**STOCK MARKET PREDICTION**

By

DEPARTMENT OF INFORMATION SCIENCES

SCHOOL OF COMPUTING AND INFORMATICS TECHNOLOGY

A Concept Paper submitted to the School of Computing and Informatics Technology

For the Study Leading to a Project Proposal in Partial Fulfillment of the

Requirements for the Award of the Degree of Bachelor of

Software Engineering of Makerere University.

Supervisor

Grace Kamulegeya

Department of Information sciences

School of Computing and Informatics Technology, Makerere University

gkamuleg@cit.mak.ac.ug, +256-41-540628, Fax: +256-41-540620

July 2019

GROUP MEMBERSHIP

|  |  |  |  |
| --- | --- | --- | --- |
| # | Names | Registration number | Signature |
| 1 | lubwama isaac | 17/U/5739/PS |  |
| 2 | kintu Arnold | 17/U/5188/EVE |  |
| 3 | olara Samuel obadia | 17/U/9634/EVE |  |
| 4 | kaluuma mahad | 17/U /1405/EVE |  |

# **Introduction**

Stock market prediction is a serious challenge for investors and corporate stockholders to forecast the daily behavior of stock market which helps them to invest with more confidence by taking risks and fluctuations into consideration. In this paper, by applying linear regression for forecasting behavior of APPLE Inc. data set, we prove that our proposed method is best to compare the other regression technique method and the stockholders can invest confidentially based on that.

Stock market prediction is the act of trying to determine the future value of a company stock or other financial instrument traded on a financial exchange. The successful

prediction of a stock’s future price will maximize investor’s gains.

This study aims to use linear regression and support vector regression models to predict price changes and evaluate different models’ success by withholding data during training and evaluating the accuracy of these predictions using known data.

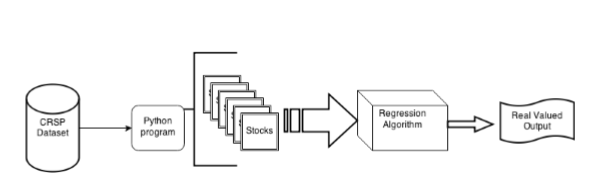
The SVM algorithm is employed to increase thee accuracy of the prediction of linear regression. Proposed model is based on the study of stocks historical data and technical indicators.

# **Background to the problem**

Data mining techniques have a more successful rate compared to traditional statistical method by discovering hidden knowledge of data.

The dataset that was used was collected from CRSP datasets as a collection of comma-separated values where each row consisted of a stock on a speciﬁc day along with data on the volume, shares out, closing price, and other features for that day in time.

The Python scientiﬁc computing library NumPy was used along with the data analysis library pandas in order to convert these CSV ﬁles into pandas Data Frames that were indexed by date. Each speciﬁc stock is a view of the master Data Frame that is ﬁltered based on that stock’s ticker. This allowed efﬁcient access to stocks of interest and convenient access to date ranges. These stock Data Frame views are then used as the data to be fed into our regression black boxes

. 

*Figure 1*: Data-ﬂow of the program showing how stock data turns into prediction value vectors

# **Problem Statement**

## **Problem Structure**

Stock market price prediction is a problem that has the potential to be worth billions of dollars and is actively researched by the largest ﬁnancial corporations in the world.

Many investors worldwide find it very hard to decide on which stocks of different companies that they should purchase and often end up making the wrong decisions and hence making losses either at a private level or at a company level. These investors are equipped with the data of the stocks but un fortunately not able to decide efficiently on where to place their money.

## **Problem Resolution**

The project is targeted to use a sample of datasets of the stock prices of random companies to analyse the data and help build a model that will help the investors in easing their decision making by giving them an overview or a bigger picture of the stocks. This is going to be achieved by analysing the increase and decrease of stock prices overtime and being able to predict the stock prices in a given period of time hence helping the investors from various organizations or private investors in deciding efficiently which stocks they would opt for.

