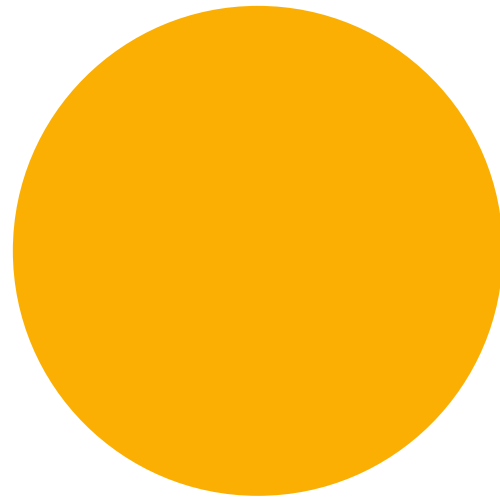


NLP application in crypto trading

Project Senta Group D



Anh Thi Dang | Tony Huang | Satoru Yamada | Eli Muse

In this Presentation

Here's what we'll cover:

Our Hypothesis

.....

Strategy

.....

Model Summary & Evaluation

.....

Findings

.....

Conclusion

.....

To explore for optimised performance

.....



Senta, Defined

Project Senta aims to predict future movement of crypto prices using sentiment scores from public social forum - Reddit & Search engine - Google

Strategy

01

Use trading API to find dates where there were significant price movement on a crypto asset

02

Extract prior days commentary on reddit, and google search and store in a dataframe.

03

Use TF- IDF to analyse text in dataframe for uniqueness.

04

Perform sentiment analysis on the stored text content from google and reddit.

05

Train ML models and choose algo with higher performance in predicting positive returns.

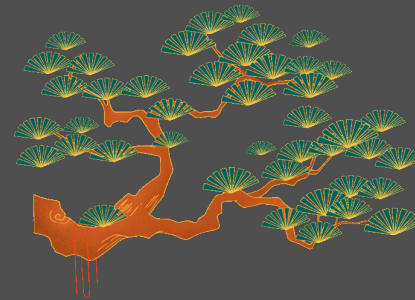
Algorithms Employed

Rationale behind our choice



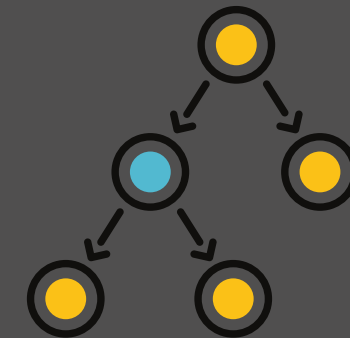
KNN

For simplicity and accuracy for classification problems.



