GALGOTIAS UNIVERSITY

(Established under Galgotias University Uttar Pradesh Act No. 14 of 2011)

STOCK MARKET ANALYSIS AND PREDICTION

A Report for the Evaluation 3 of Project 2

Submitted by

ASHISH SINGH

(1613101191)

in partial fulfilment for the award of the degree

of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING

Under the Supervision of

PRATYUSH KUMAR DEKA

(Assistant Professor)

APRIL / MAY- 2020

GALGOTIAS UNIVERSITY

(Established under Galgotias University Uttar Pradesh Act No. 14 of 2011)

SCHOOL OF COMPUTING AND SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

Certified that this project report "**STOCK MARKET ANALYSIS AND PREDICTION**" is the Bonafide work of "**ASHISH SINGH (1613101191**)" who carried out the project work under my supervision.

Signature of Head

Dr. Munish Sabbharwal

PhD(Management),PhD(CS)

Professor & Dean,

School of computer Science

& Engineering

Signature of Supervisor

Jayakumar Vidhyashankar

Professor,

School of computer Science

&Engineering

TABLE OF CONTENTS

Abstract -

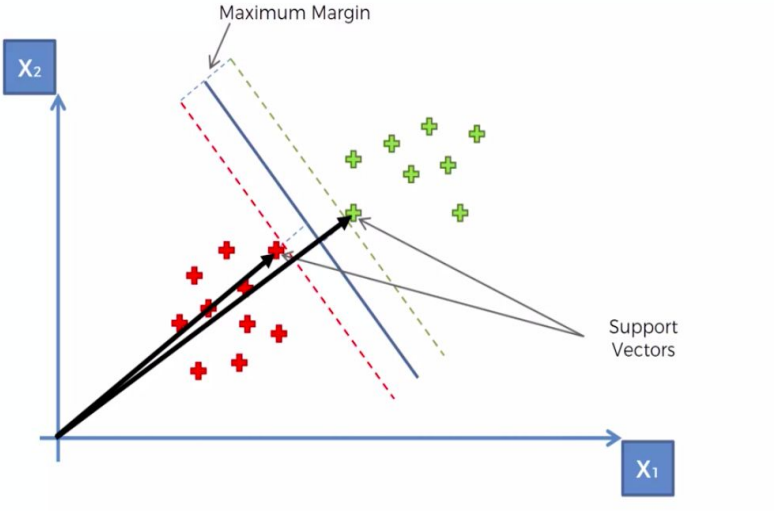
Time series forecasting has been widely used to determine the future prices of stock, and the analysis and modelling of finance time series importantly guide investors' decisions and trades. This work proposes an intelligent time series prediction system that uses sliding-window optimization for the purpose of predicting the stock prices. The system has a graphical user interface and functions as a stand-alone application. The proposed model is a promising predictive technique for highly non-linear time series, whose patterns are difficult to capture by traditional models. Within the finance world stock trading is one among the foremost important activities. stock exchange predictions an act of trying to see the longer-term value of a stock other financial instrument traded on a financial exchange. This paper is self-explanatory for the prediction of any stock using the techniques of Machine Learning. The technical and fundamental or the statistical analysis is employed by the foremost of the stockbrokers while making the stock predictions. The artificial language is employed to predict the stock exchange using machine learning is Python. during this paper we propose a Machine Learning (ML) approach which will be trained from the available stocks data and gain intelligence then uses the acquired knowledge for an accurate prediction. during this context this study uses a machine learning technique called Support Vector Machine (SVM) to predict stock prices for the big and tiny capitalizations and within the three different markets, employing prices with both daily and up-to-the-minute frequencies.

Key Words:-

(A) Stock Exchange- A stock exchange does not own shares. Instead, it acts as a market where stock buyers connect with stock sellers. Stocks are often traded on one or more of several exchanges like the NY stock market (NYSE) or the Nasdaq. Although you'll presumably trade stocks through a broker, it's important to know the connection between exchanges and corporations, and therefore the ways during which the wants of different exchanges protect investors.

