

CMPS 287 Project

Team K Members:

Hiba Lubbad 201902178

Hamza El Sheikh 201904899

Dany Tarabay 201903983

20 April 2021

Phase One

1 Topic

Cryptocurrency Price Prediction using real-time data, official reports, influential figures, and market ambiance.

2 Cryptocurrencies

Due to time limitations and the complex nature of the project, we will narrow our focus to two cryptocurrencies: Bitcoin and Ethereum. These are the two largest cryptocurrencies in terms of market capitalization representing over **1.225 trillion dollars** in combined value.

3 Problem

The cryptocurrency market is an unstable market due to the rapid changes and some affective factors. News is one of the factors that can change the whole market, for example, Wall street statements, Elon Musk tweets, European Central Bank statements, and other public figures. Another important factor is the adoption of these currencies, or in other words the public ambiance towards this technology.

4 Impact/Significance

Achieving a good accuracy for stock market prediction will have a significant impact on the decisions of cryptocurrency investors leading to potential increased profits. Additionally, our model can be used by the government in policymaking to regulate the market and financial researchers towards more informed studies

of the cryptocurrency market behavior. Finally, low and middle-class people are more involved in investing due to the increased adoption of on-demand investing applications like Robinhood, which lowers the barriers for such communities to invest. However, these communities are more vulnerable to market changes, increasing the importance of our project.

5 Resources

- Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis
- Investigating the Problem of Cryptocurrency Price Prediction: A Deep Learning Approach
- Stock Forecasting using Machine Learning
- Short-term stock market price trend prediction using a comprehensive deep learning system
- Stock Price Prediction Using Machine Learning and LSTM-Based Deep Learning Models

Phase Two

6 Determinants that have an impact on the BTC price

According to the following research paper titled “Forecasting the price of Bitcoin using deep learning” by Mingxi Liu et al.[1], the authors have presented research that correlated 40 determinants to the price of bitcoin. For example, they cited the findings of Alstyne and Marshall (2014) that the Dow Jones index, the euro-dollar exchange rate, and the oil price have significant impacts on the price of Bitcoin. We base our collection of datasets based on a chosen 27 determinants that we found most relevant and eliminated factors such as “Exchange rate of US dollar to Norwegian krona”. For each of the determinants below, we collected daily data from 2015 till the present that we will merge to create a finalized dataset for our project. The determinants are listed below.

7 Dataset

Bitcoin info (8): Date - Bitcoin price - Open - Close - High - Low - Trading Volume - Market Cap

Mining info (3): Mining difficulty - Mining award - Average Hashrate

Public attention (4): Google Trends - Twitter tweets - Baidu index - Reddit posts

Macroeconomics Environment (12): Brent Crude futures - CBOE volatility index - NYSE Copper Futures - Dow Jones industrial average - New York Stock Exchange Gold futures - HuShen 300 index - Korea composite index - Singapore Straits Times index - Standard & Poor’s 500 index - FTSE 100 index - Tokyo’s Nikkel index - US dollar index - WTI oil future

We’re going to have a dataset of 27 columns.

8 Twitter Data

Our long-term goal is to scrape through Google trends, Twitter, Top 10 news headlines (Optional), and Baidu. Each row should be iterated over an algorithm (TextBlob, Haven on demand, Tweepy) that generates the Subjectivity, Objectivity, Positivity, Negativity, and neutral ratios. However, for our current plan, we will start by scraping Twitter data.

→ After Getting these ratios we have to add them into the dataset as a new indicator that may impact the price of BTC. Then we have to divide our dataset into training, testing, and validation datasets.

9 Model Selection

According to the Literature referenced below, the models having the highest accuracy and widely used to predict Bitcoin prices are LSTM and ARIMA. After finalizing our dataset, we look forward to verify and choose one of those models for our project.

