Stock Market Prediction and Performance Analysis Using Machine Learning Algorithms

Roopalakshmi S[1], Shilpa Reddy K[2], Noor Basha[3]

[1,2,3] *Assistant Professor in CSE, Vemana Institute of Technology, Bengaluru, Karnataka, India.*

Abstract: Stock market price data is generated in huge volume and it changes every second. Stock market is a complex and challenging system where people will either gain money or lose their entire life savings. In this research work, an attempt is made for prediction of stock market trend.The predicting of exchange movements may be a big drawback in Stock market. Social media is utterly representing people's sentiment and opinions regarding the current events. Twitter has played a big role in attracting loads of attention from researchers for finding out people sentiments. Exchange prediction supported public sentiments expressed on Twitter associated with alternative social media have been the main field of analysis. To achieve the goal, this research work primarily uses machine learning techniques to discover various factors linked to stock market. Machine learning approaches have a successful outcome in extracting information by building together stock market predicting models from stock datasets. To predict stock market, the quarrying of information from these data can be useful. The work is to look at how well the changes available i.e. Cost of corporation, the ups and downs, are correlating with the public opinions being expressed in tweets of the company.

Keywords: **Stock market, Machine learning, Prediction**

I. INTRODUCTION

The stock market refers to the collection of markets and exchanges where regular activities of buying, selling, and issuance of shares of publicly held companies take place. Such financial activities are conducted through institutionalized formal exchanges held over the counter (OTC), marketplaces which operate under a defined set of regulations. A stock market is a similar designated market for trading various kinds of securities ina controlled, secure and managed environment. Since the stock market brings together hundreds of thousands of market participants who wish to buy and sell shares, it ensures fair pricing practices and transparency in transactions. While earlier stock markets used to issue and deal in paper-based physical share certificates, the modern-day computer-aided stock markets operate electronically. Stock market prediction is the act of trying to determine the future value of a company stock or other financial instrument traded on an exchange [1]. The successful prediction of a stock's future price could yield significant profit. The efficient market hypothesis suggests that stock prices reflect all currently available information and any price changes that are not based on newly revealed information thus are inherently unpredictable. Others disagree and those with this viewpoint possess myriad methods and technologies which purportedly allow them to gain future price information.

Predicting how the stock market will perform is one of the most difficult things to do. There are so many factors involved in the prediction – physical factors vs. psychological, rational and irrational behavior, etc. All these aspects combine to make share prices volatile and very difficult to predict with a high degree of accuracy. Using features like the latest announcements about an organization, their quarterly revenue results, etc., machine learning techniques have the potential to unearth patterns and insights we did not see before, and these can be used to make unerringly accurate predictions. Machine learning (ML) is the study of computer algorithms that improve automatically through experience and using data. It is seen as a part of Artificial Intelligence. Machine learning algorithms build a model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications, such as in medicine, email filtering, speech recognition, and computer vision,where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks. Machine learning involves computers discovering how they can perform tasks without being explicitly programmed to do so. algorithms. In practice, it can turn out to be more effective to help the machine develop its own algorithm, rather than having human programmers specify every needed step.

II. LITERATURE REVIEW

Badre Labiad et al [1] systems use different machine learning techniques to come up with the wanted investment decision. Artificial Neural Networks (ANN) are among the most used techniques for stock markets forecasting. In this work, we present an end-to-end system to implement ANNs for intraday forecasting of Moroccan stock market. Different ANN architectures are implemented and tested. Experimental results show that ANN can be used to make satisfactory intraday predictions.

Shou-Hsiung Cheng et al [2] proposed a method for forecasting the change of intraday stock price by utilizing text mining news of stock based on text mining techniques coupled with rough sets theories and support vector machine classifier. The method can handle without difficulty unstructured news of Taiwan stock market through pre-processing, feature selection and mark. The method also extracts the core phrases by using rough sets theories after the unstructured information has been transformed into structured data. Then, a prediction model is established based on support vector machine classifier. The empirical results show that the proposed model can predict accurately the ups and downs of a stock price within one hour after the news released. The method presented in the study is straight forward, simple and valuable for the short-term investors. Can predict accurately the rise and the fall of an individual stock price.

