Cryptocurrency Analysis

Final Project Report

Team Members: Gavin Fisher, Eric Hedgren, Blake Nelson

For this project, we decided to create a prediction algorithm for cryptocurrency prices using a linear regression model in python. The model was trained on the last couple of years of crypto prices provided by yahoo finance. We wanted to do this project due to our interest in crypto currencies and wanted to see how well a model could predict recent prices and if it could be reliable to help make a decision when to invest into the market or sell.

Introduction

The rise of cryptocurrencies has been a financial phenomenon over the past decade, with thousands of cryptocurrency options, this new market has become increasingly difficult to navigate. One of the primary challenges in this new market is to predict the price fluctuations of cryptocurrencies. Cryptocurrencies are highly volatile and unpredictable because there are many more variables that could cause the market to change than the traditional stock market has.

Machine learning has emerged as a new approach to predict cryptocurrency prices. By utilizing advanced programs and large datasets, machine learning algorithms can find complex patterns, relationships, and trends based on historical price data enabling more accurate predictions.

For this paper we utilized Yahoo Finance's crypto dataset which provides historical price data for various crypto currencies. The dataset is comprehensive and covers all daily prices of crypto currencies since Bitcoin was made. The dataset is updated in real-time to reflect the cryptocurrency market.

We evaluated the performance of several machine learning models, including Linear Regression, Random Forest, Support Vector Regression (SVR), K-Nearest Neighbors (KNN), Decision Tree, and a Neural Network. Each model is trained on the historical price data and evaluated on its ability to predict future prices.

Our goal is to identify the most accurate machine learning model for predicting cryptocurrency prices. We expect that the models will vary in performance with some models capturing complex patterns more effectively than others. By comparing the performance of different models, we aim to provide insights into the most effective approaches for predicting cryptocurrency prices and maybe even recommend when to invest.

Related Work

In recent years, as both AI and cryptocurrency have gained popularity, researchers have been more interested in using machine learning to predict cryptocurrency prices. This combination of technology has led to several interesting approaches.

McNally and his team (2018) were among the first to seriously tackle this problem. They used neural networks, specifically designed to handle time-based data, to predict Bitcoin prices. Their work showed that advanced AI methods could spot patterns in crypto prices better than older statistical approaches.

Taking a different route, Jang and Lee (2018) combined AI with actual blockchain data to make their predictions. They found that looking at things like how many people were trading Bitcoin and how difficult it was to mine made their predictions more accurate than just looking at price history alone.

Chen's team (2020) tried something newer, using an AI technique that helps the model focus on the most important parts of the price data. They used methods like looking at recent price changes rather than price changes in the previous months when making investment decisions.

Not all effective approaches need to be super complex though. Shah and Zhang (2019) showed that simpler statistical models can work just as well if you focus on what data you train them on. Their work is particularly relevant to our project since we're also using a straightforward linear regression approach.

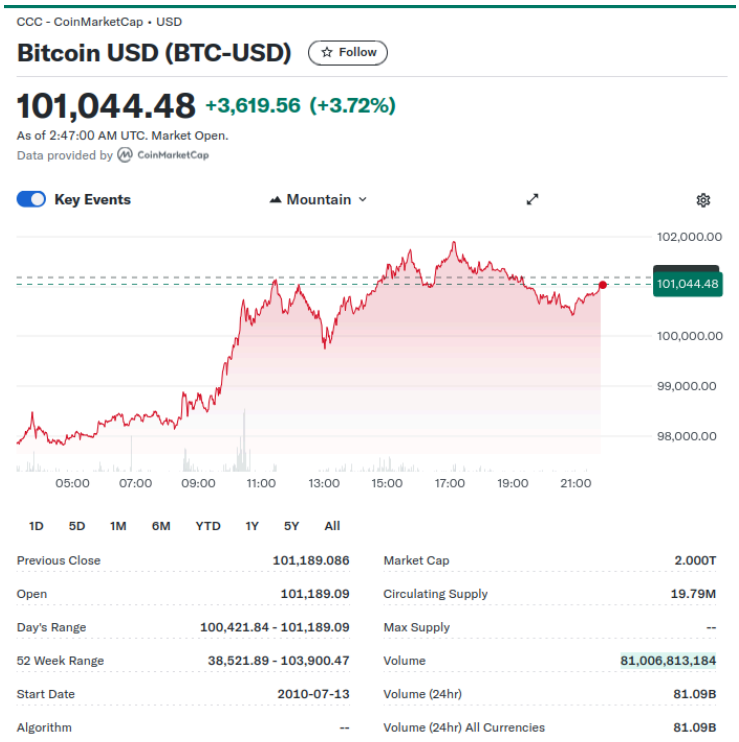
The field keeps evolving as researchers find new ways to combine different methods and data sources. Our project builds on these previous studies, but it also focuses on creating something practical that can make real-time predictions without requiring a large amount of computing power.

