

Classification of Teachers and Lecturers Engagement On Webinar During The Pandemic Using The Utilization of Big Data

Ida Afriliana^{1*}, Nurohim^{1*}

¹ Computer Engineering, Politeknik Harapan Bersama Indonesia.

* Corresponding author:

Email: idaafriharahap@gmail.com

Abstract.

The Pandemic had a big impact on education in Indonesia and also in the world. In early 2020, during this pandemic, face-to-face meetings have turned into virtual or online meetings for both the learning process and seminars or workshops. The rapid development of technology supports this change in the world of education, this can be seen from the number of online seminars conducted to improve the competence of lecturers or teachers. The development of this online seminar allows the circulation of information that is increasingly large, fast, and almost unlimited by time and space. This causes a large amount of information to be scattered in the virtual world in various fields. With this very fast information technology, trillions of bytes of data are created every day from various sources such as on social media, especially those related to applications that are often used in website-based seminar media. This is called unstructured big data. In this study, big data will be implemented to classify educators' engagement of online seminar participants during an early pandemic. The activity stages in big data management and data processing support are acquired, accessed, analytical, and applied. The method for this study is the Adaptive Neuro-Fuzzy Inference System (ANFIS) to classify the engagement of teachers and lecturers an online seminar. The results of the training error obtained from ANFIS are 0.273482 with the ANFIS structure 4-12-12-12-1 or 4 inputs, 12 hidden layers, and 1 output.

Keywords: *classification engagement; Fuzzy Logic; ANFIS*

1. INTRODUCTION

In early 2020, the Pandemic period had a huge impact on education in the world and especially in Indonesia. During this pandemic, face-to-face meetings are more difficult than virtual ones. And the face-to-face meetings have turned into virtual or online meetings for both the learning process and seminars or workshops. The rapid development of technology supports changes in the world of education, this can be seen from the number of online seminars or webinars that are held to improve the competence of teachers dan lecturers. The development of the internet in the technological era advanced allows the circulation of information more and more, fast and almost unlimited by space and time[1].

Likewise, the development of this webinar enables the circulation of information that is faster, more abundant, and almost unlimited by space and time in the virtual world. This webinar is a platform to respond to situations for someone who

<http://ijstm.inarah.co.id>

works online, on mobile, and at very tight working hours. And with webinars, one can have face-to-face meetings but the participants are in different locations, so that interactions can be done directly via pictures or video or chat, or text [2]. This causes a lot of information to be scattered in the virtual world in various fields. With this very fast information technology, trillions of bytes of data are created every day from various sources such as on social media, especially those related to applications that are often used in website-based seminar media. This ocean of data leads to Big data leading to big data terminology[1].

The term "big data" was used for the first time in 1997 by Michael Cox and David Ellsworth in a paper presented at the IEEE conference describing data visualization and the challenges it poses for computer systems [3]. Big data terminology has can handle a wide variety of data. With the existence of this big data, it can add to an excellent contribution in data management so that this data can be used to perform analysis in various fields, such as in technology, where big data is related to infrastructure and tools in Big data operations, such as computational and analytical techniques, as well as storage (storage)[2].

The opinion can be used as material for sentiment analysis to determine the assessment to deep land public transportation services city whether positive or negative, as well as what opinion factors often arise[3]. This research uses Naïve Bayes Algorithm because the Naïve Bayes algorithm is using a statistical approach in taking a decision and based on theorem Bayes that all the attributes contribute that equally important and mutually independent in certain classes[4].

The holding of many webinars during a pandemic has increased rapidly, and this webinar makes unstructured big data. This research will focus on big data obtained from webinars conducted by teachers and lecturers at several higher education institutions, institutions or by several resource persons who play an active role in webinars in the world of education. The webinars that will be used as data sources are webinars related to education which will be conducted from February 2020 to August 2020. The unstructured big data is all comments at the webinars. A collection of comments or chats on the recorded webinars will be classified according to the participation of the participants in the webinar.