Zhen Hu et al [3] 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, a theoretical and empirical framework to apply the Support Vector Machines strategy to predict the stock market. Predicts a high percentage of the outcomes for various stocks.

Noor Basha et al [4] made use of many classification algorithms to predict the severe heart syndromes based on risk rate, where the author specifically used machine learning approach. The approach can be used for various datasets to predict the conditional requirements. Support Vector machine approach to predict the syndrome.

III. Proposed System

To achieve our target research methodology involves a few stages accrual of stock prediction dataset with relevant stock attributes, pre-processing numerical value attributes, smear multiple machine learning techniques and conforming predictive analysis using these data. Thereafter, we delegate these phases fleetingly. The architectural configuration procedure is concerned with building up a fundamental basic system for a framework. It includes recognizing the real parts of the framework and interchanges between these segments. The beginning configuration procedure of recognizing these subsystems and building up a structure for subsystem control and correspondence is called construction modeling outline and the yield of this outline procedure is a portrayal of the product structural planning. The proposed architecture for this system is given below in fig 1.1. It shows the way this system is designed and brief working of the system.

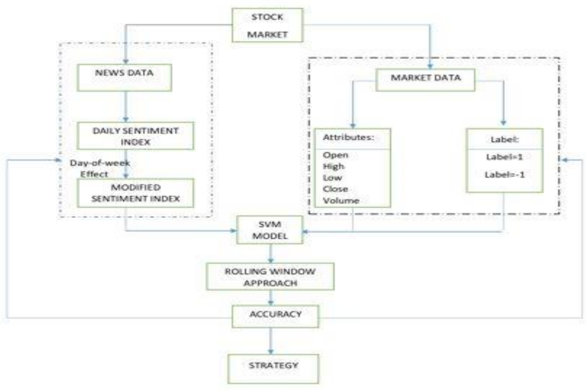

Fig 1.1 System architecture and working of stock prediction model

3.1 Investor sentiment is made up of three steps

Step 1 Web crawler: In this step, we aim to build a web crawler to automatically download the targeted textual documents from the Internet and store them to a database for further processing. The web crawler begins with the seeds in the form of a list of URLs. The scheduler manages the queue of URLs, deciding the priority and eliminating duplicate parts. Next, the downloader is responsible for acquiring the web pages from the Internet and providing them to the spider, which is used to parse the pages and extract the targeted contents. Next to obtain comprised of two sections: one is the textual news with the date from the websites, and then we store the precious data into the database; the other is the URLs contained in the pages, and then the URLs are transported to the scheduler. The procedures are repeated until we get hold of all the targeted textual documents. Each of the documents is displayed as time, headline, and contents in the database.

Step 2 Daily sentiment: A sentence-based sentiment analysis approach is used to process the textual data during a specific period. We regard a sentence as a unit to interpret the meaning of the whole document instead of a single word because a sentence can express a relatively complete meaning and help address the ambiguity problem. As a result, a document is divided into sentences first. Next, we segment the sentences into separate words, then project the words onto the sentiment space, count the number of positive and negative words, assign a specific sentiment value, and decide the polarity of each sentence based on How Net and Chinese Sentiment Analysis Ontology Base. How Net is an online common-sense knowledge base unveiling inter conceptual relationships and inter attribute relationships of concepts as connoted in lexicons of the Chinese and their English equivalents. Chinese Sentiment Analysis Ontology Base is constructed by Dalian University of Technology and depicts words and phrases from various aspects containing part of speech, polarity, and sentiment intensity.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 06 | June - 2022 Impact Factor: 7.185 ISSN: 2582-3930

Step 3 Modified sentiment: The day-of-week effect is one of the most well-known financial anomalies, which means that the average return on a Monday is much lower than that on the other days of the week. The reason includes that large amount of news is reported on the weekend or on Friday just after the market is closed. With such considerable and valuable information to deal with, investors are very likely to change their mind and take actions on Mondays. Furthermore, corporations also tend to release important news on the weekend to ensure the stability of the stock and boost the public image. If it is bad news, investors will have enough time to digest and accept it, whereas if it is good news, companies can continuously spread out news to make it known by more and more people and expand their coverage.

