Stock Price Prediction Using Machine Learning Algorithms

Kate Catalena
UIN: 421003524

Zebo Xiong
UIN: 128007734

Mohd Faisal Khan
UIN: 727006629

Nader Ghasemi
UIN: 121002001

***Abstract*—Predicting stock market trends has been and still is a very popular area of research. Stock market prices represent changes in the economy. Many factors go into determining what a stock market price will do in the future, which makes predicting future trends nearly impossible. There is a lot of machine learning research devoted to this idea because it is believed that if you can identify underlying trends, you can better anticipate what a stock market price will do down the road. In this research, we plan to take a survey of current machine learning techniques and apply them to financial data to analyze which techniques perform better in terms of predicting future stocks. A survey of the most commonly used methods and how well they perform on the same dataset would be useful in understanding each method, including their strengths, weaknesses, and overall performance in predicting future stocks.The objective here is to, based on the historical daily/hourly data, we wish to make certain predictions on the next days's trends.**

***Index Terms*—Machine Learning, Stocks, SVR, SVM, LTSM, ANN**

I. Introduction

Globalization deepens the interaction between the financial markets around the world. Shock wave of US financial crisis hit the economy of almost every country and debt crisis originated in Greece brought down all major stock indices. Through this study, we hope we can make a little contribution or voice on helping people get noticed by the possible economic down trends.

II. Related Work

In previous studies, the features selected for the inputs to the machine learning algorithms are mostly derived from the data within the same market under concern. Such isolation leaves out important information carried by other entities and make the prediction result more vulnerable to local perturbations.

Efforts have been done to break the boundaries by incorporating external information through fresh financial news or personal internet posts such as Twitter. These approaches, known as sentiment analysis, replies on the attitudes of several key figures or successful analysts in the markets to interpolate the minds of general investors. Nowadays, no financial market is isolated. Economic data, political perturbation and any other oversea affairs could cause dramatic fluctuation in domestic markets.

We also propose the use of global stock data in associate with data of other financial products as the input features to machine learning algorithms such as SVM. In particular, we are interested in the correlation between the closing prices of the markets that stop trading right before or at the beginning of US markets. Oversea markets that closes right before or at the beginning of the US market trading should provide valuable information on the trend of coming US trading day, as their movements already account for possible market sentiment on latest economic news or response to progress in major world affairs.

Commodity prices foreign currency data are also listed as potential features, as different financial markets are interconnected. For instance, slowdown in US economy will definitely cause a drop in US stock market. But at the same time, USD and JPY will increase with respect to its peers as people seek for asset havens. Such interplay implies the underlying relationship between these financial products and the possibility of using one or some of them to predict the move of the other ones

Leong et al. looked at incorporating a support vector machine to predict movements in future stock prices [1]. To solve the SVM optimization problem, they used minimum graph cuts, which allowed their SVM to better predict more complex tree structures. With a 3-fold cross-validation method, they were able to achieve 78% accuracy with their SVM, which indicates their model was successful in learning and predicting future stock market prices without overfitting. Huang et al. used a neural-network method to better predict Chinese stock prices. More specifically, they used a Bayesian-LTSM neural network method [2]. They use six indicators for input into their LTSM. For their data, they looked at 26 years' worth of data and then broke down each year into a week. With this model, they reported that, on average, their Bayesian-LTSM predicted stock market 25% percent better than a regular LTSM approach.

III. IMPLEMENTATION

For the purposes of this research, we propose testing different methods such as, an SVM, Bayesian-LTSM, multiple linear regression, XGboost, etc on stock market data to compare and contrast them with the hope we will be able to identify which method outperforms the others. For our research, we are going to focus on one specific sector, specifically the technology sector. We will look at historical data for companies Apple, Amazon, Google, Microsoft, and Facebook. We plan to use data from Yahoo Finance/Quandl and history data. Instead of looking at just daily stock market data to train our models, we hope to look at hourly or even minute-by-minute data. Many of the other papers reviewed looked at data on a daily or even weekly basis. We suspect that looking at a smaller increment

of time will better train the models, and, in the end, allow for higher accuracy rates across the board.

Together with selected features and provided latest effective algorithms, we also will make final evaluation of current machine learning techniques and apply them to financial data in order to analyze which techniques perform better in terms of predicting future stocks.

IV. CHALLENGES

Considering the four of us do not have any formal experience with machine learning, we expect that implementing some of these methods will be quite difficult. We are most concerned with the new aspect of time series with our data set. We are excited about completing this project since it will equip us with additional machine learning knowledge that we can use in our own academic and professional careers.

V. REPOSITORY

GITHUB:
https://github.tamu.edu/dior/CSCE_633_Machine_Learning

ACKNOWLEDGMENT

We would like to thank Dr. Bobak Mortazavi for his contributions to our understanding and the relevant background knowledge needed in order to successfully decide how to implement the information discussed throughout this proposal. We would also like to thank Arash Pakbin for his guidance and advice. Lastly, we would like to thank our classmates who offered their advice and support throughout the proposal generation lifespan and for their contributions to discussions in class that aided in our understanding of the material. We would like to continue getting their aid and support while working on this project.

REFERENCES

[1] Carson Kai-Sang Leung, Richard Kyle MacKinnon, and Yang Wang. 2014. A machine learning approach for stock price prediction. In Proceedings of the 18th International Database Engineering & Applications Symposium (IDEAS '14), Ana Maria Almeida, Jorge Bernardino, and Elsa Ferreira Gomes (Eds.). ACM, New York, NY, USA, 274-277. DOI: https://doi.org/10.1145/2628194.2628211

[2] Biao Huang, Qiao Ding, Guozi Sun, and Huakang Li. 2018. Stock Prediction based on Bayesian-LSTM. In Proceedings of the 2018 10th International Conference on Machine Learning and Computing (ICMLC 2018). ACM, New York, NY, USA, 128-133. DOI: https://dl.acm.org/citation.cfm?id=3195170