A Survey on Stock Market Prediction Using SVM

Sachin Sampat Patil¹, Prof. Kailash Patidar² Asst. Prof. Megha Jain ³

SSSIST, Sehore, Madhya Pradesh, India

Abstract—A lot of studies provide strong evidence that traditional predictive regression models face significant challenges in out-of-sample predictability tests due to model uncertainty and parameter instability. Recent studies introduce particular strategies that overcome these problems. Support Vector Machine (SVM) is a relatively new learning algorithm that has the desirable characteristics of the control of the decision function, the use of the kernel method, and the sparsity of the solution. In this paper, we present a theoretical and empirical framework to apply the Support Vector Machines strategy to predict the stock market. Firstly, four company-specific and six macroeconomic factors that may influence the stock trend are selected for further stock multivariate analysis. Secondly, Support Vector Machine is used in analyzing the relationship of these factors and predicting the stock performance. Our results suggest that SVM is a powerful predictive tool for stock predictions in the financial market.

Keywords—Stock Classification; Data Mining; Support Vector Machine; Forecasting

1. INTRODUCTION

The macroeconomic environment and the financial market are complex, evolutionary, and non-linear dynamical systems. Before we study the historic volatile days of the ten years, let us first know what are:

a) Stock Markets
b) Stock exchanges

a) Stock Markets:
Stock Market is a market where the trading of company stock, both listed securities and unlisted takes place. It is different from stock exchange because it includes all the national stock exchanges of the country. For example, we use the term, "the stock market was up today or the stock market bubble.

b) Stock Exchanges:
Stock Exchanges are an organized marketplace, either corporation or mutual organization, where members of the organization gather to trade company stocks or other securities. The members may act either as agents for their customers, or as principals for their own accounts. Stock exchanges also facilitates for the issue and redemption of securities and other financial instruments including the payment of income and dividends. The record keeping is central but trade is linked to such physical place because modern markets are computerized. The trade on an exchange is only by members and stock broker do have a seat on the exchange [15].

1.1 Support Vector Machine

The support vector machine (SVM) is a training algorithm for learning classification and regression rules from data, for example the SVM can be used to learn polynomial, radial basis function (RBF) and multi-layer Perception (MLP) classifiers. SVMs were first suggested by Vapnik in the 1960s for classification and have recently become an area of intense research owing to developments in the techniques and theory coupled with extensions to regression and density estimation. SVMs arose from statistical learning theory; the aim being to solve only the problem of interest without solving a more difficult problem as an intermediate step. SVMs are based 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 key 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 equivalent to maximizing the margin between two classes. Mathematically, we
will obtain a quadratic programming problem where the number of variables is equal to the number of observations.

2. HISTORY ABOUT STOCK MARKET

In 12th century France the countries they change were concerned with managing and regulating the debts of agricultural communities on behalf of the banks. Because these men also traded with debts, they could be called the first brokers. A common misbelieve is that in late 13th century Bruges commodity traders gathered inside the house of a man called Van der Beurze, and in 1409 they became the "Brugse Beurse", institutionalizing what had been, until then, an informal meeting, but actually, the family Van der Beurze had a building in Antwerp where those gatherings occurred. Until then, the family Van der Beurze had a building in Antwerp where those gatherings occurred. [10] the Van der Beurze had Antwerp, as most of the merchants of that period, as their primary place for trading. The idea quickly spread around Flanders and neighboring countries and "Beurzen" soon opened in Ghent and Rotterdam [15]. In the middle of the 13th century, Venetian bankers began to trade in government securities. In 1351 the Venetian government outlawed spreading rumors intended to lower the price of government funds. Bankers in Pisa, Verona, Genoa and Florence also began trading in government securities during the 14th century. This was only possible because these were independent city states not ruled by a duke but a council of influential citizens. Italian companies were also the first to issue shares. Companies in England and the Low Countries followed in the 16th century. The Dutch East India Company (founded in 1602) was the first joint-stock company to get a fixed capital stock and as a result, continuous trade in company stock occurred on the Amsterdam Exchange. Soon thereafter, a lively trade in various derivatives, among which options and repos, emerged on the Amsterdam market. Dutch traders also pioneered short selling a practice which was banned by the Dutch authorities as early as 1610. There are now stock markets in virtually every developed and most developing economies, with the world's largest markets being in the United States, United Kingdom, Japan, India, Pakistan, China, Canada, Germany (Frankfurt Stock Exchange), France, South Korea and the Netherlands [15].

