Predictive Analysis of Non Fungible Token Price Using Deep Learning

by

Mohammad Redwan Arefin Rifat
17241020
Tasfim Ahmed Rupok
13101088

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
School of Data and Sciences
Brac University
September 2023

Declaration

It is hereby declared that

1. The thesis submitted is our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Mohammad Redwan Arefin Rifat
17241020

Tasfim Ahmed Rupok
13101088

Co-Supervisor's Full Name & Signature:

A. N. M. Sajedul Alam
BRACU Alumnus
Department of Computer Science and Engineering
BRAC University

Approval

The thesis titled "Predictive Analysis of Non Fungible Token Price Using Deep Learning" submitted by

1. Mohammad Redwan Arefin Rifat (17241020)
2. Tasfim Ahmed Rupok (13101088)

Of Spring, 2022 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on september 17, 2023.

Examining Committee:

Supervisor:
(Member)

Moin Mostakim

Moin Mostakim
Senior Lecturer
Department of Computer Science Engineering
BRAC University

Program Coordinator:
(Member)

Md. Golam Rabiul Alam, PhD
Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Sadia Hamid Kazi, PhD
Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Abstract

A form of digital asset called Non-fungible tokens can represent a wide range of objects, such as pieces of art, collectibles, and in-game items. Non-fungible tokens are also commonly referred to by their acronym, NFTs. They are often kept within smart contracts that are hosted on a blockchain and are traded over the internet, where cryptocurrency is frequently used. The year 2021 has seen a meteoric rise in the acceptance of NFTs, which has resulted in exceptional sales in the market. Despite this, we still have little grasp of the overall structure of this market and how it evolved over time. Within the scope of this investigation, we investigate a dataset that contains 6.1 million transactions that involve 4.7 million non-fungible tokens and runs from 23rd June 2017 to 27th April 2021. The Ethereum and WAX blockchains are the primary sources of this information. Our analysis aims to achieve several objectives. In the first step of this process, we look into the statistical characteristics of the NFT market. In the second step of our process, we build a network that illustrates the relationships between different traders.We have noticed that traders frequently specialize in NFTs that are related with comparable objects, and they typically establish cohesive clusters with other traders who are involved in the trading of similar objects.Thirdly, we use clustering algorithms to organize the items that are associated with NFTs according to the visual qualities that distinguish them from one another. Our findings demonstrate that collections tend to consist of visually consistent objects.Finally, we investigate whether or not NFT sales may be forecasted by utilizing certain fundamental machine learning methods. According to the findings of our investigation, an NFT's sales history and, to a lesser extent, its aesthetic characteristics can each serve as credible predictors of the price of that NFT. We are confident that these realizations will act as a driving force behind additional research on the creation, adoption, and trading of NFTs in a variety of different settings.

Keywords: Gated Recurrent Unit (GRU), Linear Regression Model (LRM), Non-Fungible Token (NFT), Long Short-Term Memory (LSTM), Bored Ape Yacht Club (BAYC), Recurrent Neural Network (RNN).

Acknowledgement

I would like to express my sincere gratitude to my thesis supervisor, Mr. Moin Mostakim, for his invaluable guidance, support, and encouragement throughout the course of this work. His expertise and insights have been instrumental in shaping the direction and outcome of this research. I would also like to thank the member of the thesis committee, Md. Golam Rabiul Alam, Ph.D, my thesis co-supervisor A. N. M. Sajedul Alam for their valuable feedback and suggestions. I am grateful to BRAC University for providing me with the resources and facilities necessary to conduct this research. I would also like to thank my friends and family for their unwavering support and encouragement throughout this process. Last but not least, I would like to express my appreciation to all the participants who generously gave their time and effort to contribute to this research. Without the support and guidance of all of these individuals, this thesis would not have been possible.

Table of Contents

List of Figures

List of Tables

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

LRM Linear Regression Model

$LSTM$ Long Short-Term Memory

NFT Non-fungible tokenl

RNN Recurrent Neural Network

Chapter 1

Introduction

1.1 Introduction

The podcast episode of Art Newspaper entitled "What exactly are NFTs? Why cryptocurrency is taking over the art industry,"[7] transmitted on February 21, 2021, illustrates the considerable impact that Non-Fungible Tokens have had on art industry despite the general public's lack of knowledge with the concept behind NFTs. Nevertheless, the influence of NFTs extends beyond the art market, as multiple industries, such as digital content production including audio and video, investigate the potential of this technology. Notably, in the first four months of 2021, NFT volume surpassed 2 billion USD. A noteworthy tenfold increase compared to the entire 2020's NFT trading volume. So NFTs—what are they? It is a unit of information kept in a blockchain that functions as an unchangeable certification for a digital asset. NFTs offer a distinctive digital proof of ownership and the "provenance" of a digital asset responds to questions regarding ownership, previous ownership, creator attribution, and identifying the original among multiple copies. Digital media such as images, movies, and music can all have NFTs attached to them. Current applications include the commercialization of digital artifacts in the fields of gaming, sports and art collectibles. Initially, NFTs were primarily associated with the Ethereum blockchain; however, other blockchains have developed their own variants. The earliest and best-known example of non-fungible tokens is CryptoKitties, an accumulation of images of virtual cats used in an Ethereum-based game. Due to the extreme popularity of the cryptocurrency-themed game CryptoKitties, the Ethereum network became extremely congested in December 2017.[3] However, the NFT market started to gain traction in July 2020,[20] and it received a lot of press in March 2021 when an NFT of an artwork by the artist known as Beeple sold at Christie's for a record-breaking $69.3 million.[11] After Jeff Koons and David Hockney, this auction brought in the third-highest price ever for a work by a living artist.[10] Several subsequent deals set new benchmarks,including the selling of three Cryptopunks, each containing 10,000 distinct digital characters, for millions of dollars.[9] Furthermore, the very first tweet was purchased for $2.9 million. Also, an NFT titled "Auction Winner Picks Name" with a music video and performance recording was purchased for $1.33 million. Famous people have generated their very own NFTs since they are profitable, and collectibles from NBA and well-known football players fetch several hundred thousand dollars.[6] Studies on non-fungible tokens remain limited, emphasizing technical details such as components, copyright laws,[4] protocols, desired characteristics, standards and blockchain-based protocols for identifying tangible products and the influence that non-fungible tokens have had on the creative industry particularly the ability to share royalties with artists from secondary sales. As NFTs continue to reshape numerous industries, additional research is required to comprehend their potential and influence. In the past, empirical studies of the NFT market have narrowly focused on specific NFT markets, such as SuperRare[5] and Decentraland[14] or on specific NFT collections, like Axie,[15] CryptoKitties[12] and Cryptopunks. According to these studies, the value of digital games has decreased because of the ubiquity of NFTs.[13] The NFT marketplace is also vulnerable to speculation[14] because it has been noticed that the prices of NFTs are affected by the prices of cryptocurrencies.[15] Expert valuations of non-monetary transactions are associated with a higher likelihood of a successful outcome, according to the research.[5] The co-ownership network of NFTs is extremely centralized and exhibits small-world-like traits according to a study depending on the sales of 16,000 NFTs on the SuperRare marketplace.[1] The purpose of this paper is to provide a quantitative overview of the NFT market. To accomplish this, we analyze a large dataset containing 6.1 million transactions of 4.7 million NFTs across 160 cryptocurrencies, most notably Ethereum and WAX. The range of the data set is from 23rd June, 2017 to 27th April, 2021. We commence our analysis by examining the market's overarching statistical characteristics and tracing its development over time. Next, we conduct an in-depth analysis of the interconnections between NFT traders and the NFT asset network. In addition, we use clustering techniques based on visual characteristics to classify NFTs. Lastly,here, we show the outcomes of regression and classification models developed to foretell secondary markets for NFTs and their associated pricing. Non-Fungible Corporation,[8] that tracks NFT sales, and OpenSea,[19] the largest NFT marketplace, have proposed a system of categorization that we use to assure a full study, and we apply this system manually. However, it is essential to observe that the precise classification of NFTs used in different contexts is beyond the scope of this paper. Certain works of art can be classified as mementos, whereas certain objects from games can be classified as works of art due to their pleasing aesthetic and cultural aspects. In quantitative finance, one of the main challenges is Price prediction. The primary focus of this paper is NFT price prediction, but a machine learning approach using a neural network framework is also presented as a solution to the price prediction problem. In two instants, the framework is executed.In this study, two neural network types were employed to predict prices: a fundamental Recurrent Neural Network (RNN) and a Long Short-Term Memory (LSTM) network capable of handling longer dependencies. The aim is to improve prediction accuracy by implementing neural network theory and LSTM networks.

