STOCK MARKET PREDICTION USING MACHINE LEARNING: AN APPROACH

Nargis I khan
Student, Department of Computer Science and Engineering
Rajiv Gandhi College of Engineering Research & Technology, Chandrapur, India, 442401

Prof. Nitin Janwe
Associate Professor, Department of Computer Science and Engineering
Rajiv Gandhi College of Engineering Research & Technology, Chandrapur, India, 442401

Abstract: Prediction of stock market is the act of trying to investigate the future value of a company stock or other financial instrument traded on a financial exchange. The successful forecasting of a stock's future price will maximize investor's gains. This paper proposes a machine learning model to predict financial stock market prices. The proposed algorithm integrates calculations of technical parameters and linear support vector machine (SVM). Support Vector Machine (SVM) is comparatively new learning algorithm that has the desirable characteristics to control the decision function and the use of the kernel method. Proposed model is based on the study of stocks historical data and technical parameters. The proposed model was applied and evaluated using Indian stock (NSE) datasets. Firstly, two leading technical parameters that are Moving average convergence/Divergence (MACD) and Stochastic oscillator are calculated from the raw datasets. Secondly, Support Vector Machine is used in analyzing and predicting the stock performance. It is expected that Our Experimental results will suggest that SVM is a powerful predictive tool for stock predictions in the financial market.

Keywords*: - *Stock market, machine learning; SVM;

1. INTRODUCTION

Predicting the future has always been an adventurous and attractive task for the probing individuals. This kind of prediction becomes more fascinating when it involves money and risk like predicting stock market. Goal behind making any financial investment is to achieve above average return for invested money while maintaining certain level of involved risks but as the stock market is a very complex, volatile and non-linear dynamical system, stock market prediction has become a tough challenge for researchers and investors. Lot of Research has been done on stock market prediction by researchers of different fields including the business and computer science. Researchers have tried different approaches for market prediction including different techniques and algorithm and different combination of attributes. The attributes that makes a prediction model depends upon the factor on which market performance can depend. Different methods have been developed to forecast the behavior of stock market prices based on historical data. As per recent survey Stock prices prediction with conventional statistical methods has proven to be less effective because of non-linear characteristics of financial time series. On the contrary, machine learning-based methods, such as Artificial Neural Networks (ANNs), support vector machine (SVM) and data mining system offer useful tools for forecasting noise environments like stock market.

The reminder of the paper is organized as follows, in section 1 we discussed about the stock market. In section 2 we discussed about Machine learning algorithms . In section 3 we briefly examine the previous work. In section 4 we discussed the proposed model. In section 5 we analyze the expected outcome. Finally, in section 6 we summarize our conclusion.

1.1 STOCK MARKET:

Stock market is an organized and regulated financial market where securities such as bonds, notes and shares are bought and sold at price governed by the forces of demand and supply. Stock market act as, market where Corporations, governments, municipalities and other incorporated bodies can raise capital by channeling saving of the investors into Productive ventures. Also, market where investor can sell their securities to other investors for cash, thus reducing the risk of investment and maintaining liquidity in the system.

Stock is a share in the ownership of the company that means you own a part of company. Stocks are partial ownership of businesses, which can be traded in the stock market. If company ownership is divides in 100 parts, and we are the investor purchase one part which is equal to 1 share then we own 1% of that company. Stock exchanges uses a trading system which is order driven automated matching system. Stocks prices are defined on the basis of at any time how many buyers and sellers For the same stock in the market. If numbers of buyers are more than sellers then stock prices become high and if number of sellers higher than buyers, then stock prices become low. If order does not find the match, then it remains in the system and waiting for the fresh orders or updating of previous orders which are already present. Previously stock market was known as stock exchange denotes a place where actual buying

and selling takes place. With computerized trading and electronic communication networks like NASDAQ and BATs, manual involvement has been reduced. Automated trading platforms uses computer algorithms which gives high frequency trading thus forecasting model requires a robust technique which gives accurate prediction of prices so that profit can be maximized.

