

Stock Market Analysis using Supervised Machine Learning

Gorrela G N Durga Kishore

PG scholar, Department of MCA, DNR college, Bhimavaram, Andhra Pradesh.

B.S.Murthy

(Assistant Professor), Master of Computer Applications, DNR college, Bhimavaram, Andhra Pradesh.

Abstract *One of the most intricate and sophisticated forms of trade is the stock market, also referred to as the stock exchange. Small businesses, brokerage firms, and the banking industry all rely on this one entity to generate income and distribute risks; it's a very intricate arrangement. This article discusses utilising machine learning algorithms to anticipate the future stock price on the exchange using open-source libraries and already-existing algorithms in order to make this unstable business model a little more predictable. We'll see if this straightforward implementation yields respectable outcomes. The result is entirely dependent on math and makes numerous axiomatic assumptions that may or may not be true at the time of prediction.*

Keywords— *Supervised Machine Learning, Stock Market, Linear Regression, Implementation, Fundamental, Data Analysis, Basics.*

I. INTRODUCTION

One of the first ways for a regular individual to trade stocks, make investments, and profit from businesses that sell a piece of themselves on this platform is through the stock market. If used carefully, this strategy demonstrates its potential as an investment scheme. Although this platform's prices and liquidity are highly unpredictable, we use technology to our advantage in this situation. One such tool that assists us in achieving our goals is machine learning. The essential elements of this study are succinctly explained in the following three paragraphs:

As everyone is aware, the stock market is a crucial trading venue that has an impact on everyone both locally and nationally [2]. The fundamental idea is really straightforward: Businesses will list their shares as tiny commodities called stocks on the stock exchange. To raise money for the business, they act in this manner. At this offer price, the corporation will sell its stock and raise capital.

After then, the owner owns the stock and is free to sell it to a buyer at any price on a stock exchange like the Bombay Stock Exchange or BSE.

The company merely gets to keep the money produced during the IPO; traders and buyers continue to sell these shares at their own price. After each profitable transaction, the price of the particular share rises as a result of the continual passing of shares from one party to another in an effort to gain more money. In a word, this phenomenon is what causes people to be afraid of investing in the stock market and what causes stock values to increase and fall.

Now, is it really difficult to forecast the future event on the graph if we try to graph the stock exchange price over a period of time? By simply staring at the graph for a short while, the human brain is quite capable of extending it by a few coordinates. [1] And if we crowd compute, that is, have a bunch of random people try to extend the graph by a set period of time, we'll obtain a pretty good approximation of the answer for a real-world graph.

Since a lot of minds will attempt to decipher the guess and pattern, and since this sort of activity has been shown to be far more successful in practise than it initially appears to be in theory, [5] having stated that, crowdsourcing is the best approach for estimating the stock's genuine value. In statistics, we can identify the dependant and independent variables by looking at the values and characteristics of a problem represented as a graph, and then we can try to establish or find an existing link between them [3][4].

Statistics refers to this method as linear regression, and because of how straightforward and useful it is, it is often utilised. Your classifier's

accuracy directly relates to the attributes you choose and the amount of data you feed it.

II. LITERATURE SURVEY

The stock market, often known as the share market, is one of the most complex and well-developed forms of collaboration. This body is used by small businesses, organisations, and the financial sector to generate income and separation opportunities, which is a very confusing concept. In this paper, a more useful method to predict stock development with greater accuracy will be presented and reviewed. The dataset of stock market costs from the prior year is what this study primarily represents. The dataset was prepared in advance and modified for accurate analysis. From this point forward, the focus of our paper will also be on the rough dataset's data pre-processing. The suggested article also examines concerns with the accuracy of the general attributes as well as using the expectation framework in natural settings. [1]

One of the financial system's most adaptable areas, the stock market, is crucial to the advancement of the economy. Investors can buy and sell stocks, bonds, and other financial products on the stock market. To put it another way, the stock market acts as a barrier-free trading environment for a range of assets and derivatives. Via public offerings, many businesses list their business ventures on the stock market. Long-term investors are currently investing in companies on the stock market to make money. There are three listed stock exchanges in India: the Calcutta Stock Exchange (CSE), the Bombay Stock Exchange (BSE), and the National Stock Exchange (NSE). The Indian stock market is dominated by these three. A statistical measurement of volatility is the dispersion of returns for a certain security or market index. The risk attached to a security is often higher the higher the volatility. [2]