Therefore, this research is entitled Classification of Teacher and Lecturer Engagement on Webinar Participation during the Pandemic with the Utilization of Big Data. With the background that has been stated above, that there are many webinars on the competence of teachers and lecturers during this pandemic, the formulation of problems in this research is as follows:

1. What is the big data structure generated from the webinars?
2. How is big data implemented to classify the engagement of webinar participants during a pandemic?

The purpose of this research on the classification of engagement or involvement of teachers and lecturers who take part in webinars to increase

competence during a pandemic. The method of classification of engagement or involvement of educators who take part in webinars to increase competence during a pandemic with ANFIS. The advantages during this pandemic include that all activities can be carried out remotely, be it a learning process or a webinar, so sometimes a person can do more than one activity at a time. For example, following a webinar while doing online teaching activities or even other activities. This results in participants only including their name or picture on the webinar without being involved in it. The involvement of the participants in this webinar can be seen from the comments given by the participants. These comments become data in the form of text which will later be classified as his involvement in the webinar that he participates in.

II. METHODS

The frame of mind proposed in this study is as shown in figure1. The dataset in this study is the comments contained in several webinars. The dataset is taken from some webinars at Ekoji Channel, Aptikom, and TV Andi. There are a lot of comments at 31 webinars where the dataset is in the text. This raw data is obtained from comments on webinars uploaded on several youtube channels.

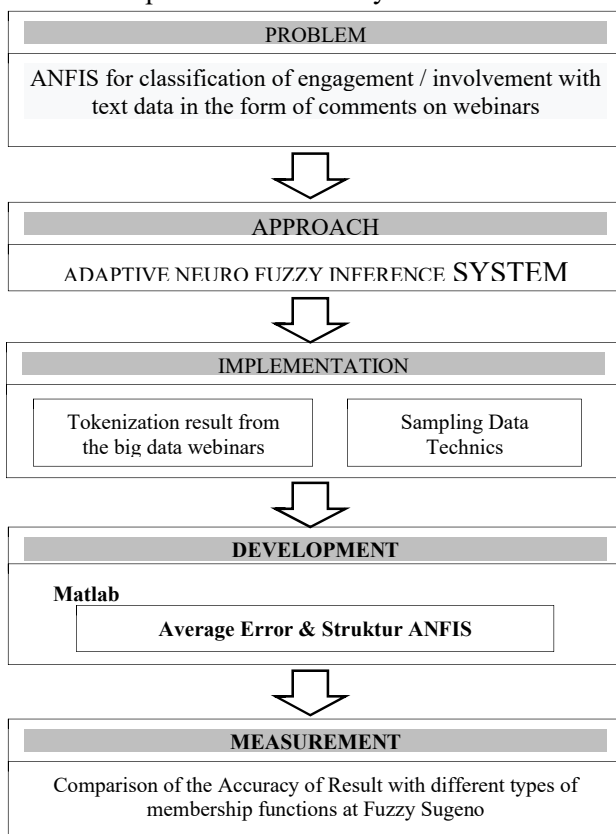


Fig 1. The frame of mind proposed

This text data is included in the big data category and can be taken as a source of data for this research. Then this text data is tokenized to can be classified using fuzzy Sugeno. After the dataset goes through the tokenized stage, then classification with Sugeno Fuzzy is carried out. Then proceed to the Adaptive Neuro-Fuzzy Inference System (ANFIS) stage to find out the accuracy of classification.

Experiment with fuzzy use the Matlab 2015a where the purpose of this study was to classify the extent to which the webinar participants were involved, seen from the comments on the webinar. The webinar taken in the sampling of this study were webinars for teachers and lecturers which were held from April to August 2020. The sampling technique used was the entire population of the data, then the dataset was divided into training and testing. The test measurement is seen from the comparison of the average error of different membership types. The result is this ANFIS can be used to process the classification or participant engagement or involvement in the webinars.

