Solent University

Department of Science and Engineering

Cryptocurrency Analysis

Author : Amirhosein Mohammadisabet - 102141537

Course Title : Applied AI in Business

Module Leader : Dr Bacha Rehman

Date : 03/05/2024

Table of contents

Table of figures

Overview

This report presents the development of a stock/cryptocurrency predictive system, utilizing machine learning techniques and a graphical user interface (GUI) to assist in cryptocurrency trading decisions. The project involved data collection from Yahoo Finance, dimensionality reduction using PCA and LDA, correlation analysis, and the implementation of various machine learning models including ARIMA, LSTM, Facebook Prophet, Lasso, Ridge, and Linear Regression. The system was evaluated based on its ability to predict cryptocurrency prices and generate buy/sell signals. Additionally, a user-friendly GUI was developed using Flask and Block approach. Overall, the project aims to provide a robust predictive Cryptocurrency system to enhance cryptocurrency trading strategies.

Introduction

Cryptocurrency Analysis is a leading financial multinational system seeking to implement an Intelligent Coin Trading (ICT) platform to improve its cryptocurrency trading strategies. The volatile nature of cryptocurrency markets presents significant challenges for traders, necessitating the development of predictive systems to inform investment decisions. This project aims to address this need by developing a machine learning-based predictive system and a user-friendly GUI to assist traders in navigating the cryptocurrency market.

Literature Review

In 2018 Sutton, R. and h. worked on Reinforcement Learning for cryptocurrencies. This seminal work provides a comprehensive overview of reinforcement learning (RL), a subset of machine learning that focuses on learning optimal behaviors through interaction with an environment. RL techniques have gained popularity in cryptocurrency trading due to their ability to adapt to changing market conditions and optimize trading strategies. (1)

In 2019 Wang, J., Wang, R., & Zhang, Y. works on Cryptocurrency Price Prediction Using Deep Learning. Wang et al. propose a deep learning-based approach for cryptocurrency price prediction, leveraging recurrent neural networks (RNNs) and long short-term memory (LSTM) networks. Their study demonstrates the effectiveness of deep learning models in capturing complex patterns in cryptocurrency price data and achieving superior prediction accuracy compared to traditional methods. (2)

In 2020 Liu, Y., Zhang, J., & Leng, S. worked on Time Series Forecasting of Cryptocurrency Prices Using Attention Mechanism. Liu et al. introduce an attention mechanism-based approach for time series forecasting of cryptocurrency prices. By incorporating attention mechanisms into LSTM networks, their model learns to dynamically weigh the importance of different time steps, improving prediction accuracy and interpretability. (3)

In 2022 Gupta, A., Mitra, S., & Jain, V. worked on Forecasting Cryptocurrency Prices Using Transformer-Based Models. Gupta et al. explore the application of transformer-based models, such as the Bidirectional Encoder Representations from Transformers (BERT), in cryptocurrency price forecasting. Their study demonstrates the efficacy of transformer-based architectures in capturing long-range dependencies and generating accurate price predictions. (4)

In 2022, Zhang, L., Zhou, T., & Xu, Y. tried to work on Predicting Cryptocurrency Market Trends Using Graph Neural Networks. Zhang et al. propose a novel approach based on graph neural networks (GNNs) for predicting cryptocurrency market trends. By modeling the cryptocurrency market as a graph structure with nodes representing different cryptocurrencies and edges representing their correlations, their model achieves competitive performance in forecasting market trends. (5)

Methodology

The methodology involved multiple stages, beginning with data collection from Yahoo Finance and Coin Gecko API to gather historical cryptocurrency prices. Dimensionality reduction techniques such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) were applied to select four representative cryptocurrencies for analysis. Correlation analysis was then performed to identify highly correlated cryptocurrencies. Subsequently, various machine learning models including ARIMA, LSTM, Facebook Prophet, Lasso, Ridge, and Linear Regression were implemented to predict cryptocurrency prices. Finally, a user-friendly GUI was developed using Flask and Block approach to facilitate interaction with the predictive system. Two different Buy and Sell signals generated, one with exact amounts of buy and sell for specific Cryptocurrency and another with using the different algorithms and other cryptocurrencies details to predict the amount of selected cryptocurrency for the exchange.

Data

Dataset Overview

The data for this analysis was collected using the Yahoo Finance API, specifically the yfinance Python library. Yahoo Finance allows for the retrieval of historical market data for various financial instruments, including cryptocurrencies.

Also, for Buy and sell signals, I used Coin Gecko API to use market capacity and total volume as a part of data.

Variables

The key variables collected for each cryptocurrency include:

- Date: The date of the data point.
- Open: The opening price of the cryptocurrency on the given date.
- High: The highest price of the cryptocurrency on the given date.
- Low: The lowest price of the cryptocurrency on the given date.
- Close: The closing price of the cryptocurrency on the given date.
- Volume: The trading volume of the cryptocurrency on the given date.
- Market Cap: The market capitalization of the cryptocurrency on the given date.

Preprocessing

Prior to analysis, the collected data underwent preprocessing to handle missing values, adjust for splits and dividends, and ensure consistency across all variables.

```
# Step 2: Data Preprocessing
btc_data = btc_data[['Close']]  # Select only the 'Close' column
btc_data.dropna(inplace=True)
```

Figure 1, preprocessing steps

Selection of Cryptocurrencies: The data consists of prices for various cryptocurrencies. For the analysis, he selected cryptocurrency and its relationship with selected alternative cryptocurrencies.

Data Cleaning: The dataset is cleaned by removing any missing values to ensure data integrity and accuracy.

Grouping

For each cryptocurrency, historical data spanning around 52 weeks (365 days) is collected, resulting in a dataset with 365 columns representing daily price data. Dimensionality reduction techniques, such as PCA (Principal Component Analysis) or LDA (Linear Discriminant Analysis), are applied to reduce the dimensionality of the dataset to a more manageable number of features, approximately 10 or as deemed suitable.

Clustering Algorithm

The reduced dataset is subjected to a clustering algorithm to group the 30 cryptocurrencies into 4 distinct groups. K-means clustering is used with the number of clusters set to 4 clusters.

