Jurnal Mantik, 6 (3) (2022), ISSN 2685-4236 (Online)

IOCSCIENCE

Published by: Institute of Computer Science (IOCS)

Jurnal Mantik

Journal homepage: www.iocscience.org/ejournal/index.php/mantik

Sentiment Analysis of Non-Fungible Token (NFT) on Twitter Social Media Using Support Vector Machine Method

Suci Br. Kembaren[1], Siti Saidah[2], Arifin Nur Rahmad[3], Luky Fabrianto[4], Novianti Madhona Faizah[5]

[1,2,3]Fakultas Ilmu Komputer dan Teknologi Informasi, Program Studi Sistem Informasi, Universitas Gunadarma, Indonesia

[4] Fakultas Ekonomi dan Bisnis, Program Studi Bisnis Digital, Universitas Nusa Mandiri, Indonesia

[5] Fakultas Teknik, Program Studi Ilmu Komputer, Universitas Tama Jagakarsa, Indonesia

ARTICLE INFO

Article history:

Received Nov 03, 2022
Revised Nov 16, 2022
Accepted Nov 30, 2022

Keywords:

Non-Fungible Token
Sentiment Analysis
Social Media
Support Vector Machine
Tweets

ABSTRACT

Social media Twitter is used as an expression of opinion for the global community, especially in Indonesia. Non-Fungible Token (NFT) also has public opinion, sentiment from opinion can be classified into two classes, negative sentiment and positive sentiment. Using the keyword "NFT Indonesia" in API from Twitter, search for tweets data obtained 2204 tweets, through the manual labeling and pre-processing stages, 1462 tweets were obtained. After tweets/data has through cleansing stage, data are separated into training data and test data and then Support Vector Machine method is used to form a model that can classify positive or negative sentiment. The results of the sentiment analysis are visualized using a pie chart. The results obtained from opinion of Indonesian netizen regarding that Non-Fungible Token (NFT) have a positive trend with a percentage of 95.90% and for negative sentiment is 4.10% with an accuracy of success in this sentiment analysis is 86%.

Corresponding Author:

Luky Fabrianto,
Economy and Business Faculty, Digital Business Study Program,
Universitas Nusa Mandiri,
Jl. Raya Jatiwaringin No.2, RT.8/RW.13, Cipinang Melayu, Kec. Makasar, Kota Jakarta Timur, Daerah Khusus Ibukota Jakarta 13620
Email: luky.lfb@nusamandiri.ac.id

1. INTRODUCTION

Growth of digital era now makes the creative industry one of the fields of work that provides benefits and opportunities for creative workers. Non-Fungible Token or commonly known as NFT is one of media that can be utilized by creative industry players. NFT itself means a token that cannot be exchanged which is issued in the form of a crypto token, which could be accessed and traded on blockchain. The difference with other crypto assets is that NFT has no exchange rate, the ownership of NFT can only be owned by one user (Tengland & Ask, 2022).

NFT became a trending topic in early 2022, when Ghozali peaked due to uploading photos to NFT marketplace with various appearances but in same facial expressions. At that time, Twitter netizen was busy discussing NFTs because a tweet from Ghozali that conveyed confusion about his NFT of more than 230 photos could be sold.

Twitter data could be freely accessed using Twitter API (Dwiarni & Setiyono, 2020). Utilization Twitter API, tweets will be processed into data that searched according to keywords. NFT is still a topic of discussion regarding its presence in Indonesia which has a positive or negative effect, therefore NFT sentiment analysis can be performed using a hyperplane which could separate data from two different classes using Support Vector Machine method (Wu & Zhou, 2006).

Related studies, (Qian et al., 2022) research using PPMCC with 8 emoticon scales (Anger, Anticipation, Disgust, Fear, Joy, Sadness, Surprise, and Trust) found Tweets majorly contained positive sentiment 72%. (Anreaja et al., 2022) compared the results of the accuracy of Naïve Bayes method with Support Vector Machine, it was found that Support Vector Machine had higher accuracy. From a social network perspective, the most popular collection of NFTs can be thought of as a single community around NFTs (Casale-Brunet et al., 2022). The results of Indonesian public opinion on metaverse technology show 66% neutral, 17% negative and 16% positive, results of test using Support Vector Machine algorithm show that Support Vector Machine performance results are 87% better than using tree algorithm which has a performance of 71% (Ahmad & Gata, 2022) and Sentiment Analysis Regarding Non-Fungible Tokens using Multinomial Naïve Bayes, obtaining an accuracy of 84% (Lesmana, 2022).

