Computers, Materials & Continua

Tech Science Press

DOI: 10.32604/cmc.2023.037553
Article

Prediction of NFT Sale Price Fluctuations on OpenSea Using Machine Learning Approaches

Zixiong Wang, Qiuying Chen and Sang-Joon Lee*

Interdisciplinary Program of Digital Future Convergence Service, Chonnam National University, Gwangju, 61186, Korea
*Corresponding Author: Sang-Joon Lee. Email: s-lee@chonnam.ac.kr
Received: 08 November 2022; Accepted: 29 December 2022

Abstract: The rapid expansion of the non-fungible token (NFT) market has attracted many investors. However, studies on the NFT price fluctuations have been relatively limited. To date, the machine learning approach has not been used to demonstrate a specific error in NFT sale price fluctuation prediction. The aim of this study was to develop a prediction model for NFT price fluctuations using the NFT trading information obtained from OpenSea, the world's largest NFT marketplace. We used Python programs to collect data and summarized them as: NFT information, collection information, and related account information. AdaBoost and Random Forest (RF) algorithms were employed to predict the sale price and price fluctuation of NFTs using regression and classification models, respectively. We found that the NFT related account information, especially the number of favorites and activity status of creators, confer a good predictive power to both the models. AdaBoost in the regression model had more accurate predictions, the root mean square error (RMSE) in predicting NFT sale price was 0.047. In predicting NFT sale price fluctuations, RF performed better, which the area under the curve (AUC) reached 0.956. We suggest that investors should pay more attention to the information of NFT creators. We anticipate that these prediction models will reduce the number of investment failures for the investors.

Keywords: NFT; sale price fluctuation; OpenSea; AdaBoost; Random forest

1 Introduction

The non-fungible token (NFT) market has seen some standouts in early 2021. According to statistics reported by White et al., NFT sales show a significant increase in two years −41% from 2019 to 2020 and a staggering 3,857% from 2020 to 2021 [1]. By the end of May 2022, the trading volume of these new markets for digital assets grows to approximately $800 million [2]. This trade growth matched the significant increase in Google NFT trends [3]. An NFT is a right to a digital property recorded on the blockchain, which can be any digital asset: an image, a video, a song, or virtual land, etc. [4]. Although transacted through cryptocurrencies, NFTs have distinct characteristics from cryptocurrencies that must be kept in mind when trying to understand them. The primary purpose of cryptocurrencies is to serve as currencies [5]. Unlike cryptocurrencies, which are equivalent, undifferentiable, and replaceable coins, NFTs are considered to be pure digital properties that can't be traded like for like. Therefore, NFTs are irreplaceable and unique [6]. An NFT can provide indisputable answers to questions such as who created this NFT, who owns it, and which of the many copies is the original [7]. There are numerous trading markets for various types of NFTs, with OpenSea and Axie Marketplace being the largest markets, and both of them have accumulated approximately 2.1 million traders [8]. However, while the public attention on NFTs has exploded, our understanding of their overall structures and market evolution is still inadequate. Nadini et al. built and studied traders and NFT networks and found that most traders are specialized [9]. Furthermore, according to a report by NonFungible Corporation, more and more speculative buyers' investments start to fail as the year progressed [10]. Therefore, it is crucial to help investors predict the fluctuations of NFT prices to reduce investment failures.

Thus far, studies on NFT prices have been relatively limited. Most commonly, authors attempt to examine trends in some specific collection [11] or link the price of a specific sought-after NFT to other financial assets [12,13]. However, these famous NFTs are limited, which implies that most investors cannot benefit from these specific NFTs. Investors need to consider the thousands of common NFTs that have not received much attention. Therefore, our paper presents a suitable method for the investors to use publicly available data to predict the NFT sale prices and their fluctuations.

In the field of financial investment, researchers have proposed many prediction methods, among which time series models are commonly employed in the study of price prediction. For instance, Mallqui et al. demonstrated that the blockchain information of Bitcoin is an important predictor of Bitcoin price fluctuations [14]. Kim et al. reported that in addition to macro economy factors, the Ethereum blockchain information is strongly associated to the Ethereum price [15]. However, unlike stocks or currencies, there is no daily price for NFTs, because the number and time of transactions are different for each NFT, making it difficult to use time series models for NFT sale-price prediction. Therefore, in this study, we employed machine learning regression models to predict the sale prices of NFTs, screened out the NFTs with multiple trading records, used classification models to predict the fluctuations of NFT sale prices.

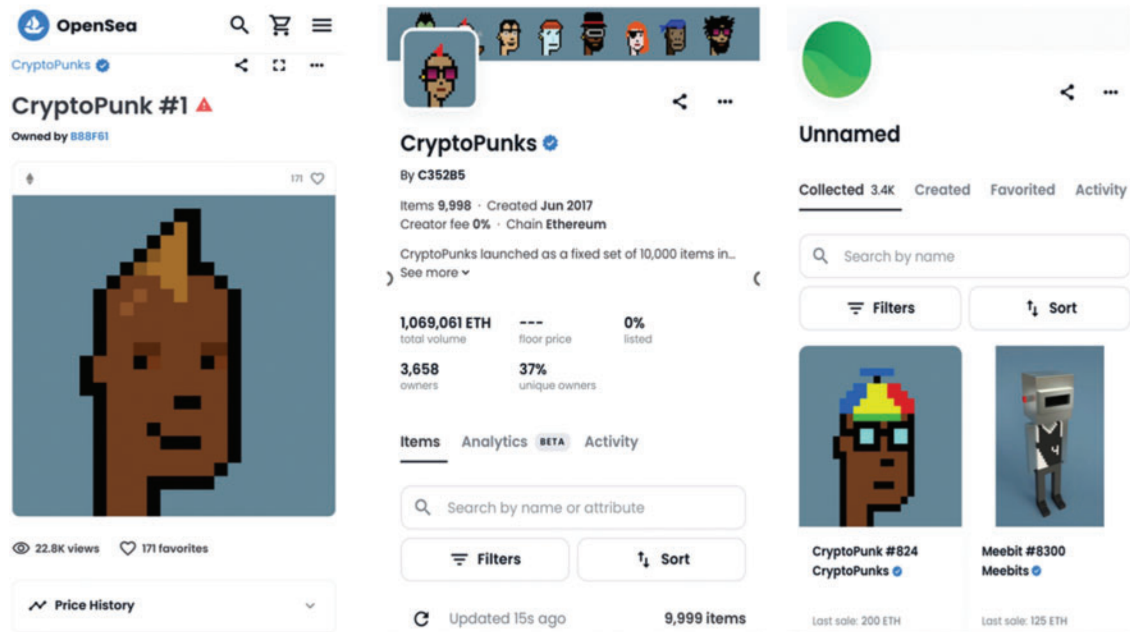
The specific purposes of this study are as follows: First, we introduced the machine learning approach for predicting the sale prices as well as the rise or fall in the prices of NFTs. Second, we analyzed and identified the features that aid in predicting the fluctuations of NFT sale prices with a high accuracy. Third, we developed realistic implications and recommendations for investors.