1.2 Problem Statement

Non-Fungible Tokens (NFTs) are a relatively new development that have had a profound impact on the digital asset environment. These tokens have made it possible for investors, collectors, and artists to take advantage of previously unimaginable prospects. However, due to the very unpredictable nature of NFT prices, those who wish to successfully navigate this market have a significant obstacle. This difficulty is the motivation behind the development of the main objective of this thesis, which is to create a predictive analysis framework that makes use of deep learning techniques in order to forecast NFT pricing with increased accuracy and dependability. The issue at hand is that there are not enough reliable prediction models available that are able to precisely project what the value of NFTs will be in the future. Existing methods for predicting the price of non-fungible tokens frequently rely on standard statistical methods or fail to grasp the complex underlying patterns and dynamics that drive price movements of non-fungible tokens. Additionally, the nature of the NFT market, which is always shifting in response to the rapid expansion of the industry, necessitates a complex methodology that is able to adjust to shifting trends and utilize a variety of data sources. This study will employ the potential of deep learning algorithms, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to analyze large-scale datasets containing non-fungible token (NFT) transactional records, metadata, social media sentiment, and other pertinent factors. This will allow the researchers to circumvent the limitations that have previously been identified. The proposed methodology tries to capture the deep correlations that exist between a variety of predictors and NFT price variations by making use of the temporal dependencies and spatial features that are embedded within the data. In addition, the thesis will investigate the potential for enhancing the prediction capabilities of the model through the incorporation of additional data sources from the outside world, such as macroeconomic indicators and market sentiment. The newly established framework will aim to provide a more comprehensive perspective of the NFT market in order to increase the accuracy of price forecasts. This will be accomplished by combining the extra criteria listed above. Numerous parties, such as artists, collectors, and investors, stand to gain major benefits from the successful development and deployment of an accurate price-predictive model for NFTs that is based on practical applications of deep learning. The suggested framework has the ability to inform investment decisions, optimize pricing strategies, and assist in risk management within the NFT market by offering more reliable insights into prospective price movements. In conclusion, this thesis tackles the need for a dependable and versatile predictive analysis framework that makes use of deep learning methods to effectively estimate NFT pricing. The purpose of this project is to produce a useful tool that will assist users in navigating the dynamic and complicated environment of the NFT market by identifying the underlying patterns and correlations that exist within NFT data, combining multiple data sources, and including information that is both temporal and spatial.Non Fungible Token (NFT) market is greatly diversified. 1 out of 100 NFTs are sold for more than a thousand USD but rest are sold only a few cents or more. NFT market places are dynamic, fast-moving and volatile in nature. To understand the NFT market places properly we have to be able to respond to changes in the market quickly. So, for individual financial security and for making investment, forecasting future values of NFTs are

important. We have seen that a JPEG was sold for $69 Million. So how do we use machine learning models to accurately predict the prices of various NFTs? In this report we have tried to figure it out. We analysis correlations between NFT valuations and their attributes from 3 primary categories:
1.The data of Public Market,
2.Metadata of NFT,
3.The data of Social Trends In order to establish connections to more conventional investing classes, we used these data sources. Using search engine and social networking trends, we also try to quantify relevant information, such as sentiment and engagement.

1.3 Research Objectives

This thesis's primary objective is to utilize the strength of deep learning techniques to develop an accurate and robust predictive model for forecasting the prices of Non-Fungible Tokens (NFTs). In recent years, NFTs have garnered considerable attention as unique digital assets that can represent ownership of digital or physical objects in a variety of domains, including art, gaming, collectibles, and virtual real estate. NFT markets are highly volatile, making it difficult for investors, collectors, and other market participants to accurately predict future price movements. The purpose of this study is to examine past data relating to many components of NFT transactions in order to find a solution to this problem. These aspects include: NFT features; transaction volume; creator reputation; rarity; and external influences including market sentiment and economic indicators. Market trends, correlations, and crucial features in the NFT space will be revealed by deep learning algorithms. In order to capture the intricate nonlinear links and dynamic patterns that drive NFT prices, this research utilizes historical data to train the prediction model. This research intends to capture the complex nonlinear linkages and dynamic patterns that influence the price dynamics of NFTs by training the predictive model using historical data. In addition, this study will investigate the possibility of incorporating additional relevant data sources, such as social media sentiment and market news, to improve the predictive model's accuracy and robustness. By considering a comprehensive set of features and employing advanced deep learning architectures, the study seeks to provide an in-depth knowledge of the NFT market's dynamics and improve the accuracy of price forecasts. The outcome of this research will have significant ramifications for a variety of NFT ecosystem stakeholders. Investors and collectors can use the predictive model to make informed decisions regarding the purchase, sale, and retention of NFTs, thereby potentially maximizing investment returns. In addition, platforms and marketplaces dealing with NFTs can optimize their pricing strategies and improve the user experience with the knowledge obtained. This study contributes to the broader field of financial analysis by employing cutting-edge deep learning methodologies to the emerging domain of non-fungible tokens, paving the way for enhanced market insights and strategies in the digital asset space. To achieve these goals, the research will acquire and preprocess a substantial quantity of historical NFT data from reputable sources. The predictive model will be trained and evaluated using advanced deep learning technique, named recurrent neural networks (RNNs). The performance of the proposed deep learning-based model will be evaluated using appropriate evaluation metrics, and comparisons will be made with extant forecasting approaches to demonstrate its efficacy and superiority. By conducting this research, we hope to advance the academic and applied comprehension of NFT price dynamics, advance the application of deep learning in financial analysis, and facilitate more informed decision-making in the emerging and rapidly evolving NFT market. Our goal is to find out more efficient way of Machine Learning to predict NFT prices with great accuracy. We have multiple options to choose from when it comes to deep learning. We fed the data through the various machine learning models with each having individual goals and perspectives. To determine the characteristics that are most closely correlated with NFT's valuation, we employed a linear regression model. According to J. Brownlee, (2016) [2] another justification for using the linear regression model was to find redundant features. In an effort to forecast future NFT prices, we also looked into several RNN iterations, such as LSTM, GRU and Multivariate variants.

Chapter 2

Literature Review

Since the NFT space is still relatively new, only a small quantity of previous research has been conducted in this field. In addition, journals with peer reviews for the social aspects of the blockchain (not cryptography) are not yet available. As a consequence of this, the works that we cite will be collected from the sources with the highest number of reviews, but they will also originate from blogs and company postings written by respectable authors.

In a study on NFTs, A. Baronchelli, M. Nadini, F. Di Giacinto,L. Alessandretti, M. Martino, and L. M. Aiiello (2021) [21] collected data on token attributes and event records, including trading, minting, and wallet-owner information.. They looked at information mostly from the Ethereum and blockchains, which covered around $6.2 million in trading and $4.9 million in NFTs from 2017 to 2021. Since still photos, movies, and gifs make up the majority of NFTs, they also employed CNN to digitize images, which is a vital component of NFT in the modern era. Their study established a methodology for calculating the market's evolution while also aiming to deflate the Non-Fungible Token market's overall structure. They made use of visual features and collections, and the main focus of their study was on the market's statistics and characteristics, a web of interactions between traders and products. In their article, they also put out a feature-based linear regression model for forecasting NFT pricing.

