

Analysis of Indonesian Public Opinion Sentiment on Policy on Twitter Social Media “PPKM” Using K-Nearest Neighbor

Aulia Khoirunnita *

Amikom University, Yogyakarta, 55281, Indonesia
aulia.1270@students.amikom.ac.id

*Corresponding author

Kusnawi 

Amikom University, Yogyakarta, 55281, Indonesia
khusnawi@amikom.ac.id

Received: 2021-08-12; Revised: 2021-09-26; Accepted: 2021-11-19; Published: 2021-12-01

Abstract—COVID-19 or Coronavirus disease 2019 is currently a pandemic that is spreading very quickly throughout the world, including Indonesia. Various handling and policies have been carried out, one of which is called “PPKM” policy or what can also be called the Enforcement of Restrictions on Community Activities issued by the Indonesian government. “PPKM” is currently one of the topics that is often discussed by the public, one of which is on the Twitter social media platform. The existence of opinions given by the community, it is necessary to have a sentiment analysis. Sentiment analysis is an analytical process obtained from various social media platforms and the internet. The aim is to find out how the public's sentiment towards the implementation of “PPKM” policies in Indonesia is through tweets and comments on the Twitter social media platform. In this study, the process of analyzing public opinion regarding the “PPKM” policy will be carried out by classifying opinions into 3 sentiments, namely positive, negative or neutral. Classification is done using the K-Nearest Neighbor algorithm. The K-Nearest Neighbor (K-NN) algorithm is a classification method for a set of data based on previously classified data learning. Included in supervised learning, where the results of the new query instance are classified based on the majority of the distance proximity of the categories in K-NN. The results of data preprocessing and sentiment classification, in the first test positive sentiment 37.6% of 261 data, negative sentiment 65.9% of 636 data and neutral sentiment 9

Keywords—Sentiment Analysis, K-Nearest Neighbor, PPKM, Covid-19

I. INTRODUCTION

At the beginning of 2020 until now, the world's citizens infected with Covid-19 are increasing every day. Based on KOMPAS.com data, the number of patients infected with corona in the world, until March 1, 2021 + 114 million cases. (Toharudin et al, 2021) Data that are positive for Covid-19, amounting to hundreds of thousands of patients who died and 64.4 million were declared cured. The development of Covid-19 in

Indonesia is still increasing in number. As a result of the increase in Covid-19 cases, the government implemented various policies in the form of “PSBB” (Mukhtadi et al, 2021), lock down, restrictions on a limited scale, social distancing, new normal, obeying health protocols and other policies. (Napitu et al., 2021).

This policy has the impact of worsening economic conditions and other multidimensional impacts. Realizing that this policy cannot overcome the Covid-19 pandemic, the government from mid-February 2021 implements a policy called the Implementation of Micro Community Activity Restrictions “PPKM” and seeks mass vaccines for all citizens in stages throughout Indonesia. (Ministry of Home Affairs Regulation (Permendagri), 2020). This community service activity aims (Imam et al, 2021), overcoming the spread of Covid-19 which is still increasing, at the village and sub-district level, optimizing restrictions on community activities in various aspects of life up to the village/district level by implementing restrictions on community activities “PPKM” Micro, and Increase public awareness to implement and comply with health protocols, social distancing and restrictions on micro community activities “PPKM” (Miharja et al, 2021).

