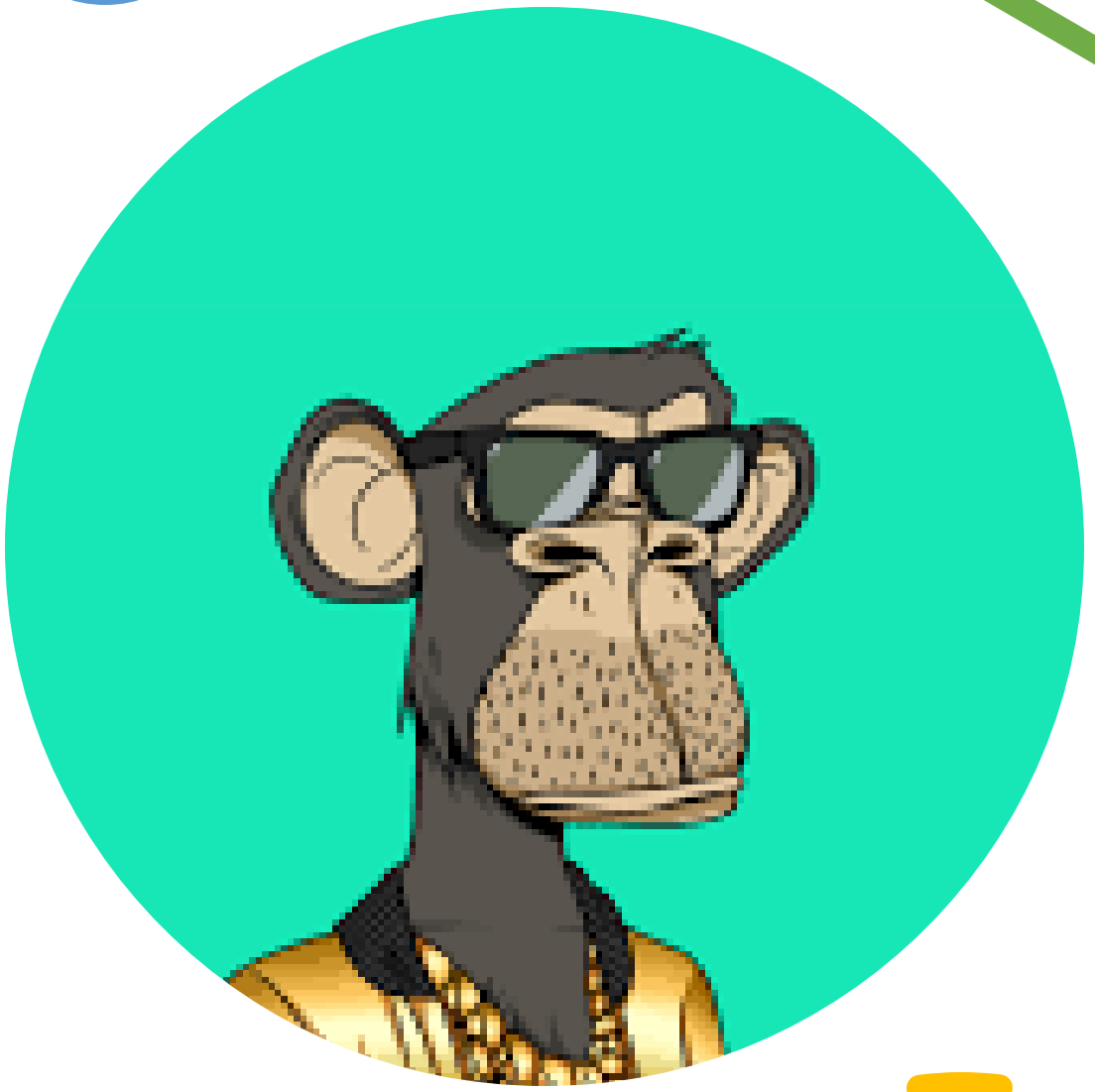


Nfts Sentiment Analysis

- Name: Raheem Nasirudeen Adeleye
- Title – Data Scientist
- Tools: Python, Twint, Pandas, NLTK, Lime, Scikit-Learn, Xgboost
- Bio – I am experienced data scientist with 4+ years, implementing machine learning solutions to solve business problems.

