



Non-fungible token categorisation using machine learning classifiers

by: Manazir Najib
Student ID: B720039

Loughborough University

21COP328: Data Science Project

Supervisor: Dr Mohamad Saada

Submitted on 12th May 2022

COP328: MSc Data Science Project Research Proposal

Introduction

Background

Non-Fungible Tokens (NFTs) are trending globally, especially with the passing of COVID-19. NFTs are a part crypto advancement that are catching the attention of individuals who have become exhausted with basic cryptocurrency exchanges presented by Bitcoin, Ethereum and any other digital currencies.

The idea of digital ownership gives new excitement to the ever-growing digital population. Numerous compilations of digital assets and artwork have become vastly popular, an example is the CryptoPunk NFT collection (which is a collection of pixels containing digital avatars, which all have different assets and traits). This collection is situated on the Ethereum blockchain and is presently priced at an excess of \$10 billion. [1]

Due to its exponential increase popularity over the recent years, there are vast number of transactions, and each transaction is publicly displayed, so the data available is in surplus. Therefore, the data can be analysed very well, initially to develop machine learning algorithms that can be used to classify and find trends of the chosen NFT collection. Likewise, further analysis can be used to extract rare NFTs within a collection and what traits make them rare, this can be developed into a rarity tool.

Aims

The research aims to give overview of the history and knowledge on crypto domain. Highlight the basics of NFTs. Give an overview of problems facing with NFT transactions and highlight utilities of NFTs. The project will ideally collect NFT data of a certain collection (like CryptoPunk collection) and use them to analyse the behaviours and trends of the tokens. This can be analysed using machine learning algorithms for classifications and predictions. Finally create a rarity tool that can give filter out rare NFTs, which possess unique traits.

Objectives

The overall Part T project objectives are as follows:

- Give insight of NFTs and Crypto domain
- Collect dataset from CryptoPunk NFT collection using API
- Analysis of NFT data
- Develop machine learning based of collected data for classification
- Provide a rarity tool to highlight assets and traits of NFT collection

Preliminary literature review

From initial readings of multiple journals and articles, few fundamentals' topics were the highlights of the papers, therefore in-depth research is needed to gather a further concrete understanding of the functionality of NFTs and how the data can be processed.

Currently the following topics have been briefly researched history of centralised transactions, fundamentals of blockchains and how it uses a digital public ledger to create and execute smart contracts.

History -Centralised Exchange

The common method of everyday transactions initialised by a centralised exchange system. The fundamental concept behind a centralised system is that the transactions and exchange of money is observed by a 'third-party.' Third parties are like banks or online platforms, which have ledgers that store and monitor all purchases being made. [2]

Moreover, in a centralised exchange system where transactions and trades arise, the public authority has set a series of regulations and guidelines, this establishes the trust between the 'third-party' mediators and clients. An example of a law is anonymity, where clients contact data are stored on a ledger by the third-party (the ledger holds all exchanges, timestamped between parties and companies, in crypto exchange this is often a digital ledger).

Public Ledger

In the previous era public ledgers were utilised in towns and villages for tracking exchanges between all parties of the village. Parties include blacksmith, ranchers and daily merchants. This was a validation strategy which works by exploiting every trade, exchange and news to the public view. [2]

With the rise of trade with digital currency, a blockchain system was developed that introduces a likewise system for verification of each trade. The use of this new digital ledger system built into the blockchain, has been phenomenally successful therefore it has become the norm. This also ensures the identity, credit and records of all transactions from a client is always retained between the client and the platform where the purchase occurs. [2]

Blockchain

Blockchain was initially introduced by Bitcoin, this blockchain system utilises a computerised public record for crypto exchange. The blockchain is a copied data construct scattered and divided between the clients of an organisation.

Individual block can be distinguished by its 'cryptographic hash,' the past update of the block has a hash reference indented on it. The chain of block is subsequently made by the basis of connections between each block. Apart from the initial block, every block that follows it after trade and transaction, records its history of handlers. The blocks are immutable, this guarantees trust for the dispersion of NFTs between clients without a mediator. [3]

Smart Contracts

Smart contracts the foundation of NFT transactions and the leap forward in blockchain system. In 1990, when the first smart contract was issued, all agreements of the trade was executed by a computerised exchange procedure. The few lines of codes guarantee that after the correct circumstances have been met, the agreement is completely executed. These codes ensure that any breach of contract is archived immediately followed by extermination of the contract, consequently punishing the offender. Each settlement of the contract is documented as an unalterable exchange which is executed and retained in the blockchain. [4][5]

Gaps / Problems

The research on NFTs is in its early development so a lot of information is unknown and can be researched from first principles. Uncertainty of determining the market price of NFTs are a noticeably big challenge as it is very reliant on uniqueness and attributes. NFT values are very volatile because they are influenced by real world events. The value is almost impossible to determine, as it is also influenced by sudden exciting of general population which can derive from random trends. [6]

Methodology

Collecting Data

After research and gathering the fundamental understating of the topic of NFTs, a dataset is required for experimentations, obtaining results, and making predictions.