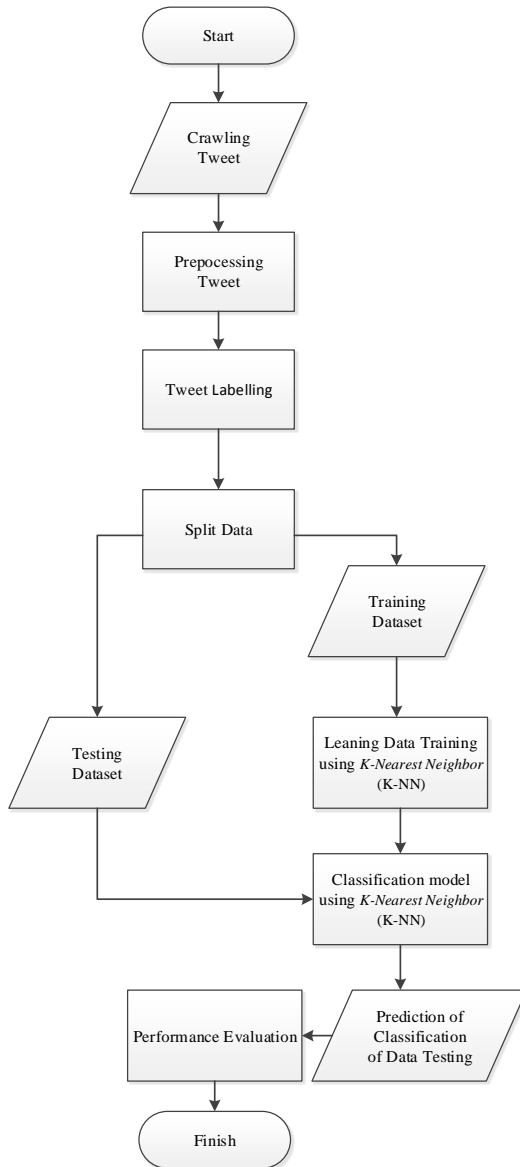
Sentiment analysis to determine the trend of public opinion on the impact of the Corona virus (Astari et al., 2020). Twitter is a social networking service website that is in great demand by internet users as a medium of communication and getting information (Nurjanah et al., 2017). Twitter is one of the platforms used by the public to express the latest conditions after the Corona virus spread. The purpose of this study is to obtain an analysis of text documents to obtain positive, negative or neutral public sentiments. The data used is a tweet document from Twitter about the impact of the Corona virus. The data collected is divided to be used as training data and test data for the classification process. The method used for classification in this study is the Naive Bayes Classifier Method. Given the existing problems, it is necessary to have a sentiment analysis to find out public opinion about PPKM policies by

classifying opinions into 2 sentiments, namely positive or negative. Classification is done using the K-Nearest Neighbor (K-NN) Algorithm.(Sari, 2020). The K-Nearest Neighbor (K-NN) algorithm is also a classification method for a set of data based on previously classified data learning. Included in supervised learning, where the results of the new query instance are classified based on the majority of the distance proximity of the categories in K-NN.

II. RESEARCH METHOD

A. Materials and Methods

This research was conducted using the K-Nearest Neighbor algorithm, namely the data collection was carried out using the tweet crawling technique, tweet preprocessing, classification and performance evaluation. With the design of the research flow can be seen in Picture 1



Picture 1. Research Stage

A. Crowd Tweet

Crawl is one of the methods used to collect or download data from a database (Gabelkov et al, 2016). In collecting tweet data to be downloaded from the Twitter database, it must go through an authentication process through the Twitter Application Programming Interface (API) (Zheng et al, 2013). Twitter API is a program or application provided by Twitter to make it easier for other developers to access information on the Twitter website (Bošnjak et al, 2012).

B. Preprocessing Tweet

Preprocessing carried out to eliminate incomplete, noisy and inconsistent data with the aim of normalizing the data. (Alasadi et al, 2017; Nayak et al, 2016) The process carried out in preprocessing tweets is case folding, namely changing all capital letters to lowercase, eliminating URLs, eliminating mentions/usernames, namely removing words starting with the "@" symbol, eliminating hashtags, namely removing words starting with the "#" symbol, eliminating symbols such as punctuation marks and emoticons/emojis, then removes Retweets, namely tweets that are copied by other users and then reposted, usually starting with "RT" at the beginning of the tweet (Son et al, 2019; Abrar et a; 2019; Astari et al., 2020). In addition, the tokenizing process or what can be called parsing is also carried out, namely cutting the input string in a document based on each word that composes it, in this case the data for each tweet is separated by each word (Pratama et a;, 2019). The tokenizing process aims to simplify the process of eliminating stop words. The last process is eliminating stop words, namely words that are most often used in sentences but have no meaning, in English the words that are used as stop words such as pronouns, prepositions, conjunctions and others (Salim, 2020).

C. K-Nearest Neighbors

KNN is a classification algorithm that uses an approximation between the value of K in the surrounding conditions (neighbors-system) (Ali et al, 2020). K-NN in the classification process must divide the dataset into training and testing data so that the calculation of the closest distance is easier. K-NN in the calculation of the closest distance using the Euclidean method from testing data to training data. The following is the formula for calculating the Euclidean distance (1).

$$Q^{(A,B)} = \sqrt{\sum_{k=1}^N (A_k - B_k)^2} \quad (1)$$

Variables (1):

- Q = Data distance of two variables A and B
 A = Training Data
 B = Testing Data
 N = Data dimension
 K = Data variables

D. Performance Evaluation

The evaluation results will be obtained from the classification process from KNN to predict sentiment has been carried out and obtain accuracy results from testing the K value by using a confusion matrix to compare the results of the K test continuously (2) (Chikh, 2012). The following is the formula for calculating the confusion matrix in the assessment of accuracy, precision and recalculation at the final stage of K-NN (3) (4)

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (2)$$

$$\text{precision} = \frac{TP}{FP + TP} \times 100\% \quad (3)$$

$$\text{Recall} = \frac{TP}{FN + TP} \times 100\% \quad (4)$$

III. RESULT AND DISCUSSION

A. Data collection

In collecting data obtained from Twitter related to PPKM from July 1, 2021 to July 14, 2021 with 964 data, it can be seen in table 1.

