Bitcoin Price Forecast and Sentiment Tool

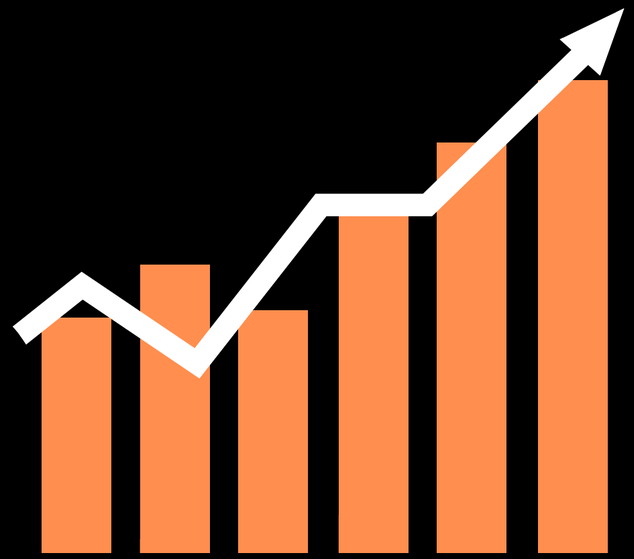
Presented by Mark Benhaim

PROBLEM STATEMENT

Project Focus: This project aims to forecast the price of Bitcoin by examining correlations between technical indicators and public sentiment from Bitcoin-related tweets, leveraging machine learning for optimal model efficiency. The project utilizes historical Bitcoin price data and a large collection of tweets about Bitcoin to build predictive models.

ESTIMATE IMPACT OF SOLUTION

- Uncover the weight of public sentiment on Bitcoin prices
- Equip investors and traders with models offering a comprehensive market view incorporating sentiment through twitter data
- Enrich understanding of social media's role in financial market dynamics.



DATASET
AFTER CLEANING

	Colmn Name	Data Type	Description
0	Time	String	Time in days
1	Open	Float	Open price of bitcoin
2	Percent change in price (close/open)	Float	The percent change of the open price to the cl...
3	Close	Float	Close price of bitcoin
4	BTC Dominance Open	Float	The opening percent marketshare of Bitcoin
5	Percent change in dominance (close/open)	Float	The percent change in Bitcoin marketshare each...
6	BTC Dominance Close	Float	The closing percent marketshare of Bitcoin
7	Volume	Float	The total bitcoin volume per day on the exchan...
8	RSI (relative strength index)	Float	The Relative Strength Index (RSI) is a well ve...
9	HV (historical volatility)	Float	Historical volatility is a statistical measure...
10	Cleantext	String	Text of each individual tweet
11	bitcoin_tweetcount	Integer	Total tweets including the keyword bitcoin per...
12	btc_tweetcount	Integer	Total tweets including the keyword btc per day
13	crypto_tweetcount	Integer	Total tweets including the keyword crypto per day
14	cryptocurrency_tweetcount	Integer	Total tweets including the keyword cryptocurre...
15	project_tweetcount	Integer	Total tweets including the keyword project per...
16	total count_tweetcount	Integer	Total tweets per day
17	Sentiment Polarity_tweetcount	Float	The polarity score of the tweets per day