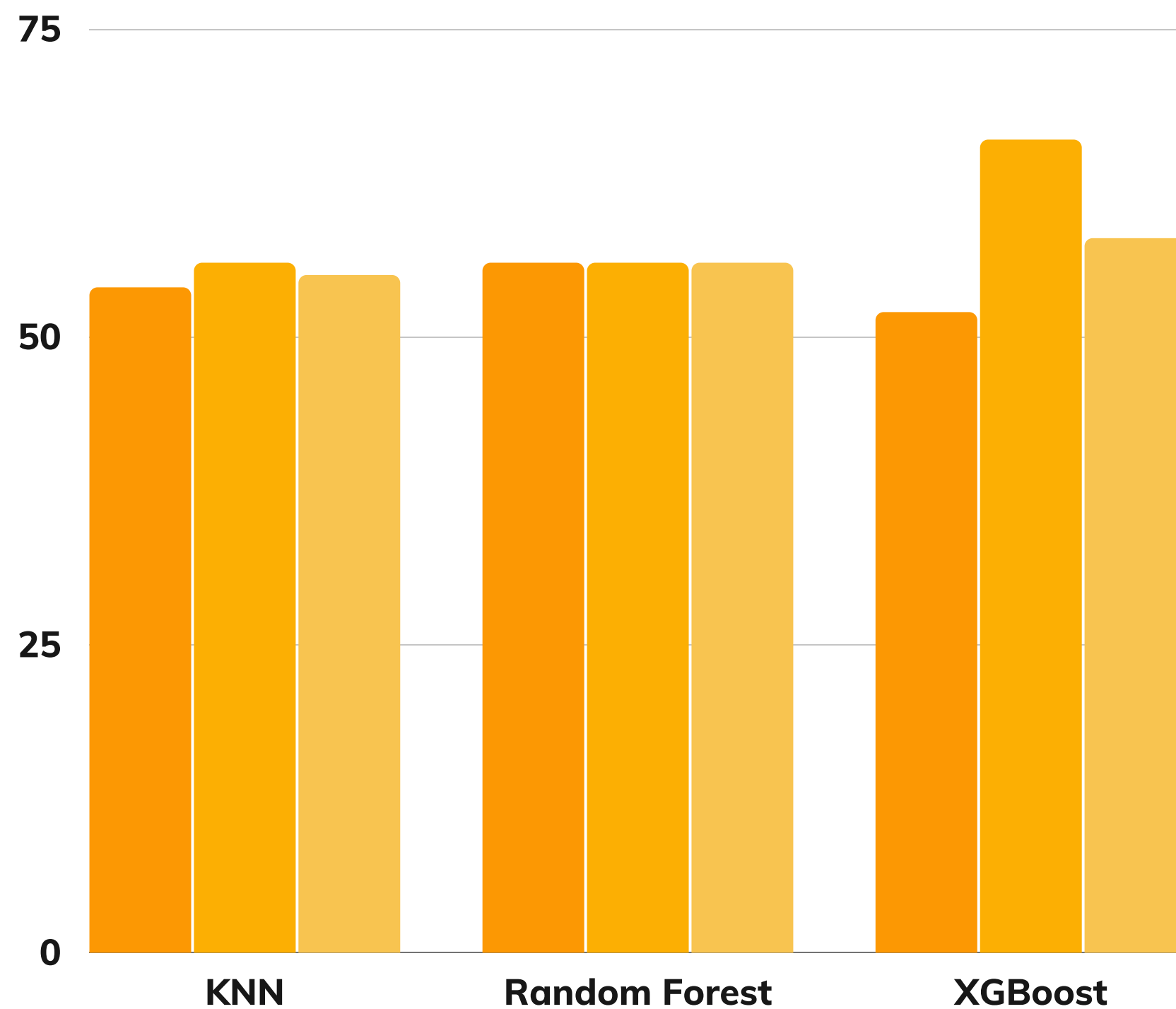
Random Forest

Efficiency at selecting and ranking features by importance



XGBoost

Speed and accuracy



Model Comparison

Summary of performance

XGBoost model scored higher in recall and F1 - score on predicting positive returns based on sentiment scores.



Findings

How sentiment is
impacting crypto assets

Features Importance

- Random Forest and XGBoost ranked google trend and google buzzword score higher than other features.
-

