chips	*Positive*

In Table 2is the input data for testing the modeling algorithm that will be used. The label on the testing data will be determined based on the workings of the algorithm used and based on the knowledge label of the training data that has been manually labeled, here is an example of the testing data used.

Table 2. Example of Model Input Data

Text	*Screenname*
Have any of your parents ever got a wa from Ruangguru like this or not, but you don't have a Ruangguru, I don't use a teacher's room even though. this is the chat like an ordinary number, not the Ruangguru number, what is it?	Pipelluv
It's really confusing, even if you open the teacher's room and continue to work on the discussion, what's wrong with the brain...	Ditnocure
On the other twitter, there is Zenius on this twitter, this is the teacher's room =))	Potato chips

Data Processing

At this stageAlso known as preprocessing. Raw data in the form of a collection of tweets will be converted into data that has a weight value, so that it can be processed at a later stage. At the data processing stage until classification is done using RapidMiner version 9.7 software. The data needed for this classification are learning data (training data) and validation data (testing data). Some of the steps taken are:

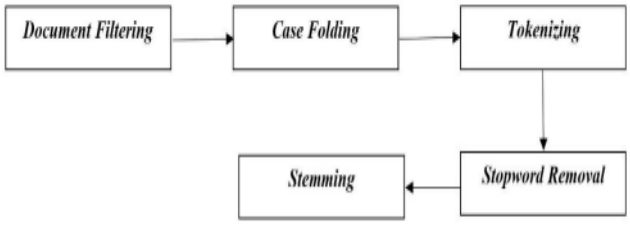

Fig 4. Stages of the Data Processing Process

In general, the series of preprocessing processes on Twitter sentiment can be seen in Figure 4.3 below based on the system.

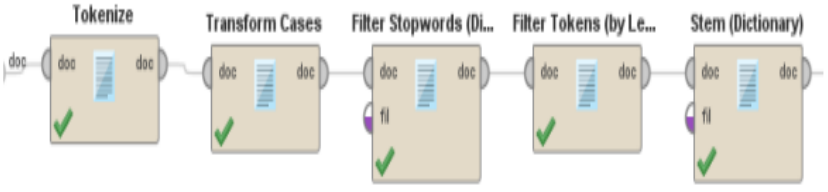

Fig 3. The main process series Preprocessing sentiment

TF-IDF Weighting

At this stage count the number of vocabulary (terms) in each document (comments) in the dataset. The example of term matrix data is taken only 3 documents that are classified manually. The results of the 3 document term matrix data can be seen in Figure 4.4 using equation (3)

D1 = I think Zenius material is cooler than the teacher's room
D2 = the teacher's room appears hopefully it will still help students understand the lesson
D3 = join the teacher's room lessons
D4 = Am I the only one who feels that the teacher's room is more difficult?

Table 3. Weighting Results

Term	TF				DF	IDF		W= TF * IDF		
	D1	D2	D3	D4		log (D/DF)	D1	D2	D3	D4
Menurut	1	0	0	0	1	0,602	0,602	0	0	0
Aku	1	0	0	1	2	0,301	0,301	0	0	0,301
Ya	1	0	0	1	2	0,301	0,301	0	0	0,301
Materi	1	0	0	0	1	0,602	0,602	0	0	0
zenius	1	0	0	0	1	0,602	0,602	0	0	0
lebih	1	0	0	1	2	0,301	0,301	0	0	0,301
keren	1	0	0	0	1	0,602	0,602	0	0	0
daripada	1	0	0	0	1	0,602	0,602	0	0	0
ruang	1	1	1	1	4	0	0	0	0	0
guru	1	1	1	1	4	0	0	0	0	0
muncul	0	1	0	0	1	0,602	0	0,602	0	0
semoga	0	1	0	0	1	0,602	0	0,602	0	0
tetap	0	1	0	0	1	0,602	0	0,602	0	0
bantu	0	1	0	0	1	0,602	0	0,602	0	0
siswa	0	1	0	0	1	0,602	0	0,602	0	0
paham	0	1	0	0	1	0,602	0	0,602	0	0
pelajaran	0	1	0	0	1	0,602	0	0,602	0	0
ikutan	0	0	1	0	0	0,602	0	0	0,602	0
les	0	0	1	0	0	0,602	0	0	0,602	0
apa	0	0	0	1	1	0,602	0	0	0	0,602
cuma	0	0	0	1	1	0,602	0	0	0	0,602
merasa	0	0	0	1	1	0,602	0	0	0	0,602
susah	0	0	0	1	1	0,602	0	0	0	0,602

The IDF values of the sample data are:

