JPPIPA 8(5) (2022)

Jurnal Penelitian Pendidikan IPA

Journal of Research in Science Education

http://jppipa.unram.ac.id/index.php/jppipa/index

Fuzzy Sugeno Method for Opinion Classification Regarding Policy of *PPKM* and Covid-19 Vaccination

Djihan Wahyuni[1*], Eni Sumarminingsih[1], Suci Astutik[1]

[1] Departement of Statistics, Universitas Brawijaya, 65145, Malang, Indonesia

Received: August 3, 2022
Revised: October 18, 2022
Accepted: November 27, 2022
Published: November 30, 2022

Corresponding Author:
Djihan Wahyuni
djihandije@gmail.com

DOI: 10.29303/jppipa.v8i5.1958

Abstract: The Indonesian government has implemented various interventions to overcome the impact of the Covid-19 pandemic, including those written in Minister of Home Affairs Instructions on *PPKM* (Community Activities Restrictions Enforcement) and Covid-19 vaccination policies. This policy are not at least reaping the pros and cons, so it is necessary to monitor public opinion to be able to provide solutions or become an evaluation of future policies. The aim of this study is to determine the polarity of public opinion regarding PPKM and Covid-19 vaccinations policies on Twitter, as well as to determine the implementation of FIS Sugeno in classifying textual data. There are several stages carried out, i.e. data collection, data pre-processing, data labeling, data weighting, identification of membership functions, determination of fuzzy sets, formation of a classification system, and evaluation of classification results. In this study, the performance of FIS Sugeno in classifying tweets was quite good with an average accuracy of 89.13%. Meanwhile, public opinion regarding the PPKM and Covid-19 vaccination policies tends to be balanced with 36.92% of tweets classified as a positive sentiments, 22.85% being negative sentiments, and another 40.23% belonging to neutral sentiments.

Keywords: Sentiment Analysis; FIS Sugeno; PPKM; Covid-19 Vaccination

Introduction

Various policies have been implemented to minimize the impact of the Covid-19 pandemic. Community participation is one of the keys to success in overcoming the Covid-19 pandemic. However, the policies offered by the government have many pros and cons. Given the importance of the PPKM and Covid-19 vaccination policies, it is necessary to monitor public opinion to be able to provide solutions or become an evaluation of future policies.

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

How to Cite:
Wahyuni, D., Sumarminingsih, E., & Astutik, S. (2022). Fuzzy Sugeno Method for Opinion Classification Regarding Policy of PPKM and Covid-19 Vaccination. *Jurnal Penelitian Pendidikan IPA*, *8*(5), 2210–2215. https://doi.org/10.29303/jppipa.v8i5.1958

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). The topics raised in this research are Indonesian public opinion regarding *PPKM* and Covid-19 vaccinations policies. The performance of the resulting model can be a consideration for policies to overcome the Covid-19 pandemic in the future.

Fuzzy logic is a branch of an artificial intelligence system that packs human thinking abilities into algorithms which are then run by machines. This logic is generally applied to problems that contain elements of uncertainty and imprecision (Thakkar, et al., 2021). Fuzzy is a method of counting using variables in the form of words (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 much 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).

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 certain products, services, or activities (Liu, 2012; Ravi & Ravi, 2015). According to KBBI, a sentiment means an opinion or view that is based on feelings towards something. Sentiment itself can be interpreted as an opinion. An analysis of sentiment or opinion is important to be used as an evaluation material for certain groups (Xia, et al., 2021). For example in the business world, sentiment analysis is useful for determining consumer satisfaction with marketed goods and services, so that it can be considered in improving the quality of production (Birjali, et al., 2021).

Text mining is a technology that is able to analyze semi-structured and unstructured text data (Han, et al., 2011). Text data generally has some characteristics, such as high dimensions, noise, and bad textual structure (Wang & Lo, 2021). 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 methods in text mining is to do text preprocessing. Text preprocessing aims to repair the data structure in order to be properly by the computer (Kumar & Prakash, 2019).

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).

