CryptoTitans – Informed Bitcoin Trading
Olabode Faleye, Kazuomori Lewis, Vicente Izquierdo, Pedro Pablo Correa

Bitcoin is one of the most lucrative investment opportunities of our generation, however its high volatility makes it difficult for even seasoned investors to predict market trends and maximize their return on investment (ROI). To this end, our team aimed to develop strategies to predict market trends to inform even a novice investor when it is optimal to buy, sell, or simply hold Bitcoin. To develop informed trading strategies, we utilized Bayesian Regression to predict prices on the order of minutes, Recurrent Neural Networks for prediction on the order of days, and the more traditional Bollinger Bands for trading strategies that follow long term market trends. By taking these different approaches, we hope to appeal to investors of any given engagement level with informed trading strategies to help maximize their ROI.

Investment in Bitcoin offers an incredibly lucrative opportunity to make financial returns at a rate that has never before been observed. However, at this point in time the value of Bitcoin is more tied to a speculative value than any realized one, and as a result there lies incredible volatility within the market. News stories, technological breakthroughs or setbacks, as well as government announcements about regulation all have the ability to affect the price in greater than 30% swings over the course of a day. To this end, our team chose three different models over three different timescales to make generalized but informed decisions about optimal times to buy, sell, or hold Bitcoin. The three timescales were on the order of seconds to minutes, days, and a longer-term strategy that follows market trends. We chose these three strategies to diversify our options during periods of high volatility. If market swings are in excess of tens of percentages per day, it may be better to sit back and follow generalized market trends to maximize ROI. In the converse, if there is very little movement in price, our shorter timescale models will still provide trading insight to allow for a profit.

As a first approach, we took analytics about the prices of bitcoin from january 2017 - october 2017. It was crucial to plot the behavior of the prices and analysis such as Simple Moving Averages (SMA), volume of trading and Time Series analysis (correlation and ACF/PACF) to understand the nature of the price movements. This led us to first focus on Bollinger Bands, method that works with confidence intervals of SMA and create buying and selling strategies based on the moments that the price escape from the confidence interval. As we developed a deeper understanding of prediction methods and began to learn about Neural Networks and different types of regressions in class, we thought there was an opportunity to combine these methods to make a more robust model. As development on each model it became clear they had distinct solutions that were best kept separate.

Our first approach of Bollinger Bands had the “art” of testing different windows of time to model the curves. The longer the window size, the smoother the curve, but this came with a cost of blindness for short time trends. In the case of very short window size the model couldn’t create an actual trend. We ended up having an optimal window size of fifty days with a profit margin of 1.5 in comparison with holding the cryptocurrency.

In the case of bayesian regression, the main difficulty was that the future predictions were made in the matter of seconds. Luckily, the first days of december of this year companies such as CME Group announced that they will start to trade Bitcoin future options, which give the opportunity for high frequency trading of the token. The model itself had a high accuracy with a MSE of about 0.9.

In the case of our neural network implementation, we found the model to be relatively accurate in predicting trends of the market. The absolute values of price were slightly off, but the general price trends predicted by the model were consistent with that of actual data. Much optimization remains to be done in this area to further improve the accuracy of the model, potentially to the point where absolute prices can be reasonably predicted.

We first began by pulling bitcoin price data from HitBTC, one of the largest Bitcoin trading exchanges, and blockchain transaction data from blockchain.info. Luckily, due to simplicity and thoroughness of the data sources data cleansing was not necessary for these sets of data. A web scraping script using Beautiful Soup was built to automatically pull a year's worth of data from the blockchain.info site. This initially yielded 31 different metrics regarding the blockchain but we narrowed these down to the 15 that we believed we most relevant for the RNN, deleting features such as block size and number of bitcoins in circulation as these features simply increase with time regardless of changes in price.

For the intraweek trading strategy, the recurrent neural network used was a simplified Long Short Term Memory Model based on work by David Sheehan (dashee87.github.io). It relies on a sliding window of memory of a particular length (we found ~3 days to be optimal), and uses both bitcoin exchange prices and blockchain transaction data from this window to predict the price of Bitcoin the following day. The value in this model vs the other two is that it considers the blockchain information when predicting bitcoin price. To the best of our knowledge, this is the first model of it’s kind to integrate this type of information. However, there lies a large degree of variability in the accuracy of our model to predict future price trends based on the length of the sliding window of memory, number of neurons, epochs used to train, as well as the way that the data was normalized. Optimally, we found that using a window of memory of 3 days, 15 neurons, training for 30 epochs, and normalizing all data by the first value in the table led to the best results. However, significant optimization can still be done to improve the accuracy of the model.

The bayesian regression strategy, is based on a simplified version of *Bayesian regression and Bitcoin,* Devavrat Shah, Kang Zhang (https://arxiv.org/abs/1410.1231), implemented by Abhinandan Ramesh (https://github.com/abhinandanramesh/). It uses a dataset containing three different time scales, to predict price variations at short time (20 seconds in this case) in a binary fashion (i.e. Does the price go up or down?). The model has promising applications for investment decisions in the case of higher frequency trading (e.g. Futures trading).

The Bollinger model, on the other side, aims to include a traditional financial approach to cryptocurrency trading, by analyzing the behavior of the price curve compared with a confidence interval of a SMA curve (flexible window of days). Considering the high-volatility context in which cryptocurrencies operate, the Bollinger Bands strategy give a bigger picture of the behavior of the price. It works as a “brake” for the other methods, which are more blind for historical prices.

For future development, we will work in better user interface. We propose a web-based application, using more robust tools to develop a tool over the models described before. . Another step in the development of the trading strategies we propose automated trading execution. For this, we have to include API hooks for the different cryptocurrency exchanges, to execute the orders from the strategies automatically. Finally, to optimize transactions fees we want to include slippage and transaction costs. In the case of slippage it's impossible to measure, so the use of estimation algorithms might be the solution. In the other hand for transaction costs there are different in each exchange, so the challenge might be bigger here.

For the optimal workload for each member we divided the different algorithms and their focus to finally compare the results of each one independently. Olabode and Kazuomori worked in research, development and adaptation of the RNN LSTM model (including data preprocessing, normalization and optimization of the model) and writing of the final report. Pablo worked in research, development and adaptation of the bayesian regression algorithm, backtesting for other models and writing of the final report. Finally, Vicente worked in research, development and adaptation of the traditional financial strategy of Bollinger Bands algorithm and backtesting of it, and writing of the final report.

Regarding the mentoring opportunities of the class, we took feedback from the teaching staff about longer term improvements upon the project, but unfortunately did not have a dedicated mentor. Had we the choice, we would have selected someone with a strong background in finance who also implemented machine learning strategies for trading.