JOIN (Jurnal Online Informatika)
Volume 6 No. 2 | December 2021: 225-232
DOI: 10.15575/join.v6i2.827
p-ISSN: 2528-1682
e-ISSN: 2527-9165

Chatbot for Signaling Quranic Verses Science Using Support Vector Machine Algorithm

Undang Syaripudin[1], Deden Suparman[2], Yana Aditia Gerhana[3], Ayu Puji Rahayu[4], Mimin Mintarsih[5], Rizka Alawiyah[6]
[1,3,6]Department of Informatics, UIN Sunan Gunung Djati Bandung, Indonesia
[2]Faculty Science and Technology, UIN Sunan Gunung Djati Bandung, Indonesia
[4]Islamic Education, STAI Darul Falah Bandung Barat
[5]Sharia Banking Ma'soem Universuty Bandung

Article Info

Article history:

Received November 09, 2021
Revised November 27, 2021
Accepted November 29, 2021
Published December 22, 2021

Keywords:

Chatbot
Quran
Science
Support vector machine (SVM)
Testing

ABSTRACT

The many verses in the Qur'an encourage finding the right way how to understand it thematically. The purpose of the research is to develop a chatbot application that can be used to explore and elaborate the content of verses in the Qur'an that hint at science. The support vector machine (SVM) algorithm classifies question and answers datasets in chatbot applications. The number of data sets used is 76, with test data as much as 10%. The test results show that the SVM algorithm is quite good in classifying, with an accuracy value of 87.5%. While the user test results obtained an average MOS of 8.4, which means the chatbot application developed is very effective in understanding the Qur'an, which implies science. This research is expected to provide an overview of the explanation of the Qur'an about science and technology.

Corresponding Author:

Yana Aditia Gerhana
Department of Informatics,
State Islamic University of Sunan Gunung Djati,
Jl. AH Nasution 105, Bandung, Indonesia
Email: yanagerhana@uinsgsd.ac.id

1. INTRODUCTION

Industrial revolution 4.0 brings changes in various fields of life, including education. Education 4.0, a reflection of the Industrial 4.0 era, refers to a period where paradigms, approaches, and educational technology have been changed [1]. Education 4.0 is closely related to the presence of integrated e-learning with several Artificial Intelligence (AI) applications, which provide personal learning spaces [2]–[7]. One of the learning profiles 4.0 is the learning process that takes place adaptively supported by AI, which regulates the learning process according to the profile of students in real-time [3], [8]–[10].

The chatbot is one of the AI technologies that has been used in various learning applications [11][12]. The chatbot is a computer program built to replace human expertise or human operators to communicate with users (humans), working alone without any supervision or assistance from humans[13]. Chatbot applications are currently used in thematic learning of the Qur'an or Hadith. The Qur'an is a holy book that explains many aspects of life, including those that hint at science and technology.

The development of the chatbot application was carried out by Muhammad et al. for media for learning to read the Qur'an with the Tsaqifa method. The results of testing by 20 respondents using the System Usability Scale method obtained a value of 73.3, which means that the chatbot application can be accepted by users [14]. Yusuf Nur et al. developed a chatbot on the instant messaging application ChatAja. This chatbot was created to make it easier for users to perform worship in the month of Ramadan. There are prominent features of the adhan, one day one father, Ramadan quiz, and reminder. The user can well receive the chatbot, evidenced by the average value of the respondent's test results, 84.4% [15]. While Kurniawan et al. also developed a chatbot about learning Islamic law using the naive Bayes algorithm and Bayesian network. The result of dataset classification, nave Bayes is more accurate than the Bayesian network [13].

Smutny and Schreiberova reviewed the Facebook Messenger educational chatbot to support learning. Of the 47 educational chatbots using the Facebook Messenger platform, it shows that chatbots that are part of

p-ISSN: 2528-1682
e-ISSN: 2527-9165

the instant messaging application are still in the early stages of becoming artificial intelligence teaching assistants [16]. A chatbot for educational needs in universities was developed by Nguyen et al. under the name NEU-Chabot. Students can instantly get daily curriculum updates, new admissions, tuition fees, IELTS writing assignment scores, etc. NEU-chatbot was developed using the Deep Learning model, where the test results obtained an accuracy value of 97.1%[17].