2. MACHINE LEARNING ALGORITHMS

This section contains Introduction to machine learning and work to evaluation of the machine learning techniques. Some of the algorithms we have considered are as follows:
Machine learning is a concept that provides systems the ability to learn automatically and improve from experience without being explicitly programmed. Machine learning concepts focuses on the development of computer programs that can access data and use it learn for themselves. The learning operation starts with observing the data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the sample data that we provide. The primary goal is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly. Machine learning algorithms are often categorized as unsupervised and supervised.

A) Unsupervised learning algorithm

The training of machine in this learning technique is done by using information which is neither classified nor labeled and allowing the algorithm to work on that information without guidance. The ultimate goal of machine is to group unsorted information according to similarities, patterns and differences without any prior training of data. Different from supervised learning, no teacher is provided that means no training will be given to the machine. That's why machine is restricted to find the hidden structure in unlabeled data by our-self. Unsupervised learning classified into two categories of algorithms:

1. Clustering: A clustering problem is where we want to discover the inherent groupings in the data, such as grouping customers by purchasing behavior.

2. Association: Association learning problems where we want to find rules that describe big portions of your data, such as people that buy X also tend to buy Y.

B) Supervised learning algorithm

In Supervised learning, as the name indicates there is a supervisor as teacher. Learning in which we train the machine using data which is well labeled that means some data is already tagged with correct answer is known as Supervised learning. Again, machine is supplied with new set of examples (data) so that supervised learning algorithm analyses the training data (set of training examples) and produces a correct outcome from labeled data. Supervised learning algorithms increases consistently with the data. It is a type of inductive learning. Supervised learning classified into two categories of algorithms:

1. Classification: A classification problem is where the output variable is a category, such as "Red" or "blue" or "disease" and "no disease".

2. Regression: A regression problem is where the output variable is a real value, such as "dollars" or "weight".

2.2. Support vector machine (SVM)

Vapnik and his co-workers in 1963 developed Support vector machine (SVM). SVM is a supervised machine learning algorithm, generally used for regression and classification analysis. It is also known as probabilistic binary linear classifier. It can also perform non-linear classification with the help of different types of kernels. SVM can be used to resolve many real-world problems such as text and hypertext categorization, classification of images, permutation test, weather forecasting, sales forecasting etc. SVM is widely recognized due to its remarkable advantages such as effective in high dimensional spaces, memory well organized, flexible as uses different kernel functions for different decision functions. Kernel selection is an important task in support vector machines.

SVM appears from statistical learning theory; the goal is to solve only the problem of interest without solving a more difficult problem as an intermediate step. The foundation of SVMs is on the structural risk minimization principle, closely related to regularization theory. This principle incorporates capacity control to prevent over- fitting and thus is a partial solution to the bias-variance trade-off dilemma. Two important elements in the implementation of SVM are the techniques of mathematical programming and kernel functions. The parameters are found by solving a quadratic programming problem with linear equality and inequality constraints; rather than by solving a non- convex, unconstrained optimization problem.SVM algorithm developed by Vapnik are based on statistical learning theory. SVM can be used for both classification and regression task. In classification case we try to find

an optimal hyper plane that separates two classes. In order to find an optimal hyper plane, we need to minimize the norm of the vector w, which defines the separating hyper plane. This is same as maximizing the margin between two classes.

In SVM , we plot each data as a point in n-dimensional space (where n is number of features we have) with the value of each feature being the value of a particular coordinate. second, we perform classification by finding the hyper-plane that classifies the two classes very well. Below figure shows the 2-dimensional case where the data points are linearly separable.

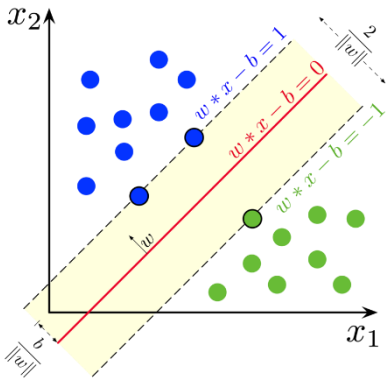

Figure 2: Support Vector Machine

3. LITERATURE SURVEY

In this section we will review SVM for forecasting stock prices. In recent survey it has been observed that including indicators as inputs along with dataset in the system gives better prediction results. This section will present some of the works that has been carried out by researchers in adopting and applying SVM in financial stock market prediction.

