

# Using Twitter for NFT asset valuation

CS 328 : Introduction to Data Science (2022)

Devanshu Thakar [18110174]  
Jayesh Salunkhe [18110072]  
Hardik Mahur [19110086]

Instructor : Prof. Anirban Dasgupta

# Introduction

---

- Upsurge of Web 3.0

Transition from Web 1.0 → Web 2.0 → Web 3.0

Web 3.0 - Decentralization and token-based economics

- Introduction to blockchain technology

A publicly distributed ledger

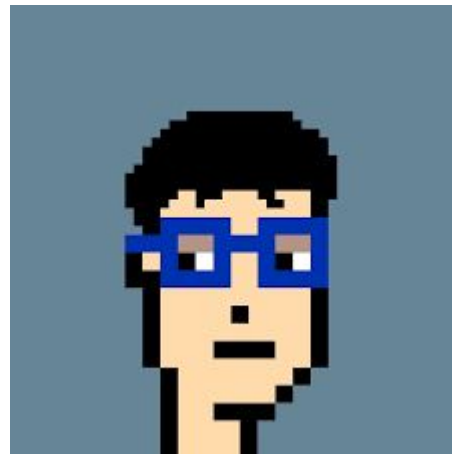
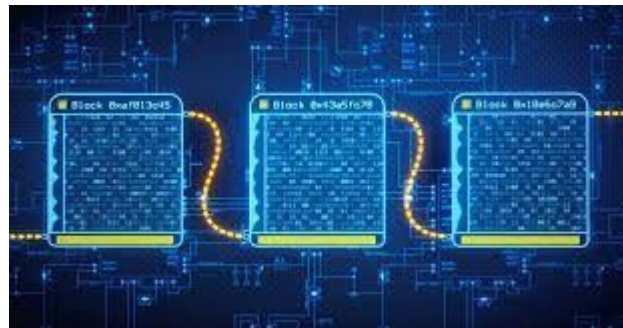
Crucial role in cryptocurrency systems

- What is an NFT?

a token that certifies a digital asset to be unique

Can store anything that can be converted to digital files

Eg. Music, images, tweets, videos etc.



# NFT : Facts

---

- Exponential growth of NFTs in 2021
- NFT transactions worldwide jumped 21,350% to **more than \$1.5 billion** in 2021, from \$82.5 million in 2020.
- Jack Dorsey sells his first tweet ever as an NFT for 2.9M \$.
- ‘Disaster Girl’ - a popular internet meme sold for 495,000 \$
- NFT asset valuation
  - Value of an NFT is based on the perception of buyers
  - Most NFTs sold online can be downloaded and shared publicly for free.
  - The Rarity, Utility, ownership history and associated social proof decide the value of NFT.



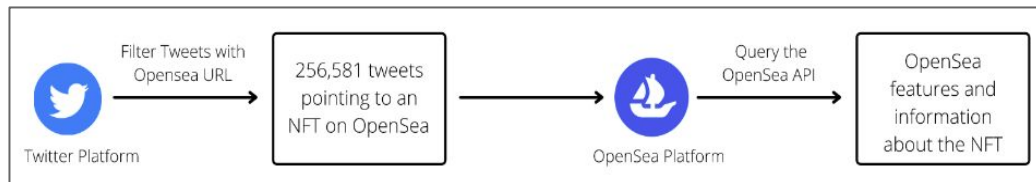
# Importance and Relevance

---

- The Non-Fungible Tokens NFT's made to headlines in 2021. NFT market made \$17.6 billion in sales in 2021, which was just \$82.5 million in 2020. A 200 fold increase displays peoples rising interest. [1]
- Based on blockchain technology, NFTs are digital assets with proof of ownership. They can be traded, collected and created by anyone. It derives its value not majorly from its content but major from its popularity.
- Various social media platforms are building NFT integrations to attract customers to their platform and facilitate interaction between traders.
- Thus NFT's asset value are influenced by various social media platforms. We plan to study Twitter interactions and their influence on asset price.

# Dataset Description

- Dataset originate from Twitter and OpenSea marketplace.
- Collected by scraping tweets which contains link to NFT asset in OpenSea marketplace. A total of 2,45,159 tweets were analysed and 62,997 OpenSea asset links were extracted. This data was collected in time period of Jan 1, 2021, to March 30, 2021. Features extracted from this data includes tweet properties (number of likes, retweets etc) and information of users which are involved.



Pipeline for data collection process[2]

- Other important data was extracted from OpenSea market by use of OpenSea API . It contains asset features like verified asset, pre-sale option, transfer, bidding information etc.
- In total Dataset consists of 77 Twitter features and 19 OpenSea Features.