3.2 Support Vector Machine

An SVM model is basically a representation of different classes in a hyperplane in multidimensional space. The hyperplane will be generated in an iterative manner by SVM so that the error can be minimized. The goal of SVM is to divide the datasets into classes to find a maximum marginal hyperplane (MMH). The main goal of SVM is to divide the datasets into classes to find a maximum marginal hyperplane (MMH) and it can be done in the following two steps:

First, SVM will generate hyperplanes iteratively that segregates the classes in best way.

Second, it will choose the hyperplane that separates the classes correctly.

3.3 Data Description: The intention is to explore the trend of a very important index in China, the SSE 50 Index, not only by using stock market data but also exploiting news documents related to it and its constituents. The SSE 50 Index is a primarily blue-chip stock index on the Shanghai stock market, and it is made up of the 50 largest stocks of good liquidity and representativeness. Conventional time series data include opening price, closing price, high for the day, low for the day, trading volume in number of shares, trading volume in RMB, change in RMB, and change in percentage. We download such data of the SSE 50 Index and its 50 constituents from the Wind Economic Database, which is the market leader in China's financial information service industry. Accordingly, we applied the web crawler we built to download all the posts and documents of the 51 shares from the Sina stock forum and East money stock forum over the period between June 17th, 2014 and June 7th, 2016, including 485 trading days. The two forums are widely regarded as active and mainstream communities in China. The number of reviews of each stock is 37 855 on average, peaking at 23 236 and reaching the lowest point at 7797. The details are illustrated in the second column of Table I. The total number of the reviews on the Sina stock forum and East money stock forum is 1 930 592 after filtering and demonizing during the given period.

3.4 Sentiment Calculation: It is not utilized directly to explore the stock market trend, but select 8 sentiment indicators based on 51 sentiment indexes. The reasons include that the number of the features of market data and sentiment indexes needs to be balanced. Although we have already computed 51 sentiment indexes, the number of market data is around 10, so it is unfair for the market attributes to some extent. Next, too many variables are inclined to cause the problem of overfitting. Furthermore, we find that using 8 sentiment features achieves a better result in forecasting the SSE 50 Index than using all 51 indexes.

3.5 Prediction: First, we need to label the data according to the following equation:

Label = 1, Closet−1 < Closet −1, Closet−1 > Closet

where Closet denotes the close price of the SSE 50 Index, and Closet−1 stands for the close price on the previous day. Besides, 1 also means buy order as it indicates the increase, whereas −1 means sell order as it implies the decline. Next, we implement two experiments to predict the index movement direction. Experiment 1 is to use market data, which include opening price, closing price, high for the day, low for the day, trading volume in number of shares, trading volume in RMB, change in RMB, and change in percentage. A fivefold cross-validation approach is adapted to train an SVM model. Eventually, we find the proper parameters and the kernel functions to achieve the best performance. Panel A of Table III sheds light on the prediction results. For Experiment 1, the accuracy can be 79.96%, and we use RBF kernel function, C = 256, γ = 0.9942; for Experiment 2, the accuracy can be as high as 97.73%, and we employ RBF kernel function, C = 181.0193, γ = 0.005524.

