

COVID-19 Vaccination and PPKM Policy with the Implementation of the Fuzzy Sugeno Method to Income Classification

Djihani Wahyuni^{1*}, Eni Sumarminingsih², Suci Astutik³

^{1,2,3}Departement of Statistics, Universitas Brawijaya, 65145, Malang, Indonesia

djihandije@gmail.com¹, eni_stat@ub.ac.id², suci_sp@ub.ac.id³

ABSTRACT

Article History:

Received : 24-07-2022

Revised : 08-09-2022

Accepted : 12-09-2022

Online : 08-10-2022

Keywords:

Sentiment Analysis;

FIS Sugeno;

PPKM;

Covid-19 Vaccination.

This study aims to determine the implementation of Fuzzy Sugeno in classifying textual data obtained from Twitter so as to determine the polarity of public opinion regarding PPKM policies and Covid-19 vaccinations. This study uses primary data via Twitter related to COVID-19 vaccination and PPKM policies in Indonesia starting from February 9, 2021 to January 17, 2022. There are several stages carried out, namely data collection, data pre-processing, data labeling, data weighting, identification of membership functions, determination of fuzzy sets, formation of classification systems, and evaluation of classification results. The results of this study explain that Fuzzy Sugeno's performance in classifying tweets is quite good with an average accuracy of 89.13%. Meanwhile, public opinion regarding PPKM policies and Covid-19 vaccinations tends to be balanced with 36.92% of tweets classified as positive sentiments, 22.85% negative sentiments, and another 40.23% classified as neutral sentiments. In addition, the fuzzy set that is formed based on the data observation method is very well done because it is able to adjust the frequency of the data in each category. This really helps improve the performance of the built classification system.



<https://doi.org/10.31764/jtam.v6i4.10096>



This is an open access article under the **CC-BY-SA** license

A. INTRODUCTION

Various policies have been implemented to mitigate the impact of the Covid-19 pandemic. Community participation is one of the keys to successfully defeating the Covid-19 pandemic. However, the guidelines provided by the government have many advantages and disadvantages. Given the importance of PPKM and Covid-19 vaccine guidelines, it is necessary to monitor public opinion to propose solutions or evaluate future guidelines (Djalante et al., 2020).

Currently, many people use social media to express their opinions, among other things on Twitter. Twitter is a medium for sharing messages and interacting globally, therefore Twitter is an abundant source of data (Twitter, 2021). Data obtained from Twitter can be in the form of tweets from certain topics which if further processed will provide information about emotions and can be grouped into positive, negative, and neutral opinions (Paroubek & Pak, 2010).

There are previous studies related to sentiment analysis. Villavicencio, et al apply naïve bayes for monitoring sentiment Covid-19 vaccines on twitter. This study identified positive, negative, and neutral sentiments and obtained a model accuracy of 81.77% (Villavicencio, et al. 2021). Efrilianda, et al conducted a sentiment analysis related to the Covid-19 pandemic

through hashtag on twitter with 212 tweets of which 48% were in the positive category, 30% were negative, and 22% were neutral tweets (Efrilianda, et al., 2021). Setiawan, et al applied Support Vector Machine and Naïve Bayes in sentiment analysis related to post-Covid-19 online lectures, the accuracy of the SVM algorithm was greater than Naive Bayes, that is 85% per 81.20% (Setiawan, et al., 2021).

In fact, sentiments cannot be categorized precisely, this is what makes the fuzzy variables input that have an element of uncertainty, specifically the tendency of sentiment levels in each tweet (Fu, & Wang, 2010). The advantages of fuzzy in defining opinion data will be combined with the selection of membership functions with data observation methods (Aryandani, et al., 2022). This combination will result in a flexible textual data classification model, even for large amounts of data. This model is properly used in monitoring public policy-related opinions because the policy time period is generally quite long and has many pros and cons, resulting in abundant data. It also needs to be offset by capable data sources, social networks such as Twitter.

