# Recommendation System with Content-Based Filtering in NFT Marketplace

Edi Surya Negara [1],*, Sulaiman [1], Ria Andryani [1], Prihambodo Hendro Saksono [1], and Yeni Widyanti [2]

[1] Data Science Interdisciplinary Research Center, Computer Science Faculty, Universitas Bina Darma, Palembang, Indonesia; Email: sulaiman@student.binadarma.ac.id (S.), ria.andryani@binadarma.ac.id (R.A.), p.h.saksono@binadarma.ac.id (P.H.S.)
[2] Economics and Business Faculty, Universitas Bina Darma, Palembang, Indonesia; Email: yeniwidyanti@binadarma.ac.id (Y.W.)
*Correspondence: e.s.negara@binadarma.ac.id (E.S.N.)

*Abstract*—**Non-Fungible Token (NFT) is a digital asset that cannot be exchanged or used, and uses Crypto currency values according to the type of digital money used, for example Bitcoin, Ethereum. The NFT Marketplace is a platform for buying and selling NFT like Tokopedia. This common problem is often encountered in e-commerce, especially in the NFT Marketplace, among other buyers often having difficulty finding products. This makes it difficult for the NFT Marketplace and sellers to promote products that match the preferences of potential buyers. A recommendation system that is very much needed in overcoming these problems, responding to these problems the author tries to make a recommendation system using the Content Based Filtering approach using the cosine similarity. The results of this study indicate that the Machine Learning model can provide Top-N recommendations from the product being sought.**

*Keywords*—**crypto currency, non-fungible token, content-based filtering, Non-Fungible Token (NFT) marketplace**

## I. INTRODUCTION

The times have made new information and technology seem unlimited and the flow is so fast like the development of various blockchain technologies such as cryptocurrencies and smart contracts [1]. This fairly rapid development process is also supported by the development of various algorithms, data processing techniques, and various highly sophisticated computing technologies [2, 3]. Developments in Cryptocurrencies today are promising, citing the publication of Commodity Futures Trading Regulatory Agency (BAPPEBTI), it is stated that the Crypto tax in Indonesia is planned to be at a rate of 0.05% and this tax rate is lower than stocks which are subject to 0.1%, this makes business activists in the field of competition. Digital competition creates a market specifically for NFT [4]. The increase in NFT transactions in Indonesia is an opportunity to boost state revenues which will enforce tax regulations [5].

NFT is a digital asset such as music, in-game items, paintings, and videos stored in smart contracts. NFT is unique which means that there are no duplicates in each NFT ownership. Since the pandemic in Indonesia, the development of NFT has increased quite rapidly, this can be seen from the increasing number of local marketplaces that release transaction services, for example, Tokomal. This happens because there are many local Indonesian creators who are competing to make extraordinary works of art. The beginning of NFT was scrutinized by the broader community after someone named Ghozali managed to sell his digital works at a fantastic price, namely at a price of USD 1 million.

NFT is a representation of digital assets that cannot be exchanged or exchanged on a par with other NFTs or of the same type, and the concept of NFT is digital authenticity that cannot be replicated [6]. Marketing and business strategies to compete in the NFT The marketplace includes a recommendation system. Recommendation systems have been present and are widely used by almost all business fields where the public or consumers need information as advice in decision-making [7, 8]. Use a recommendation system will have a profit, because the more precise the system in recommending NFT can make prospective buyers more comfortable in choosing and increase the possibility to buy more diverse products [9]. Piyadigama *et al.* also conducted research showing that the use of a recommendation system for NFT products can increase the number of sales. So that there is an increase and progress that encourages the development of the NFT recommendation system [10]. The recommendation system is useful for filtering abundant data into important and useful information for the company.

This research discusses how to create a recommendation system model with a content-based filtering approach, content-based filtering approach works by suggesting similar items based on the user's past activity or being viewed in the present to the user. The more information the user provides, the better the recommendation system's accuracy.