NFTs such as Crypto Punks is commonly traded on OpenSea this platform stores all transactions of NFTs, therefore all data is stored on this website. With the help of an API, it is possible to extract the data from the websites fairly easily.

Application Programming Interface (API) is a software mediator that permits two applications to engage in data exchange. API software can be used freely that can target any NFT on the OpenSea platform and extract the data with specified filters, such as time stamps. The NFT API enables the tracking, confirmed ownership, every transaction, value and more data, without sorting through and engaging in a smart contract. This data can be obtained into Python through simple importing codes, or it can be put into a database in SQL Server, which can be later assessed for data analysis. [7]

Pre-processing/ Cleaning Data

Like all with data science projects, the data obtained is raw, therefore needs cleaning and refining, to give best results and reliable predictions. Ensuring missing data is removed as an entirety, all dates and times are in the same format, elimination unwanted or unusable columns are essential for this research. Simplifying the dataset, allows for better understating for the machine learning algorithms and less computational effort.

Analysing the Data

After the data has been reconstructed and structured appropriately, this can now undertake data analysis. Various analysis can be achieved, the possibilities are almost endless. One aspect that can be researched is the sales and price statistics. This can be observed and compared between two time periods, during and after COVID-19, this can give insight of the growth of the crypto domain in isolation. Another analysis would be simple price evaluation, rarity and traits can be identified for each CryptoPunk pixel. Another idea would be to determine the volume of purchase dependent on the valuation of Ethereum coin through out the collected time period. [8]

Machine Learning

Several algorithms and pipelines will be created initially, involving supervised and unsupervised methods before approaching to create a full complete model for furthering the research.

The CryptoPunk NFT collection has 9999 individual NFTs, each stored in a pixel. The NFT have different classifications with different traits of appearances. Some of the classifications are base male, base female, zombies, aliens and apes. These can be put through a classification algorithm that can verify the CryptoPunk. This can be further developed using Convolutional-Neural network (CNN), that can process and identify based on the image analysis of each pixel.

Another model could be created that highlights the uniqueness of individual NFT, and the assets and traits held by the specific NFT. This can be analysed and later a prediction model can be generated, after fine tuning, of appraisal of the NFTs depending on its qualities and abilities. Thus, creating a rarity tool, which can offer insight of NFTs that have a potential of flourishing in value in the future for its unique traits. [1][8][9]

Software

OpenSea API – free online software used for the extraction of CryptoPunk data

SQL Server – storing the data/cleaning

Python – Data Analysis/ Development of machine learning models and CNN

Project plan

The following Figure 1 shows a Gantt chart, which gives a periodical breakdown of task that need to be completed to fulfil the deliverables and objectives of this project. The chart highlights key dates and deadlines. Some key dates are

- Project Proposal - 13th of May
- Final Dissertation - 26th August

The chart gives rough estimates with how long certain task should take; however, some have contingencies. These tasks may take longer to complete to its difficulty. They have been highlighted in grey. Some tasks may overrun the time estimated, or some may need to be approached earlier than initially thought as it is a more demanding task, the Gantt chart is a first update and will be changed regularly.

Progress of the project will be determined by the progress of the tasks, as most of the task cannot be achieved unless the prior task has been completed, this gives good indication if the project is feasible or not.

Further work is an optimistic goal, outside the scope of the deliverables required by this project. However, it can be beneficial for development of this project in the future and a starting point for the future. This will be only done if the progress of the project is flowing well, and enough time is available to produce good methods and results for the further work.

