COSC 3337 - Project Report

Hasan Qureshi

Daniel Bonilla

Svapnarajsinh Bhadola

Matthew Klayme

BTC Price Prediction Analysis

Introduction:

The purpose of this project is to see if we can create a model that is able to predict the closing price of bitcoin using the trading history attributes from the previous years. The bitcoin dataset used for training and testing covers the trading history from Sept 2014 to Oct 2021. In detail, the set contains the data of the date, opening price, high price, low price, close price, adjusted closing price, and volume of bitcoin within the time frame. Each entry in the dataset covers the prices of bitcoin for that trading day.

During the construction and testing of the models, we were able to produce interesting results on the trends of bitcoin and test the capabilities of the models. From experimenting, the regression models seemed to fit the data too well and encouraged us to try a forecasting model to compare results. The forecast models also fit the data well and was more interesting overall. The forecast models also provided more information on the trends and seasonality of the prices over the past years.

Though the results look promising, this project is for informational purposes only, you should not construe any such information or other material as legal, tax, investment, financial, or other advice. Nothing contained on our report constitutes a solicitation, recommendation, endorsement, or offer for any cryptocurrencies or anything of the nature.

Method:

Two regression models and two forecast models were used to predict the prices of bitcoin. The two regression models are a linear regression model and a neural network regression model. For the neural network regression model, there were three hidden layers with each layer using 64 neurons. This model used RELU activation with an ADAM optimizer.

The two forecast models were a Facebook prophet model and a long short term memory model (LSTM). The Facebook prophet model is a model that can do well with nonlinear data and is able to provide detailed plots on seasonality and trends. The prophet model only uses one attribute and dates to identify trends to aid in predictions.

The LSTM model is another neural network model but instead of making a regression line it can make price predictions using a set number of previous days. The number of previous days to analyze was set to 100 for this model. The LSTM model is a stacked model with multiple layers of LSTM used with 50 neurons each. ADAM optimizer was also used with a mean squared error as loss. For the parameters used, epochs are set to 100 and batch size is set to 64.

For training and testing, the data is split into the following, the first 80% is all values up to 5/15/2020 with the remaining 20% continuing from this point to 10/16/2021. The data did contain some null values but were resolved using the ffill() function. For some of the models the date was converted to ordinal to use this attribute for predictions.

The main libraries used include pandas and numpy for data handling, sklearn with keras and tensorflow for neural network/regression models, and the fbprophet library for the prophet model. Other libraries include seaborn and datetime for more data handling and representation of the data.

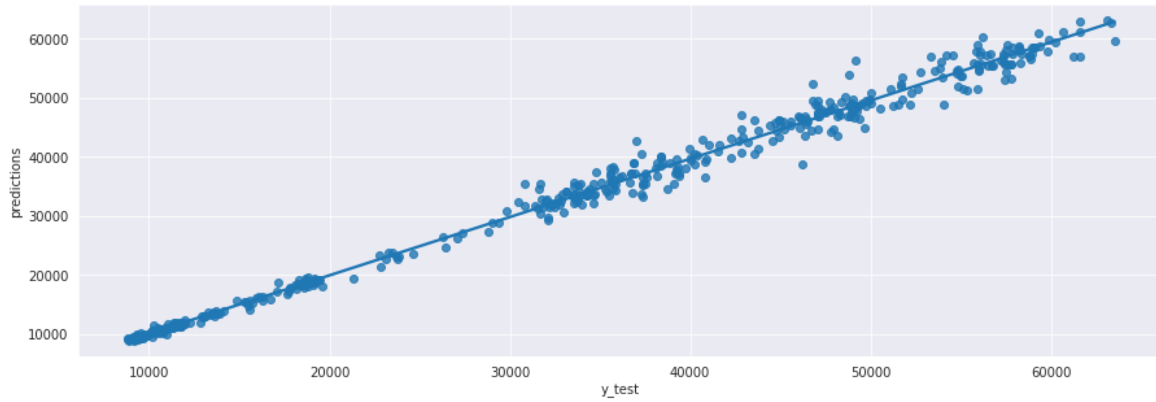
Results and Discussion:

