Deep Learning for Dynamic Price Prediction of NFTs

Mingxuan He *

December 6, 2023

Abstract

In this proposal, I outline the motivation, data, methods, and preliminary results of my project. The goal of this project is to predict the price of NFTs using classical machine learning and deep learning methods. Training and testing data include transactions data from the Ethereum blockchain via Dune Analytics, and NFT collection data from OpenSea, the largest NFT marketplace. The results will be presented in a paper and a presentation. The code will be written in Python utilizing packages including scikit-learn and tensorflow.

Keywords: NFTs, machine learning, onchain data, predictive modeling

I Introduction

A non-fungible token (NFT) is a digital asset stored on a blockchain. The uniqueness of each NFT asset is verifiable through the unique identification assigned by the blockchain on which they exist. NFTs can represent photos, videos, audio, and other types of digital files. NFTs are most commonly bought and sold with cryptocurrencies via digital marketplaces, with the Ethereum blockchain being the most popular market.

Recently, picture NFTs have emerged as a popular investment vehicle. For example, the NFT of a digital artwork by Beeple was sold for $69 million in March 2021. The NFT market has grown rapidly since 2019, with the total trading volume exceeding $12 billion in Q1 2022 alone.

Similar to their fungible counterpart (cryptocurrencies), a picture NFT can vary from a few dollars to millions of dollars. The price of a picture NFT is determined by a variety of factors, most notably the possession of "rare traits". However, similar to the traditional arts market, most NFT prices are also heavily dependent on market taste and sentiment, such as the current popularity of the collection and/or the artist.

In this paper, I apply machine learning methods to predict the price of picture NFTs. Specifically, I train models on market prices and rarity data obtained from the Ethereum blockchain. The goal is to estalish a model framework that provides accurate estimates of NFT prices. Such a model has various applications, including but not limited to onchain trading and bidding contracts, risk management for crypto portfolios, and fair valuation of staked/collateralized NFTs for decentralized finance

*mingxuanh@uchicago.edu, Chicago, IL, United States of America

(DeFi) applications. In addition, the model can be used to price large batches of NFTs quickly, which is useful for AI-generated NFTs.

The rest of this paper is structured as follows. Section **??** discusses the existing literature on this topic and the novelty of this paper. Section **??** provides an overview of the data used in this project. Section **??** outlines the general methodology and the machine learning models used in training. Section **??** presents the results. Section **??** draws conclusions.

II Literature Review

The fast-growing market of NFTs has attracted interest from various academic disciplines. In the economics and finance literature, various authors have discussed the price dynamics of the NFT market. A comprehensive work by Nadini et al. (2021) mapped out important statistical features of NFT markets, revealing that mean prices, sales per asset, sales per collection all follow power-law distributions. Ante (2022) found that NFT sales are triggered by price shocks in Bitcoin (BTC) and active NFT wallets are reduced by price shocks in Ether (ETH). Similarly, Dowling (2022) found co-movement between cryptocurrencies and NFTs through wavelet coherence analysis. On the buyer side, Kong and Lin (2021) that well-connected and experienced investors generally pay lower prices for NFTs. These results support the general belief that the NFT market is not fully efficient, and there is room for improvement in pricing NFTs.

The literature agrees on that rarity is one of, if not the most important factors in determining the price of NFTs. Using data from the CryptoPunks collection, Kong and Lin (2021) built a hedonic regression model and highlighted rarity as a key determinant of price premium in the cross-section. Mekacher et al. (2022a) used a custom-built rarity score and showed that rarer NFTs sell for higher prices, are traded less frequently, guarantee higher returns, and are less risky. Taking a network clustering approach, Nadini et al. (2021) extracted vector representations of the visual features of NFTs and analyzed their cosine distance network using the pre-trained convolutional neural network AlexNet. These results inspired the inclusion of trait features and the OpenRarity rarity rank as a key set of features in my model.

In the applied machine learning literature, there have been notable attempts to apply regression models and neural networks to predict the price of NFTs. Nadini et al. (2021) used a linear regression model with network features such as the buyer's and seller's degree and PageRank centrality, principal components of the NFT's visual features, as well as market data such as the past median price of primary and secondary sales.

