Machine Learning Engineer Nanodegree

Capstone Proposal

Sui Watchorn
August 31, 2018

Proposal

Domain Background

Harry Markowitz, the 1990 Nobel Laurate for Economics, dedicated his work to the Efficient Market Hypothesis (EMH) maintaining all securities are efficiently priced in the long run and drastic changes are attributed to random fluctuations given the investors behave logically. Under this theory, the stock market is not predictable.

Before embarking in a career in IT, my undergrad studies were focused in Economic & Finance Management and finance or specifically, stocks, have remained a personal interest of mine. Machine Learning (ML) wasn't available to the wider public at the time of my undergraduate studies and if it was, it was more in research and development to define the theories and algorithms that comprise the foundation of ML.

Technological advancement of hardware and software have since enabled the rapid development of ML and it has been widely adopted and used to do predict random stock market fluctuations. These predicts are used to try to beat the market by assigning levels of risk to reward or returns in the stock market. ML strategies from Reinforced Learning to Neural Networks have been used to build models precisely to predict the risk and returns in stock market/prices with the hopes of maximizing rewards/returns.

Problem Statement

National Association of Securities Dealers Automated Quotations (NASDAQ) of the United States was founded in 1971 and it makes up part of the stock market. It is an electronic stock exchange with more than 3,300 company listings trading on the exchange. Taking into consideration with approximately 253 trading days per year for over 3,300 companies the data would be substantial for daily and/or hourly trades. This

project will narrow down the scope and explore the use of Machine Learning in stock prediction for daily trading specifically to the AT&T stock (T). Reinforced Learning with Q learner will be used to identify buy or sell opportunities and compare it with actual market performance to gauge model performance or whether T beat the actual market performance.

Datasets and Inputs

Global and domestic stock market information, including NASDAQ, is widely available for download via API or CSV from sources such as Yahoo! Finance, Quandl, Google Finance, etc. The project will focus primarily on T and its daily performance. The company has been publicly traded since 1983 with total of 9,020 days of trading up to the present, not including holidays and weekends. Additional data is available for hourly trades and transactions but for the scope of this project, it will focus primary on daily pricing for 1-yr.

The datasets used in this project will be sourced from Yahoo! Finance, Quandl, Google Finance, etc. These datasets will be used for comparison and feature engineering throughout this project.

Solution Statement

The datasets with open-high-low-close daily data will be used to develop a predictive model to determine the daily pricing for buy or sell positions of the selected equity—T. In the interest of this project, dividends, fees, and taxes will not be considered.

Reinforced Learning strategies will be used to identify the positions and a Q learner will be developed to evaluate the metrics. Success or failure will be determined by the returns or rewards.

Benchmark Model

ML strategies will be employed in the beginning to determine a benchmark model that will be used to test against the predictive modeling. The model will then be used in a set time period to track daily performance as a gauge to determine trading position of buy or sell. The best quantitative representation in the stock market is returns or rewards in the form of money made. We can use an initial capital amount to purchase the selected stock and track the returns thru the trial period. The performance will be tracked and charted daily.

Evaluation Metrics

Direct comparison between the model and actual time-series open-high-low-close information for T will be used to evaluate the model. A benchmark with standardized results will show which strategy is more successful—actual market performance vs the predictive model.

Project Design

This project intends to use a technical approach to stock predictions using ML strategies. Hopefully, it can be used to enhanced personal understanding of market trends and prove random fluctuations to be applicable for everyday trading, if it is determined to be a viable model.

Below is a high-level overview of project outline:

1. Download the needed data from either and/or a combination of the sources from Yahoo! Finance, Quandl, Google Finance, etc.
2. Extract, explore, and process the data to refine/fit the project objectives and establish the timeframe of the trial.
3. Develop and calculate identifiers for feature development.
4. Use the identifiers for the Q learner to determine patterns and execute the appropriate actions of the identified patterns.
5. Train the model using Q learner to filter out outliers—random fluctuations.
6. Test the model using initial capital to refine Q learner.
7. Compare the model for rewards or returns to determine success or failure.
8. Identify lessons learned and areas for improvement.