MANTECH Publications

Journal of Business Analytics and Information System Management
Volume 1 Issue 1

Stock Market Analysis and Prediction

Adarsh S B[1], Anunay Kumar[2], J Aditya[3], Piyush Rai[4]

Students[1, 2, 3, 4]

Department of f Information Science and Engineering

National Institute of Engineering Mysuru, Karnataka, India

Corresponding Author's email id: *anunaykumar73@gmail.com[2]*

Abstract

A stock market, equity market or share market is the aggregation of buyers and sellers (a loose network of economic transactions, not a physical facility or discrete entity) of stocks (also called shares), which represent ownership claims on businesses; these may include securities listed on a public stock exchange as well as those only traded privately. Analysis of these stocks and their volume movements has always been a critical task. A lot of money is invested in knowing the nature of stock and predicting its movements. By using some simple yet effective set of algorithms, our aim is to analyze and predict the stocks of various companies targeting any one stock exchange. Since prediction can help investors by providing an insight, we aim to provide more and more accurate predictions.

Keywords: *Equity, Stock Market*

INTRODUCTION

Prediction is a process to make assumption of future based on existing data. The more precise the prediction, the more it could be easier to make decision for future. Prediction of stock exchange trends has been an interesting topic in the field of pattern recognition and machine learning because of its possible monetary profit. A stock market is an organized set-up with a regulatory body and has registered members who can buy or sell shares. It's a public

market, where different companies invest high capital and do trading of their shares. Stock market prediction provides information about stock market, which can help the shareholders to make decision about trading. It may serve as warning system for long-term shareholders while for short-term investors may serve as recommender system. One of the problems with Stock market data is that the available data is highly volatile with very high dimensions. Many of the attributes are highly correlated and makes prediction of stock market a highly challenging, complicated and daunting task. In this paper we have shown the analysis of stocks and then the prediction using Monte Carlo based on the sampling dataset used. Different algorithms give different results and use their own proportion of datasets. Here Linear Regression and Vector Regression methods were put to use. The percentage of dataset used by them is in between 62-89 percent. Hadoop is an open source, Java-based programming framework that supports the processing and storage of extremely large data sets in a distributed computing environment. But using Hadoop compared to python packages was complex and consumes more memory. Quandl python package was used instead to create data sets and organize them. For experimental purpose we have considered here the stock data of Amazon, Apple, Google and Microsoft. The past values are used to build a classification function and then uses this function to predict about future values.

II. LITERATURE REVIEW

Predictive analytics encompasses a variety of statistical techniques from predictive modelling, machine learning, and data mining that analyze current and historical facts to make predictions about future or otherwise unknown events. There are various data analysis and prediction models available but stock market deals with very large data sets. D. Kumar and S. Murugan [1] have introduced a new method by combining time-series data with Artificial Neural Network (ANN). This prediction model is based on a feedforward ANN with back propagation. K. Abhishek, A. Khairwa, T. Pratap and S. Prakash [2] have developed a forecasting model for the Microsoft Corporation. This forecasting model includes a two-step procedure. In first-step, ANN go through training with the genetic algorithm. This algorithm teaches the prediction model to identify the new trends. In a second step the results are

obtained by analyzing the trained prediction model to the dataset. Experimenting with real time data required a more dynamic approach than Artificial Neural networks.

Linear regression is perhaps one of the most well-known and well understood algorithms in statistics and machine learning. Linear regression [3] was developed in the field of statistics and is studied as a model for understanding the relationship between input and output numerical variables but has been borrowed by machine learning. It is both a statistical algorithm and a machine learning algorithm. In linear regression, the relationships are modeled using linear predictor functions whose unknown model parameters are estimated from the data. Such models are called linear models. Most commonly, the conditional mean of y given the value of X is assumed to be an affine function of X; less commonly, the median or some other quantile of the conditional distribution of y given X is expressed as a linear function of X. Like all forms of regression analysis, linear regression focuses on the conditional probability distribution of y given X, rather than on rather than on the joint probability distribution of y and X, which is the domain of multivariate analysis

Support Vector Machine [4] can also be used as a regression method, maintaining all the main features that characterize the algorithm (maximal margin). The Support Vector Regression (SVR) uses the same principles as the SVM for classification, with only a few minor differences. First of all, because output is a real number it becomes very difficult to predict the information at hand, which has infinite possibilities. In the case of regression, a margin of tolerance (epsilon) is set in approximation to the SVM which would have already requested from the problem.