M. Nadini, L. Alessandretti, F. Di Giacinto, M. Martino, L. M. Aiello, and A. Baronchelli (2021) [8] sought to explain the general structure of the Non-Fungible Token (NFT) marketplace and provide an outline for evaluating this market's growth. Between 2017 and 2021, the authors analyzed information gathered primarily from Ethereum concerning 6.1 million transactions and 4.7 million NFTs. and blockchains based on WAX. The results of this study contain a statistical analysis of the market, the formation of a network of relationships among traders (with links between buyers and sellers), and the grouping of things based on visual characteristics and collections. Additionally, the paper presents a clustering of objects based on collections. In addition to that, the research puts forward a linear regression model with features that are derived from these findings in order to forecast NFT pricing. Among the top five offerings on the NFT market, the NFTs that are traded the most frequently are those that belong to the categories Games, Collectibles, and Art, which account for 44%, 38%, and 10% of transactions respectively. The Art sector has dominated the market in terms of volume, providing over 71% of the entire transaction volume. When looking at the relationship between traders, it can be seen that the highest 10% of traders, as determined by the number of purchases and sales they make, account for 85% of all transactions. transactions. Moreover, traders who specialize in a collection tend to purchase and sell non-financial tokens (NFTs) among other traders who concentrate on the same collection. AlexNet, a kind of convolutional neural network, was used to generate dense vector illustrations of images, and principal component analysis (PCA) was applied to figure out the degree of similarity between 1.25 million unique NFTs. It was determined that the first five principal components contribute to 38.3% of the total variation and are used to evaluate the ability of visual characteristics to forecast NFT sales. A linear regression model with a least-squares loss was developed for calculating the prices of the primary and secondary sales. Using this method, the primary offering price of the NFT was estimated. The NFTs were divided into three groups based on eleven distinguishing features. Researchers discovered that the collection's median sale price can account for as much as 50% of the variation between primary and secondary market prices for NFTs. Furthermore, predictions were found to be more accurate if the median sale price was calculated for a more recent time period prior to the primary transaction. This is due to the fact that the middle price of the NFT's recent sales is more indicative of its true market value right now. Together, the buyer and seller centrality measures in the trader network and the visual qualities of the object associated with the NFT account for about 20% of the variance. The data for this analysis originated from NFT marketplaces as opposed to the WAX or Ethereum blockchains directly, which means that independent NFT producers were not taken into consideration. This is one of the study's several drawbacks. The investigation did not take into account other factors that can have an effect on market behaviour, such as social media.

C. Pinto-Guti errez, S. Gait an, D. Jaramillo, and S. Velasquez, (2022) [18] emphasized the use of vector autoregressive models (VARs), in order to show that Ether (ETH) and Bitcoin (BTC) are the most prevalent cryptocurrencies to forecast future NFT prices. To predict what an NFT will be worth in the future, this group looks at things like the Google search volume, SP 500, and the value of cryptocurrencies. This team shows that data on Google search trends is linked to large gains in cryptocurrency prices and NFT accumulations. In order to study the connection between cryptocurrency returns and NFT focus, this team uses VARs and wavelet coherence techniques. This study found no association between Ether returns and NFT curiosity, but did uncover a link between Bitcoin and NFT forecasting.

A. Kapoor, D. Guhathakurta, M. Mathur, R. Yadav, M. Gupta, and P. Ku- maraguru, (2022) [17] investigated two primary issues, the first of which is: a) What relationship exists between Twitter activity and OpenSea's price? b) Is it possible to estimate the value of NFT by using signals acquired from Twitter and OpenSea, and if so, can the features that most strongly influence the forecast be identified? This work aims to do two things: address this question and create one of the first NFT datasets to incorporate information from both OpenSea and Twitter. First determining if the NFT will be successful using a binary classification model, and then categorising profitable NFTs into different price bands using a multi-classification model.

Kapoor et al. concluded that including Twitter data in their feature set improved their model's accuracy by 6% when compared to a model that solely used data from NFT platforms (like OpenSea). This was the case when contrasting their model to one that relied solely on information gathered from NFT systems. This research offers new perspectives on additional training procedures and attributes that might be incorporated into our predictive model.

Chapter 3

Work Plan

3.1 Overview

In our model, first, we have performed RNN. Several RNNs, including vanilla GRUs, RNNs, and LSTMs, were developed to estimate the typical price of a BAYC NFT before the final model was constructed. An NFT's primary and secondary sales prices were calculated using a linear regression model, which approximates a linear connection between the dependent variable (secondary sales price) and a collection of independent variables.

3.2 Workflow Chart

A workflow chart for NFT (non-fungible token) price prediction would typically include the following steps:
1.Data collection: Gather historical NFT sales data from various marketplaces and platforms.
2.Data cleaning and preprocessing: Handling missing values, outliers, and inconsistencies, we cleanse and prepare the data for analysis.
3.Feature engineering: Extract relevant features from the data that can be used to predict NFT prices, such as the artist's name, the NFT's rarity, and the number of bids on the NFT.
4.Model selection and training: Select a suitable machine learning model, such as a neural network, and train it on the preprocessed data.
5.Model evaluation: Utilize measures like root mean squared error to assess the trained model's performance.
6.Model deployment: Deploy the trained model in a web or mobile application for users to access and make predictions on NFT prices.
7.Continuously monitor the model performance and update it when necessary.

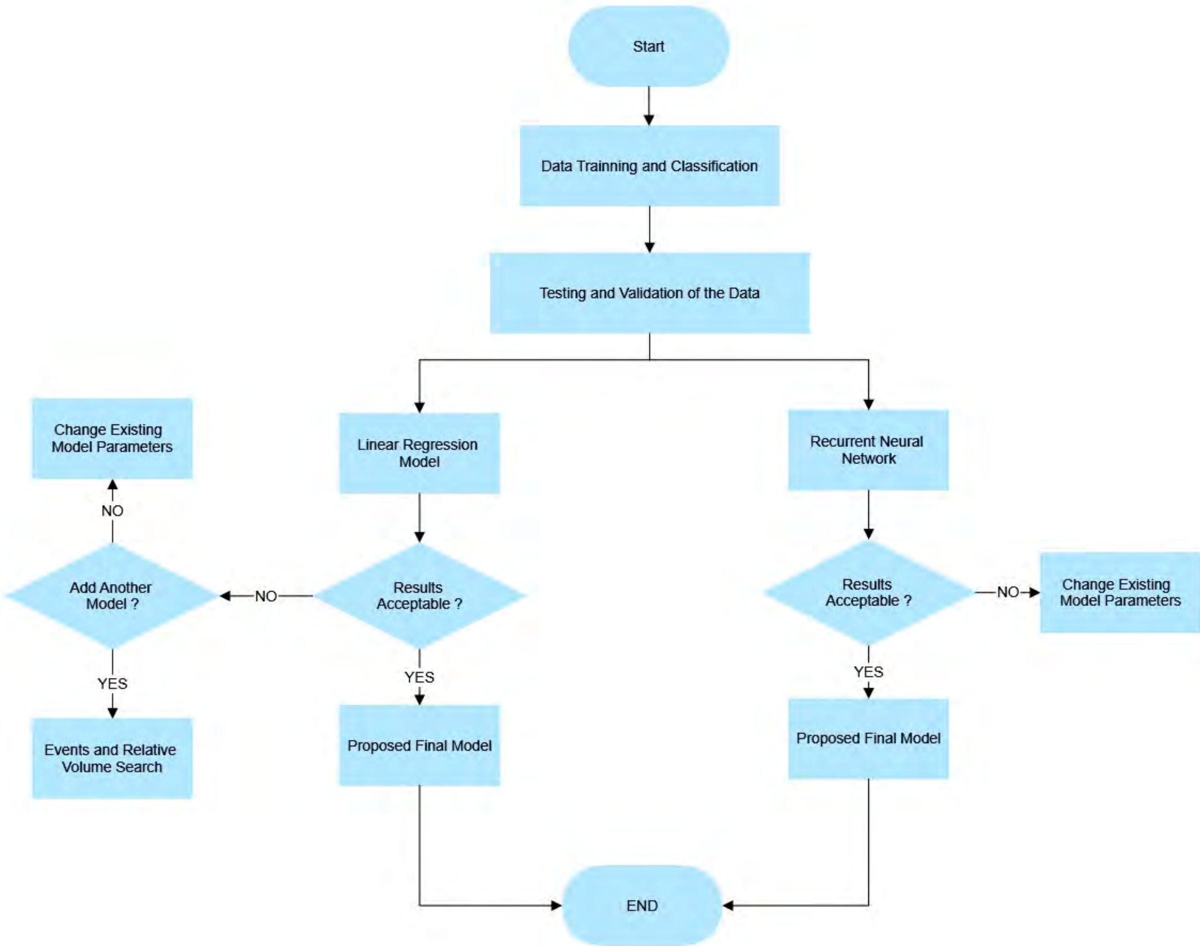

Figure 3.1: Workflow Chart

Chapter 4

Methodology

4.1 Data Collection and Processing

NFT Data:
In contrast to the websites of 2010, contemporary platforms such as OpenSea and other NFT exchanges exhibit a higher prevalence of non-fungible tokens (NFTs). Because of the volume and variety of data types, querying and searching for NFT-related data has become a difficult task.
Although indexing historical data connected to digital assets on the Ethereum blockchain might appear simple, we discovered that using current APIs to do so was more difficult than using standard APIs for the finance market (Yahoo Finance, for example). Currently, Ethereum nodes require the use of an indexer to store transactions by wallet address because they do not do so natively. The indexers that are currently available for querying the Ethereum blockchain are either too costly for our project, too fragmented, or have a high developer friction level.
In our project proposal, we were required to do a more extensive investigation of application programming interfaces (APIs) than initially anticipated in order to obtain the NFT-related data that we were eager to gather.We looked into the following APIs:

- NFT API for OpenSea
- API for historical data on Covalent NFT
- API for Etherscan
- API for Coingecko
- Moralis API

The Covalent API has emerged as a prominent candidate for indexing blockchain data across multiple chains, including Ethereum, Binance, Polygon, Solana, Ronin, and others. The team has a higher level of confidence in the data employed for this project due to Covalent's affiliation with prominent crypto venture capital firms and its well-regarded standing within the crypto development community. We obtained following data using the Covalent API:

- date

- The mean value of the NFT for a specific day.
- The quantity of NFTs that were purchased on a specific day.

