3. PROJECT ARCHITECTURE

The goal of CryptOL is to use machine learning to predict the trend of Bitcoin prices. The project team selected three machine learning models for the experiment: linear regression, autoregressive integrated moving average (ARIMA), and long short-term memory (LSTM). To evaluate the success of the models, the project examines the directional accuracy of the predictions. That is, if a model predicts the price of Bitcoin will rise and it does rise, then a successful prediction has occurred. In addition, the focus of CryptOL is on short-term price predictions 15 minutes, 1 hour, and 1 day into the future. Cryptocurrency markets are highly volatile. Therefore, the project team concluded that it would be more difficult to arrive at long-term predictions.

3.A PROJECT DESIGN

While CryptOL uses three separate machine learning models, the project exhibits a singular design to tie the components together. The figure below illustrates the general design of the project.

TEST SET

TRAINING SET

DATA GENERATION

MODEL TRAINING

- LINEAR REGRESSION

- ARIMA

- LSTM

USER INTERACTION

EVALUATION

PREDICTIONS

3.A.i DATA GENERATION

CryptOL obtains its data using the Yahoo Finance API. The number of historical Bitcoin prices the project collects depends on the price prediction interval selected by the user. 15-minute predictions collect prices from the previous 60 days. This results in a data set with 5,760 price points ((1,440 minutes in a day X 60 days) / 15-minute intervals). 1-hour predictions collect prices from the previous 60 days resulting in a data set of 1,440 price points. Finally, the 1-day predictions collect prices from the previous 365 days. Unfortunately, this results in a data set with only 365 price points.

3.A.ii SPLITTING THE DATA

Machine learning algorithms generally require that the data collected be split into at least two separate sets before a model is trained. One set is called the training set and the other is the test set. A machine learning model is trained using the training set. Predictions are made by applying the test set to the newly trained model.

CryptOL uses 80% of the data for training sets and 20% of the data for tests sets.

3.A.iii MODEL TRAINING, PREDICTIONS & EVALUATION

This will be covered in the IMPLEMENTATIONS section of this paper.

3.A.iv USER INTERACTION

The user interacts with CryptOL through a graphical user interface. The user selects an algorithm and a time interval if the option is offered. After some time, a prediction is returned to the user. At this point the user can decide whether to utilize the prediction result of the algorithm.

ADD MORE HERE – SEQUENCE DIAGRAM OF CLIENT/SERVER

3.B IMPLEMENTATION

3.B.i LINEAR REGRESSION

CryptOL implemented a simple linear regression model. Linear regression makes it possible to predict the the value of an independent variable based on the values of one or more independent values, or features. For this project, the dependent variable is the price of Bitcoin in 15 minutes, 1 hour, or 1 day. The dependent variables are the features that the project team selected.

53 features were selected for the model. Three of these features are the prices of Bitcoin, Ether, and Dogecoin. The remaining 50 features include the 50 previous of Bitcoin at the selected time interval. The linear regression model trained using these features was able to predict the direction of the market with an accuracy as high as 72% sometimes.

3.B.ii ARIMA (TIME SERIES)