The organization of the rest of the paper is as follows: Section 2 reviews the related works. Section 3 presents our research design and dataset overview. In Section 4, we evaluate the prediction accuracy of each model and identify the variables that result in a highly accurate NFT price prediction. Section 5 discusses the theoretical and practical implications. In Section 6, we summarize the study results and discuss the limitations.

2 Related Works

2.1 OpenSea

OpenSea is the largest marketplace for NFT trading in the world, and according to statistics, it has amassed over $32B in trading volume and over 2.1M traders [8]. NFTs on OpenSea are mainly based on four blockchains, namely Ethereum, Klaytn, Polygon, and Solana, and thus far, more than 95% of the transactions are performed using the Ethereum blockchain. Further, there are many kinds of NFTs on OpenSea, including images, audio and video. The NFT collections can be divided into nine categories: Art, Domain Names, Collectible, Music, Sports, Photography, Trading Cards, Virtual Worlds, and Utility. Each NFT usually has three kinds of related pages, and the information on these pages can be summarized as: NFT information, collection information, and related account information. These three website pages of the NFTs are presented in Fig. 1 to illustrate OpenSea.



(a)NFT information page (b)Collection information page (c)Account information page

Figure 1: NFT related pages (for example, CryptoPunk #1)

2.2 Fluctuations in NFT Prices

The existing studies on NFT price fluctuations are primarily focused on specific sought-after collections. Dowling found that the pricing sequence of virtual lands in Decentraland is characterized by inefficiencies and appreciation in value [4]. In another study, Dowling studied Cryptopunks, Decentraland, and Axie Infinity and found a limited volatility transmission effect between the pricing of these NFTs and Bitcoin prices [12]. From other perspectives, Aharon and Demir used a time-varying parametric vector autoregressive approach to analyze the correlation between the NFT market and financial assets such as stocks [13]. However, we cannot use the sales trends of these well-known collections as a basis for studying the NFT market. That's because these sought-after collections, like Decentraland virtual lands and Cryptopunks, aren't representative of the general market.

To date, no research has used machine learning approach to predict NFT sales price fluctuations, and no reports are available on the demonstration of a specific NFT sale-price prediction error, except that of Nadini et al., who studied the predictability of NFT sales using a linear regression model. Subsequently, they used the AdaBoost algorithm to predict whether an NFT will be traded multiple times. However, they did not report the prediction error and used variables that are difficult to access by most investors, such as centrality of buyers and sellers in the network of NFT transactions, and so on [9].

We performed an empirical study with thousands of common NFTs as subjects to generalize the results. We attempted to predict NFT sale prices and fluctuations accurately by using publicly available and easily accessible data as independent variables of our machine learning models. We expect that this study will help the investors understand the NFT sale prices and their fluctuations more quickly and make appropriate investments.

2.3 Machine Learning

Machine learning can be divided into supervised learning, unsupervised learning, and reinforcement learning [16], and the machine learning algorithms are usually used for clustering, classification, and regression. Supervised learning is the prediction based on the input data, resulting in common decision rules [17]. Machine learning has been used in many areas, including demand forecasting, price prediction, and illness diagnosis [15,18,19].

2.3.1 Price Prediction with Machine Learning

To date, price prediction has been attempted by various researchers using a multitude of approaches. Weng et al. used random forests (RF), neural networks, support vectors, and boosted regression trees to form an expert system to predict stock prices. They found that features collected online improved the accuracy of predictions [20]. Antipov et al. found that the RF algorithm has an advantage over other methods in terms of accuracy when evaluating the price of residential apartments [21]. Čeh et al. compared the prediction accuracy of the RF algorithm and hedonic model when predicting apartment transaction prices. Their results also showed that RF produced significantly better prediction results [22].

In the field of cryptocurrency, the area most closely related to NFTs, Jang et al. determined the blockchain information and macroeconomic indicators are important to predicting Bitcoin's price [23]. Abraham et al. collected Google trend and Twitter data to predict cryptocurrencies prices, and they found that tweet volume is an important predictor of Bitcoin and Ethereum price fluctuations [24]. Valencia et al. compared RF, neural networks, and support vector machines in predicting cryptocurrencies prices. They also reported that Twitter data can be used to predict the price of cryptocurrencies and found that neural networks have higher accuracy [25].

2.3.2 AdaBoost

Adaptive Boosting (AdaBoost) was developed by Yoav Freund and Robert Schapire. It solved several optimization problems associated with boosting algorithms. This classification meta-algorithm uses the usual principles of boosting by building linear compositions of assumptions and executing the design through iterations of weak learning algorithms based on probability profiles computed using the results of prior iterations [26]. This method allows the learner to concentrate on difficult-to-learn examples [27]. Fig. 2 illustrates the modeling process of AdaBoost.

