

# K-Nearest Neighbors Algorithm to Student Opinion of the Online Learning Method at Wira Wacana Sumba Christian University

Andry A.P. Tanggu Mara<sup>1</sup>, Eko Sedyono<sup>2</sup>, Hindriyanto Purnomo<sup>3</sup>

Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana, Salatiga, Indonesia  
<sup>1</sup>972019011@student.uksw.edu, <sup>2</sup>eko@uksw.edu, <sup>3</sup>hindriyanto.purnomo@uksw.edu

Accepted 01 June 2021

Approved 30 July 2021

**Abstract**— The education sector is one of the areas that has felt the major impact of the Covid-19 pandemic. The impact that arises is the teaching and learning process must be carried out from home using the online learning method. This teaching and learning method raises a variety of responses from students. This is what makes researchers analyze these views, both in the form of positive opinions or negative opinions. The analysis process is carried out by applying sentiment analysis or opinion mining from the comment on Facebook, text mining is processed using the preprocessing method, labeled it to positive and negative. Based on the available data, a classification process is carried out using the K-Nearest Neighbors algorithm. Rapid Miner is used to experiment text data with the KNN algorithm in order to find the value of accuracy, precision and recall. From the results of research, it was obtained a value of 87.00% for accuracy and 0.916 for the AUC value. The values are high enough for the classification of student opinion against this pandemic so that this research is classified as Excellent Classification.

**Index Terms**— Covid-19; Online Learning Mode; K-Nearest Neighbors; Sentiment Analysis; RapidMiner

## I. PRELIMINARY

The outbreak of the Corona Disease Virus 2019 or also known as Covid19 has had a huge impact on the teaching and learning process at all levels of education in all parts of the world. Specifically in Indonesia, the teaching and learning system is carried out from home, this is based on government regulations and recommendation through the Indonesian Minister of Education and Culture (Mendikbud) for all students to study from home (BDR) and even the 2020 national exam was cancelled. The study from home policy is implemented with the aim of limiting and reducing physical contact as an effort to prevent virus transmission [1].

Wira Wacana Christian University Sumba is one of the universities affected by the Covid 19 outbreak so that all student activities are "homed", learning to teach from home using online or online learning models. As of March 2020 where the government's recommendation regarding Social Distancing during the Covid 19 pandemic was issued by Unkriswina through the Chancellor's Circular Number 019/EDR-R/2020 explaining that it is necessary to pay attention to measures to prevent the spread of Covid 19 in the campus environment so that lectures are carried out online. or online(Unkriswina, 2020). The education system is carried out using an online method where every learning activity is carried out virtually through the media presented on the internet. Of course there is an impact, whether it supports or even harms the students. The positive impacts obtained are being safe from the Covid 19 outbreak, learning is more practical and flexible, saving time and energy, as well as a more personal learning approach(Pakpahan, R., & Fitriani, 2020). However, there are several problems that arise as a result of the application of the learning method from home using online media as an example of an unstable and adequate network to be accessed properly considering that Indonesia is an archipelagic country whose internet infrastructure development is uneven. There are obstacles for students and students with low economics considering that it takes money to access the internet. The effectiveness of online learning also depends on the maturity and readiness of the school, in this case the teaching teacher. There are many teachers who are not able to convey material effectively through online learning systems [2].

Based on the existing impacts, the focus to be studied is students using the Sentiment Analysis method which is a process to find out a person's views or opinions on events that occur, whether it is a positive or negative view. Opinions and views of students can be in written or oral form. K-Nearest Neighbors is an algorithm used to perform the opinion classification process resulting from sentiment analysis. The advantage of the K-Nearest Neighbor algorithm is the high accuracy value in calculations that have been

proven and have been applied in applications [3]. The aim is to classify student opinions in the form of positive and negative opinions and the accuracy value will be calculated using the K-NN algorithm approach. The results can be used as evaluation material regarding online learning models during the pandemic.

Previous research related to the classification of sentiment analysis is classification of wikipedia articles by Hardiyanto and Rahutomo (2016). The Indonesian Wikipedia article classification is intended for the classification of articles on the Indonesian Wikipedia website in text form by using a text pre-processing model and then forwarded by TF-IDF weighting. Based on this weighting, the articles in the Indonesian Wikipedia are classified using the K-Nearest Neighbor Algorithm. The results of the manual test show the accuracy of the truth with a value of 60% [4]. Comparison with this research is the classification process which is carried out manually without the help of classification tools or tools.