The literature also identified other features such as search trends (Jain et al., 2022; Kaneko, 2021), social media influence e.g. Twitter (Kapoor et al., 2022).

III Background & Data

3.1 Institutional background

The NFT market is an emerging market with unique aspects. In this section, I provide a brief overview of the background of NFT markets, including relevant terminology.

- **NFT Marketplaces:** NFT Marketplaces are platforms built with automated blockchain contracts (known as "smart contracts") to facilitate transactions using cryptocurrencies. Most popular marketplaces include OpenSea, Blur, Rarible, SuperRare, etc.
- **NFT Collections:** Many NFTs are created in collections, in which all NFTs share a common theme. For example, the CryptoPunks collection consists of 10,000 unique pictures of pixelated faces.
- **Traits:** Each NFT in the collection has a unique combination of traits, including color, background, face/expression, and accessories. The traits typically differ by rarity. For example, there are more than 2,000 CryptoPunks with the trait "Earring", but only 44 with the trait "Beanie".
- **Rarity Score & Rarity Rank:** The current industry standard for calculating the rarity of an NFT within a collection is the OpenRarity Standard [1], where the rarity of an NFT is evaluated on the rarity of its traits. The calculations for this metric is outlined in Appendix **??**. By comparing the rarity score of all NFTs in a collection, we can rank the rarity of each NFT, with 1 being the rarest.

3.2 Data sources

I obtained data from two sources: Dune Analytics and OpenSea.

3.2.1 Dune Analytics

Dune Analytics is a platform for querying public databases from the Ethereum blockchain. For each collection, I query the NFT trades database and gather information on all NFT transactions from issuance to Sep 30, 2023.

3.2.2 OpenSea

OpenSea is the largest NFT marketplace focusing on NFTs based on Ethereum and Ethereum's Layer-2 ecosystem. The data from OpenSea's public API include data on NFT collections. In particular, I extracted data on the traits and rarity of each unique NFT in the collections. The set of available traits varies by collection, but generally include color, background, face/expression, and accessories.

3.3 Features and target variables

Here I outline the features and target variables used in this project. The target variable is the price of the NFT. The features are divided into four categories: market, base traits, extra traits, and last trade. The features are summarized in Table **??**.

[1] https://www.openrarity.dev/

Feature Category	Feature	Description
Market	volume_eth	daily market volume (ETH)
	price_p5_eth	daily 5-percentile price (ETH)
	price_max_eth	daily highest price (ETH)
	price_min_eth	daily minimum (floor) price (ETH)
Base Traits	rarity_rank	rarity rank measured by OpenRarity
	Background_count	number of items with the same *Background* trait
	Mouth_count	number of items with the same *Mouth* trait
	Eyes_count	number of items with the same *Eyes* trait
	Fur_count	number of items with the same *Fur* trait
Extra Traits	Hat_count	number of items with the same *Hat* trait
	Clothes	number of items with the same *Clothes* trait
	Earring_count	number of items with the same *Earring* trait
Last Trade	last_trade_timediff	time since the last time this item was traded
	last_trade_price	price at the last time this item was traded

Table 1: Feature Columns Description (Bored Ape Yacht Club)