3. LITERATURE SURVEY

How can we predict whether the price of a particular stock will go up or down in the upcoming year? In the modern techniques, one way is to develop a predictor based on the information in the historical data. First of all we should select some major factors that may influence the performance of the stocks; we can further discover an interesting model from our dataset to predict the future performance of any stocks. To predicting the stock market status various algorithm was developed out of which the Artificial Neural Network is the famous technology to predict the stock market. In this scenario we implement the algorithm which predicts the stock market using SVM (Support Vector Machine) which give an output very efficiently. In 2013, the name of scientist Zen Hu, Jie Zhu, Ken Tse3 publish the Paper with name “Stock Market Prediction Using Support Vector Machine” in IEEE journal “6th international conference on Information, Innovation Management and Industrial Engineering” with page no, 978-1-4799-0245-3/13/ in 2013 in which they are introducing the prediction of stock market with the help of Support vector Machine. These three scantiest is working together for developing the first idea for which we predict the stock market value with the help of support vector machine. Support Vector Machine (SVM) is a relatively new learning algorithm that has the desirable characteristics of the control of the decision function, the use of the kernel method, and the sparsity of the solution. They present a theoretical and empirical framework to apply the Support Vector Machines strategy to predict the stock market. Firstly, four company-specific and six macroeconomic factors that may influence the stock trend are selected for further stock multivariate analysis. Secondly, Support Vector Machine is used in analyzing the relationship of these factors and predicting the stock performance. They giving results suggest that SVM is a powerful predictive tool for stock predictions in the financial market.[1].The base paper which is used by Zen Hu, Jie Zhu, Ken Tse for demonstrating the result is that 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,they investigate the 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. Their experimental result shows that SVM out forms the other classification method further they combining model by integrating SVM with other classification method and the result after combining model gives the best result forecasting method. [2].The another paper which is help to giving
an idea about the building such huge concept is that “A Tutorial on Support Vector Machines for Pattern Recognition”, Data Mining and Knowledge Discovery, in 1998. C. J. C. Burges. He describes the Support vector machine for separable and non-separable data, working through a non-trivial example details. He describes a mechanical analogy and discusses when SVM solution is unique and when they are global. Also he describe the how support vector training can be practically implemented and giving details the kernel mapping technique which is used to construct SVM Solution which are in nonlinear data. They proof that SVM have very large radial basis function kernel.[3].In 1995 C. Cortes and V. Vapnik on the topic “Support Vector Networks” they are working on the non-separable training data because the previous work is done on the training data which is separated by error. They also compare the support vector network to other learning algorithm. [4] The M. Pontil and A. Verri, “Properties of SVM", Technical Report, Massachusetts Institute of Technology, 1997.E.E. Osuna, R. Freund and F. Girosi, “Support Vector Machines: raining and Applications”, Technical Report, Massachusetts Institute of Technology, Artificial Intelligence Laboratory, AI Memo 1997.[5] A Tutorial on Support Vector Regression Alex J. Smola and Bernhard. In this tutorial they give an overview of the basic ideas underlying Support Vector (SV) machines for function estimation. Furthermore, they include a summary of currently used algorithms or training SV machines, covering both the quadratic (or convex) programming part and advanced methods for dealing with large datasets. Finally, they mention some modifications and extensions that have been applied to the standard SV algorithm, and discuss the aspect of regularization from a SV perspective.[10] Zhu et al. (2008) investigated whether trading volume can significantly improve the prediction performance of neural networks under short, medium and long term forecasting horizons. Support vector machines (SVMs) have been extensively researched in machine learning community for the last decade and actively applied in different areas. SVMs are basically used for regression and classification functions, which are called “classifying SVM”, support vector regression (SVR), respectively. Support vector machines have been applied successfully in many problems such as speech recognition (Changxue et al., 2001), signal recognition (Gexiang et-al.,2004),text categorization (Pan et al., 2009), gene selection (Zhang Q. 2007), intrusion detection (Zhenguo and Guanghua, 2009), spam filtering (Amayri and Bouguila,2009), forecasting (Shen et al., 2006; Guo-Rui et al.,2007; Liu et al., 2009; Shu-xi and Wang, 2006; Tian et al., 2009), medical image classification (Bai and Tian,2009; Zaim et al., 2007), classification (Changsheng et al., 2003; Jing et al., 2009; Zai-Wen et al., 2009; Reljin and Pokrajac, 2008) [12]. Several papers have been published assessing performance of SVM against some statistical and machine learning algorithms in financial applications. 2001: Tay and Lijuan (2001) examined the feasibility of SVM in financial time series forecasting by comparing it with a multi-layer back-propagation (BP) neural networks. Analysis of the experimental results proved that it is advantageous to apply SVMs to forecast financial time series. Fan and Palaniswami, (2001) used SVMs for classification for stock selection on the Australian Stock Exchange and significantly outperformed the benchmark. [12]2002: Yang et al. (2002) tried varying the margins in SVM regression in order to reflect the change in volatility of financial data and also analyzed the effect of asymmetrical margins so as to allow for the reduction of the downside risk. The former approach produced the lowest total error when predicting the daily closing price of Hong Kong’s Hang Seng Index (HSI). Calvo et al. (2002) compared the performance of NN, Naive Bayes and SVM algorithms for the automatic categorization of corporate announcements in the Australian Stock Exchange (ASX). They performed tests on two categorization tasks: “market sensitivity” which indicates whether an announcement will have an impact on the market, and “report type” which classifies each announcement into one of the report categories defined by the ASX. [12]2003: Sansom et al. (2003) evaluated utilizing neural networks (NNs) and support vector machines (SVM) for wholesale (spot) electricity price forecasting. The SVM required less time to optimally train than the NN, whilst the SVM and NN forecasting accuracies were found to bevery similar. Gavrishchaka and Ganguli, (2003) used SVMs for forecasting the volatility of foreign-exchange data. Their preliminary benchmark tests indicated that SVMs can perform significantly better than or comparable to both naive and GARCH models. [12]2004: Huang et al. (2004) used SVM for credit rating analysis and back propagation neural network (BNN) as a benchmark. They obtained prediction accuracy around 80% from both BNN and SVM methods for the United States and Taiwan markets. However, only slight improvement of SVM was observed. Zhou et al. (2004) applied an accurate online support vector regression (AOSVR) to forecast the prices of the electric-power markets. Results showed
that it was effective in forecasting the prices of the electric – power market. [12]2005: Shin et al. (2005) demonstrated that SVMs perform better than back-propagation neural networks when applied to corporate bankruptcy prediction and the accuracy and generalization performance of SVM that is