The research of Siti Ernawati and Risa Wati in 2018, The application of the K-Nearest Neighbors algorithm in the sentiment analysis of travel agent reviews shows the following: processing 100 positive and negative review data with the K-Nearest Neighbor (K-NN) algorithm related to sentiment. Experiments and results show that by using the K-Nearest Neighbor (K-NN) algorithm, it achieves high accuracy results and is classified as the best accuracy value of 87.00% and the AUC point of 0.916[5]. There are several things that are lacking from this research, namely the KNN algorithm is applied theoretically while the formulation or algorithm equation in finding distance or proximity data as a classification process is not applied.

Research related to the use of online learning models conducted by Toni Limbong (2020) shows the following: as an effort to support the Government's decision regarding the spread of the Corona Virus, the Catholic University of Santo Thomas Medan applies an online or online-based learning model. The researcher applied the Multi Attribute Utility Theory Method to a case study of the effectiveness of online learning using the Zoom and Edmodo applications at the Faculty of Computer Science, Santo Thomas Catholic University, Medan and obtained objective results with the assessment of Theory (0.88) as the highest assessment, followed by Theory and Practicum courses (0.70) , Practicum courses (0.42) and Field courses (0.20). The conclusion that the researcher obtained and became a reference for the decision if the university leadership would make an online or online exam policy, then the form of questions and the nature of the exam were in the form of theory, such as: multiple choice, essay and also analysis [6]. This study uses pseudocode and a combination of opinions so that the conclusions generated are solely the views of the researchers.

Some of the basic things that will be used in this research are the application of sentiment analysis and the K-Nearest Neighbors algorithm as a method for classifying Unkriswina students' views on the use of online learning during the Lock Down period due to the Corona Virus outbreak. The application of pseudocode or distance and proximity calculation algorithms as part of the classification process will be applied. Of course, the main target is how to find out the views of students while participating in online learning, which of course their views vary so that they need to be classified and separated positive opinions and negative opinions.

## II. RESEARCH MATERIALS AND METHODS

### A. Sentiment Analysis

Sentiment analysis or also called opinion mining is a computational or computational study in finding and identifying opinions, attitudes, emotions, sentiments, evaluations, subjectivity, judgments contained in a text. Sentiment analysis is intended to find the percentage value of positive labeled sentiment and negative labeled sentiment towards a person, object or in a certain condition. Sentiment analysis has 3 values that are generally used, namely: positive sentiment, negative sentiment and neutral sentiment [7].

The steps of sentiment classification analysis of text-mining data or text data are as follows: 1). Initial stage: Collecting datasets such as public opinion, ratings of restaurants or products and others. 2). Pre-processing: stages in text-mining to convert raw data into important information where the stages include: Tokenization, Stopwords Removal, and Stemming. 3). Transformation: Weighted text data. 4). Feature Selection: The stage of limiting and reducing data that is not needed. 5). Classification: Classification stages such as: Naive Bayes, K-Nearest Neighbor, Support Vector Machine and others. 6). Interpretation/Evaluation: The evaluation stage is to calculate the accuracy value and the Area Under the Curve value[8].

### B. K-Nearest Neighbors

In the book Data Mining Algorithm, Kusri explains that the K-Nearest Neighbors Algorithm is an approach to finding cases by calculating the proximity between a new case and an old case through a weight matching process from a number of available features [9]. In another view, it is stated that K-Nearest Neighbors is an algorithm for classifying objects based on the data that is closest to the object. Data is illustrated in many dimensional spaces, where each dimension reflects a feature of the object. Accurate k values for this algorithm depend on data with high k values [10].

The main purpose of this algorithm is to classify an object based on the attributes and training sample. The K-NN model applies a classification that refers to the

proximity of the points of existing objects as the approximate value of the new sample [11].