In this research, a fuzzy inference system is used in classifying tweets. This algorithm was selected because of more adaptive and efficient (Kaur, 2012). The Sugeno method provides output (consequent) in the form of a constant or linear function which is easily understood as compared to the Mamdani method that produces output of membership function (MathWorks, 2014). According to Iwendi et al. (2022) explaining the application of individual COVID-19 classification using adaptive neuro-fuzzy inference. Meanwhile, this research implements fuzzy Sugeno on the impact of PPKM and Covid-19. According to Civelek (2020) implementing fuzzy Sugeno optimizers on wind turbines. While Rauf et al. (2020) used fuzzy sugone as an adaptive inertial weighting bat algorithm. Based on previous research, this study raises the opinion of the Indonesian people regarding PPKM policies and Covid-19 vaccinations. The performance of the resulting model can be a consideration for policies to overcome the Covid-19 pandemic in the future.

B. METHODS

The concept of mindset in this research is using fuzzy logic method, Fuzzy Inference System (FIS) Sugeno, sentiment analysis, and text analysis. The method is formed in one unit to get the results of grouping the PPKM and Covid-19 Vaccine policies. Fuzzy logic is a branch of artificial intelligence systems that package human thinking abilities into algorithms which are then run by machines. Thakkar et al. (2021) explain that this logic is generally applied to problems that contain elements of uncertainty and imprecision. Fuzzy is a method of counting by using variables in the form of words can be called linguistic variables instead of variables in the form of numbers. It is stated that the number variable is more precise than the word variable, but words will be closer to human intuition (Serrano-Guerrero et al., 2021).

The Sugeno or Takagi-Sugeno-Kang method was introduced by Prof. Tomohiro Takagi and Prof. Michio Sugeno in 1985 in an effort to build a systematic approach with rules represented in the form of IF - THEN which is more adaptive optimization and more efficient terms of computation (Hu, et al., 2021). According to the KBBI, sentiment means an opinion or view that is based on feelings towards something. understanding based on Liu (2012) regarding sentiment analysis is part of the field of data mining. The purpose of sentiment analysis is to

analyze a person's attitudes, emotions, opinions, and evaluations of a particular product, service or activity. Sentiment itself can be interpreted as a response or an opinion. according to Xia et al. (2021) Sentiment analysis is important to be used as evaluation material for certain groups. one example in the business world, sentiment analysis is useful for determining consumer satisfaction with marketed goods and services, so that it can be taken into consideration in improving the quality of production (Birjali et al., 2021).

According to Han et al. (2011), text mining is a technology that is able to analyze semi-structured and unstructured text data. based on the explanation of Wang & Lo (2021) high dimensions, noise and poor text structure are one of the characteristics of text data in general. Text mining has an important role in processing text data automatically by computers. The initial idea of making text mining is to find information that can be extracted from a text (Thangaraj & Sivakami, 2018). The first step in applying the method in text mining is to do text pre-processing. Text pre-processing aims to improve the data structure so that it can be repaired properly by the computer (Kumar & Prakash, 2019).

This study used primary data obtained from Twitter. The data used is the content of tweets related to PPKM and Covid-19 vaccination in Indonesia during the period of Inmendagri No. 3 of 2021 until the period of Inmendagri No. 2 of 2022, starting on February 9, 2021 until January 17, 2022 with provisions in Indonesian, uploaded in the region Indonesia, and use the keywords "PPKM", "vaccine COVID", and "vaccine" in the search column. The analysis steps in this study are as shown in Figure 1.