Shunrong Shen and Tongda Zhang [6] proposed an algorithm which employed the notion of temporal correlation among numerous products and global markets to make prediction of next day stock market trend. Prediction accuracies were measured using Support Vector Machine (SVM) algorithm. Wen Fenghua et [7] used singular spectrum analysis (SSA) to decompose stock price in diverse terms with varying economic features and made price prediction by introducing these features into the support vector machine (SVM). Yanshan Wang [8] proposed an empirical study on the Korean and Hong Kong stock market. Principal Component Analysis was

used for features selection and Support Vector Machine for stock market prediction.

Rohit Choudhry [9] proposed a hybrid machine learning system to predict stock prices by using correlation among stock prices of different companies. Hybrid system was based on Genetic Algorithm (GA) and Support Vector Machine (SVM). Karazmodeh [10] employed Particle Swarm Optimization [PSO] and Support Vector Machine to propose a computationally more efficient model to predict stock price by using correlation among stock prices of various companies. Monte Carlo method is a broad class of computational algorithms that rely on repeated random sampling to obtain numerical results. Their essential idea is using randomness to solve problems that might be deterministic in principle. They are often used in physical and mathematical problems and are most useful when it is difficult or impossible to use other approaches. Monte Carlo methods are mainly used in three distinct problem classes: optimization, numerical integration, and generating draws from a probability distribution. By the law of large numbers, integrals described by the expected value of some random variable can be approximated by taking the empirical mean of independent samples of the variable. The central idea is to design a judicious Markov chain model with a prescribed stationary probability distribution. These flows of probability distributions can always be interpreted as the distributions of the random states of a Markov process whose transition probabilities depend on the distributions of the current random states.

III. CLASSIFICATION APPROACH EMPLOYED FOR PREDICTING STOCK EXCHANGE DATA

A. Linear Regression

Linear regression is a linear approach used to express the relationship between two variables; one is called independent variable X which can be more than one and the other scalar variable Y dependent on X. In linear regression, relationships between variables are modeled using linear predictor functions and values for unknown variables are quantified using the known variables. Such models are called linear models. Regression starts with inclusion of a large data set and a predictive model can be built by training this data. In the training process, a regression algorithm estimates the value of concerned variable as a function of the predictors for each case and model is developed which can predict the output

value of the target variable for some new input data.

B. Monte Carlo Simulation

Monte Carlo simulation is a computerized mathematical technique that allows people to account for risk in quantitative analysis and decision making. The technique is used by professionals in such widely disparate fields as finance, project management, energy, manufacturing, engineering, research and development, insurance, oil & gas, transportation, and the environment. This simulation performs risk analysis by building models of possible results by substituting a range of values—a probability distribution—for any factor that has inherent uncertainty. It then calculates results over and over, each time using a different set of random values from the probability functions. Monte Carlo simulation provides a number of advantages over deterministic, or "single-point estimate" analysis:

1. Probabilistic Results: Results show not only what could happen, but how likely each outcome is.

2. Graphical Results: Because of the data a Monte Carlo simulation generates, it's easy to create graphs of different outcomes and their chances of occurrence. This is important for communicating findings to other stakeholders.

3. Sensitivity Analysis: With just a few cases, deterministic analysis makes it difficult to see which variables impact the outcome the most. In Monte Carlo simulation, it's easy to see which inputs had the biggest effect on bottom-line results.

4. Scenario Analysis: In deterministic models, it's very difficult to model different combinations of values for different inputs to see the effects of truly different scenarios. Using Monte Carlo simulation, analysts can see exactly which inputs which values had together when certain outcomes occurred. This is invaluable for pursuing further analysis.

5. Correlation of Inputs: In Monte Carlo simulation, it's possible to model interdependent relationships between input variables. It's important for accuracy to represent how, in reality, when some factors go up, others go up or down accordingly.

IV. EXPERIMENTAL SETUP

We will take the experimental values of four companies, Amazon, Apple, Microsoft and Google. The stock market represents these companies by ticker symbols, AMZN, APPL, MFST and GOOGL. The steps are as follows: Basic Analysis of Stock Information: Starting, we handle requesting stock information with pandas, and analyze basic attributes of a stock.

	AAPL	AMZN	GOOGL	MSFT
Date				
2016-10-25	0.005100	-0.003472	-0.008603	-0.000164
2016-10-26	-0.022495	-0.015075	-0.007785	-0.005903

Figure 1 List of stock details

Now that we've seen the DataFrame, we go ahead and plot out the volume and closing price of the AAPL(Apple) stocks.

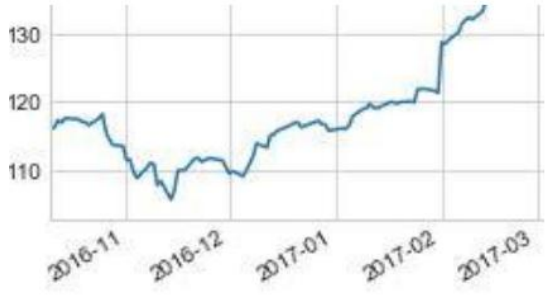

Figure 2 Volume and closing price of APPL

After getting the visualizations for the closing price and the volume traded each day for AAPL stock, we caculate the moving average for the AAPL stock.