The NFT collections selected for inclusion in our project proposal include of the following: Bored Ape Yacht Club (BAYC), Cryptopunks, Doodles, Deadfellaz, Gutter Cat Gang, Sup Ducks, Cyber Kongs, Creature World, Cool Cats, Lazy Lions, Azuki.

Public Market Data:
We used the Yahoo Finance API to get information on the stock market, which is widely recognised and referenced in academic literature. The Yahoo Finance API is a frequently employed resource for obtaining financial data, leading to its extensive acceptance by developers owing to its user-friendly nature. As articulated in our project proposal, our primary aim is to integrate NFT-specific data with publicly accessible market data to prognosticate the prospective worth of an NFT. To compile all market information into a single data set, we must first incorporate all tickers, which are the abbreviated symbols used to identify publicly traded shares of a certain firm on a given stock exchange. We can successfully combine the aforementioned dataframe with our NFT information.

Google Trends Data:
The primary determinant of value for an NFT is often the level of hype generated around a specific project.It is imperative for us to have a process for quantifying this statistic.One way that can be employed is utilizing Google Trends data to monitor the temporal changes in the search volume of a given collection's name. The PyTrends API enables the extraction of this information to be combined using our data collection.

Significant Event Data:
Extra variables, such recent press on the collections in question, can be difficult to assess when making predictions about them.For instance, when the Bored Ape Yacht Club announces the public availability of a new derivative, or when a fraudster sets his sights on the collection, we witness comparable impacts on price, which can be either positive or negative in nature. The aim is to quantify this phenomena by incorporating an Event characteristic whose values (if positive) indicate a favorable correlation with prices and (if negative) a negative correlation with those same values.

Aggregated Data Dictionary:
By utilizing the non-fungible token (NFT), publicly available market data, and data from Google Trends, we have constructed a comprehensive data dictionary. This data dictionary is designed to appropriately format the data for input into a neural network.

Variable	Description	
opening_date	Date of which information is being pulled.	X (input data)
average_volume_quote_day	Average price of the NFT as of opening_date.	
unique_token_ids_sold_count	The number of NFTs from a given collection sold in one day.	
ETH_USD	ETH token value at closing.	
BTC_USD	Bitcoin value at closing.	
GC=F	gold value at closing.	
^GSPC	S&P value value at closing.	
^DJI	Dow Jones value at closing.	
^NDX	Nasdaq 100 value at closing.	
MSFT	Microsoft stock price value at closing.	
AAPL	Apple stock price value at closing.	
NFLX	Netflix stock price value at closing.	
TSLA	Tesla stock price value at closing.	
AMZN	Amazon stock price value at closing.	
FB	Meta stock price value at closing.	
Relative Search Volume	Relative google search volume for collection name on a scale from 0-100	
Events	-1,0,1 indicating bad news, no news, and good news respectively	
Gas	A measure of network traffic, which indicates the transaction fee of purchase	
average_volume_quote_day[opening_date+1]	Average price of the NFT as of opening_date + 1 (the next day's price).	Y (labels)

Table 4.1: The provided table serves as a comprehensive data dictionary, encompassing all pertinent information required for input into both linear and recurrent models.

Data Adaptations:
Additional changes were required to assure accurate alignment of data, owing to the inherent discrepancies between the cryptocurrency market and the traditional stock market. The NFT and cryptocurrency markets exhibit continuous operation, functioning round the clock, seven days a week, in contrast to conventional markets that adhere to a restricted schedule from Monday through Friday during normal business hours.To better incorporate our NFT data with stock information, we've decided to only use the NFT data related to days that also have corresponding stock data. For datasets that include daily information, such as those obtained from Google Trends, the aforementioned modification was judged necessary.

Limitations:

- `Twitter API`: The purpose of our research was to integrate Twitter data about a certain collection so that the level of interest in that collection could be measured. Although we were able to successfully obtain access to the Twitter API, we encountered difficulties in obtaining authorization for the specific endpoint(s) necessary to obtain the desired volume statistics.To retrieve specific details of individual tweets, it is necessary to have privileged access to the Twitter /search/all endpoint.The objective of our study was to gain further insights into the frequency of tweets related to collections. To achieve this, we utilised the endpoint and substituted Google Trends data as a proxy. One potential alternative application for this access is to perform sentiment analysis on tweets pertaining to the collection; however, we encountered difficulties in proxying this task.
- `OpenSea Events API`: One statistic we wanted to look at more thoroughly using the OpenSea Events API was "owner similarity," which describes how similar owners are to one another across various NFT collections. A small group of skilled NFT traders exhibits a regular bimodal pattern, as noted by Nadini et al. (year). This pattern is defined by the traders' engagement in either high-volume transactions involving costly assets or comparatively few deals involving inexpensive ones.

4.2 Dataset Management

The following was used to collect our selected NFTs,
Sales Data:
Some of the Ethereum Sales data from more than 230000 are given in table 4.2

Date	Price Floor
2021-04-30 00:00	0.16
2021-05-30 00:00	0.5
2021-06-27 00:00	2.95
2021-07-22 00:00	6.1
2021-08-16 00:00	12.2425
2021-09-29 00:00	37.12
2021-10-30 00:00	37.729
2021-11-30 00:00	43.3785999999997
2021-12-29 00:00	20.384000000000007
2022-01-25 00:00	93.96000000000001
2022-02-25 00:00	0.10161
2022-03-20 00:00	117.57867225518
2022-04-23 00:00	150
2022-06-10 00:00	87.56

Table 4.2: Ethereum Sales data

4.3 Dataset Information

Dataset Includes:
-collection_twitter_statistics.csv
-collection.csv
-nfts_prediction.csv
-nfts_train.csv
-submission_form.csv
Data Information:

- `collection_twitter_statistics.csv`: Statistics Collected from various creators from NFT industry from their account.
- `collection.csv`: Dataset which are available in the training and prediction dataset.
- `nfts_prediction.csv`: Data and statistics about the NFTs from all the selected sources and used in the trained model.
- `nfts_train.csv`: Data and statistics about the NFTs from all the selected sources alongside the information to train the price prediction model.
- `submission_form.csv`: A test file to score the submitted information to go through scoring system.

collection _id	n_tweets _in _range	avg _replies	min _retweets	min _likes	min _replies	max _likes	max _replies	max _retweets
0	3	183.000000	6	183.0000	8.500000	3	14	76
1	15	3.562500	0	36.37500	3.562500	0	21	34
2	1	9.000000	1	15.00000	9.000000	9	9	1
3	525	0.306084	0	1.178707	0.306084	0	8	13
4	25	0.884615	0	8.153846	0.884615	0	6	12

Table 4.3: Social Media Data Frame (Twitter)

collection _id	Global index	NFT ID	Collection ID	Rarity Score	Last Sale Date	max _likes
0	21934	0	45	2.0000	2022-11	1.266732
1	34621	0	49	98.0772	2021-08	1.200357
2	33622	1	45	138.3455	2022-08	3.180572
3	33654	2	45	127.7546	2021-08	3.502910
4	32624	3	45	8.153846	2022-09	4.009567

Table 4.4: Social Media Data Frame 2

Sales Data (Yahoo Finance): Majority of the financial dataset we fetched from Yahoo! Finance which has all the information required and publicly available and accessible. We fetched a total of 2801 data providing NFTs price for 2801 days which is around 9 years. We were mainly focusing on the closing price for the future dates.