## II. LITERATURE REVIEW

### A. Non-Fungible Token (NFT)

NFT is a practice that makes value in a digital artwork the only or pure one, the value in the artwork cannot be duplicated so the value will be very unique [11]. NFT itself is a digital file in the form of a token and non-exchangeable which mostly uses the Ethereum blockchain technology to identify ownership of digital assets which can be music, videos, pictures, collections, or other digital files such as equipment or in-game characters [12].

NFT can be valued at very expensive because it has its unique and historical value, apart from that, because the artistic taste of the owner. Ownership of an NFT is evidenced by an immutable and cryptographically secured record on the blockchain that is meant to be taken as proof by others in the cryptosp here that someone is the owner of the underlying asset, which is similar to a digital certificate of title or stamp of authenticity [3]. The method of communication on a decentralized and distributed blockchain, where each block contains a cryptographic hash to form a network.

NFT stored in the blockchain, every transaction occurring on the NFT will be recorded in a Smart contract which will issue a unique code and store it. Smart contracts themselves can simply be called the history of transactions [1], smart contracts aim to make the transaction process easier, more flexible, and more efficient. With the Smart contract, it is possible to carry out credible transactions without a third party [14]. Smart contracts are created using a series of programming codes, where there are parties who enter into contracts automatically in the blockchain system.

Lennart Ante researched the Non-Fungible Token (NFT) Market and Its Relationship with Bitcoin and Ethereum [15]. This research obtained results where the price of bitcoin triggers an increase in NFT sales and the value of cryptocurrency in the market affects growth in the NFT market but there is no opposite effect between NFT and bitcoin and cryptocurrency. Borri *et al.* conducted research on The Economics of Non-Fungible Tokens [16]. This research obtained results where NFT has low exposure but significantly predicts returns on the NFT market.

The main reason the author doesn't use the collaborative filtering approach in collaborative filtering requires user likes or rating data, while in this research the data has an imbalance where a lot of data do not have likes or ratings compared to the other way around. If use this approach, there will be inaccuracies in the model in the prediction process.

### B. Recommendation System

Recommendation system is a software tool and technique that provide suggestions for items that are most likely to be of interest to a particular user [17, 18]. The recommendation system predicts a user's rating or preference for a particular item. These recommendations are made based on past user behavior or other user behavior [19]. So, the system will recommend something to the user based on behavioral data or preferences over time.

The recommendation system does not recommend specific items but recommends a number of items that may match the user's preferences, the output of the recommendation the system is a Top-N recommendation. The main purpose of the recommendation system is to increase product sales as well as sell various items (see Fig. 1).



Figure 1.  Recommendation system.

### C. Content Based Filtering

Content-Based Filtering uses the availability of content (often also referred to as features, attributes, or characteristics) of an item as a basis for providing recommendations [20, 21]. The main reason for making a recommendation system using this approach is that there is little data available and the resulting recommendations make users gain an understanding of why an item is considered relevant to them. This method will work by sorting items by top similarity [22]. While the drawback of this approach is that the accuracy of the model depends on the keywords entered [23] (see Fig. 2).
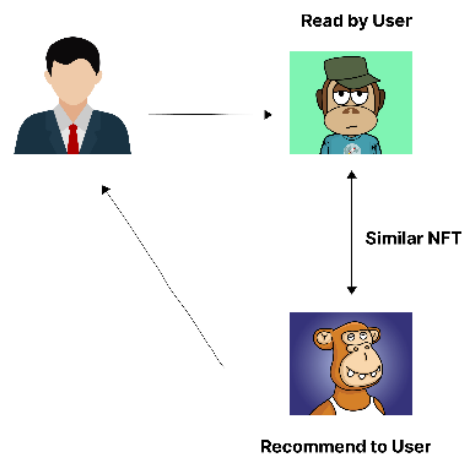


Figure 2.  Content-based filtering.