The Support Vector Machine (SVM) algorithm has been widely used in chatbot applications. SVM is used in the irrigation decision method for the winter wheat growing period in additional irrigated areas. The results showed that decision-making accuracy in the wintering, regreening, and jointing periods was 89.4%, 95.7%, and 93.6%. These results indicate that the proposed method can effectively make optimal decisions for water-saving irrigation in the winter wheat season in additional irrigation areas [18]. SVM is also used in the predictive model of the rate of biodegradation with multiple linear regression. The results of the comparison of several models, the results of the SVM model have a satisfactory fit, robustness, and external predictability [19].

Mean opinion score (MOS) is used in subjective assessment of user experience in using chatbot applications. MOS in previous studies was used in speech recognition assessment and management quality assessment [20], [21]. This study will provide additional knowledge about the development of chatbots using SVM and an explanation of the Koran that hints at science.

2. METHOD

This research begins with a literature study, namely the collection of written data obtained from literature studies. The book used in this study is used as a dataset on the chatbot. The book used is "Ayat-Science: Secrets of Real and Unseen Greatness in the Qur'an" by Nurrohaman. In addition to the book, the data was obtained from expert sources on the Qur'an from the State Islamic University of Sunan Gunung Djati, Bandung. Support vector machine (SVM) is used to classify datasets, either questions or user answers, on chatbots.

The architecture of the chatbot application system is generally shown in Figure 1. if the user enters a question, the chatbot will process the question. It will go through the text preprocessing stages, namely cleaning

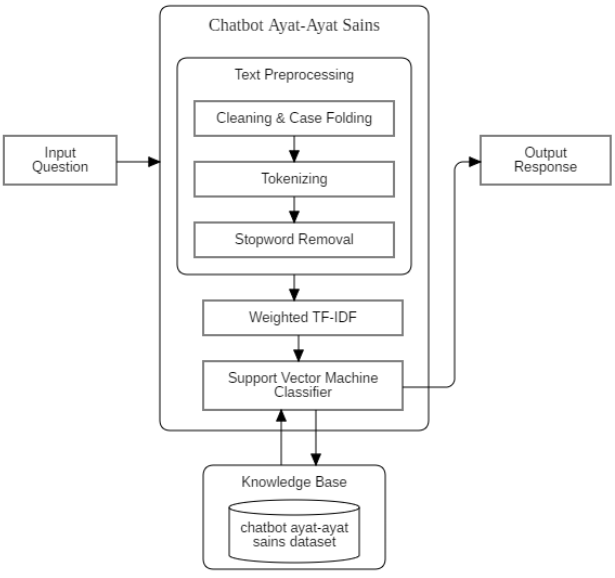

Figure 1. System Architecture

2.1. The Qur'an and Science

The philosophical basis for developing science or knowledge can be studied and extracted from the Qur'an. Like other holy books, the Qur'an has miracles. Among its miracles, the Qur'an explains a lot about verses that hint at science. These verses contain scientific information that can be proven true from time to time by scientific developments. Even the Qur'an invites all humans to study and prove the verses of science that are in the Qur'an. This needs to be done as proof that the Qur'an is a revelation from Allah SWT, which is called 'Ijaz al-ilmi' [22].

2.2. Chatbot

The chatbot is a computer program designed to simulate interactive communication or conversation between users (humans) through text, audio, or video. The response obtained is the result of keywords in the input made by the user and produces a response that is considered the most appropriate so that the communication that occurs seems to be carried out between two humans who communicate with each other [23]. There are two chatbot models, namely generative-based-model and retrieval-based-model.

Generative-based models are chatbot models that can carry out human-style conversations. Bots like this need millions of examples to train to get conversational quality like human style. One of the bots that implement this model is Siri. Retrieval-based models are another chatbot model that is fairly easy to build and also gives more predictable results. API is needed to build this chatbot model [24]. In this study, the chatbot that will be built is a chatbot with a Retrieval model.

2.3. Classification

Classification is a process of examining data objects to group them into certain classes from several available classes. In classification, two main tasks must be done: building a model as a prototype that will be stored as memory. The second task is to use the model so that recognition/classification/prediction can be made on another data object to know which class the data object is in. in the stored model [25]. This classification method has the primary purpose, namely, to assist in understanding the grouping of data.