2.3.3 Random Forest

RF is an ensemble learning algorithm which was developed by Breiman. It incorporates the properties of many decision tree algorithms to classify or predict the values of the exploited variables [28]. RF is a new training set generated by repeated random selection of k samples from the raw training N with substitution, and then k classification trees are created based on the bootstrap set to form a random forest [29]. By assembling multiple weak classifiers, the results are polled or averaged so that the overall model results have high accuracy and a generalization performance [30]. Fig. 3 illustrates the modeling process of RF.

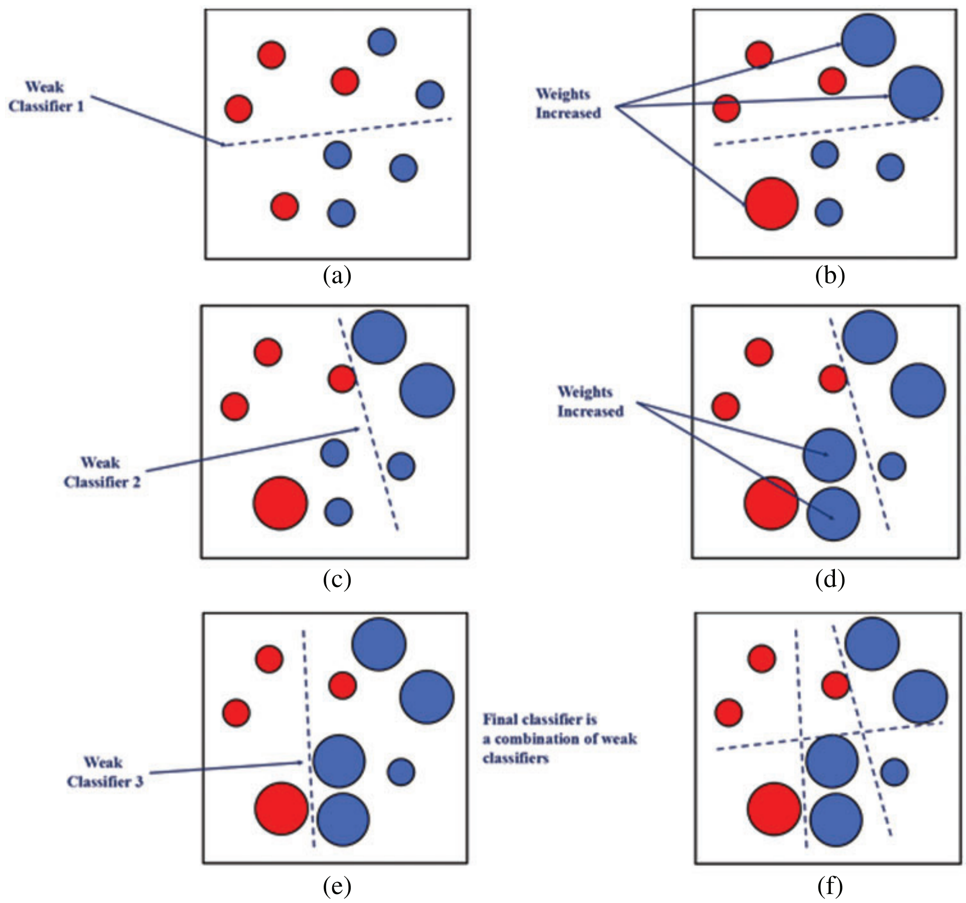


Figure 2: Modeling process of AdaBoost

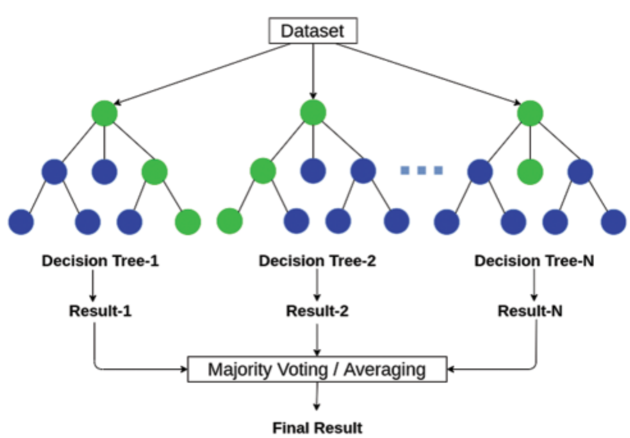


Figure 3: Modeling process of RF

3 Research Design

In this section, we present a method to predict the NFT sale prices and price fluctuations. Fig. 4 shows the flow of the presented method. In this study, we implemented supervised learning, which is commonly used to analyze the training data and result in an extrapolated function that can be used to map new instances [21,22].

