Share Market Analysis and Prediction

CHETAN BHAGAT
*Department of Information
Technology*
G.H.Raisoni(GHRCEM)
Pune, India

MAYURKUMAR BORSE
*Department of Information
Technology*
G.H.Raisoni(GHRCEM)
Pune, India

MANOJ CHAVAN
*Department of Information
Technology*
G.H.Raisoni(GHRCEM)
Pune, India

DNYANESHWAR JADHAV
*Department of Information
Technology*
G.H.Raisoni(GHRCEM)
Pune, India

MS. MINAXI DOORWAR
*Department of Information
Technology*
G.H.Raisoni(GHRCEM)
Pune, India

Abstract—Nowadays, the prediction of share market prices and conditions has become a major researched topic amongst the data scientists, investment bankers, and stock brokers. As, the behavior of share market is very nonlinear and volatile in nature, it makes a very high-risk investment. Consequently, a lot of researchers have came up with their efforts to forecast the share market and average movement. Researchers have used various methods in computer science and economics in their illustrate to gain a piece of this volatile information and make great fortune out of the share market investment. The approaches like data mining and machine learning approaches can incorporate into Business Intelligence (BI) systems to help users for decision support in many real-life applications. This paper presents the brief survey of application of machine learning in share market prediction and investigates various techniques for the share market prediction using like Artificial Neural Network (ANN) and Support Vector Machine (SVM). ANN is non- linear and non- parametric classifier which is viable for forecasting of share prices. Support Vector Machine focuses on marginal values rather than average values for the classification predicting model. The aim of this paper is to provide a review of the application of machine learning in share market prediction to determine what can be done in the future.

Keywords—Share Market, Machine Learning

I. INTRODUCTION

Predicting the future of share price has always been an adventurous and attractive task for the investors individuals. This kind of prediction becomes more fascinating when it involves money and risk like predicting share 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 share market is a very complex, volatile and non-linear dynamical system, share market prediction has become a tough challenge for researchers and investors. Lot of Research has been done on share market prediction by researchers of different fields including the business and computer science. Researchers have tried different approaches for share market prediction.

The attributes that makes a prediction model depends upon the factor on which market performance can be depending. Different types of methods have been developed to forecast the behavior of stock market prices based on previous historical data. As per recent survey Share prices prediction with conventional statistical methods has proven to be less effective because of nonlinear characteristics of financial time series. On the contrary, machine learning-based methods, such as Artificial Neural Networks (ANNs), support vector machine (SVM) ,linear regression, recent trends, LSTM model and data mining system offer useful tools for forecasting noise environments like share market.

II. RELATED WORK

A. Stock/Share Market

A place where shares of public listed companies are traded is known as Stock Market. It is a regulated financial and well organized market. In stock market, securities such as bonds, notes and shares are bought and sold at price and this prices governed by the forces of demand and supply. Stocks are bunches of shares. Stock market serves as, market where corporations, governments, municipalities and other incorporated bodies can raise their capital by taking saving of the investors into productive ventures. Also, Share market is a place where stock investor can sell their securities to other stock investors for cash. And this reduces the risk of investment and maintaining liquidity in the system.

B. Machine Learning Algorithms

This section continues Introduction to machine learning and evaluation of the machine learning techniques. Some of the algorithms we had considered are as follows:

Computer algorithms Studies that improve automatically through experience and by the use of data is known as Machine learning. It is part of Artificial Intilligence. This concept focuses on the computer development programs that can access given data and use it learn for themselves. The concept start 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 main goal of machine

learning is to allow the computers learn automatically without human interaction or assistance and adjust actions accordingly. Machine learning algorithms were categorized as unsupervised and supervised.

Flask Framework

Flask is written in Python and it is a micro web framework. It is classified as a micro-framework because Flask does not require particular tools or libraries that's why it is classified as micro-framework. Flask does not have database abstraction layer form validation or any other components where pre-existing third-party libraries provide common functions.

