Capstone Proposal

January 2020

Shamus Clifford

Machine Learning Engineer

Capstone Project

Domain Background

The stock market is a place where you can buy, sell, and trade stocks. A stock is a share of a company. So, buying a stock in a company is buying part of that company. If you buy a stock and the price of the company goes up, you will make money. If the price of that same company goes down, you will lose money. Being able to predict whether the price of a stock will rise or fall would be lucrative to any investor.(Amadeo 2019)

The **Random Walk Hypothesis** is a financial theory that states stock market prices are random, that the changes in price are unpredictable.(Smith 2019) The Random Walk Hypothesis is consistent with the **Efficient Market Hypothesis** (EMH). EMH is an investment theory that states that the price of a stock reflects all information made available to market participants at any given time.(Van Bergen 2012) Based on Random Walk and EMH, it would be impossible for an investor to predict future prices.

I disagree with these theories. I believe an investor, armed with machine learning models, can predict the price of a stock better than the market can. Technical analysis or "identifying trading opportunities by analyzing statistical trends gathered from trading activity," can help an investor predict future prices.(Hayes 2019) Machine learning models can aid in technical analysis.

Problem Statement

I plan on proving that EMH is false, that the market is not efficient and an investor can predict the price of a stock, meaning a stock's future price is not totally random as the Random Walk Hypothesis suggests. I will analyze a market sectors price movements over a 5 year span to predict the stock price in the future (1 day, 7 days, 28 days). Success will be measured against the Naive Method. The Naive Method "forecasts the next values using the last value observed."(Ramírez 2017) Using the Naive Method, the price of the stock on January 3 would be the same price it was at the end of January 2. This method works well for a random walk forecast.(Hyndman & Athanasopoulos 2018) I will test this against the AWS DeepAR Forecasting algorithm which is a "supervised learning algorithm for forecasting scalar (one-dimensional) time series using recurrent neural networks (RNN)."(Mishra 2019) Using DeepAR, I will predict stock prices better than the Naive Method.

Datasets and Inputs

For this project I will use stocks from the Information Technology Sector. DeepAR performs better when stocks belong to a cluster.(Das 2018) The list of stocks that will be used are:

- Apple (AAPL)
- Microsoft (MSFT)
- Intel (INTC)
- Cisco (CSCO)
- Adobe (ADBE)
- Salesforce (CRM)
- NVIDIA (NVDA)
- Accenture (ACN)
- PayPal (PYPL)
- Oracle (ORCL)

These stocks were chosen from the Vanguard Information Technology ETF. I chose the top ten weighted stocks in the fund (ignoring Visa and Mastercard which I do not believe fit perfectly in the Information Technology sector). Historical data for these stocks comes from Yahoo Finance. I will use the daily adjusted close price of each stock over a 5 year span (January 1, 2014 - December 31, 2018) as training data and 1 year of test data (January 1, 2019 - December 31, 2019). I chose a 5 year span because I wanted DeepAR to detect any trends in seasonality of the stocks.

Solution Statement

The proposed solution involves applying deep learning techniques that have proved to be successful for forecasting time series. The first step would be to get the data from Yahoo Finance. This could be done in a variety of ways. You could get the data using the API available on rapidAPI, download it directly from Yahoo Finance to upload, or by web scraping. Then, the data needs to be formatted for DeepAR.

After that, we will use DeepAR and the adjusted close price for stocks to train an Estimator and create a Predictor which will predict the future stock price. The model will be determined successful if it can predict future prices better than the Naive Method.

Benchmark Model

The benchmark model will be the Naive Method. The Naive Method sets all forecasts to be the value of the last observation (the previous days adjusted close price). This method works well for random walk forecasts.(Hyndman & Athanasopoulos 2018)

Evaluation Metrics

The evaluation metric will be Mean Absolute Percentage Error (MAPE). I will compare the MAPE for the Naive Method and compare it to the DeepAR model. The method with a smaller MAPE did a better job at predicting the future stock price.(Halimawan & Sukarno 2013)

Project Design

First, I will collect the data from Yahoo Finance. Then, I will clean the data removing details that are irrelevant. The details I will focus on are date, stock ticker, and adjusted close price for the 6 year span January 1, 2014 - December 31, 2019. Next, I will prepare my data by splitting it into train and test sets. The training data will be from January 1, 2014 - December 31, 2018. The test data will be from January 1, 2018 - December 31, 2019. Finally, I will prepare my data by turning it into a JSON Lines format for the DeepAR Algorithm.

After my data has been prepared, I will train the DeepAR model using the training data. Once the model has been trained, I will be able to evaluate my model and compare my results to the Naive Method. At this point I can make adjustments to my model if it is not performing as expected.

Once my model has been refined, I will deploy it and create a simple web app to predict future stock prices for the 10 stocks listed previously. Creating the web app will involve having a deployed model, a Lambda function and an API for my for my web app to interact with.

References

Amadeo, Kimberly. "Before You Invest in the Stock Market, Make Sure You Know What It Is." The Balance, The Balance, 25 July 2019, www.thebalance.com/what-is-the-stock-market-how-it-works-3305893.

Das, Binoy. "Aws-Samples/Amazon-Sagemaker-Stock-Prediction." GitHub, 20 Nov. 2018, github.com/aws-samples/amazon-sagemaker-stock-prediction/blob/master/notebooks/dbg-deepar.ipynb.

Halimawan, Alam Akbar, and Subiakto Sukarno. "STOCK PRICE FORECASTING ACCURACY ANALYSIS USING MEAN ABSOLUT DEVIATION (MAD) AND MEAN ABSOLUTE PERCENTAGE ERROR (MAPE) ON SMOOTHING MOVING AVERAGE AND EXPONENTIAL MOVING AVERAGE INDIKATOR." The Indonesian Journal of Business Administration, vol. 2, 13 Nov. 2013, pp. 1613–1623., https://media.neliti.com/media/publications/68283-EN-stock-price-forecasting-accuracy-analysi.pdf.

Hayes, Adam. "Technical Analysis Definition." Investopedia, Investopedia, 18 Aug. 2019, www.investopedia.com/terms/t/technicalanalysis.asp.

Hyndman, Rob J, and George J Athanasopoulos. Forecasting: Principles and Practice. 2nd ed., OTexts: Melbourne, Australia, 2018.

Mishra, Abhishek. "Machine Learning in the AWS Cloud: Add Intellegence to Applications with Amazon SageMaker and Amazon Rekognition." Amazon, John Wiley & Sons, Inc., 2019, docs.aws.amazon.com/sagemaker/latest/dg/deepar.html.

Ramírez, Rubén Guerrero. Different Methods for Forecasting Time Series Tutorial, 2 Dec. 2017, rstudio-pubs-static.s3.amazonaws.com/336602_0a55f2341eab4637b27766f6619af477.html.

Smith, Tim. "Random Walk Theory." Investopedia, Investopedia, 18 Nov. 2019, www.investopedia.com/terms/r/randomwalktheory.asp.

Van Bergen, Jason. "Efficient Market Hypothesis: Is The Stock Market Efficient?" Forbes, Forbes Magazine, 1 May 2012, www.forbes.com/sites/investopedia/2011/01/12/efficient-market-hypothesis-is-the-stock-market-efficient/.