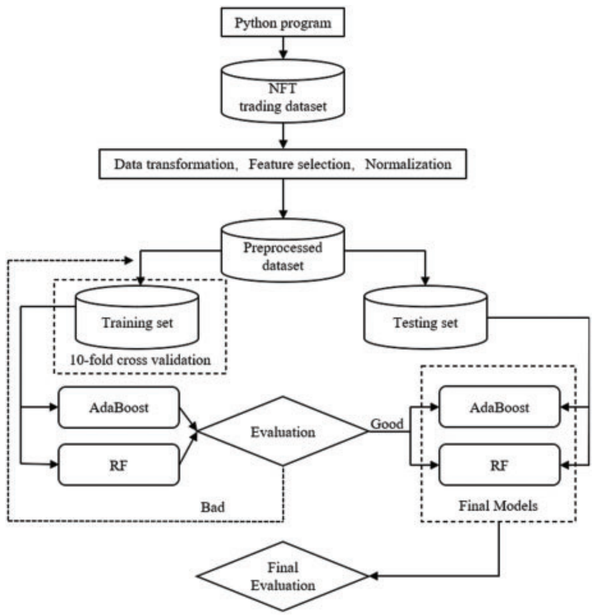


Figure 4: Flowchart of the method

3.1 Data Collection

Due to the changes in OpenSea policies, the application programming interface provided by the website has some limitations. Therefore, instead of using a web crawler to collect the data, we employed an automated testing tool to simulate the behavior of the people visiting the webpages (by using a Python program) and to collect the visited Hyper Text Markup Language links. Finally, XPath analysis was performed to collect data and obtain key information. Our publicly available Python program can be visited using the following uniform resource locator: https://github.com/wzx333222111/ZIXIONG-WANG. We collected 15,000 pieces of NFT information with sale records on the activity page between July 4 and July 7, 2022. We selected the ether chain NFTs for the data collection because of two reasons: (1) to unify the unit of value, and (2) because ether chain is the blockchain used by most NFTs. Various variables expected to be related to the target variables are shown in Table 1.

Table 1: Attributes of the dataset

Category	Attribute	Description	Type
NFT information	Favorited	Numbers of this NFT favorited by user	Numerical
	View	Numbers of this NFT viewed by user	Numerical

(Continued)

Table 1: Continued

Category	Attribute	Description	Type
	All time avg. price	Numbers of an NFT's all-time avg. price	Numerical
	Listings	Whether there are listings to this NFT	Categorical
	Offers	Whether there are offers to this NFT	Categorical
NFT collection information	Items	Numbers of NFTs in this collection	Numerical
	Owners	Numbers of owners of this collection	Numerical
	Floor price	Numbers of the floor price of this collection	Numerical
	All time avg. price	Numbers of all time avg. price	Numerical
	All time volume	Numbers of all time volume	Numerical
	Verify status	Whether the collection is a verified collection	Categorical
Creator information	Collected	Numbers of NFTs collected by user	Numerical
	Favorited	Numbers of NFTs favorited by user	Numerical
	Identification status	Whether the user is a verified account	Categorical
	Created status	Whether the user has created an NFT	Categorical
	Activity status	Whether the user has transfer or sale an NFT	Categorical
	Offers received	Whether the user has received an offer	Categorical
Owner information	Collected	Numbers of NFTs collected by user	Numerical
	Favorited	Numbers of NFTs favorited by user	Numerical
	Created status	Whether the user has created an NFT	Categorical
	Activity status	Whether the user has transfer or sale an NFT	Categorical
	Offers received	Whether the user has received an offer	Categorical
Target variables	Sales price	Numbers of sale price of an NFT	Numerical
	Fluctuation	Fluctuation of NFT sale price −1 for rise −0 for fall	Categorical

3.2 Data Preprocessing and Ensemble Modeling

Next, we preprocessed the collected data, i.e., performed reduction, cleaning, transformation, and normalization of the collected data to improve the prediction accuracy of the model.

Among the 15,000 collected data, we finally selected 8586 datapoints and split them into a training set (7634 data) and a testing set (952 data) for predicting NFT sale prices. The standard deviation and mean values of the variables are presented in Table 2. In addition, we filtered the NFTs with multiple sale price records, finally selected 6136 datapoints, and divided them into training (5358 data) and testing sets (778 data) to predict the fluctuations of the NFT sale prices.

Table 2: Descriptive statistics of the variables

Category	Attribute	Mean	S.D.
NFT information	Favorited	7.514	190.956
	View	61.223	27.317
	All time avg. price	0.515	1.149
	Sale price	0.519	1.182
NFT collection information	Items	10871.583	17941.408
	Owners	25923.708	86507.56
	Floor price	0.419	1.149
	All time avg. price	0.769	1.962
	All time volume	12804.826	41944.275
Creator account information	Collected	133.375	1653.986
	Favorited	3.616	50.883
Owner account information	Collected	761.711	11920.806
	Favorited	62.160	328.5221

Then, we used regressor and classifier to develop predictive models. We used Orange (Ver. 3.32.0) as the data analysis tool, which integrates many machine learning algorithms, such as RF, Tree, and Gradient Boosting. However, after comparing the preliminary prediction results of these algorithms available in Orange, we selected the RF and AdaBoost algorithms for this study. The schematic of this tool's operation is depicted in Fig. 5. To improve the reliability of the regression model, we used K-fold cross-validation, which was performed 10 times. Cross-validation is a method in which one model validation technique, among a large number of similar ones, is used to assess how the results of statistical analyses can be generalized to independent datasets [31,32]. In prediction problems, the model is usually provided a known dataset for training (training set) and an unknown dataset to test the model (testing set) [33]. The purpose of cross-validation is to reduce selection bias and overfitting problems [34].

3.3 Predictor Phases for Regression Model

We developed four phases to estimate the predictive performance of NFT information, collection information and related account information in the regression model. Table 3 shows the variables included in each phase. In Phase 1, we input NFT information as predictors. We expect that using only NFT information as predictors yields high prediction accuracy because these variables provide the most basic information about NFT. The related account information and collection information are added to the variables considered in Phase 2 and Phase 3, respectively. We expect that the addition of these variables should lead to an improvement in prediction accuracy, albeit with a much smaller overall contribution than that of the variables in the first phase. Phase 4 includes all the three types of information, and we expect to get the best prediction results in this phase.

