

# STOCK MARKET ANALYSIS USING SUPERVISED LEARNING

Dr. M V RATHNAMMA<sup>1</sup>, Dr. V VENKATA RAMANA<sup>2</sup>

<sup>1,2</sup>Professor, Dept. of CSE

<sup>1</sup>K.L.M College of Engineering for Women, Kadapa,

<sup>2</sup>KSRM College of Engineering, Kadapa.

## ABSTRACT

For almost fifty years, academics and analysts have been frustrated by the inherent uncertainty in predicting stock market trends. In addition, our research effort is focused on observing stock price variations for the benefit of people who are eager to engage in the stock market but lack the confidence to do so owing to the danger of volatility in stock. In this study, we propose four prediction algorithms that may be used to forecast the stock market based on past data. K-Nearest Neighbour (KNN), Random Forest (RF), Support Vector Machine (SVM), and Linear Regression are the suggested supervised techniques. Yahoo Finance, Kaggle, Towards Data Science, and NSE India will be mined for their own historical data sets. The outcomes of the various supervised algorithms used in the four separate algorithms. After collecting and analysing data from a variety of companies, we found that the random forest algorithm had the greatest overall performance for making predictions (to the tune of 95.8% accuracy).

## 1. INTRODUCTION

The STOCK MARKET is one of the first venues for the average individual to buy and sell stocks, make investments, and profit from the sale of ownership interests in corporations. If used properly, this technique may be a lucrative investment opportunity. However, the platform's pricing and liquidity are quite volatile, therefore we rely on technology to assist us out. One of these tools that helps us go where we want to go is machine learning. The following three

paragraphs provide a synopsis of the paper's main points:

As is well-known, the stock market is a crucial marketplace that has far-reaching effects on both individuals and nations [2]. The main idea is that businesses will sell pieces of themselves to the public in the form of stocks. They are trying to boost the company's coffers by doing so. The price at which a firm first sells its shares to the public is known as the IPO price. This is the price at which the corporation is willing to sell shares of stock in order to generate capital. After that, the stock is legally the owner's to sell on an exchange like the BSE (Bombay Stock Exchange) at any price he sees fit. Investors and dealers keep selling these shares at whatever price they see fit, while the corporation merely keeps the money gained during the IPO. Shares are constantly being traded from one investor to another in the hopes of generating more profits, which causes the share price to rise after each successful trade. Traders stand to lose money if the stock market price drops because the corporation issues more shares at a lower initial public offering price. The dread of investing in the stock market and the ups and downs of stock prices may both be traced back to this phenomena.

Because lots of minds will attempt to decode the pattern and make an educated estimate, and that's been shown to work a lot better in practise than it does in theory. [5] However, crowd computing is the most accurate way for estimating the stock's real worth in the future. We aim to utilise a computer to mimic such an

example using a more scientific and mathematical method since it is almost certain that crowd computing is a very sluggish activity.

Graph analysis is a technique used in statistics for determining the connection between two or more variables by plotting their values and characteristics against one another. [3][4]. In statistics, this method is known as linear regression and is widely used owing to its ease of use and high level of accuracy. The machine learning version of this approach uses characteristics to educate a classifier, which then makes predictions about the label's value, the veracity of which may be evaluated both during training and testing. You need to choose the proper features and have enough data to train your classifier if you want it to be accurate. Your classifier's efficacy is proportional to the data you feed it and the characteristics you choose.

## 2. LITERATURE SURVEY

The authors of this report, Sneha Kalra et al. in 2019, investigated how stock prices moved in response to the release of materially new business articles. Based on the daily news variation in the social media data, they employed the classifier Naive Bayes to differentiate negative or positive remarks for prediction purposes; blogs data may be included for future study.

Based on their findings, Aditya Menon et al. (2019) conclude that the long short term memory algorithm for predicting the economic information in confluence into the trendy era should be prioritised as the best forecasting algorithm [2].

In 2017, a group of researchers led by Ashish Sharma conducted a review of regression techniques for stock prediction using stock market data and discovered that regression analysis is widely utilised for predicting trends

in the stock market. Future work may benefit from the incorporation of more numerical variables [3].

This year's writers Mu yen chen et al. [4] used the deep learning technique LSTM (long short-term memory) to estimate how much of an effect news stories had on stock prices.

## 3. SYSTEM ANALYSIS

### EXISTING SYSTEM

For your mathematical efforts to bear fruit, you must first settle on a suitable model. The input data must be considered in the process of selecting or developing the model. If the incorrect model is used for the right data, or if the right model is used for the wrong data, you'll end up with a trash model. In order to handle your data, you need to check whether a supportable SVM is available.

It's also a good idea to test out many models concurrently to see which one yields the best results. Furthermore, implementation is the quickest and least time-consuming of the processes, saving us time that might be better spent on other, more crucial, activities.