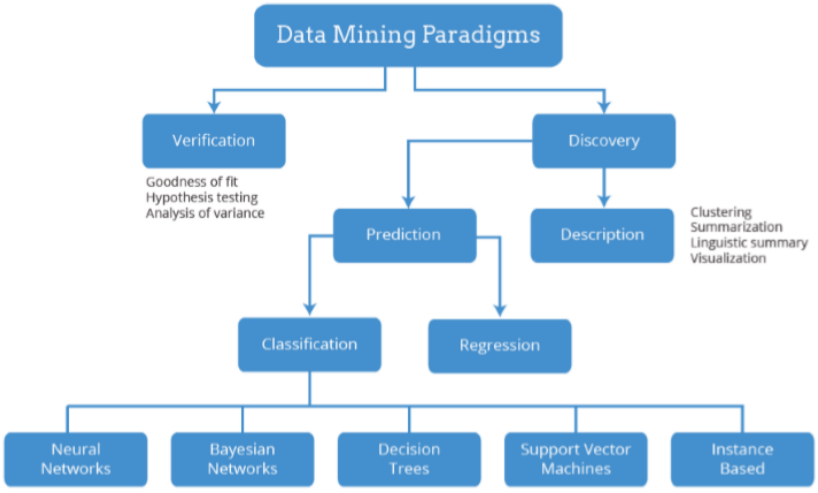

Figure 2. Data Mining Classification

Figure 2 shows that data mining consists of two types: verification-oriented data mining, which is oriented to verify user data, and discovery-oriented data mining, which helps find new rules or independent patterns from a dataset. Data mining discovery has two methods, namely prediction, and description. The description method is oriented to data interpretation which focuses on understanding with visualization. At the same time, the prediction-oriented method aims to automatically build a behavioral model that obtains a new and invisible sample and can predict the values of the variables associated with the sample. In addition, predictions are developed to obtain patterns that form knowledge understandably and quickly to operate. Data mining prediction consists of classification and regression techniques. Algorithms used for classification include Neural Network, Naive Bayes Classification, Decision Trees, Support Vector Machine, and Instance-Based [25].

2.4. Support Vector Machine (SVM)

Support Vector Machine is a relatively new and powerful method in classification and pattern recognition [26]. Support Vector Machine (SVM) is a technique for making predictions, both in the case of classification and regression [27]. Initially, the SVM algorithm was created to assist in performing optimal binary (two classes) classification. However, later, this algorithm was extended to regression and clustering problems [28]. In geometric terms, the SVM binary classifier can be viewed as a hyperplane in the feature space that separates points representing positive instances from points representing negative occurrences, as in Figure 3.

p-ISSN: 2528-1682
e-ISSN: 2527-9165

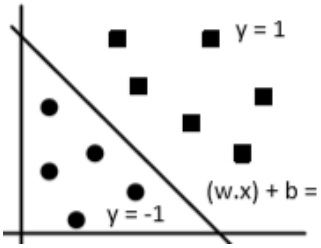

Figure 3. Linear dividing plane

The following equation can determine the class of data to be predicted or test data.

with:
x = data to be predicted class (data testing)
xi = data support vector, i=1, 2, ...l
l = many data support vector

The algorithm is designed for binary classification, and when dealing with k class, it can be changed to k(k-1)/2 binary classification.

SVM was developed for two-class classification problems, then redeveloped for multi-class classification. In multi-class case classification, more than one hyperplane is formed. One of the approach methods used is One vs. All, where each i-th classification model is trained with all of the data to find a solution. Because the data in the class cannot be separated linearly (non-linearly), a Kernel Trick will be used on the initial data features of the dataset using the RBF (Radial Basis Function/ Gaussian) kernel, which is defined by $K(x, y)=\exp(-\gamma| |x - y||2), >0$.

2.5. Data Augmentation

Experiments from research can be evaluated by testing the Confusion Matrix. The confusion matrix is a method that uses a matrix table as shown in the table below; if the dataset only consists of two classes, one class is considered positive and the other negative [29].

Table 1. Confusion Matrix Model

		True Class	
		Positive	Negative
Predicted Class	Positive	True positive count (TP)	False negatives count (FP)
	Negative	False positives count (FN)	True negatives count (TN)

After the test data is entered into the confusion matrix, the accuracy value can be calculated. Calculate the accuracy value using the following equation.

$$\text{accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FN+FP}} \quad (2)$$

3. RESULTS AND DISCUSSION

The developed chatbot application can be used in thematic learning of Al-Qur'an verses that hint at science

3.1. Display of Application

Figures 4 and 5 display the thematic learning chatbot application of Al-Qur'an verses that hint at science.

Figure 4. Main Interface

Figure 5. Conversation Interface

Figure 4 shows the initial view when the user first opens the chatbot. A chat bubble will appear containing greetings and explanations about the chatbot. Figure 5 is an interface that shows the conversation between the user and the chatbot. The left side of the chat bubble is the chatbot, and the right is the user.