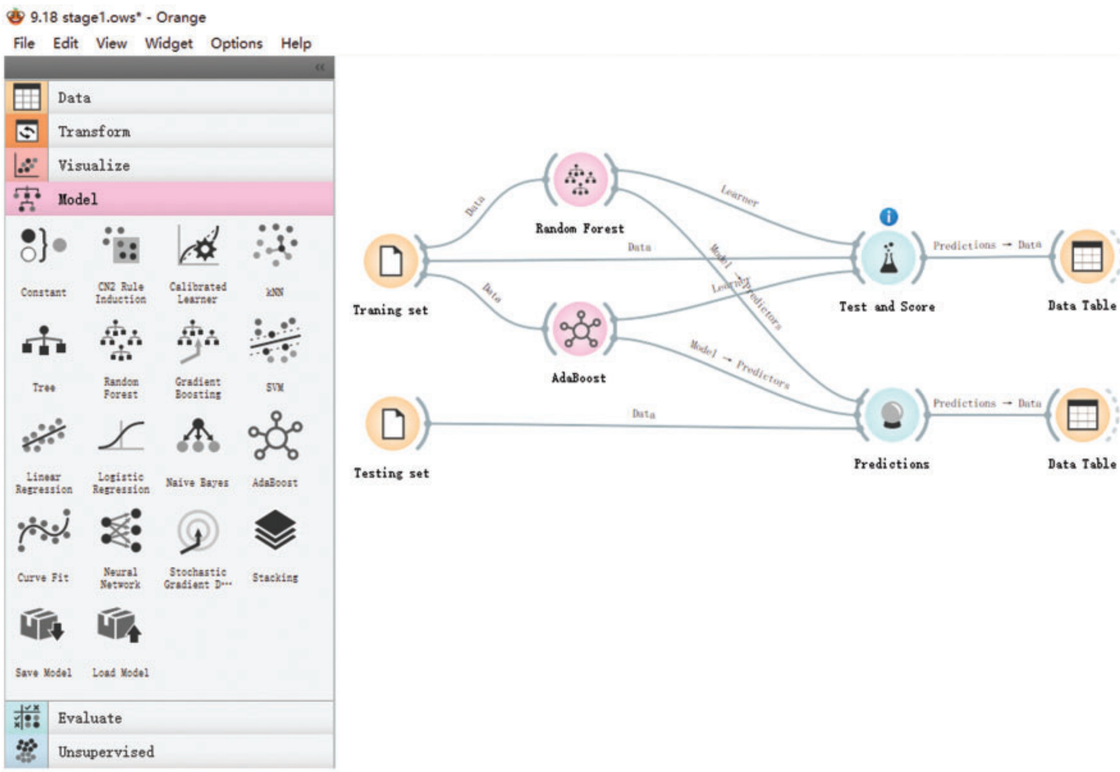


Figure 5: Orange operation schematic

Table 3: Variables included in each phase

	Phase 1	Phase 2	Phase 3	Phase 4
NFT information	√	√	√	√
Account information		√		√
Collection information			√	√

4 Empirical Results

4.1 Regression Model Predictive Result Analysis

In this study, we used root mean square error (RMSE), mean absolute error (MAE) and R-square (R^2) to evaluate the performance of the regression models [15,24,35].

We performed a step-by-step analysis from Phase 1 to 4. Phase 1 includes only the NFT information variables, and RF performs better in this phase of the prediction (RMSE = 0.119, MAE

= 0.042, and R^2 = 0.990) (see Table 4). In Phase 2 with account information, we found that AdaBoost achieved better prediction results, and RMSE, MAE and R^2 (RMSE = 0.047, MAE = 0.007, and R^2 = 0.998) (see Table 5) showed significant improvements over their Phase 1 counterparts. The prediction results of Phase 3, which combined NFT and collection information, were slightly more accurate than those of Phase 1 (RMSE = 0.075, MAE = 0.009, and R^2 = 0.996) (see Table 6). However, compared with Phase 2, the prediction error in this phase became larger. Finally, we input all the variables into the prediction model and observed that the results of Phase 4 were better than those of Phase 1 and 3, whereas the RMSE value was slightly larger than those of Phase 2 (RMSE = 0.053, MAE = 0.007, and R^2 = 0.998) (see Table 7). These results show that the relevant account information contains information directly related to the NFT sale price.

Table 4: Results of data analysis for phase 1

Input	Algorithm	RMSE	MAE	R^2
NFT information (number of inputs: 5)	AdaBoost training	0.179	0.051	0.977
	AdaBoost testing	0.135	0.015	0.987
	RF training	0.197	0.074	0.972
	RF testing	0.119	0.042	0.990

Table 5: Results of data analysis for phase 2

Input	Algorithm	RMSE	MAE	R^2
NFT information (number of inputs: 5)	AdaBoost training	0.116	0.032	0.990
Account information (number of inputs: 11)	AdaBoost testing	**0.047**	**0.007**	**0.998**
	RF training	0.225	0.093	0.964
	RF testing	0.186	0.063	0.976

Table 6: Results of data analysis for phase 3

Input	Algorithm	RMSE	MAE	R^2
NFT information (number of inputs: 5)	AdaBoost training	0.099	0.024	0.993
Collection information (number of inputs: 6)	AdaBoost testing	0.075	0.009	0.996
	RF training	0.133	0.048	0.987
	RF testing	0.077	0.027	0.996

4.2 Evaluation of Classification Model

In the classification model, we evaluated the model using F1-score, precision, and recall [36]. We also performed the receiver operating curve (ROC) analysis. The ROC is established by graphing the relationship between the true positive and the false positive rates under different settings of the threshold. The false positives can also be called as false alarm probabilities [37]. The results of the training set are shown in Table 8. The RF classification model exhibited a peak area under the curve (AUC) value of 0.945, whereas the AdaBoost model showed a peak AUC value of 0.853. In addition, considering the precision, recall, and F1 scores, the RF algorithm outperformed the AdaBoost algorithm in determining the fluctuations of NFT sale prices. The ROC analysis results are shown in Table 9.