## III. METHODOLOGY

This research uses experimental research methods, where the experimental method is carried out intentionally by researchers by giving certain treatments to research subjects in order to evoke an event or condition that will be studied and the consequences [24]. This research is divided into several stages including data collection, data preprocessing, proposed model, testing and evaluation. The research stage can be seen in Fig. 3.
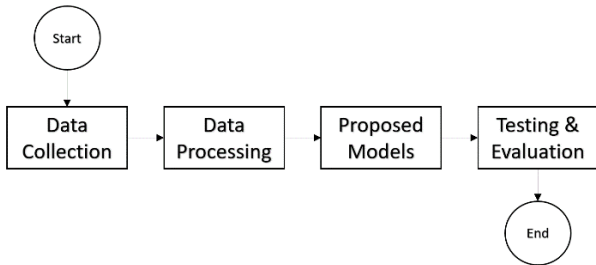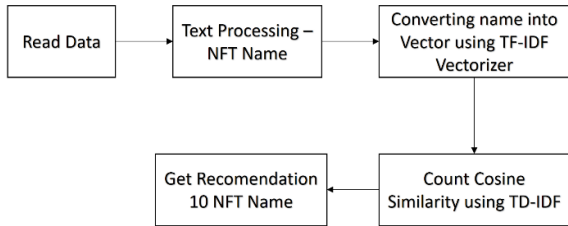


Figure 3. Research stage.



Figure 4. Count cosine similarity using the output of TF-IDF.

Fig. 4 is the process or method carried out in obtaining recommendations, first converting the data into vector, generated cosine similarity, and then getting recommendations.

## IV. EXPERIMENTAL AND RESULT

This research uses NFT art collection 2021 data sourced from Kaggle.com [25]. In this dataset there are 4 data formats, including; gif, image, video, and CSV. The author does not use the entire dataset, but only uses tabular data in CSV format where in the dataset there are 4189 rows and 15 columns, example datasets are shown in Fig. 5. In the preprocessing stage, several steps are carried out, converting the data series into list form, creating a dictionary to determine key-value pairs from nft_name and nft creator, and transforming to vector using TF-IDF (see Fig. 1).

| title | Name | creator | art_series | price | sysmbol | type | likes | nstm | tokens |
|---|---|---|---|---|---|---|---|---|---|
| 38 min Drawings | Giant Frog | kristyglas | Kristyglas_38-windrawings_giant-frog | 50 | SWAP_HIVE | PHOTO | 2 | False | 38 |
| Experimental Video | Biospeciemens | juliakpansford | Juliakpansprod_experimental-video_biospecimens | 500 | SWAP_HIVE | VIDEO | 0 | False | 1 |
| Sexy Art | long legs | badsexy | Badsexy_sexy-art_long-legs | 10 | SWAP_HIVE | PHOTO | 0 | True | 2 |
| Dream World | A Guide in my Dreams | yoslehz | Yoslehz_dream-world_a-guide-in-my-dreams | 20 | SWAP_HIVE | PHOTO | 1 | False | 2 |
| Dream World | Silent Observer | yoslehz | Yoslehz_dream-world_silent-observer | 20 | SWAP_HIVE | GIF | 0 | False | 2 |
| Dream World | Joy and Chaos | yoslehz | Yoslehz_dream-world_joy-and-chaos | 20 | SWAP_HIVE | GIF | 0 | False | 2 |

Figure 5. Sample dataset.

Next stage is the process carried out from the EDA (Exploratory Data Analysis) process.