The method used is by observing the discussion and comment forms in the Facebook group which is then used as a dataset to be processed as research subjects. The stages of the research are as follows:

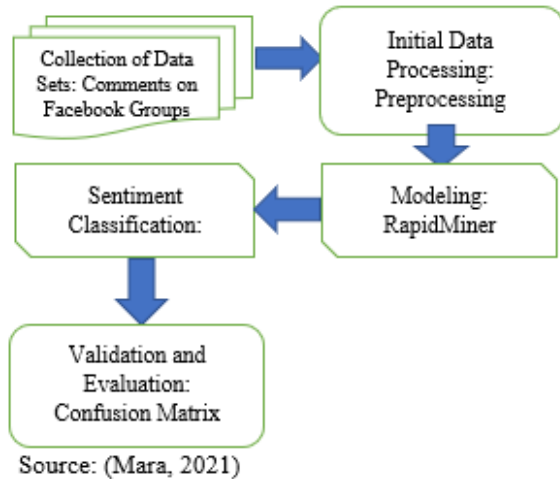


Figure 1. Proposed Method

Figure 1, are the stages applied in this study with the following explanation.

#### C. Collection of Data Sets

The first stage is collecting data by creating a Facebook group account consisting of active Unkriswina students and providing space to answer questions related to online learning in the midst of the Covid 19 pandemic. The data in question are opinions or views from Wira Wacana Sumba Christian University students as targets. study. The data obtained is still in the form of a collection of opinions so that it must be processed first into a dataset.

#### D. Initial Data Processing

The data sample used for training is as much as 200 data. At the initial processing of data through pre-processing as follows:

##### i. Case Folding

*Case Folding* is the process of converting all letters of text data to lowercase.

##### ii. Tokenization

*Tokenization* is the stage of separating words, phrases, punctuation marks or symbols.

##### iii. Stopword Removal

Then next is the process of eliminating the stopword list which is a list of connecting words between sentences.

##### iv. Stemming

*Stemming* aims to change or replace tokens that have affixes into basic words. For example, the word remove is changed to replace [12].

#### E. Modeling

Experiments in processing text data in this study using RapidMiner 8.2. The training data used are opinions Unkriswina Sumba students obtained from Facebook Social Media and grouped into 2 parts, namely: positive opinions and negative opinions.

#### F. Sentiment Classification

K-Nearest Neighbors is a classification method used in this research. Equation 2 below is an equation to calculate the proximity between two cases:

$$\text{Similarity}(T,S) = \frac{\sum_{i=1}^n f(T,S) * w}{w} \quad (1)$$

Information:

T : New Object

S : Old Object

n : Number of Attributes On Each Object

i : Individual Attribute 1 Until N

f : Function Similar Attribute I between T and S

w : i-th Attribute Weight

Generally, Similarity is at a value between 0 to 1, a value of 0 is that the two objects are absolutely not similar, while a value of 1 indicates that the object is similar to absolute [13].

#### G. Validation and Evaluation

The validation stage is carried out by applying 10-fold-cross validation. The validation process has two sub-processes, namely: training set and testing set. The training data sub-process is intended to be used in RapidMiner modeling which will then be tested. The evaluation or testing of the results of the K-NN classification uses the Confusion Matrix.

TABLE 1. CONFUSION MATRIX

		PREDICTION	
		POSITIVE	NEGATIVE
CURRENT	POSITIVE	TP	FN
	NEGATIVE	FP	TN

Source: (Han, Kamber, & Pei, 2012)

i. TP (True Positive): The positive value obtained corresponds to the actual value.

ii. TN (True Negative): The negative value obtained corresponds to the actual value.

iii. FP (False Positive): Positive value but does not match the actual value.

iv. FN (False Negative): Negative value but does not match the actual value.

### III. RESULT AND DISCUSSION

The data is obtained from the comments in the Facebook group discussion form: <https://www.Facebook.com/groups/465615134612746> with more than 200 respondents and 291 comment data. The data were then given positive and negative labels so as to produce 200 comment data, these labeled data were used as research datasets. Rapidminer version 8.2 is used with the aim of obtaining a model that suits research needs.

#### A. Opinion Document Collection

The Opinion Document used is the result of collecting the opinions of Unkriswina students, as many as 200 opinion documents consisting of 26 positive opinions and 174 negative opinions. These 200 positive and negative opinion data are used as training documents and there are 50 testing documents.

TABLE 2. OPINION DOCUMENTS

No.	Student Opinion	Training Opinion	Opinion Testing
1	Positive	25	-
2	Negative	175	-
3	Total	200	50

The Opinion Document in table 1 will be processed with the Preprocessing approach and the application of opinion classification by prioritizing the data normalization stage.