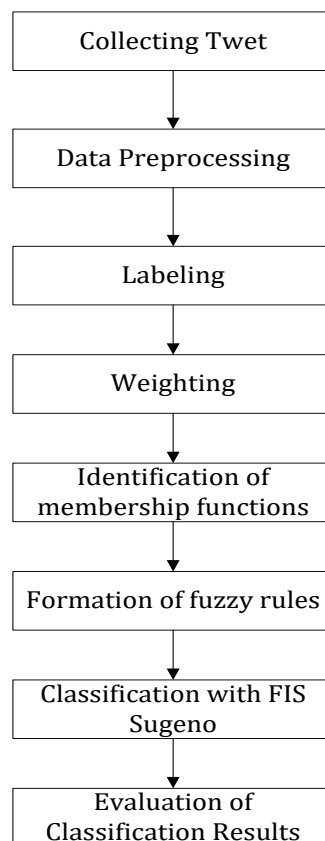


Figure 1. The analysis steps in this study

1. Collecting Twet: In this process, namely collecting tweets discussing PPKM policies and the covid-19 vaccine;
2. Data Pre-processing: Data preprocessing is an early stage in processing text data. The purpose of this process is to make text data more structured. The stages in this process are case folding, tokenizing, stemming, spelling normalization, and removing stopwords (Wang, et al., 2021);
3. Labeling: the process of labeling data;
4. Weighting: The weighting method used is term frequency. This unsupervised feature selection method is used because it is known to be simple and has low computation time (Pedersen, 1997). Weighting using term frequency focuses on the number of terms that appear in each document (Torkayesh, et al, 2021; Şahin, 2021);
5. Identify membership function: know the membership process;
6. Formation of fuzzy rules: the process of forming fuzzy rules;
7. Classification with FIS Sugeno: The data classification process uses fuzzy Sugeno;
8. Evaluation of Classification Results: Evaluation of classification results commonly used is k-fold cross validation, this method used for evaluating the performance of a classification algorithm with a limited amount of data. The way k-fold cross validation is to calculate the average success, this begins by dividing the data into n-folds. The distribution of data will form n partitions with the same size D1, D2, ..., Dn, iteration of i in the Di partition will be the test data while the rest will be the training data. For the sake of validity testing, it is recommended to use the number of 10-fold cross validation in the model. Figure 2 shows iterations on 10-fold cross validation, as shown in Figure 2.

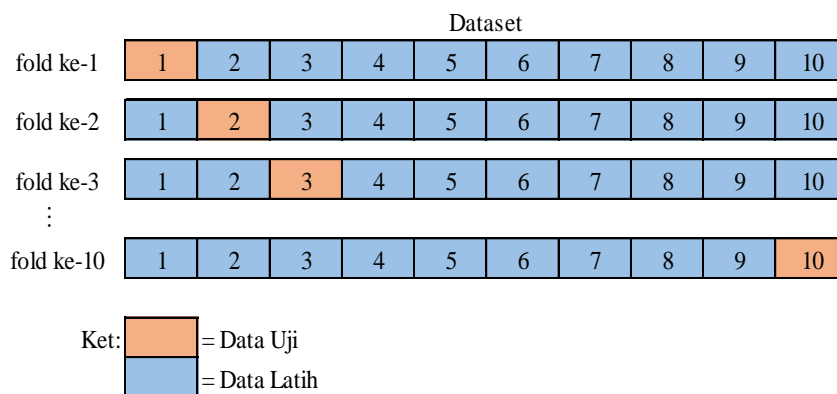


Figure 2. 10-Fold Cross Validation

In each iteration, accuracy will be calculated using a confusion matrix that describes the performance of a specific model or algorithm. This matrix contains predictions which will then be compared with the actual classification results (Gorunescu, 2011), as shown in Table 1.

Table 1. Confusion Matrix

Actual class	Predicted class		
	Positive	Negative	Netral
Positive	True positive	False negative	False Netral
Negative	False positive	True negative	False Netral
Netral	False Netral	False Netral	True Netral

The table above shows nine possibilities that can occur from the classification results. Through all the possibilities above, the accuracy of the model can be obtained with this formula. $Accuracy = true\ prediction / total\ data$. Accuracy calculations are carried out up fold to k. Then the performance of the model is obtained from the average value of the iterations performed.

C. RESULT AND DISCUSSION

Data retrieval uses web scraping with the sncrape library in Python software because it allows retrieval of large amounts of data. Retrieved raw data of 776.484 which includes tweets related to PPKM and Covid-19 Vaccination. The data obtained from twitter is unstructured text data, it is necessary to transform the data in order to structure and facilitate further processes. Examples of original data or not yet going through the pre-processing listed in Table 2.