The exploration and modelling of the spatial patterns of residential crowding over space have received relatively little attention, despite the rising body of research on the consequences of residential crowding on the public health and housing market concerns. In the neighbourhoods of Alexandria, Egypt, this research focuses on examining the geographical correlations between residential

density and socio-demographic factors. In a GIS-based platform, global and local geostatistical approaches were used to identify the geographical changes of home crowding variables. The global ordinary least squares (OLS) model makes the assumption that interactions between the explanatory and response variables are uniformly distributed throughout the study region. Hence, it is unable to take into consideration the variability of spatial interactions. [3]

One of the most intricate and sophisticated methods of conducting business is the stock market. This body controls the allocation of income and risk for small assets, brokerage businesses, and the banking industry; it is a very complex model. To help turn this business model from unpredictable to predictable, this document suggests utilising a machine learning algorithm to forecast the future cost of shared resources for exchange using open source libraries and already-existing techniques. We'll see if this straightforward application yields respectable outcomes. The outcome is purely based on numbers and presupposes a huge number of assumptions that may or may not hold true at the time of prediction in the real world. [4]

One of the most challenging and complicated ways to conduct business is the stock market. Small businesses, trading companies, and the banking industry all depend on the same body to distribute revenue and risk; a more complicated model is required. Yet, this book suggests employing pre-existing algorithms along with open source libraries to anticipate future prices of shared resource exchange in order to help develop this business structure from uncertainty. We'll test whether this application can produce respectable results with ease. Completely dependent on numbers, with a number of axioms that might or might not be true at the time of the forecast. [6]

The stock market offers the greatest potential for significant returns, making it the ideal place to invest your hard-earned money. But, even with the technology at hand, it is a risky deed because it is difficult to comprehend unexpected market developments and evaluate data correctly. We suggest "Prediction of stock market deviation using ARIMA algorithm," a real-time risk prediction tool

that takes market interests into account, in order to simplify the investment process and improve awareness. To understand historical data, it is based on the ARIMA (Auto Regressive Integrated Moving Average) method, a parametric time series analysis technique. [7]

III. EXISTING METHOD

As we all know, the stock market is a significant trading venue that has an impact on everyone, both locally and nationally. The fundamental idea is simple to comprehend. Businesses will offer their stock, a small commodity known as stocks, on the stock market. To raise money for the business, they act in this manner. People, on the other hand, occasionally lose money and occasionally make more money in the stock market since we are unable to determine which products will sell the best in the future.

IV. PROPOSED METHOD

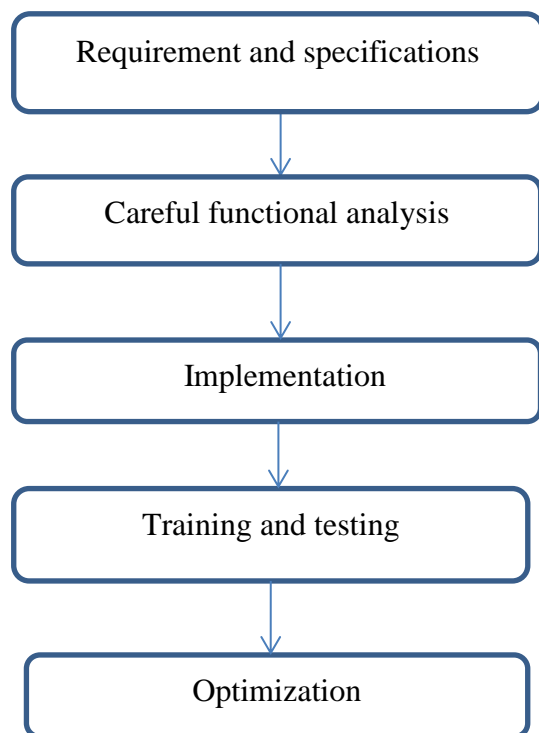


Fig.1. Flow chart for proposed system

A. Requirements and Specification

As a first step, you must thoroughly understand the precise problem requirements as well as the machine and throughput specifications. Analyse the situation thoroughly, look into the background, learn everything there is to know about it, then decide what you really want and make it your aim.

B. Careful function Analysis

Because they directly affect the prediction model, the features that are derived from the data must be extracted with extreme caution. All of them must be directly related to the labels.

C. Implementation

In order for your math to yield results, you must choose a suitable model. The input data must be considered when choosing or designing the model. A bad model built or picked for inappropriate data, or vice versa, will result in a garbage model that is entirely useless. To handle your data, you must look for compatible SVMs or other available techniques.