In general, this research consist of two stages, namely tokenization and data mining method with ANFIS. Sometimes some webinars that do not record digital traces for comments from participants, so at first stage, must get the webinar was carried out with a sufficiently adequate dataset for this research. And from figure 1 can we have seen that the stages in this study are: preprocessing dataset, generating fuzzy Sugeno; determination of train dataset and test dataset; experiments and testing; and the last stage is research evaluation and validation

A. Pre Processing

The second stage is generating the fuzzy Sugeno, then entering the ANFIS process. The initial stage before conducting data preprocessing. In this preprocessing stage, tokenization of the text dataset was carried out, then four variables were selected from each of the resulting big data.

B. Generating Fuzzy Sugeno

After preprocessing the data, then the data is ready to be generated using fuzzy Sugeno, because only fuzzy Sugeno can be processed by ANFIS. At this stage each set and its membership for each variable. Then made rules with if-clause. Then we can determine the ANFIS measurement value for each experiment.

C. Determination of Test Data and Test Data

The training and testing data in this study were the results of the test data from 30 webinars and then the tokenization was carried out from participant comments on the webinar. The training data and test data were carried out experimentally using ANFIS.

D. Experiment and Testing

After the fuzzy Sugeno has been generated, the next experimental stage is ANFIS. The fuzzification process that was generated by the first experiment, second experiment, third experiment, and fourth experiment then carried out the ANFIS process, with two different optimization methods, namely the hybrid optimization method and the backpropagation optimization method. Then the generated FIS is

also created with two different ANFIS structures. So after the training process or training, you will get different average error results. The results of the average error will be analyzed.

E. Research Evaluation and Validation

In this study, as an evaluation of the proposed model, namely the ANFIS method as well, with data testing, then through the ANFIS process for testing and checking. You will get a graph showing the difference between training and testing points.

III. RESULT AND DISCUSSION

The text data from the big data database is tokenized and then the data preprocessed is carried out. Among them are eliminating blank data and determining variables that will be used as classification indicators. In this study, there are four variables in the classification of engagement or involvement of webinar participants, namely in table 1.

Table 1. Table of research variables on the classification of webinar participant engagement

	Variabels
X1	Number of Subscribers
X2	the number of watched
X3	likes
X4	Engagement or those who are interested
X5	output - participant interest

The number of webinars conducted for this research sampling was 31 webinars. Where the dataset is arranged in tables X₁, X₂, X₃, and X₄, as shown in table 1, namely X₁ is the number of subscriptions of webinar participant or non-participants, X₂ is the number of webinar participants who watched, X₃ is the number of webinar participant who like or like, and X₄ is a webinar participants who is involved in the webinar. Participants involved can be seen from the comments by providing responses from the interviewees' questions and answers, asking questions and also giving positive responses about the material in the webinar. Then attendance comments and greetings are not included in the webinar. Lots of only comments provide emoticons too, and this includes the sentiment analysis, which was carried out in this study. So the initial dataset is text data and emoticon data that will be tokenized.

In table 2, you can see the dataset for this study, and all data has been prepositioned so that all datasets can be done at a later stage. The dataset is the comments and the comment was tokenized, the text which appropriates or shown engagements to the webinars. The greeting's comments were not tokenized, because only the positive comments can be tokenizing for this study. The dataset is not just

comments, the emoticons shown positive respects to engagement can be tokenized too. This study was tokenized carried out by reading all comments and classify them.

Table 2. engagement classification webinar participant dataset

No	x1	x2	x3	x4
1	2,400,000	3,386	218	24
2	2,400,000	7,829	587	453
3	2,400,000	5,288	384	365
4	2,400,000	6,399	366	38
5	2,400,000	4,134	206	154
6	1,170	364	13	11
7	362	461	34	4
8	36	501	28	21
9	796	3,546	97	84
10	16,400	2,429	230	51
11	16,400	960	118	8
12	16,400	1,876	174	4
13	16,400	1,112	91	1
14	16,400	1,106	105	12
15	16,400	1,029	122	40
16	16,400	1,709	73	25
17	16,400	512	44	1
18	16,400	2,847	165	385
19	16,400	3,694	207	674
20	4,460	335	24	12
21	4,460	4,524	225	345
22	4,460	1,522	102	37
23	4,460	339	32	4
24	4,460	270	25	9
25	4,460	734	87	10
26	4,460	1,063	75	41
27	4,460	443	37	1
28	4,460	670	32	17
29	4,460	608	65	13
30	4,460	230	22	2
31	4,460	341	23	7

The next stage is the implementation of fuzzy logic for the dataset. This stage is to form the set and set members of each variable. Set and set members as shown in Table 3.