Fig. 6 shows of likes on NFT is lots of blanks compared to an NFT that gets likes, with the above visualization deciding to use a content-based filtering method instead of a collaborative filtering method. Because the collaborative filtering method requires rating data from users to get recommendations (see Fig. 7).
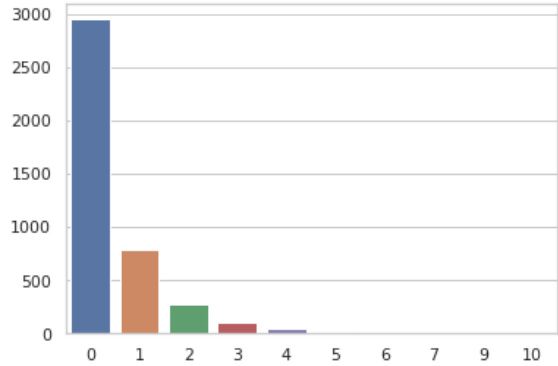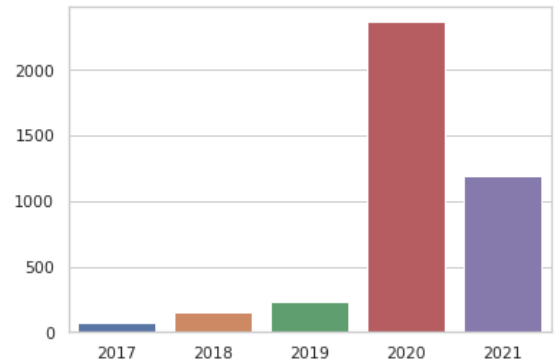


Figure 6. Likes on NFT.



Figure 7. NFT frequency.

The next stage is conversion the dataset using TF-IDF where TF-IDF is a representation scheme commonly used in information retrieval and document extraction systems that are relevant to certain queries. The technique is to find the representation of the important features of each NFT name category using the TF-IDF vectorizer() function from the Sklearn library. The TF-IDF technique in this research is a process in feature engineering to find text representations and convert data to vector form which is carried out before the modeling stage.

TF-IDF is defined by two quantities TF and IDF, where TF (Term Frequency) works by measuring how often a word or term appears in a particular text and normalizes by dividing the number of occurrences by the length of the document. The TF calculation formula can be seen in Eq. (1). Whereas IDF (Inverse Document Frequency) works by measuring the importance of terms across the corpus, IDF considers terms that are very common throughout the document and weighs rare terms. The IDF calculation formula can be seen in Eq. (2), and the last step is to multiply the TF value with IDF to get the TF-IDF score Eq. (3).

$$TF(X, d) = \frac{appereance\ term\ X}{amount\ of\ terms\ in\ document} \qquad (1)$$

$$IDF(X) = log\frac{amount\ of\ text\ in\ document}{amount\ of\ text\ that\ has\ term\ x} \qquad (2)$$

$$weight(X, d) = TF(X, d) \times IDF(X) \qquad (3)$$

The results obtained in vectorization can be seen in the Fig. 8 below.

```
matrix([[0., 0., 0., …, 0., 0., 0.,],
        [0., 0., 0., …, 0., 0., 0.,],
        [0., 0., 0., …, 0., 0., 0.,],
        . . . ,
        [0., 0., 0., …, 0., 0., 0.,],
        [0., 0., 0., …, 0., 0., 0.,],
        [0., 0., 0., …, 0., 0., 0. ]])
```

Figure 8.　Vectorization results using TF-IDF.

The next stage is modeling using the process of calculating the degree of similarity between NFTs with cosine similarity, cosine similarity can find similarities between one nft_name and another. The results obtained in the process of calculating the degree of similarity using TF-IDF can be seen in Fig. 9.

```
matrix([[1., 0., 0., …, 0., 0., 0.,],
        [0., 1., 0., …, 0., 0., 0.,],
        [0., 0., 1., …, 0., 0., 0.,],
        . . . ,
        [0., 0., 0., …, 1., 0., 0.,],
        [0., 0., 0., …, 0., 1., 0.,],
        [0., 0., 0., …, 0., 0., 1. ]])
```

Figure 9.　Vectorization results using TF-IDF.

Next step is to create the name recommendations' function (see Fig. 10).

```
def name_recommendations(nft_name, similarity_data=cosine_sim_df, ite
ms=data[['nft_name', 'nft_creator']], k=10):
    index = similarity_data.loc[:,nft_name].to_numpy().argpartition(
        range(-1, -k, -1))

    closest = similarity_data.columns[index[-1:-(k+2):-1]]

    closest = closest.drop(nft_name, errors='ignore')

    return pd.DataFrame(closest).merge(items).head(k)
```

Figure 10. Name_recommendation function.