# **Objectives**

* The project should help us derive a model that will diagrammatically structure the data from the datasets into graphs e.g. bar graphs and pie charts to make it easier for the investors to understand the data.
* The project should help the investor know the assumed stock prices of any given time from now in order for him/her to predict or plan accordingly.
* The project should be able to help the investor deduce the most appropriate stock to choose from after viewing the analysis done on the datasets that were chosen.
* To study and improve the algorithms and steps used to predict the stock price

# **Methodology**

## **Data Representation**

The dataset that was used was collected from the [APPL Inc.](https://finance.yahoo.com/quote/AAPL/) as a collection of commas separated values where each row consisted of a stock on a specific day along with data on the volume, shares out, closing price, and other features for that day in time. The Python scientific computing library NumPy was used along with the data analysis library pandas in order to convert these CSV files into pandas Data Frames that were indexed by date. Each specific stock is a view of the master Data Frame that is filtered based on that stock’s ticker. This allowed efficient access to stocks of interest and convenient access to date ranges.

## **Predication method**

Regression predicts a numerical value. Regression performs operations on a dataset where the target values have been defined already. And the result can be extended by adding new information. The relations which regression establishes between predictor and target values can make a pattern. This pattern can be used on other datasets which their target values are not known. Therefore, the data needed for regression are 2-part, first section for defining model and the other for testing model. In this section we choose linear regression for our analysis. First, we divide the data into two parts of training and testing. Then we use the training section for starting analysis and defining the model. 70% data used for training purpose and 30% data used for testing purpose.

**Types of Regression Models:**

The price prediction function provides a few regression models that can be chosen to perform the prediction. This includes

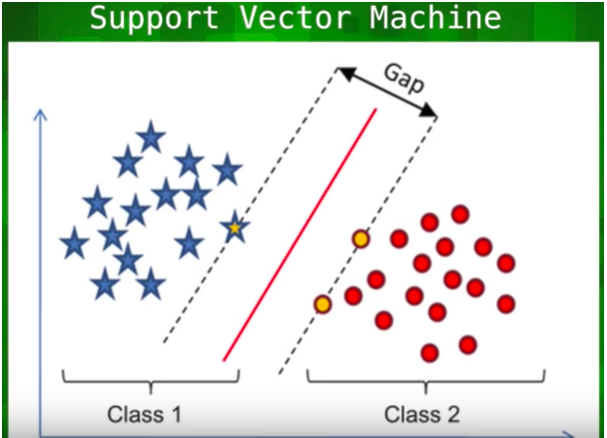
1) Linear Regression

2) Support Vector Regression (SVR)

In Machine learning, support vector machines are supervised learning models with learning algorithms that analyses data used for classification and regression analysis.

Given a set of training examples, each example belonging to one of two categories, an SVM training algorithm builds a **model** that **assigns new examples** to one category or the other, making it a non-probabilistic binary linear classifier.

An SVM model is a representation of the examples as **points in space clear gap,** graphed so that the **examples of the separate categories are divided** by a that is as wide as possible. New examples are then mapped into that **same space and predicted** to **belong**to a **category based on which side of the gap they fall under.**



.

*Figure 2.* The examples would fall closer to the gap than the rest of the examples.

# **Outcome**

In this paper, it is discovered that Stock Market Prediction is an important issue for financial investors to decide which stocks one should buy and sell. Comparative performance efficiency of different techniques based on methodologies and datasets are beneficial in modifying them with the current methodologies for efficiently predicting the stock market prices. This study will help researchers in understanding the different machine learning approaches used till now along with pros, cons and their performance efficiency. This concise information will help them to explore other possibilities

# **References**

Altieri, G.V., P.M.M. Manoharan, R.M. Pai and A. Nayak, 2015. Stock market prediction: A big data approach. Proceedings of the IEEE Region 10th Conference on TENCON 2015, November 1-4, 2015, Macao, China, pp: 1-5.  
  
Bache, A. and M.B. Chanda, 2016. Stock market prediction using text opinion mining: A survey. Int. J. Adv. Res. Compute. Sci. Software Eng., 6: 566-569.  
  
  
Cheng, C.H. and Y.S. Chen, 2007. Fundamental analysis of stock trading systems using classification techniques. Proceedings of the International Conference on Machine Learning and Cybernetics, Volume 3, August 19-22, 2007, Hong Kong, China, pp: 1377-1382.