# Machine Learning Engineer Nanodegree Capstone Proposal

## Build a Stock Price Indicator

### 1. Domain Background

Many investment firms have been using financial models for predicting stock prices to yield profits.

#### 1.1 Random Walk Theory

Random walk theory states that the changes in stock prices are unpredictable regardless of the past movements which imply that the uncertainty of a stock price going up is equal to going down(Malkiel 1973; “Eugene Famaâs Efficient Market Hypothesis,” n.d.)(Malkiel 1973). This important financial theory got its name from a book (A Random Walk Down Wall Street) which was written by Burton Malkiel in 1973.

#### 1.2 Efficient Market Hypothesis

Random walk theory was later superseded and explained by the efficient market hy(Malkiel 1973)pothesis. According to the efficient market hypothesis, it is impossible to "beat the market" because the stocks are priced accurately and reflect all available information (“Eugene Famaâs Efficient Market Hypothesis,” n.d.). As per this hypothesis, investors are taking a gamble when they buy or sell a security as there is no guarantee they could buy a stock at a bargain price and sell at a much higher price to make a profit. This implies that investors have to buy a higher risk investment to make profits. However many investors (most notably Warren Buffet) have made profits consistently over a substantial span of time.

#### 1.3 Fundamental Analysis

Fundamental Analysis is a complete top to the bottom approach where the analysts tend to study all factors that could influence the stock's value. They are the overall economy, country's economy, sector analysis and company analysis. With the help of fundamental analysis, investors should be able to compare the current price of a stock against their own calculated intrinsic value which could possibly give them a good indication of whether a stock is undervalued or overvalued.

#### 1.4 Technical Analysis

The purpose of doing technical analysis is to predict a stock's price based on its past movements. Technical analysis assumes that the fundamental analysis has already been accounted for a company's current stock price. This type of investment strategy is generally used by short-term investors, where holding periods could range anywhere from seconds to weeks.

### 2. Problem Statement

To build a stock price predictor that makes use of past trading data to produce projected

estimates. For this purpose, I would be using supervised learning regression algorithms because the output to be predicted is a continuous value. The dataset that would be used to develop the machine learning model will predict the Adjusted Close price of a stock, which is a numerical value. The model's accuracy could be evaluated by using the actual adjusted close price.

3. Datasets and Inputs

The dataset would contain multiple metrics ( like Open, Close) including Adjusted Close Price that provides enough information about a particular stock. Using these metrics we can build the features that are necessary to predict the outcome. These metrics are numerical values. To get the input dataset, Yahoo Finance is being considered as it is easy to download the dataset for a particular stock. We only need to type the stock ticker symbol and then choose historical data tab to download the dataset.

4. Solution Statement

Using the metrics that come with the dataset we can build the features that are necessary to train the model. For features, I would be considering all the metrics that are in the dataset except Adjusted Close price. As mentioned earlier, I would be using many supervised learning regression algorithms to train the model as output to be predicted is a continuous value.

5. Benchmark Model

We can make use of S&P 500 stock market index as it reflects major sectors of the market. I would be using the stock data of the companies that are listed in this index to train and test the machine learning model. The purpose of using a stock market index is to gauge the changes in a portfolio of stocks representing a sample of the overall market instead of tracking thousands of stocks that are being traded.

6. Evaluation Metrics

Accuracy score could be calculated using the predicted adjusted close price and the actual adjusted close price to quantify the performance of both the benchmark model and the solution model.

7. Project Design

To build stock predictor, I would develop training and prediction interfaces. Both the interfaces would accept a list of stock ticker symbols and date ranges as inputs. The training interface would build the necessary the model while the prediction interface would predict the outcome. Date ranges that should be given for the prediction interface should fall after date ranges given in the training interface.

References

“Eugene Famaâs Efficient Market Hypothesis.” n.d. In *The Efficient Market Hypothesists*.
Malkiel, Burton Gordon. 1973. *A Random Walk Down Wall Street*.