The name recommendation function has the meaning of taking the highest number of k values from the similarity data (in this case: cosine sim DF data frame). Then, take the data from the highest to the lowest weight (level of similarity). This data is entered into the closest variable. Next, it is necessary to remove the NFT name that is being searched for so that it does not appear in the list of recommendations.

The recommendation system made using a content-based filtering approach produces a very satisfactory recommendation, where the results of recommendation show that the name of the NFT and the inputted words have similarities in 10 output names in the recommendation system. Testing is done by calling the name recommendation function shown in Fig. 11.

The main limitation of this research may occur when using only NFT item names for recommendations. Using only names for recommendations will not work when the main NFT collection (For example: Bored Ape Yacht Club) is used in most cases, as most have the same name as the number.



```
[ ]  # get recomendations
     name_recommendations('Love')
```

| | nft_name | nft_creator |
|---|---|---|
| 0 | Love night ❤ | andrianna |
| 1 | dragon love | pollyfoxxy |
| 2 | A Love Eternal | andrewmusic |
| 3 | endless love | elgeko |
| 4 | Animal Love | desro |
| 5 | power of love | noahfromearth |
| 6 | Rasta Love | evacortez |
| 7 | ASIA LOVE | forlenza |
| 8 | Love D Project | jeanravel |
| 9 | Love and Peace | eve66 |

Figure 11. Result recommendation system using a content-based filtering.

## V. Conclusion

Based on several scenarios starting from the analysis to the testing stage, the following conclusions can be drawn, the method proposed this research obtains results that can recommend NFT names appropriately, and the approach method used this research is suitable to be applied to the recent NFT marketplace or new NFT marketplace, when compared collaborative filtering method that requires rating data from user. Recommendation system is made based on the level of similarity. Common similarities can be measured in the recommendation system, preferences and tastes. In addition, similarities can also be found through other data and information such as user demographics and social status. Suggestions from the author for further research to apply Exploratory Data Analysis (EDA) in more detail, as well as the use of larger datasets that can be obtained from the results of crawling data from the NFT marketplace.

### CONFLICT OF INTEREST

The authors declare no conflict of interest.

### AUTHOR CONTRIBUTIONS

Edi Surya Negara was the research lead carried out research design development and supervised all research development and lead the writing of manuscripts. Sulaiman wrote program code to test the research model

REFERENCES

[1] E. S. Negara, A. N. Hidayanto, R. Andryani, *et al.*, "Survey of smart contract framework and its application," *Information*, vol. 12, no. 7, 2021.

[2] N. Nurhachita and E. S. Negara, "A comparison between deep learning, naïve bayes and random forest for the application of data mining on the admission of new students," *IAES International Journal of Artificial Intelligence*, vol. 10, no. 2, 2021.

[3] E. S. Negara, K. Keni, and R. Andryani, "Data prediction for coffee harvest using least square method," in *Proc. IOP Conference Series: Materials Science and Engineering*, IOP Publishing, 2020, vol. 1007, no. 1, 012193.

[4] Suryani Suyanto & Associates. (Apr. 2021). Bappebti explained about the tax imposition plan on crypto. [Online]. Available: https://www.ssas.co.id/bappebti-jelaskan-soal-rencana-pengenaan-pajak-pada-kripto/ (in Indonesia)

[5] D. Liana, "Tracing the potential of NFT as a source of tax revenue," Center for Budget Studies, Agency for Expertise *DPR RI*, 2022.

[6] A. Popescu, "Non-Fungible Tokens (NFT)-innovation beyond the craze," in *Proc. 5th International Conference on Innovation in Business, Economics and Marketing Research*, 2021.

[7] M. Irfan, A. D. Cahyani, *et al.*, "Recommendation system: Online book with collaborative filtering method," *Jurnal Teknologi Technoscientia*, pp. 76–84, 2014.