- IDF (by) = log (4/1) = 0.602
- IDF(i) = log(4/2) = 0.301
- IDF (yes) = log (4/2) = 0.301
- IDF(Material) = log(4/1) = 0.602
- IDF (zenius) = log (4/1) = 0.602
- IDF (over) = log (4/2) = 0.301
- IDF (cool) = log (4/1) = 0.602
- IDF(than) = log(4/1) = 0.602
- IDF(space) = log(4/4) = 0
- IDF (teacher) = log (4/4) = 0
- IDF (appear) = log (4/1) = 0.602
- IDF (hopefully) = log (4/1) = 0.602
- IDF (fixed) = log (4/1) = 0.602
- IDF(auxiliary) = log(4/1) = 0.602
- IDF (students) = log (4/1) = 0.602
- IDF (understand) = log (4/2) = 0.301

- IDF (lesson) = log (4/1) = 0.602
- IDF (follow-up) = log (4/1) = 0.602
- IDF (les) = log (4/1) = 0.602
- IDF(what) = log(4/1) = 0.602
- IDF (Only) = log (4/1) = 0.602
- IDF(felt) = log(4/1) = 0.602
- IDF (hard) = log (4/1) = 0.602

The formula for TF-IDF is in equation 3, where tf represents the value of term frequency and log (N/df) represents the value of IDF with N being the number of data.

Algorithm Application

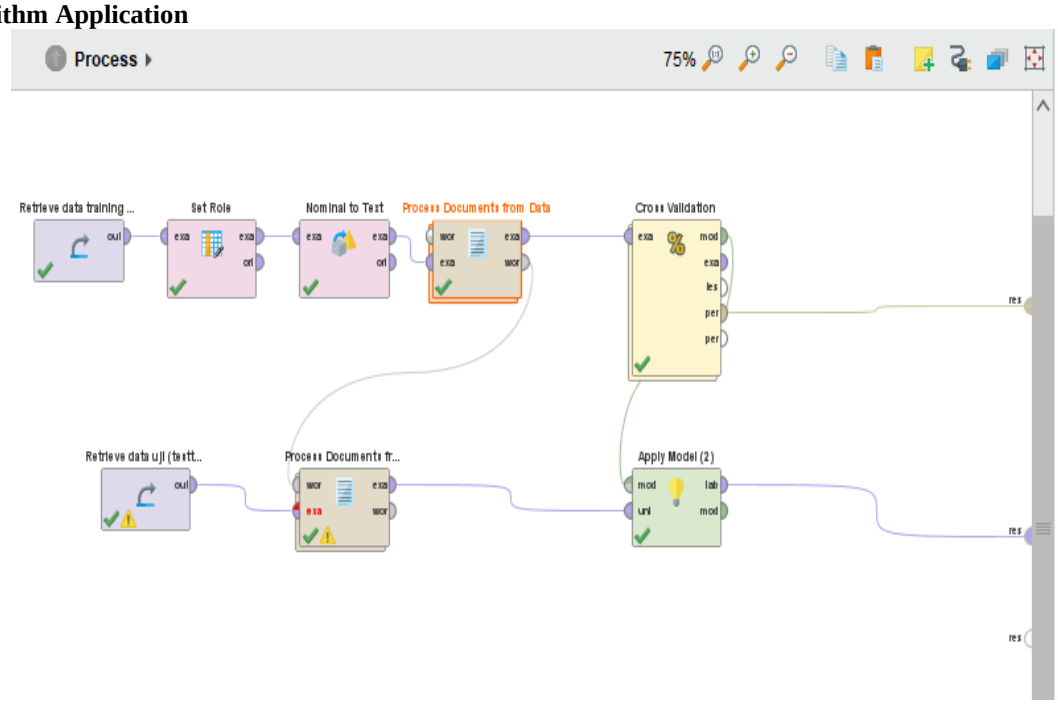

Fig 4. Algorithm Implementation in RapidMiner

K-NN Algorithm Manual Calculation

In the classification process using K-NN, the first thing to do is determine the value of k, K-NN uses the concept of classification with its nearest neighbor which is used as a predictive value for new data. The closest distance is needed to determine the number of similarities that the new data has. Some of the processes in K-NN are as follows:

1. Determine the value of K, for example K=1. Then the closest 1st document is used as a classification result
2. Calculate the distance between the new data in each data label with the distance of the training data. To calculate the distance, the Euclidean Distance equation will be used. As the formula for equation 1 as follows:

$$D\ (X,Y) = \sqrt{\sum_{k=1}^{n}(Xk - Yk)^2} \qquad (1)$$

3. Then sort the results based on the provisions of the K value you want to use. If K = 2, then the 2 closest distances will be selected as the classification results.