The initial stages of this research are data acquisition, manual labeling, pre-processing (Yang, 2013)includes: data cleaning, data split and data weighting (Herwijayanti et al., 2018). Followed by process of training the data to form a sentiment analysis classification model using Support Vector Machine method, after the model is formed, the test data is applied to the model and evaluated, final stage is visualization in pie chart.

2. RESEARCH METHODOLOGY

The first stage is data acquisition or data retrieval by accessing Twitter Developer website link https://developer.twitter.com/, of course an API key, API secret key, access token and access token key from Twitter are needed to connect with the tools used, in this research tools used is Python programming language.

Furthermore, pre-processing is carried out which consists of several steps, including: manual labeling, data cleaning, data splitting and data weighting. First step is manual labeling, that if a sentence is considered positive, then it is labeled = Positive, and if a sentence is considered negative, then it is labeled = Negative. Next step is data cleaning to clean tweets data from unnecessary and find potential words that affect sentiment analysis process. Stages of this process include case folding, cleansing, removing duplicates, stop word removal, word steaming, manual cleaning, and tokenizing. Output of this process is clean data. By using clean tweets, next step is split the data into two, training data and test data (Vrigazova, 2021). The training data is used to train Support Vector Machine algorithm in performing sentiment analysis, and the test data is used to test performance of Support Vector Machine method classifies a tweet.

After data cleaning and data splitting steps have been carried out, followed by word weighting, the aim of this step is to calculating the weight of each word in a tweet using TF-IDF (Term Frequency – Inverse Document Frequency) method. The purpose of TF-IDF is to find the number of words known to TF (Term Frequency) multiplied by number of tweets in which a word appears IDF (Inverse Document Frequency) (Herwijayanti et al., 2018). TF-IDF method is calculating weight by integration between Term Frequency (TF) and Inverse Document Frequency (IDF). TF-IDF will produce a value *(w)* for a word in the document at Corpus. The TF-IDF equation is as follows:

$$w_{i,j} = TF_{i,j} x log \left(\frac{n}{DFi} \right) \quad (1)$$

$w_{i,j}$ is weights of word i in j^{th} tweet/document, n is total number of tweets, TF is number of emergence of the word i in tweet j, and IDF is number of tweets j that contain word of i. purpose of TF-IDF method is to make data could be analyzed using a Support Vector Machine.

After the tweets have through the pre-processing stage, the tweets or data are processed with training the data using a Support Vector Machine to produce the best hyperplane to separate two different classes, Support Vector Machine model formed then tested using test data. Figure 1. below is an illustration of Support Vector Machine method.

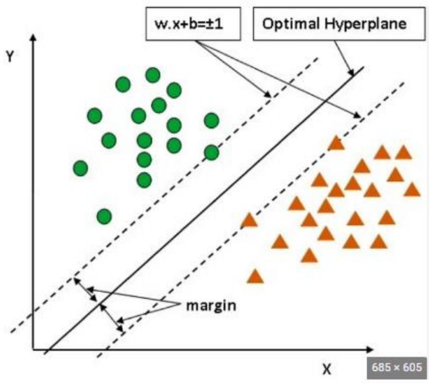

Figure 1. Illustration of Support Vector Machine method (Pathak et al., 2021)

If a case contains data with two classes that are linearly separated, then equation for find hyperplane could be written as:

$$w \,.\, x + b = 0 \quad (2)$$

w is weight, x is input variable, and b is bias value. x is normal weight of the hyperplane and $\frac{b}{w}$ is distance from hyperplane to point 0. Data that is included in class A and has label 1 fulfills this equation:

$$w \,.\, x + b \geq 1 \quad (3)$$

Data that is included in class B and has label -1 fulfills this equation:

$$w \,.\, x + b \leq -1 \quad (4)$$

Furthermore, after tweets through stages of acquisition, pre-processing and process, an evaluation is carried out, namely by utilizing the test data tested on the model from Support Vector Machine method formed, evaluation measurement of the predictions found is value of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) (Luque et al., 2019). After these values are obtained, accuracy, precision, recall and F-measure calculations could be performed (Chicco et al., 2021).