Preprocessing Overview

Step 1

Date Alignment of both datasets

Step 2

Incorporate a moving average to eliminate null Values

Step 3

Create Sentiment Scores for Tweets Data

Step 4

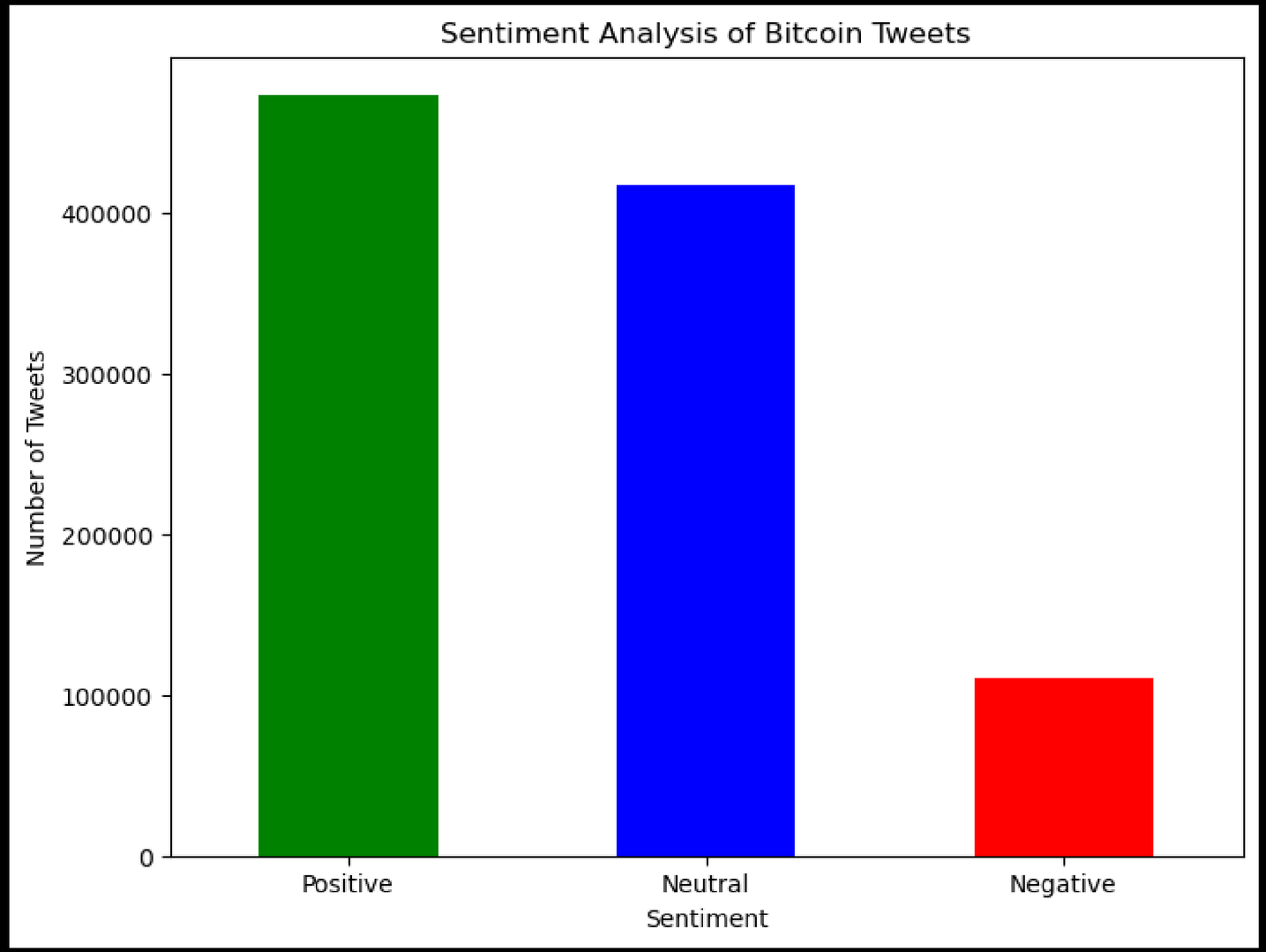
Convert columns in percentages from strings to floats