Table 7: Results of data analysis for phase 4

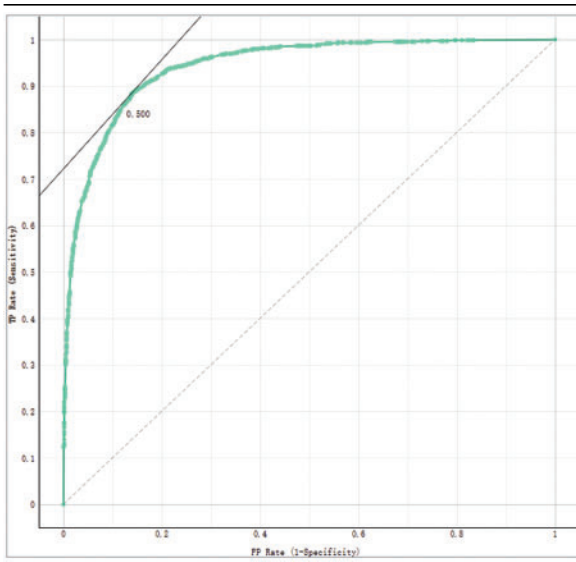
Input	Algorithm	RMSE	MAE	R^2
NFT information (number of inputs: 10)	AdaBoost training	0.072	0.020	0.996
Account information (number of inputs: 14)	AdaBoost testing	0.053	0.007	0.998
Collection information (number of inputs: 5)	RF training	0.179	0.066	0.977
	RF testing	0.103	0.039	0.993

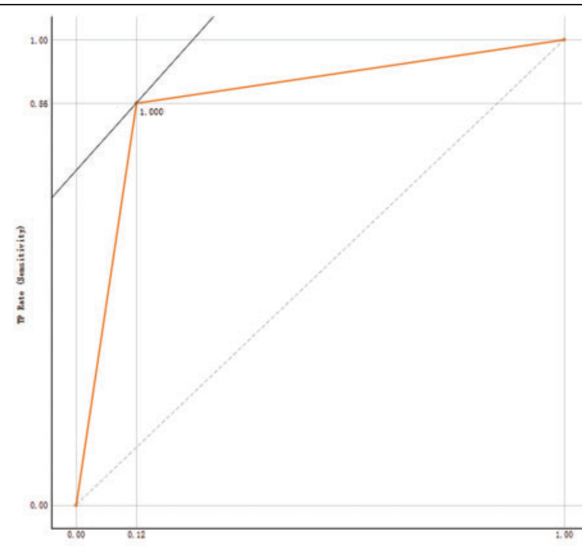
Table 8: Training set results of classification model

Algorithm	AUC	F1	Precision	Recall
AdaBoost	0.853	0.854	0.853	0.855
RF	0.945	0.883	0.882	0.884

Table 9: ROC of each model

Random forest (AUC: 0.945)	AdaBoost (AUC: 0.853)





For further evaluation, we selected 778 datapoints as the testing set, and again predicted the fluctuations of NFT sale prices using the RF model. Table 10 shows results of the testing set. Among the 778 datapoints in the testing set, 461 and 317 points exhibited actual sale price fall and rise, respectively. The RF classification model correctly predicted the price volatility for 700 samples and incorrectly predicted price rise and fall for 36 and 42 samples, respectively. The results of the testing set are shown in Table 11.

Table 10: Confusion matrix for RF

		Predicted		$\sum$
		0	1	
Actual	0	419	42	461
	1	36	281	317
$\sum$		455	323	778

Table 11: Prediction performance of RF

Algorithm	AUC	F1	Precision	Recall
RF	0.956	0.878	0.870	0.886

4.3 Evaluation of Variables for Classification Model

After conducting a learning process to determine the fluctuations of NFT sale prices, we further analyzed the variables collected from OpenSea to identify the ones that contained more information for determining the fluctuations of NFT sale prices based on the RF model. Fig. 6 shows the contribution levels of the variables that affected the predictions. NFT all time avg. price (0.025) was the most influential variable for predicting price rise or fall, followed by the number of NFT favorites (0.022), number of creator's favorites (0.020), creator's activity status (0.020), and number of owner's collection (0.011). Among them, NFT favorite means that other users browsed this NFT and clicked the favorite button, which to some extent reflects the recognition and liking of other users for this NFT. Creator's favorites indicate those NFTs for which the creator of the NFT clicked the favorite button. Creator's activity indicates whether there has been recent NFT transfers or transaction activities. Both these variables reflect the activity level of the NFT creator. The owner's collection represents the number of NFTs owned by the owner.

			#	Gini
1	N	all time price		0.025
2	N	NFT favorite		0.022
3	N	creator favorite		0.020
4	C	creator activity	2	0.020
5	N	owner'scollection		0.011
6	N	creater's collection		0.011
7	N	collection floor price		0.009
8	C	creator creat	2	0.008
9	N	NFTview		0.008
10	N	collection all time avg. price		0.008
11	C	NFT List	2	0.006
12	N	collection all time volume		0.004
13	C	owner create	2	0.004
14	N	ownerfavorite		0.004
15	C	NFT offer	2	0.004
16	C	collection identification	2	0.003
17	N	collection owners		0.003
18	C	creator identification	2	0.002
19	C	owner activity	2	0.002
20	N	collection items		0.001
21	C	owner offer received	2	0.001
22	C	creator offer received	2	0.000

Figure 6: Predicted contributions of the variables

5 Implications and Discussion

5.1 Theoretical Implications

The results of this study have important theoretical implications. First, we determined whether publicly available information on OpenSea helps to predict the price fluctuations of NFTs. According to previous studies, specific famous NFT prices are associated with cryptocurrencies and other financial assets [12,13]. However, these studies were performed without considering the thousands of ordinary NFTs sold publicly in the NFT market. We identified a significant correlation between the sale price fluctuations of these ordinary NFTs and the information publicly available on OpenSea, and herein, suggest that this information can be used to predict the fluctuations of NFT prices.

