NFT Floor price prediction

Sharath Giri (SG153) and Anitesh Reddy (AS361)
May 3, 2023

Abstract

Usage of Machine Learning technique have been used in stock price predictions for a few years. However, very limited work is done in predicting NFT floor prices because of its volatile price nature. In this project, we focus on NFT floor price predictions which can be used by crypto hedge funds to lend popular crypto currencies in open market. Focus of the project is to improve the accuracy of price prediction by taking into account not only the NFT floor price trends but also macro-economy trends, sentiment trends of the collection and other market features. Brief conclusion to be added

1. Introduction

NFT lending ecosystem has grown to close to USD 23 Million in the beginning of 2023 [1] overall, this space is expected to grow exponentially in the coming five years. With significant increase in market cap, for the ecosystem to remain sustainable it would require robust collateral future price predictions. Primary focus is to get accurate future floor price predictions for 7 Days, 14 days and 28 days out. These are the typical window of operations for hedge fund managers.

1.1. Broad Overview

Our hypothesis, NFT floor price is heavily influenced by the market sentiments on that particular collection and the base cryptocurrency trends (Bitcoin, Ethereum and Polygon) along with the other key financial indicators and NFT collection features. High level break-down of the project would be:

- Data Gathering:
  - Read block-chains to gather trends of NFT floor price, percentage NFTs listed from the collection, number of unique owners, percentage of listings from whales, Buys/sell ratio, transaction trends, spread trends
  - Discord / Twitter discussions
  - Stock market, base cryptos (Bitcoin, Ethereum, Polygon), other macroeconomic factor trends
- Project will explore various models such as LSTM, GRU, CNN and GNN to predict the future price points. Idea here is to get the best accuracy and also improve these models by boosting mechanism. (ARIMA)
- Most of the current work is only focused on top 10 NFT collections as they are relatively more stable. We want our model to be able to predict even for collections with low NFT floor price and low buy/ listing ratio

1.2. Why Accurate Pricing of NFTs Matters

Extreme price and volume fluctuations are common occurrences within the NFT market. For example, Bored Ape Yacht Club's floor price went from 102.3 ETH on 1 April 2022, to 139 ETH in a short span of 30 days; a 35.9% increase. However, by the end of May 2022, this floor price plummeted to 83.6 ETH; a 39.9% drop. In a similar period, BAYC #8537 made a profit of 70.7 ETH within the span of 25 days, a 45.7% increase in the owner's purchase price.[2]. These fluctuations results in very inefficient NFT lending ecosystem.

2. Related Work

- There is a prediction model released using time-series ARIMA model which only takes into account recent. It has the ability to use past prices to predict future prices, using components such as auto-regressive terms (past prices), non-seasonal differences of the past prices, and the lagged forecast errors.[3]
- Another ARIMA model using only the historic floor price trends which produces high accuracy for next day prediction using only floor price. [4]
- Nansen Price Estimates v2.0 model, which uses NFT transaction data as well along with floor price. Still does not use other attributes. [2]

3. Goal & Data Sourcing

Goal & Hypothesis: We believe there are other attributes of a NFT collection make an impact on its floor price along with other financial indicators. We want to capture these correlations using various models and improve the forecast which has not been done yet. We are gathering our data from various sources as listed below:

Required Datasets		Status	Source
NFT collections		Completed	NFTfloorprice.com
NFT KPIs	Floor price	Completed	NFTfloorprice.com
	Buy/sell Ratio	In Progress	Covalent
	% of unique Owners	In Progress	Covalent
	Transaction trend	Completed	NFTfloorprice.com
	Listing Ratio	In Progress	Covalent
	Whale Listing	In Progress	Covalent
Base Crypto Currency trend	Polygon	Completed	Investing.com
	Bitcoin	Completed	Investing.com
	Ethereum	Completed	Investing.com
Stockprice trends	US - S&P500	Completed	Investing.com
	India -BSE30	Completed	Investing.com
	Japan -Nikkei225	Completed	Investing.com
	UK - FTSE100	Completed	Investing.com
	China - Proxy - 1322	Completed	Investing.com
Macroeconomic trends	Quaterly GDP - US	Completed	
	TIPS - US Bond	Completed	Investing.com
	Gold	Completed	Investing.com
	Monthy Inflation - US	Completed	
Twitter Sentiment		In Progress	Twitter
Discord Sentiment		In Progress	Discord