3.4 Exploratory analysis

3.4.1 Price history

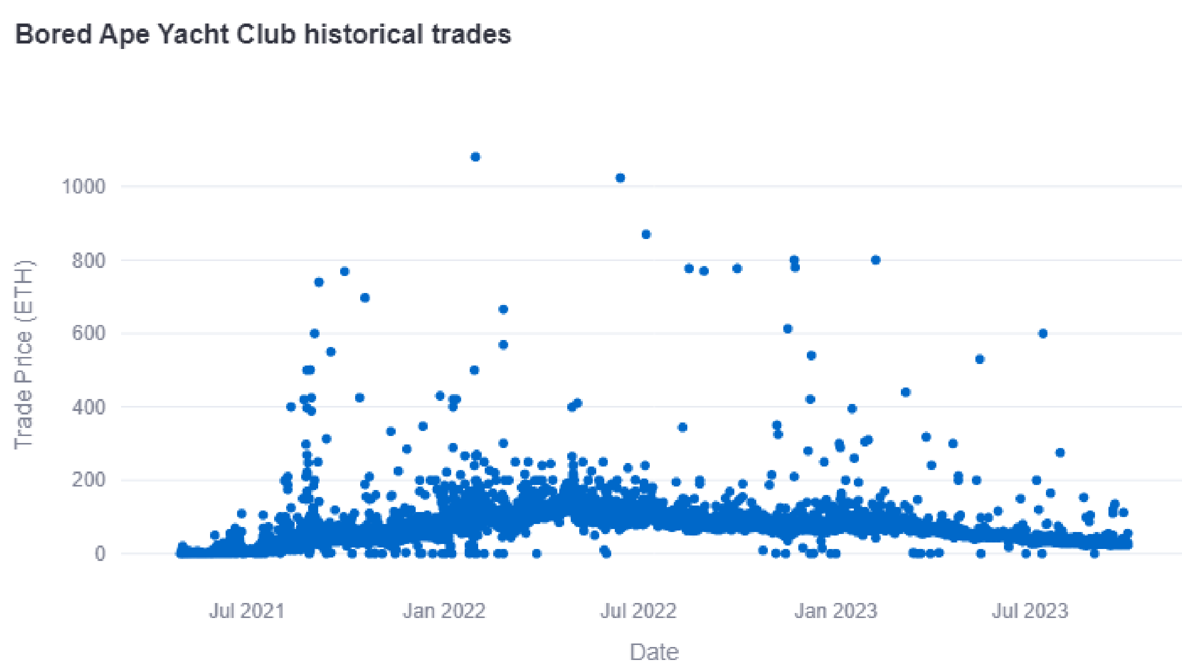

Figure 1: Trade Price by Date

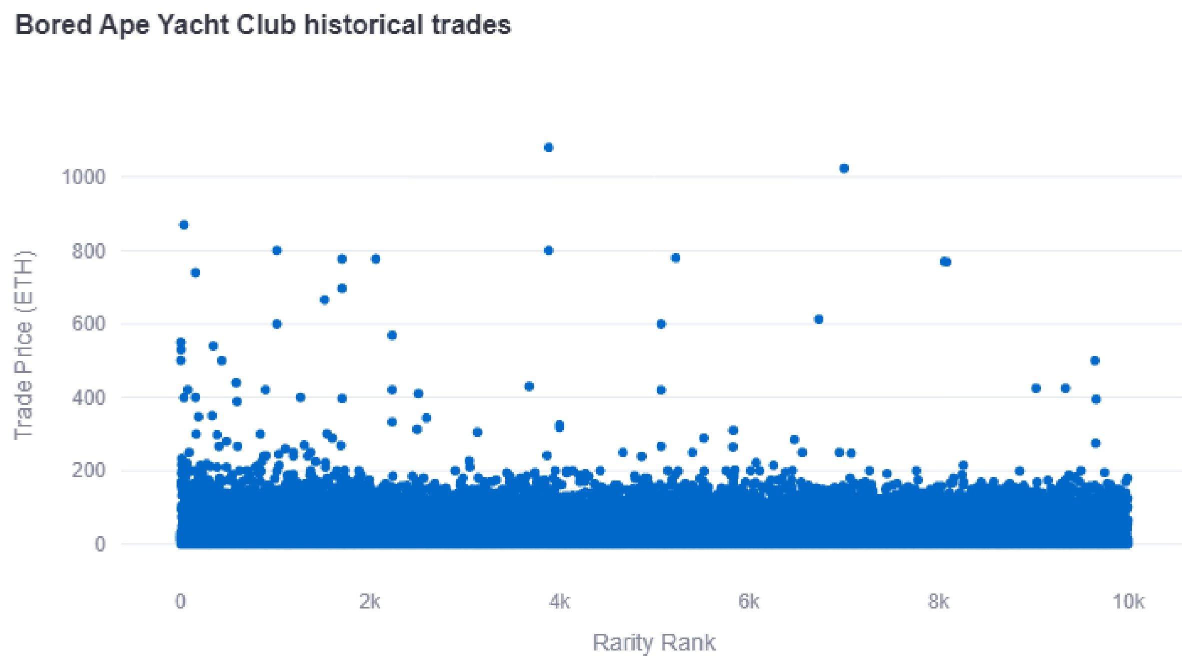

