SEIS 763 – 03

Machine Learning

Team Project Final

12/5/2021

**Analyzing and Predicting Bitcoin Price Using Machine Learning**

|  |  |  |  |
| --- | --- | --- | --- |
| Anderson, Emily  SEIS-763  Machine Learning  University of St. Thomas  MN, USA  [ande56352@stthomas.edu](mailto:ande56352@stthomas.edu) | Nie, Summer  SEIS-763  Machine Learning  University of St. Thomas  MN, USA  [nie09208@stthomas.edu](mailto:nie09208@stthomas.edu) | Lee, Byunghun  SEIS-763  Machine Learning  University of St. Thomas  MN, USA  [lee08271@stthomas.edu](mailto:lee08271@stthomas.edu) | Johnson, Matthew  SEIS-763  Machine Learning  University of St. Thomas MN, USA  [john2285@stthomas.edu](mailto:john2285@stthomas.edu) |

1. **Project Title.**

Analyzing and Predicting Bitcoin Price Using Machine Learning

Keywords – Bitcoin, Bitcoin Prediction, Blockchain, Crypto Currency, Linear Regression, Logistic Regression, SVM (Support Vector Machine).

1. **Description of data source and web link(s). / Ongoing work and achieved results**

Kaggle bitcoin historical data - https://www.kaggle.com/mczielinski/bitcoin-historical- data

1. **Number of records & Number of attributes with description of each attribute.**

The dataset has 4,857,377 records and 8 variables ranging from January 1, 2012, through October 31, 2021.

The data set includes seven attributes: 'Timestamp', 'Open' (Open Price), 'High' (High Price), 'Low'(Low Price), 'Close' (Close Price), 'Volume\_(BTC)', 'Volume\_(Currency)', 'Weighted\_Price'.

1. **Some general statistics of the dataset.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Data Type | Data Name | Data Min Date | Data Max Date | Average | Maximum | Minimum |
| Integer | Timestamp | 01/01/2012 | 10/31/2021 | --- | --- | --- |
| float | Open | --- | --- | 6009.02 | 61763.56 | 3.8 |
| float | High | --- | --- | 6013.36 | 61781.83 | 3.8 |
| float | Low | --- | --- | 6004.49 | 61673.55 | 1.5 |
| float | Closing | --- | --- | 6009.01 | 61781.80 | 1.5 |
| integer | Volume\_(BTC) | --- | --- | TBD | TBD | TBD |
| float | Volume\_(Currency) | --- | --- | TBD | TBD | TBD |
| float | Weighted\_Price | --- | --- | 6008.93 | 61716.21 | 3.8 |

1. **Tools / methods your team (plan to) use in your study.**

Our team plans to use the following tools and methods:

Tools

* MATLAB software platform.

Methods / Proposed Work

* Our team plans to use Open, High, Low, Volume\_(Currency), Weighted\_Price and Volume\_(BTC) as the predictor variables, i.e., the set X of independent variables, and use Closing as the response variable, i.e., the set Y of dependent variables, in one model
* Our team extracts 2-day, 7-day, and 14-day data from March 2013, March 2016, March 2019, and March 2021 to build and test linear regression, logistic regression, and support vector machine models.
* The 2-day, 7-day, and 14-day data ranges are March 17-18, March 17-23, and March 17-30, respectively.
* We compare, rank, and pick the best fitted model for future price prediction.

1. **Detail description of what problems/questions your team plans to predict / study.**

We chose to split our data set up into groups of date ranges to see if the datasets are proper for building models, how adjusting model parameters will impact a model, and how well different models perform in different datasets. Some main questions would be:

Do the machine learning methods show a difference between time periods and predictability?

What models provide the best performance in predicting bitcoin price?

What level of predictability does each method of machine learning offer?

Chart, line chart, histogram

Description automatically generated

* Simulation study
  + Design I: Linear Regression
  + Design II: logistic regression model selection
  + Design III: SVM

**Presentation Write-up**

Our presentation consists of six main sections.

First, Byunghun will introduce the background of our project, including our purposes of doing this research, the history of bitcoin, and our motivations for choosing bitcoin price prediction as a subject to study. Also included in this section are introductions of the bitcoin historical dataset details, data analysis process and model building process.

Emily will talk about Linear Regression starting from slide 12. Matt is responsible for the logistic regression, and Summer will discuss the SVM models. Byunghun will then take over and present his findings of linear regression, logistic regression and SVM models using Python.

Our team will wrap up our presentation with a summary with interesting findings and talk about further improvements that can be made.

*Resources:*

*Bitcoin's Price History, JOHN EDWARDS, September 21, 2021,* [*Bitcoin's Price History (Investopedia.com)*](https://www.investopedia.com/articles/forex/121815/bitcoins-price-history.asp)

*Bitcoin: A Peer-to-Peer Electronic Cash System, Satoshi Nakamoto,* [*A Peer-to-Peer Electronic Cash System (Bitcoin.org)*](https://bitcoin.org/bitcoin.pdf)

*Bitcoin USD (BTC-USD), Yahoo finance,* [*Bitcoin USD (Yahoo finance)*](https://finance.yahoo.com/quote/BTC-USD/chart?p=BTC-USD#-)

*Bitcoin Price Prediction using Linear Regression, SIDHARTH PANDITA, May 15, 2021,* [*Bitcoin Price Prediction (Medium)*](https://medium.com/hackerdawn/bitcoin-price-prediction-using-linear-regression-94e0e5a63c42)

*Discrete-time, overlap-add, FIR filter, MathWorks,* [*Discrete-time, overlap-add, FIR filter (MathWorks)*](https://www.mathworks.com/help/signal/ref/dfilt.fftfir.html)

*Fit linear regression model, MathWorks,* [*Fit linear regression model (MathWorks)*](https://www.mathworks.com/help/stats/fitlm.html)

*Lasso or elastic net regularization for generalized models, MathWorks,* [*Lasso or elastic net regularization for generalized linear models (MathWorks)*](https://www.mathworks.com/help/stats/lassoglm.html)

*Standardized z-scores, MathWorks,* [*Standard z-scores(MathWorks)*](https://www.mathworks.com/help/stats/zscore.html)

*Other References:*

*Quandl – https://data.nasdaq.com*

*CoinMarketCap - https://coinmarketcap.com*

*Crypto data world -* [*https://www.cryptodatadownload.com*](https://www.cryptodatadownload.com)

*Features and Definitions:*

|  |  |
| --- | --- |
| Features | Definitions |
| Block Size | Average block size in MB |
| Total bitcoins | Total number of bitcoins mined |
| High, Low | Highest and lowest values of different days |
| Number of transactions | Total number of unique Bitcoin transactions per day |
| Trade Volume | USD trade volumes from the top exchanges |