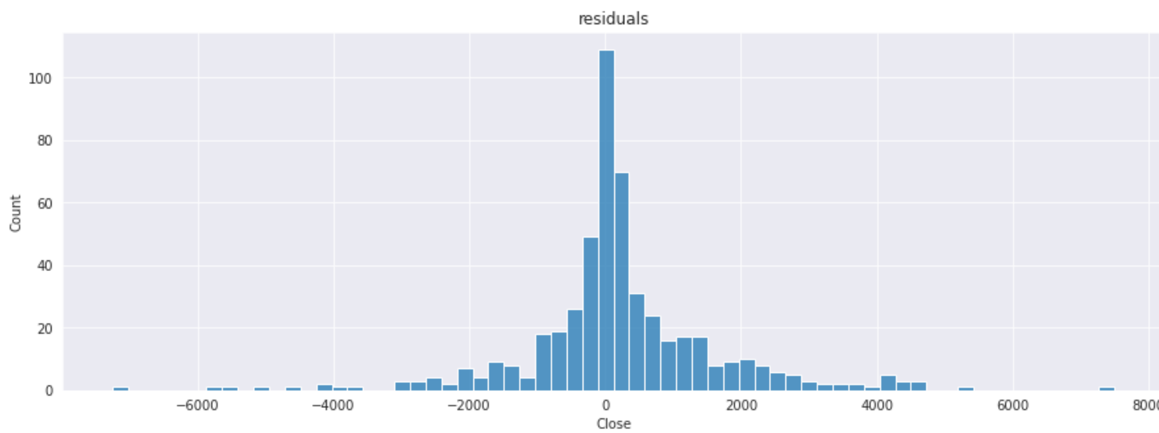
Before the models were used some data analysis was done just to see the overall graph of the closing price for bitcoin.

Since the first 80% of the data is used for training, we were interested to see how the models would predict the values past 2020. With only the training data to work with, these models would only have 1 experience with a massive spike in price back in 2018. The spike in early 2021 will easily be 3 times as large as the spike in 2018. We were expecting these models to do somewhat poorly in terms of predicting this massive spike when testing.

Linear Regression:

The first model used is linear regression. The data was trained with the open prices and date up to 80% to predict the closing price. The line of best fit as well as the residuals were produced with this model.