Better than that of BPN as the training set size is getting smaller. Huang et al. (2005) investigated the predictability of financial movement direction with SVM by forecasting the weekly movement direction of NIKKEI 225 index. To evaluate the forecasting ability of SVM, we compare its performance with those of linear discriminate analysis; quadratic discriminate analysis and Elman back propagation neural networks. The experiment results show that SVM outperforms the other classification methods. Pai and Lin (2005) proposed a hybrid methodology that exploits the strength SVM model in forecasting stock prices problems. [12]2006: Chen et al. (2006) compared SVMs and back propagation (BP) neural networks when forecasting the six major Asian stock markets. Both models perform better than the benchmark. Autoregressive model in the deviation measurement criteria, whilst SVMs performed better than the BP model in four out of six markets Han et al.(2007) proposed a method by using SVM with financial statement analysis for prediction of stocks. Author [12] methodologies to select equities based on soft-computing models which focus on applying fundamental analysis for equities screening and compared the performance of three soft-computing models, namely multi-layer perceptions (MLP), adaptive neuro-fuzzy inference systems (ANFIS) and general growing and pruning radial basis function (GGAP-RBF). Tirkayi and Ahlatcioglu (2005) proposed a new method for group decision making in fuzzy environment and demonstrates the usefulness of fuzzy methodology in financial problems. They accomplished it by making some modification of Chen’s method and it is shown how their method can be used for stocks selection on ISE (Istanbul Stock Exchange). [12]2009:Xidonas et al. (2009) presented an expert system methodology for supporting decisions that concern the selection of equities, on the basis of financial analysis. They proposed a methodology that is employed for selecting the attractive equities through the evaluation of the overall corporate performance of the corresponding firms and validity of the proposed methodology that is tested through a large scale application on the Athens Stock Exchange. Horton (2009) examined Japanese Candlestick methods of technical analysis for 349 stocks. Using more data and alternative tests, the study contradicts an earlier article in the literature, finding little value in the use of candlesticks and providing more support for the weak form of the efficient markets hypothesis. [12]2010: Maciel et al. (2010) analyzed neural networks for financial time series forecasting, specifically, their ability to predict future trends of North American, European, and Brazilian stock markets