Figure 1. Data sourcing and status

There is a lot of pre-processing that needs to be done on this data as they are collected from various sources. Firstly, we don't get data for all the dates, so modelling decision is to be taken in regards to what is to be done with missing data. A lot of the NFT collection data points are corrupted, such as listing ratios: It is a common practice for owners of NFT to list there NFT as a very high unreasonable price and they leave the NFT listing there very well knowing the fact it won't be bought. We need to ensure such corrupt data points are removed. Please do note, since stock market data points are not available during weekends and holidays we have interpolated the data instead of deleting the data points. We have listed below few uncommon data points along with its description and hypothesis around why we believe these features can impact NFT floor price:

Trend	Description & Hypothesis
NFT Floor Price	Description: Trend of the lowest price of NFT listed from the collection.
% of NFTs listed	Description: % of NFTs listed from the over all collection Hypothesis : If this % grows it shows that the confidence on the NFT is coming down, bear market for the collection
% of unique Owners	Description: % of unique Owners from the over all owners Hypothesis: If limited owners are owning majority of the NFT price, they can easily influence the market price
% of Listings from Whales	Description: % of NFTS listed from the sub collection of Whale owners Hypothesis: If someone who is owning quite a few NFTs from the collection and he/she is listing the NFT, expectation is that floor price will go down
Buy/ Sell Ratio	Description : Ratio of Buy postings/ Sell Postings Hypothesis: higher the value more bullish the market is and hence the NFT floor price should go up
Spread trend	Description : Trend of NFT price spread in a collection Hypothesis: is the spread is high the impact of other NFT features should play a smaller role on floor price movement
Transaction trend	Description : No of transactions made in a day Hypothesis: Higher this value more activity in the market and floor price can change faster

Figure 2. Key feature description

4. Model

We initially started work with one NFT collection (BAYC) and used LSTM with two layers and built and model. This produced very poor results. When we shifted to GRU as expected the results have turned out to be better. Following this, we used our learning from assignment 3 and introduced feature engineering and added floor price (power of (2,3 & 5), sin, cos and exponential) values to our data. Introduction of these features have increased the accuracy for next day prediction tremendously. (96% prediction with 10% error). Since this model seemed to be reasonable, we built our initial model architecture as shown in figure 3.

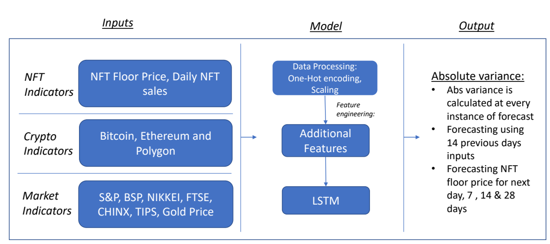

Figure 3. First Model Architecture

As an experiment, instead of concatenating all the inputs as vector representation, we plan to use GNN to figure out input representation. Idea is to represent each day as a graph and all the input attributes of the collection as its nodes. Firstly, we find out correlation between all the input attributes to figure out the edges to be assigned in the graph. Figure 4 represents the overall model architecture.