3.2 Testing

In addition to testing the accuracy value, this study also tested the level of user satisfaction with the developed application using the Mean Opinion Score (MOS) framework.

3.2.1 Accuracy Value

The level of chatbot curation in answering user questions can be known through testing the confusion matrix method. The accuracy value is known using the split percentage of the dataset test scenario and divides it into two parts, namely training data and test data. We randomly divided 76 datasets (questions and answers on science verses) into 90% training data and 10% test data in this test. The following is a confusion matrix of the tests carried out.

Table 2. Confusion Matrix Testing

		Label of Prediction		
		Physics	**Biology**	**Math**
Previous of Label	**Physics**	4	1	0
	Biology	0	1	0
	Math	0	0	2

p-ISSN: 2528-1682
e-ISSN: 2527-9165

The table above explains that the value in the row in each label is the amount of test data used by the label, while the value in the label column is the predicted value from the test data. As in the first line, the physics label shows that the test data used are five, but only 4 data are correctly predicted, and one other data is predicted on the biology label. In the second line, the biological label uses 1 test data, and all of them are predicted to be correct on the biological label. In addition, the mathematical label uses 2 test data, and both are correctly predicted on the mathematical label. After getting the confusion matrix, the accuracy value can be known by calculating using equation (2). The details of the calculation are as follows

$$\text{accuracy} = \frac{4+1+2}{8} = 0.875 = 87.5\%$$

From the calculation results above, it can be seen that the accuracy value of the test by dividing the dataset into 90% training data and 10% test data produces an accuracy value of 87.5%.

3.2.2 MOS (Mean Opinion Score) Testing

In addition to accuracy, this study also assessed user satisfaction with the use of the chatbot application. The assessment framework uses the Mean Opinion Score (MOS), using a questionnaire as an assessment instrument. Respondents' assessment was on a scale of 1-10 with a statement of 10-8 being a very effective score, 7-6 being effective, 4-2 being quite adequate, while 2-0 being ineffective, and the number of questions being 18 samples. Respondents who filled out the assessment were 20 students of informatics department State Islamic University of Sunan Gunung Djati Bandung. Testing is carried out repeatedly to obtain consistency of judgment from respondents. The assessment has four aspects: clarity of questions, clarity of answers, speed of answers, and color and layout. The assessment results show an average question clarity value of 8.5, an average clarity of 8.1, an average answer speed of 8.1, and an average and layout of 7.5. The results explain that, in general, the application components according to the respondents are very effective, except for the color and layout. Figure 6 describes the results of the average respondent's assessment.

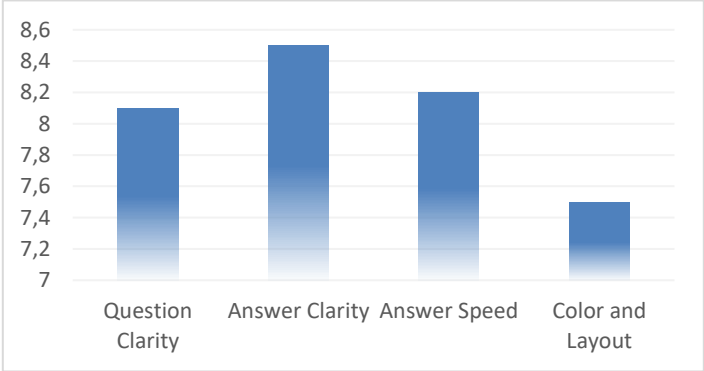

Figure 6. Diagram of Average Respondent Satisfaction Assessment

The results of the first test obtained a value of 8.3. The second test improved the color and application layout aspects and obtained a MOS value of 8.4. The third test was carried out by improving the color and layout and the clarity of the answers, and the MOS score was 8.55. The overall average value of the MOS test is 8.42. Figure 7 describes the increase and average respondents' satisfaction with using the Quranic verse chatbot application that hints at science. The overall user satisfaction assessment results show that they are delighted with the chatbot application developed.