.COMPARISON OF LITERATURE

Sr no	Published Year	Methodology	Description
1	(2008) [1]	Studies Using Genetic Algorithms with Other Techniques to Analyze Stock Markets	Within the planned learning paradigm, a genetic rule is 1st want to choose input options for LSSVM learning. Then another GA is employed for parameter improvement for the LSSVM. Finally, the evolving LSSVM learning paradigm with the simplest feature set, optimum parameters, and a mixed kernel is employed to predict stock exchange movement direction in terms of historical knowledge series. For analysis functions, testing is completed mistreatment knowledge from 3 stock indices – the S&P five hundred, stock index Industrial Average, and also the big apple stock market Index. the whole knowledge set of monthly values covers the amount from Gregorian calendar month 1926 to Gregorian calendar month 2005 with a complete of 960 observations. Experimental results reveal that the planned evolving LSSVM will manufacture some prediction models that area unit a lot of simply taken as a result of they use a smaller range of prophetic options and area unit a lot of economical than different parameter improvement strategies.
2	(2009)	Studies Using Support Vector Machin	A unique study by Schumaker and Chen used an SVM in conjunction with matter analysis observing the impact of stories articles on stock costs. They developed a prognosticative machine learning
	es to Analyze Stock Markets		approach for monetary newspaper article analysis exploitation many completely different matter representations: Bag of Words, Noun Phrases, and Named Entities. Through this approach, they investigated an oversized range of economic news articles and stock quotes covering stocks listed on the S&P500 throughout a five-week amount from Oct twenty six, 2005 to Gregorian calendar month twenty-eight, 2005. They calculable a separate stock value twenty minutes when a newspaper article was free. exploitation AN SVM by-product specially tailored for separate numeric prediction, and models containing completely different stock-specific variables, they showed that the model containing each article terms and stock value at the time of article unleash provided the nearest estimate to the particular future stock value, identical direction of value movement because the future value, and therefore the highest come back employing a simulated trading engine.
3	(2009)	Studies Using Genetic Algorithms with Other Techniques to Analyze Stock Markets	A dynamic fuzzy model is projected by Chiu and Chen in combination with a SVM to explore stock exchange dynamics. The fuzzy model integrates input variables exploitation factors with variable degrees of influence. A GA adjusts the prestigious degree of every input variable dynamically. The SVM is then wont to predict exchange dynamics. A multi-period experiment is meant to simulate exchange volatility. The sixty-one input variables within the study embody exchange technical indicators, futures exchange technical indicators, and economic science variables. to judge the performance of the new integrated model, they compare it with ancient forecast ways.
4	(2011)	Studies Using Support	Yeh, Huang and Lee address issues that arise once mistreatment support vector regression to forecast stock

#	Year	Category	Description
		Vector Machines to Analyze Stock Markets	exchange values once handling kernel operate hyperparameters. Typically, a hyperparameter may be a parameter whose price is ready before the training method begins. In their system, blessings from totally different hyperparameter settings will be combined and overall system performance will be improved. They develop a two-stage multiple-kernel learning formula by incorporating consecutive token optimization and therefore the gradient projection technique.
5	(2013)	Studies Using Artificial Neural Networks to Predict Stock Market Values)	Chavan and Patil contribute to our understanding of ANN stock exchange prediction by measure totally different model input parameters found in 9 printed articles. They plan to realize the foremost necessary input parameters that manufacture higher model prediction accuracy. supported their survey, they realize that almost all cc techniques build use of technical variables rather than elementary variables for a specific stock worth prediction, whereas political economy variables are principally wont to predict exchange index values. additionally, hybridized parameters manufacture higher results in comparison with the employment of solely one input variable kind. .
6	(2015)	Studies Using Hybrid or Other AI Techniques to Analyze Stock Markets	The study by Patel, Shah, Thakkar and Kotecha compares four Indian stock exchange prediction models: ANN, SVM, random forest, and naive-Bayes with 2 approaches for model input. the primary approach for input files involves computation of 10 technical parameters victimization stock commerce data (open, high, low and shut prices), whereas the second approach focuses on representing these technical parameters as trend settled knowledge. They assess the accuracy of every of the prediction models for every of the 2 input approaches. analysis is predicated on knowledge from 2 stocks and 2 stock value indices -
			CNX smashing (50 of the most important stocks found on the National exchange (NSE) of India) and therefore the S&P Mumbai exchange (BSE) Sensex for the amount Jan 2003 to December 2012.
7	(2016)	Studies Using Hybrid or Other AI Techniques to Analyze Stock Markets	Dash and Dash introduce a unique decision network using a computationally economical practical link artificial neural network (CEFLANN) and a rule set to a lot of effectively generate trading choices. They read the stock mercantilism call as a classification downside with 3 potential values – get, hold or sell. The CEFLANN network employed in the choice network produces a collection of continuous mercantilism signals by analyzing the nonlinear relationship that exists between some well-liked technical indicators. The output mercantilism signals are accustomed track trends and to provide mercantilism choices supported that trend mistreatment mercantilism rules. this can be a unique approach centered on profitable stock mercantilism choices through integration of the training ability of the CEFLANN neural network with the technical analysis rules.
8	(2017)	Studies Using Artificial Neural Networks to Predict Stock Market Values	Chong, Han and Park analyze deep learning networks for stock exchange analysis and prediction. Deep learning networks extract options from an oversized set of information while not hoping on previous information of predictors that makes it helpful for prime frequency stock exchange prediction. they supply Associate in Nursing objective assessment of each the benefits and disadvantages of deep learning algorithms for stock exchange analysis and prediction. exploitation high-frequency intraday stock returns as computer file, they examine the results of 3 unattended feature extraction methods—principal element analysis,

			autoencoder, and also the restricted physicist machine—on the network's overall ability to predict future market behavior.