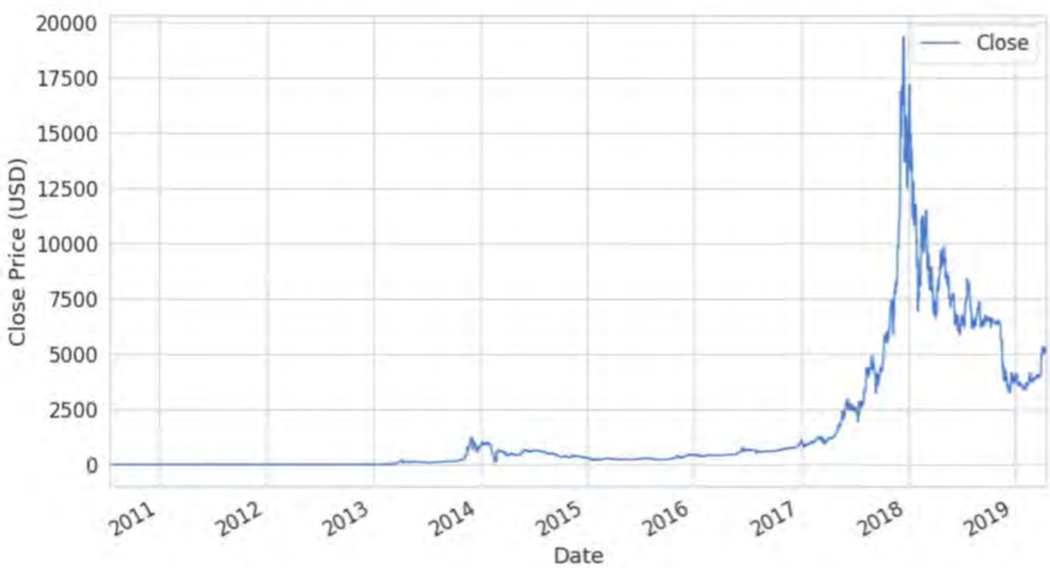

Figure 4.1: NFTs price for 2801 days

Chapter 5

Baseline Model Implementations and Result

5.1 Recurrent Neural Network

In the beginning, a recurrent neural network (RNN) was built to make predictions about the average value of a Bored Ape Yacht Club (BAYC) NFT. Recurrent Neural Networks (RNNs) are a type of deep learning models optimized for handling sequential information. This characteristic renders them highly suitable for performing time series predictions.Prior to arriving at the ultimate model, a variety of recurrent neural networks (RNNs) were constructed in order to forecast the mean price of a BAYC NFT. Among these are LSTMs and GRUs.

We fed the market data and timestamp to find out the price floor of BAYC NFTs. In the table below the increase of price floor is observed throughout the selected period given in Table 5.1

Date	Price Floor
2021-04-30 00:00	0.16
2021-05-30 00:00	0.5
2021-06-27 00:00	2.95
2021-07-22 00:00	6.1
2021-08-16 00:00	12.2425
2021-09-29 00:00	37.12
2021-10-30 00:00	37.729
2021-11-30 00:00	43.3785999999997
2021-12-29 00:00	20.384000000000007
2022-01-25 00:00	93.96000000000001
2022-02-25 00:00	0.10161
2022-03-20 00:00	117.57867225518
2022-04-23 00:00	150
2022-06-10 00:00	87.56

Table 5.1: Time Stamp and Price Floor of BAYC NFTs

The top sales within the chosen period were one of the prerequisite statistics that we needed. Our data is displayed in Table 5.2 below:

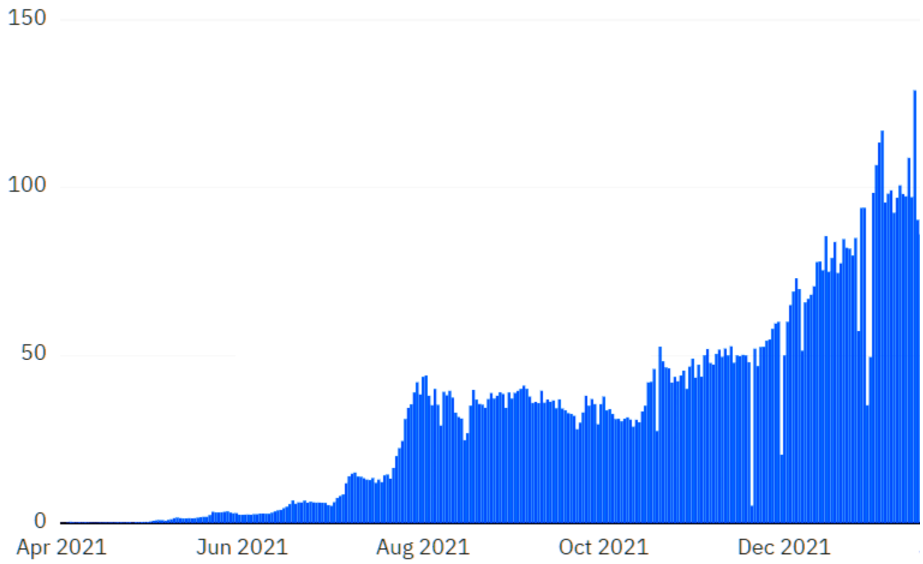

Figure 5.1: BAYC Price Floor

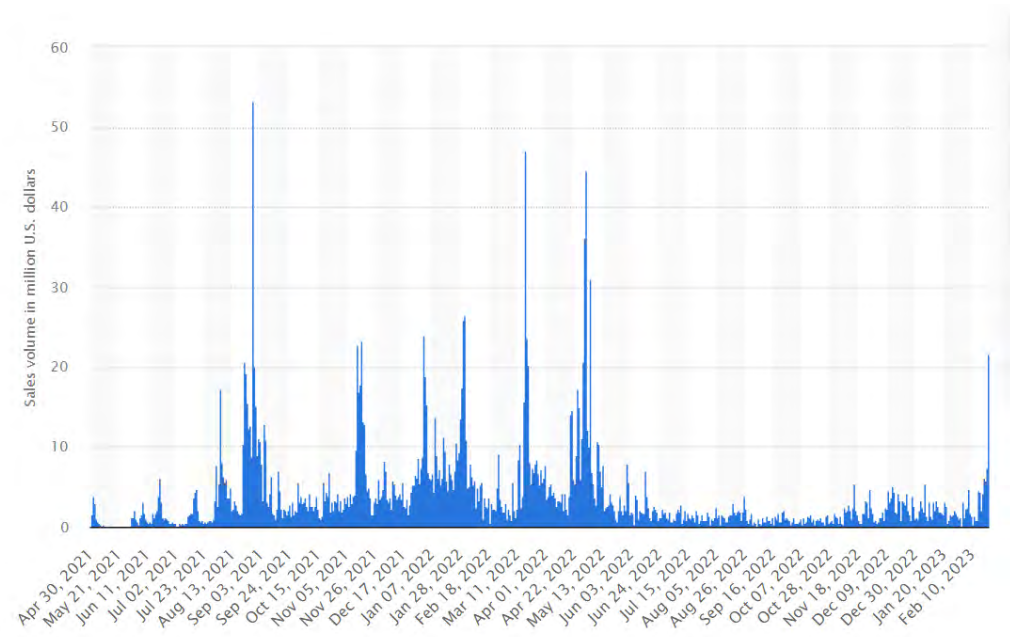

Figure 5.2: BAYC VOLUME GRAPH

Rank	ID	Ethereum Price (Eth)	US Price $	Time
1	4218	39000.00	55,213,470	2022-05-11
2	4256	38000.00	53,797,740	2022-05-11
3	3217	36969.00	52,338,122	2022-05-11
4	7537	1024.00	1,449,708	2022-06-14
5	4400	800.00	1,132,584	2022-01-31
6	1734	800.00	1,132,584	2022-01-20
7	5233	800.00	1,132,584	2022-01-30
8	2087	769.00	1,088,696	2021-09-30
9	3749	740.00	1,047,640	2021-09-06
10	8585	696.97	986,720	2021-10-19
11	4580	666.00	942,876	2022-02-25
12	3928	626.00	886,247	2021-09-13
13	7090	600.00	849,438	2021-09-02
14	1837	569.00	805,550	2022-02-25
15	8135	550.00	778,652	2021-09-17