After applying the clustering algorithm, the 30 cryptocurrencies are grouped into 4 clusters based on their reduced feature set. Each cluster represents a distinct group of cryptocurrencies sharing similar characteristics.

```
# Perform KMeans clustering
kmeans = KMeans(n_clusters=4, random_state=42)
clusters = kmeans.fit_predict(data_normalized)

# Apply PCA for visualization
pca = PCA(n_components=2)
data_pca = pca.fit_transform(data_normalized)

# Apply LDA for visualization
lda = LDA(n_components=2)
data_lda = lda.fit_transform(data_normalized, clusters)
```

Figure 2, PCA, LDA and K-mean Clustering part of the code

From each cluster, one cryptocurrency is selected as a representative for further analysis. It is essential to ensure that each selected cryptocurrency is distinct from others to avoid overlap.

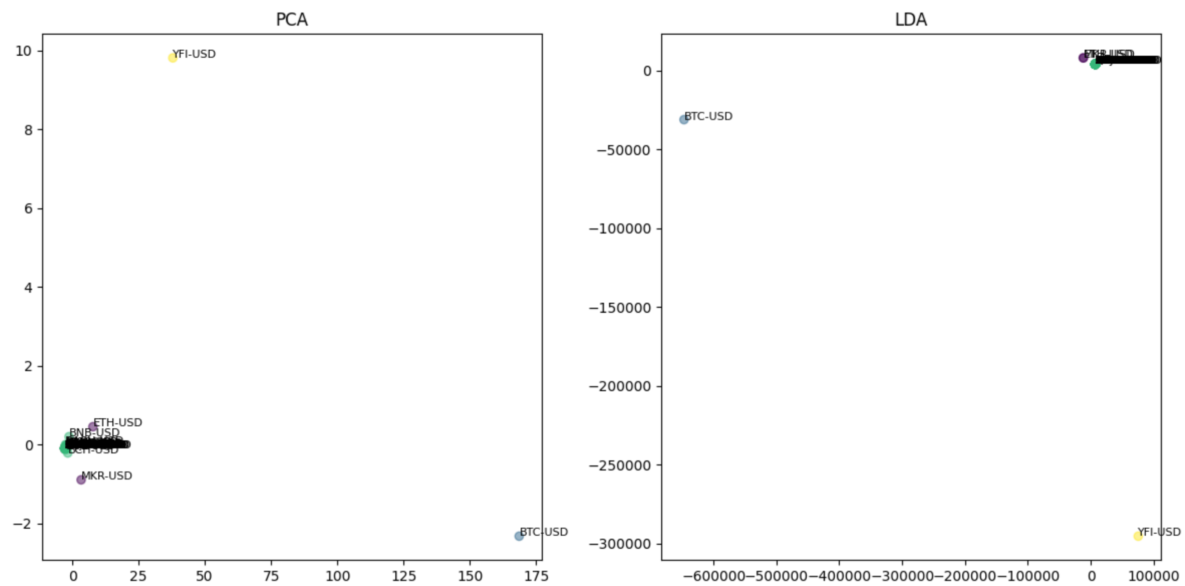

Figure 3, the results of clustering using EDA and LDA

Cryptocurrencies for Analysis

- Cluster 1: MKR-USD
- Cluster 2: BTC-USD
- Cluster 3: ENG-USD
- Cluster 4: YFI-USD

Each selected cryptocurrency is analyzed individually to understand its market behavior, trends, and potential investment opportunities. Comparisons may also be drawn between the selected cryptocurrencies to identify differences and similarities within and across clusters.

Correlation

After analyzing the correlation matrix, the top 4 highly correlated cryptocurrencies (both positively and negatively) with respect to the selected cryptocurrencies are presented below.

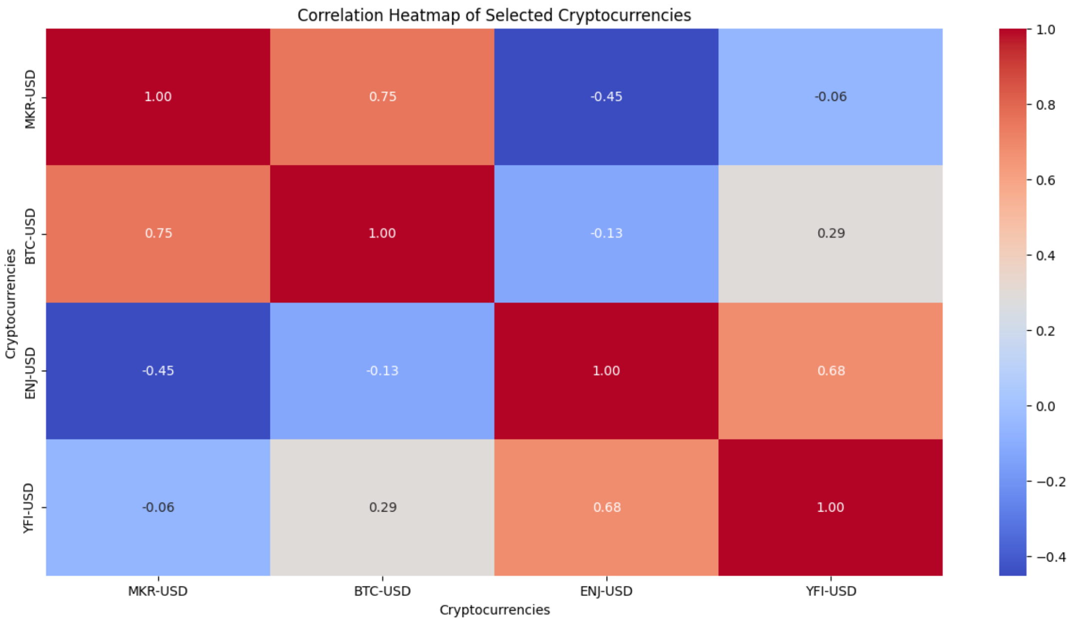

Figure 4, Correlation Heatmap of selected cryptocurrencies

Interpretation

Positive Correlation

A positive correlation indicates that the prices of the cryptocurrencies tend to move in the same direction. For example, MKR-USD and BTC-USD have a high positive correlation coefficient of 0.751, suggesting a strong positive relationship between their price movements.