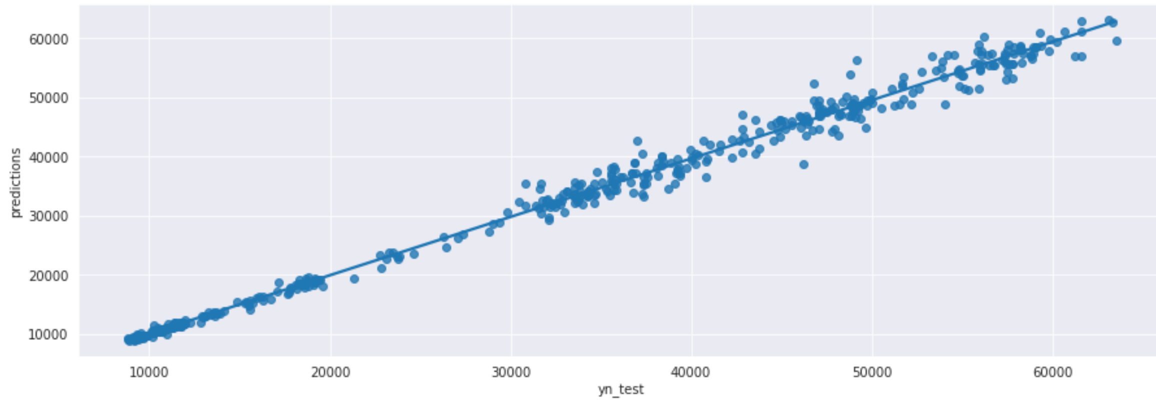


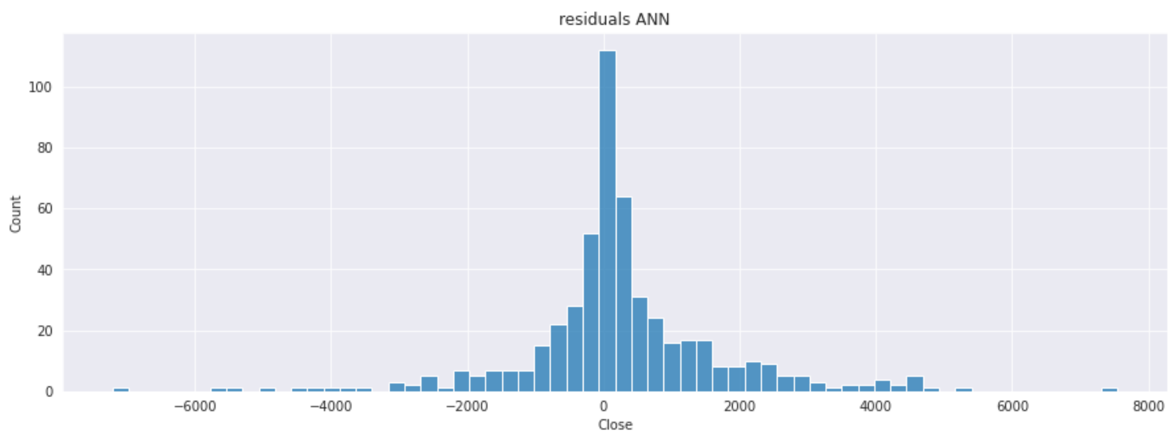


The R^2 value was 0.99287 with MAE, MSE, RMSE being 933.79, 2224363.11, and 1491.42 respectively. Looking at the line, the graph did well in the beginning then started to lose accuracy as the days to predict got larger. The model did well against the massive spike in 2021. Overall, the extremely high R^2 values tell us that the data used might be too much to work with and may have caused some over fitting problems. To be sure that it's not just a linear model having problems we created a neural network regression model to compare. Using this model also assumes that the dataset is in linear fashion. Though, there spikes and dips in the dataset trend, our R^2 value somewhat confirms the linear growth in the dataset. It's obviously not linear, but this would be a naïve implementation of this model.

Deep Neural Multilayer Perceptron (Neural Network Regression)(MLPRegressor):

The set up for this model was mostly the same as the last and used date and open price to predict the closing price. The difference is the data was scaled with standard scaler to fit the model for a better reading. The results below were produced.

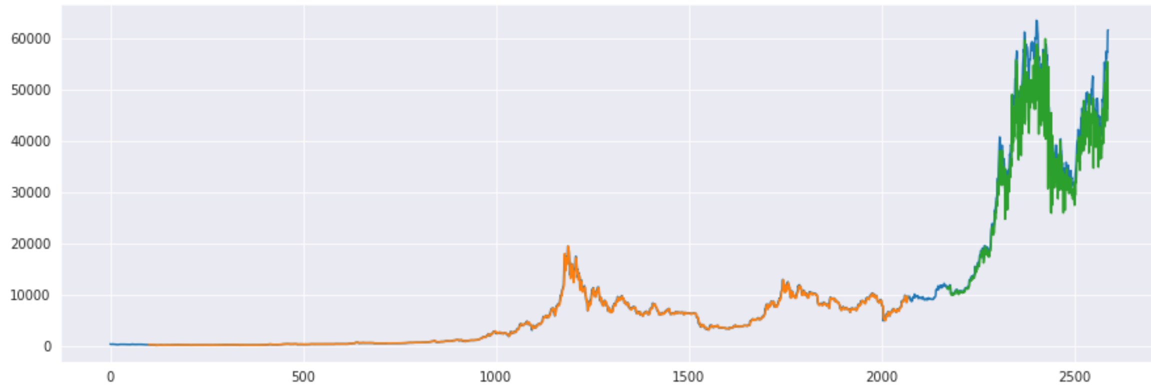




The R^2 value was 0.99266 with MAE, MSE, RMSE being 941.20, 2247874.36, and 1499.29 respectively. The results are almost the same as the linear regression model therefore have the same analysis. Since the results are almost the same, this model also has the same overfitting concerns. At this point we decided that due to the data, regression models might not be the best for this project and decided to try forecast models.

Long Short Term Memory (LSTM):

To combat the problems with the regression models, we used a stacked LSTM forecast model that takes 100 previous days from the data to predict the next day. Since the model is a univariate model, it only takes one attribute and uses it to predict the next value. Due to this univariate specification, the model was fed only closing prices of bitcoin to predict future closing prices. Using the 80% training and 20% testing split the model was able to come up with the following graph.