Table 2. Twitter Raw Data Example

Tweet 1	Efforts to prevent the spread of covid 19, Wonosegoro Koramil, Wonosegoro Police and Kec. Wonosegoro night patrols to prevent crime and socialization of progress during PPKM regulations. https://t.co/usehAslKtz
Tweet 2	5 Schools in Klaten Will Hold Face-to-face Learning, Teachers Get COVID-19 Vaccines First Dose https://t.co/lkTLmVRJsC
Tweet 3	Weird, why is it only now that Sinovac's Covid-19 vaccine is said to be low in effectiveness #Vaksin #Corona #Sinovac https://t.co/LBafh734mZ

Table 2 show the raw data that will go through data pre-processing, which includes the stages of case folding, tokenizing, stemming, removing stopwords, and spelling normalization.

1. Case folding is a process of uniforming all tweet data into a certain form. This stage consists of several parts, namely:

- a. Changing all words to lowercase, this is done at the researcher's discretion so that it is easier to observe a lot of text data.
- b. Delete mentions contained in tweets. This is done because the mention contains the name of a twitter user that has no significant relationship with the opinions contained in the tweet. For example: @infomalang.
- c. Delete the hashtags contained in each tweet. This is based on the researcher's direct observation of hashtags in each tweet, where there is no space between words in the hashtag and does not describe certain emotions. For example: #ppkm #vaccinocovid19.
- d. Delete the web address listed in each tweet. For example: <http://t.co/RXTPHm>.
- e. Delete all characters other than letters in each tweet. This will remove punctuation marks, numbers, and emoticons in each record. For example: COVID-19 → covid.
- f. Removing excessive repetition of letters in each word. Excessive repetition of letters will complicate the next data pre-processing process, especially in the process of standardizing words (spelling normalization). For example: no → no.
- g. Deleting repeated data

This case folding process reduces 164,741 data, this number shows that there is a lot of repetitive data from the results of web scrapping Twitter data. After going through the case folding process, the remaining 604,333 data can be continued to the next process. The examples of the results obtained from the case folding process are written in Table 3.

Table 3. Case Folding Results

<i>Tweet</i>	<i>Case Folding Results</i>
Efforts to prevent the spread of covid 19, Wonosegoro Koramil, Wonosegoro Police and Kec. Wonosegoro night patrols to prevent crime and socialization of progress during PPKM regulations. https://t.co/usehAslKtz	efforts to prevent the spread of covid, Wonosgoro Koramil, Wonosegoro Police, and Wonosegoro Sub-district Transtib, night patrols to prevent crime and socialization of health procedures during PPKM regulations
5 Schools in Klaten Will Hold Face-to-face Learning, Teachers Get COVID-19 Vaccines First Dose https://t.co/lkTLMVRJsC	schools in klaten will hold face-to-face learning, teachers get the first dose of covid vaccine
Weird, why is it only now that Sinovac's Covid-19 vaccine is said to be low in effectiveness #Vaksin #Corona #Sinovac https://t.co/LBafh734mZ	It's strange why only now the Sinovac covid vaccine is said to be of low effectiveness

2. Tokenizing

The data that has gone through the case folding stage is then continued with the tokenizing process. This process is in the form of splitting sentences into chunks of words called tokens. In this study, word cutting is based on a single word, where each single word will have a certain value. At this stage, the frequency for each single word that appears is also calculated, this is done to be able to see all the words used more easily. This process is carried out with the Counter library where there are 223,186 variables from all the data contained as shown in Table 4.

Tabel 4. Tokenizing Results

No	Variable	Frequency
1	Covid	518.715
2	Vaccine	495.610
3	PPKM	403.997
4	In	247.432
5	Wich	151.935
:	:	:
2231.186	Agree	1

3. Stemming

This process serves to reduce research variables by changing each word into its basic word. At this stage, the Sastrawi library is used which is able to process the Indonesian language. This library has been specially designed to handle high quality and well documented textual data. Also used is a collection of basic Indonesian words that refer to the KBBI to support the stemming process. Examples of stemming results can be seen in Table 5.

Tabel 5. Streaming Results in Indonesian

Initial Words	Streaming Results in Indonesian
<i>Menyetujui</i>	<i>Setuju</i>
<i>Melacak</i>	<i>Lacak</i>
<i>Melihat</i>	<i>Lihat</i>
<i>Dibilang</i>	<i>Bilang</i>

4. Removing Stopwords

This process aims to reduce stopwords. Stopwords are a collection of meaningless words in a sentiment analysis series. The disposal of these words is done to get maximum performance results, where the words that will be analyzed are only words that have a certain meaning. In this study, the Indonesian language stopwords provided by the Satrawi library were used as many as 124 words.