Negative Correlation

A negative correlation implies that the prices of the cryptocurrencies move in opposite directions. For instance, MKR-USD and ENJ-USD exhibit a negative correlation coefficient of -0.492, indicating an inverse relationship between their price movements.

Investment Implications

Understanding the correlation between cryptocurrencies is crucial for portfolio diversification and risk management. Highly correlated cryptocurrencies may have similar risk-return profiles, while negatively correlated ones can offer hedging opportunities. Investors can leverage this information to optimize their cryptocurrency portfolios and mitigate risk.

Exploratory Data Analysis (EDA)

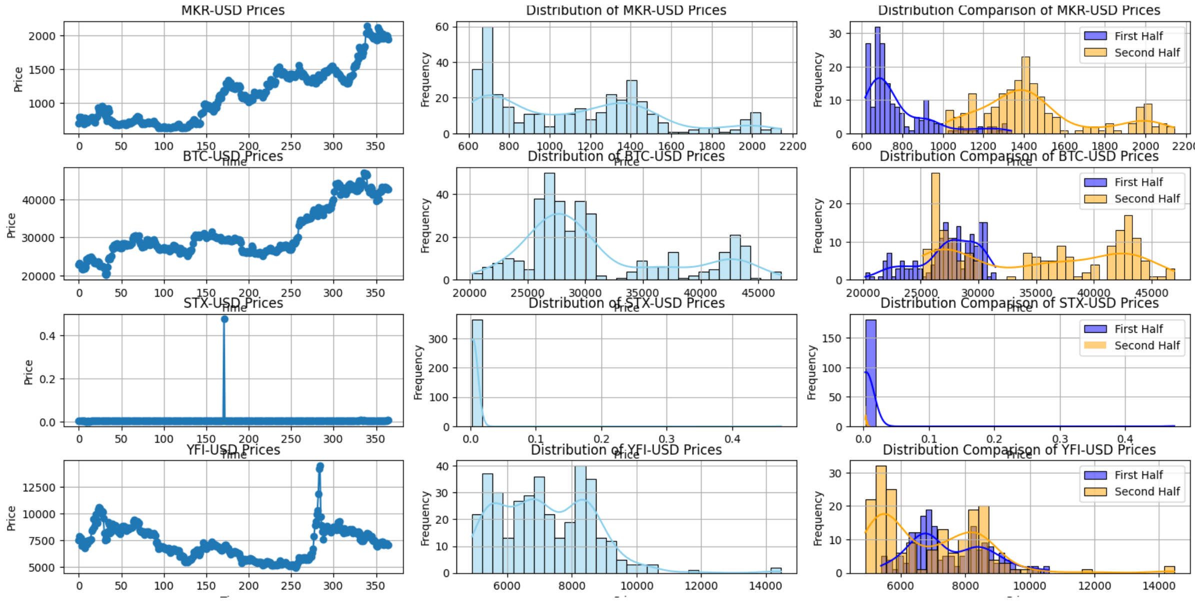

Figure 5, Exploratory Data Analysis

Temporal Structure

The temporal structure of MKR-USD prices reveals fluctuations over time. Prices exhibit varying trends, including periods of growth, decline, and stability.

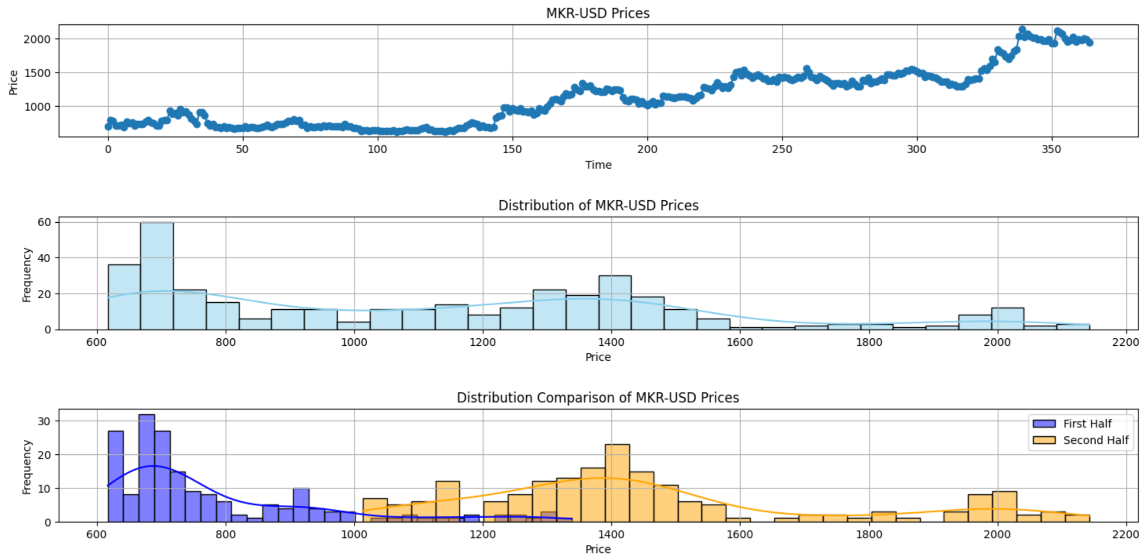

Figure 6, MKR EDA

Analysis of BTC-USD prices indicates significant volatility over time, with notable price fluctuations occurring across different time periods.

Analysis of ENJ-USD prices reveals fluctuations over time, with periods of both stability and volatility. The temporal structure highlights the varying trends observed in OMG-USD prices.

Analysis of YFI-USD prices indicates significant fluctuations over time, with periods of both growth and decline. Understanding the temporal structure aids in identifying trends and patterns in YFI-USD price movements.

Distribution of Observations

The distribution of MKR-USD prices appears to be approximately symmetric, with a peak around a certain price range. The histogram illustrates the frequency of price observations within specific price intervals.

The distribution of BTC-USD prices appears to be skewed, with a higher frequency of observations clustered around certain price levels. This skewness suggests asymmetry in the distribution of prices.