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 (Yang & Pedersen, 1997). Weighting using term frequency focuses on the number of terms that appear in each document (Torkayesh, et al, 2021; Şahin, 2021).

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 1 shows iterations on 10-fold cross validation.

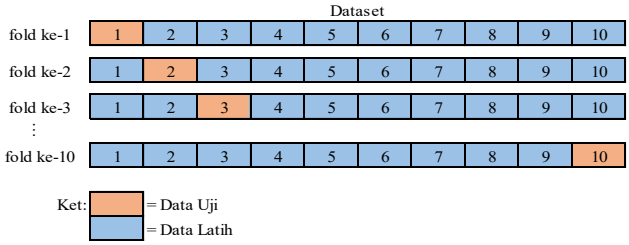

Figure. 1. 10-Fold Cross Validation

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).

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 calculations are carried out up fold to k. Then the performance of the model is obtained from the average value of the iterations performed.

Method

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", "vaccinecovid", and "vaccine" in the search column. The analysis steps in this study are as follows:
Collecting tweet

- Data Preprocessing
- Labeling
- Weighting
- Identification of membership functions
- Formation of fuzzy rules
- Classification with FIS Sugeno
- Evaluation of Classification Results

Result and Discussion

Data retrieval uses web scraping with the snscrape 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. Raw Data

Twitter	Topics
Tweet 1	Upaya cegah sebaran covid19 Koramil wonosrgoro, polsek wonosegoro dan trantib kec. Wonosegoro patroli malam cegah timdak kriminal dan sosialisasi prokes di saat aturan PPKM. https://t.co/usehAslKtz
Tweet 2	Aneh, Knapa Baru Sekarang Vaksin Covid-19 Sinovac Dibilang Efektivitas Rendah #Vaksin #Corona #Sinovac https://t.co/LBafh734mZ

Preprocessing data started with case folding, this is a process of uniforming all tweets into a certain form. It consists of several parts, which turn whole letter into smaller (lowercase), deleting mentions (@), deleting hashtags (#), deleting web addresses, deleting characters besides the alphabet, and removing excessive letter repetition. This process uses Regex library in Python Software.

Tokenizing is a process of splitting sentences into chunks of words called tokens. In this study, word-cutting is based on a single word, where every single word will later become a variable. This process is done with the Counter library in Python Software. In this study, there were 223.186 variables consisting of the most common words and words that only appeared once from the entire data. Stemming is the process of mapping and parsing the form of a word. This process serves to reduce research variables by changing affixed words into basic words. Spelling normalization is a process of standardization of words is carried out, either correcting misspellings of words or words written in abbreviated form. For example the word "tidak" which has many forms of writing such as "gak", "tdk", "g", "ga", and so on. The last step of preprocessing is removing stopwords. Stopwords are collections of meaningless words. This word removal 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, stopwords which had been provided by the library Sastrawi. In this study, the labeling process was not carried out manually, but using the sentiment lexicon dataset or emotion dictionary. 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 that have a value less than 0 will be labeled as negative sentiment and tweets with a value greater than 0 will be labeled as positive sentiment. Table 3 displays the results of labeling all tweets.

Table 3. Labelling Result

Month	Positive Sentiment	Negative Sentiment	Netral Sentiment
PPKM (Micro)			
Feb	9.291	2.495	8.322
Mar	6.469	1.409	4.733
Apr	4.748	1.089	3.005
May	3.356	899	2.518
Jun	8.524	4.114	9.465
Jul	12.408	8.808	19.225
Covid-19 Vaccination (Micro)			
Feb	4.061	2.020	4.538
Mar	8.866	4.488	9.150
Apr	4.909	3.272	5.514
May	6.957	5.961	6.702
Jun	12.607	8.278	11.083
Jul	17.380	13.181	14.398
PPKM (National)			
Jul	4.464	3.106	2.815
Aug	11.774	9.706	22.487
Sep	12.776	8.734	19.282

Month	Positive Sentiment	Negative Sentiment	Netral Sentiment
Oct	8.171	5.512	11.397
Nov	7.698	4.426	11.385
Dec	6.102	6.921	11.615
Jan	2.578	2.077	4.134
	Covid-19 Vaccination (National)		
Jul	6.235	4.678	5.678
Aug	16.220	10.969	15.078
Sep	17.235	9.704	12.829
Oct	7.919	4.586	6.744
Nov	5.885	3.378	4.890
Dec	4.959	3.076	4.668
Jan	11.544	5.185	11.470
Total	223.136	138.072	243.125

The weighting uses term frequency 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.