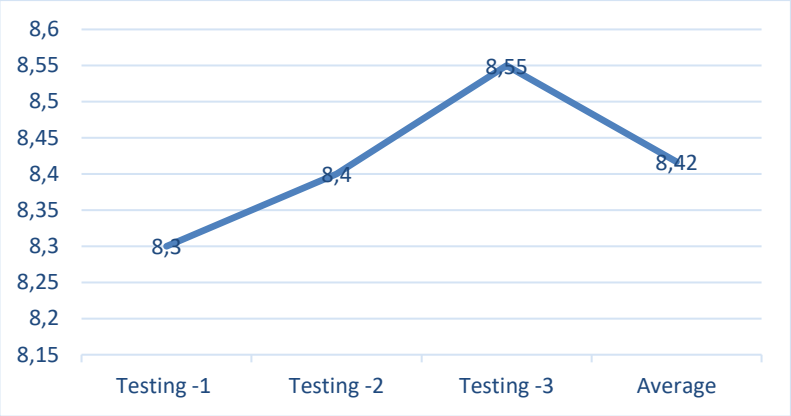

Figure 7. Mean Opinion Score (MOS) Test Results

4. CONCLUSION

Implementing the support vector machine (SVM) algorithm on the chatbot application about the verses of the Koran that hints at science have succeeded in classifying questions and answers and predicting the labels of questions entered by users. The performance of the support vector machine algorithm on the chatbot dataset classification model of Al-Quran verses that hints at science is quite good, as evidenced by the high accuracy value of 87.5%. While testing from the user side, using MOS, the results obtained that the chatbot application developed was very effective in understanding the Qur'an, which implies science. Future research can use the Naïve Bayes algorithm or K-Nearest Neighbors with more data sets. As for testing, you can use techniques other than MOS. Tests using a statistical approach or big data analysis can be used for data with a large enough amount.

ACKNOWLEDGEMENTS

This research was supported by the Research Institutions and Community Service and Department of Informatics State Islamic University of Sunan Gunung Djati who gave the opportunity for us to do this research.

5. REFERENCES

[1] H. Keser and A. Semerci, "Technology trends, Education 4.0 and beyond," *Contemp. Educ. Res. J.*, vol. 9, no. 3, pp. 39–49, 2019, doi: 10.18844/cerj.v9i3.4269.
[2] C. C. Chea, J. Tan, and J. Huan, "Higher Education 4.0: The Possibilities and Challenges," *J. Soc. Sci. Humanit.*, vol. 5, no. 2, pp. 81–85, 2019.
[3] C. Damartini and B. Lorenzo, "Do Web 4.0 and Industry 4.0 Imply Education X.0?," *IT Prof.*, vol. 19, no. 3, pp. 4–7, 2017, doi: 10.1109/MITP.2017.47.
[4] M. Rafiq, A. Bashar, and A. Shaikh, "Innovative Trends in Knowledge Management: A Cloud Computing Perspective," *Proc. First Middle East Conf. Glob. Business, Econ. Financ. Bank.*, no. October, pp. 1–12, 2014.
[5] A. A. Shahroom and N. Hussin, "Industrial Revolution 4.0 and Education," *Int. J. Acad. Res. Bus. Soc. Sci.*, vol. 8, no. 9, pp. 314–319, 2018, doi: 10.6007/ijarbss/v8-i9/4593.
[6] G. A. Makrides, "The Evolution of Education from Education 1.0 to Education 4.0: Is it an evolution or a revolution?," no. March, 2019.
[7] R. Cerchione and E. Esposito, "Using knowledge management systems: A taxonomy of SME strategies," *Int. J. Inf. Manage.*, vol. 37, no. 1, pp. 1551–1562, 2017, doi: 10.1016/j.ijinfomgt.2016.10.007.
[8] R. Cerchione and E. Esposito, "International Journal of Information Management Using knowledge management systems : A taxonomy of SME strategies," *Int. J. Inf. Manage.*, vol. 37, no. 1, pp. 1551–1562, 2017, doi: 10.1016/j.ijinfomgt.2016.10.007.
[9] T. Kabudi, I. Pappas, and D. H. Olsen, "AI-enabled adaptive learning systems: A systematic mapping of the literature," *Comput. Educ. Artif. Intell.*, vol. 2, no. December 2020, p. 100017, 2021, doi: 10.1016/j.caeai.2021.100017.
[10] Y. Atif and L. Zhang, "Network resource management in support of QoS in ubiquitous learning," *J. Netw. Comput. Appl.*, vol. 41, no. 1, pp. 148–156, 2014, doi: 10.1016/j.jnca.2013.11.003.
[11] K. Zhang and A. B. Aslan, "AI technologies for education: Recent research & future directions," *Comput. Educ. Artif. Intell.*, vol. 2, p. 100025, 2021, doi: 10.1016/j.caeai.2021.100025.
[12] C. W. Okonkwo and A. Ade-Ibijola, "Chatbots applications in education: A systematic review," *Comput. Educ. Artif. Intell.*, vol. 2, p. 100033, 2021, doi: 10.1016/j.caeai.2021.100033.
[13] K. Jamal, R. Kurniawan, A. S. Batubara, M Z A Nazri, F. Lestari, and P. Papilo, "Text Classification on Islamic Jurisprudence using Machine Learning Techniques," *ICCAI*, 2019.
[14] F. M. Aulia, M. Sudarma, and I. M. A. Suyadnya, "PEMANFAATAN INSTANT MESSAGING UNTUK