10 References and Citations

- Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis
- Investigating the Problem of Cryptocurrency Price Prediction: A Deep Learning Approach
- Stock Forecasting using Machine Learning
- Short-term stock market price trend prediction using a comprehensive deep learning system
- Stock Price Prediction Using Machine Learning and LSTM-Based Deep Learning Models
- Mingxi Liu, Finance Research Letters

11 Dataset Links

- STI Index
- GSPC index
- BTC KRW Index
- CSI 300 Index
- VIX Index
- Baidu Historical Price
- F2pool info (Mining info)
- Bitcoin Market Capitalization
- BTC Historical Data

- FTSE 100 Index
- S&P 500 Index
- Brent Crude Oil Prices
- Dow Jones Industrial Index
- Gold Price
- S&P 500 Futures Index
- Tokyo's Nikkei Index
- US Dollar Index
- WTI oil futures

Phase Three

12 LSTM

Traditional neural networks are not capable of having persistent learning while an event is happening. It's not clear how these neural networks could use its reasoning about previous events to inform later ones. Recurrent neural networks (RNN) address this issue. They are networks with loops in them, allowing information to persist. Long Short Term Memory networks (LSTMs) are a special kind of RNN, capable of learning long-term dependencies, the issue of long-term dependency is discussed later on when LSTMs are compared with regular RNNs. LSTMs work by averaging what information is going to be thrown away from previous knowledge, like the gender of a previous state, with new information that is going to be used.

The inputs for the LSTM are: Long term memory (LTM), short term memory (STM), and the input. Whereas the outputs are the updated long and short term memories, as well as the prediction of the event in question. Both updated memories are dependent on the three inputs mentioned above. LSTMs architecture contains the forget, learn, remember, and use gate. LTM goes to the forget gate, where it forgets anything that is not considered useful in that instance. STM and the input are joined in the learn gate, containing information recently learned while removing any unnecessary information. LTM and the information in the learn gate are joined together in the remember gate, updating the LTM. The use gate merges information we previously knew and learned from the input to make a prediction, and the output becomes the STM. This is done recurrently as inputs are added as a function of time.

13 LSTM and other models when dealing with cryptocurrency

13.1 LSTM and Simple Recurrent Neural Networks RNNs

Recurrent Neural Networks are improved neural networks that take current input as well as previously perceived input into account. At any time step of the computation, the current weights along the previous weights of time steps are used to compute the network activation. Moreover, the gradient at each timestamp is calculated and the contributions of each time step to the gradient are accumulated. However, there are a few issues that arise from this model. This process can quickly become computationally expensive as the length of the sequence of data increases. Previous research has shown that as the length of the sequence increases the RNNs become unable to connect the information properly needed for prediction. This was called the long-term dependency problem. Furthermore, as the number of time steps increases the gradient can

either explode or decay completely due to the number of multiplications involved in calculating gradients of very early time steps. Therefore, LSTMs were introduced to solve this problem with more complex architecture. LSTMs are designed to remember information for long periods of time without the struggle RNNs face.

13.2 LSTM and ARIMA

ARIMA is widely regarded as the most efficient forecasting technique in social science and is used extensively for time series. The use of ARIMA for forecasting time series is essential with uncertainty as it does not assume knowledge of any underlying model or relationships as in some other methods. ARIMA essentially relies on past values of the series as well as previous error terms for forecasting. However, ARIMA models are relatively more robust and efficient than more complex structural models in relation to short-run forecasting.

Although ARIMA is considered one of the best models to predict stock market prices, we found some differences between ARIMA and LSTM since we're going to predict Bitcoin price with other determinants. After doing our research we noticed that LSTM works better with larger amounts of data than ARIMA, wherein the dataset we have to keep forecasting old and new data. LSTM can take good care of the old data and wait for new data to be forecasted.

Advantages of LSTM:

1. No pre-requisites (stationarity, no level shifts)
2. Can model non-linear function with neural networks
3. Can handle lots of data efficiently

In conclusion, we're going to use LSTM for our Bitcoin prediction model since it's more precise and can handle our data more efficiently.

14 References and Citations

- Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis
- Investigating the Problem of Cryptocurrency Price Prediction: A Deep Learning Approach
- Stock Forecasting using Machine Learning
- Short-term stock market price trend prediction using a comprehensive deep learning system
- Stock Price Prediction Using Machine Learning and LSTM-Based Deep Learning Models

- Mingxi Liu, Finance Research Letters
- Understanding LSTM Networks
- LSTM Recurrent Neural Network for Cryptocurrency Price Prediction
- Long Short Term Memory Original Paper