### DISADVANTAGES

- 1) DECREASED ACCURACY
- 2) POOR PERFORMANCE

### PROPOSED SYSTEM

This information is now ready to be fed into a classifier. The Sklearn library of the Scikit-learn package provides the simplest classifier, therefore we'll be utilising it. This classifier is ideal for our needs and was chosen for its ease of use and complexity reduction. To analyse data and make predictions, linear regression is often used. It relies heavily on the interdependencies among important traits to make predictions about relationships between variables. [9] Supervised machine learning

refers to this kind of forecasting. Labelled data, in which the characteristics are linked to their respective labels, provides the basis for supervised learning. Here, we train the classifier to recognise the patterns that indicate which of 199 possible permutations of features produces a given label. Here, the classifier takes a look at the characteristics, consults its label, and stores the information. A few days later, it recalls the feature combination together with the label, which in this instance is the stock price. Then it learns the structure that the features are using to generate the label. In this way, supervised machine learning achieves its results.

## ADVANTAGES

1. EXTREMELY RELIABLE
2. EXTREMELY EFFICIENT

## ALGORITHM USED

**K-Nearest Neighbours:** Because it does not follow the usual pattern, K closest neighbour is sometimes referred to be a sluggish learner. However, the division is computed by the easy engagement of the bulk of the near spots on the opposite side, balancing the training data circumstances.

**SVM:** The support vector machine that allows for the representation of data sets like space marks divides each segment into as many clearly defined spaces as is possible. According to where space would carry off into, it was predicted that they would fall into a certain group when the new model was created in that gap.

## 4. MODULES

### VIEW USER

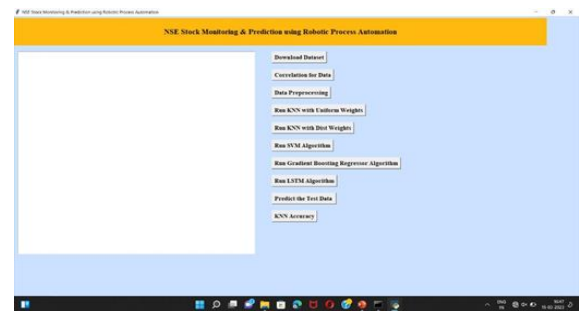
The list of people who have registered may be seen by the administrator.

The administrator may examine the user's information, including username, email, and address. enables the users.

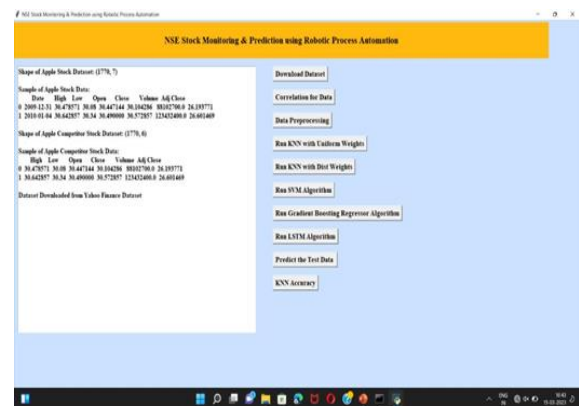
## REMOTE USER

There are now n users online. Before doing any activities, the user should register. Once a user registers, the database will record their information. After successfully registering, he must log in using an authorised username and password.