The final stage of this research is visualization of results to display the results of sentiment analysis classification using a pie chart (Thanuj Kumar et al., 2021) to describe the conclusions of results of positive and negative sentiments.

3. RESULT AND DISCUSSION

Acquired data consist of 2204 tweets from time span from 11 January 2022 to 31 May 2022, at pre-processing stage there were only 1462 tweets left which were used in this study, data separated by composition of 70% training data and 30% test data, there were 1023 tweets of training data and 439 tweets of testing data.

Furthermore, word weighting is carried out, Table 1. below is an example of Training Data document and Table 2. is an example of Test Data document.

Tabel 1. Example of Training Data document

Kalimat	Dokumen
['ghozali', 'hasil', 'kenal', 'nft', 'indonesia']	D1
['balpil', 'jadi', 'serial', 'animasi', 'indonesia', 'pertama', 'hadir', 'nft', 'keren']	D2
['platform', 'nft', 'opensea', 'akses', 'gara', 'gara', 'ghozali', 'indonesia']	D3

Tabel 2. Con Example of Testing Data document

Kalimat	Dokumen
['mantap', 'keren', 'keren', 'hasil', 'nft', 'indonesia']	?

Results of TF-IDF word weighting of training data using equation *(1)* is describe in Table 3. below.

Tabel 3. Word weighting of training data

Term/word	*W = TF x IDF*		
	D1	D2	D3
akses	0	0	0,477
animasi	0	0,477	0
balpil	0	0,477	0
gara	0	0	0,954
ghozali	0,176	0	0,176
hadir	0	0,477	0
hasil	0,477	0	0
indonesia	0	0	0
jadi	0	0,477	0
kenal	0,477	0	0
keren	0	0,477	0
nft	0	0	0
opensea	0	0	0,477
pertama	0	0,477	0
platform	0	0	0,477
serial	0	0,477	0

The last table is result of weight calculating of each word in training data, if the less word appears in entire document, so, weight of the word is higher. To calculate weights for the test data, steps are same as for training data, but IDF values used are derived from IDF values for training data.

After that, data training is carried out, process is obtaining a classification model using training data with Support Vector Machine method, cleaned data and weighted data will be calculated to find the best hyperplane that could separate two different classes. To find hyperplane (b) is need vector (a) value and vector weight (w). Figure 2. is a block diagram of data training process using Support Vector Machine.

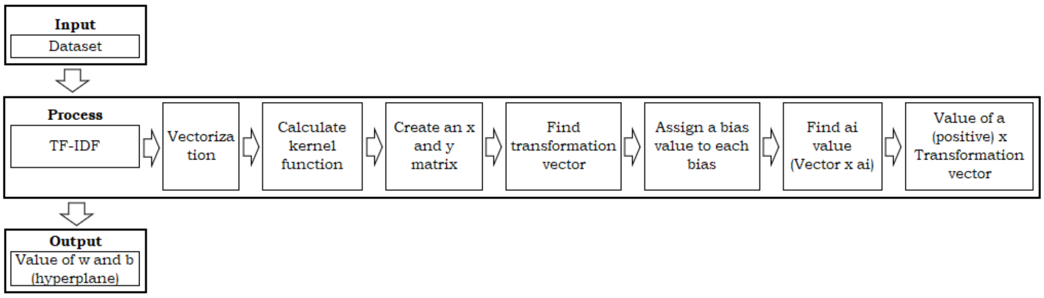

Figure 2. Data training process Block Diagram

Based on calculations from training data using kernel formula, results are obtained with equations *(2)*, *(3)*, and *(4)* and then *w* and *b* values can be obtained with the following calculations:

$$w = 0{,}528 \times \begin{bmatrix} 1{,}596 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0{,}843 \\ 0{,}528 \\ 0{,}528 \end{bmatrix}$$

From results above could be concluded that:

$$w = \begin{bmatrix} 0{,}843 \\ 0{,}528 \end{bmatrix} \text{ dan } b = 0{,}528$$

The *b* value is hyperplane used to classify positive and negative sentiment classes, hyperplane of training data is 0,528. Furthermore, after vector weight values *(w)* and hyperplane *(b)* from the classification model are obtained, testing is carried out. Figure 3. is a block diagram of data testing.