IV. EXPERIMENT RESULTS AND DISCUSSION

AI is computed based on the stock points. For example, if we buy a stock at the price of 100 and sell it at 150, then we earn 50 stock points and AI is 50 stock points; after that, we short the equity at 150 and liquidate the position at 120, then we make 30 stock points and AI becomes 80 stock points. (EMD) is an estimate of the maximum losses average, based on a geometric Brownian motion assumption. MDD and EMD are regarded as indicators of mechanism is allowed and there are no market frictions. The prediction results of Experiment 1 and Experiment 2 are, respectively, utilized to compute AI and MDD. AI compared with the trend of the closing price of the SSE 50 Index, and highlight the MDD district of AI simultaneously. It can be seen from the line graphs that the AI of Experiment 2 (916.6264 stock points) is more than two times than that of Experiment 1 (404.8598 stock points). Moreover, although both the methods fail to detect the dramatic decline at first, Experiment 2 predicts the trend afterward, and is able to uncover the following rise. The sharp decrease is known as the Chinese stock market crash in 2015. Besides, the MDD of Experiment 2 is 0.3770, whereas the MDD of Experiment 1 is 0.4073. Similarly, the EMD of the next 30 days for Experiment 2 (0.1882) is also lower than that of Experiment 1 (0.2546).

Experiment 2 can help investors make higher profits with the same risk. Finally, we try to discover whether we can achieve a better performance based on the prediction results of Experiment 2. Hence a stop-loss order strategy is applied to limit the potential losses. We set the stop order to be 95 stock points, which means that we would stop to trade if 95 stock points had already been lost in a trading day. The strategy accomplishes a much better performance, and all of the measures that are displayed in the third row of Table IV improve significantly. From Figs. 9 and 10, we can point out that a stop-loss order can put an end to a losing period, but it cannot turn it into a win. However, the reduced loss also means an increase in a final AI. As a result, we can deduce that our approach can be of great benefit to investors if combined with a proper strategy, for example a stop-loss order strategy. In other words, the method is useful to decision-making processes that are pervasive phenomena of nature downside risk. Sharpe ratio (SR) is a way to gauge the performance of an investment by calculating the adjusted-risk return, which is defined as $\mathbf{SR = ra - rf\ \sigma a}$ where **ra** is the mean of the asset returns, **σa** is the standard deviation of the asset returns, and **rf** denotes the risk-free rate and set to be 0 in this paper. We have previously mentioned that the label 1 means buy order, whereas –1 means sell order. Accordingly, we follow the prediction results to buy or sell, then find if it is beneficial to support investment decisions and reduce financial risk. We postulate that short selling.

A graph representing the accumulated income (real price) and predicted price of the company, as shown in Fig 4.1

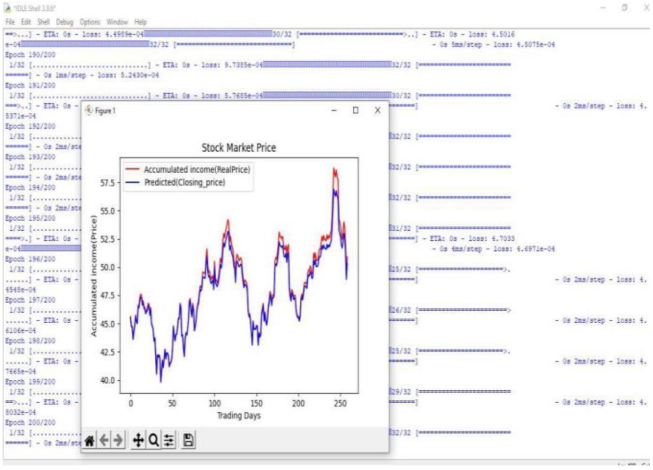

Fig 4.1 Graph of actual price and predicted price.

The below Fig shows the graph of actual and predicted price of another company and display the same along with mean square error value.