Table 1. Tweet Crawling Results

#	Class	Total
1	Negative	636
2	positive	261
3	Negative	67
Grand Total		964

B. Preprocessing Tweet

Preprocessing carried out to eliminate incomplete, noisy and inconsistent data with the aim of normalizing the data. The process carried out in preprocessing tweets is changing all capital letters to lowercase, removing URLs, eliminating mentions/usernames, namely removing words starting with the "@" symbol, eliminating hashtags, namely removing words starting with the "#" symbol, removing symbols such as punctuation marks and emoticons/emoji, then removes Retweets, namely tweets that are copied by other users and then re-posted, usually starting with "RT" at the beginning of the tweet. In addition, the tokenizing process or what can be called parsing is also carried out, namely cutting the input string in a document based on

each word that composes it, in this case the data for each tweet is separated by each word.

C. K-Nearest Neighbor Results

Testing accuracy with the K-NN algorithm to find out the results of the classification process carried out by dividing the data into 80% training data and 20% testing data. In the testing process using K-NN by providing validation K = 3, K = 5, and K = 7 and the accuracy results obtained will be compared with the confusion matrix can be seen in tables 2 and 3.

Table 2. Confusion matrix 1 with testing data 20 and training 80

#	K . value	Accuracy Score	Precision Score	Recall Score
1	K= 3	85.49	0.75	0.85
2	K= 5	85.49	0.55	0.56
3	K= 7	81.87	0.64	0.77

Table 3. Confusion matrix 2 with testing data 30 and training data 70

#	K . value	Accuracy Score	Precision Score	Recall Score
1	K=3	89.66	0.82	0.87
2	K=5	85.17	0.71	0.78
3	K=7	82.76	0.60	0.72

From the two tables, the validation accuracy values are different from the first data and the second validation of the K test results in the best accuracy from other K tests. Then the best results obtained are K = 3 85.49% than the other K from the first data and the second data produces the best K = 3 89.66%

V. CONCLUSION

With the application of the K-Nearest Neighbor classification algorithm, the results of the sentiment analysis of Indonesian public opinion on the application of PPKM can be concluded in data preprocessing, sentiment classification and the accuracy of the K-NN Algorithm. Based on data preprocessing and sentiment classification, in the first test positive sentiment was 37.6% from 261 data, negative sentiment was 65.9% from 636 data and neutral sentiment was 9.6% from 67 data. For the accuracy of the K-NN algorithm, the best results obtained are K = 3 85.49% than the other K from the first data and the second data produces the best K = 3 89.66%. Based on the results of the sentiment analysis of Indonesian public opinion on the implementation of PPKM, negative opinions are greater than positive opinions. So that the PPKM system needs to be evaluated and the PPKM system improved. For further research, it is expected to

increase accuracy in applying different algorithms to get more accurate results in sentiment analysis.