Sugeno Fuzzy design is as follows:

1. selection of variables from a tokenized dataset.
2. create a set and set members of each variable.
3. creating a typical variables curve by doing several scenarios.
4. the scenario classification engagement of the webinar participant study was carried out in 8 experiments.
5. experiment 1 to experiment 4, the membership curve used is a triangular curve by distinguishing the number of rules from 4 rules to 16 rules. Then the analysis of the variable's output results is carried out.
6. for the 5th to the 8th experiment, the membership curve used is the Gauss curve by distinguishing the number of rules from 4 rules to 16. Then an analysis of the output results is carried out.

Table 3. Generating Sugeno Fuzzy

	Variabels	Set fuzzy	set members of each variable
X1	Number of Subscribers	Little; Enough; a lot	0-100; 101-500; >=500
X2	the number of watched	Little; Enough; a lot	0-500; 500-1000; 1000-5000
X3	likes	Little; Enough; a lot	0-100; 100-300; 400-800
X4	Engagement or those who are interested	Little; Enough; a lot	0-10; 10-100; 100-500
X5	output - participant interest	Little; Enough; a lot	0-10; 10-100; 100-500

The numerical range created has been adjusted to the existing dataset. Where is the dataset of the highest and lowest values from the dataset in Table 1.

The next step is to do fuzzy logic with the Sugeno fuzzy method, by doing 8 experimental scenarios for the same training data, but with different membership curves, as in table 4 and table 5

Table 4. Scenarios 1 to 4 of Sugeno fuzzy experiments for classification of webinar participant engagement

Himpunan Semesta	Nama Variabel	Eksperimen			
		R10	R12	R14	R16
Sedikit 0-1000	Jumlah Subscribe	Trimf [-2000 0 2000]	Trimf [-2000 0 2000]	Trimf [-2000 0 2000]	Trimf [-2000 0 2000]
Sedang 1001-2000		Trimf [500 2500 4500]	Trimf [500 2500 4500]	Trimf [500 2500 4500]	Trimf [500 2500 4500]
Banyak 2001-5000		Trimf [3000 5000 7000]	Trimf [3000 5000 7000]	Trimf [3000 5000 7000]	Trimf [3000 5000 7000]
Sedikit 0-500	Jumlah yang nonton	Trimf [-2000 0 2000]	Trimf [-2000 0 2000]	Trimf [-2000 0 2000]	Trimf [-2000 0 2000]
Cukup 500-1000		Trimf [500 2500 4500]	Trimf [500 2500 4500]	Trimf [500 2500 4500]	Trimf [500 2500 4500]
Banyak 1000-5000		Trimf [3000 5000 7000]	Trimf [3000 5000 7000]	Trimf [3000 5000 7000]	Trimf [3000 5000 7000]
Sedikit 0-100	Jumlah like	Trimf [-320 0 320]	Trimf [-320 0 320]	Trimf [-320 0 320]	Trimf [-320 0 320]
Cukup 100-300		Trimf [80 400 7200]	Trimf [80 400 7200]	Trimf [80 400 7200]	Trimf [80 400 7200]
Banyak 400-800		Trimf [480 800 1120]	Trimf [480 800 1120]	Trimf [480 800 1120]	Trimf [480 800 1120]
Sedikit 0-10	Engagment	Trimf [-200 0 200]	Trimf [-200 0 200]	Trimf [-200 0 200]	Trimf [-200 0 200]
Cukup 10-100		Trimf [50 250 450]	Trimf [50 250 450]	Trimf [50 250 450]	Trimf [50 250 450]
Banyak 100-500		Trimf [300 500 700]	Trimf [300 500 700]	Trimf [300 500 700]	Trimf [300 500 700]
Baik 0-10	Output-keterlibatan peserta				
Cukup 10-100					
Kurang 100-500		0,5	0,5	0	1

At table 4 can be seen that scenarios 1 to 4, fuzzy set use triangle curves by changing the number of rules. Generating the different number of rules in every scenarios.