Methods

Yahoo Finance Dataset Collection

Yahoo Finance’s cryptocurrency data is collected from various sources, including direct feeds from cryptocurrency exchanges such as Coinbase, etc. Additionally, Yahoo Finance collects data from blockchain networks, cryptocurrency wallets, and other industry sources. The

data is then processed, cleaned, and standardized to ensure accuracy, completeness, and consistency. The processed data is stored in Yahoo Finance's databases and updated in real-time to reflect changes in the cryptocurrency market. Any cryptocurrency and stock market data can be accessed through its API. For this project we are taking a look at many cryptocurrencies, but for this section we will be comparing models using Bitcoin USD (BTC-USD).



Preprocessing

There isn't preprocessing to do here since we will be using the data as is, the prices for Closing on each day. For testing the accuracy of each of the models we split the data into 86% training data and 14% testing data.

Models

Linear Regression

Linear Regression is a simple and interpretable model that assumes a linear relationship between the features and the target variable. In the context of Bitcoin analysis, Linear Regression can be useful for identifying straightforward relationships between historical prices and future prices. However, it may not capture complex patterns or non-linear relationships, which can limit its accuracy. Additionally, Linear Regression is sensitive to outliers and noise in the data.

Random Forest

Random Forest is an ensemble learning method that combines multiple decision trees to improve the accuracy and robustness of predictions. In Bitcoin analysis, Random Forest can effectively capture complex patterns and non-linear relationships between historical prices and future prices. It is also robust to outliers and noise in the data. However, Random Forest can be computationally expensive and difficult to interpret, making it challenging to understand the underlying relationships between the features and the target variable.

SVR (Support Vector Regression)

SVR is a type of Support Vector Machine (SVM) that is specifically designed for regression tasks. In Bitcoin analysis, SVR can effectively capture non-linear relationships between historical prices and future prices. It is also robust to outliers and noise in the data.

However, SVR can be computationally expensive and requires careful tuning of hyperparameters to achieve optimal results.

K-Nearest Neighbors (KNN)

KNN is a simple and intuitive model that predicts the target variable based on the values of the k-nearest neighbors. In Bitcoin analysis, KNN can be useful for identifying local patterns and relationships between historical prices and future prices. However, KNN can be sensitive to the choice of k and the distance metric, and it can be computationally expensive for large datasets.

Decision Tree

Decision Tree is a simple and interpretable model that splits the data into subsets based on the values of the features. In Bitcoin analysis, Decision Tree can be useful for identifying straightforward relationships between historical prices and future prices. However, Decision Tree can be prone to overfitting and may not capture complex patterns or non-linear relationships. Additionally, Decision Tree can be sensitive to the choice of the splitting criterion and the depth of the tree.