Step 5

Merge the dataframes

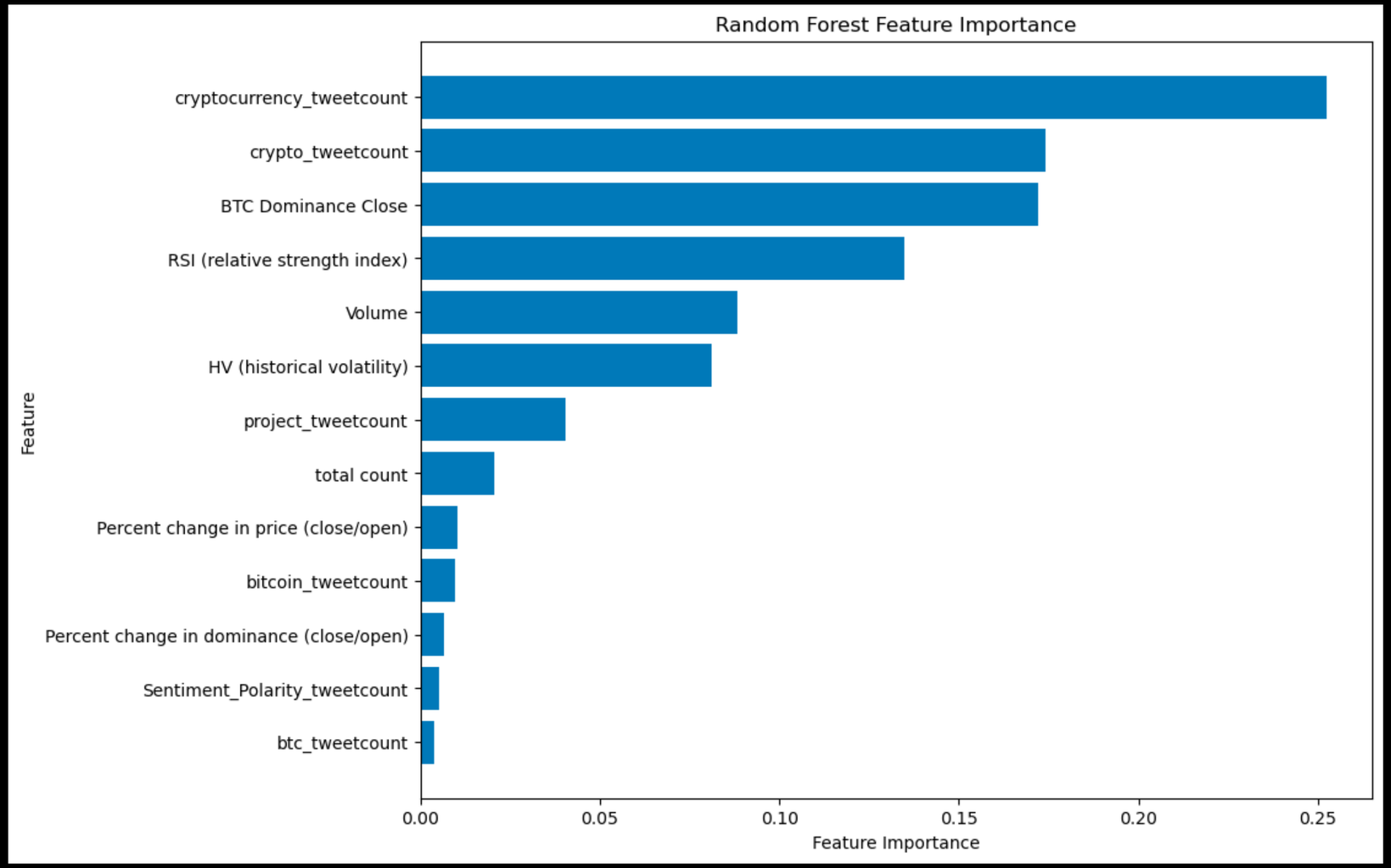
DATA VISUALIZATIONS

Sentiment Analysis of Bitcoin Tweets



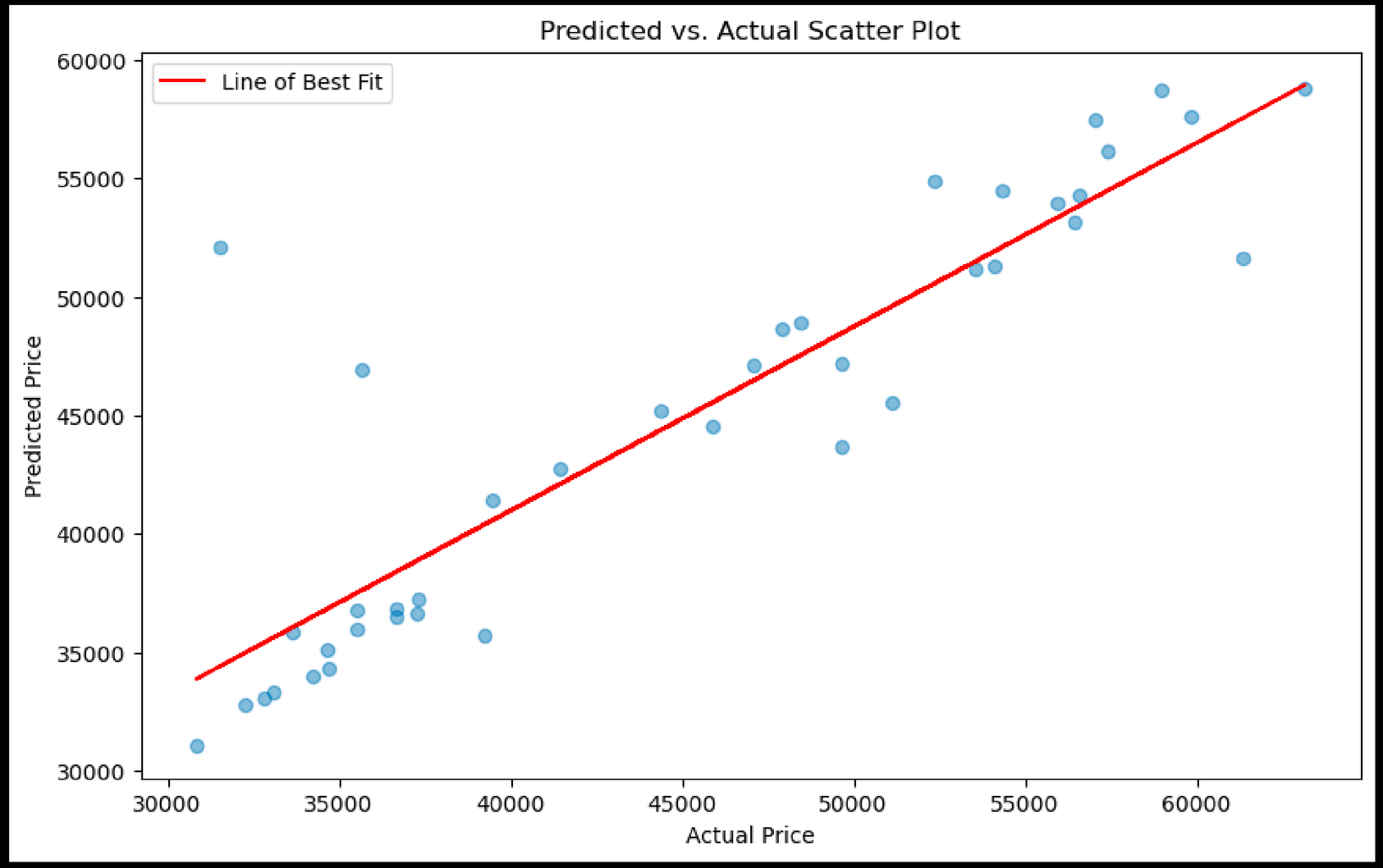
DATA VISUALIZATIONS

Random Forest Feature Importance



DATA VISUALIZATIONS

Predicted vs. Actual Scatter Plot



Model Comparison

R-squared Scores:
Decision Tree: 0.7099
Random Forest: 0.7967
XGBoost: 0.8460
Neural Network 0.6368
Mean Squared Error Scores:
Decision Tree: 28958357.8903
Random Forest: 20295357.6543
XGBoost: 15368174.1812
Neural Network 36254296.1425
Mean Absolute Error Scores:
Decision Tree: 2706.7700
Random Forest: 2427.7511
XGBoost: 2117.9323
Neural Network 3654.3021

Model Performance

XGBoost: highest R-squared score and lowest MSE and MAE, closely followed by Random Forest.

This suggests that these two models have the best predictive performance among the tested models for this dataset.