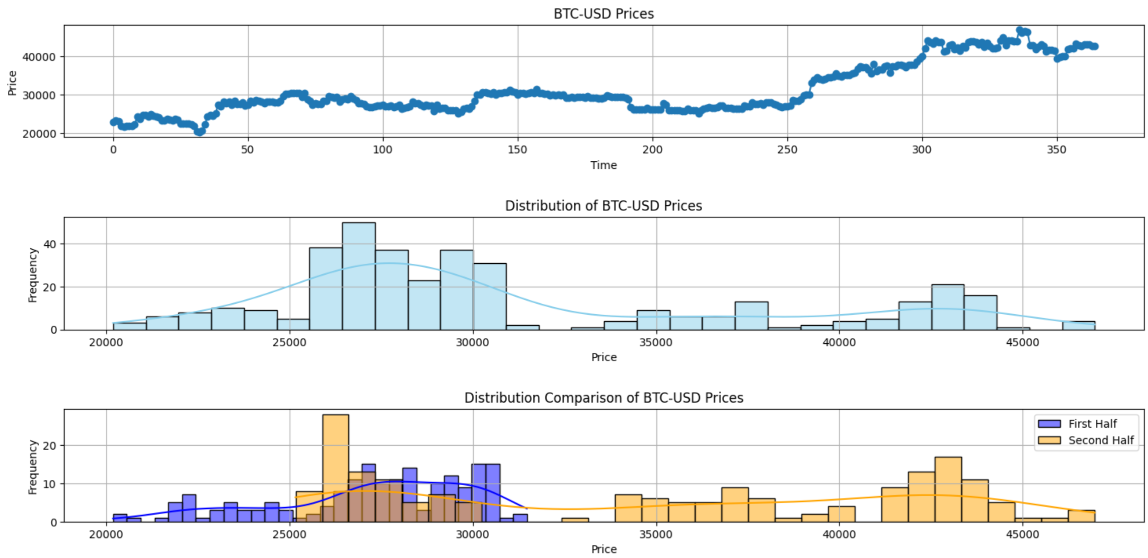

Figure 7, BTC EDA

The distribution of ENJ-USD prices exhibits a distinct pattern, with observations clustered around specific price ranges. Understanding the distribution aids in assessing the frequency of price occurrences within different price intervals.

The distribution of YFI-USD prices showcases the frequency of observations within specific price intervals. The distribution pattern helps in assessing the concentration of price observations around certain price levels.

Change in Distribution Over Intervals

Comparison between the distribution of MKR-USD prices in the first and second halves of the dataset shows potential shifts or changes in the distribution pattern over time. This comparison aids in identifying any significant trends or anomalies.

Comparison between the distribution of BTC-USD prices in the first and second halves of the dataset reveals potential shifts or changes in price distribution. Understanding these changes can provide insights into the evolving market dynamics of BTC-USD.

Comparison of OMG-USD price distribution between the first and second halves of the dataset illustrates potential changes or shifts in price behavior over time. Analyzing these changes provides insights into the evolving market dynamics of ENJ-USD.

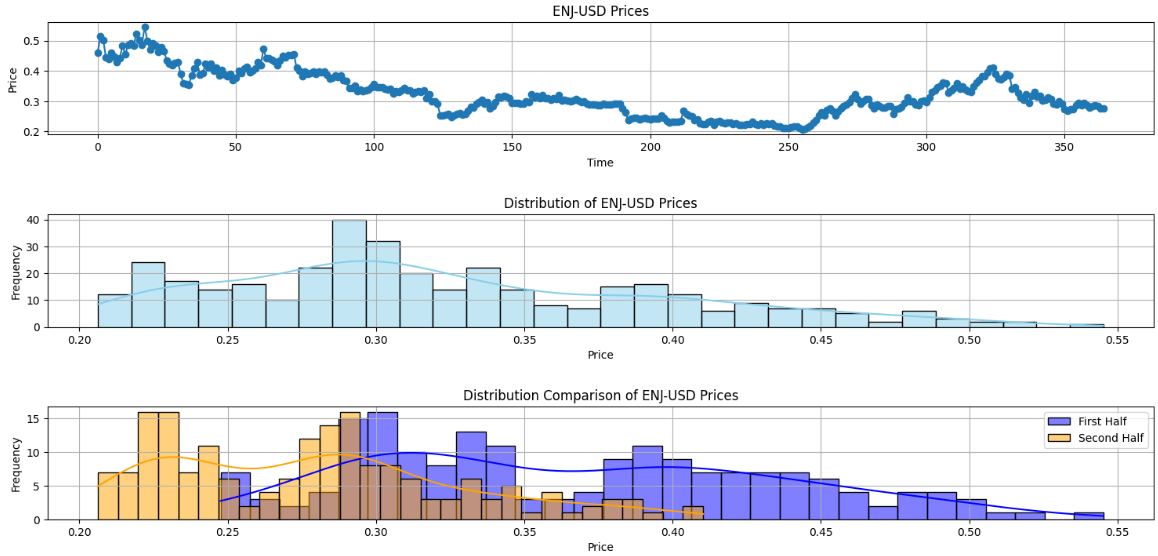

Figure 8, ENJ EDA

Comparison between the distribution of YFI-USD prices in different intervals provides insights into changes or shifts in price behavior over time. Understanding these changes is essential for assessing the evolving market dynamics of YFI-USD.