TABLE 3. OPINION DOCUMENTS BEFORE PRE-PROCESSING

No.	Opinion Document
1	Studying from home can be while playing with friends
2	Online learning is not at all good, I mostly don't understand
3	BASICALLY if this online learning model is implemented seriously by the lecturer and we are STUDENTS... I think everything will be fine 😊 and we also get SCIENCE with EFFECTIVE !!!
4	Online lectures are very inconvenient, costs a lot to buy packages, the network is so unstable here
5	I don't have a cellphone let alone a laptop, online lectures for me have to find more money to buy a cellphone
6	More of us are required to learn on our own actually
7	The lectures are ok but I don't know the lecturers and friends in class, I can only see they have photos
8	Online lectures but if there is no internet is the same as lying, it is difficult.
9	Most of the lecturers teach not clear, suddenly give assignments. Few materials are taught, a myriad of tasks are given
10	I don't concentrate when studying online, not to mention if the network has been disrupted, it's already bad

Table 3 is the initial opinion documents on training data that have not been preprocessed. The following is a preprocessing stage with a case folding, tokenizing, stopword removal approach in opinion documents in table 3.

#### B. Pre-processing Comment Data

Before the dataset is classified using the K-Nearest Neighbors method, as an initial stage, pre-processing will be carried out as follows:

TABLE 4. PREPROCESSING STUDENT OPINIONS

No.	Student Opinion
1	study from home while playing with friends
2	online learning is not good I don't understand
3	the basic online learning model is applied seriously by lecturers and students, all will be good and get effective knowledge
4	online college is a hassle to buy an unstable network package
5	I don't have a laptop, I have to find money to buy a cellphone
6	we need to learn on our own
7	college is ok, don't know the lecturers, friends, see their photos foto
8	online college no internet is the same as a lie
9	The teaching lecturer is not clear, suddenly gives material assignments, a little teaching warehouse assignments
10	lack of concentration online lectures, network, severe interference

Next is the determination of the frequency term in the training data resulting from the preprocessing approach, as shown in table 5 below:

TABLE 5. TERM FREQUENCY OF TRAINING DATA

No.	term	Id_document	Amount
1	studying	1	2
2	from	1	1
3	House	1	1
4	while	1	1
5	play	1	1
6	friend	1	1
7	study	2	1
8	on line	2	2
9	no	2	2
10	good	2	1
11	no	2	2
12	understand	2	1
13	studying	4	2
14	on line	4	2
15	Troubled	4	1
16	go out	4	1
17	cost	4	1
18	many	4	1
19	buy	4	1
20	package	4	1

The result of the term frequency of training data is in the form of word tokens which are then carried out in the classification stage, but first class labeling is carried

out on each opinion with the aim that the tools used can identify class documents from student opinions.

TABLE 6. LABELING OF TRAINING DATA

No.	Student Opinion	Class Sentiment
1	study from home while playing with friends	Positive
2	online learning is not good I don't understand	Negative
3	the basic online learning model is applied seriously by lecturers and students, all will be good and get effective knowledge	Positive
4	online college is a hassle to buy an unstable network package	Negative
5	I don't have a laptop, I have to find money to buy a cellphone	Negative
6	we need to learn on our own	Negative
7	college is ok, don't know the lecturers, friends, see their photos foto	Negative
8	online college no internet is the same as a lie	Negative
9	The teaching lecturer is not clear, suddenly gives material assignments, a little teaching warehouse assignments	Negative
10	lack of concentration online lectures, network, severe interference	Negative

Table 6 is the stage of labeling student opinion training data which will then be tested with the K-Nearest Neighbors algorithm.

C. Classification Using the K-Nearest Neighbors Method

Before the classification is carried out, first the calculation of the proximity of the distance is carried out using the existing equations, namely:

$$d = \sqrt{\sum_{i=1}^n (a_i - b_i)^2} \tag{2}$$

The comparison is on sample data and test data. Sample data used for example id\_document = 1 with test data id\_document = x. Then the application of the formula is as follows:

$$d_1 = \sqrt{\sum_{i=1}^p (1)^2 + (-0,07175)^2 + (0,5)^2 + (1)^2 + (0,2222222)^2 + (1)^2} = 1.83494708 \tag{3}$$