Table 4. Training Data and Test Data

	Doc	Tweet	Sentiment
Training Data	1	I think Zenius material is cooler than the teacher's room	*negative*
Training Data	2	The teacher's room appears again on TV, hopefully it will help students understand the lesson	*Positive*
Training Data	3	join the teacher's room lessons	*Positive*
Testing Data	4	Am I the only one who feels that the teacher's room is more difficult to understand?	?

To classify whether document 4 is positive or negative, a calculation is carried out, for example K = 3 as follows:

D1 and D4

$$= \sqrt{\begin{matrix}((0{,}602-0)^2\,x\,5)+((0{,}301-0{,}301)^2 x\,3)\\ +((0-0)^2\,x\,11)+((0-0{,}602)^2\,x\,4)\end{matrix}}$$

$= \sqrt{1{,}82+0+0+1{,}448}$

$= \sqrt{3{,}258}$

= 1.804

D2 and D4

$$= \sqrt{\begin{matrix}((0-0)^2\,x\,9)+((0-0{,}301)^2 x\,3)\\ +((0{,}602-0)^2\,x\,7)+((0-0{,}602)^2\,x\,4)\end{matrix}}$$

$= \sqrt{0+\ 0{,}27+2{,}534+1{,}448}$

$= \sqrt{4{,}252}$

= 2.062

D3 and D4

$$= \sqrt{\begin{matrix}((0-0)^2\,x\,14)+((0-0{,}301)^2 x\,3)\\ +((0{,}602-0)^2\,x\,3)+((0-0{,}602)^2\,x\,4)\end{matrix}}$$

$= \sqrt{0+\ 0{,}27+0{,}724+1{,}448}$

$= \sqrt{2{,}442}$

= 1.562

K = 3

K=1	1,562 (Document 3 = Positive)
K=2	1,804 (Document 1 = Negative)
K=3	2,062 (Document 2 = Positive)

So, because the closest distance is at K = 1 and sees the dominance of its neighbors, document 4 is included in the positive category.

Table 5. Experiments changing the value of K on Rapidminer

K . value	Accuracy	Precision	Recall	AUC
1	87.50%	62.97%	61.67%	0.500%
2	87.50%	62.97%	61.67%	0.843%
3	87.32%	66.14%	55.00%	0.895%
4	87.50%	66.40%	55.00%	0.904%
5	86.61%	64.40%	51.67%	0.907%
6	86.61%	64.06%	48.33%	0.909%
7	87.32%	65.46%	53.78%	0.910%
8	87.14%	65.93%	52.67%	0.914%
9	88.04%	70.70%	57.11%	0.915%
10	87.68%	68.45%	52.67%	0.913%
11	87.86%	65.89%	60.44%	0.913%
12	86.96%	65.70%	52.17%	0.912%
13	**88.21%**	**68.71%**	**59.33%**	**0.915%**
14	87.68%	67.90%	51.67%	0.916%
15	87.86%	70.98%	49.56%	0.913%

Based on the changes in the K value that have been made, the best accuracy results are in position K=13 with an accuracy value of 88.21% and an AUC value of 0.915%.

Table 6. The results of the accuracy confusion matrix using the K-Nearest Neighbor algorithm with a value of K=13

Accuracy: 88.21% +/- 4.78% (micro average: 88.21%)			
	True positive	True negative	Class precision
Pred. positive	440	37	92.24%
Pred. negative	29	54	65.06%
Class recall	93.82%	59.34%	

The accuracy value of the confusion matrix with equation (4) is:

Accuracy = (TN+TP) / (TN+FN+TP+FP)
= (54+440) / (54+29+440+37)
= 494 / 560
= 0.8821
= 88.21%

The following picture is a terminal in carrying out the implementation stages with the K-NN Algorithm on Rapid Miner

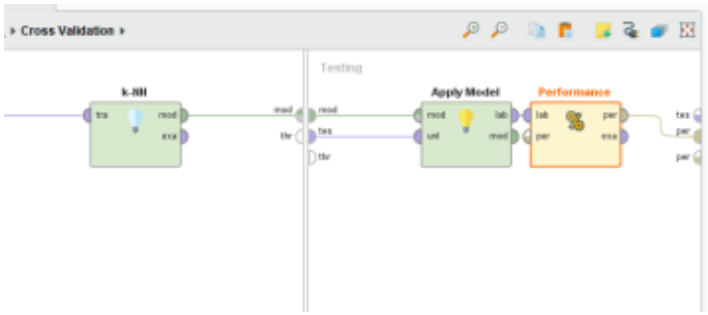

Fig 5. Algorithm use in Rapidminer

Research Results (Output)

The results obtained from the research carried out are that the percentage tendency of twitter user sentiment assessment towards Ruangguru tends to be positive with an accuracy value of 88.21% with 15 trials the K parameter is at K = 13.