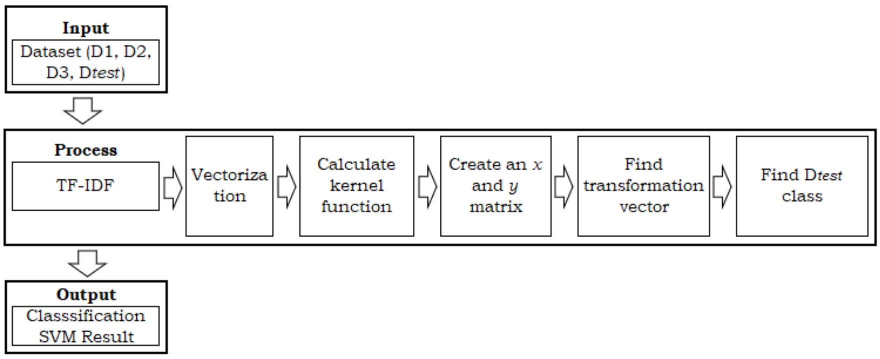

Figure 3. Block diagram of data testing.

After calculating kernel functions for $x_j^T x_j$and $y_j^T y_j$, next step is obtaining x and y values by adding up matrix values for each row in $x_j^T x_j$and $y_j^T y_j$ matrix.

$$y(D_{uji}) = w^T \times \emptyset\,(D_{uji})$$

$$y(D_{uji}) = \begin{bmatrix} 0{,}843 \\ 0{,}528 \end{bmatrix} \times \begin{bmatrix} 1{,}821 \\ 0 \end{bmatrix}$$

$$y(D_{uji}) = 1{,}535$$

The results of test data are 1.535. It could be concluded that test data has y value greater than hyperplane *(b)*, this test data belongs to positive class. The test data used in testing process using a classification model produces 422 positive tweet predictions and 17 negative tweet predictions. Table 4. is result of confusion metric obtained.

Table 4. Result of TP, TN, FP, dan FN

		Label	
		TRUE	FALSE
Prediction	TRUE (Positif)	361	60
	FALSE (Negatif)	17	1

Table 5. is result of accuracy, precision, recall, F-measure obtained based on calculations from values in table 4.

Tabel 5. Result of Accuracy, Precision, Recall and F- measure

	Positif	Negatif
Accuracy	0,86	
Precision	0,86	0,94
Recall	1,00	0,22
F-measure	0,92	0,36

The results of classification of tweets obtained from test data show that positive sentiment tends to have a greater percentage than percentage of negative sentiment, in breakdown of 95.90% for positive sentiment with test data classified as 421 tweets and 4.10% for negative sentiment and test data classified as 18 tweets. The accuracy in this sentiment analysis research is 86%. Visualization is final step of this sentiment classification research, served using pie chart as shown in Figure 4 below.

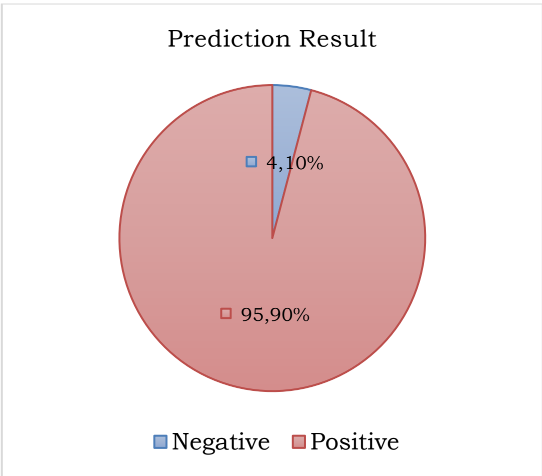

Figure 4. Pie Chart of prediction result

4. CONCLUSION

Based on results of this sentiment analysis research, it could be concluded that Indonesian netizen sentiment towards the presence of Non-Fungible Tokens on Twitter social media mostly gave a positive response. According to this research, this positive result is caused by many public opinions agreeing with benefits of Non-Fungible Token presence.