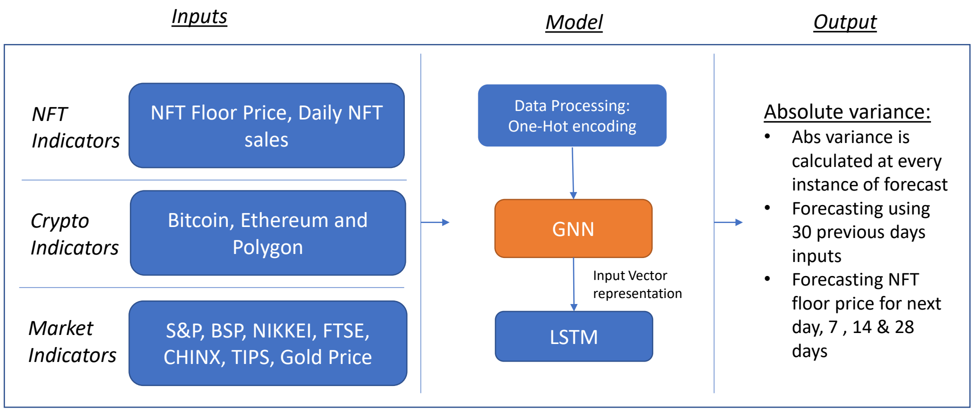

Figure 4. GNN + LSTM Model Architecture

First, we did a correlation analysis to figure out what should be the graph architecture, as shown in figure:5 . Then we used networkX library to build the graph, as shown in figure:6. After which these graphs where trained using Node2Vec with dimensions=64, walk length=30, num walks=200. One these vector representations were learned, they were fed to the same GRU model as before.