Based on the results of the research that has been done, it can be concluded that K-Nearest Neighbor (K-NN) can be applied to conduct sentiment analysis with an accuracy value above 70%. By dominating the positive category on the effectiveness of twitter users towards Ruangguru. Below is a graph of the AUC parameter k=13.

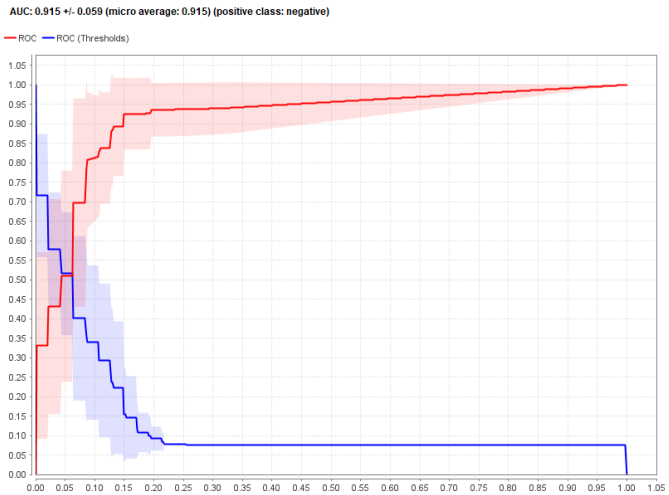

Fig 6. Graph of AUC parameter k=13

Based on the results of the calculation with the confusion matrix, it was found that the prediction results made 54 negative data that were correctly predicted and 440 positive data that were correctly predicted, using the K-Nearest Neighbor algorithm with a value of k = 13 which got the best accuracy results on a value scale. k= 1-15.

4. Conclusions

The conclusions obtained from this research, namely:

1. By applying the K-Nest Neighbor algorithm, the results of community sentiment through Twitter social media display sentiments that show public opinion towards Ruangguru tends to be positive, although there are also negatives seeing the impact of the current pandemic period in Indonesia, the tweet is divided into 440 sentiments that are positive and 54 negative sentiments.
2. Testing sentiment analysis in Indonesian with the K-Nearest Neighbor method produces the best accuracy at k=13 value of 88.21%.

REFERENCES

Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, ED, Gutierrez, JB, & Kochut, K. (2017). A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques. http://arxiv.org/abs/1707.02919

Bibi, S. (2015). Effectiveness of Application of Blended Learning Algorithm and Programming Courses. 1, 274–286.

Dewi Onantya, I., Indriati., & Pandu Adikara, P. (2019). Sentiment Analysis in Reviews of Using BCA Mobile Applications. 3(3), 2575–2580.

Dewi Onantya, I., Indriati, I., & pand. (2019). Sentiment Analysis on BCA Mobile Application Reviews Using BM25 and Improved K-Nearest Neighbor. Journal of Information Technology and Computer Science Development, 3(3), 2575–2580.

Fauziah, S., Sulistyowati, DN, & Asra, T. (2019). Optimization of the Vector Space Model Algorithm with the K-Nearest Neighbor Algorithm in Searching for Journal Article Titles. Journal of Pilar Nusa Mandiri, 15(1), 21–26. https://doi.org/10.33480/pilar.v15i1.27

Karami, A. (2020). Twitter and Research : A Systematic Literature Review Through Text Mining. IEEE Access, 8, 67698–67717. https://doi.org/10.1109/ACCESS.2020.2983656

Mentari, ND, Fauzi, MA, & Muflikhah, L. (2018). Analysis of 2013 Curriculum Sentiment on Twitter Social Media Using 2013 Curriculum Sentiment Analysis on Twitter Social Media Using the K-Nearest Neighbor Method and Feature Selection Query Expansion Ranking. August.

Nofitri, R., & Irawati, N. (2019). Data Analysis Results Advantages Using Rapidminer Software. V(2), 199–204.

Nurjanah, WE, Perdana, RS, & Fauzi, MA (2017). Sentiment Analysis on Television Shows Based on Public Opinion on Twitter Social Media using the K-Nearest Neighbor Method and Sentiment Analysis on Television Shows Based on Public Opinion on Twitter Social Media using M. October.

Subhan, A., Sediyono, E., & Dwi, O. (2015). Ontology-Based Sentiment Analysis at Sentence Level to Measure Product Perception. 02, 84–97.