Figure 2: Trade Price by Rarity Rank

3.4.2 Principal component analysis

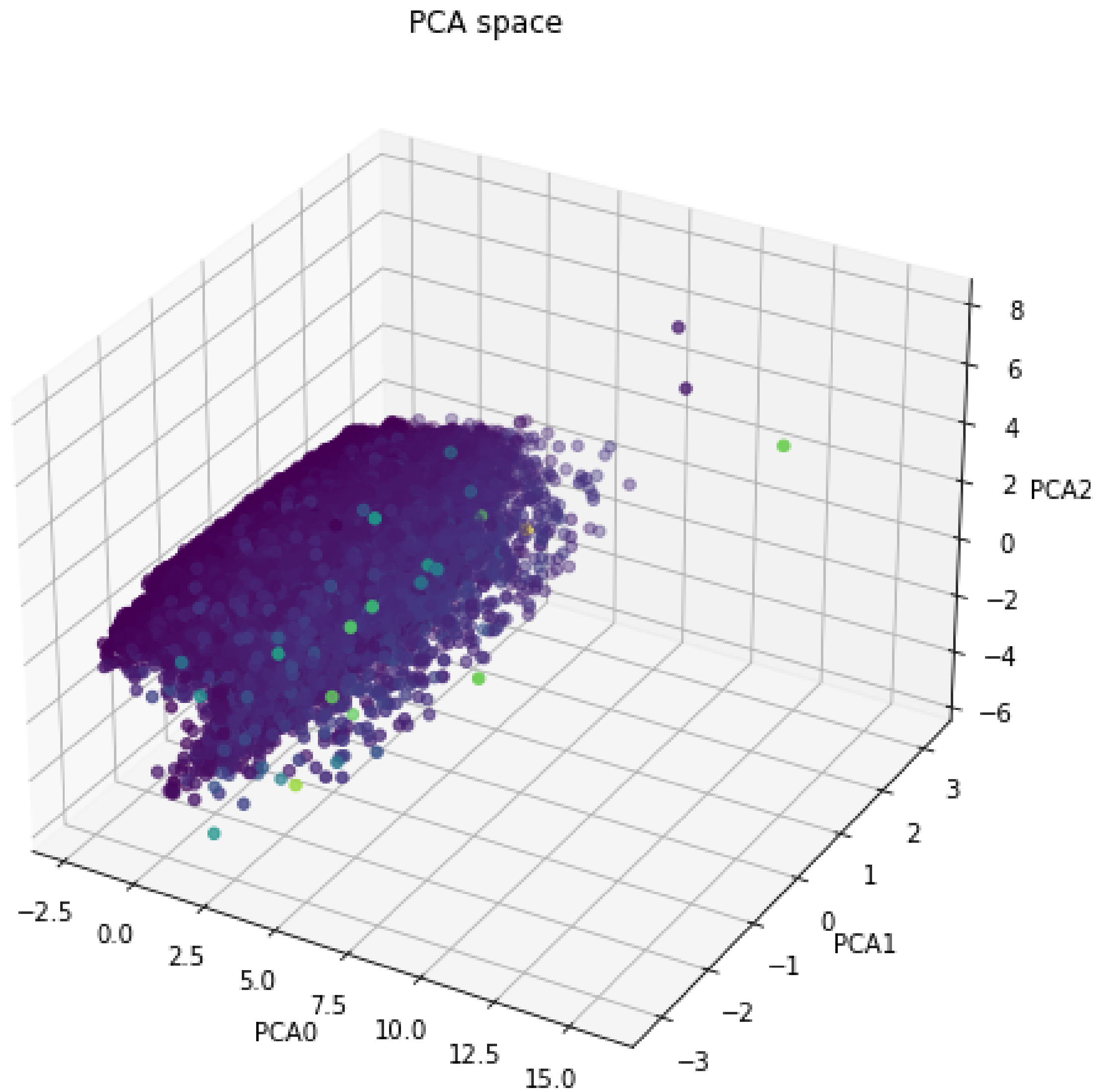

Figure 3: PCA(3) plot of NFT prices (Bored Ape Yacht Club)