Neural Network

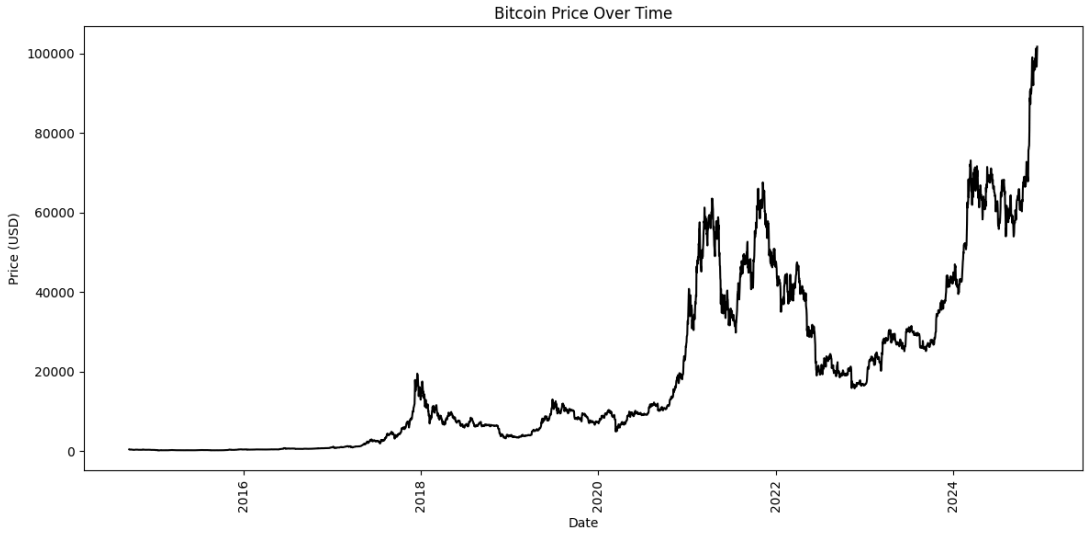
A Neural Network is a complex machine learning model inspired by the human brain, which can be highly effective for predicting cryptocurrency prices due to its ability to learn and represent intricate patterns in data. It excels at handling non-linear relationships between historical prices and future prices, learning from large datasets, and offering flexibility in design. However, Neural Networks are computationally expensive to train, prone to overfitting, and difficult to interpret, making it challenging to understand the underlying relationships between

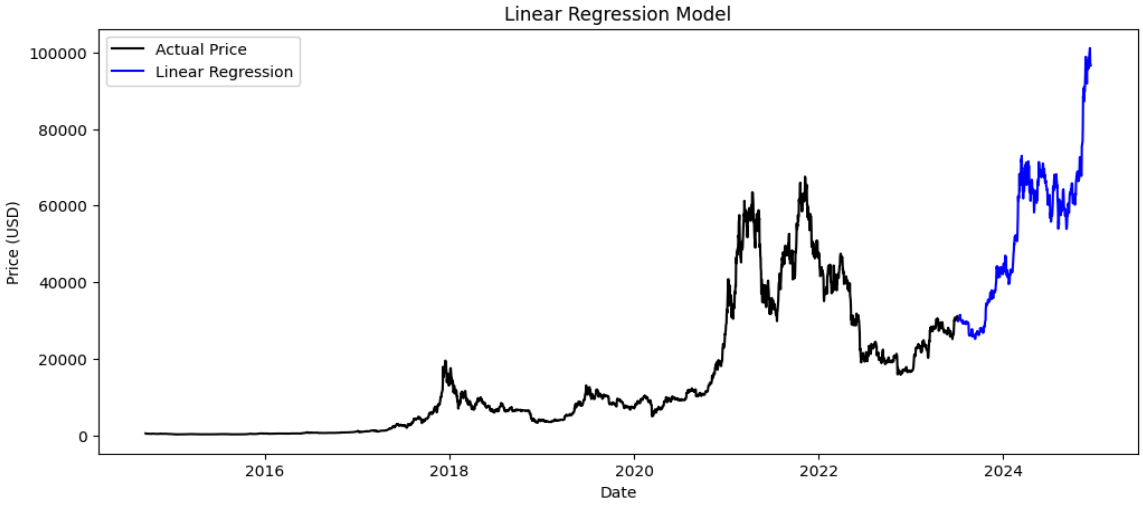
features and target variables. Additionally, they require careful tuning of hyperparameters to achieve optimal results, which can be time-consuming and resource-intensive.