(B) Machine Learning- A stock exchange does not own shares. Instead, it acts as a market where stock buyers connect with stock sellers. Stocks are often traded on one or more of several exchanges like the ny stock market (NYSE) or the Nasdaq. Although you'll presumably trade stocks through a broker, it's important to know the connection between exchanges and corporations, and therefore the ways during which the wants of different exchanges protect investors.

(C) Support Vector Machine- Support Vector Machine (SVM) may be a supervised machine learning algorithm capable of performing classification, regression and even outlier detection. The linear SVM classifier works by drawing a line between two classes. All the info points that fall on one side of the road are going to be labeled together class and every one the points that fall on the opposite side will be labeled as the second. Sounds simple enough, but there's an infinite amount of lines to settle on from. How can we know which line will do the simplest job of classifying the data? This is where the LSVM algorithm comes in to play. The LSVM algorithm will select a line that not only separates the 2 classes but stays as distant from the closest samples as possible. In fact, the "support vector" in "support vector machine" refers to 2 position vectors drawn from the origin to the points which dictate the choice boundary.



1. **INTRODUCTION: -** Basically, quantitative traders with tons of cash from stock markets buy stocks derivatives and equities at an inexpensive price and afterward selling them at high price. The trend during a stock exchange prediction isn't a replacement thing and yet this issue is kept being discussed by various organizations. There are two types to research stocks which investors perform before investing during a stock, first is that the fundamental analysis, during this analysis investors check out the intrinsic value of stocks, and performance of the industry, economy, political climate etc. to decide that whether to invest or not. On the opposite hand, the technical analysis it's an evolution of stocks by the means of studying the statistics generated by market activity, like past prices and volumes.

In the recent years, increasing prominence of machine learning in various industries have enlightened many traders to apply machine learning techniques to the field, and some of them have produced quite promising results.

This paper will develop a financial data predictor program during which there'll be a dataset storing all historical stock prices and data are going to be treated as training sets for the program. The main purpose of the prediction is to scale back uncertainty associated to investment deciding.

Stock Market follows the stochastic process, which means that the simplest prediction you'll have about tomorrow's value is today's value. Indisputably, the forecasting stock indices is extremely difficult due to the market volatility that needs accurate forecast model. The stock market indices are highly fluctuating and it effects the investor's belief. Stock prices are considered to be a very dynamic and susceptible to quick changes because of underlying nature of the financial domain and in part because of the mix of a known parameters (Previous day's closing price, P/E ratio etc.) and the unknown factors (like Election Results, Rumors etc.). There have been numerous attempts to predict stock price with Machine Learning. The focus of each research projects varies a lot in three ways.

(1) The targeting price change can be near-term (less than a minute), short-term (tomorrow to a few days later), and a long-term (months later),

(2) The set of stocks can be in limited to less than 10 particular stock, to stocks in particular industry, to generally all stocks.

(3) The predictors used can range from a global news and economy trend, to particular characteristics of the company, to purely time series data of the stock price.

Prediction of stock trend has long been an intriguing topic and is extensively studied by researchers from different fields. Machine learning, a well-established algorithm in a wide range of applications, has been extensively studied for its potentials in prediction of financial markets. Popular algorithms, including support vector machine (SVM) and reinforcement learning, have been reported to be quite effective in tracing the stock market and help maximizing the profit of stock option purchase while keep the risk low. However, in many of these literatures, 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. Despite its success in some occasions, sentiment analysis may fail when some of the people are biased, or positive opinions follow past good performance instead of suggesting promising future markets.

One of the most prevailing and exciting supervised learning models with associated learning algorithms that analyze data and recognize patterns is Support Vector Machines (SVMs). It is used for solving both regression and classification problems. However, it is mostly used in solving classification problems. SVMs were first introduced by B.E. Boser et al. in 1992 and has become popular due to success in handwritten digit recognition in 1994. Before the emergence of Boosting Algorithms, for example, XG Boost and AdaBoost, SVMs had been commonly used. If you want to have a consolidated foundation of Machine Learning algorithms, you should definitely

have it in your arsenal. The algorithm of SVMs is powerful, but the concepts behind are not as complicated as you think.

