# Prediction of Stock Market Using AI & MI Algorithms

Mrs Bhagya M[1], Akshay Kumar T A[2], Anjali N E[3], Naveen kumar N[4], Sneha B[5]
[1,2,3,4,5]*Computer Science and Engineering, M.S. Engineering College, Bangalore, Karnataka*

**Abstract: The nature of stock market movement has always been ambiguous for investors becauseof various influential factors. This study aims to significantly reduce the risk of trend prediction with machine learning and deep learning algorithms. Four stock market groups, namely diversified financials, petroleum, non-metallic minerals and basic metals from Tehran stock exchange, are chosen for experimental evaluations. This study compares nine machine learning models (Decision Tree, Random Forest, AdaptiveBoosting (Adaboost), eXtreme Gradient Boosting (XGBoost), Support Vector Classifier (SVC), Naïve Bayes, K-Nearest Neighbors (KNN), Logistic Regression and Artificial Neural Network (ANN)) and two powerful deep learning methods (Recurrent Neural Network (RNN) and Long short-term memory (LSTM). Ten technical indicators from ten years of historical data are our input values, and two ways are supposed for employing them. Firstly, calculating the indicators by stock trading values as continuous data, and secondly converting indicators to binary data before using. Each prediction model is evaluated by three metrics basedon the input ways. The evaluation results indicate that for the continuous data, RNN and LSTM outperform other prediction models with a considerable difference. Also, results show that in the binary data evaluation,those deep learning methods are the best; however, the difference becomes less because of the noticeable improvement of models' performance in the second way.**

## I.INTRODUCTION

The task of stock prediction has always been a challenging problem for statistics experts and finance. The main reason behind this prediction is buying stocks that are likely to increase in price and then selling stocks that are probablyto fall. Generally, there are two ways for stock market pre- diction. Fundamental analysis is one of them and relies on a company's technique and fundamental information like mar-

The associate editor coordinating the review of this manuscript and approving it for publication was Zhe Xiao ket position, expenses and annual growth rates. The second one is the technical analysis method, which concentrates on previous stock prices and values. This analysis uses historicalcharts and patterns to predict future prices [1], [2].

Stock markets were normally predicted by financial experts in the past time. However, data scientists have startedsolving prediction problems with the progress of learn- ing techniques. Also, computer scientists have begun using machine learning methods to improve the performance of prediction models and enhance the accuracy of predictions. Employing deep learning was the next phase in improving

## II. EXISTING SYSTEM

A Financial forecast can be elaborated as a forecast regarding the future business circumstances that are expected to affect a company, organization, or a country. A financial forecast visualizes the movements in relevant historical data and then projects these movements in order to help the decision makers by providing information regarding the forthcoming financial status of the company. Simply we can say that, a financial forecast is a business plan or budget for a business. It is basically considered as an estimate of two vital forthcoming financial outcomes of a business–the projected revenue and the costs. Prediction of the financial state of a business is never an easy task; with most of the forecasts go wrong. But still it is a better idea to have an educated guess about the future than to not forecast at all, since ''Best'' educated guesses about future are more valuable for purpose of planning and budgeting.

## III. PROPOSED SYSTEM

We propose the concept of time series analysis and forecasting in the perspective of Indian economy. The major downfall of the Indian rupee in the recent times has led to the critical need for stock market

prediction so as to safeguard the interest of the investors. This system tries to build an efficient ARIMA model to predict the Indian stock market volatility. The publicly available time series data of Indian stock market has been used for this project. The predicted time series has been compared with the actual time series, which shows roughly a deviation of 5% mean percentage error for both Nifty and Sen sex on average.

## IV. SYSTEM ARCHITECTURE

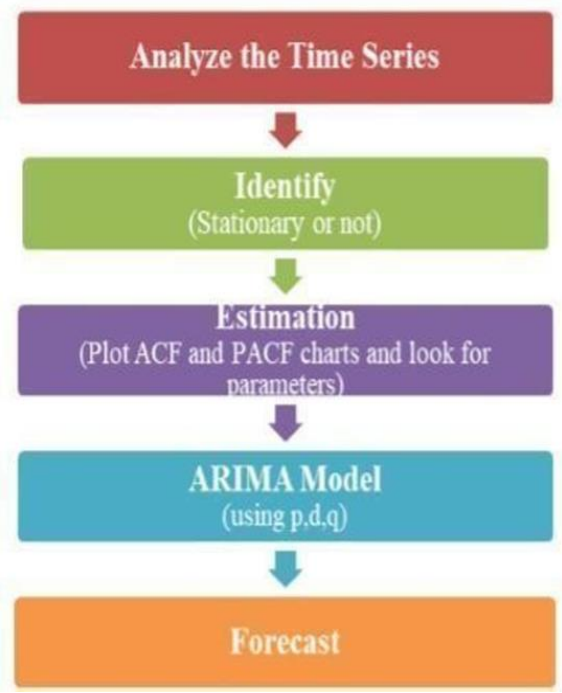



Fig 4.1 System Architecture

The stock market prediction neural network architecture comprises of three layers, input layer, hidden layer and output layer. There are three neurons in the input layer, two neurons in the hidden layer and one neuron in the output layer. Sigmoid transfer function is used for hidden layer and output layer neuron.

## V. METHODOLOGY

The idea is to weigh out the importance of recent and older data and determine which parameters affect the "current" or "next" day prices the most. The machine learning model assigns weights to each market feature and determines how much history the model should look at to predict future stock prices.

