NFT Price prediction system

Shashank Rustagi
MT21081
MTech CSE, IIITD
shashank21081@iiitd.ac.in

Jahnvi H Kadia
MT21123
MTech CSE, IIITD
jahnvi21123@iiitd.ac.in

Drishya Uniyal
MT21119
MTech CSE, IIITD
drishya21119@iiitd.ac.in

Shashank Daima
2019106
IIITD
shashank19106@iiitd.ac.in

Giridhar S
MT21026
Mtech CSE, IIITD
giridhar21026@iiitd.ac.in

1 ABSTRACT

We aim to predict the price of the new image based on the trained images for the new users to get a fair chance to profit by investing in NFTs. NFT(s) are assets with hexadecimal ids assigned to them and stored in the blockchain. This makes it impossible for two duplicates to exist in the NFT marketplace. But it's quite difficult to find an NFT, which might be rare and cheap. Our report/Project is a small action in this direction. We compare NFT based on Image's Bitmaps and its features to predict price categories such as A(90 ETH+), B(80-90 ETH), etc. This will help naives to get some indication of the actual rareness of NFT. Hence, Naives can make better decisions. We used CNN, VGG19, Mobile-Net V2 to predict the prices. And comparing these approaches to find the best NFT-recommendation system.

2 INTRODUCTION

. We aim to predict the price of the new image based on the trained images for the newbies to get a fair chance to profit by investing in NFTs. Firstly, we will predict the new images by building different types of CNN models with just the image as input. Our methodology is based on deep learning techniques to find similar NFTs to previously boomed NFTs. The recent rise in popularity in the blockchain and crypto space is well known. In the high demand for crypto and Dapp (decentralised applications) space, one name that comes to notice is NFT or Non-fungible tokens. NFTs are digital items like art, music, sports, and more that blockchain technology can buy and sell and is famous for their high inflation and celebrity status. The new concept interests us in classifying the images based on Ethereum (an open-source blockchain with smart contract functionality). Curious about how much will an image cost? Our proposal is shown in the image on right side.

. If we consider the images on the right, we can see their price.Our aim says that our model should predict a new image based on the trained images to check how much our image will cost.

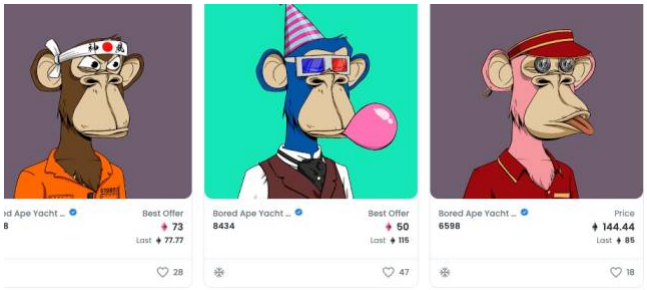


Figure 1: The Bored Ape Yacht Club is a collection of 10,000 unique Bored Ape NFTs (https://opensea.io/collection/boredapeyachtclub).

Figure 2: CNN model that be used for predicting the price

. This way, a new image will be compared to the trained images using different types of Convolutional Neural Network models, which pass the image through its layers, makes a feature map, and then predicts the prices. . This way, a new image will be compared to the trained images using different types of Convolutional Neural Net-work models, which passes the image through its layers, makes a feature map, and then predicts the prices. This way, when we get a new image, we will be able to predict its price brackets.

3 LITERATURE REVIEW

"[1] Mapping the NFT revolution: market trends, trade networks, and visual features". In this paper the author first

characterizes statistical properties of the market. Second, the author builds the network of interactions, show that traders typically specialize on NFTs associated with similar objects and form tight clusters with other traders that exchange the same kind of objects. Third, the author cluster objects associated to NFTs according to their visual features and show that collections contain visually homogeneous objects.

. Finally, the author investigates the predictability of NFT sales using simple machine learning algorithms and find that sale history and, secondarily, visual features are good predictors for price. We anticipate that these finding's will stimulate further research on NFT production, adoption, and trading in different contexts.

"[2] Non-Fungible Tokens: Blockchains, Scarcity, and Value". In this paper the author explains the mechanisms of NFT. The mechanism for NFT creation is firstly upload a file onto an NFT auction market, secondly the file is recorded on the digital ledger as an NFT. Lastly, NFTs are ready to be sold out or can be purchased. The NFT creation by an artist describes it's work that solely belongs to that artist and can never copyright that work to reproduce more NFTs with the similar art. Hence, someone who purchases the file does not have exclusive access to the file and so cannot have access to the original digital file.