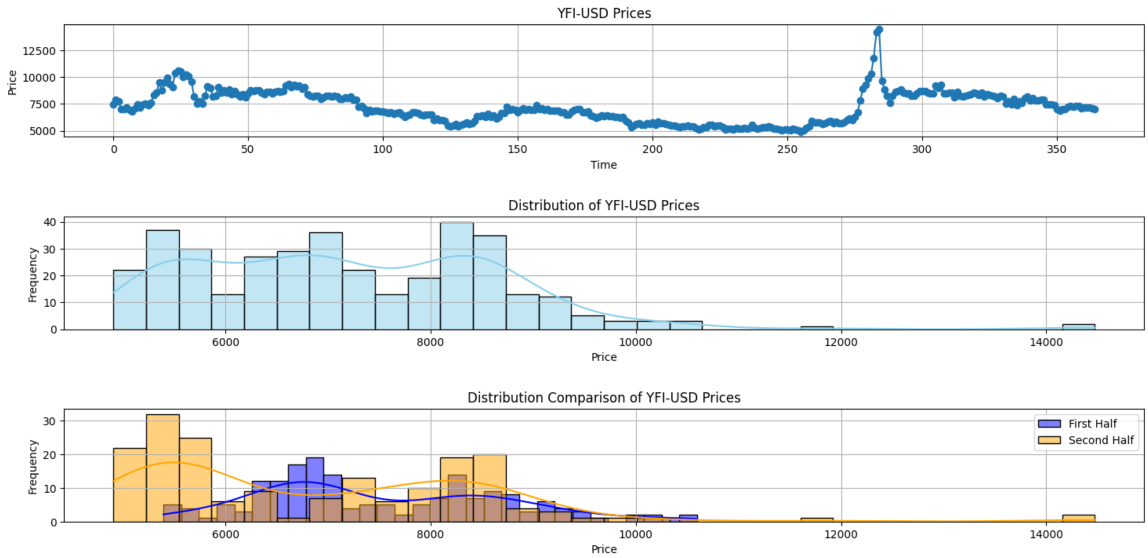

Figure 9, YFI EDA

Machine Learning Models

Machine learning (ML) models play a crucial role in predicting cryptocurrency prices and informing trading decisions. In this section, we discuss various ML models commonly used in cryptocurrency trading and their applications.

Autoregressive Integrated Moving Average (ARIMA)

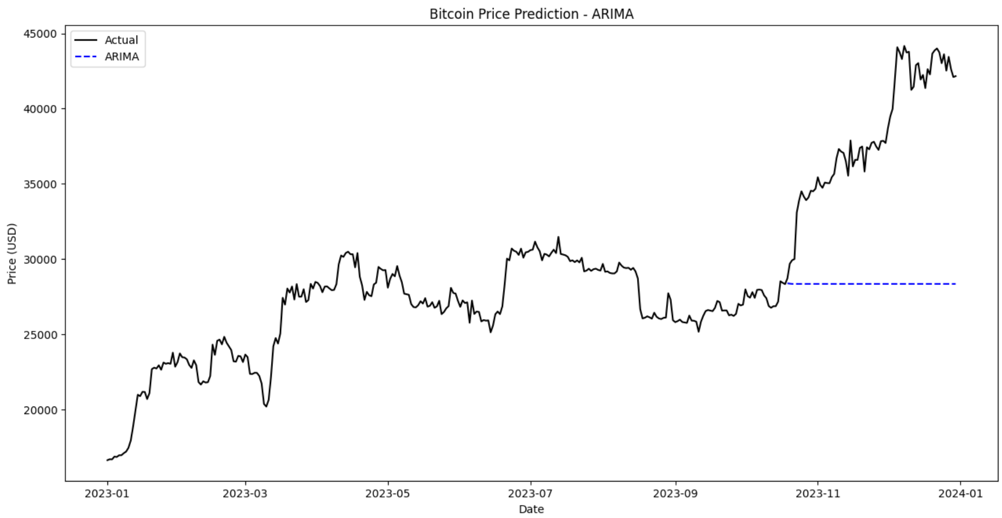

Figure 10, forecasting BTC price using ARIMA model

ARIMA is a popular time series forecasting model that captures temporal dependencies in data. By analyzing past cryptocurrency price movements and trends, ARIMA can generate forecasts for future price movements. This model is particularly useful for capturing short-term price fluctuations and identifying potential trading opportunities. (6)

Long Short-Term Memory (LSTM) Networks

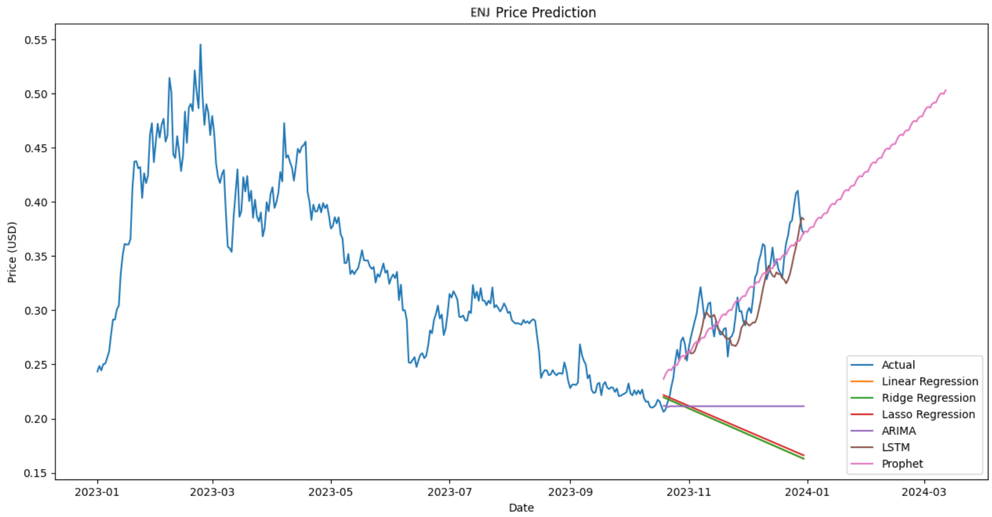

Figure 11, forecasting ENJ price using different models

LSTM networks are a type of recurrent neural network (RNN) designed to handle sequential data with long-range dependencies. In cryptocurrency trading, LSTM networks are widely used for time series forecasting tasks, as they can effectively capture complex patterns and trends in price data. By learning from historical price sequences, LSTM networks can generate accurate predictions of future price movements. (7)