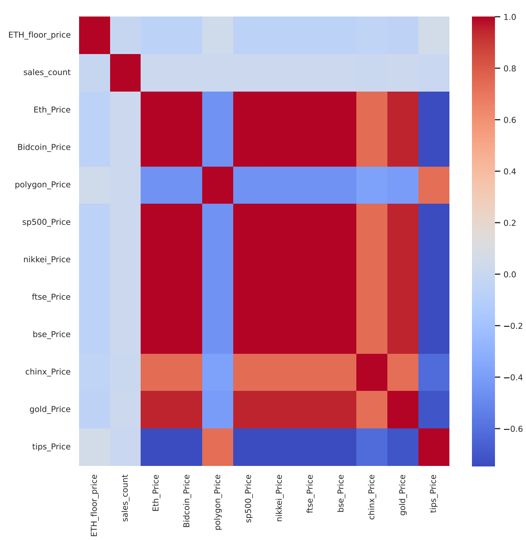

Figure 5. Correlation matrix of the variables

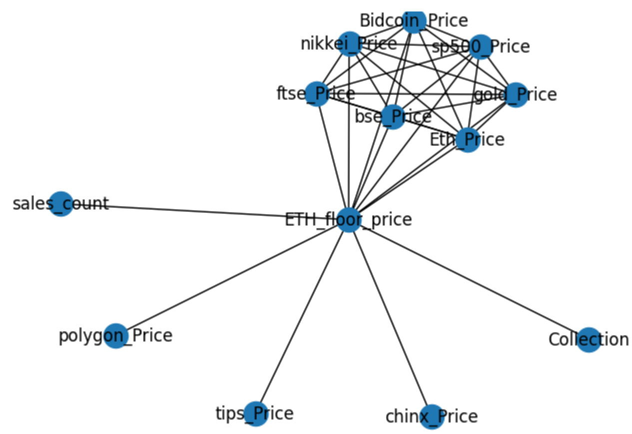

Figure 6. Graph structure used to learn the node embedding

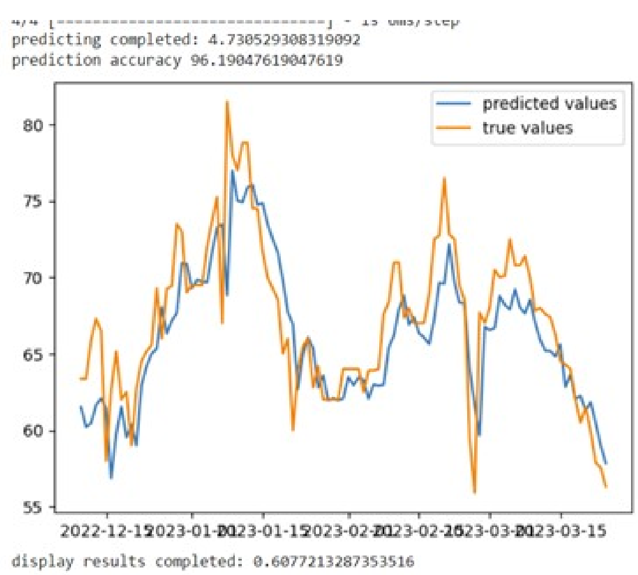

Figure 7. Initial Results of running only BAYC on GRU

5. Experiments & Results:

As mentioned earlier, we initially have only tested LSTM and GRU for BAYC data-set for next day prediction which is resulting in 96% accuracy. As shown in Figure 7:

Using this as a bases we built our first model and the results from the first model to predict price after 14 days are shown in figure 8 for various collections.

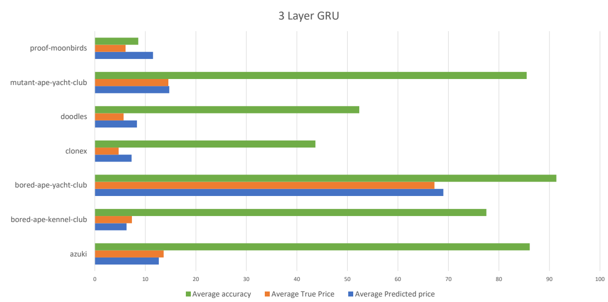

Figure 8. Figure displays results from the first model architecture

We have predicted NFT floor price for next day, future 7th, 14th and 28th day using 3 layer GRU model and the results are compared in figure 10. It is interesting to note that for some unstable NFTs 14 days prediction is better than 7 day prediction. To try and understand if we can solve this issue we experimented with taking last 7 and 30 days as input instead of 14 days.

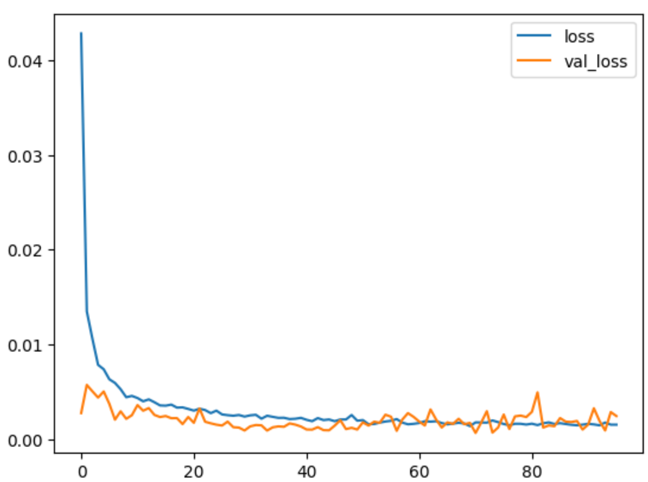

Figure 9. Training & Validation accuracy 30 days rolling for next 14th day prediction for GRU model

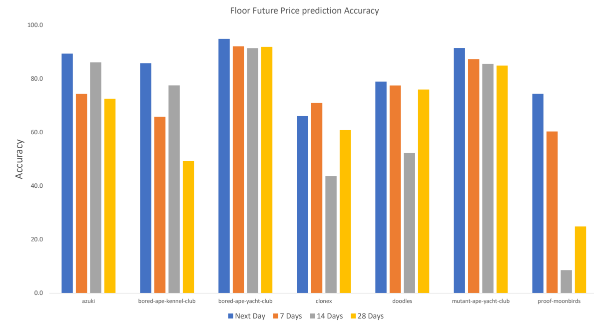

Figure 10. Figure compares results for next day, 7th , 14th and 28th day prediction for GRU model

The results of this experiment are shown in Table 1. Here again, results are not intuitive, if we increase number of past dates used to predict, accuracy does not increase always. But overall the model accuracy increases by 7% by using last 30 days instead of 14 days. So we will use last 30 days rolling window as input to the model.