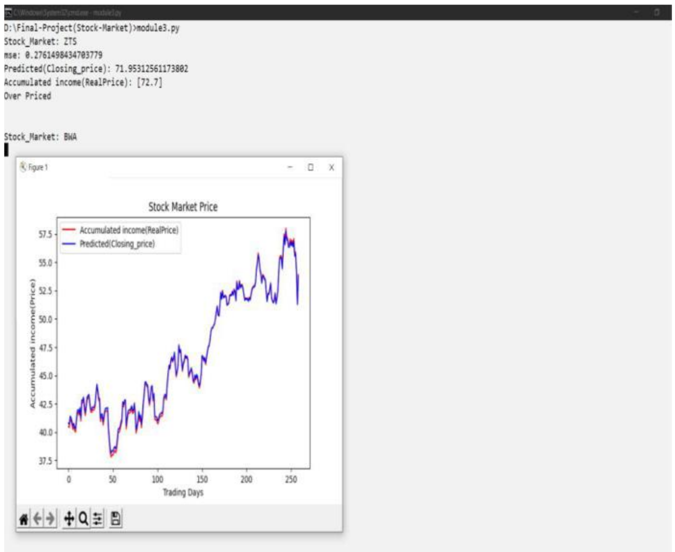

Fig 4.1 Graph of actual price and predicted price with mean square error value.

V. CONCLUSION

Establishing a forecasting framework to predict the prices of stocks. We processed stock data and tweets to predict the stock market prices used recurrent neural network to predict the stock price, with excellent results. Our proposed model has a better fitting degree and improved accuracy of the prediction results. Therefore, the model has broad application prospects and is highly competitive with existing models. Our future work has several directions. Our work has found that RNN and Sentiment Analysis have more predictive outcomes for price prediction than other methods. However, simply considering the impact of historical data on price trends is too singular and may not be able to fully and accurately forecast the price on a given day. Therefore, we can add data predictions related to stock-related news and basic information, so as to enhance the stability and accuracy of the model in the case of a major event.

REFERENCES

[1] Badre LABIAD, Abdelaziz Berrado, LoubnaBenabbou "INTELLIGENT SYSTEM FOR INTRADAY STOCK MARKET FORECASTING" 2019

[2] Shou-Hsiung Cheng "Forecasting The Change Of Intraday Stock Price By Using Text mining News Of Stock", Conference: International Conference on Machine Learning and Cybernetics, ICMLC 2010, Qingdao, China, July 11-14, 2010.

[3] Zhen Hu ,Jie Zhu , and Ken Tse "STOCK MARKET PREDICTION USING SUPPORT VECTOR MACHINE" 2013

[4] Noor Basha, Ashok Kumar P.S, P Venkatesh, "Early Detection of Heart Syndrome Using Machine Learning Technique," 4th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT) Pages 387-391.

[5] M. Umami, S. H. Adil, K. Raza and S. S. A. Ali, "Stock market prediction using machine learning techniques", 2016 3rd International Conference on Computer and Information Sciences (ICCOINS), pp. 322-327, 2016.

[6] K. Raza, "Prediction of Stock Market performance by using machine learning techniques", 2017 International Conference on Innovations in Electrical Engineering and Computational Technologies (ICIEECT), pp. 1-1, 2017.

[7] H. Gunduz and Z. Cataltepe and Y. Yaslan, "Stock market direction prediction using deep neural networks", 2017 25th Signal Processing and Communications Applications Conference (SIU), pp. 1- 4, 2017.

[8] H. L. Siew and M. J. Nordin, "Regression techniques for the prediction of stock price trend", 2012 International Conference on Statistics in Science Business and Engineering (ICSSBE), pp. 1-5, 2012.

[9] K. V. Sujatha and S. M. Sundaram, "Stock index prediction using regression and neural network models under non normal conditions", INTERACT-2010, pp. 59-63, 2010.

[10] Fama, E. 1970, 'Efficient Capital Markets: A Review of Theory and Empirical Work',Journal of Finance, vol. 25, pp. 383-417.