Membership functions used based on the subjectivity of the researcher and the probability of fuzzy in each category in a fuzzy. Figures 2 and 3 show the membership functions formed for the two input variables.

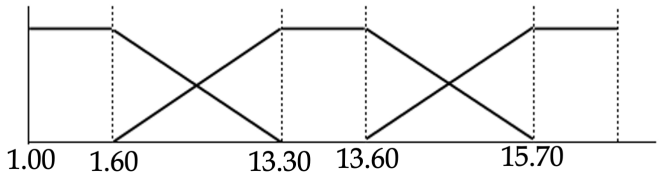

Figure 2. Fuzzy Set of TF Positive

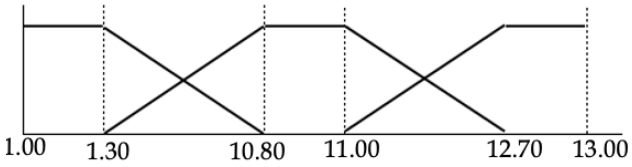

Figure 3. Fuzzy Set of TF Negative

Based on the two figures above, the input variables formed in this study can be described in Table 4.

Table 4. Input Variables

Variable	Category	Domain
TF Positive	Low	(1.00, 1.60, 13.30)
	Medium	(1.6, 13.3, 13.6, 15.72)
	High	(13.60, 15.72, 16.00)
TF Negative	Low	(1, 1.30, 10.80)
	Medium	(1.3, 10.8, 11.05, 12.74)
	High	(11.05, 12.74, 13.00)

In performing classification with a fuzzy inference system, input data, membership functions and fuzzy rules are required. Fuzzy rules used are as described in Table 5.

Table 5. Fuzzy Rules

TF Positive (a)	TF Negative (b)	Z Value
Low	Low	$0{,}2a - 0{,}1b$
Low	Medium	$0{,}2a - 0{,}3b$
Low	High	$0{,}2a - 0{,}5b$
Medium	Low	$0{,}4a - 0{,}1b$
Medium	Medium	$0{,}4a - 0{,}3b$
Medium	High	$0{,}4a - 0{,}5b$
High	Low	$0{,}6a - 0{,}1b$
High	Medium	$0{,}6a - 0{,}3b$
High	High	$0{,}6a - 0{,}5b$

Evaluation of the classification system was used with 10-fold cross validation. Figure 4 below shows the results of the classification in this study.

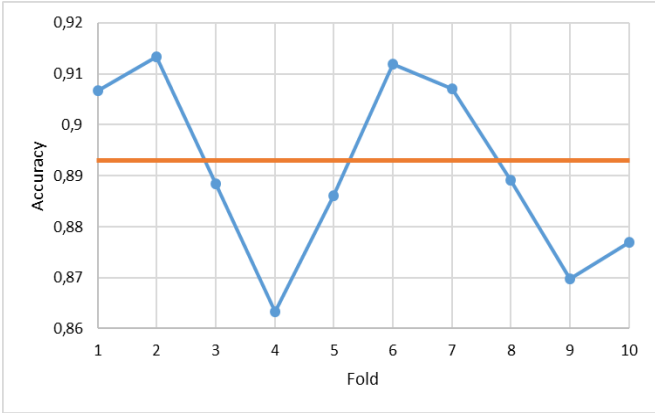

Figure 4. 10-Fold Cross Validation Result

From Figure 4, it is known that the average accuracy of the sentiment classification model using the Sugeno FIS on tweets is 89.127%.