IV Methodology

For benchmarking, I will use an OLS model. For the main model, I will use a neural network with a fully connected layer. I will also explore other deep learning methods, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs). I choose Mean Squared Error (MSE) as the objective function to minimize. This is currently work in progress.

V Results

I present my preliminary results.

5.1 Benchmarking

MSE metric for benckmark models:

Model	**MSE**
OLS	455.8
Random Forest	220.1

References

Alessandretti, L., ElBahrawy, A., Aiello, L. M., & Baronchelli, A. (2018). Anticipating cryptocurrency prices using machine learning. *Complexity*, *2018*, 1–16.

Ante, L. (2022). The non-fungible token (nft) market and its relationship with bitcoin and ethereum. *FinTech*, *1*(3), 216–224.

Branny, J., Dornberger, R., & Hanne, T. (2022). Non-fungible token price prediction with multivariate lstm neural networks. *2022 9th International Conference on Soft Computing & Machine Intelligence (ISCMI)*, 56–61.

Costa, D., La Cava, L., & Tagarelli, A. (2023). Show me your nft and i tell you how it will perform: Multimodal representation learning for nft selling price prediction. *Proceedings of the ACM Web Conference 2023*, 1875–1885.

Dowling, M. (2022). Is non-fungible token pricing driven by cryptocurrencies? *Finance Research Letters*, *44*, 102097.

Henriques, I., & Sadorsky, P. (2023). Forecasting nft coin prices using machine learning: Insights into feature significance and portfolio strategies. *Global Finance Journal*, 100904.

Horky, F., Rachel, C., & Fidrmuc, J. (2022). Price determinants of non-fungible tokens in the digital art market. *Finance Research Letters*, *48*, 103007.

Jain, S., Bruckmann, C., & McDougall, C. (2022). Nft appraisal prediction: Utilizing search trends, public market data, linear regression and recurrent neural networks. *arXiv preprint arXiv:2204.12932*.

Kaneko, Y. (2021). 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)*, 222–227.

Kapoor, A., Guhathakurta, D., Mathur, M., Yadav, R., Gupta, M., & Kumaraguru, P. (2022). Tweetboost: Influence of social media on nft valuation. *Companion Proceedings of the Web Conference 2022*, 621–629.

Kong, D.-R., & Lin, T.-C. (2021). Alternative investments in the fintech era: The risk and return of non-fungible token (nft). *Available at SSRN 3914085*.

Mekacher, A., Bracci, A., Nadini, M., Martino, M., Alessandretti, L., Aiello, L. M., & Baronchelli, A. (2022a). Heterogeneous rarity patterns drive price dynamics in nft collections. *Scientific reports*, *12*(1), 13890.

Mekacher, A., Bracci, A., Nadini, M., Martino, M., Alessandretti, L., Aiello, L. M., & Baronchelli, A. (2022b). How rarity shapes the nft market. *arXiv preprint arXiv:2204.10243*, 9.

Nadini, M., Alessandretti, L., Di Giacinto, F., Martino, M., Aiello, L. M., & Baronchelli, A. (2021). Mapping the nft revolution: Market trends, trade networks, and visual features. *Scientific reports*, *11*(1), 20902.

A Rarity Score and Rarity Rank

I use OpenSea's OpenRarity standard to calculate the rarity of each NFT within its collection. The rarity score is defined as follows:
For an NFT x with traits $i \dots n$, its rarity score is

$$R(x) = \frac{I(x)}{\mathbb{E}[I(x)]}, \text{where } I(x) = \sum_{i=1}^{n} -\log_2 P(trait_i).$$