Table 5. Scenario 5 to 8 of Sugeno fuzzy experiment for classification of webinar participant engagement

Himpunan Semesta	Nama Variabel	Eksperimen			
		R10	R12	R14	R16
Sedikit 0-1000	Jumlah Subscribe	Gaussmf [849.5 3.469e-14]	Gaussmf [849.5 3.469e-14]	Gaussmf [849.5 3.469e-14]	Gaussmf [849.5 3.469e-14]
Sedang 1001-2000		Gaussmf [849.5 2500]	Gaussmf [849.5 2500]	Gaussmf [849.5 2500]	Gaussmf [849.5 2500]
Banyak 2001-5000		Gaussmf [849.5 5000]	Gaussmf [849.5 5000]	Gaussmf [849.5 5000]	Gaussmf [849.5 5000]
Sedikit 0-500	Jumlah yang nonton	Gaussmf [849.5 3.469e-14]	Gaussmf [849.5 3.469e-14]	Gaussmf [849.5 3.469e-14]	Gaussmf [849.5 3.469e-14]
Sedang 500-1000		Gaussmf [849.5 2500]	Gaussmf [849.5 2500]	Gaussmf [849.5 2500]	Gaussmf [849.5 2500]
Banyak 1000-5000		Gaussmf [849.5 5000]	Gaussmf [849.5 5000]	Gaussmf [849.5 5000]	Gaussmf [849.5 5000]
Sedikit 0-100	Jumlah Like	Gaussmf [135.9 5.551e-15]	Gaussmf [135.9 5.551e-15]	Gaussmf [135.9 5.551e-15]	Gaussmf [135.9 5.551e-15]
Sedang 100-300		Gaussmf [135.9 400]	Gaussmf [135.9 400]	Gaussmf [135.9 400]	Gaussmf [135.9 400]
Banyak 400-800		Gaussmf [135.9 800]	Gaussmf [135.9 800]	Gaussmf [135.9 800]	Gaussmf [135.9 800]
Sedikit 0-10	Engagment	Gaussmf [84.95 3.47e-15]	Gaussmf [84.95 3.47e-15]	Gaussmf [84.95 3.47e-15]	Gaussmf [84.95 3.47e-15]
Sedang 10-100		Gaussmf [84.95 250]	Gaussmf [84.95 250]	Gaussmf [84.95 250]	Gaussmf [84.95 250]
Banyak 100-500		Gaussmf [84.95 500]	Gaussmf [84.95 500]	Gaussmf [84.95 500]	Gaussmf [84.95 500]
Sedikit 0-10	Output-keterlibatan peserta				
Sedang 10-100					
Banyak 100-500		5.37e-05	1.51e-05	1	0.847

At table 5 can be seen that scenarios 5 to 8, fuzzy set use gauss curves by changing the number of rules. Generating a different number of rules for every scenario.

The fuzzy membership curve can be seen in Figure 1 to Figure 3.

Figures 1 to 4 show the gauss curve implemented in the engagement classification of webinar participant teachers and lecturers, where the curve used is gauss. Of the 4 experiments, an analysis of the output value was carried out, in the range of 0 to close to 1 and even 1. So this Gauss curve is more suitable to be implemented for the engagement classification of webinar participants. There are 12 rules used in the fuzzy set.

Therefore, the experimental results with this Gauss curve will be continued to the next stage, namely ANFIS.