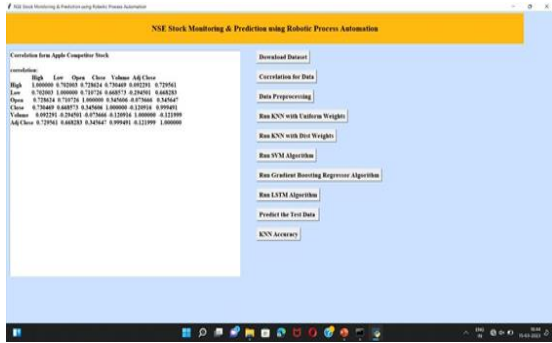
## 5. RESULTS



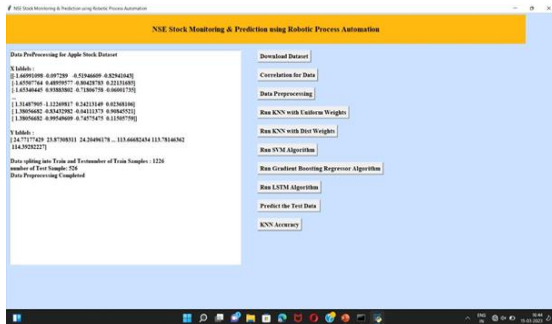
### Stock Trend Prediction Using KNN



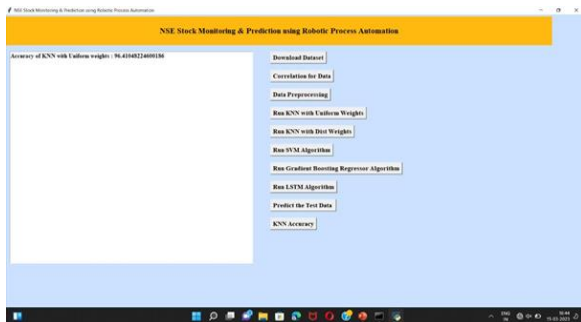
### Download Dataset



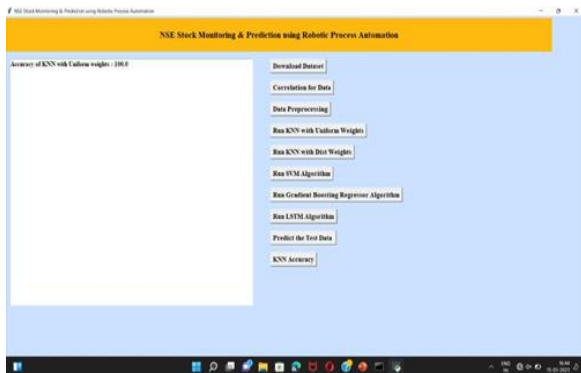
Correlations for Data



Data Pre-processing

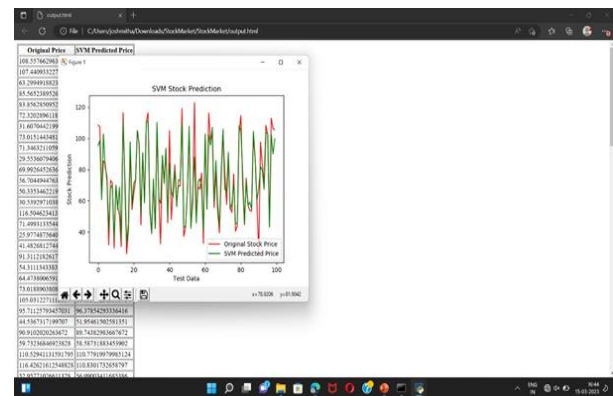
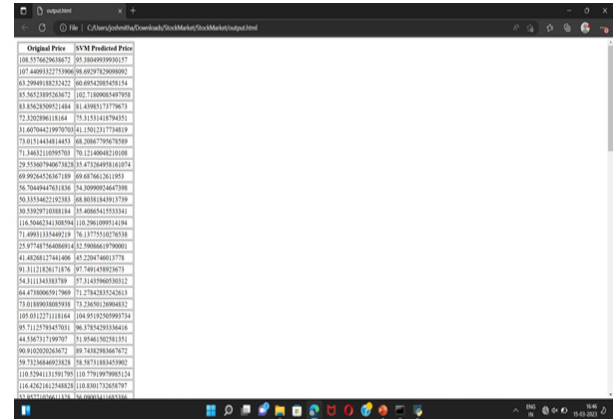


Run KNN with Uniform weight

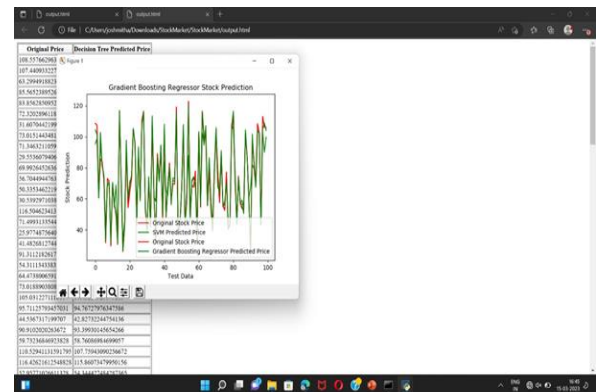


Run KNN with Dist. Weight

With KNN uniform weights, we are receiving 100% accuracy on the screen above. Now, click the "Run SVM Algorithm" button to receive the result shown below.



Run SVM Algorithm



Run Gradient Boosting Regressor Algorithm

The green lines in the gradient boosting algorithm's prediction in the previous screen are





[http://epublications.bond.edu.au/infotech\\_pubs/13/](http://epublications.bond.edu.au/infotech_pubs/13/), 2003, pp. 211–216.

[4] Monic Tirea and Viorel Nehru, Intelligent Stock Market Analysis System - A Fundamental and Macro-economical Analysis Approach, IEEE, 2014.

[5] Kian-Ping Lim, Chee-Wooi Hooy, Kwok-Boon Chang, and Robert Brooks, Foreign investors and stock price efficiency: Thresholds, underlying channels and investor heterogeneity, *The North American Journal of Economics and Finance*, Vol. 36, <http://linkinghub.elsevier.com/retrieve/pii/S1062940815001230>, 2016, pp. 1–28.

[6] Lamartine Almeida Teixeira and Adriano Lorena Inácio de Oliveira, A method for automatic stock trading combining technical analysis and nearest neighbor classification, *Expert Systems with Applications*, <http://linkinghub.elsevier.com/retrieve/pii/S0957417410002149>, 2010, pp. 6885–6890.

[7] Banshidhar Majhi, Hasan Shalabi, and Mowafak Fathi, FLANN Based Forecasting of S&P 500 Index. *Information Technology Journal*, Vol. 4, Issue 3, <http://www.scialert.net/abstract/?doi=itj.2005.289,292>, 2005, pp. 289–292.

[8] Ritanjali Majhi, G. Panda, and G. Sahoo, Development and performance evaluation of FLANN based model for forecasting of stock markets, *Expert Systems with Applications*, Vol. 36, Issue 3, <http://linkinghub.elsevier.com/retrieve/pii/S0957417408005526>, 2009, pp. 6800–6808.