. The problem with this is that anyone can upload the artwork onto an NFT, without proving whether they are the original creator of it or not. Also, the person who is buying an NFT is not an automatic owner of original objects and also there isn't any means that clarifies that the file is not reproduced or used by anyone else. All these cases can lead to risk For the ownership of a digital asset, an NFT forms a blockchain with a unique signature. For the creative works, in the real-world scenario one would autograph their work, this can be done through a cryptographic hash function.

"[3] Influence of Social Media on NFT valuation". Twitter is used as a vehicle for NFT under and over valuation. This research paper is taking social media features into account to understand the effect of OpenSea and Twitter on the sales of the NFT.OpenSea is the dataset for NFT. NFT has been divided into multiple asset classes like Unprofitable NFT, very low value NFT, low value NFT, and so on. OpenSea has a term named ASSET which is a unique digital term stored as an NFT on the blockchain. Buyer can resell the NFT at a higher price if he wishes to.

. The author performs temporal analysis using both of the Datasets. Spearman's correlation coefficient was calculated and it was showing a strong positive correlation. Time delay was analyzed between posting the NFT on OPEN SEA and then tweet posting time. The time delay was following a log normal distribution. Average number of followers and asset value is plotted and the plot is log log plot.

. Binary Classification and Multi-class Classification is used for model evaluation. Multiple models like DTree and XGBoost classifier and XGBoost worked better on Binary classification as compared to multi-class classification. Twitter + Opeansea collectively performed better than individually with an overall accuracy of nearly 63

"[4] A Time-series Analysis of How Google Trends Searches Affect Cryptocurrency Prices for Decentralized Finance and Non-Fungible Tokens". This paper talked about decentralized finance and non fungible tokens. The Cryptocurrency boom came after 2020. Defi financial transactions have a great future because there is a huge cost cutting on the operating cost. A huge boom in the NFT market is also seen after the Twitter CEO tweeted about his first NFT. Price information was gathered from google data for both the NFT and DeFi.

. Google trend searches were equalized and normalized. In this research paper, three models were constructed. A linear gaussian state space model was used. First model was the bitcoin model in which there was a second order model for the trend component. The Ethereum model is almost similar to the bitcoin model but just one change from the previous day trend also affected today's trend.

. The third model revolves around the closing price of the NFT and the model is named as Model for De-Fi and NFT related coins. All the three models are compared in the end, and found out that the previous day price is a very good price forecasting variable.

"[5] Model for improvement of nonprofit student organizations funding by machine learning and digital transition to blockchain". Most student organizations can be classified as nonprofits. Sponsorships are their main source of income and fundraising is a key activity responsible for allocating sufficient funds required for project realization. Typically, companies are the principal outside stakeholders concerned with the success of student organizations' projects and their main sponsor. Student organizations cannot be successful and build long term relationships with enterprises without being transparent in how the donated funds are being spent. By introducing the blockchain-based smart contracts model a student organization could publicly declare the amount of capital required for every project, its distribution as well as fundraising deadlines. As the data on the blockchain is completely transparent it can be effortlessly utilized as an input data set for building machine learning models.

. The first and most important thing to predict is the overall funding of the given workshop. This can help the cost management and all the activity planning for the workshop.

Due to its specificity and adaptation to this task, a multiple linear regression model is selected.

. As sponsors seek brand visibility, having been listed as NFT holders for various workshops goes hand in hand with getting attention from freshmen students and young graduates. The very nature of blockchain protects from delisting or hiding a credible project's sponsors. In the same way, phony sponsorship will not be an option as cryptography prevents them from creating fake evidence

4 NOVELTY

There has not been much studies about prediction of prices using model on NFTs. We propose to train the data using different models and then compare these models to find which performs better. As far as we have studied, there has not been any comparison of models in predicting the prices for nfts.

5 METHODOLOGY OF PROPOSED METHOD

In this project, we split the nfts according to prices into 7 categories, for which distribution is given below:

Label	Count
A+	1611
A	904
B	903
C	1021
D	811
E	794
F	785

. Thus we apply different models to classify nfts according to the given categories.

5.1 Data Collection

We collected our data manually by web scraping every entry of some sample NFT-collections (e.g. BoredApeYachtClub, CryptoChicks etc.) and currently have around 6800 NFTs. We used a selenium chrome driver to scrap NFT's id, imageUrl, price, currentTime, collectionName.

5.2 Models

1. VGG19: It contains 16 convolution layers in total.

2. MobileNetV2: It contains 53 convolution layers, 1 Avgpool with nearly 350 gflops.

.