D. Training & Testing

The process of training a model is relatively simple. The only thing you need to check is that the data is plentiful, consistent, and coherent. A larger training data set results in a classifier that is more powerful and accurate, which eventually improves overall accuracy. Moreover, testing is a relatively simple procedure. Be certain that the size of your test data is at least 20% that of your training data.

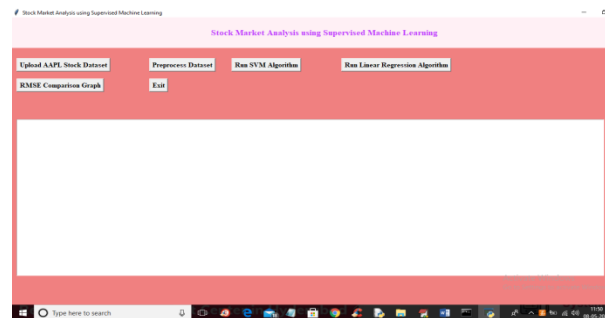
E. Optimization

We must constantly optimise because it is nearly impossible to construct a versatile classifier all at once. There is always room for advancement. Always bear in mind the fundamental specifications and accepted practises when optimising.

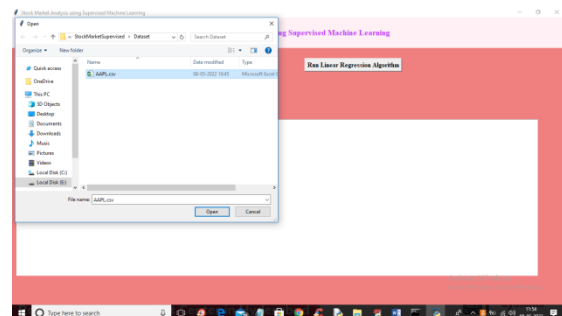
V. RESULT

To implement this project we have used AAPL dataset.

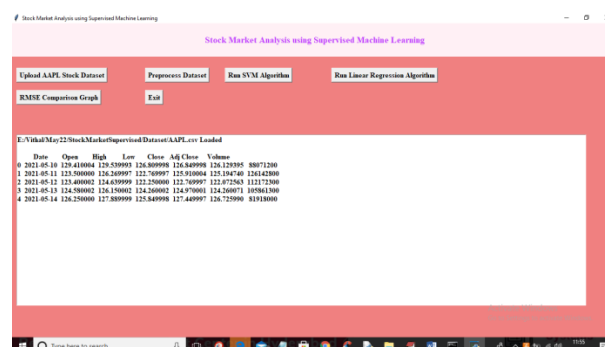
To run project double click on 'run.bat' file to get below screen



In above screen click on 'Upload AAPL Stock Dataset' button to upload dataset

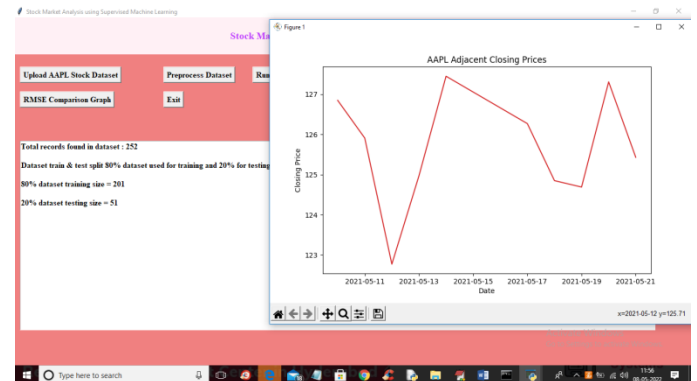


In above screen selecting and uploading 'AAPL.csv' dataset file and then click on 'Open' button to load dataset and to get below output