5. Spelling Normalization

Spelling normalization or normalization of pronunciation means the standardization of the words contained in this study. This is important because in the data in the form of tweets it is not uncommon to find words in the form of abbreviations, colloquialisms, slang (slang), and even very diverse regional languages. The pronunciation normalization process is done by matching words that have gone through the stemming process and removing stopwords. The number of words processed in this stage is 222,997 words. Each word is matched with the KBBI to separate words that are standard and words that are not standard. Furthermore, the researcher made an additional dictionary manually for words that were not standardized, as many as 2,620 words.

The labeling process is the stage of giving the initial sentiment class to each tweet before being tested using the FIS Sugeno classification method. To meet the need for a classification method, tweet data must be grouped into three sentiment classes, namely positive sentiment, negative sentiment, and neutral sentiment. In this study, it is necessary to label 604,333 tweets that have gone through pre-processing of data.

Basically, labeling can be done manually by providing the right sentiment according to the available tweets, but this will take a long time especially with a large amount of data. In this study, automatic labeling was carried out with the help of an emotion dictionary or sentiment lexicon dataset. The emotion dictionary contains a collection of words (variables) that refer to a particular class.

The labeling process is based on the number of positive and negative variables in each tweet, where each negative variable is worth -1 and each positive variable is +1. Then all values are added up, so that tweets with a value of less than 0 will be labeled as negative sentiment, tweets with a value greater than 0 will be labeled as positive sentiment, and tweets with a value of 0 will be labeled as neutral sentiment. This process is done with Python Software, where the results of data splitting can be explained in Table 6.

Tabel 6. Data Split Results

Type	Topics	Sentiment			Amount of Data
		Positive	Negative	Neutral	
Micro	PPKM	44796	18814	47268	110878
	Vaccine Covid-19	39138	37200	51385	127723
National	PPKM	53563	40482	83115	177160
	Vaccine Covid-19	69997	41576	61357	172930

Based on Table 6, it is explained that the type of micro PPKM has a positive sentiment of 44796 where the data is obtained from February 2021 to July 2021. The meaning of positive sentiment is that the data supports the PPKM policy. In contrast to negative sentiment, namely data that contradicts or contradicts with PPKM as much as 18814. While neutral sentiment is news that is in the middle in the sense that it does not support PPKM and also does not conflict with PPKM as much as 47268. Micro regarding PPKM is 110878. The results obtained are not much different from the statement on the covid-19 vaccine which has a total of 127723 data with positive sentiments amounting to 39138, negative sentiments 37200 and neutral sentiments 51385. For the National type, it can be seen that PPKM has positive sentiments of 53563, negative sentiments of 40482 and neutral sentiments of 83115. These results can be seen that the total number of data on National PPKM is 177160. While for the Covid-19 Vaccine as many as 172930 with positive sentiment sharing as much as 69997, negative sentiment is 41576 and neutral sentiment is 61357.

The weighting in this study uses term frequency (TF) which focuses on the number of terms that appear in each tweet. The weighting is intended to determine the value that will be the input variable in the FIS. It is known that the input variables used are positive term frequency and negative term frequency. The terms used are adjusted to the needs of the input variables, namely positive and negative words that appear in each tweet. This is related to the labeling process so that it can take advantage of the labeling algorithm to get weight data from each tweet. The results of the study were visualized to be able to describe the opinion of the Indonesian people towards the PPKM and Covid-19 Vaccination policies. Based on Table 6, the results of the study can be re-described into a picture of Indonesian public opinion on the PPKM and Covid-19 Vaccination policies both as a whole, adjusted to the type of policy, and reviewed by month. Public opinion as a whole starting from February 2021 – January 2022 is depicted in Figure 3.