4. IMPORTANCE OF STOCK MARKET

The stock market is one of the most important ways for companies to raise money, along with debt markets which are generally more imposing but do not trade publicly. This allows businesses to be publicly traded, and raise additional financial capital for expansion by selling shares of ownership of the company in a public market. The liquidity that an exchange affords the investors enables their holders to quickly and easily sell securities. This is an attractive feature of investing in stocks, compared to other less liquid investments such as property and other immovable assets. Some companies actively increase liquidity by trading in their own shares. History has shown that the price of stocks and other assets is an important part of the dynamics of economic activity, and can influence or be an indicator of social mood. An economy where the stock market is
on the rise is considered to be an up-and-coming economy. In fact, the stock market is often considered the primary indicator of a country's economic strength and development.

![Diagram of Financial Market](image)

**Figure 1. Hierarchical Structure: Financial Market**

Rising share prices, for instance, tend to be associated with increased business investment and vice versa. Share prices also affect the wealth of households and their consumption. Therefore, central banks tend to keep an eye on the control and behavior of the stock market and, in general, on the smooth operation of financial system Functions. Financial stability is the raison d'être of central banks. Exchanges also act as the clearinghouse for each transaction, meaning that they collect and deliver the shares, and guarantee payment to the seller of a security. This eliminates the risk to an individual buyer or seller that the counterparty could default on the transaction. The smooth functioning of all these activities facilitates economic growth in that lower costs and enterprise risks promote the production of goods and services as well as possibly employment. In this way the financial system is assumed to contribute to increased prosperity, although some controversy exists as to whether the optimal financial system is bank-based or market-based Recent events such as the Global Financial Crisis have prompted a heightened degree of scrutiny of the impact of the structure of stock markets (called market microstructure), in particular to the stability of the financial system and the transmission of systemic risk. The field of financial forecasting is characterized by data intensity, noise, non stationary, unstructured nature, and hidden relationships. Predicting financial indicators is therefore a difficult task. However, forecasting is important in the sense that it provides concrete data for investment decisions. How can we predict whether the price of a particular stock will go up or down in the upcoming year? In the modern techniques, one way is to develop a predictor based on the information in the historical data. First of all, we should select some major factors that may influence the performance of the stocks; we can further discover an interesting model from our dataset to predict the future performance of any stocks. That is to say, we need to learn a model that can map those factors into the class attribute which indicates the whole performance of stocks. Support vector machine (SVM) is a machine learning technique that can be used for this purpose of classification. Established on the unique theory of the structural risk minimization principle to estimate a function by minimizing an upper bound of the generalization error, SVM is shown to be very resistant to the over-fitting problem, eventually achieving a high generalization performance. Another key property of SVM is that training SVM is equivalent to solving a linearly constrained quadratic programming problem so that the solution of SVM is always unique and globally optimal, unlike neural networks training, which requires nonlinear optimization with the danger of getting stuck at local minima. Although SVM approach has been widely applied in financial forecasting, little analysis is extended into the stock market of China, which is the second biggest economic entity and also one of the most important emerging markets in the world. Additionally, while indices, the main indicators of countries’ economic condition; have been the efficient instrument for both hedgers and speculators in traditional and derivative market, less work is performed in this area than the analysis on individual stocks. Besides, model inputs in prior research typically involve price and volume data, and may also include a selection of well-known technical Indicators but few papers apply fundamental indicators in the model. Compared to the technical analysis, which is the study of collective market sentiment mainly reflected in the price and volume, fundamental analysis focuses more on the intuitive physical interpretation and attempts to find the intrinsic value of the assets. Fundamental variables selected and included in the model generally have intuitive justification and have certain connection with the target; while it may be difficult to explain a technical analytical model.