TABLE 7. DATA DISTANCE BETWEEN TEST DATA

Test Data	Sample Data	Distance
Id_document = x	Id_document = 1	1.83494708
	Id_document = 2	1.45308990
	Id_document = 3	1.17634227
	Id_document = 4	0.08900544
	Id_document = 5	1.07443387
	Id_document = 6	0.45009112
	Id_document = 7	1.09231099
	Id_document = 8	1.52096677

	Id_document = 9	1.69230019
	Id_document = 10	0.03670674

The distances in the test data as shown in table 7 can be sorted into the closest as follows:

TABLE 8. RESULTS OF NEAREST DISTANCE COUNT

Test Data	Sample Data	Distance
Id_document = x	Id_document = 10	0.03670674
	Id_document = 6	0.45009112
	Id_document = 4	0.08900544
	Id_document = 5	1.07443387
	Id_document = 7	1.09231099
	Id_document = 3	1.17634227
	Id_document = 2	1.45308990
	Id_document = 8	1.52096677
	Id_document = 9	1.69230019
	Id_document = 1	1.83494708

The following are the stages of data processing using RapidMiner tools from the results of pre-processing in the early stages.

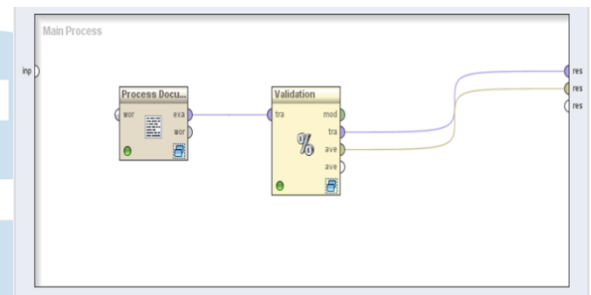


Figure 2. Input Validation (Mara,2021)

Figure 2, is the process of data validation into Rapidminer tools for later classification.

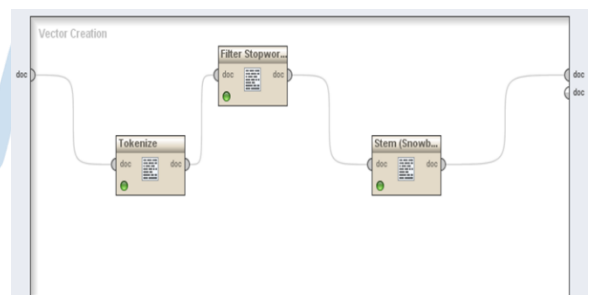


Figure 3. Pre-processing stage (Mara,2021)

Figure 3, is the pre-processing stage of the data set that has been determined, step by step is applied to obtain the expected final result.

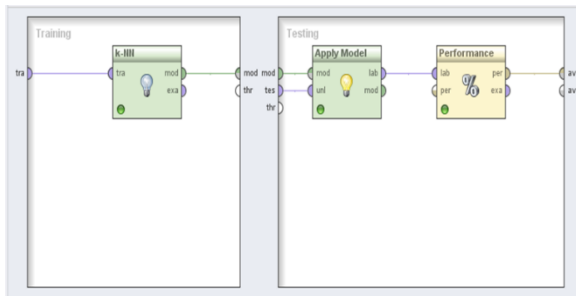


Figure 4. Application of K-Nearest Neighbors (Mara,2021)

Figure 4, is the data classification stage based on the data set that has been preprocessed.

TABLE 9. EVALUATION OF TRAINING DOCUMENTS WITH KNN

N O	CLASS SENTIMENT	WEIGHT   ACCURACY	PRECISION	RECALL
1	POSITIVE	26   13%	1.0	1.0
2	NEGATIVE	174   87%	1.0	1.0
	TOTAL/AVERAGE	GE100%	1.0	1.0

Table 9 is the result of the K-Nearest Neighbors sentiment training process with an accuracy value of 13% for positive sentiment and very large accuracy on negative sentiment reaching 87% of the total 100% accuracy. While the precision results in the positive sentiment class are worth 1 and the negative sentiment class is worth 1, thus the precision results are very accurate. Meanwhile, the recall results are the same as the precision results, with the results of positive and negative sentiment classes being worth 1. It means that the results of the training data sentiment are correct for all sentiment classifications.