Second, NFT is a digital asset with considerable public interest, but few researchers have explored the price-prediction approach. Although Nadini et al. obtained the NFT trading information from Ethereum and WAX blockchains and analyzed the predictability of NFT sales using a linear regression model, they used some variables that are difficult to obtain by most investors, such as centrality of buyers and sellers in the network of NFT transactions and did not report their prediction error [9]. In the present study, we extended the study of Nadini et al. by using Python programs to collect NFT information that is readily available to most investors and used this collected NFT information as independent variables. Further, we demonstrated the prediction error by empirical analysis (RMSE = 0.047, and MAE = 0.007).

Third, in addition to the NFT information of interest, such as NFT price history and numbers of favorited and most viewed NFTs, we also evaluated the prediction value provided by the collection and related account information. The results of the classification model for predicting the fluctuations in the NFT sale prices demonstrated the importance of the relevant account information and revealed that the number of creator's favorites, activity status, and number of owner's collections primarily contribute to the prediction results. Collectively, the results of the regression and classification models indicate that these three features are crucial for predicting the NFT prices, and thus, future studies on NFT price fluctuations should be conducted by considering these three features.

5.2 Practical Implications

In this study, publicly available data and simple machine learning algorithms were used. As a result, the methodology and results of this study can be easily replicated by academics and practitioners aiming to predict future NFT price fluctuations.

The results of this study also have practical implications: First, we identified some important variables on OpenSea to predict price fluctuations of NFTs. The investors can use these NFT and account information to understand the NFT prices more simply and make appropriate investments. More specifically, we demonstrated that the prediction value of the relevant account information is higher than that of the collection information. Thus, investors can rapidly and accurately predict the NFT prices. Further, using only the NFT and related account information can prevent unnecessary data collection and reduce the calculation time. We suggest that investors can focus on the number of creators' collections, status of their activity, and number of owners' collections, because the NFTs created by more active creators have a greater potential. Thus, in simple words, investors need to find such active creators.

Second, NFT practitioners can produce more attractive NFTs and promote them to ensure more people click the favorite button. In addition, practitioners should pay attention to the role of account activity frequency, because the NFTs created by more active accounts will attract more attention and likes. Therefore, it is necessary to increase the number of favorites given to other NFTs as well as increase the NFT trading activities.

Third, we compared the accuracy of AdaBoost and RF, in predicting the NFT prices and their fluctuations and observed that the practitioners can use the AdaBoost and RF algorithms for predicting the NFT prices and price fluctuations, respectively.

6 Conclusion and Limitations

In this study, machine learning algorithms were used for predicting the NFT prices and their fluctuations. We used regression models to predict the NFT sale prices and classification models to predict the NFT sale price fluctuations. The regression (RMSE = 0.047, MAE = 0.007) and classification (AUC = 0.956, F1 = 0.878, and Precision = 0.870, Recall = 0.886) models provided accurate predictions using the AdaBoost and RF algorithms, respectively.

The results indicate that, in addition to the historical average price of the NFTs, the relevant account information provides the most predictive power. The results obtained using the classification model indicate that the number of favorites and activity status of the creator are essential for the prediction performance as they attract more public attention and likes. This shows that the investors can pay more attention to the creator's activity frequency to assess the investment potential of the NFTs.

The limitations of this study are as follows:

1. Only the publicly available information on OpenSea was considered in this study. However, the website is continuously updated as the market develops, and we cannot predict what new information will be publicly available on the site in the future. In future studies, the updates on the website should be followed to obtain updated results.
2. We only considered the factors available on the most popular website, OpenSea. However, there are many other NFT trading marketplaces, such as Binance.com. Further, as indicated in previous studies, blockchain information has a positive effect on Ethereum price prediction [9]. Thus, the impact of blockchain information on the price of NFT, which is another product of blockchain technology, will be examined in our future study.
3. We cannot distinguish between normal and wash trading. The NFT market, as an emerging market, does not have tenable trading rules, which leads to the phenomenon of wash trading in various trading markets.

Acknowledgement: This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the Innovative Human Resource Development for Local Intellectualization support program (IITP-2022-RS-2022-00156287) supervised by the IITP (Institute for Information & communications Technology Planning & Evaluation). This work was supported by Institute for Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2022-0-01203, Regional strategic Industry convergence security core talent training business).

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] B. White, A. Mahanti and K. Passi, "Characterizing the OpenSea NFT marketplace," in *Companion Proc. of the Web Conf. 2022*, Lyon, France, pp. 25–29, 2022.
[2] NonFungible Corporation, NFT Market Report Q2–2022, 2022. [Online]. Available: https://nonfungible.com/market-tracker
[3] Google Trends, 2022 (2022. [Online]. Available: https://trends.google.com/trends/explore?q=nft&geo=KR

[4] M. Dowling, “Fertile LAND: Pricing non-fungible tokens,” *Finance Research Letters*, vol. 44, no. 4, pp. 1–5, 2022. https://doi.org/10.1016/j.frl.2021.102096

[5] D. G. Baur, K. H. Hong and A. D. Lee, “Bitcoin: Medium of exchange or speculative assets?” *Journal of International Financial Markets, Institutions and Money*, vol. 54, no. 4, pp. 177–189, 2018.

[6] Q. Wang, R. Li, Q. Wang and S. Chen, “Non-fungible token (NFT): Overview, evaluation, opportunities and challenges,” arXiv preprint arXiv:2105.07447, 2021.

[7] T. W. Lounge, “Choosing the right blockchain for your NFT,” 2020. [Online]. Available: https://medium.com/phantasticphantasma/choosing-the-right-blockchain-for-your-nft-d1df2bebae91

[8] DappRadar, NFT Marketplace, 2022. [Online]. Available: https://dappradar.com/nft/marketplaces

[9] M. Nadini, L. Alessandretti, D. Giacinto, M. Martino, L. M. Aiello *et al.,* “Mapping the NFT revolution: Market trends, trade networks, and visual features,” *Scientific Reports*, vol. 11, no. 1, pp. 1–11, 2021.