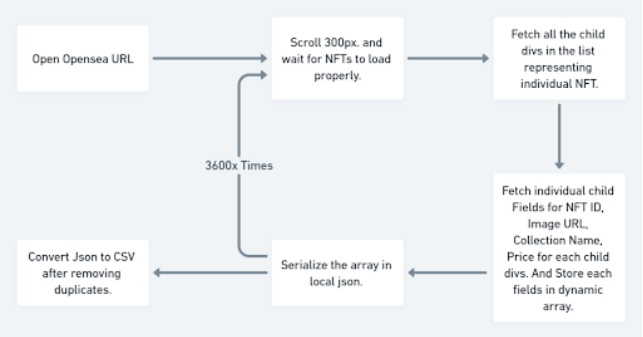


Figure 3: Data collection process

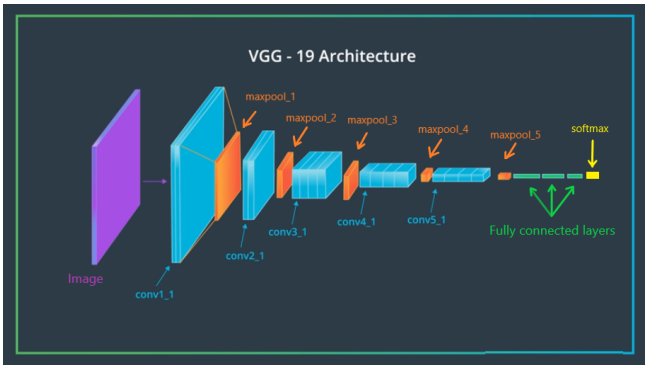


Figure 4: VGG19 (https://medium.com/@princyjune66/malaria-detection-using-transfer-learning-7cb5798a7f30).

6 BASELINE MODEL

The following is the metrics obtained after training and testing the data after using CNN.

Train ACC	VAL ACC	Test ACC
74.86 %	53.61 %	53.46 %

7 EVALUATION

MobileNetV2 is better than CNN because instead of using a single 3*3 convolution layer which in turn is followed by batch norm and ReLU, MobileNetV2 splits the convolution layers into 3*3 depthwise convolution and 1*1 pointwise convolution. Also, MobileNet is comparatively faster as well as a smaller model. MobileNetV2 gives better performance than CNN.

VGG19 is the type of CNN model considered by far the best computer vision model. VGG19 is one of the most well-known algorithms for image classification. VGG19 uses 3*3 small filters, and CNN uses only a single 3*3 convolution layer. Hence VGG19 gives better performance than CNN.

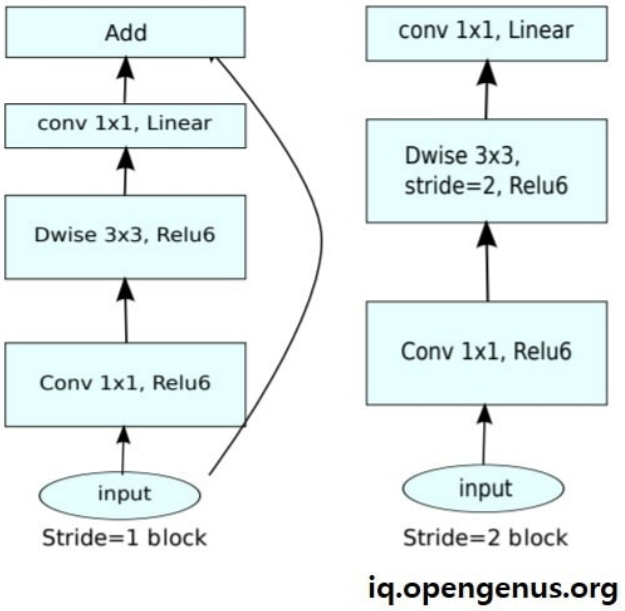


Figure 5: MobileNetV2 (https://iq.opengenus.org/mobilenetv2-architecture/).

Model	Train ACC	Val ACC	Test ACC
CNN	74.86	53.61	53.46
ResNet	67.6	58.08	56.2
MobileNetV2	59.44	54.64	57.56 %
VGG19	60.46	56.7	57.85

. Among the models used VGG19 performs better than the other models, maybe this can be because VGG19 has many layers and has been pretrained on a large data. VGG19 is a deep network and also helps in classification and object detection.

8 COMPARISON

Comparing with the existing models, mostly the prediciton of prices by judging the book cover images have been done. One can predict the price of a book by its book cover. Working on the same methodology, we proposed our model. For the baseline we initially predicted the pirces of the nft's on a particular dataset, but as the dataset changed and multiple collections were added, we focused on giving a class to prices as mentioned. We classifed the nft's into classes.
The existing model uses CNN as the base model and achieves an accuracy of 53.74 percent on the test data. Only CNN has been used here by the author and we have compared different models for better result.

	precision	recall	f1-score	support
0	0.66	0.70	0.68	136
1	0.86	0.88	0.87	258
2	0.39	0.20	0.26	135
3	0.36	0.46	0.40	155
4	0.41	0.28	0.34	116
5	0.46	0.54	0.49	113
6	0.38	0.46	0.42	112
accuracy			0.55	1025
macro avg	0.50	0.50	0.49	1025
weighted avg	0.55	0.55	0.54	1025