D. Change in k Value

The following is an experimental process by changing the k value to determine the accuracy, precision, recall, and AUC values:

TABLE 10. EXPERIMENTAL CHANGES IN THE VALUE OF K

K. VALUE	ACCURACY	PRECISION	RECALL	AUC
1	80.05	77.58	84.00	0.500
2	83.50	89.88	74.00	0.833
3	81.50	78.48	85.00	0.854
4	82.00	84.47	79.00	0.814
5	82.50	76.77	85.00	0.857
6	83.50	88.59	78.00	0.918
7	84.00	86.54	83.00	0.911
8	<b>87.00</b>	<b>92.94</b>	<b>81.00</b>	0.916
9	85.00	85.05	84.00	0.926
10	84.50	87.42	81.00	0.919
11	86.50	87.16	86.00	0.912
12	85.50	88.51	81.00	0.914
13	86.00	87.70	84.00	0.912
14	85.00	89.90	79.00	0.911
15	83.50	85.08	81.00	0.912

Based on table 10, namely the change in the value of k, the highest accuracy results with a value of 87.00% and the AUC value of 0.916 is at the value of k=8.

E. Validation and Evaluation

Based on table 10, Accuracy with the KNN method produces a Negative class recall value of 93.00% and Positive of 81%, Negative Precision class values of 83.04% and Positive of 92.05%. Accuracy with the KNN method in table 10 is derived into the Confusion Matrix. The results are as follows:

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

$$Accuracy = \frac{(93+81)}{(93+7+81+19)} \tag{4}$$

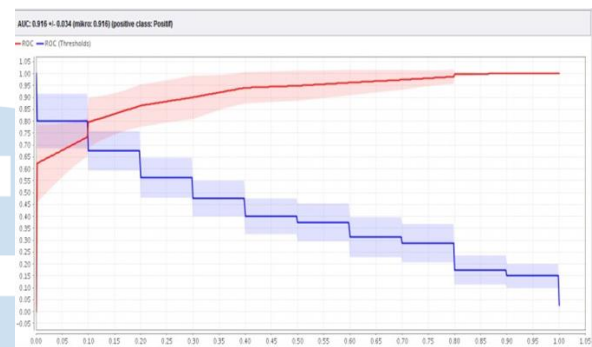


Figure 5. KNN on the AUC Graph (Mara, 2021)

Based on the accuracy value, an AUC graph can be made as shown above. Figure 5 shows the AUC graph with the application of the K-Nearest Neighbors method, resulting in the Area Under Curve (AUC) = 0.916. So from the values that have been obtained, it is concluded that the accuracy classification in this study is included in the Excellent Classification which can be seen in the guide to the accuracy of the AUC value as below:

- i. 0.90 - 1.00: Excellent Classification,
- ii. 0.80 - 0.90: Good Classification,
- iii. 0.70 - 0.80: Fair Classification,
- iv. 0.60 - 0.70: Poor Classification,
- v. 0.50 - 0.60: Failure

Source: (Han et al., 2012)

IV. CONCLUSION

Research has been carried out by applying the classification of comment data from Facebook social media by taking into account the views of Wira Wacana Sumba Christian University students on the use of online or online learning models during the Covid-19 pandemic by utilizing the K-Nearest Neighbors method. The data of 26 positive reviews and 174 negative reviews were used as a dataset and then classified with the results of an accuracy value of

87.00% and an AUC value of 0.916 so that these results were used as a reference for classifying the classification group, namely Excellent Classification. Based on the results of the study obtained a very good accuracy value and obtained a fairly large AUC value. In relation to the online learning process, the value generated is in the form of a negative opinion presentation of 87.00% indicates that most of the student population considers online learning that has been implemented so far to be ineffective. The presentation of the negative opinion is supported by an accuracy value based on the K-Nearest Neighbors algorithm of 87%, which means that the presentation has a very high validity value. So it can be concluded that there needs to be an evaluation of the online learning model at Unkriswina Sumba, both the competence of lecturers in teaching online, the handling of uneven internet networks for students, solutions for students who are economically limited and unable to study online. The presentation of the negative opinion is supported by an accuracy value based on the K-Nearest Neighbors algorithm of 87%, which means that the presentation has a very high validity value. So that it can be concluded that there needs to be an evaluation of the online learning model at Unkriswina Sumba, both the competence of lecturers in teaching online, the handling of the uneven internet network for students, solutions for students who are economically limited and unable to study online. The presentation of the negative opinion is supported by an accuracy value based on the K-Nearest Neighbors algorithm of 87%, which means that the presentation has a very high validity value. So it can be concluded that there needs to be an evaluation of the online learning model at Unkriswina Sumba, both the competence of lecturers in teaching online, the handling of uneven internet networks for students, solutions for students who are economically limited and unable to study online.