We have run GCN+GRU model for one collection using Node2Vec embedding, it resulted in 94% accuracy which is similar to that on only GRU. below figure 11 shows the training graph.

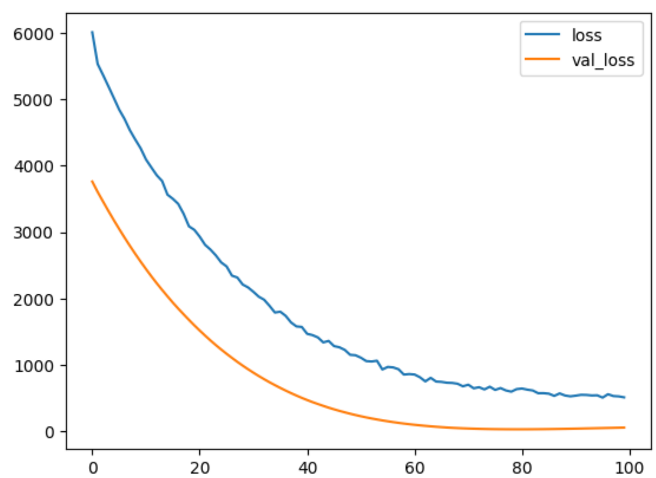

Figure 11. Training & Validation accuracy 30 days rolling for next 14th day prediction for GCN+GRU model

6. Future Goal:

We intend to use other NFT collction attributes such as buy-sell ratio, Whale listing, Unique users, listing ratios and even sentiment analysis from Discord & Twitter and will create a strong forecasting model which captures real time data and has a very user friendly API which used by individuals for lending crpyto currencies with NFT as collateral. Also, instead of using Node2Vec to figure out node embedding, would like to build and train customized GCN model.

NFT Collection	Last 7 Days	Last 14 days	last 30 days
azuki	89%	82%	74%
bored-ape-kennel-club	78%	82%	65%
bored-ape-yacht-club	90%	91%	92%
clonex	34%	45%	70%
doodles	31%	68%	77%
mutant-ape-yacht-club	84%	88%	87%
proof-moonbirds	2%	24%	60%

Table 1. Experimental results showing the accuracy for future 14th day prediction using various past days lengths

We would finally like to create a package product which can be used by market space providers such as X2Y2, Drops, Pine, NFTfi, BendDAO etc to assist their crypto loan products.

7. Learnings & Conclusions:

Definitely we understood gathering and processing the data is a major task. We have spent almost 80% of time on gathering the data and there is still more to do. It was very interesting to see how scaling the features play a major role in predicting accurate results. Merging various datasets is not very trivial due to different units and date formats. Preparing data to feed to an RNN gave us much better understanding of how RNNs work. With various experiments, it was evident that with increase in one-hot encoding size, model parameters have to increase.

References

[1] `https://www.nftgators.com/nft-lending-ecosystem-experiences-rapid-growth-as-platforms-like-benddao-and-nftfi-emerge-as-key-pla #:~:text=Since%20October%202022%2C%20the%20NFT,borrow%20volume%20to%20%2423.4M.`

[2] `https://www.nansen.ai/research/nft-price-estimates-machine-learning-Model#:~:text=In%20February%202022%2C%20Nansen%20launched,available%20for%20any%20given%20NFT`, 2022. Accessed: October 5, 2022.

[3] `https://ethglobal.com/showcase/nft-floor-price-prediction-c1giu`, 2022.

[4] `https://mirror.xyz/elenahoo.eth/ga8iZLiQPeqLBDXgaAPGVa0TUGfzKJqeAKqfCk3vslI`, 2022.