Table 5.2: Top 15 NFTs Sale and Costs

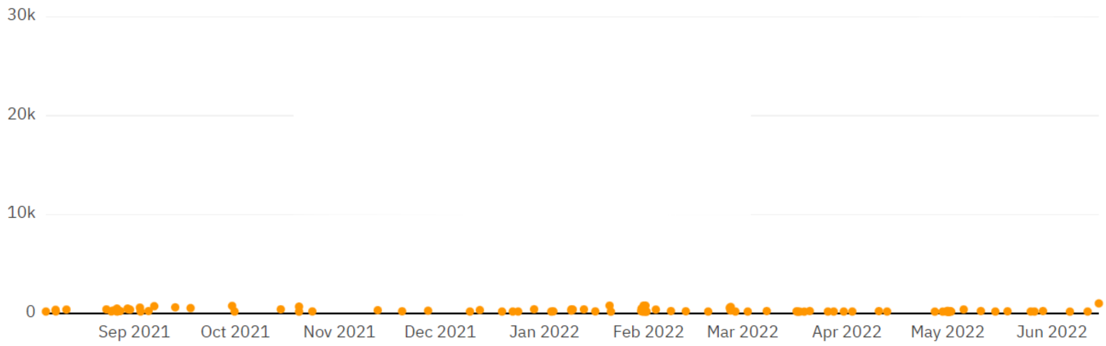

Figure 5.3: Bored Ape Yacht Club Top Sales

We plot the data into our Recurrent Neural Network system in time series.

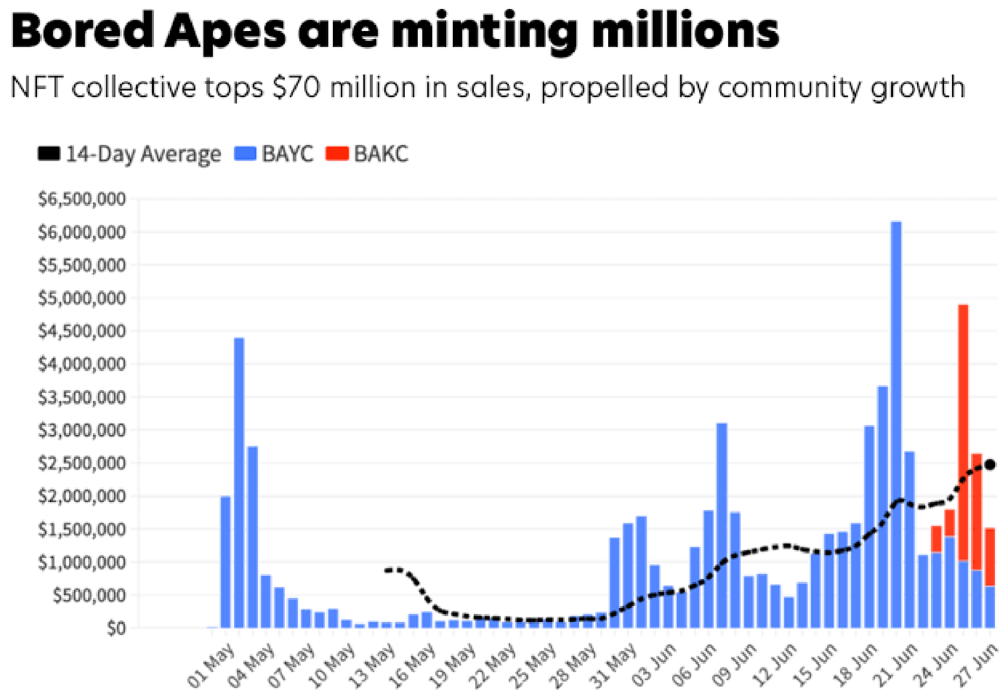

Figure 5.4: A model of the time series information that was used to train the RNN.

Among several types of recurrent neural networks (RNNs) that were evaluated, it was observed that the RNN employing Gated Recurrent Unit (GRU) cells had the highest validation accuracy. The ideal hyperparameters were determined by a grid search and are shown in the table below:

GRU Blocks Number	Size of Hidden Dimension	Rate of Learning	Length of Sequence
3	16	0.01	5

Table 5.3: RNN hyperparameters

An RNN and were used to make long-term price predictions for the BAYC NFT. The model exhibited a typical error rate of 23% for the training dataset, 21% for the validation dataset, and 20% for the test dataset after training and validation. Figure 5.4 depicts the RNN's projected training pricing and the actual NFT pricing over time.

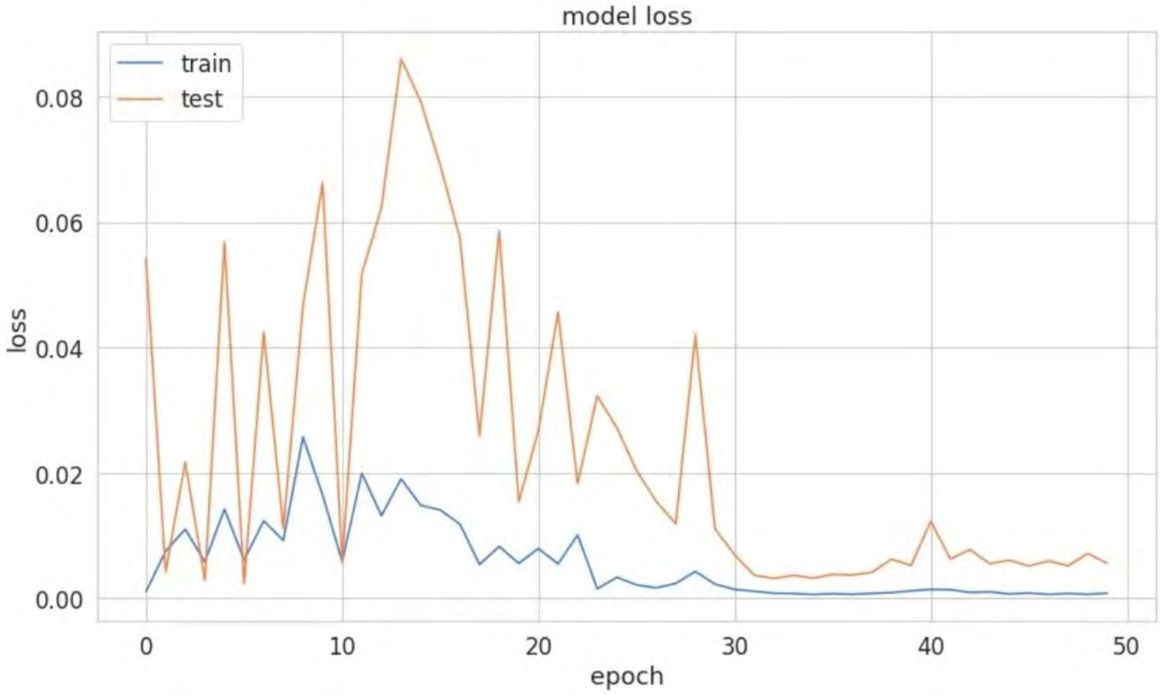

Figure 5.5: Loss in Data training

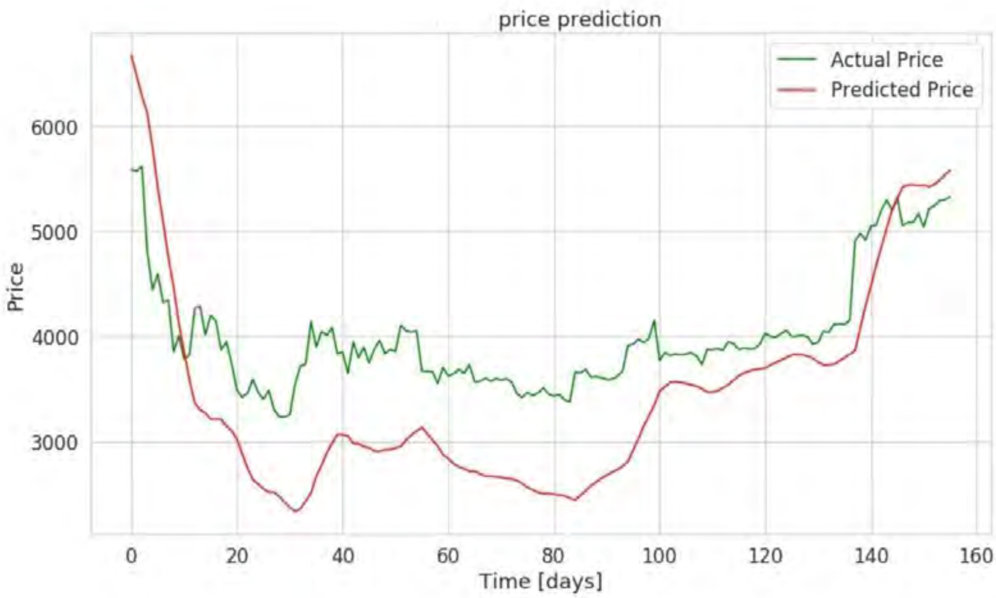

Figure 5.6: Accuracy of the tested model

Although NFT prices are highly unstable in nature,[16] we are not satisfied with the result. Consequently, following a thorough examination of RNN for price prediction, we decided to find out the key factors that influence NFT prices rather than trying to figure out the exact NFT prices. And to that we use Linear Regression Model.