Support vector machine can be applied to forecast the stock exchange by using some company specific parameters such as price per earnings ratio of stock, net income, net revenue diluted earnings per share etc. SVM is acutely explicit kind of learning algorithm indicated as the capacity control of decision function and usage of kernel function. The maximal-margin classifier shows how SVM works. SVM can be used to learn polynomial, multi-layer perceptron (MLP) and Radial basis function (RBF) classifiers. SVM works upon the principle of structural risk minimization that leads to the prevention of over-fitting problem. SVM can be used to solve numerous real world problems such as text and hyper text categorization, classification of images, permutation test, weather forecasting, sales forecasting etc

In [1] Authors has used Support vector Machine for predicting stock market. They present a theoretical and empirical framework to apply the Support Vector Machines strategy to predict the stock market. Four Parameters are company-specific And Six macroeconomic factors that may influence the stock trend has selected for further stock multivariate analysis. They applied Support Vector Machine for analyzing the relationship of these factors and predicting the stock performance. According to experimental results SVM is a powerful predictive tool for stock predictions in the financial market. The base paper which is used by authors for demonstrating the result is [2] Where investigation of predictability of financial movement direction with SVM by forecasting the weekly movement with those of linear discriminate analysis, quadratic discriminate analysis and Elam Back propagation Neural Network has been done. As per experimental result SVM out forms the other classification method. In [3] support vector machine technique is combined with variable selection method for stock market prediction. A variable selection is applied to select the significant variables to feed to the SVM model. 11 variables have been used in the process of variable selection implemented by cross-correlation based method and by peeling method. Accordingly three models have been proposed. Based on experimental results, both hybridized models shows improvement in the prediction accuracy compared to SVM model. The trend prediction accuracy is presented in table 1. The accuracy increases from 82.5% to 84.166% by using two different variable selection methods combine with SVM. Paper [4] proposed a model in which they integrates particle swarm optimization (PSO) algorithm and LS-SVM for stock market prediction using financial technical indicators. Comparative study has been done on LS-SVM ,LS-SVM optimized by PSO and ANN-BP. Levenberg- marquardt(LM) algorithm is used as a benchmark to compare LS-SVM and LS-SVM-PSO. According to results LS-SVM optimised with PSO outperforms with lowest error value as compared to LS-SVM, while ANN-BP is the worst one. Indicators include relative strength index, money flow index, exponential moving average, stochastic oscillator and moving average convergence/divergence. Paper [5] Studies and compare four machine learning algorithms that are Single layer perceptron, multilayer perceptron, Radial basis function and Support vector machine. 9 parameters(i.e oil rates, gold rates, silver rates, FEX ,SMA, ARIMA, KIBOR, NEWS, Twitter) has been used as inputs for prediction model. The results of all four algorithms over training set and testing is presented in table 2. According to results shown SVM performed best on training set while MLP algorithm did well on test data set.

As test data is completely new and unseen instances. Therefore, MLP seems to be more efficient in predicting the stock market. MLP outperforms the remaining three algorithms. In [6] The SVM forecasting model has been compared with back-propagation neural network (BPN) and Case-based reasoning (CBR) and the result shows that SVM is the best model out of these three models. In [7] 12 technical indicators has taken as input for SVM, ANN, random forest (RF) forecasting model and SVM outperforms after comparison. [8] Support vector machine has been compared with ANN, ANFIS and SVM outperforms.

Support vector machines has been applied successfully in many problems such as speech recognition, signal recognition, text categorization, gene selection, intrusion detection, spam filtering, forecasting, medical image classification, classification. Several papers have been published assessing performance of SVM against some statistical and machine learning algorithms in financial application.

4. THE PROPOSED MODEL

The proposed model is based on the study of historical data of stocks, technical indicators and Linear-SVM algorithm to be used in the prediction of stock prices. The proposed model architecture consists of input block which represent the stocks historical data files. Next technical parameters (MACD and stochastic) will be calculated for each file in input processing block 1 and 2 respectively. Next two SVMs are use to analyze and predict the stocks .Lastly, one output represents whether to BUY or SELL the stocks.