REFERENCES

- Abrar, M. F., Arefin, M. S., & Hossain, M. S. (2019, February). A framework for analyzing real-time tweets to detect terrorist activities. In 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE) (pp. 1-6). IEEE.
- Alasadi, S. A., & Bhaya, W. S. (2017). Review of data preprocessing techniques in data mining. *Journal of Engineering and Applied Sciences*, 12(16), 4102-4107.
- Ali, M., Jung, L. T., Abdel-Aty, A. H., Abubakar, M. Y., Elhoseny, M., & Ali, I. (2020). Semantic-k-NN algorithm: an enhanced version of traditional k-NN algorithm. *Expert Systems with Applications*, 151, 113374.
- Astari, NMAJ, Dewa Gede Hendra Divayana, & Gede Indrawan. (2020). Sentiment Analysis of Twitter Documents Regarding the Impact of the Corona Virus Using the Naive Bayes Classifier Method. *Journal of Systems and Informatics (JSI)*, 15(1), 27–29. <https://doi.org/10.30864/jsi.v15i1.332>
- Bošnjak, M., Oliveira, E., Martins, J., Mendes Rodrigues, E., & Sarmiento, L. (2012, April). Twittrecho: a distributed focused crawler to support open research with twitter data. In *Proceedings of the 21st international conference on World Wide Web* (pp. 1233-1240).
- Chikh, M. A., Saidi, M., & Settouti, N. (2012). Diagnosis of diabetes diseases using an Artificial Immune Recognition System2 (AIRS2) with fuzzy K-nearest neighbor. *Journal of medical systems*, 36(5), 2721-2729.
- Gabielkov, M., Ramachandran, A., Chaintreau, A., & Legout, A. (2016, June). Social clicks: What and who gets read on Twitter?. In *Proceedings of the 2016 ACM SIGMETRICS international conference on measurement and modeling of computer science* (pp. 179-192).
- Imam, Z. (2021). Strategy Analysis UKM of Government Policy in Implementing Restrictions of Community Activities (PPKM) to Handle the Covid-19 Pandemic. *Budapest International Research and Critics Institute (BIRCI-Journal): Humanities and Social Sciences*, 4(3), 6151-6160.
- Miharja, M., Salim, E., Nachrawi, G., Putranto, R. D., & Hendrawan, A. (2021). Implementation of Emergency Public Activity Restrictions (PPKM) in Accordance With Human Rights and Pancasila Principles. *Budapest International Research and Critics Institute (BIRCI-Journal): Humanities and Social Sciences*, 4(3), 6855-6866.
- Mukhtadi, M., Prayitno, A., & El Harryy, P. S. F. (2021). Management strategy of the government of Indonesia in handling the Covid-19 pandemic. *Technium Social Sciences Journal*, 22, 455-462.
- Napitu, U., Corry, & Matondang, MKD (2021). Socialization of restrictions on the implementation of micro community activities "PPKM" in the bah Kapul sub-district. *Community Development Journal*, 2(2), 232–241.
- Nayak, A. S., Kanive, A. P., Chandavekar, N., & Balasubramani, R. (2016). Survey on pre-processing techniques for text mining. *International Journal of Engineering and Computer Science*, 5(6), 16875-16879.
- Nurjanah, WE, Perdana, RS, & Fauzi, MA (2017). Sentiment Analysis of Television Shows Based on Public Opinion on Twitter Social Media using the K-Nearest Neighbor Method and Weighting the Number of Retweets. *Journal of Information Technology and Computer Science Development (J-PTIHK) Universitas Brawijaya*, 1(12), 1750–1757.
- Pratama, B., Saputra, D. D., Novianti, D., Purnamasari, E. P., Kuntoro, A. Y., Gata, W., ... & Budilaksono, S. (2019, May). Sentiment Analysis of the Indonesian Police Mobile Brigade Corps Based on Twitter Posts Using the SVM and NB Methods. In *Journal of Physics: Conference Series* (Vol. 1201, No. 1, p. 012038). IOP Publishing.
- Regulation of the Ministry of Home Affairs (Permendagri). (2020). Regulation of the Ministry of Home Affairs (Permendagri) concerning Amendments to Regulation of the Minister of Home Affairs Number 54 of 2019 concerning Funding for the Election of Governors, Regents, and Mayors Sourced from Regional Revenue and Expenditure Budgets. <https://peraturan.bpk.go.id/Home/Details/143496/permendagri-no-41-tahun-2020>
- Salim, J. N., Trisnawarman, D., & Imam, M. C. (2020, July). Twitter users opinion classification of smart farming in Indonesia. In *IOP Conference Series: Materials Science and Engineering* (Vol. 852, No. 1, p. 012165). IOP Publishing.
- Sari, R. (2020). Sentiment Analysis in the Review of Fantasy World Tourism Objects Using the K-Nearest Neighbor (K-Nn) Algorithm. *EVOLUTION: Journal of Science and Management*, 8(1), 10–17. <https://doi.org/10.31294/evolution.v8i1.7371>
- Son, J., Lee, H. K., Jin, S., & Lee, J. (2019). Content features of tweets for effective communication during disasters: A media synchronicity theory perspective. *International Journal of Information Management*, 45, 56-68.
- Toharudin, T., Pontoh, R. S., Caraka, R. E., Zahroh, S., Kendogo, P., Sijabat, N., ... & Pardamean, B. (2021). National Vaccination and Local

Intervention Impacts on COVID-19
Cases. *Sustainability*, 13(15), 8282.

Zheng, Q., Wu, Z., Cheng, X., Jiang, L., & Liu, J.
(2013). Learning to crawl deep web. *Information
Systems*, 38(6), 801-819.