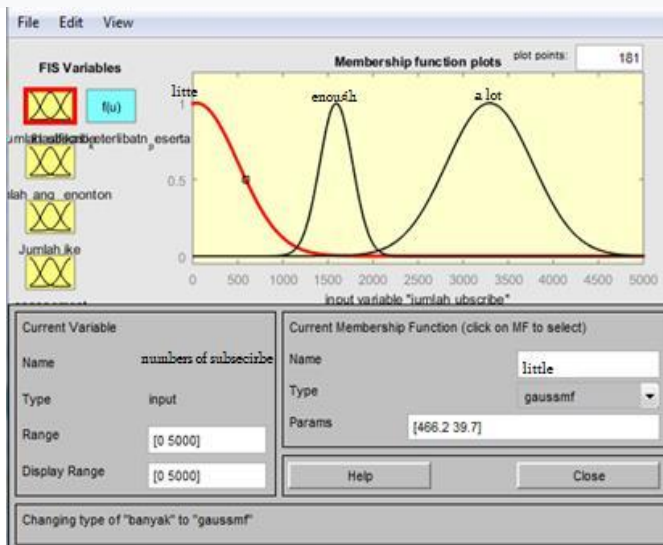


Fig. 2. A set of gauss Curve from number of subscribe

At figure 2 can be seen the set gauss curves from numbers of subscribers. The fuzzy set is little, enough, and a lot. After the fuzzy set has been done, then generating the rules, this study generates 12 rules.

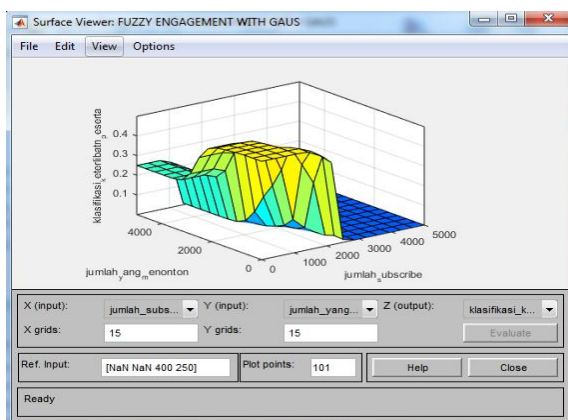


Fig. 3. A Plot of Sugeno fuzzy rules for engagement classification of webinar participant

From this experiment, the rules with fuzzy Sugeno set are obtained as shown in figure 3. Fuzzy set with the gauss curves and 12 rules and one output, namely are low engagement, quite an engagement, very engagement.

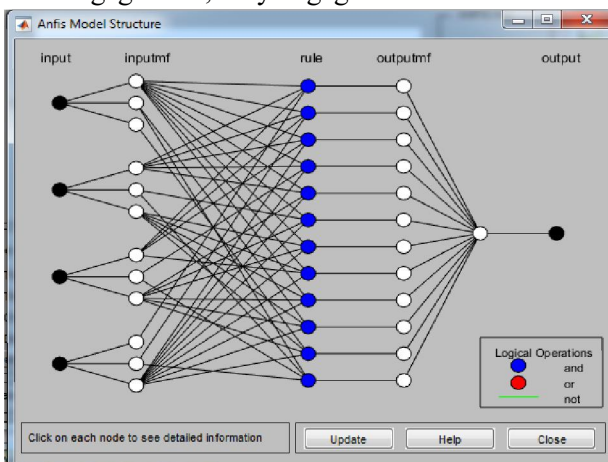


Fig. 4. ANFIS structure engagement classification of webinar participant

After this fuzzy Sugeno has been completed, then the dataset is implemented with the Adaptive Neuro-Fuzzy Inference System (ANFIS). The ANFIS design carried out in this study is a follow:

1. Conducting experiments o fuzzy Sugeno with output results up to 1. In this study, fuzzy with Gauss membership curves with 12 rules ware selected.
2. Conducting an experimental process by making a difference in the number of epoch, the optimization method and the allowable error tolerance value.
3. The training dataset is carried out like the second stage repeatedly until the error rate is the smallest.

4. The final step is testing the dataset with ANFIS.

The experiment for this dataset get the ANFIS structure engagement classification of webinar participant can be seen at figure 4, where structure engagement classification of webinar participant are 4-12-12-12-1 and error tolerance is 0.273.