Figure 6: ResNet

	precision	recall	f1-score	support
0	0.68	0.76	0.72	136
1	0.95	0.87	0.91	258
2	0.42	0.22	0.29	135
3	0.34	0.46	0.39	155
4	0.49	0.32	0.39	116
5	0.42	0.44	0.43	113
6	0.44	0.61	0.51	112
accuracy			0.57	1025
macro avg	0.53	0.53	0.52	1025
eighted avg	0.58	0.57	0.57	1025

Figure 7: MobileNet

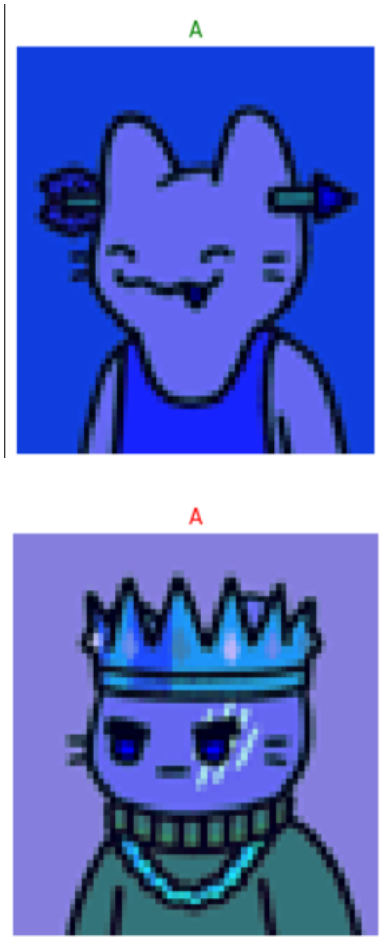
Model (Base)	Test Accuracy
Existing Model	53.74
Our Model	53.46

9 CONCLUSION

Our goal with this project is provide direction for Application of Deep Learning Techniques in NFT and Blockchain space. We chose to include baseline Convolutional Neural Network(CNN), VGG-16 and MobileNet V2 as part of this project. We find out that MobileNet V2 gave best performance as compared to baseline CNN and VGG19. Given the amount of data we manually scraped from Opensea Marketplace, we were able to predict the actual price range of NFT.

	precision	recall	f1-score	support
0	0.68	0.76	0.72	136
1	0.92	0.89	0.91	258
2	0.40	0.47	0.43	135
3	0.42	0.10	0.17	155
4	0.53	0.30	0.38	116
5	0.40	0.53	0.46	113
6	0.41	0.77	0.53	112
accuracy			0.58	1025
macro avg	0.54	0.55	0.51	1025
weighted avg	0.59	0.58	0.56	1025

Figure 8: VGG19



Current NFT landscapes can benefit from such a project but there is scope for future improvements. For example, Many NFT collections have celebrity status attached to them. In such a scenario, we must include possible social network sentiments in account, predicting the prices and rarity of individual NFT collections. Although prediction based on social network sentiments requires a much bigger timeframe for data collection. We chose not to include it in our project. This overview provides a starting point for the community of DL being interested in the field of DL and NFT. Furthermore, more research would be allowed to decide the more suitable direction of work to be taken in order to provide more accurate alternatives to our study.

10 REFERENCES

[1]. Nadini, M., Alessandretti, L., Di Giacinto, F. et al. Mapping the NFT revolution: market trends, trade networks, and visual features. Sci Rep 11, 20902 (2021). https://doi.org/10.1038/s41598-021-00053-8

[2]. Chohan, Usman W. and Chohan, Usman W., Non-Fungible Tokens: Blockchains, Scarcity, and Value (March 24, 2021). Critical Blockchain Research Initiative (CBRI) Working Papers, 2021 , Available at SSRN: https://ssrn.com/abstract=3822743 or http://dx.doi.org/10.2139/ssrn.3822743

[3]. Arnav Kapoor and Dipanwita Guhathakurta and Mehul Mathur and Rupanshu Yadav and Manish Gupta and Ponnurungam Kumaraguru, TweetBoost: Influence of Social Media on NFT Valuation. https://doi.org/10.48550/arXiv.2201.08373

[4]. Y. Kaneko, "A Time-series Analysis of How Google Trends Searches Affect Cryptocurrency Prices for Decentralized Finance and Non-Fungible Tokens," 2021 International Conference on Data Mining Workshops (ICDMW)

[5]. Vasilj, Matea Skender, Sven Horvat, Marko. (2021). Model for improvement of nonprofit student organizations funding by machine learning and digital transition to blockchain.