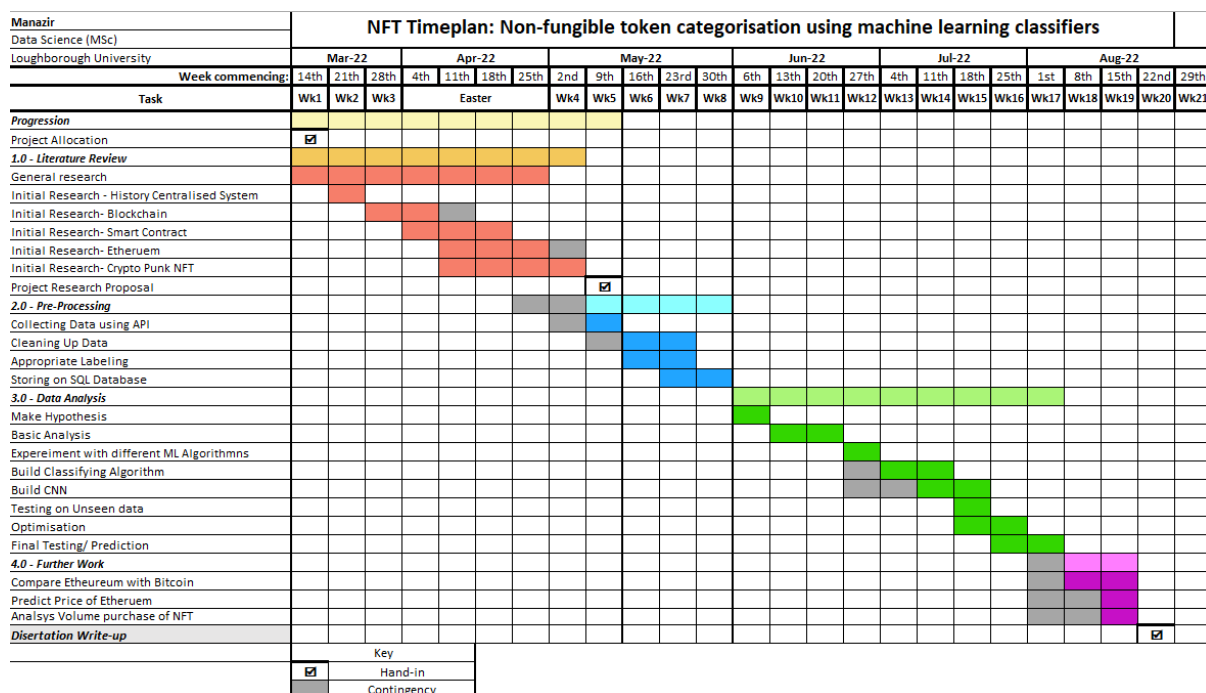


Figure 1- Gantt Chart

References

1. Medium. 2022. *Deep Learning w/ CryptoPunk*. [online] Available at: <<https://medium.com/analytics-vidhya/deep-learning-w-cryptopunks-6b27dc767fab>> [Accessed 12 May 2022].
2. Investopedia. 2022. What Are Centralized Cryptocurrency Exchanges? [online] Available at: <https://www.investopedia.com/tech/what-are-centralized-cryptocurrency-exchanges/> [Accessed 7 April 2022].
3. Investopedia. 2022. What is a Cryptocurrency Public Ledger? [online] Available at: <<https://www.investopedia.com/tech/what-cryptocurrency-public-ledger/>> [Accessed 7 April 2022].
4. Christidis, K. and Devetsikiotis, M., 2016. Blockchains and Smart Contracts for the Internet of Things. *IEEE Access*, 4, pp.2292-2303.
5. Zheng, Z., Xie, S., Dai, H., Chen, W., Chen, X., Weng, J. and Imran, M., 2020. An overview on smart contracts: Challenges, advances and platforms. *Future Generation Computer Systems*, 105, pp.475-491.
6. Mazur, M., 2021. Non-Fungible Tokens (NFT). The Analysis of Risk and Return. *SSRN Electronic Journal*,
7. MuleSoft Blog. 2022. *What are APIs and how do APIs work?* | *MuleSoft Blog*. [online] Available at: <<https://blogs.mulesoft.com/learn-apis/api-led-connectivity/what-are-apis-how-do-apis-work/#:~:text=API%20stands%20for%20%E2%80%9CAppl%20Programming,the%20response%20back%20to%20you.>> [Accessed 12 May 2022].
8. Zzun.app. 2022. *NFT-Price-Prediction-CNN - Using visual feature extraction, prices of NFTs are predicted via CNN (Alexnet and Resnet) architectures.* | *PythonRepo*. [online] Available at: <<https://zzun.app/repo/furkanoruc-NFT-Price-Prediction-CNN>> [Accessed 12 May 2022].
9. Shahriar, S. and Hayawi, K., 2022. *NFTGAN: Non-Fungible Token Art Generation Using Generative Adversarial Networks*. [online] arXiv.org. Available at: <<https://arxiv.org/abs/2112.10577>> [Accessed 12 May 2022].