05/10/2021

Please use this template. Retain the gray text. Your new materials—in black 12-point Times New Roman—should not exceed 5 additional pages excluding references and figures. Use the Appendix for bulky material that will be read on an as-needed basis only. Note the evaluation criteria, and leave plenty of time for editing.

# ASSIGNMENT 2P: PROJECT PROPOSAL PLUS

Aidan Duffy

## 2.0 WHAT’S CHANGED (if applicable)

Provide no more than a page of 12-point type explaining what has been changed or added since assignment 1P (if you selected it). Include in this whether and how the material in module 2 influenced this, or refer to reading that you did in working on this assignment (#2).

N/A – did Assignment 1T.

## 2.1 SUMMARY DESCRIPTION, VERSION 2 (as applicable)

One- or two-paragraph overall description of your proposed term project.

I want to develop a machine learning program, called CRYPTOFORECAST, that will help develop price predictions for the cryptocurrency market. Many programs are already tailored to predicting fluctuations in the stock market, and I did not want to reinvent the wheel. Additionally, given the rise of the popularity of investing cryptocurrencies like Bitcoin and Ethereum (and yes, even Dogecoin!), CRYPTOFORECAST will use the same fundamental financial indicators and technical analysis but rely on learning from a dataset that analyzes the largest tokens currently in circulation. This program may help future investors in DCA, or dollar cost averaging, to lower their cost bases, as well as realizing when the end of a “bull run” may be approaching.

## 2.2 I/O EXAMPLES, VERSION 2

At least two specific examples of projected output for designated input. You will not be held to this exactly.

Here is the I/O demonstrating the price history of a given input token:

Graphical user interface, text, application

Description automatically generated

## 2.3 FUNCTIONAL REQUIREMENTS, VERSION 2

Separate your requirements into two approximately even categories (select modest “definite” requirements, otherwise “nice-to-do”). This organization allows you to first attain readily do-able goals without getting bogged down, and then move on to other goals as you are able. State requirements in declarative language as advised in assignment 1. Giving each requirement a label (e.g., “(Recognize 0-9):”) helps with clarity and readability.

Your response replaces this.

### 2.3.1 Definite Functional Requirements (first priority)

### 2.3.1.1 Definite #1: Read Datasets

CRYPTOFORECAST shall read in and format all of the datasets.

### 2.3.1.2 Definite #2: Filter and Normalize Datasets

CRYPTOFORECAST shall sort data, apply necessary filters, and normalize the filtered datasets.

### 2.3.2 Nice-to-do Functional Requirements (second priority)

### 2.3.2.1 Nice-to-do #1: Integrate Visualizer for Datasets

CRYPTOFORECAST shall provide visualizations and graphs for the price history of cryptocurrencies in the datasets.

## 2.4 V2: HOW SUCCESS WILL BE ASSESSED

Explain, as specifically as possible (quantification is ideal) how success of the project should be assessed.

Luckily, given the short feedback cycle of financial markets, success can be assessed almost immediately. I will be able to cross reference my model’s price predictions daily with actual market prices.

## 2.5 V2 TECHNOLOGY EXPLANATION

Explain what two specific machine learning technologies you intend to use--and exactly why you feel they apply to your particular project. One of the two may be emphasized as the implementation and the other as an alternative or as a complement—discussed but not implemented if need be. Include a discussion of neural nets, whether you select them or not.

prices. After doing research on similar projects that have already been completed, I will attempt to utilize a recurrent neural network, or RNN. More specifically, I will attempt to implement an LSTM, or long short term memory, setup. This type of neural network was selected most often because of its ability to store past information for sequence prediction problems, making it a most powerful tool in a machine learning developer’s toolbox. Since the dataset will be sequential, the data from one hour or day is directly related to the data point from the previous unit of time, again reaffirming the importance of this style of neural network. As a proposed alternative, in my research I found an algorithm called “XGBoost”(Extreme Gradient Boosting), which a tree based supervised learning algorithm. It is often used as well as it is very efficient. The main reason it is an alternative is because all of the data would first need to be preprocessed and transformed into a supervised learning problem first.

## 2.6 CODE EXAMPLES

Show fragments of code execution. For example, if you are using TensorFlow, show that you have run some relevant code. The more relevant to your project, the better. This does not need to complete at this stage.

This is the code addressing functional requirement number one:

Table

Description automatically generated

Here is code snippet 2 that addresses the graphing functionality of price histories:

Graphical user interface, text, application, email

Description automatically generated

## 2.7 DATA SOURCES V2

Explain whether or not your project requires data. If so, describe were you will obtain it. Be careful about this because you won’t have a project if it needs data and you have to spend too much time hunting and gathering it.

This project most definitely requires a dataset, and as of writing this, I am using one acquired from Kaggle [here](https://www.kaggle.com/sudalairajkumar/cryptocurrencypricehistory). If I determine this data is insufficient, there are ample other sources of data available on Kaggle that I will be able to utilize.

## 2.8 REFERENCES FOR PROPOSAL V2

Fill in, and also cite each of the following (e.g., “[2]”) within the text. References can include specific places in the notes and textbook. You are free to include references used in the prior assignment version. Keep in mind that this “use of resources” is a whole evaluation criterion.

[1] <https://towardsdatascience.com/lstm-time-series-forecasting-predicting-stock-prices-using-an-lstm-model-6223e9644a2f>

[2] <https://cs230.stanford.edu/projects_winter_2020/reports/32066186.pdf>

[3] <https://medium.com/swlh/a-technical-guide-on-rnn-lstm-gru-for-stock-price-prediction-bce2f7f30346>

## 2.9 Evaluation of Assignment 2



2.10 Appendix