SEIS 763

Machine Learning

Team Project

Bitcoin Price Prediction

12/05/2021

Software User’s Guide

## **Table Contents**

MATLAB

MATLAB Instructions 3

Python

Python Instructions I 4

Python Instructions II 5

## **MATLAB Instructions**

The instructions provide a step-by-step process for uploading, processing, executing, and analyzing machine learning models in MATLAB. When uploading the source data, follow the instructions in the “Uploading the source data” section; when processing the source data, follow the instructions in the “Processing the source data” section; when executing the source data, follow the instructions in the “Executing the source data” section; and when analyzing the source data, follow the instructions in the “Analyzing the source data” section.

### Uploading the source data

1. Locate bitcoin-dataset.csv file.
2. Save bitcoin-dataset.csv file to Final Submission Package directory.
3. Insert bitcoin-dataset.csv file into the readtable function.
4. Execute script.

.

### Processing the source data

1. Convert Unix timestamp to date time structure.
2. Convert dataset to array.
3. Remove NaNs from dataset.
4. Convert array to timetable.
5. Select date range.
6. Convert response variable to down equal to 0 and up equal to 1.
7. Execute z-score.
8. Add predictor and response variables to new variable.
9. Split data into 80/20, train/test, structure.
10. Assign predictor and response data to train and test variables.

### Executing the source data

1. Execute machine learning model with training-testing split.
2. Predict model.

### Analyzing the source data

1. Execute CFM stats, if necessary.
2. Execute graphs and plots, if necessary.
3. Execute ROC curves, if necessary.

## **Python Instructions I**

The instructions provide a step-by-step process for uploading, processing, executing, and analyzing machine learning models in Python. When uploading the source data, follow the instructions in the “Uploading the source data” section; when processing the source data, follow the instructions in the “Processing the source data” section; when executing the source data, follow the instructions in the “Executing the source data” section; and when analyzing the source data, follow the instructions in the “Analyzing the source data” section.

### Uploading the source data

1. Get Kaggle API Token (Credentials) named Kaggle.json.
2. Get Kaggle username and token.
3. Upload Kaggle.json into Google drive.
4. Create a new Colab notebook environment.
5. Mount drive to Colab notebook.
   1. Get authorization code using prompted URL.
   2. Provide an empty box.
   3. Execute config path to Kaggle.json.
6. Got to Kaggle.
7. Copy API command to download dataset.
8. Unzip dataset.
9. Load dataset.

### Processing the source data

1. Convert Unix timestamp to date time structure.
2. Remove NaNs from dataset.
3. Index data time.
4. Create binary columns down = 0 and up = 1.
5. Observe integers to make predictions.
6. Scale features
7. Split data into 80/20, train/test, structure.
8. Assign predictor and response data to train and test variables.
9. Use recursive feature for each machine learning model.

## **Python Instructions**

### Executing the source data

1. Execute machine learning model with training-testing split.
2. Predict model.

### Analyzing the source data

1. Execute CFM stats, if necessary.
2. Execute graphs and plots, if necessary.
3. Execute ROC curves, if necessary.

Note:

Google Colab was used for the python interface and implemented for the bitcoin price prediction project. It was optimal for python to build and publish data by installing libraries and other dependencies while creating machine learning models.

The main body of the code was utilized through the Google Colab Python script that displays the linear regression, logistic regression, and support vector machine models. The recursive feature selection process tested features and the R-square adjusted outputs informed the accuracy of the models and correlation factor of the variables in the data. The model was tested using randomly selected data frames based on first model (linear regression) price prediction graphs predictions using different time frames. The Kaggle dataset was the data source and dataset in Google Colab and the Python scripts were fetched by following steps stated above in the Python instructions.