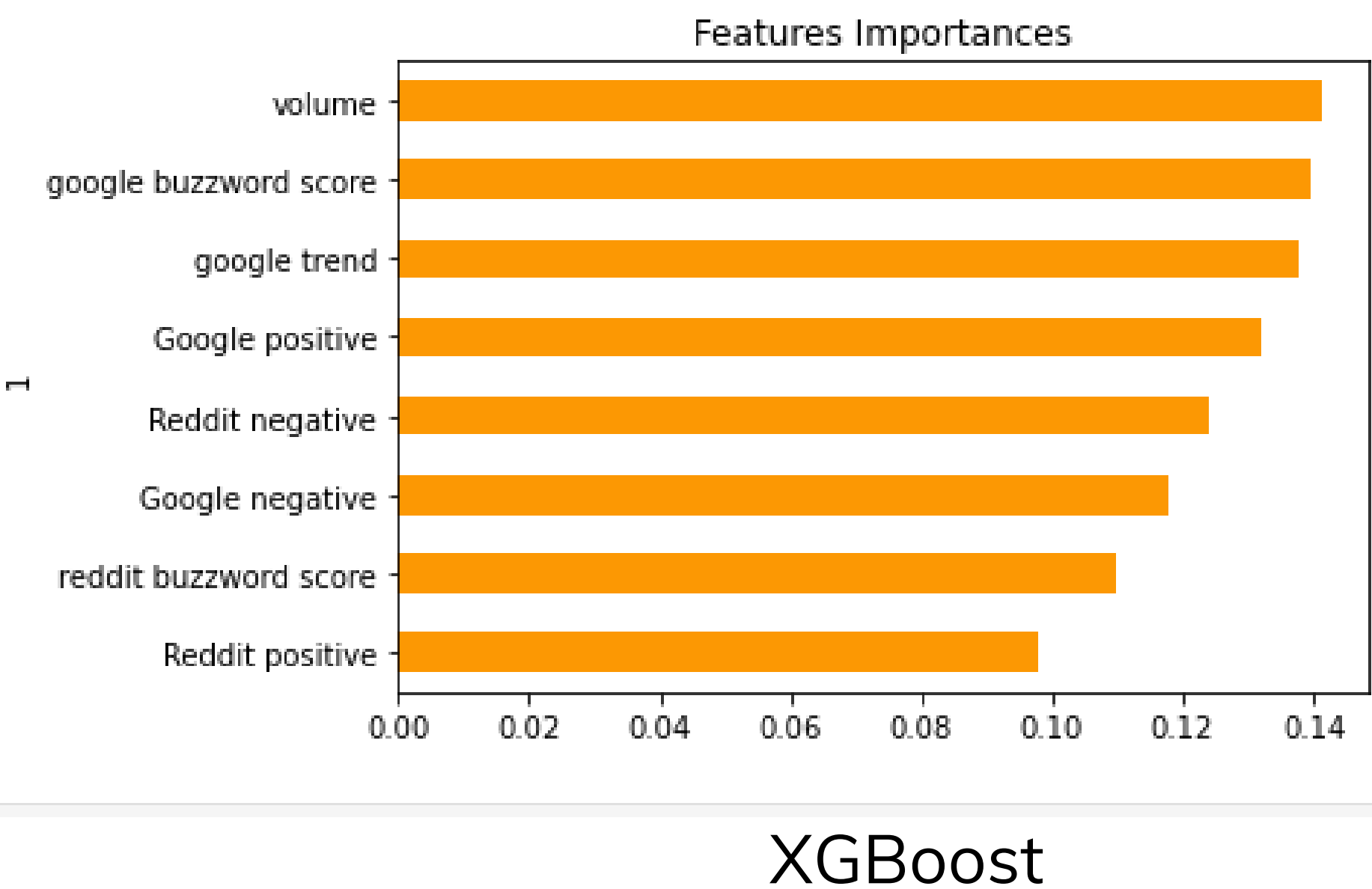
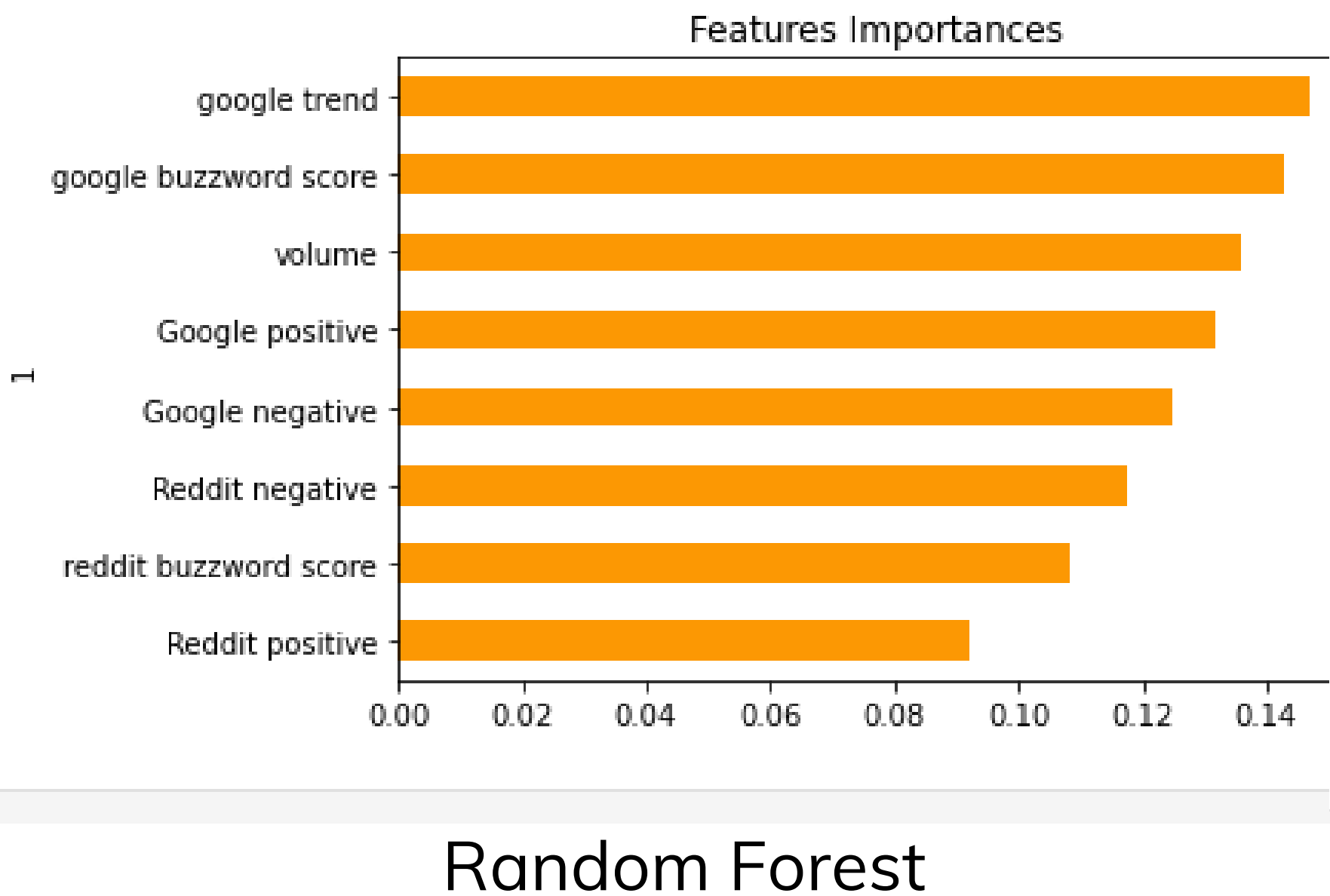
Limitations

- 90% of features used are sentiments score from reddit and google.
-

To Explore

- Price features
- More data from other social media and news platforms
- Volatility

Features Importance



Conclusion

Sentiment effect on crypto price

Based on the performance metric observed in Random Forest and XGBoost, our analysis reveal that public sentiments does have effect on crypto price. However, our model is limited to data collected from Reddit and Google.



To explore for optimized model performance.

1

Effect of price data on model performance

2

Data from other social media platforms and news articles i.e twitter, facebook, new api etc

3

Volatility