Bitcoin Price Prediction & Sentiment Tool

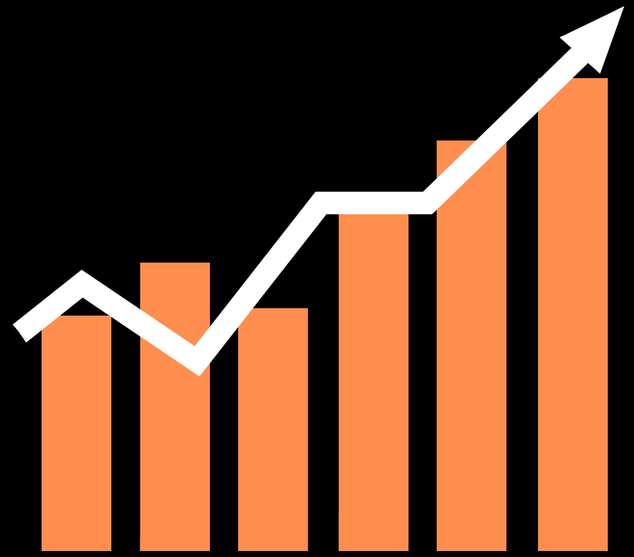
Presented by Mark Benhaim

PROBLEM STATEMENT

Project Focus: This project aims to forecast the price of Bitcoin and analyze sentiment related to Bitcoin on Twitter using machine learning techniques. The project utilizes historical Bitcoin price data and a large collection of tweets about Bitcoin to build predictive models.

ESTIMATE IMPACT OF SOLUTION

This study has the potential to reveal insights about the influence of public sentiment on Bitcoin prices. The predictive models built from this research could be beneficial to investors, providing a more holistic view of market conditions by incorporating public sentiment. This project also contributes to the broader field of study regarding the influence of social media on financial markets.