Introduction

This rising cryptocurrency niche recorded over $23 billion in t rading volumes as per the latest DappRadar report. Currently, NFT-
related active wallets account for close to 50% of the total cry pto industry usage, a statistic that will likely increase given th e continued interest in 2022.

Before jumping into the developments and prospects, it is wor th understanding why NFTs are gaining traction across the bo ard. Well, there are many factors behind the sudden surge but the most significant one is the indistinguishable nature of NFT s. Each NFT token has a unique value, making them a suitable on-chain asset to represent digital collectibles such as in-game items or off-
chain assets like property and tokenized stocks.

Methodologies

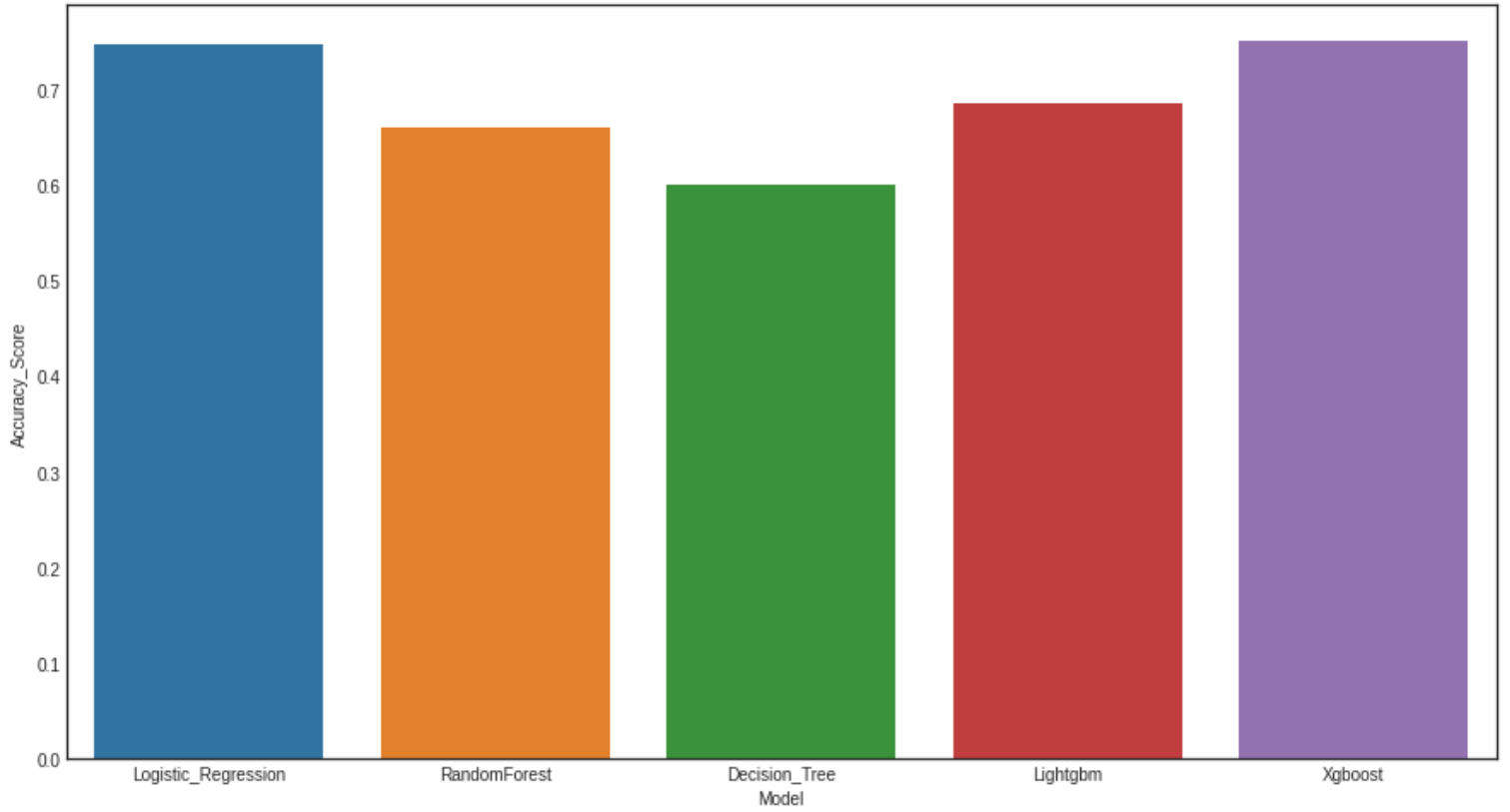
- Data Gathering
- Data Labelling
- Features Engineering
- Exploratory Data Analysis
- Data Cleaning
- Word Cloud Analysis
- Feature Extraction
- Machine Learning
- Interpretable Machine Learning