# Problem Description

---

- Given features from tweets mentioning about NFT and data from from opensea Marketplace estimate value of a NFT asset.
- Classify NFT into classes based on avg. selling price such that
  - Class 1 → \$10 to \$100
  - Class 2 → \$100 to \$1000
  - Class 3 → \$1000 to \$10,000
  - Class 4 → \$10,000 to \$100,000
  - Class 5 → \$100,000 to \$1000,0000
- Given the same features estimate the average selling price of NFT asset.



# Proposed approach

---

## Classification

- Nominal multi-class classification
- Ordinal Classification
  - Naive Bayes
  - KNN
  - Decision tree
  - LightGBM
  - XGBoost

## Regression models

- Linear
- Decision tree
- Ridge
- Lasso
- Gradient boosting

# Results : Multi-class Classification

---

	Multi-class Classification			
Metric	KNN	Decision tree	LightGBM	XGBoost
Recall	0.8526	1.0000	0.9277	0.7917
Precision	0.8928	1.0000	0.9307	0.8038
mean accuracy (train)	0.7193	0.9999	0.8587	0.7358
mean accuracy (test)	0.5040	0.6091	0.7114	0.6801
RMSE (train)	0.6436	0.0095	0.3886	0.6723
Accuracy (train)	0.7502	0.9999	0.8587	0.5985
RMSE (test)	0.8767	0.7407	0.5615	0.6695
Accuracy (test)	0.4854	0.6091	0.7114	0.5994



# Results : Ordinal Classification

---

	Ordinal Classification Model				
Metric	Naive Bayes	KNN	Decision tree	LightGBM	XGBoost
Recall	0.7442	0.8278	1.0000	0.9370	0.8545
Precision	0.7833	0.8841	1.0000	0.9389	0.8601
mean accuracy (train)	0.5444	0.7502	0.9999	0.8587	0.7357
mean accuracy (test)	0.5567	0.4854	0.6091	0.7114	0.6866
RMSE (train)	0.7978	0.7172	0.0095	0.3604	0.5512
Accuracy (train)	0.5383	0.7193	0.9999	0.8761	0.7156
RMSE (test)	0.7750	0.8613	0.6934	0.5504	0.5865
Accuracy (test)	0.5467	0.5040	0.6263	0.7149	0.6822

# Results : Reduced features vs. All features

---

linear regression	all features	reduced features
R2-score	1.0000	1.0000
Explained variance score	1.0000	1.0000
Max Error	1,348.4387	7.92E-07
MAE	1.4937	1.4937
MSE	500.5298	9.37E-08
MedAE	1.4937	1.44E-14
RMSE (train)	26.4463	1.18E-07
RMSE (test)	22.3725	1.20E-07

# Results : Regression

---

	Regression Models				
metrics	linear regression	decision tree regression	ridge	lasso	gradient boosting
Max Error	7.92E-07	205,397.0000	8.74E-07	59.6187	213,299.7731
MAE	9.37E-08	121.3339	3.41E-09	0.1877	97.7814
MSE	1.44E-14	13,184,878.2310	5.85E-16	1.6687	13,220,561.8642
MedAE	9.37E-08	121.3339	3.41E-09	0.1877	97.7814
r2 score	1.0000	0.9264	1.0000	1.0000	0.9262
RMSE (train)	1.18E-07	2.72E-14	1.23E-08	0.7596	42.1238
RMSE (test)	1.20E-07	3,631.0988	2.42E-08	1.2918	3,636.0091

# Conclusion

---

- In this project we used classification and regression models on the Twitter -OpenSea dataset.
- Important features that models use are majorly attributed to popularity of NFT assets on social media platform than its actual content.
- The Twitter interactions are very effective in predicting the price of an asset.
- From our results we can report that for classification task *LightBGM* Model and for regression task *Ridge Regression* gives the best results.
- We were able to achieve better accuracy in classification task than authors of the original papers and dataset[2].

# Future Work

---

- The NFT related data is volatile and price of an NFT asset can shown large fluctuations with high frequency.
- A possible future work can be to develop a scrapper which can collect recent data related to NFTs.
- Markets other than OpenSea can also be considered for data curation.
- Similarly a real-time Twitter scrapper can also be developed.
- Our classification and regression models can be applied to real time data for NFTs.

# References:

---

- 1) Crypto traders sold \$17.6 billion worth of NFTs in 2021, including Bored Apes and cryptopunks | Fortune
- 2) Arnav Kapoor, Dipanwita Guhathakurta, Mehul Mathur, Rupanshu Yadav, Manish Gupta, and Ponnurangam Kumaraguru. Tweetboost: Influence of social media on NFT valuation. CoRR, abs/2201.08373, 2022. URL: <https://arxiv.org/abs/2201.08373> , arXiv:2201.0837
- 3) How can blockchain principles help improve ESG systems? | World Economic Forum
- 4) Spottie Wifi Creator Story: The World's First Cryptopunk Rapper