Orange represents the training values, blue represents original values, and green represents the predicted values. The numbers on the bottom represent the number of days passed from the original starting point (9/17/2014). The MSE for the training is 5754.90 while the MSE for the testing is 35220.77. Similarly, to the regression models, the predictions start off with a higher accuracy and loses accuracy as more time passes. The mean squared error being high is understandable as the predictions in the graph become less accurate as it reaches the last date on the data. The model did generally keep up with the original values as it faced the massive spike following 2020. The predictions also looked as if it mostly undershot its prediction than overshot.

Facebook Prophet:

To compare with the LSTM model, another forecast model was needed, and we decided to use Facebook's prophet model. This model has different restrictions than the LSTM model and is instead fed the dates and attribute to predict. Dates is a given and the attribute used for prediction is closing price of bitcoin. The model was able to produce a prediction graph as well as some very useful information about the trends, seasonality, and events about the data. The graphs below is the prediction only using 80% of the data.

The black points represent the actual values, the blue line represents what the model is predicting, and the shades represent the variance in which the actual value could reside in. Looking at the graph, the model does start off more accurate when the days to predict is closer to the actual data and it does lose its accuracy later. The variance does show that the data could either rise or drop a tremendous amount and has the predicted values showing a slight increase in value. Compared to the original data, this model missed the massive spike in closing prices but did generally catch that there could be an increase with the shaded variance. The model also showed the trends of the training data with the graphs shown below.

The model recognized the spike in 2018 and identified the points in time bitcoin was raised. These trends were accounted for when it created the prediction graph. After seeing these trends, we ran the model again with the full dataset to get a better sense of the trends in the data without any gaps. The results of the new graphs are shown below.

Looking at the updated trends, the spike in 2021 completely overshadows the trends in 2018 and sheds some light into the new closing patterns. For weekly, Friday still stays on top for bitcoin closing price increasing and shows that Tuesday and Thursday are usually when the price decreases. Monday has price staying neutral while Wednesday looks like it has a slight decrease. In terms of yearly trends, the extra 20% of data completely changed the graph as it went from July and September providing the highest increase in closing price to March and May now reigning on top. Instead of May being the point in time where closing price went up, January is the new time where bitcoin closing price could go up.

Conclusion:

Out of the 4 models used, we found that Facebook prophet was the most useful for our analysis. The model not only provided a prediction graph, but also information on trends and seasonality through easy-to-read graphs. The prediction from the prophet model seemed more realistic to us as it had a harder time picking up the massive increase in bit coin during early 2021. The LSTM model used was also interesting to see as it tracked the 2021 event quite well given the lack of preparation for events. The regression models didn't quite work out as we expected but nonetheless gave us an idea of how a generally increasing data set can affect them. For predicting bitcoin prices, forecasting models showed to be more reliable in shorter bursts of time. This makes sense since predicting a year in the future could cause some problems with accuracy. While these graphs do seem to be generally accurate, they aren't accurate enough to help aid in purchasing bitcoin.

Once again, this project is for informational purposes only, we highly advise against using these models in a way to help purchase bitcoin.

Resources:

- OddAsparagus11. "Bitcoin Price Dataset - from Sept 2014 to Oct 2021." *Kaggle*, 16 Oct. 2021, https://www.kaggle.com/oddasparagus11/bitcoin-price-dataset-from-sept-2014-to-oct-2021.
- Naik, Krish. *Stock Price Prediction And Forecasting Using Stacked LSTM- Deep Learning*, YouTube, 25 May 2020, https://www.youtube.com/watch?v=H6du_pfuznE.
- Choudhury, Kaushik. "Deep Neural Multilayer Perceptron (MLP) with Scikit-Learn." *Medium*, Towards Data Science, 31 Aug. 2020, https://towardsdatascience.com/deep-neural-multilayer-perceptron-mlp-with-scikit-learn-2698e77155e.
- "Legal Disclaimer." *Harrington Investments*, 23 July 2018, https://www.harringtoninvestments.com/legal-disclaimer/.