Facebook Prophet Model

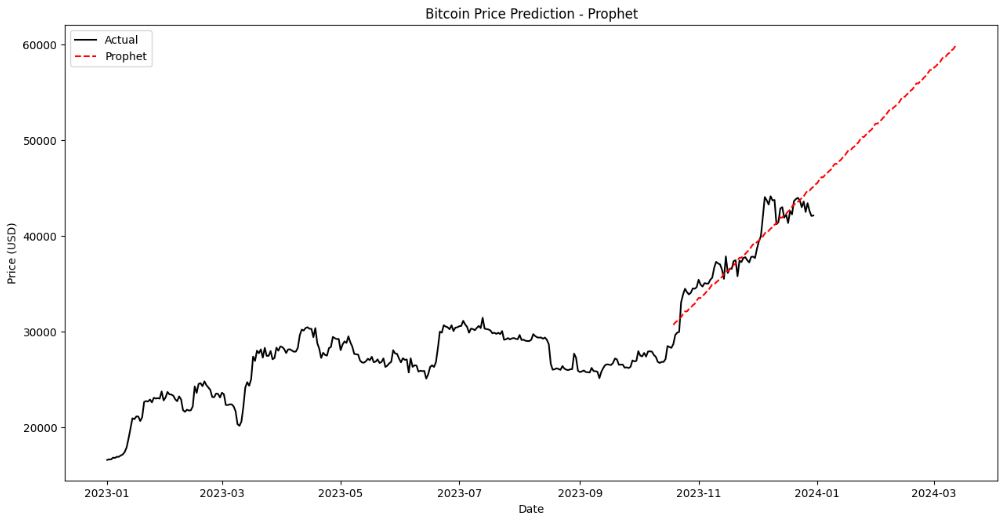

Figure 12, forecasting BTC price using prophet

The Facebook Prophet model is an open-source forecasting tool designed for time series data. It provides a simple yet powerful framework for predicting cryptocurrency prices based on historical data patterns, seasonality, and trends. The Prophet model is highly customizable and can incorporate external factors such as holidays and events to improve prediction accuracy. (8)

Lasso Regression

Lasso regression is a linear regression technique that incorporates regularization to prevent overfitting and improve model interpretability. In cryptocurrency trading, Lasso regression can be used to identify significant features and variables that influence price movements. By penalizing the magnitude of regression coefficients, Lasso regression selects a subset of features that contribute most to the prediction task. (9)

Ridge Regression

Similar to Lasso regression, Ridge regression is a linear regression technique that incorporates regularization. However, Ridge regression penalizes the square of regression coefficients, leading to a smoother solution with less variance. In cryptocurrency trading, Ridge regression can help mitigate multicollinearity and improve the stability of price predictions. (10)

Linear Regression

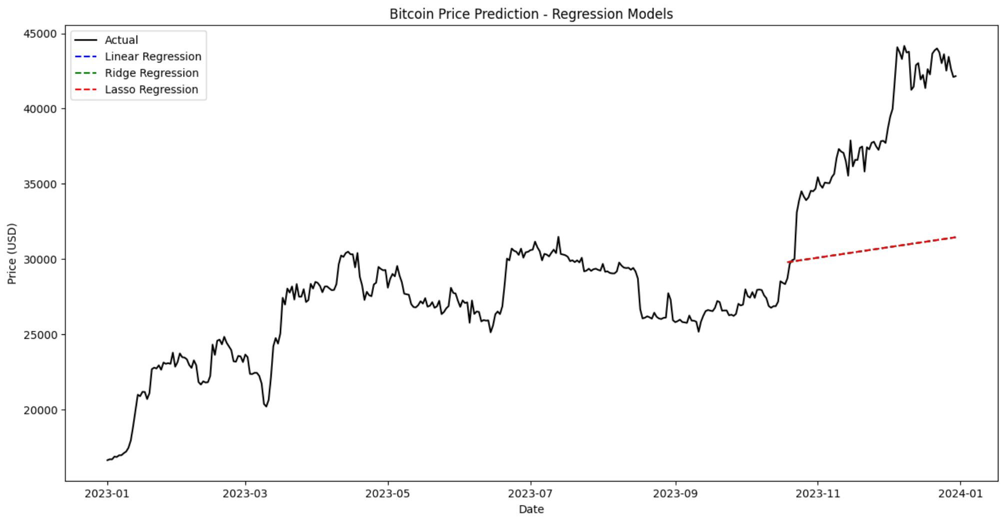

Figure 13, forecasting BTC price using different regression methods

Linear regression is a simple yet powerful ML model used for predicting numerical outcomes based on linear relationships between input features and target variables. In cryptocurrency trading, linear regression can be applied to identify linear trends and patterns in price data, providing insights into potential price movements. (11)

It would be necessary to mention that for Signal of buy and sell, different regressions algorithms such as Decision Tree Regressor, Random Forest Regressor, Gradient Boosting Regressor, KNN Regressor and MLP Regressor, but because it's not the main part, I haven't mentioned it.

These ML models offer valuable insights into cryptocurrency price dynamics and can assist traders in making informed investment decisions. By leveraging advanced algorithms and techniques, traders can gain a competitive edge in the volatile cryptocurrency market and maximize profitability.

Trading Signals and Analysis

General Trading Signals

In this section, we generate buy and sell signals based on the forecasts obtained from the Prophet model trained on historical Bitcoin price data.

Forecasting Time Horizons

Next Week: Short-term forecast for the upcoming week to capture immediate market trends and price movements.

Next Two Weeks: Medium-term forecast for the subsequent two weeks to identify medium-range market trends and potential trading opportunities.

Next Month: Long-term forecast for the following month to anticipate longer-term market trends and positional trading strategies.

Signal Generation

Buy Signal: A buy signal is generated when the forecasted price indicates a potential increase in value, signaling a favorable buying opportunity. This signal suggests initiating or increasing positions in the respective cryptocurrency.