III. PROPOSED METHODOLOGY

A. Yahoo Data

Stock market prediction seems a complex problem because there are many factors that have yet to be addressed and it doesn't seem statistical at first. But by proper use of previous data for machine learning techniques to make appropriate assumptions and the current data and train the machine to learn from it.

We used dataset for analysis was picked up from Yahoo Finance. The dataset consisted of approximately 1 lakh records of the required stock prices and other relevant values. At certain time intervals for each day of the year the data returns the stock prices. It made up of various sections namely low, symbol, open, date, high and volume. The data of only one company is considered for the purpose of simulation and analysis. First All the data is read which is available in a file of csv format and transformed into a data-frame using the keras, tensorflow in Python. From this, the data for one particular company is extracting by segregating data on the basis of the symbol field. Following this normalization of the data was performed through usage of the scikit library in Python and the data was divided into two parts i.e testing and training sets. The test set was kept as 20% of the available dataset.

Machine learning has many models but this paper focuses on four of the most important amongst them and made the predictions using these.

B. Regression Based Models

In general, the Regression based Model is used for predicting continuous values through some given independent values. Regression uses a given linear function for predicting continuous values:

$$V = a + bK + error \qquad [1]$$

Where, V is a continuous value; K represents known independent values; and, a, b are coefficients. Fig. 1 Flow Chart for Regression Based Mode.

The paper utilizes for predicting correct values by minimizing the error function as given in Fig.1 the gradient descent linear regression algorithm is uses. Linear Regression algorithm as governed by equation [1] is performed on the data and then the relevant predictions are made. Low, open, high, close and volume are he factors considered for the regression.
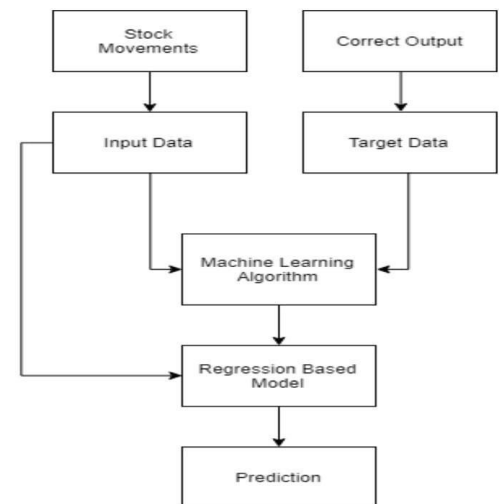

Fig. 1 Flow Chart for Regression Based Model

C. Long Short Term Memory(LSTM) Network Based Model

Long Short Term Memory is the advanced version of Recurrent-Neural-Networks (RNN) where the information belonging to previous state persists. These are different from RNNs as They involve long term dependencies and RNNs works on finding the relationship between the recent and the current information so this are different from RNNs. it is indicates that the interval of information is relatively smaller than that to LSTM.

The predictions depends on large amounts of data and are generally dependent on the long term history of the market it is the main purpose behind using LSTM model in stock market prediction. LSTM regulates error by giving an aid to
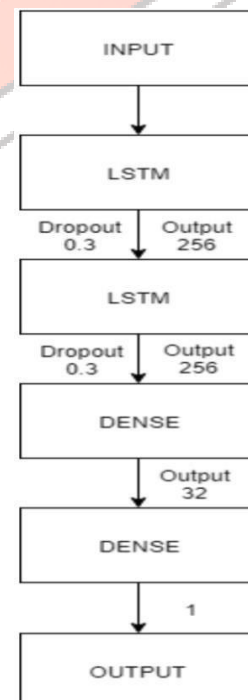

Fig 2 Flow Chart for LSTM

the RNNs through retaining information for older stages making the prediction more accurate Thus proving itself as much more reliable compared to other methods.

Stock market involves processing of huge data, the gradients with respect to the weight matrix may become very small and may degrade the learning rate. This corresponds to the problem of Vanishing Gradient. LSTM prevents this from happening. The LSTM consists of a remembering cell, input gate, output gate and a forget gate. The cell remembers the value for long term propagation and the gates regulate them.