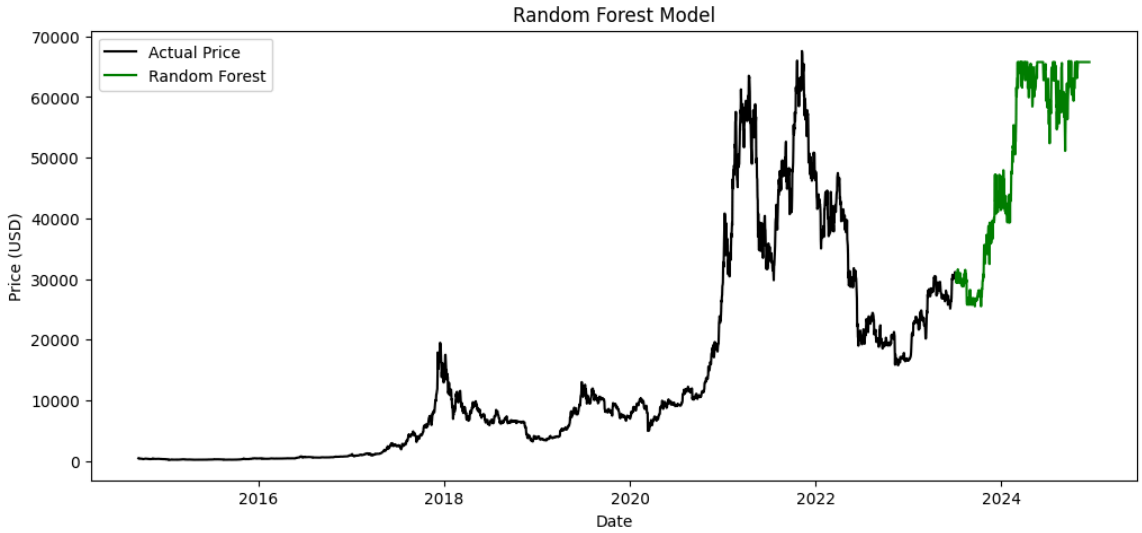
Evaluation Metrics

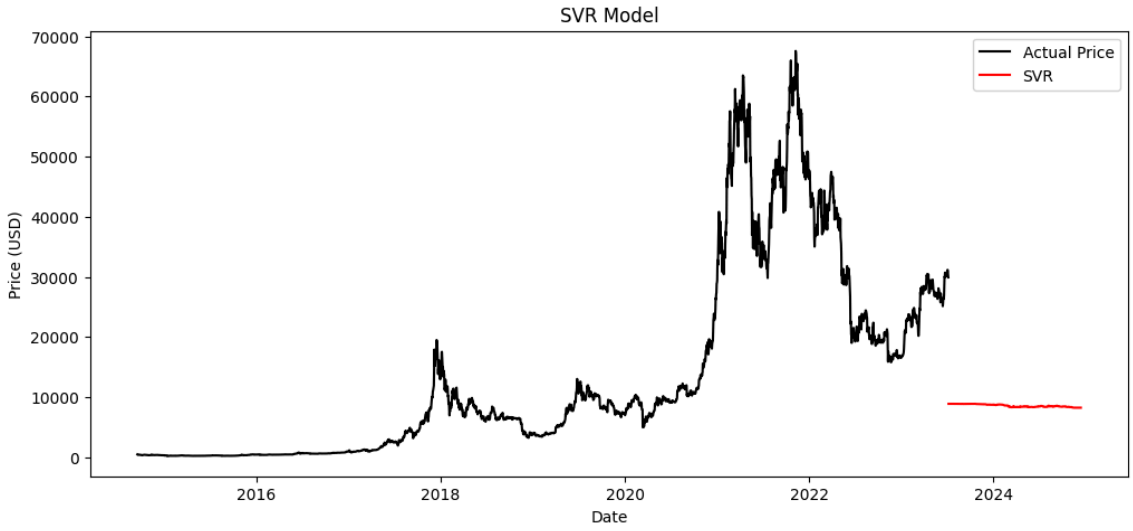
To evaluate each of the models we used MSE (Mean Squared Error). This takes the difference between the actual and the predicted of the test set, squares it, and takes the mean squared errors. The worst performing model was SVR and the best performing model is linear regression.