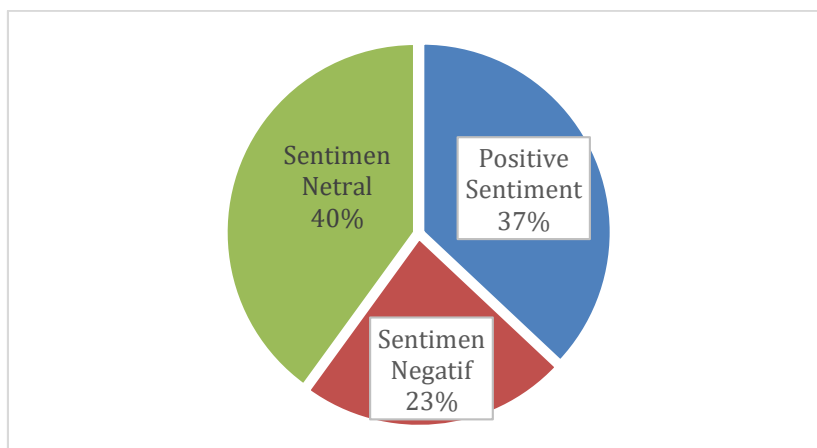


Figure 3. Aggregate Sentiment Polarity

As explained in the previous discussion, public sentiment regarding the PPKM and Covid-19 vaccination policies tends to be balanced for positive and neutral sentiment. As for the negative sentiment, it was less, which was as much as 23%. Based on direct observation of research data, most of the tweets contained support for the PPKM policy, considering the high number of Covid-19 virus infections which was quite worrying at the time. As for negative sentiments, they are mostly found in tweets related to the Covid-19 Vaccination policy, which in fact has received a lot of criticism because it is a new vaccine in the world.

D. CONCLUSION

Based on the analysis of the main functional components that have been carried out, it can be concluded that: First, the Sugeno FIS can be used in the classification of opinion data. In this study, Indonesian opinions regarding PPKM and Covid-19 vaccination policies were used and the results showed 36.92% of tweets were classified as positive sentiments, 22.85% negative sentiments, and another 40.23% classified as neutral sentiments. Second, the form of the fuzzy set based on the data observation method is very good because it is able to regulate the frequency of data in each category. This is very helpful in improving the performance of the built classification system. third, the performance of the classification system using the Sugeno FIS is quite good with an average accuracy of 89.13%. Fourth, it is useful as a consideration for the government in implementing PPKM and vaccination policies if there are visions that are as dangerous as COVID-19. Fifth, suggestions for further research are to apply fuzzy Sugeno in the business, economic, and social fields of grouping or classifying a problem. Sixth, the suggestion for further research is that in the Twitter data retrieval process there are still many tweets that are taken repeatedly, this can be overcome by improving the web scrapping algorithm, ensuring a stable internet network during the data retrieval process, and performing data retrieval periodically. Seventh, further research can determine the membership function using other methods such as particle swarm optimization, k-means, EM clustering, hyper cone, and so on. Of course, this generator needs to be adapted to research needs.

ACKNOWLEDGEMENT

A very big thanks to the Faculty of Mathematics and Natural Science Universitas Brawijaya for all the support. I also would like to thank to Gusti Pangestu and Diyan Nur Rakhman for fruitful discussion as well as their support and encouragement.

REFERENCES

- Aryandani, A., Solimun, Nurjannah, Rinaldo Fernandes, A.A. & Efendi, A. (2022). Implementation of Fuzzy C-Means in Investor Group in the Stock Market Post-Covid-19 Pandemic. *WSEAS Transactions on Mathematics*, 21, 415–423.
- Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems*, 226, 107134.
- Civelek, Z. (2020). Optimization of fuzzy logic (Takagi-Sugeno) blade pitch angle controller in wind turbines by genetic algorithm. *Engineering Science and Technology, an International Journal*, 23(1), 1–9.
- Djalante, R., Lassa, J., Setiamarga, D., Sudjatma, A., Indrawan, M., Haryanto, B., ... & Warsilah, H. (2020). Review and analysis of current responses to COVID-19 in Indonesia: Period of January to March 2020. *Progress in Disaster Science*, 6, 100091.