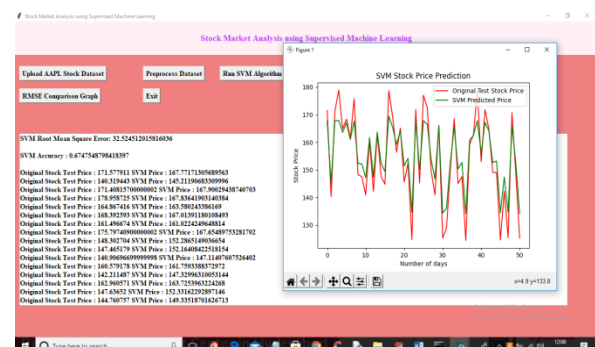


In above screen dataset loaded and now click on 'Preprocess Dataset' button to normalize dataset values and then split dataset into train and test part

for training and testing machine learning algorithms

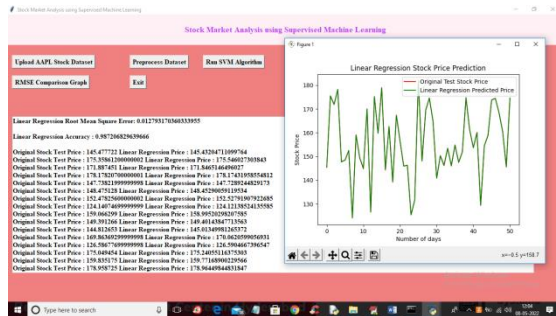


In above screen in text area we can see dataset contains 252 total records and using 201 records for training and 52 records for testing and in graph x-axis represents DATE and y-axis represents closing values on those dates and now close above graph and then click on 'Run SVM Algorithm' button to train SVM on training data and then test on test data and get below output

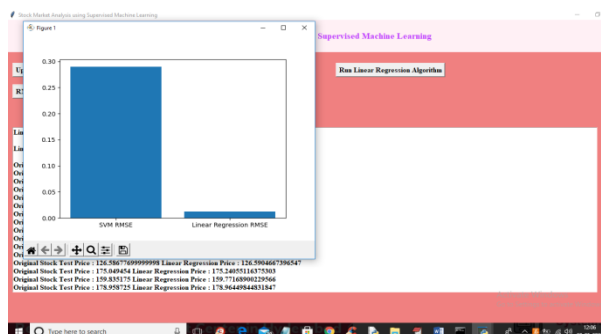


In above screen we got SVM RMSE as 32 and accuracy as 67 and then we can see original stock test PRICES and SVM predicted TEST prices and you can see difference between original and predicted prices and in graph x-axis represents number of days for future and y-axis represents stock prices and green line represents PREDICTED prices and red line is the original prices and SVM is not good in prediction so its accuracy is less and in graph we can see huge gap between predicted and

original prices. Now close above graph and then click on 'Run Linear Regression Algorithm' button to get below output



In above screen with Linear regression we got RMSE as 0.012 and accuracy as 98% and in graph we can see both lines are fully overlapped so there is not much difference between original test and predicted prices so Linear regression is best in stock prices prediction and in text area you can see both original price and linear regression predicted price and you won't find much difference between them. Now close above graph and then click on 'RMSE Comparison Graph' button to get below output



In above graph x-axis represents algorithm names and y-axis represents RMSE values and in both algorithms Linear regression got less RMSE error so linear regression is best.

VI. CONCLUSION

As we've seen thus far, machine learning is a very powerful technique with many practical uses. We have already seen how strongly machine learning is based on data. Realising that data is incredibly valuable and that, despite how easy it may seem, data analysis is a challenging task is crucial. Deep learning and neural networks, which are extensions of machine learning, are largely based on the same core idea. This study provides a thorough grasp of how to use machine learning. There are a variety of strategies, tactics, and techniques available to us that can be used to handle and address various problems in all imaginable situations. This article focuses solely on supervised machine learning and makes an effort to lay out the fundamentals of this complex process.

REFERENCES

- [1] Andrew McCallum, Kamal Nigam, Jason Rennie, Kristie Seymore "A Machine learning approach to Building domain-specific Search engine", IJCAI, 1999 – Citeseer
- [2] Yadav, Sameer. (2017). STOCK MARKET VOLATILITY - A STUDY OF INDIAN STOCK MARKET. Global Journal for Research Analysis. 6. 629-632.
- [3] Montgomery, D.C., Peck, E.A. and Vining, G.G., 2012. Introduction to linear regression analysis (Vol. 821). John Wiley & Sons.
- [4] Draper, N.R.; Smith, H. (1998). Applied Regression Analysis (3rd ed.). John Wiley. ISBN 0-471-17082-8.
- [5] Robert S. Pindyck and Daniel L. Rubinfeld (1998, 4th ed.). Econometric Models and Economic Forecasts
- [6] "Linear Regression", 1997-1998, Yale University <http://www.stat.yale.edu/Courses/1997-98/101/linreg.htm>
- [7] Agarwal (July 14, 2017). "Introduction to the Stock Market". Intelligent Economist. Retrieved December 18, 2017.