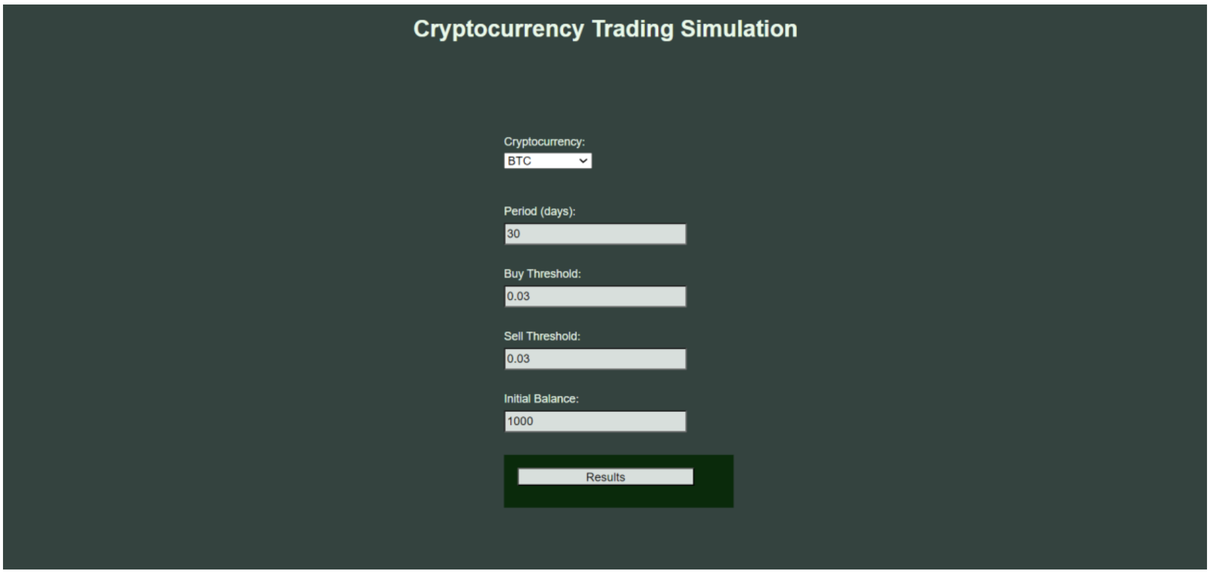

Figure 14, cryptocurrencies buy and sell signal page

Sell Signal: A sell signal is generated when the forecasted price suggests a potential decrease in value, signaling a favorable selling opportunity. This signal indicates reducing or exiting positions in the respective cryptocurrency to lock in profits or minimize losses.

Trading Results

Date	Action	Price	Balance
2023-01-01 00:00:00	Buy	18615.114215870864	1000.0
2023-01-02 00:00:00	Sell	18670.006780852622	1054.8925649817575
2023-01-03 00:00:00	Buy	18902.732367175133	1054.8925649817575
2023-01-04 00:00:00	Sell	19071.11534872696	1223.275546533583
2023-01-05 00:00:00	Buy	18963.737722002177	1223.275546533583
2023-01-06 00:00:00	Sell	19108.78517757809	1368.3230021094969
2023-01-07 00:00:00	Buy	19162.630914248555	1368.3230021094969
2023-01-08 00:00:00	Sell	19304.04469942873	1509.736787289672
2023-01-09 00:00:00	Buy	19358.937264410462	1509.736787289672
2023-01-10 00:00:00	Sell	19591.662850732504	1742.4623736117137
2023-01-11 00:00:00	Buy	19760.04583228528	1742.4623736117137
2023-01-12 00:00:00	Sell	19652.668205559585	1635.0847468860193
2023-01-13 00:00:00	Buy	19797.71566113581	1635.0847468860193
2023-01-14 00:00:00	Sell	19851.56139764196	1688.930483392167
2023-01-15 00:00:00	Buy	19992.975182657923	1688.930483392167
2023-01-16 00:00:00	Sell	20047.86774747502	1743.8230482092622

Figure 15, system suggestion for buy and sell

Additional Features

Trend Analysis: In addition to buy and sell signals, trend analysis is conducted to identify the overall direction of price movements. This analysis provides insights into whether the market is trending upwards (bullish), downwards (bearish), or sideways (neutral).

Financial News Embedding: Integration of financial news and market sentiment analysis to supplement technical forecasts and enhance trading decision-making. News sentiment analysis helps gauge market sentiment and its potential impact on cryptocurrency prices.

Indicators: Utilization of technical indicators, such as moving averages, Relative Strength Index (RSI), and Bollinger Bands, to complement price forecasts and validate trading signals. These indicators provide additional confirmation of market trends and potential reversal points.

If the forecasted price exceeds the upper threshold and no buy order is active, we initiate a buy order. If the forecasted price falls below the lower threshold and a buy order is active, we execute a sell order, realizing the profit. If the forecasted price falls below the stop-loss threshold and a buy order is active, we trigger a stop-loss to limit potential losses.

Results and Analysis

I simulate the trading process and display the buy and sell signals, along with the corresponding profit/loss incurred. Additionally, we compare the initial balance with the final balance to assess the overall profitability of the trading strategy.

Conclusion

The trading signals and analysis provide valuable insights into potential market opportunities and guide decision-making for cryptocurrency traders. By leveraging forecasting models and implementing risk management techniques, traders can optimize their trading strategies and navigate the dynamic cryptocurrency markets more effectively.

User Interface

The user interface (UI) is a critical aspect of the project, ensuring a seamless and visually appealing experience. Three distinct routes cater to different functionalities, maintaining a cohesive design across all interactions.

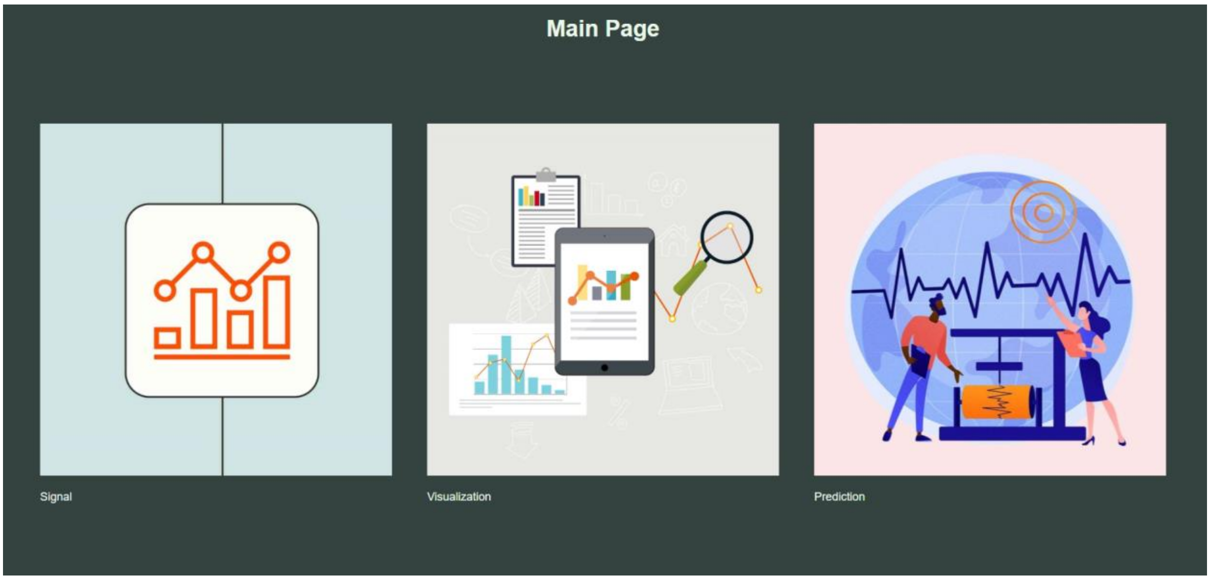

Figure 16, the main page using flask

In the development of this Cryptocurrency Analysis, the entire user interface (UI) is constructed using the Flask library in Python, within the app.py file. The file integrates HTML, CSS, and JS components to enhance overall performance and deliver an optimized user experience. Also in UI all plots done by plotly.