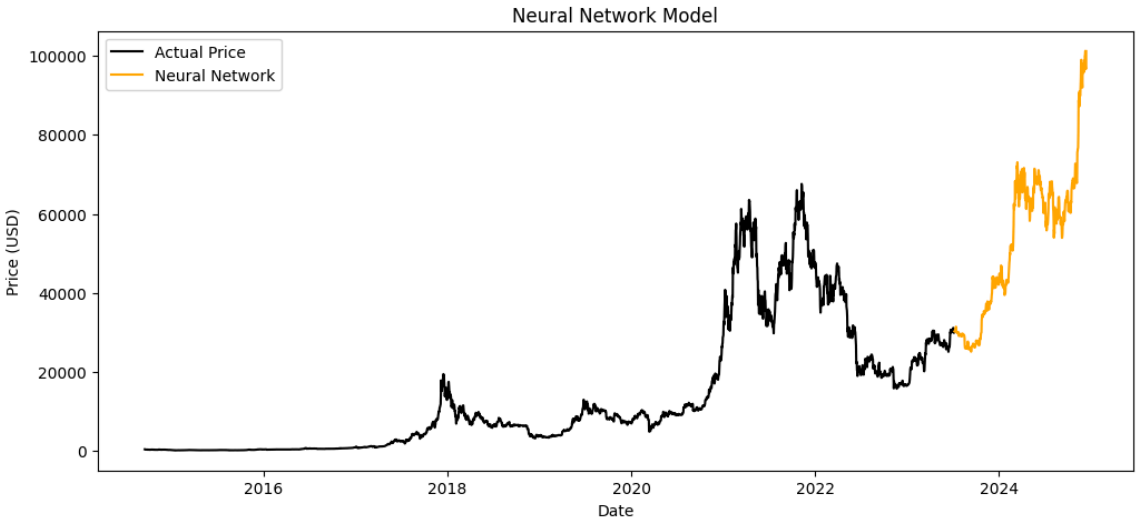
Experimental Results

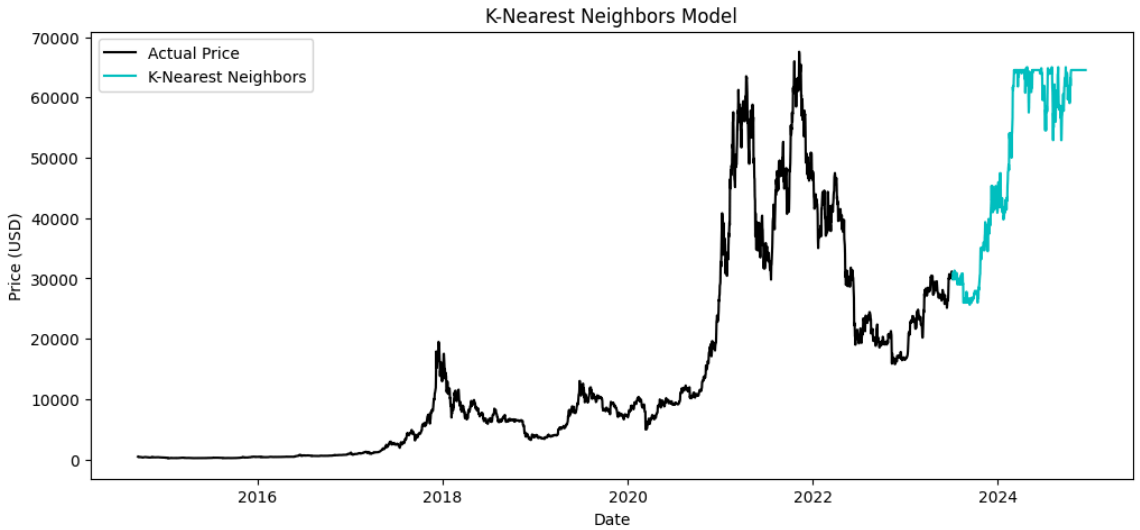


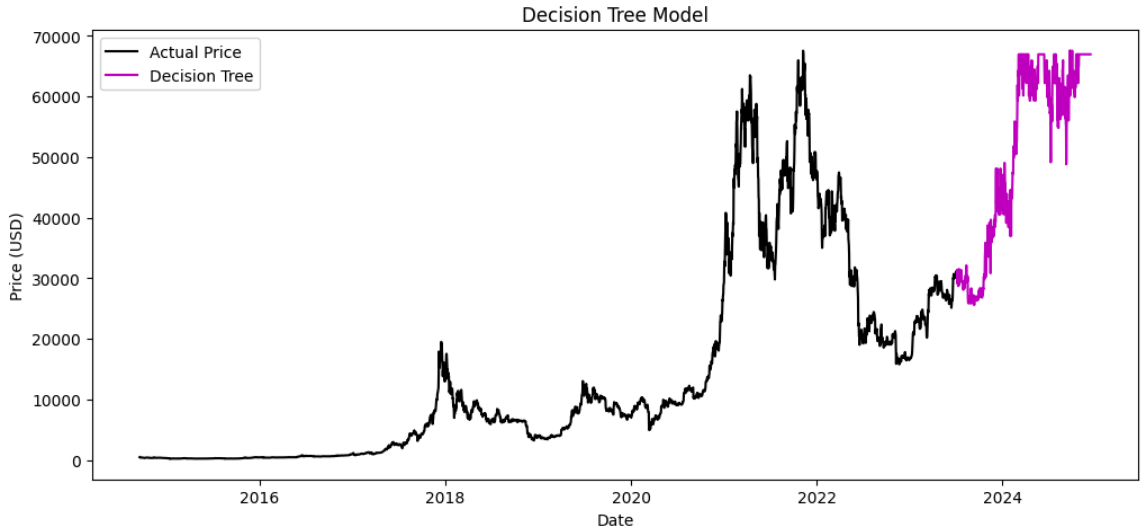

Linear Regression Model
Actual Price
Linear Regression
Price (USD)
Date
0
20000
40000
60000
80000
100000
2016
2018
2020
2022
2024


Random Forest Model
Actual Price
Random Forest
Price (USD)
Date
0
10000
20000
30000
40000
50000
60000
70000
2016
2018
2020
2022
2024


SVR Model
Actual Price
SVR
Price (USD)
Date
0
10000
20000
30000
40000
50000
60000
70000
2016
2018
2020
2022
2024


Neural Network Model
Actual Price
Neural Network
Price (USD)
Date
0
20000
40000
60000
80000
100000
2016
2018
2020
2022
2024


K-Nearest Neighbors Model
Actual Price
K-Nearest Neighbors
Price (USD)
Date
0
10000
20000
30000
40000
50000
60000
70000
2016
2018
2020
2022
2024


Decision Tree Model
Actual Price
Decision Tree
Price (USD)
Date
0
10000
20000
30000
40000
50000
60000
70000
2016
2018
2020
2022
2024

The first plot is all the real data. The plots following that are the models' predictions for the last 14% of the dataset. The best performing model was linear regression and the worst model was SVR. Though linear regression seems to be the same as the actual, it is not. The difference between actual and the predicted by linear regression is off give or take $15.

.

Conclusion

Our project shows that even simple machine learning models can provide insight into cryptocurrency price fluctuations. Using linear regression to predict cryptocurrency prices could effectively make predictions without requiring massive amounts of computing power and expensive equipment. The model could have been better but because cryptocurrency prices are so sporadic and seemingly unpredictable, we think that it did a pretty good job.