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 Procedures

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

Baseline Modelling

Sentiment Analysis

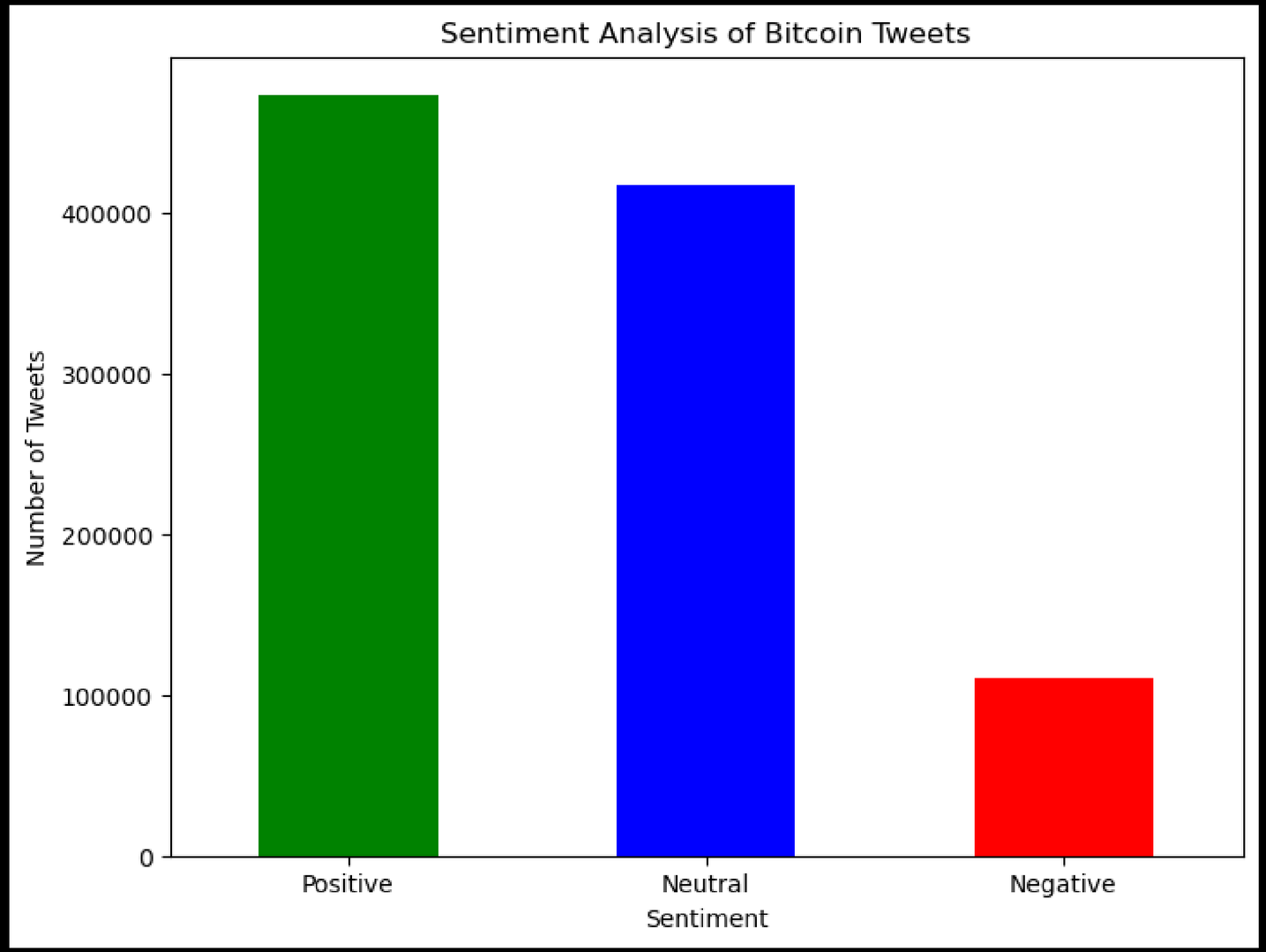
Feature Importance

Random Forests Regressor:

R-squared: 0.7966865551362414

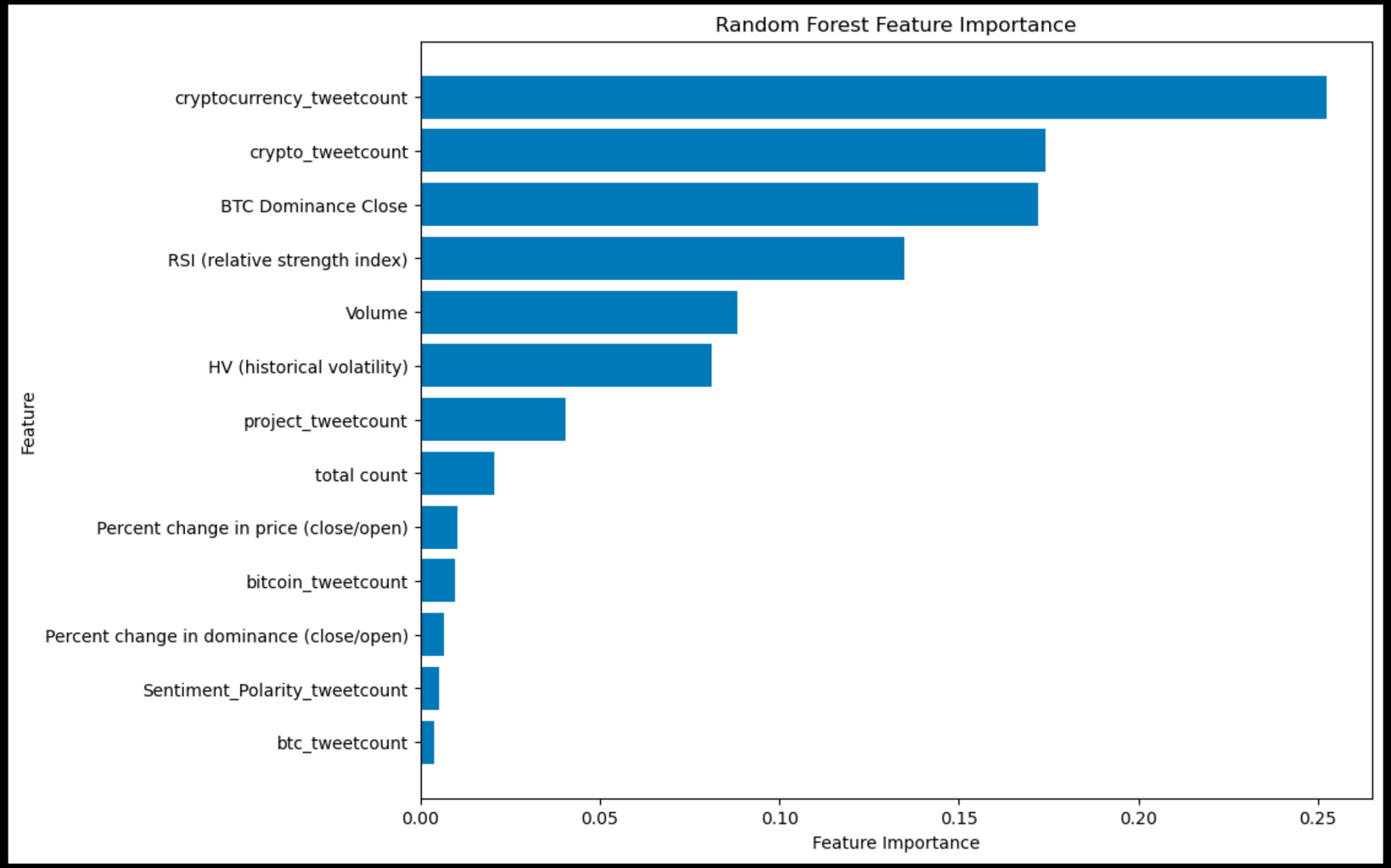
DATA VISUALIZATIONS

Sentiment Analysis of Bitcoin Tweets



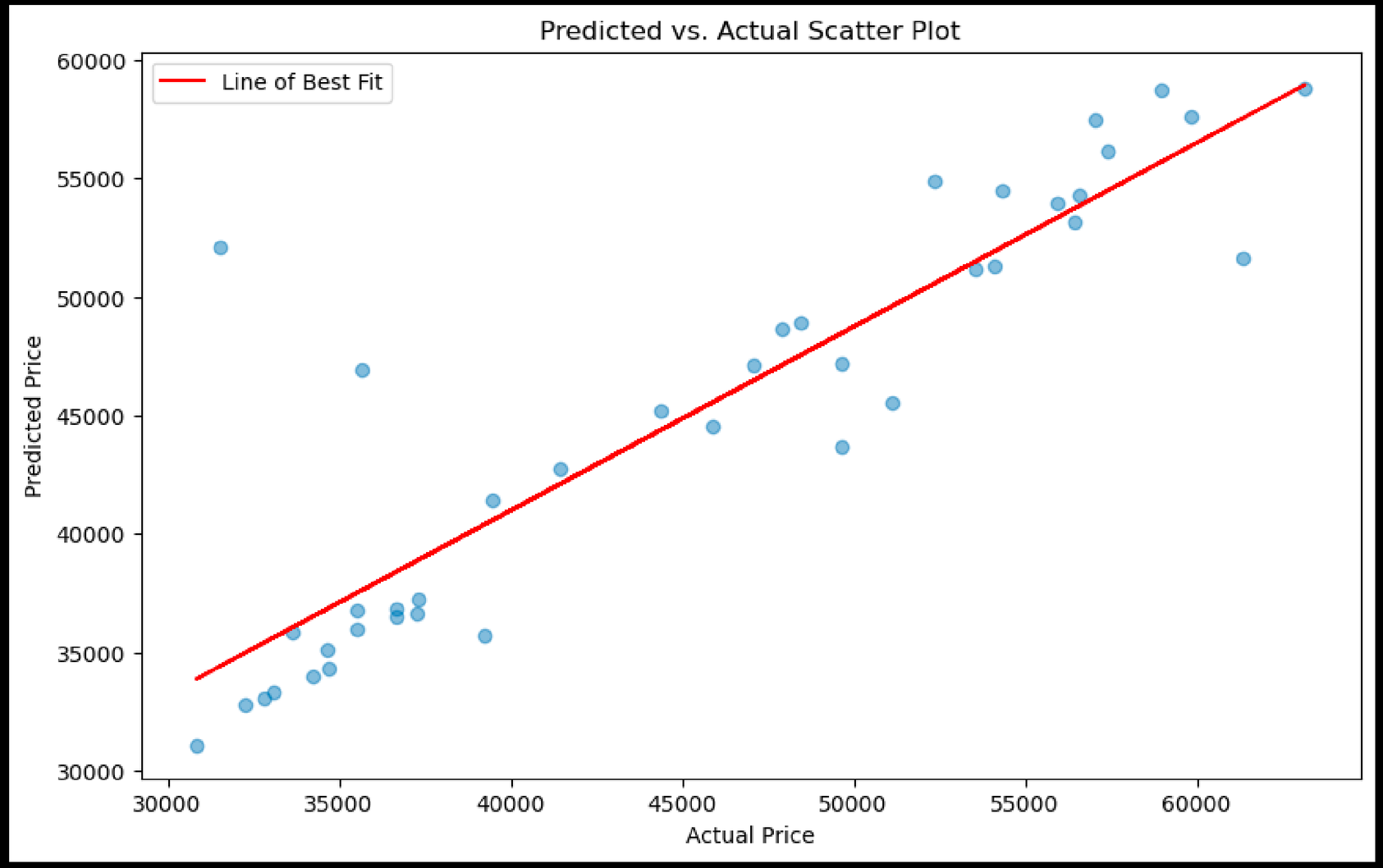
DATA VISUALIZATIONS

Random Forest Feature Importance



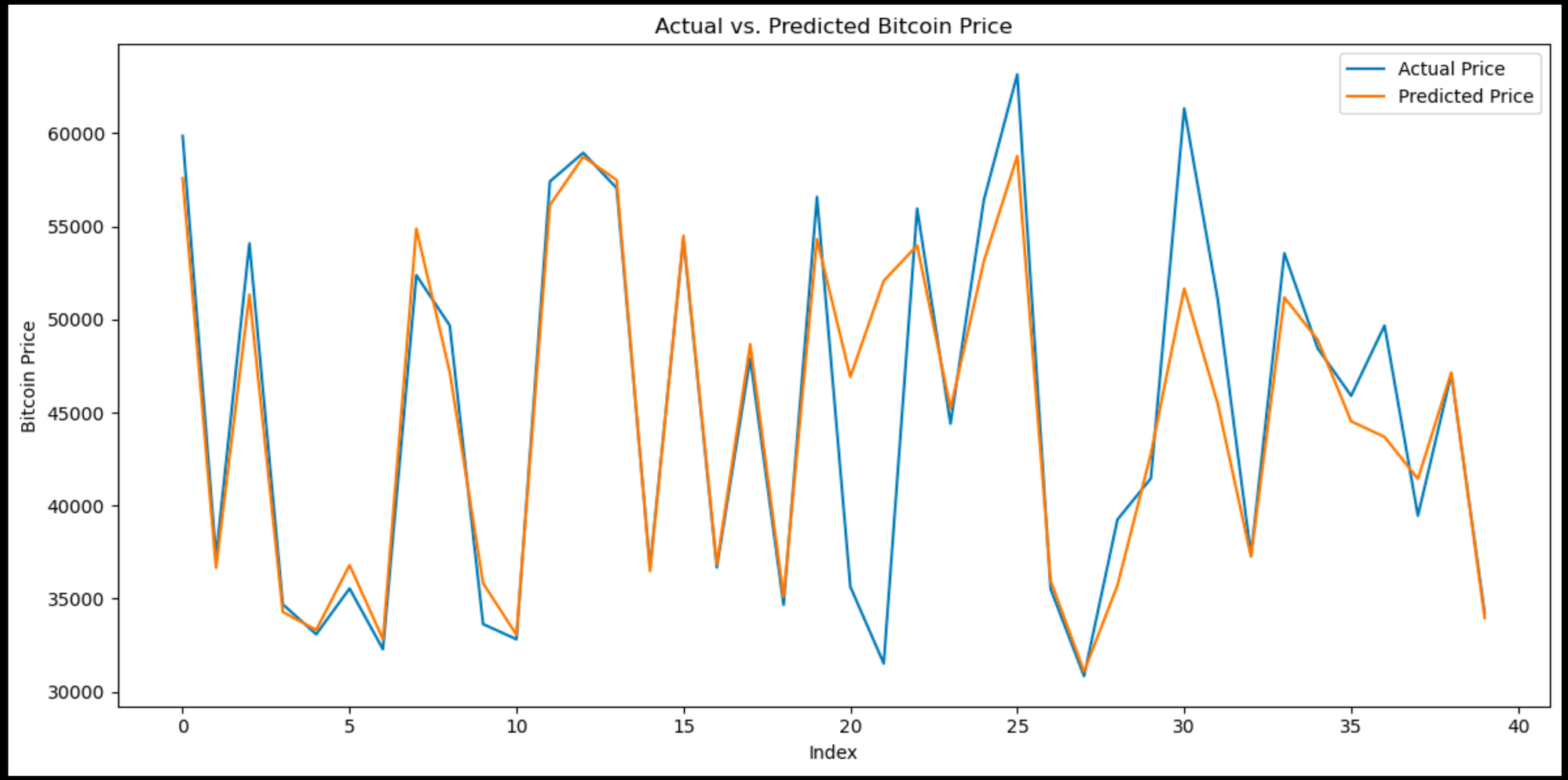
DATA VISUALIZATIONS

Predicted vs. Actual Scatter Plot



DATA VISUALIZATIONS

Actual vs Predicted Bitcoin Price from a Temporal Perspective



NEXT STEPS

- More advance ML models (neural networks)
- Model Tuning

+

- Interactive Dashboard
- Model deployment
- discussion