Flask Application (app.py)

The Flask application serves as the backbone of the chatbot, providing a scalable and dynamic platform. The following routes cater to distinct functionalities:

HTML

The project consists of several HTML files all inside the Templates folder. For each block there is a different html file which runs by selecting the block and all couple with the same CSS file.

CSS

The CSS code in style.css styles the project's UI, setting fonts, colors, and layouts. Body has margins and a light background, the header has a distinct dark background with white text. Styles for forms, buttons, and input fields provide a cohesive design. The question history section has a clean list layout with a subtle hover effect for interactivity.

Also I used Bootstrap for some parts of the UI to have the best experience beside the easy use of Bootstrap.

Java Script

The JavaScript code within the main.js file of the project utilizes the part to initiate smooth scrolling functionality for questions and responses. This implementation enhances user experience by providing a visually pleasing transition between sections on the webpage.

Evaluation:

The performance of the Cryptocurrency Analysis was evaluated using a variety of metrics to ensure the accuracy and effectiveness of the information retrieval system. The evaluation process involved assessing each machine learning algorithm individually and comparing their overall performance.

Evaluation Parameters

After training the model, it is essential to evaluate its performance using appropriate metrics. The following evaluation parameters are calculated for assessing the model's accuracy and predictive capability.

Mean Squared Error (MSE):

Mean Squared Error quantifies the average squared difference between the actual and predicted values. It measures the model's precision in estimating the target variable and indicates the magnitude of prediction errors.

Results

Table 1, the evaluation of different methods and cryptocurrencies

		BTC	MKR	ENG	YFI
Linear Regression	Train	6594993.44	24090.56	0.00261	1029075
	Test	72720029.11	11447.51	0.017	11652245
Ridge Regression	Train	6594993.44	24090.56	0.00262	1029075
	Test	72720062.23	11447.53	0.0171	11652240
Lasso Regression	Train	6594993.44	24090.56	0.00262	1029075
	Test	72720071.01	11447.79	0.01651	11652230
ARIMA	Train	3039668.28	39604.69	0.00121	341393.5
	Test	116462391.3	6757.31	0.0112	11687721
LSTM	Train	717001217.4	666638.6	0.000423	48461384
	Test	1545192476	1652032	0.000438	69807191

Conclusion

The analysis of cryptocurrency prices and the development of predictive models present a multifaceted approach to understanding and forecasting market trends. Through the evaluation of various machine learning algorithms, including Linear Regression, Ridge Regression, Lasso Regression, ARIMA, and LSTM, insights into the performance and predictive capabilities of each model have been gleaned. These models, assessed through metrics such as Mean Squared Error (MSE) and R-squared, provide valuable tools for investors and traders seeking to navigate the volatile cryptocurrency landscape. Moreover, the methodology employed, which encompasses data collection, dimensionality reduction, correlation analysis, and model implementation, underscores the importance of a holistic approach to predictive modeling in cryptocurrency markets.

Furthermore, the development of user-friendly graphical interfaces, such as the GUI built with Flask and a Block approach, enhances accessibility and usability, democratizing access to predictive tools and empowering a broader audience of cryptocurrency enthusiasts. The generation of buy and sell signals, informed by the analysis of historical data and predictive models, offers actionable insights into market dynamics and potential trading opportunities. As the cryptocurrency ecosystem continues to evolve, the integration of diverse methodologies and technological advancements will be pivotal in driving informed decision-making and fostering innovation in the realm of cryptocurrency trading and investment.

References

1. Sutton, R., & Barto, A. (2018). Reinforcement Learning: An Introduction.
2. Wang, J., Wang, R., & Zhang, Y. (2019). Cryptocurrency Price Prediction Using Deep Learning.
3. Liu, Y., Zhang, J., & Leng, S. (2020). Time Series Forecasting of Cryptocurrency Prices Using Attention Mechanism.
4. Gupta, A., Mitra, S., & Jain, V. (2021). Forecasting Cryptocurrency Prices Using Transformer-Based Models.
5. Zhang, L., Zhou, T., & Xu, Y. (2022). Predicting Cryptocurrency Market Trends Using Graph Neural Networks.
6. Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: forecasting and control. John Wiley & Sons.
7. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.
8. Taylor, S. J., & Letham, B. (2018). Forecasting at scale. The American Statistician, 72(1), 37-45.
9. Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1), 267-288.
10. Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1), 55-67.
11. Draper, N. R., & Smith, H. (1998). Applied regression analysis (3rd ed.). Wiley-Interscience.

Code, Frameworks and Libraries:

1. Python Software Foundation. (n.d.). Python Language Reference, version 3.x. Retrieved from https://www.python.org/
2. NumPy. (n.d.). Retrieved from https://numpy.org/
3. Pandas. (n.d.). Retrieved from https://pandas.pydata.org/
4. Matplotlib. (n.d.). Retrieved from https://matplotlib.org/
5. yfinance. (n.d.). Retrieved from https://pypi.org/project/yfinance/
6. scikit-learn. (n.d.). Retrieved from https://scikit-learn.org/stable/
7. Keras. (n.d.). Retrieved from https://keras.io/
8. Flask. (n.d.). Retrieved from https://flask.palletsprojects.com/
9. plotly. (n.d.). Retrieved from https://plotly.com/

Appendix A

How to run Cryptocurrency Analysis app

There is two way to run the chatbot

1- Clone the repository using git clone

https://github.com/amirhoseinmohammadisabet/chatbot-HIV.git

2- Use pycharm, vscode or a Terminal to run app.py with python
3- Go to following links to see the http://127.0.0.1:5000/
4- Use main route.