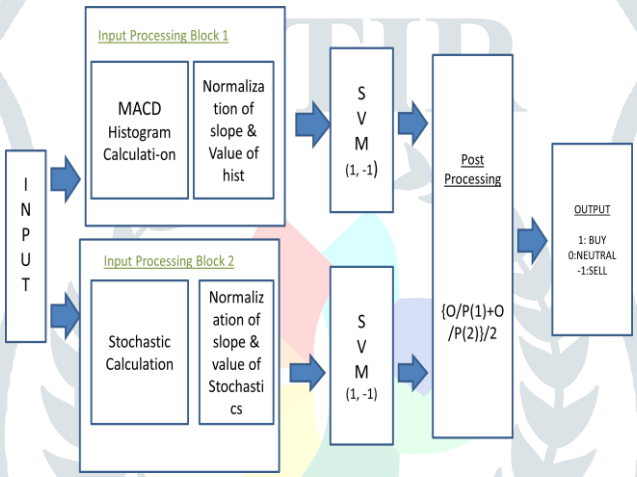

Fig. 2: The Proposed Model

Three technical indicators are calculated from the raw datasets:

- Exponential Moving Average (EMA): This indicator returns the exponential moving average of a area over a given period of time. EMA formula is as follows.
  EMA = [α *Today's Close] + [1-α* Yesterday's EMA].

- Stochastic Oscillator (SO): The stochastic oscillator is defined as the difference between the current closing price of a security and its lowest low price, relative to its highest high price for a given period of time. The formula for this computation is as follows:

%K = [(today's Close price – Lowest price) / (Highest Price – Lowest Price)] * 100

- Moving Average Convergence/Divergence (MACD): This function calculates difference between a short and a long term moving average for a area. The calculations for MACD and its signal are as follows

MACD = [0.075*EMA of close prices]-[0.15* EMA of close prices]
Signal line = 0.2*EMA of MACD

Following are the steps for prediction of stock market:

Step 1: This step is important for the download data from the site www.nseindia.com. We are predicting the future value of stock which came under nifty50. So the share values up to the closing date are downloaded from the site.

Step 2:In the next step the 70% dataset is provided to input processing block 1 at once as shown in fig 2.

Step 3: In the next step technical parameters that are MACD and Stochastic Oscillator are calculated which will be given as input to two different SVM along with the output list.

Step 4: Next the training of 70 % data is being carried out and 30 % of data will be tested.

Step 5: In next step algorithm will predict the stock for the dataset value efficiently.

Step 6:In final step algorithm displays whether to buy or sell the selected stock and that will be the mean result of two SVM .

5. EXPECTED OUTCOME

The data we use in this project comes from the site www.nseindia.com. we label each stock in our data set as a good or a poor investment. Although there is no such method for defining a market investment as "good" or "poor", we use a method that is simple and objective: if the price of a company's stock over a given month rose, it is classified as a good investment; otherwise it is classified as a poor investment. Our training sample is based on companies which comes under nifty50.