#### REFERENCES

- [1] Asmara. C. G., 2020. Para Siswa, Simak Surat Terbaru Mendikbud Nadiem Untuk Kalian. CNBC Indonesia. Retrieved from [https://www.cnbcindonesia.com/tech/2020032512\\_4101-37-147405/para-siswa-simak-surat-terbaru-mendikbud-nadiem-untuk-kalian](https://www.cnbcindonesia.com/tech/2020032512_4101-37-147405/para-siswa-simak-surat-terbaru-mendikbud-nadiem-untuk-kalian).
- [2] Unkriswina. 2020. Surat Edaran Rektor perihal Pencegahan Penyebaran Virus Corona. Retrieved from Kampus Unkriswina website: <https://unkriswina.ac.id/informasi-kampus-unkriswina/surat-edaran-rektor-perihal-pencegahan-penyebaran-virus-corona-18-september-2020>
- [3] Ernawati. S., W. R., 2018. Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Review Agen Travel. *Jurnal Khatulistiwa Informatika*: Vol. 4 No. 1.
- [4] Fatyanosa, T. N., Bachtiar, F. A., 2017. Classification method comparison on Indonesian social media sentiment analysis. *IEEE, 2017 International Conference on Sustainable Information Engineering and Technology (SIET)*.
- [5] Han, J., Kamber, M., Pei, J., 2012. Data Mining: Concepts and Techniques. In *Data Mining: Concepts and Techniques* (3rd ed.). Amsterdam: Elsevier. Retrieved From <https://doi.org/10.1016/C2009-0-61819-5>
- [6] Jiahua, W., Shuai, M., Sahisnu, L. B., 2020. A Knowledge-Driven Approach to Classifying Object and Attribute Coreferences in Opinion Mining. to appear in Findings of 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP-2020, findings). Retrieved From <https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html>.
- [7] Kurniawan, S., Gata, W., Puspitawati, D. A., -, N., Tabrani, M., & Novel, K., 2019. Perbandingan Metode Klasifikasi Analisis Sentimen Tokoh Politik Pada Komentar Media Berita Online. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 3(2), 176–183. Retrieved From <https://doi.org/10.29207/resti.v3i2.935>
- [8] Kusriani, K., & Luthfi, E. T., 2018. *Algoritma Data Mining* (1st ed.). Yogyakarta: Penerbit Andi. Retrieved from <http://andipublisher.com/produk-0907003050-algoritma-data-mining.html>
- [9] Lidya, K., S., Opim, S., Syahril, E., 2018. Sentiment Analysis Pada Text Bahasa Indonesia Menggunakan Support Vector Machine (SVM) dan K-Nearest Neighbor (K-NN). *Seminar Nasional Teknologi Informasi Dan Komunikasi*.
- [10] Limbong T., S. J., 2020. Menentukan Matakuliah yang Efektif Belajar Daring (Belajar dan Ujian) dengan Metode Multi-Attribute Utility Theory (MAUT). *Jurnal Rekayasa Sistem Dan Teknologi Informasi (RESTI)*, 4(2), 370–376.
- [11] Morgan, H., 2020. Best Practices for Implementing Remote Learning during a Pandemic. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 3, 134–140.
- [12] Pakpahan, R., Fitriani, Y., 2020. Analisa Pemanfaatan Teknologi Informasi Dalam Pembelajaran Jarak Jauh di Tengah Pandemi Virus Corona Covid-19. *JISAMAR: Journal of Information System, Applied, Management, Accounting and Research*, 2(4).
- [13] Romadloni, N.T., Santoso, I., Budilaksono, S. 2019. Perbandingan Metode Naïve Bayes, KNN dan Decision Tree Terhadap Analisis Sentimen Transportasi KRL Commuter Line. *IKRA-ITH Informatika*, 3(2).
- [14] Sani, R. R., Zeniarza, J., Luthfiarta, A., 2016. Pengembangan Aplikasi Penentuan Tema Tugas Akhir Berdasarkan Data Abstrak Menggunakan Algoritma K- Nearest Neighbor. *Prosiding Seminar Nasional Multi Disiplin Ilmu & Call For Papers*, 103–111.