The probable stock exchange prediction target is often the longer-term stock price or the volatility of the costs or market trend. In the prediction there are two types like dummy and a true time prediction which is employed available market prediction system. In Dummy prediction they need define some set of rules and predict the longer-term price of shares by calculating the typical price. In the real time prediction compulsory used internet and saw current price of shares of the corporate.

Computational advances have led to introduction of machine learning techniques for the predictive systems in financial markets. In this paper we are using a Machine Learning technique i.e., Support Vector Machine (SVM) in order to predict the stock market and we are using Python language for programming.

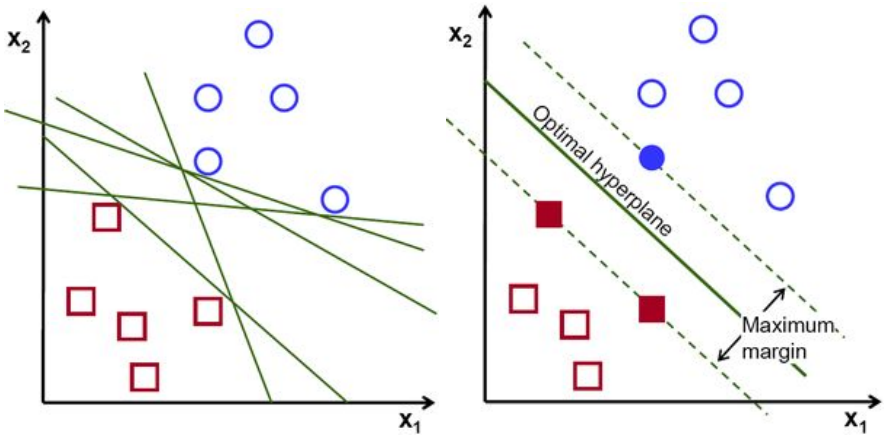
2-Methodology: -

In this project the prediction of stock market is done by the Support Vector Machine (SVM) and Radial Basis Function (RBF).

(A) Support Vector Machine: - A Support Vector Machine (SVM) may be a discriminative classifier that formally defined by the separating hyperplane. In other words, the given labeled training data (supervised learning), the algorithm outputs the optimal hyperplane which categorizes new examples. In the two-dimensional space this hyperplane is a line dividing a plane into two parts where in each class lay in either side.

Support Vector Machine (SVM) is considered to be as one of the most suitable algorithms available for the time series prediction. The supervised algorithm can be used in both, regression and classification. The SVM involves in plotting of data as point in the space of n dimensions. The objective of the support vector machine algorithm is to find a hyperplane in an N-dimensional space (N — the number of features) that distinctly classifies the data points.

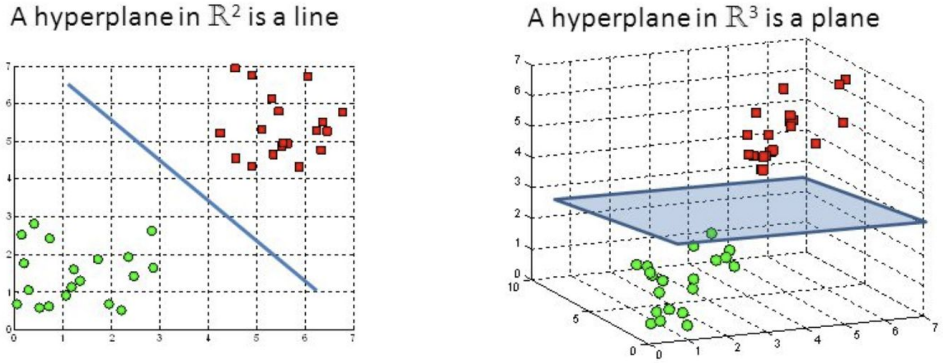
These dimensions are the attributes that are plotted on particular co-ordinates. SVM algorithm draws a boundary over the data set called as the hyper-plane, which separates the data into two classes as shown in the Fig



(Possible hyperplanes)

To separate the two classes of data points, there are many possible hyperplanes that could be chosen. Our objective is to find a plane that has the maximum margin, i.e the maximum distance between data points of both classes. Maximizing the margin distance provides some reinforcement so that future data points can be classified with more confidence.