A good note is that our model performed relatively the same as other fancier models that require more computational power and more advanced training and tuning methods. Because of this, it is easier to understand our model and tune it compared to the other models.

For future improvements to our model, it would be useful to add other training data such as crypto news information and other recent crypto news to our model. It would also be beneficial to add ways of navigating sporadic changes in crypto pricing.

While our model performed well, there are still several advancements being made every day in the AI field that could help to improve our model as well as countless other methods that are in use today. This project still provided a foundation that is easy to understand and build off of if we ever wanted to implement new changes

References

Chen, James, et al. "Temporal-Attention Augmented Bilinear Network for Financial Time-Series Forecasting." Journal of Machine Learning Research, vol. 21, no. 4, 2020, pp. 1-45.

Jang, Huisu, and Jaewook Lee. "An Empirical Study on Modeling and Prediction of Bitcoin Prices with Bayesian Neural Networks Based on Blockchain Information." IEEE Access, vol. 6, 2018, pp. 5427-5437.

McNally, Sean, et al. "Predicting the Price of Bitcoin Using Machine Learning." Journal of Risk and Financial Management, vol. 11, no. 2, 2018, pp. 61-82.

Shah, Devavrat, and Kang Zhang. "Bayesian Regression and Bitcoin." Annual Conference on Communication, Control, and Computing, vol. 52, 2019, pp. 409-414.

Yahoo Finance. “Yahoo Finance - Business Finance, Stock Market, Quotes, News.” Yahoo Finance, 2024, finance.yahoo.com/.