This sentiment analysis research, of course still has drawbacks, therefore author's suggestion for next development is to acquisition more tweets data for classification, on labeling process users are better able to understand good and correct Indonesian vocabulary, then on labeling step could be automatically, in data cleaning need more attention and research is carried out with other methods to compare from other methods besides Support Vector Machine.

REFERENCES

Ahmad, A., & Gata, W. (2022). Sentimen Analisis Masyarakat Indonesia di Twitter Terkait Metaverse dengan Algoritma Support Vector Machine. *Jurnal JTIK (Jurnal Teknologi Informasi Dan Komunikasi)*, *6*(4), 548–555. https://doi.org/10.35870/JTIK.V6I4.569

Anreaja, L. J., Harefa, N. N., Galih, J., Negara, P., Nathan, V., Pribyantara, H., & Prasetyo, A. B. (2022). Naive Bayes and Support Vector Machine Algorithm for Sentiment Analysis Opensea Mobile Application Users in Indonesia. *JISA(Jurnal Informatika Dan Sains)*, *5*(1), 62–68. https://doi.org/10.31326/JISA.V5I1.1267

Casale-Brunet, S., Zichichi, M., Hutchinson, L., Mattavelli, M., & Ferretti, S. (2022). The impact of NFT profile pictures within social network communities. In *ACM International Conference Proceeding Series* (Vol. 1, Issue 1). Association for Computing Machinery. https://doi.org/10.1145/3524458.3547230

Chicco, D., Tötsch, N., & Jurman, G. (2021). The matthews correlation coefficient (Mcc) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData Mining*, *14*, 1–22. https://doi.org/10.1186/s13040-021-00244-z

Dwiarni, B. A., & Setiyono, B. (2020). Akuisisi dan Clustering Data Sosial Media Menggunakan Algoritma K-Means sebagai Dasar untuk Mengetahui Profil Pengguna. *Jurnal Sains Dan Seni ITS*, *8*(2). https://doi.org/10.12962/j23373520.v8i2.49815

Herwijayanti, B., Ratnawati, D. E., & Muflikhah, L. (2018). Klasifikasi Berita Online dengan menggunakan Pembobotan TF-IDF dan Cosine Similarity. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *2*(1), 306–312. http://j-ptiik.ub.ac.id

Lesmana, R. Y. (2022). Jurnal Teknik Informatika, Vol. 14, No. 3, Agustus 2022. *Jurnal Teknik Informatika*, *14*(3), 135–139.

Luque, A., Carrasco, A., Martín, A., & de las Heras, A. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition*, *91*, 216–231. https://doi.org/10.1016/j.patcog.2019.02.023

Pathak, D. K., Kalita, S. K., & Bhattacharya, D. K. (2021). Hyperspectral image classification using Support Vector Machine: a spectral spatial feature based approach. *Evolutionary Intelligence 2021 15:3*, *15*(3), 1809–1823. https://doi.org/10.1007/S12065-021-00591-0

Qian, C., Mathur, N., Zakaria, N. H., Arora, R., Gupta, V., & Ali, M. (2022). Understanding public opinions on social media for financial sentiment analysis using AI-based techniques. *Information Processing & Management*, *59*(6), 103098. https://doi.org/10.1016/J.IPM.2022.103098

Tengland, A., & Ask, M. (2022). *Predicting NFT Marketplace Growth Using Frequency of Tweets Regarding Safety Concerns*. http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-320199

Thanuj Kumar, S., Vinitha Dominic, K., & Sumathi, V. (2021). Data mining and data visualization for analysing the rate of bed availability at hospitals due to COVID 19. *Archive of Biomedical Science and Engineering*, 001–004. https://doi.org/10.17352/ABSE.000023

Vrigazova, B. (2021). The Proportion for Splitting Data into Training and Test Set for the Bootstrap in Classification Problems. *Business Systems Research*, *12*(1), 228–242. https://doi.org/10.2478/bsrj-2021-0015

Wu, Q., & Zhou, D. X. (2006). Analysis of Support Vector Machine classification. *Journal of Computational Analysis and Applications*, *8*(2), 99–119.

Yang, H. (2013). Data Preprocessing-Chapter 3. *Citeseerx*. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.467.29&rep=rep1&type=pdf