- Efrilianda, D., Dianti, E., & Khoirunnisa, O. (2021). Analysis of Twitter Sentiment in Covid-19 Era Using Fuzzy Logic Method. *Journal of Soft Computing Exploration*, 1–5.
- Fu, G., & Wang, X. (2010). Chinese Sentence-Level Sentiment Classification Based on Fuzzy Sets. *Coling*, 312–319.
- Gorunescu, F. (2011). *Data Mining: Concepts, Model, and Techniques*. Springer.
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques third edition*. Morgan Kaufmann Publisher.
- Hu, Z., Wang, J., Zhang, C., Luo, Z., Luo, X., Xiao, L., & Shi, J. (2021). Uncertainty modeling for multi center autism spectrum disorder classification using takagi-sugeno-kang fuzzy systems. *IEEE Transactions on Cognitive and Developmental Systems*.
- Iwendi, C., Mahboob, K., Khalid, Z., Javed, A. R., Rizwan, M., & Ghosh, U. (2022). Classification of COVID-19 individuals using adaptive neuro-fuzzy inference system. *Multimedia Systems*, 28(4), 1223–1237.
- Kaur, A. (2012). Comparison of Fuzzy Logic and Neuro Fuzzy Algorithms for Air Conditioning System. *International Journal of Soft Computing and Engineering*, 417–420.
- Kumar, K., & Prakash, A. (2019). Developing a framework for assessing sustainable banking performance of the Indian banking sector. *Social Responsibility Journal*.
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan and Claypool Publishers.
- MathWorks. (2014). *Fuzzy Logic Toolbox User's Guide*. The MathWorks, Inc.
- Paroubek, P., & Pak, A. (2010). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. *Seventh International Conference on Language Resource and Evaluation*.
- Pedersen, Y. Y. (1997). A comparative study on feature selection in text categorization. *Proc of ICML-97*, 412–420.
- Rauf, H. T., Malik, S., Shoaib, U., Irfan, M. N., & Lali, M. I. (2020). Adaptive inertia weight Bat algorithm with Sugeno-Function fuzzy search. *Applied Soft Computing*, 90, 106159.
- Şahin, M. (2021). A comprehensive analysis of weighting and multicriteria methods in the context of sustainable energy. *International Journal of Environmental Science and Technology*, 18(6), 1591–1616.
- Serrano-Guerrero, J., Romero, F., Olivas, J. (2021). Fuzzy logic Applied to Opinion Mining: A Review. *Journal of Knowledge-Based Systems*.
- Setiawan, H., Utami, E., & S. (2021). Twitter Sentiment Analysis Post-Covid-19 Online Lecture Using Support Vector Machine and Naive Bayes Algorithm. *Journal of Computing and Informatics*, 43–51.
- T Thakkar, H., Shah, V., Yagnik, H., & Shah, M. (2021). Comparative anatomization of data mining and fuzzy logic techniques used in diabetes prognosis. *Clinical EHealth*, 4, 12–23.
- Thangaraj, M., & Sivakami, M. (2018). Text Classification Techniques: A Literature Review. *Interdisciplinary Journal of Information, Knowledge, and Management (IJIKM)*, 118–134.
- Torkayesh, A. E., Ecer, F., Pamucar, D., & Karamaşa, Ç. (2021). Comparative assessment of social sustainability performance: Integrated data-driven weighting system and CoCoSo model. *Sustainable Cities and Society*, 71, 102975.
- Twitter. (2021). <https://about.twitter.com/>.
- Villavicencio, C., Macrohon, J., Inbaraj, X., Jeng, J., & Hsieh, J. (2021). Twitter Sentiment Analysis towards Covid-19 Vaccines in the Phileppines Using Naïve Bayes. *Journal of Information*.
- Wang, L. L., & Lo, K. (2021). Text mining approaches for dealing with the rapidly expanding literature on COVID-19. *Briefings in Bioinformatics*, 22(2), 781–799.
- Wang, S., Celebi, M. E., Zhang, Y. D., Yu, X., Lu, S., Yao, X., ... & Tyukin, I. (2021). Advances in data preprocessing for biomedical data fusion: an overview of the methods, challenges, and prospects. *Information Fusion*, 76, 376–421.
- Xia, E., Yue, H., & Liu, H. (2021). Tweet sentiment analysis of the 2020 US presidential election. *Companion Proceedings of the Web Conference 2021*, 367–371.