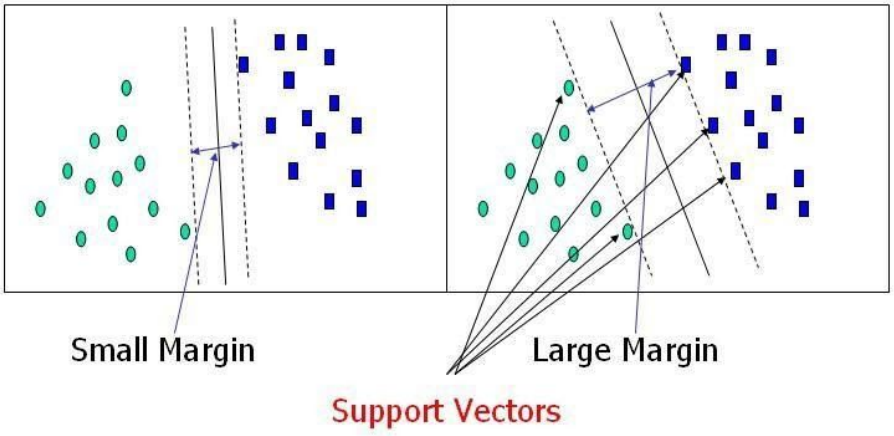
Hyperplanes and Support Vectors: -



(Hyperplanes in 2D and 3D feature space)

hyperplanes are decision boundaries that help classify the info points. Data points falling on either side of the hyperplane are often attributed to different classes. Also, the dimension of the hyperplane depends upon the amount of features.

If the amount of input features is 2, then the hyperplane is simply a line. If the amount of input features is 3, then the hyperplane becomes a two-dimensional plane. It becomes difficult to imagine when the amount of features exceeds 3.



(Support Vectors)

(B) Radial Basis Function (RBF): -In the machine learning, the radial basis perform kernel, or RBF kernel, could be a widespread kernel perform utilized in the assorted kernelized learning algorithms. specifically, it's most ordinarily utilized in support vector machine classification.
A radial basis perform is that the real-valued perform whose price depends solely on the space from the origin, thus that; or as an alternative on the space from another purpose, called a center, so that. Any function which satisfies the property is a radial function.

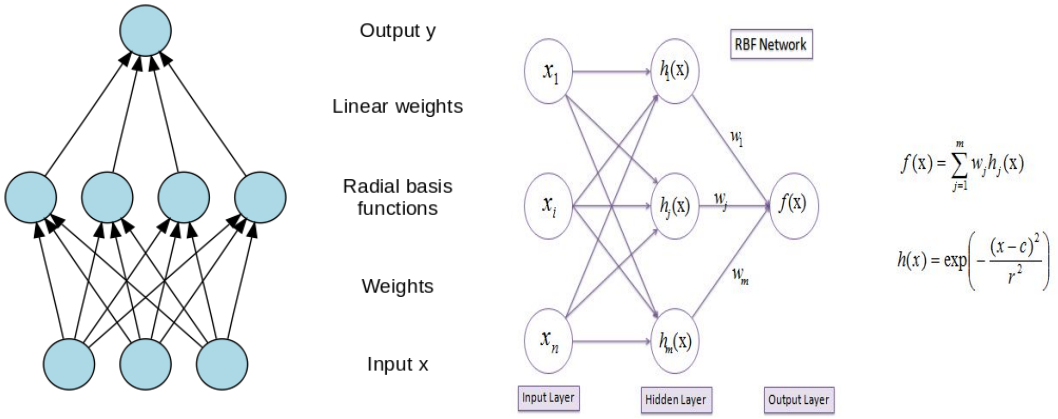
RBF = Local Response Function

The RBF Kernel is nothing more than a low-band pass filter, which is well known in Signal Processing as a tool to smooth images. RBF Kernel acts as the prior that selects out smooth solutions.

