Capstone 2 Project Proposal
John Arancio
November, 2020

BTC Price Prediction

Being that there have been numerous predictive models produced regarding the general stock market I am proposing that it might be interesting to explore possible predictions within the new "asset class" of cryptocurrency. This project will contain various approaches in order to form an accurate price prediction that combines the use of social media, traditional technical indicators used in financial markets and machine learning regression as in the form of predictive time series analysis.

Problem:
Can BTCs (Bitcoin) price be predicted for a specific timeframe utilizing various indicators and even Twitter tweets sentiment?

Stakeholders:
The stakeholders surrounding this issue within the cryptocurrency markets are any retail or institution trader or investor. It should also be considered that those not specifically involved with cryptocurrency itself but rather within traditional markets might find use of the problem being solved.

Datasets:
The data for BTCs prices will be collected using the Binance Exchanges API and it will contain the date, high, low, open price, close price and volume for each time segment within the time range. The time range will be from 2019-09-02 to 2020-11-15 which accounts for 441 complete trading days since cryptocurrency markets are open 24/7. For the Twitter data, I will be using a basic web scraper that will pull historical tweets from the time range specified. The tweets sentiment will be aggregated using mean polarity for each day.

Approach (Supervised Learning):
In order to derive predicted prices from the original BTC dataset from Binance we will use supervised regression techniques to achieve this. The supervised learning will include a simple **linear regression model** as well as a **random forest model** that will use the previous close price from the dataset as our "prediction target". By identifying the "prediction target" we will be able to graphically demonstrate performance of the models against actual price movement which will show how well the models are able to predict prices. Using MAE (Mean Absolute Error), MSE

(Mean Squared Error), and RMSE (Root Mean Squared Error) as our KPI's (Key Performance Indicator) we will be able to measure the accuracy of the price forecast prediction fairly accurately. Those measurements will allow us to see how well the model performed.

Our features will be the technical indicators (i.e. RSI, ADX, ATR, OBV etc.), the mean sentiment polarity of the day's tweets regarding $BTC, and then the close price, open price, high, low and average volume. Each feature will be integrated to work with only the "Daily" time frame and possibly move to smaller time frames (i.e. 4HR, 1HR, 30MIN etc.) along the way.

Time permitting further research into the area of deep learning shows a positive use case of solving the price prediction problem by utilizing a **LSTM (Long Short Term Memory)** deep neural net architecture. The LSTM architecture can be useful as it only remembers the information or data that is important for the desired outcome while being able to forget what is not important, that which can be effective when handling data for time series forecasting.

<u>Intended Use:</u>
Those who are interested in market outcomes as the "stakeholders" can further utilize my findings and research to make better decisions investing or trading. If my models prove to be useful and accurate in price prediction I can envision the predictive models being automated for real-time use in regards to trading. Analyzing residuals, viewing the features to see how important they are and what their impact is across the board can provide hidden insights an "investor" might not see until utilizing the presented models.

Since cryptocurrency is a relatively new "asset" it might be useful to predict price outcomes for those who are simply observing the various characteristics of BTC itself as a way to gauge the crypto market. It will also be useful to compare and contrast performance across a traditional asset such as $TSLA or even a different cryptocurrency like $ETH. By comparing the models eventually formed using BTCs data to other types of assets we might be able to see if the combination of technical indicators and tweet sentiment is consistent with performance across all markets (as in generalizability).