Symbol	Series	Date	Prev Close	Open Price	High Price	Low Price	Last Price	Close Price	Average Price	Total Traded Quantity	Turnover	No. of Trades	Deliverable Qty	% Dly Qt to Traded Qty
AMBUJACEM	EQ	6-Jun-18	202.05	201.5	206.8	200.45	204.5	204.15	204.39	2884900	589632730.7	26947	1733914	60.1
AMBUJACEM	EQ	7-Jun-18	204.15	205	207.3	204.75	206.15	206	206.12	4036226	831961052.1	34763	3106361	76.96
AMBUJACEM	EQ	8-Jun-18	206	205.3	207.6	204.75	206	206	206.28	2601050	536547202.1	14815	1980712	76.15
AMBUJACEM	EQ	11-Jun-18	206	206.95	210.6	205.15	205.45	206.05	207.26	4756149	985772989.1	25100	2782461	58.5
AMBUJACEM	EQ	12-Jun-18	206.05	205.9	207.5	204.95	207.1	206.9	206.33	2857333	589565611.7	21408	2033341	71.16
AMBUJACEM	EQ	13-Jun-18	206.9	207	211.6	206.2	210	210	210.13	4884946	1026461255	27769	3459730	70.82
AMBUJACEM	EQ	14-Jun-18	210	210.7	210.7	204.8	206.2	206.05	206.69	4024884	831911315	20633	3113483	77.36
AMBUJACEM	EQ	15-Jun-18	206.05	204.5	207.65	202	205	204.95	206.01	6832521	1407586753	37706	5449351	79.76
AMBUJACEM	EQ	18-Jun-18	204.95	204.5	205.9	202.5	202.55	202.9	203.71	803881	163758847.4	14056	289774	36.05
AMBUJACEM	EQ	19-Jun-18	202.9	202.25	203.25	198.85	201.65	200.25	200.39	1605387	321700446.7	27296	784113	48.84
AMBUJACEM	EQ	20-Jun-18	200.25	200.95	201.8	196.55	197.1	197.75	198.03	1716864	339987219	27910	841007	48.99
AMBUJACEM	EQ	21-Jun-18	197.75	198	198.25	195.55	197.1	197	196.87	2532830	498641690.2	27251	1719394	67.88
AMBUJACEM	EQ	22-Jun-18	197	197.1	206	195.65	205	204.95	202.55	3167408	641553799	29397	1439810	45.46
AMBUJACEM	EQ	25-Jun-18	204.95	205	209	204.8	205.9	206.8	207.49	3363586	697903794.9	29489	1463941	43.52
AMBUJACEM	EQ	26-Jun-18	206.8	207	215.15	206.65	211.5	211.95	212.67	7976623	1696420483	77767	4077245	51.11
AMBUJACEM	EQ	27-Jun-18	211.95	213.4	213.4	204.05	204.5	205.3	207.51	3049747	632850350.4	23473	1223619	40.12
AMBUJACEM	EQ	28-Jun-18	205.3	205.55	208.25	200.1	200.75	200.9	203.33	3785931	769776265.6	26770	2157895	57
AMBUJACEM	EQ	29-Jun-18	200.9	201.9	208	201.7	208	207.45	205.77	3749070	771458929.2	45664	2090737	55.77
AMBUJACEM	EQ	2-Jul-18	207.45	207.8	207.95	198.25	199.25	199.9	201.72	3041228	613488919.4	24172	1530769	50.33
AMBUJACEM	EQ	3-Jul-18	199.9	200.35	202.3	197.85	200.7	200.9	200.84	4134828	830425076.8	31061	2592562	62.7
AMBUJACEM	EQ	4-Jul-18	200.9	201.4	204.3	199.7	200.9	201	201.8	2262034	456475042.7	12653	1084337	47.94
AMBUJACEM	EQ	5-Jul-18	201	201.8	208.85	201.5	207.9	207.45	206.02	4657759	959613410.8	34094	2070195	44.45
AMBUJACEM	EQ	6-Jul-18	207.45	207.75	212.3	206.7	207.1	207.55	209.69	3029745	635312677.8	27179	1220387	40.28
AMBUJACEM	EQ	9-Jul-18	207.55	209.05	210	205.4	206	206.15	207.11	2007632	415790717.1	18505	989415	49.28
AMBUJACEM	EQ	10-Jul-18	206.15	208.9	208.9	203.55	204.45	204.35	205.88	3550279	730945556.7	35039	2252111	63.43
AMBUJACEM	EQ	11-Jul-18	204.35	204.6	205.1	201.2	202.65	202.55	202.45	2524353	511064003.3	17760	1334305	52.86
AMBUJACEM	EQ	12-Jul-18	202.55	203	206.2	201.25	201.95	201.7	203.76	4308979	877982410.7	30036	2371392	55.03
AMBUJACEM	EQ	13-Jul-18	201.7	202.9	203.5	196.3	197	196.9	198.32	2220084	440296970	17634	1122868	50.58

Table 3: dataset sample

Each file (Test Data) will contain 15-35 days of closing Data (Date, Open Price, High Price, Low Price, Close Price, Volume) of a particular stock.