The Radial basis function kernel, is also called as the RBF kernel, or Gaussian kernel, is a kernel that is in the form of a radial basis function (more specifically, a Gaussian function). The RBF kernel is defined as

KRBF (x, x') = exp [-y ||x -x'||2]

Where y is the parameter that sets "spread" of the kernel

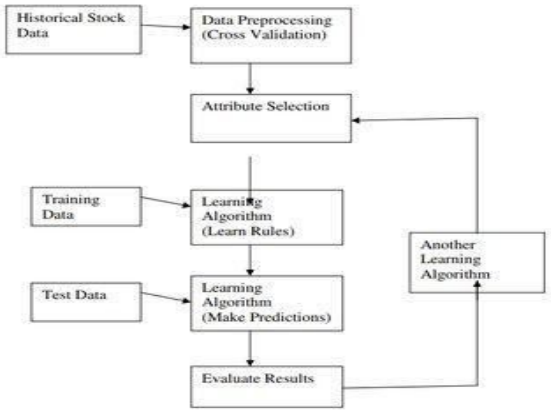


Radial Basis Function Networks (RBF)

The RBF units provide a new basis set for synthesizing the output function. The radial basis functions are not orthogonal and are overcomplete.

(C) The Learning Environment: -

The Weka and the YALE Data Mining Environments were used for carrying out the experiments. The general setup used is as follows:



(Learning Environment)

(3) Model Creation and Evaluation Methods

In this paper we focus on predicting the Stock Market using Machine Learning model i.e., Support Vector Machine (SVM) by RBF kernel.

(A) Feature Selection

In this project we use four features to predict stock price direction – price volatility, price momentum, sector volatility, and sector momentum. More details are provided in Table 1, styled in the form used by Kim [4].

Feature Name	**Description**	**Formula**
σ s	Stock price volatility. This is an average over the past n days of percent change in the given stock's price per day.	$\left\lvert \frac{\sum_{i=t-n+1}^{t} \frac{C_i - C_{i-1}}{C_{i-1}}}{n} \right\rvert$
Stock Momentum	This is an average of the given stock's momentum over the past n days. Each day is labeled 1 if closing price that day is higher than the day before, and −1 if the price is lower than the day before.	$\frac{\sum_{i=t-n+1}^{t} y}{n}$
σ i	Index volatility. This is an average over the past n days of percent change in the index's price per day.	$\frac{\sum_{i=t-n+1}^{t} \frac{I_i - I_{i-1}}{I_{i-1}}}{n}$
Index Momentum	This is an average of the index's momentum over the past n days. Each day it is labeled 1 if closing price that day is higher than the day before, and −1 if the price is lower than the day before	$\frac{\sum_{i=t-n+1}^{t} d}{n}$

Table 1: Features used in SVM

Step 1: This step is very important for the transfer information from net. we tend to square measure predicting the money value of any stock. in order that the share worth up to the point in time square measure transfer from the location

Step 2: within the next step the info worth of any stock that may be regenerate into the CSV file (Comma Separate Value) in order that it'll simply load into the rule.

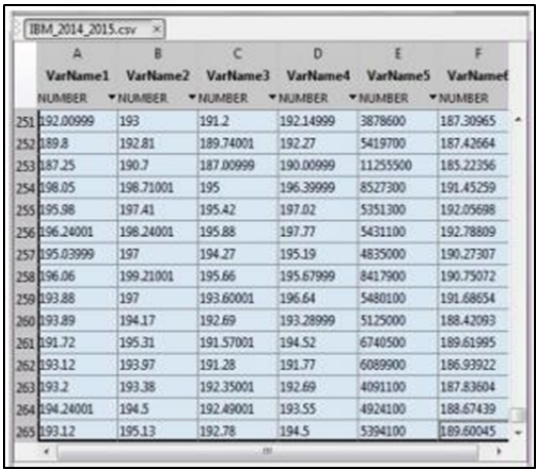
Step 3: within the next step that during which within which GUI is open and we tend to once we after we} click on the SVM button it'll show the window from which we choose the stock dataset worth file.