[8] F. Ferdiansyah, E. S. Negara, and Y. Widyanti, "Bitcoin-USD trading using SVM to detect the current day's trend in the market," *Journal of Information Systems and Informatics*, vol. 1, no. 1, pp. 70–76, 2019.

[9] W. E. Pangesti, R. Suryadithia, M. Faisal, *et al.*, "Collaborative filtering based recommender systems for marketplace applications," *International Journal of Educational Research & Social Sciences*, vol. 2, no. 5, pp. 1201–1209, 2021.

[10] D. Piyadigama and G. Poravi, "An analysis of the features considerable for NFT recommendations," arXiv preprint, arXiv:2205.00456, 2022.

[11] A. S. Ginting and S. Sadono, "Alternative exhibition space in the form of a non-fungible token art market platform," in *Proc. Art & Design*, vol. 9, no. 2, 2022.

[12] R. M. Sukmariningsih, A. Nurudin, and E. Nursanty, "Imposition of tax laws on cryptocurrencies and NFTs in Indonesia," *Owner: Riset dan Jurnal Akuntansi*, vol. 6, no. 2, pp. 1644–1654, 2022.

[13] K. B. Muthe, K. Sharma, and K. E. N. Sri, "A blockchain based decentralized computing and NFT infrastructure for game networks," in *Proc. Second International Conference on Blockchain Computing and Applications (BCCA)*, 2020, pp. 73–77.

[14] E. Negara, A. Hidyanto, R. Andryani, and D. Erlansyah, "A survey blockchain and smart contract technology in government agencies," in *Proc. IOP Conference Series: Materials Science and Engineering*, vol. 1071, no. 1, 012026, 2021.

[15] L. Ante, "The non-fungible token (NFT) market and its relationship with bitcoin and Ethereum," *FinTech*, vol. 1, no. 3, pp. 216–224, 2022.

[16] N. Borri, Y. Liu, and A. Tsyvinski, "The economics of non-fungible tokens," *SSRN*, 2022, doi: 10.2139/ssrn.4052045

[17] M. Braunhofer, M. Elahi, and F. Ricci, "User personality and the new user problem in a context-aware point of interest recommender system," in *Information and Communication Technologies in Tourism*, Springer, 2015, pp. 537–549.

[18] S. Sutjiningtyas, A. A. Dharmawan, *et al.*, "Design and build a shoe product recommendation system for an online store using the user-base collaborative filtering method," *Bulletin of Information Technology (BIT)*, vol. 3, no. 2, pp. 143–148, 2022.

[19] E. S. Negara, D. Triadi, and R. Andryani, "Topic modelling twitter data with latent Dirichlet allocation method," in *Proc. International Conference on Electrical Engineering and Computer Science (ICECOS)*, 2019, pp. 386–390.

[20] B. Tessy, S. Azvy, W. Yuwono, and I. Syarif. "Recommendation system for property search using content-based filtering method," in *Proc. 2018 International Conference on Information and Communications Technology (ICOIACT)*, pp. 25–29, 2018.

[21] A. Wijaya and D. Alfian, "The laptop recommendation system uses collaborative filtering and content-based filtering," *Jurnal Computech and Bisnis*, vol. 12, no. 1, pp. 11–27, 2018. (in Indonesian)

[22] S. Suwarno and T. Fernando, "Design a design service marketplace using the content-based filtering method," *Prosiding Seminar Nasional Ilmu Sosial dan Teknologi (SNISTEK)*, no. 4, pp. 324–330, 2022. (in Indonesian)

[23] B. Mobasher, H. Dai, T. Luo, *et al.*, "Improving the effectiveness of collaborative filtering on anonymous web usage data," in *Proc. the IJCAI 2001 Workshop on Intelligent Techniques for Web Personalization (ITWP01)*, 2001, pp. 53–61.

[24] A. Jaedun, "Experimental research methodology," Faculty of Engineering UNY, vol. 12, 2011. (in Indonesia)

[25] A. D. Haan. (2021). NFT-art-dataset. [Online]. Available: https://www.kaggle.com/datasets/vepnar/nft-art-dataset /