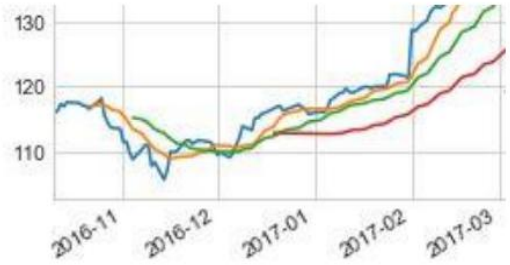

Figure 3 Moving Average (MA) graph for Apple stock

Daily Return Analysis: We will now analyze the risk of the stock. In order to do so, we need to take a closer look at the daily changes of the stock, and not just its absolute value using pandas to retrieve the daily returns for the APPL stock.

Risk Analysis: There are many ways we can quantify risk, one of the most basic ways using the information we've gathered on daily percentage returns is by comparing the expected return with the standard deviation of the daily returns(Risk). Value at risk using the Monte Carlo method:

	AAPL	AMZN	GOOGL	MSFT
Date				
2016-10-17	117.55	812.95	806.84	57.22
2016-10-18	117.47	817.65	821.49	57.66
2016-10-19	117.12	817.69	827.09	57.53

Figure 4 Closing prices for the four companies

Monte Carlo Simulation:

Figure 5 Output of Monte Carlo Simulation, VaR

This basically means for every initial AMZN stock one purchases he/she is putting about $18.13 at risk 99% of the time from our Monte Carlo Simulation.

CONCLUSION

This paper has underlined the utilization of Monte Carlo prediction algorithm to improve the performance predicting prices of stocks in near future. We have investigated the effect of input features on overall quality of algorithm sets. It is verified that Monte Carlo always does not guarantee for improved accuracy. Sometime through the use of it, some low errors can be achieved despite a major reduction of input

data which may result an overall reduction in accuracy of classification model. That can be viewed in case of Markovian chain model implementation on stock exchange dataset in presented research work. This paper can be used as a base line in future that researchers can use to understand the effect of implementing this and other algorithms on highly correlated datasets. It can also help researchers for selecting optimal principal components when applying chain model on a particular classification model having high dimensional input dataset and can also be used when investigating the automatic selection of parameters for techniques, such as the number of PCs, kernel parameters for SVM, etc.

REFERENCES

I. Kumar, D. Ashok, and S. Murugan. "Performance analysis of Indian stock market index using neural network time series model." InPattern Recognition, Informatics and Mobile Engineering (PRIME), 2013 International Conference on, pp. 72- 78. IEEE, 2013.

II. Abhishek, Kumar, Anshul Khairwa, Tej Pratap, and Surya Prakash. "A stock market prediction model using Artificial Neural Network." InComputing Communication & Networking Technologies (ICCCNT), 2012 Third International Conference on, pp. 1-5. IEEE, 2012.

III. Linear Regression- https://en.wikipedia.org/ wiki/Linear_ regression

IV. Vector Regression- http://www.saedsayad.com/support_ vector_machine_reg.htm

V. Proposed system for estimating intrinsic value of stock using Monte Carlo Simulation- http://ieeexplore.ieee.org/ document/8250558/

VI. Shunrong Shen, Haomiao Jiang, and Tongda Zhang. "Stock Market Forecasting Using Machine Learning Algorithms," International Journal of Machine Intelligence, vol. 3, pp. 17-22, 2012.

VII. W. Fenghuaa, X. Jihong, H. Zhifang and G. Xu. "Stock Price Prediction based on SSA and SVM." Proc. 2nd

International Conference on Information Technology and Quantitative Management (ITQM 14), Elsevier, 2014, pp. 645-631, doi: 10.1016/j.procs.2014.05.309.

VIII. Y. Wang and I. C. Choi, "Market Index and Stock Price Direction Prediction using Machine Learning Techniques: An empirical study on the KOSPI and HIS," International Journal of Business Intelligence and Data Mining, vol.1, pp.1-13, September 2013.

IX. R. Choudhry and K. Garg, "A Hybrid Machine Learning System for Stock Market Forecasting," International Scholarly and Scientific Research & Innovation, vol.2, pp.242-245, 2008.

X. M. Karazmodeh, S. Nasiri and S. Majid Hashemi, "Stock Price Forecasting using Support Vector Machines and Improved Particle Swarm Optimization," Journal of Automation and Control Engineering, vol. 1, pp.173-176, June 2013.