D. Autoregressive Integrated Moving Average (ARIMA)

ARIMA model is short form of Autoregressive Integrated Moving Average. It is a class of models that 'explains' a given time series based on its own past values, that is, its own lags and the lagged forecast errors, so that equation can be used to forecast future values. It converts non-stationary data into stationary data.

ARIMA is divided into three parts i.e. Auto Regression(AR), Integrated(I) and Moving Average(MA). Auto Regression shows a regresses on its own lagged, or prior, values which is changing variables. Integrated shows he differencing of raw observations to allow for the time series to become stationary, i.e., data values are replaced by the difference between the data values and the previous values. Moving Average incorporates the dependency between an observation and a residual error from a moving average model applied to lagged observations.
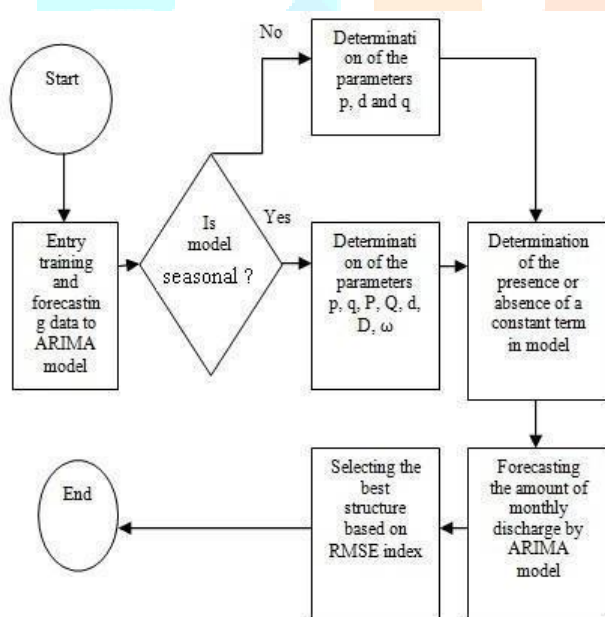
ARIMA consists of three parameters which is nothing but p, d and q. Where p stands for lag observation in model also known as the lag model, d stands for the number of times that the raw observations are differenced; also known as the degree of differencing and q stands for the size of the moving average window; also known as the order of the moving average.

Firstly we import the required libraries like numpy ,pandas etc. then we load data and convert it into time series data. Plotting of time series data is important to see the trend and seasonality in data. After that splitting the data into training and testing for evaluation purpose.

IV. CONCLUSION

A. This paper provides a review on machine learning techniques to predict share prices, In This review we have surveyed an articles that have used artificial neural network (ANN) and support vector machine (SVM) as a forecasting model for stock prices. We also used LSTM model algorithm, Linear Regression algorithm. LSTM and Regression, on the Yahoo finance dataset. Both the techniques have shown an improvement in the accuracy of predictions, thereby yielding positive results. Use of recently introduced machine learning techniques in the prediction of stocks have yielded promising results and thereby marked the use of them in profitable exchange schemes. It has led to the conclusion that it is possible to predict stock market with more accuracy and efficiency using machine learning techniques.

V. REFERENCES

[1] National Stock Exchange
https://www.nseindia.com/
[2] Bombay Stock Exchange
https://www.bseindia.com/
[3] Stock market prediction using machine learning approach
https:// www.ijies.net/finial-docs/finial-pdf/220519CS11/
[4] API Keys
https:// www.alphavantage.co2
[5] Stock Market Prediction Using Artificial Neural Networks
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.134.1114&rep=rep1&type=pdf/
[6] Stock market prediction using supervised machine learning techniques
https://www.computer.org/csdl/proceedings-article/csde/2020/09411609/1taFeSM9YBy/
[7] M. Usmani, 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)
https:// scialert.net/fulltext/?doi=jai.2018.48.54/
[8] "Regression techniques for the prediction of stock price trend," 2012 International Conference on Statistics in Science, Business and Engineering (ICSSBE)
https://ieeexplore.ieee.org/document/6396535/
[9] LSTM and Regression Methods for Stock Market Prediction
https://www.ijcrt.org/papers/IJCRT2005021.pdf
[10] Stock transaction prediction modeling and analysis based on LSTM
https://www.researchgate.net/publication/326050300_Stock_transaction_prediction_modeling_and_analysis_based_on_LSTM
[11] Stock Price Prediction Using the ARIMA Model
https://ieeexplore.ieee.org/document/7046047