IV. CONCLUSION

This research was conducted to engagement the classification of webinar participants in this pandemic. This research was conducted to see the extent to which webinar participants were involved in the themes presented in the webinar. This needs to be done because there are so many webinars that are conducted but sometimes there are participants who take part in the webinar but are not involved it. The engagement classification in this study is divide into 3, namely, participants are well involved in the webinar, participants are enough involved in the webinar, and participants are less involved in the webinar.

ANFIS can be classified this study with the result of the research being that optimization methods is backpropagation error tolerance is 0 (zero) ad the number of the epoch is 100. From the experiment, the training error 0.273. The research on engagement classification of webinar participants resulted in ANFIS sturcture, 4-12-12-12-1, which means there are four input layers, 12 hidden layers, 12-second hidden layers, 12-third hidden layers, and one output layer. This research can still develop tokenizing tools using other ones, doing text mining using the R program.

V. ACKNOWLEDGMENTS

The author is grateful to the Research and Community Service Institution Politeknik Harapan Bersama of Polytechnic, Deputy Academic Director, and the Chair of the DIII Computer Engineering Study Program, and all of its staff.

REFERENCES

- [1] D. Kusumasari and O. Rafizan, "Studi Implementasi Sistem Big Data Untuk Mendukung Kebijakan Komunikasi Dan Informatika," *Masy. Telemat. Dan Inf. J. Penelit. Teknol. Inf. dan Komun.*, vol. 8, no. 2, p. 81, 2018, doi: 10.17933/mti.v8i2.104.
- [2] E. R. E. Sirait, "Implementasi Teknologi Big Data Di Lembaga Pemerintahan Indonesia," *J. Penelit. Pos dan Inform.*, vol. 6, no. 2, p. 113, 2016, doi: 10.17933/jppi.2016.060201.
- [3] R. Y. Yanis, "Sentiment Analysis of Bpjs Kesehatan Services To Smk Eklesia and Bina Insani Jailolo Teachers," *J. Terap. Teknol. Inf.*, vol. 2, no. 2, pp. 25–34, 2018, doi: 10.21460/jutei.2018.22.105.
- [4] A. Hamzah, "Sentimen analysis untuk memanfaatkan saran kuisioner dalam evaluasi pembelajaran dengan menggunakan naïve bayes classifier", *Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST)*, 17-24, 2014
- [5] F. G. Rupilele, D. Manongga dan W. H. Utomo, "Sentiment Analysis Of National Exam Public Policy With Naive Bayes Classifier Method

- (NBC)", *Journal of Theoretical and Applied Information Technology*, Vol 58 No. 1, ISSN 1992-8645, 2013
- [6] H. Suyuti, "Webinar Sebagai Media Bimbingan Klasikal Sekolah Untuk Pendidikan Seksual Berbasis Online (Meta Analisis Pedagogi Online)," vol. 4, pp. 5–10, 2019.
- [7] D. S. Woelandari P.G, "Potential Benefits and Business Value of Big Data Analytics," *Maj. Ilm. Bijak*, vol. 15, no. 2, pp. 106–114, 2018, doi: 10.31334/bijak.v15i2.199.
- [8] I. Afriliana, "Prediksi Indeks Kualitas Air Sungai Pemali Comal Dengan Menggunakan Adaptive Neuro Fuzzy Inference System (Anfis)," pp. 58–62, 2016.
- [9] S. N. Asiyah and K. Fithriasari, "Klasifikasi Berita Online Menggunakan Metode Support Vector Machine Dan K-Nearest Neighbor," *J. Sains dan Seni ITS*, vol. 5, no. 2, pp. 317–322, 2016.
- [10] D. A. P. Wulandari, M. Sudarma, and N. Paramaita, "Pemanfaatan Big Data Media Sosial Dalam Menganalisa Kemenangan Pilkada," *Maj. Ilm. Teknol. Elektro*, vol. 18, no. 1, p. 101, 2019, doi: 10.24843/mite.2019.v18i01.p15.