Conclusion

Based on the main functional component analysis that has been carried out, it can be concluded that: FIS Sugeno can be used in the classification of opinion data. In this study, Indonesian opinions were used regarding PPKM and Covid-19 Vaccination policies and the results showed that 36.92% of tweets were classified as positive sentiments, 22.85% were negative sentiments, and the other 40.23% were classified as neutral sentiments. A fuzzy set form 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 is very helpful in improving the performance of the built classification system. The performance of the classification system using the Sugeno FIS which is made is quite good with an average accuracy of 89.13%.

Acknowledgements

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. I. S. Y. A. H., Solimun, N., Rinaldo Fernandes, A. A., & Efendi, A. C. H. M. A. D. (2022). Implementation of Fuzzy C-Means in Investor Group in the Stock Market Post-Covid-19 Pandemic. *WSEAS Transactions on Mathematics*, *21*, 415-423. DOI: 10.37394/23206.2022.21.49

Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems*, *226*, 107134. https://doi.org/10.1016/j.knosys.2021.107134

Efrilianda, D. A., Dianti, E. N., & Khoirunnisa, O. G. (2021). Analysis of twitter sentiment in COVID-19 era using fuzzy logic method. *Journal of Soft Computing Exploration*, *2*(1), 1-5. DOI: https://doi.org/10.52465/joscex.v2i1.12

Fu, G., & Wang, X. (2010). Chinese Sentence-Level Sentiment Classification Based on Fuzzy Sets. Coling, 312–319. Retrieved from https://aclanthology.org/C10-2036.pdf

Gorunescu, F. (2011). Data Mining: Concepts, Model, and Techniques. Springer. DOI: 10.1007/978-3-642-19721-5_5

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*. DOI: 10.1109/TCDS.2021.3073368

Kaur, A., & Kaur, A. (2012). Comparison of fuzzy logic and neuro-fuzzy algorithms for air conditioning system. *International journal of soft computing and engineering*, *2*(1), 417-20. Retrieved from shorturl.at/OQT38

Kumar, K., & Prakash, A. (2018). Developing a framework for assessing sustainable banking performance of the Indian banking sector. *Social Responsibility Journal*, *15*(5), 689-709. https://doi.org/10.1108/SRJ-07-2018-0162

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, *5*(1), 1-167. https://doi.org/10.2200/S00416ED1V01Y201204HLT016

MathWorks. (2014). Fuzzy Logic Toolbox User's Guide. The MathWorks, Inc.

Pak, A., & Paroubek, P. (2010, May). Twitter as a corpus for sentiment analysis and opinion mining. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. Retrieved from https://aclanthology.org/L10-1263/

Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowledge-based systems*, *89*, 14-46. https://doi.org/10.1016/j.knosys.2015.06.015

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. Retrieved from https://link.springer.com/article/10.1007/s13762-020-02922-7

Serrano-Guerrero, J., Romero, F. P., & Olivas, J. A. (2021). Fuzzy logic applied to opinion mining: a review. *Knowledge-Based Systems*, *222*, 107018. https://doi.org/10.1016/j.knosys.2021.107018

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.

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. https://doi.org/10.1016/j.ceh.2020.11.001

Thangaraj, M., & Sivakami, M. (2018). Text Classification *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. https://doi.org/10.1016/j.scs.2021.102975

Twitter. (2021). Retrieved from https://about.twitter.com/.

Villavicencio, C., Macrohon, J. J., Inbaraj, X. A., Jeng, J. H., & Hsieh, J. G. (2021). Twitter sentiment analysis towards covid-19 vaccines in the Philippines using naïve bayes. *Information*, *12*(5), 204. https://doi.org/10.3390/info12050204

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. https://doi.org/10.1093/bib/bbaa296

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. https://doi.org/10.1016/j.inffus.2021.07.001

Xia, E., Yue, H., & Liu, H. (2021). Tweet sentiment analysis of the 2020 US presidential election.

In *Companion Proceedings of the Web Conference 2021* (pp. 367-371). https://doi.org/10.1145/3442442.3452322

Yang, Y., & Pedersen, J. O. (1997, July). A comparative study on feature selection in text categorization. In *Icml* (Vol. 97, No. 412-420, p. 35).