APLIKASI PEMBELAJARAN MEMBACA AL-QUR'AN DENGAN METODE TSAQIFA," *SINTECH J.*, vol. 1, no. 2, 2018.
[15] N. Y. N. Pratama and F. Y. Al Irsyadi, "Perancangan Chatbot Islami untuk Aplikasi ChatAja," *Emit. J. Tek. Elektro*, vol. 21, no. 1, 2021.
[16] P. Smutny and P. Schreiberova, "Chatbots for learning: A review of educational chatbots for the Facebook Messenger," *Comput. Educ.*, vol. 151, p. 11, 2020.
[17] T. Thanh, N. Anh, D. LebHa, T. Hoangc, and T. Nguyend, "NEU-chatbot: Chatbot for admission of National Economics University," *Comput. Educ. Artif. Intell.*, vol. 2, 2021.
[18] Hongzheng Shen; Kongtao Jiang; Weiqian Sun; Yue Xu; Xiaoyi Ma., "Irrigation decision method for winter wheat growth period in a supplementary irrigation area based on a support vector machine algorithm," *Comput. Electron. Agric.*, vol. 182, March, 2021.
[19] Weihao Tanga; Yanying Lia; Yang Yub; ZhongyuWanga; Tong Xua; Jingwen Chena; Jun Linb; XuehuaLia;, "Development of models predicting biodegradation rate rating with multiple linear regression and support vector machine algorithms," *Chemosphere*, vol. 156, Augus, no. 1–7, 2020.
[20] J. S. M. B. R. D. R. E. C. Sreerama., "Impact of Accuracy and Latency on Mean Opinion Scores for Speech Recognition Solutions," in *Procedia Manufacturing*, 2015, pp. 1–7.
[21] Jie Xu; Liyuan Xing; Andrew Perkis; Yuming Jiang, "On the Properties of Mean Opinion Scores for Quality of Experience Management," 2011.
[22] Fatimah, "Ayat-ayat Sains dalam al-Quran (Tela'ah Balaghah)," *J. al–Hikmah*, vol. vol 5, no. 2, 2017.
[23] D. Christianto and Dkk, "Penggunaan Named Entity Recognation dan Artificial Intelligence Markup Language untuk Penerapan Chatbot Berbasis Teks," *J. Telemat. ITHB*, vol. 10, no. 2.
[24] K. Jwala, G. N. V. . Sirisha, and G. V. P. Raju, "Developing a Chatbot using Machine Learning," *Int. J. Recent Technol. Eng.*, vol. 8 ISSN: 22, no. 3, 2019.
[25] S. Defiyanti, "Integrasi Metode Clustering dan Klasifikasi untuk Data Numerik," *CITEE*, no. ISSN: 2085-6350, 2017.
[26] S. Bashiri and Dkk, "USING PCA COMBINED SVM IN THE CLASSIFICATION OF EUTROPHICATION IN DEZ RESERVOIR (IRAN)," *Environ. Eng. Manag. J.*, vol. 16, no. 9, 2017.
[27] P. A. Octaviani, Y. Wilandari, and D. Ispriyanti, "PENERAPAN METODE KLASIFIKASI SUPPORT VECTOR MACHINE (SVM) PADA DATA AKREDITASI SEKOLAH DASAR (SD) DI KABUPATEN MAGELANG," *J. GAUSSIAN*, vol. 3, no. 4, 2014.
[28] S. Widaningsih and A. Suheri, "KLASIFIKASI JURNAL ILMU KOMPUTER BERDASARKAN PEMBAGIAN WEB OF SCIENCE DENGAN MENGGUNAKAN TEXT MINING," *Semin. Nas. Teknol. Inf. dan Komun. 2018 (SENTIKA 2018)*, vol. ISSN: 2089, 2018.
[29] M. Y. H. Setyawan, R. M. Awangga, and S. R. Efendi, "Comparison Of Multinomial Naive Bayes Algorithm And Logistic Regression For Intent Classification In Chatbot," *IEEE Syst. J.*, 2018.