As the purpose of the classification model should be used to predict how a stock will perform in the upcoming days, when the result is unknown, it is necessary to test the SVM model on a test sample. 150 such files is be used in total for this project.
70% of the Test Data will be used to train ML Algorithm. 30% will be used to test/Validate result and produce conclusions. Stocks taken are from Indian Stock market. (NSE). Specifically stocks present in nifty50 has been chosen to create Test Data as they contain large volume and has less chances of price manipulation by operators.
Test Data file Name format :
CompanyName_UP/DOWN_PriceValue.csv
Ex:-AmbujACem_UP_228.csv

File Name	O/P Label	Stochastic Value	Stochastic Slope	MACD Value	MACD Slope
AmbujaCem_DOWN_207.csv	0	-0.2605842	-0.9623978	-0.25821	-0.87634
AmbujaCem_DOWN_230.csv	0	-0.6606773	-0.935094	-0.2263	-0.34932
AXISBANK_DOWN_530.csv	0	-0.0438671	-0.968065	-0.148423	0.952721
AmbujaCem_UP_228.csv	1	0.4467311	0.9134840	-0.383517	0.812373
BPCL_UP_374.csv	1	0.8051687	0.9749006	0.1165824	0.968156
BAJAJAUTO_DOWN_3164.csv	0	0.1371038	-0.9311099	-0.360292	-0.961656
GAIL_UP_386.csv	1	-0.3975044	0.9519803	-0.757913	0.596199
HCL_UP_1055.csv'	1	-0.629806	0.8838832	1.0	0.935676

Table 4: Intermediate results

Table 1: PREDICTION ACCURACY

Model	Accuracy
SVM	82.5%
CC + SVM	84.1667%
Peeling + SVM	89.1667%

Table 2: COMPARISON OF MACHINE LEARNING TECHNIQUES

Data Set used for verification	Machine Learning Algorithm			
	SLP	MLP	RBF	SYM
Training Set	83%	67%	61%	100%
Testing Set	60%	77%	63%	60%
Average	71.5%	72%	62%	80%

6. CONCLUSION AND FUTURE WORK

In this project, we proposed the use of data collected from stock market with machine learning algorithms to predict the stock index movements. Our conclusion can be summarized into following aspects: As we are using two SVM algorithms in same prediction model, the efficiency and accuracy of model will be better. SVM provide a promising tool for forecasting stock market. Also the technical parameters that are MACD and Stochastic oscillator are the two leading parameters which will efficiently suggest the future value of stocks. In This paper we have studied a procedure to predict stock index movement using support vector machine (SVM).

ACKNOWLEDGMENT

I am grateful to my guide Prof. Nitin Janwe for his unwavering support and clear guidance. Also, my sincere thanks to all authors, whose papers in the area of stock market prediction are published in various conferences proceeding and journals

REFERENCES

[1] J. Zhu, Zhen Hu and Ken Tse “Stocks Market Prediction Using Support Vector Machine”, 6th International Conference on Information Management, Innovation Management and Industrial Engineering, 2013.
[2] Y. Nakamori, Wei Huang, Shou-Yang Wang, “Forecasting stock market movement direction with support vector machine”, Computers & Operations Research, Vol. 32, Issue 10, October 2005.
[3] Hakob Grigoryan, "Stock Market Trend Prediction Using Support Vector Machines and Variable Selection Methods”: Advances in Intelligent Systems Research (AISR), volume 141International Conference on Applied Mathematics, Modeling and Statistics Application (AMMSA 2017).
[4] Osman Hegazy 1, Omar S. Soliman 2 and Mustafa Abdul Salam3, “A Machine Learning Model for Stock Market Prediction”: International Journal of Computer Science and Telecommunications [Volume 4, Issue 12, December 2013]
[5] Mehak Usmani1, Syed Hasan Adil2, Kamran Raza3 and Syed Saad Azhar Ali4, “ Stock Market Prediction Using machine learning techniques “: 2016 3rd International Conference On Computer And Information Sciences (ICCOINS)-IEEE 2016.
[6] Kim k., “Financial time series forecasting using support vector machines”, Neuro computing, 55(2003), 2003.
[7] Thenmozhi M, Manish k., “Forecasting stock index movement: A comparison of support vector machines and random forest”, Indian Institute of Capital Market Conference, Mumbai India, 2005.
[8] Kizilaslan R, Freund S., Oztekin, Iseri A., “A data analytic approach to forecasting daily stock returns in an emerging market”, European Journal of Operational Research, 253(2016), , 2016. ISSN 2278-6856
[9] Stock Market, Wikipedia.