Observations

- The tweet sentiment consist of 46.3% for neutral, 45.7% for positive and 8.0% for negative.
- Top 10 word count ranges 10-28 words while 53 is the highest word in the tweet.
- Top word character ranges 72-255 words while 270 is the highest word character in the tweet.
- Top 10 number of likes ranges 2-11 while 3121 the highest is number of likes in the tweet.
- Most tweet occur on Friday.
- Positive sentiment with the highest average of number of likes.
- Positive sentiment with the total sum of word count and word character.
- On Friday we have more Positive sentiment with 48.3%.
- On Tuesday we have more Neutral sentiment with 54.2%.
- On Saturday we have 15.2% of negative sentiment, which is the highest percentag e for negative sentiment across the days of week.

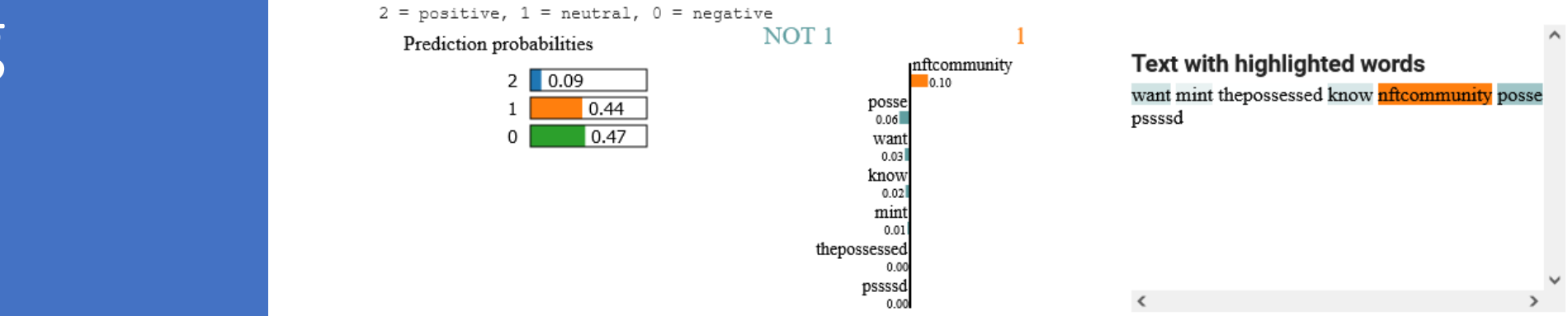
Model Comparison

- The top 2 models are Xgboost with 75.1% and Logistic Regression with 74.6%.



Interpretable Machine Learning

- This highlight the prediction accuracy and features that drives the sentiment.



Summary

The Nfts world is still growing, most ideal collections are with Ethereum blockchain.

The Nfts with the most number of likes have feature time and organization.

Textblob used to analyze the sentiment of each tweet.

Extreme Gradient boosting model (Xgboost) and Linear Model (Logistic Regression) have high impact on the sentiment analysis.

Limitations

Data Collection phase

Lack of domain expertise to guide during the projection.

Future Work

Collection of more data.

Using of various NLP Transformers

Using Shapely values for model interpretability.