5.2 Linear Regression

We utilized a linear regression model to identify the factors that have an impact on an NFT's price. By analyzing several model diagnostics, we assessed the predictor's suitability. These are the corresponding t statistics that examine the variance inflation factors (VIF) that follow predictor multicollinearity as well as the standard deviation of the regression coefficients. Table 5.4 following shows these outcomes for the basic model. It is shown in following Table 5.4 is high multicollinearity among the predictors indicated considerable relationships between them. This will inflate the variances of the predictors artificially and make the regression surface volatile which results making intervals and diagnostic tests unstable. This outcome was expected as many predictors, including the NASDAQ, have strong correlation. The table also shows that some predictors are not relevant at the 0.05 level, possibly due to multicollinearity. We can therefore conclude that some predictors account for the large majority of the model accuracy, which implies that a few of them can be eliminated.

variable used to anticipate result	Coeff	VIF	t	P>$\|t\|$	Confidence Interval
Time elapsed since release	959.7	548.9	1.166	0.246	-670.4,2587.6
Average price of NFTs	.09	41.8	1.053	0.294	-0.082, 0.267
# of NFTs sold	987.6	4.9	4.163	0.000	517.6, 1457.5
Gas	11.5	286.0	0.732	0.466	-19.7, 42.8
Ethereum to US dollar	-14.2	749.6	-0.541	0.590	-66.0, 37.7
Bitcoin to US Dollar	-0.6	705.0	-0.352	0.726	-4.2, 3.0
Gold	165.8	3309.4	1.335	0.185	-80.3, 411.9
S&P	72.6	531499.6	0.136	0.892	-987.9, 1133.2
Dow Jones	-27.7	140946.3	-0.743	0.459	-101.7, 46.3
NASDAQ	-8.0	142726.0	-0.097	0.923	-170.9, 154.9
MSFT	811.9	8842.1	0.792	0.430	-1218.0, 2841.9
AAPL	-553.0	2677.4	-0.499	0.618	-2746.4, 1640.4
NFLX	-57.5	797.6	-0.336	0.738	-396.4, 281.5
TSLA	30.0	456.9	0.368	0.713	-131.1, 191.2
AMZN	28.9	2311.5	0.598	0.551	-66.8, 124.6
FB	238.9	838.3	0.803	0.424	-350.4, 828.3

Table 5.4: For each predictor, the following were calculated: coefficient, VIF, t-statistic, p-value, and confidence interval.

After removing some variables, the accuracy have become far greater. An adjusted R^2 result of 0.7 for the first model indicates high model quality.The addition of two more predictors, Events and Relative Volume Search, increased the model's quality, reliability, and capacity to capture the impact of hype on the price of the NFT. We began to exclude predictors with high multicollinearity after including these predictors and recalibrating the VIF factors. Reassessment of the VIF factors led to the retention of predictors with high p values. The four predictors in the final model all have relatively low VIFs, which are suitable for data analysis. Both of these new predictors had statistically significant results at the 0.05 level, while the other two were significant at the 0.1 level. Out of all the predictors, the daily volume and the previous day's price had the strongest correlation with the NFT's pricing for the next day. The final model's modified R^2 score of 0.69 indicated that almost The final model retained all the features of the initial model, but eliminated the redundant predictors. The characteristics of the predictors and the final model are summarized in Table 5.5

variable used to anticipate result	Coefficient	VIF	t-statistic	P>$\|t\|$	Confidence Interval
Mean NFT Value	1168.0	6.6	11.8	0.000	972.5,1363.5
Quantity of NFTs sold	903.3	2.7	4.7	0.000	525.2,1281.4
Gas	11.0	5.3	1.9	0.056	-0.3,22.2
Search Volume in proportion	407.3	4.3	1.8	0.072	-37.5,853.4

Table 5.5: The final result prediction includes the VIF coefficient, t-statistic, p-value, and confidence interval.

Before removing some variables for increasing the accuracy, we get the following result,

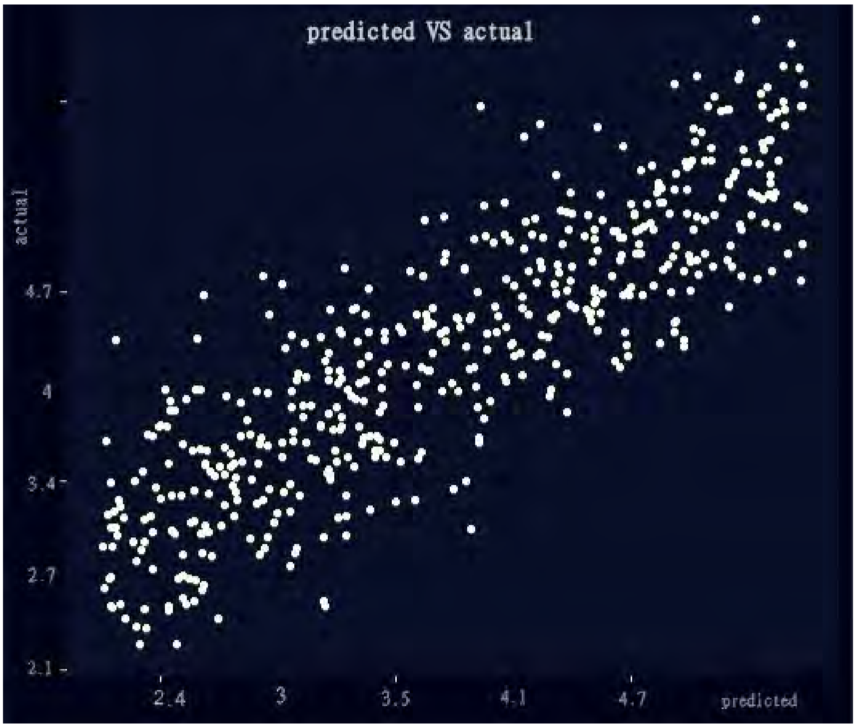

Figure 5.7: Multicollinearity and Variance inflation factor before removing variables

After removing some variables for increasing the accuracy, we get the following result,

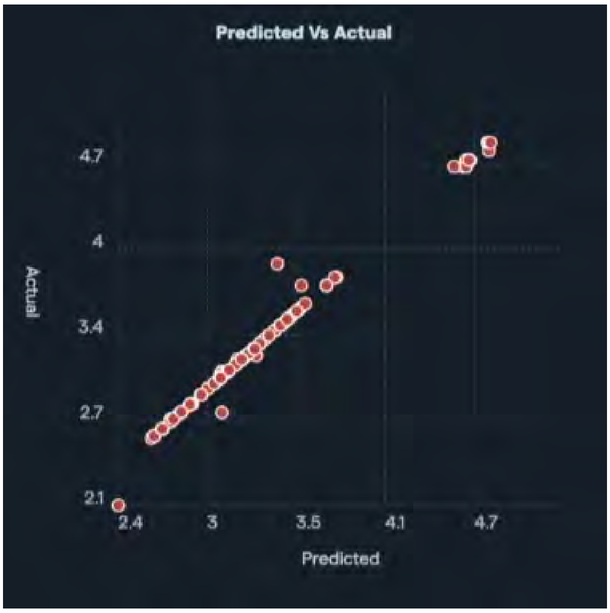

Figure 5.8: Multicollinearity and Variance inflation factor After Removing Some Variables

The model also assessed a test set. This evaluation was required to make sure that the model wasn't overfit and that the findings could be applied to another data set. 20% of the entire dataset, or 265 time series occurrences during January 2022 and January 2023, comprised up the test set. The model's modified R^2 score on the test set was 0.65, Which demonstrates that the final result has a higher degree of applicability and that only a very tiny percentage of the quality of the model was compromised.