Step 4: when choosing the stock dataset file from the folder it'll show graph Stock before mapping and stock when mapping.

Step 5: future step rule calculated the log2c and log2g worth for minimizing error. So, it'll predict the graph for the dataset worth expeditiously.

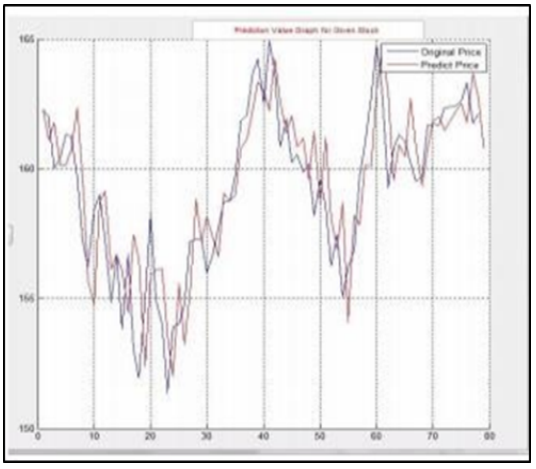
Step 6: In final step rule show expected the anticipated the expected worth graph of choose stock that shows the first worth and predicted worth of the stock.

(4) RESULTS: -



IBM_2014_2015.csv

	A	B	C	D	E	F
	VarName1	VarName2	VarName3	VarName4	VarName5	VarNamet
	NUMBER	NUMBER	NUMBER	NUMBER	NUMBER	NUMBER
251	192.00999	193	191.2	192.14999	3878600	187.30965
252	189.8	192.81	189.74001	192.27	5419700	187.42664
253	187.25	190.7	187.00999	190.00999	11255500	185.22356
254	198.05	198.71001	195	196.39999	8527300	191.45259
255	195.98	197.41	195.42	197.02	5351300	192.05698
256	196.24001	198.24001	195.88	197.77	5431100	192.78809
257	195.03999	197	194.27	195.19	4835000	190.27307
258	196.06	199.21001	195.66	195.67999	8417900	190.75072
259	193.88	197	193.60001	196.64	5480100	191.68654
260	193.89	194.17	192.69	193.28999	5125000	188.42093
261	191.72	195.31	191.57001	194.52	6740500	189.61995
262	193.12	193.97	191.28	191.77	6089900	186.93922
263	193.2	193.38	192.35001	192.69	4091100	187.83604
264	194.24001	194.5	192.49001	193.55	4924100	188.67439
265	193.12	195.13	192.78	194.5	5394100	189.60045



(Stock dataset for IBM Inc. in CSV file) : (Predicted Output by SVM for IBM Inc.)

(5) CONCLUSION: -

within the project, we have a tendency to projected the utilization of the information collected from completely different world monetary markets with machine learning algorithms so as to predict the stock market index movements. SVM formula works on the massive dataset price that is collected from completely different world monetary markets. Also, SVM doesn't provides a downside of over fitting. numerous machine learning primarily based models ar projected for predicting the daily trend of Market stocks. Numerical results counsel the high potency. the sensible commercialism models designed upon our well-trained predictor. The model generates higher profit compared to the chosen benchmarks.

(6) REFERENCES: -

[1] Zhen Hu, Jibe Zhu, and Ken Tse "Stocks Market Prediction Using Support Vector Machine", 6th International Conference on Information Management, Innovation Management and Industrial Engineering, 2013.M.

[2] Wei Huang, Yoshiteru Nakamori, Shou-Yang Wang, "Forecasting stock market movement direction with support vector machine", Computers & Operations Research, Volume 32, Issue 10, October 2005, Pages 2513–2522.

[3] N. Ancona, Classification Properties of Support Vector Machines for Regression, Technical Report, RIIESI/CNRNr. 02/99.

[4] K. jae Kim, "Financial time series forecasting using support vector machines," Neurocomputing, vol. 55, 2003.

[5] Debashish Das and Mohammad shorif uddin data mining and neural network techniques in stock market prediction: a methodological review,