### 5. MOTIVATION & OBJECTIVE

Where and how to invest one’s money is a problem that many individuals face. Despite the large amount of money and interest embedded into this problem, there is still no definitive answer on when and where money
should be invested. Due to this, most people either keep their money in the bank or hand it off to someone else to manage. This work looks to develop an automated trading system that can be used to make more money off trading securities than the traditional buy-and-hold strategy and to reduce the risk involved in making these investments. In today’s world the stock market is important concept for the well developing business. The stock means the ownership of any company so that owner sell the stock to other person or the other person parches the stock so the they are responsible or authenticate person for the shearing the profit and loss of the company. So many people have interest about the stock market and the future of the nation is totally depends upon the economy of the country so that many people invest their money in various stock exchanges and earn the profit. This is small idea to build up the project for predicting stock market value so the people get an idea about the tomorrow stock market value and safe to take decision at the time of investment. In recent years, there have been a growing number of studies looking at the direction of movements of various kinds of financial instruments. Both academic researchers and practitioners have made tremendous efforts to predict the future movements of stock market index or its return and devise financial trading strategies to translate the forecasts into profits. In the following section, we focus the review of previous studies on artificial neural networks (ANN) and support vector machines (SVM) applied to stock market prediction. Once the efficient parameter values are specified, prediction performances of ANN and SVM models can be compared to each other. This performance comparison was performed on the entire data set considering the parameter values specified using the parameter setting data set. That is, the prediction models must be re-trained Using a new training data set which must be a new part of the entire data set and must be larger than the training subset of parameter setting data set after re-training, out of sample evaluation of models must be carried out using a new holdout data set, which is the remaining part of entire data set. Therefore, the entire data set was re-divided into the training data set (70% of entire) and the holdout data set (30% of entire) for comparison experiments. This was also realized by considering the dispersion of increases and decreases in the entire data set. The objective of this work is to develop a market trading model that can successfully trade market securities for a profit, beating buy-and-hold. This was accomplished through the construction of a market trading simulator and the exploration of many different trading models. The models will be evaluated by classification rate, using SVM and ANN ant it will predict the stock market value. Following are the objective:

1. The main objective of implementation of this project is to predict the stock market current condition with reference of last day’s stock market value. It is beneficial to the use who can predict the value of the any company share value using this algorithm so it will guess that the financial condition of that company whether the share of company is satisfactory or not how it goes in increasing range or going to the decreasing. So it is forecast the stock exchange. In market so many algorithm are present for stock market prediction but the SVM provides the stock value with efficiently within a less amount of time.

2. The Second main objective of the project is that it will work on the big data value. This algorithm takes the data from the global networking financial site with specified attribute. These attribute have, Last day opening Stock Price, Last-day high Value of Stock Price, Last-day low, Value of Stock Price, Last day stock Volume, Last-day closing price. Etc. these value is available in the financial data after that these data is covert in to the .CSV (comma Separate Value) .xlsx (Excel file). These file contain large amount of data that database at the time of prediction these value are executed by algorithm and get the prediction of given stock in the market.

3. Third objective is that in given project we have to compare the prediction value by SVM and ANN so we are getting concluded which algorithm get efficient to predict the value of stock market. The Chapter 1 giving an introduction about the stock market in which we have to learn about stock, stock exchange, history of stock market also we focused on the importance of stock market, what are the different properties of stock market and what are the different function of stock market. In chapter 2 we focus on history of stock market. In chapter 3 we focus on the literature review about the topic which is the path of building the project it was a nice survey regarding the topic to predict the value of stock market by using SVM. Actually SVM have much application but the main objective of this work is to predict the stock market value with the help of support Vector Machine (SVM).In chapter 4, 5 and 6 we focus on importance of stock market, motivations and objectives of stock market respectively.
6. CONCLUSION

In the project, we proposed the use of data collected from different global financial markets with machine learning algorithms to predict the stock index movements. Our conclusion can be summarized into following aspects: SVM algorithm works on the large dataset value which collected from different global financial markets. Also SVM does not give a problem of over fitting. Correlation analysis indicates strong interconnection between the Market stock index and global markets that close right before or at the very beginning of trading time. Various machine learning based models are proposed for predicting daily trend of Market stocks. Numerical results suggest high efficiency. A practical trading model is built upon our well trained predictor. The model generates higher profit compared to selected benchmarks.

REFERENCES