5.3 Result

In order to assess the precision of the two models (RNN and Linear Regression), we conducted an analysis of metrics and these are RMSE (Root Mean Squared Error), R2 (R-squared), and RMSLE (Root Mean Squared Logarithmic Error). These measures are utilized to evaluate the efficacy of regression models. Smaller values of Root Mean Squared Error (RMSE) and Root Mean Squared Logarithmic Error (RMSLE) are indicative of higher accuracy, whereas a larger value of R-squared (R2) suggests greater precision between the model and the observed data.
For RNN we get the following results:
RMSE: 9.37,
R^2: 0.46,
RMSLE: 0.70.
For Linear Regression Model we get the following results:
RMSE: 7.49,
R^2: 0.65.
RMSLE: 0.48

5.4 Result Analysis

So the results we have found suggest the following:

RMSE (Root Mean Square Error):

- `RNN RMSE: 9.37`
- `Linear Regression RMSE: 7.49`

The Root Mean Square Error (RMSE) is a metric used to measure the average magnitude of discrepancies between anticipated and observed data. Lower root mean square error (RMSE) ratings are indicative of greater levels of accuracy. In this specific case, the Linear Regression model demonstrates a reduced Root Mean Square Error (RMSE), suggesting its higher accuracy in minimising prediction errors.

R-squared (R^2) Value:

- RNN R^2: 0.46
- Linear Regression R^2: 0.65

The amount of variance in the dependent variable (the target) that can be predicted from the independent variables (the features) is expressed as R^2. An increased R^2 value suggests that the model fits the data more accurately. In this case, the Linear Regression model is probably a better match because it explains more variance in the data, as indicated by its higher R^2 score.

RMSLE (Root Mean Squared Logarithmic Error):

- `RNN RMSLE: 0.70`
- `Linear Regression RMSLE: 0.48`

RMSLE is similar to RMSE but applies a logarithmic transformation to both the predicted and actual values before calculating the error. A smaller RMSLE indicates better accuracy. Here we also see that the Linear Regression model has a lower RMSLE value, which means it is more accurate in predicting the logarithmic errors of the data.

Based on the observed metrics, it can be concluded that the Linear Regression model demonstrates more accuracy in comparison to the RNN model when applied to the provided dataset. The model exhibits a reduced RMSE and an increased R^2, both of which suggest improved prediction capabilities.

Chapter 6

Conclusion

In summary, our thesis signifies a notable progression in the domain of price prediction for Non-Fungible Tokens (NFTs) through the utilization of deep learning methodologies, notably Recurrent Neural Networks (RNNs) and Linear Regression models. The key distinguishing factor of our research, in contrast to previous models, resides in the significant enhancement in accuracy that we have attained. By leveraging RNNs and Linear Regression, we have refined the predictive capabilities of NFT price movements, offering a more precise tool for both investors and enthusiasts in the NFT market. This heightened accuracy is paramount in an increasingly dynamic and speculative market where precise pricing information is crucial for decision-making. The results of our study not only enhance the scholarly comprehension of NFT price forecasting but also possess practical implications for stakeholders within the NFT ecosystem. Investors, collectors, and traders stand to gain advantages from our enhanced model, which offers increased precision in decision-making about the acquisition, divestment, or retention of NFTs. This has the ability to mitigate risks and optimize returns. Additionally, our study underscores the possibility of integrating diverse machine learning methodologies, such as Recurrent Neural Networks (RNNs) and Linear Regression, to augment prediction models within the domain of cryptocurrency and digital asset markets. This methodology has the potential to facilitate additional breakthroughs and progress in the respective domain. In essence, our thesis signifies a significant advancement in the field of NFT price prediction, characterized by its heightened precision in relation to pre-existing models. This accomplishment exhibits possibilities for individuals navigating the non-fungible token (NFT) market and emphasizes the capacity of machine learning in comprehending the intricacies of digital asset valuation.

Bibliography

[1] A. Barrat and M. Weigt, “On the properties of small-world network models,” *The European Physical Journal B-Condensed Matter and Complex Systems*, vol. 13, pp. 547–560, 2000.

[2] J. Brownlee, “Linear regression for machine learning,” *Machine Learning Algorithms*, 2016.

[3] J. I. Wong, “The ethereum network is getting jammed up because people are rushing to buy cartoon cats on its blockchain,” *Quartz Media, December*, 2017.

[4] T. M. Evans, “Cryptokitties, cryptography, and copyright,” *AIPLA QJ*, vol. 47, p. 219, 2019.

[5] M. Franceschet, “Art for space,” *Journal on Computing and Cultural Heritage (JOCCH)*, vol. 13, no. 3, pp. 1–9, 2020.

[6] J. Devlin, *The “insane” money in trading collectible cards*, 2021.

[7] A. S. B. Luke, A. Shaw, and H. Stoilas, “Wtf are nfts? why crypto is dominating the art market,” *The Arts Newspaper). Available at: https://www. theartnewspaper. com/podcast/wtf-nfts (Accessed: 3 April 2021)*, 2021.

[8] M. Nadini, L. Alessandretti, F. Di Giacinto, M. Martino, L. M. Aiello, and A. Baronchelli, “Mapping the nft revolution: Market trends, trade networks, and visual features,” *Scientific reports*, vol. 11, no. 1, pp. 1–11, 2021.

[9] D. Phillips and S. Graves, *The 10 most expensive nfts ever sold*, 2021.

[10] S. Reyburn, “Jpg file sells for $69 million, as `NFT mania' gather space$,” *The New York Times*, vol. 11, 2021.

[11] R. Riegelhaupt, “Results: Beeple’s purely digital nft-based work of art achieves $69.3 million at Christie's$,” *Christie’s*, vol. 11, 2021.

[12] K. Sako, S. Matsuo, and S. Meier, “Fairness in erc token markets: A case study of cryptokitties,” in *Financial Cryptography and Data Security. FC 2021 International Workshops: CoDecFin, DeFi, VOTING, and WTSC, Virtual Event, March 5, 2021, Revised Selected Papers 25*, Springer, 2021, pp. 595–610.

[13] A. Serada, T. Sihvonen, and J. T. Harviainen, “Cryptokitties and the new ludic economy: How blockchain introduces value, ownership, and scarcity in digital gaming,” *Games and Culture*, vol. 16, no. 4, pp. 457–480, 2021.

[14] M. Dowling, “Fertile land: Pricing non-fungible tokens,” *Finance Research Letters*, vol. 44, p. 102 096, 2022.

[15] M. Dowling, “Is non-fungible token pricing driven by cryptocurrencies?” *Finance Research Letters*, vol. 44, p. 102 097, 2022.

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

[17] A. Kapoor, D. Guhathakurta, M. Mathur, R. Yadav, M. Gupta, and P. Kumaraguru, "Tweetboost: Influence of social media on nft valuation," in *Companion Proceedings of the Web Conference 2022*, 2022, pp. 621–629.

[18] C. Pinto-Gutiérrez, S. Gaitán, D. Jaramillo, and S. Velasquez, "The nft hype: What draws attention to non-fungible tokens?" *Mathematics*, vol. 10, no. 3, p. 335, 2022.

[19] J. van Slooten, "Predictive value of tweet sentiment on the bored ape yacht club's trading volume and floor price," Ph.D. dissertation, Master's thesis. Erasmus University Rotterdam, 2022.

[20] J. Rahman and Y. Jin, "The control of tax corruption: Evidence from non-fungible token market in china," *Journal of Money Laundering Control*, 2023.

[21] M. Nadini, L. Alessandretti, F. Di Giancinto, M. Martino, L. Aiello, and A. Baronchelli, "Non-fungible tokens: Can we predict the price they'll sell for," *Obtenido de The Alan Turing Institute: https://www. turing. ac. uk/blog/non-fungible-tokens-can-wepredict-price-theyll-sell*,