[10] NonFungible Corporation, Non-fungible tokens quarterly report Q3–2021, 2021. [Online]. Available: https://nonfungible.com/market-tracker

[11] X. J. Jiang and F. L. Xiao, “Cryptokitties transaction network analysis: The rise and fall of the first blockchain game mania,” *Frontiers in Physics*, vol. 9, no. 57, pp. 1–12, 2021. https://doi.org/10.3389/fphy.2021.631665

[12] M. Dowling, “Is non-fungible token pricing driven by cryptocurrencies?” *Finance Research Letters*, vol. 44, no. 534, pp. 1–6, 2022. https://doi.org/10.1016/j.frl.2021.102097

[13] D. Y. Aharon and E. Demir, “NFTs and asset class spillovers: Lessons from the period around the COVID-19 pandemic,” *Finance Research Letters*, vol. 47, pp. 1–9, 2021. https://doi.org/10.1016/j.frl.2021.102515

[14] D. C. Mallqui and R. A. Fernandes, “Predicting the direction, maximum, minimum and closing prices of daily Bitcoin exchange rate using machine learning techniques,” *Applied Soft Computing*, vol. 75, no. 2, pp. 596–606, 2019.

[15] H. M. Kim, G. W. Bock and G. Lee, “Predicting Ethereum prices with machine learning based on blockchain information,” *Expert Systems with Applications*, vol. 184, no. 3, pp. 1–8, 2021. https://doi.org/10.1016/j.eswa.2021.115480

[16] J. D. Kelleher, B. M. Namee and A. D’arcy, *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies*. Cambridge, Massachusetts: MIT press, 2020.

[17] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis and D. I. Fotiadis, “Machine learning applications in cancer prognosis and prediction,” *Computational and Structural Biotechnology Journal*, vol. 13, pp. 8–17, 2015.

[18] J. Soni, U. Ansari, D. Sharma and S. Soni, “Predictive data mining for medical diagnosis: An overview of heart disease prediction,” *International Journal of Computer Applications*, vol. 17, no. 8, pp. 43–48, 2011.

[19] S. C. K. Tekouabou, “Intelligent management of bike sharing in smart cities using machine learning and Internet of Things,” *Sustainable Cities and Society*, vol. 67, pp. 1–14, 2021. https://doi.org/10.1016/j.scs.2020.102702

[20] B. Weng, L. Lu, X. Wang, F. M. Megahed and W. Martinez, “Predicting short-term stock prices using ensemble methods and online data sources,” *Expert Systems with Applications*, vol. 112, no. 3, pp. 258–273, 2018.

[21] E. A. Antipov and E. B. Pokryshevskaya, “Mass appraisal of residential apartments: An application of random forest for valuation and a CART-based approach for model diagnostics,” *Expert Systems with Applications*, vol. 39, no. 2, pp. 1772–1778, 2012.

[22] M. Čeh, M. Kilibarda, A. Lisec and B. Bajat, “Estimating the performance of random forest versus multiple regression for predicting prices of the apartments,” *ISPRS International Journal of Geo-information*, vol. 7, no. 5, pp. 1–16, 2018. https://doi.org/10.3390/ijgi7050168

[23] H. Jang and J. Lee, “An empirical study on modeling and prediction of bitcoin prices with bayesian neural networks based on blockchain information,” *IEEE Access*, vol. 6, pp. 5427–5437, 2017.

[24] J. Abraham, D. Higdon, J. Nelson and J. Ibarra, “Cryptocurrency price prediction using tweet volumes and sentiment analysis,” *SMU Data Science Review*, vol. 1, no. 3, pp. 1, 2018.

[25] F. Valencia, A. Gómez-Espinosa and B. Valdés-Aguirre, "Price movement prediction of cryptocurrencies using sentiment analysis and machine learning," *Entropy*, vol. 21, no. 6, pp. 1–12, 2019. https://doi.org/10.3390/e21060589

[26] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, 1997.

[27] S. Mishra, D. Mishra and G. H. Santra, "Adaptive boosting of weak regressors for forecasting of crop production considering climatic variability: An empirical assessment," *Journal of King Saud University-Computer and Information Sciences*, vol. 32, no. 8, pp. 949–964, 2020.

[28] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[29] A. Géron, *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. Sebastopol, California: O'Reilly Media, Inc press, 2019.

[30] V. Rodriguez-Galiano, M. Sanchez-Castillo, M. Chica-Olmo and M. J. O. G. R. Chica-Rivas, "Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines," *Ore Geology Reviews*, vol. 71, pp. 804–818, 2015.

[31] M. Stone, "Cross-validatory choice and assessment of statistical predictions," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 36, no. 2, pp. 111–133, 1974.

[32] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," *IJCAI*, vol. 14, no. 2, pp. 1137–1145, 1995.

[33] A. Galkin, "What is the difference between test set and validation set," 2018.

[34] G. C. Cawley and N. L. Talbot, "On over-fitting in model selection and subsequent selection bias in performance evaluation," *The Journal of Machine Learning Research*, vol. 11, pp. 2079–2107, 2010.

[35] S. C. K. Tekouabou, W. Cherif and H. Silkan, "Improving parking availability prediction in smart cities with IoT and ensemble-based model," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 3, pp. 687–697, 2020.

[36] A. C. Müller and S. Guido, *Introduction to Machine Learning with Python A Guide for Data Scientists*. Sebastopol, USA: O'Reilly Media, Inc., 2016.

[37] R. A. Kohavi, "Study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. of the Fourteenth Int. Joint Conf. on Artificial Intelligence*, Montréal, Canada, pp. 1137–1143, 1995.