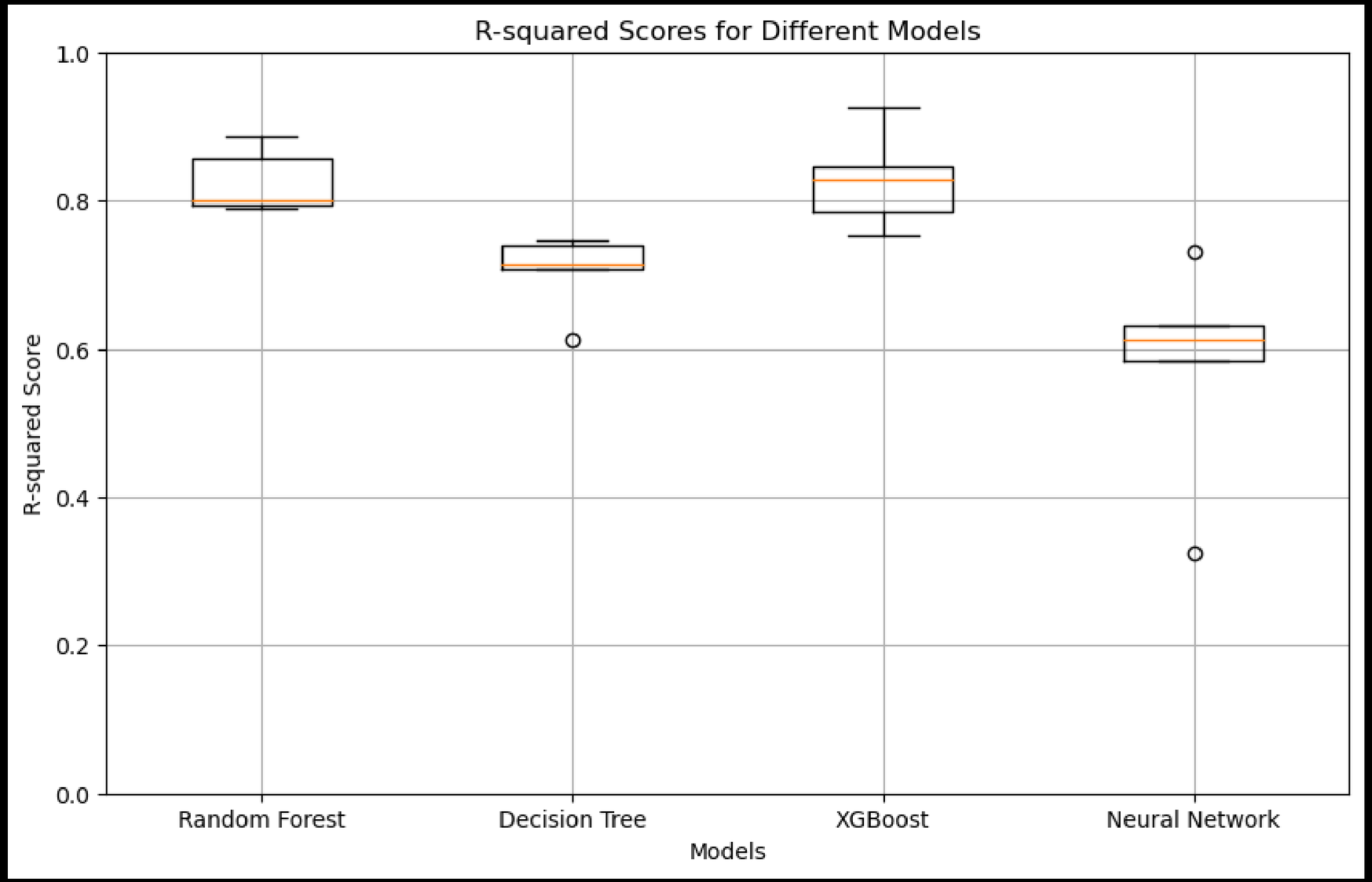
Decision Tree: moderately high R-squared score but lags behind XGBoost and Random Forest.

Neural Network: lowest R-squared score, indicating it has the weakest predictive power among the four models for this dataset.

K-Fold Cross Validation: similar results same outcome

DATA VISUALIZATIONS

R-squared Scores for Different Models



Neural Network Vulnerability

Data Size:
- Neural Networks thrive with ample data.
- Our dataset's limited size hinders the model's ability to capture intricate relationships.

Result:
- Reduced R-squared values
- Increased MSE
- Increased MAE

Recommendation:
- Further investigation into the model is crucial as we gather more data.

Model Interpretation

Key Takeaways:

- Strong correlation found between Bitcoin tweets' sentiment and price patterns
- Twitter sentiment acts as a real-time indicator, potentially preceding traditional indicators

Recommendations:

1. Twitter Data: Integrate into analysis for an edge in market movement predictions
2. XGBoost: Prioritize for immediate forecasting
3. Neural Network: Refine model through enlarging the dataset

Final Thought:

- Blending financial metrics with Twitter sentiment data can provide deeper insights, potentially maximizing returns and minimizing risks in the ever-evolving cryptocurrency landscape