## VI. IMPLEMENTATION

Implementation is the stage of the project where the theoretical design is turned out into a working system. Thus, it can be considered to be the most critical stage in achieving a successful new systemand in giving the user, confidence that the new system will work and be effective. The implementation stage involves careful planning, investigation of the existing system and its constraints on implementation, designing of methods to achieve changeover and evaluation of changeover methods.

Modules:

- Data Collection
- Stationarity Of A Time Series
- Determination Of ACF And PCF
- Price Forecasting

Module Description:

Data Collection:

In this project began with the analysis of Indian stock market data related to Sensex and Nifty. The publically available Stock market data sets contain historical data and Live data about all the stocks has been collected from the NSE. The dataset specifies the ‘‘opening price, lowest price, closing price, highest price, adjusted closing price and volume’’ against each date. The historical data of the Indian stock market collected through a span of five years beginning from ‘‘January 2017 to today date’’ has been taken into consideration for this work. The data has been divided into two parts–‘‘the training part and the testing part’’. The ‘‘training part’’ from the time series data is used for formulation of the model while the ‘‘testing part’’ is used for the validation of the proposed model.

Stationarity Of A Time Series

A time series needs to be lacking trend and seasonality, in order to be stationary. Such type of time series are characterized by having a constant variance and constant mean over a given period of time. The ‘‘trend and seasonality’’ component may affect a time series at different instants. As ARIMA model takes into account, the earlier values of the series to model its prediction, so modeling a steady series with regular properties involves little in security. In order to design a model that is efficient

in predicting future values of series, the primary time series has to be Stationary one. There are certain tests that assist in checking whether the series is stationary or not. Some of these include ‘‘Auto-correlation function (ACF)’’, ‘‘Partial auto-correlationfunction (PACF)’’. A ‘non-stationary time-series’ needs to be corrected by means of differencing. An easy way to alter a non stationary time series into stationary one needs to compute the differences between consecutive observations.

Determination Of ACF And PACF

An important step while selecting the model is the determination of ideal parameters for the model. Plotting the ACF and PACF against the consecutive time lags for the series is one simple approach to choose the parameters of the model. ‘‘ACF and PACF’’ are statistical methods that signify the relation between the observations in a time series with one another [11], [12]. They aid in defining the parameters of ‘‘AR and MA’’ terms. The general form of ACF is as:

$$\frac{\text{Covariance}(X_t, X_t - h)}{\text{Std. dev}(X_t).\ \text{Std.dev}(X_t - h)} = \frac{\text{Covariance}(X_t, X_t - h)}{\text{Variance}(X_t)}$$

The above equation gives the ACF between xt and xt−h, Where xt ‘denoted the value of time series at time t for h values=1,2,3...’ etc. In case of ‘AR models’, ACF for time series will shrink exponentially, so, PACF is implemented to identify the order of‘p’.‘‘Partial auto correlation plots (PACF)’’ signify the correlation between a variable and its lags and are very beneficial for describing the order of the ‘AR(p) model’. The common form is of PACF is as:

$$\frac{\text{Covariance}(y, X3|X1, X2)}{\text{Variance}(y|X1|X2)\text{Variance}(X3|X1, X2)}$$

where: ‘‘y = response variable’’, ‘‘X1, X2, and X3 are the predictor variables’’.

Price Forecasting

The first thing to do for analysis of the time series data requires plotting the data as below. The next important thing that needs to be done involves the decomposition of the time series data intoits essential constituents. Since the data under consideration is the time series data from Indian stock market, the components of time series data related to stock market. Decomposing the time series data helps in revealing a lot of hidden patterns inside the time series. A time series in general consists of four components. We are using ARIMA model to forecast the stock price for another 10 days.

## VII. DATA FLOW DIAGRAM

A data-flow diagram is a way of representing a flow of data through a process or a system (usually an information system). The DFD also provides information about the outputs and inputs of each entity and the process itself. A data-flow diagram has no control flow — there are no decision rules and no loops. Specific operations based on the data can be represented by a flowchart

Level:0

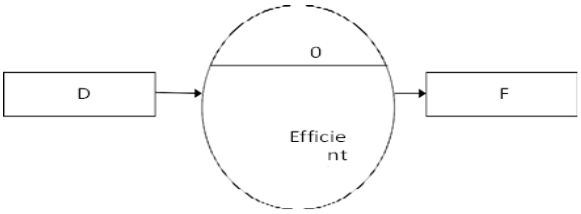



Fig 7.1level 0

Level 0 Describes the overall process of this project. we are passing NSE datasetas a input the system will forecast the stock values using ARIMA model

Level: 1

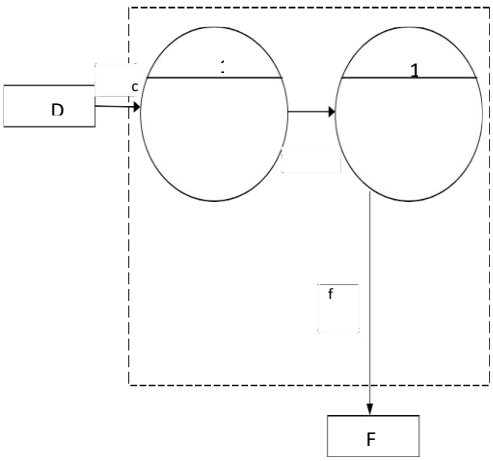



Fig 7.2level 1

Level 1 Describes the first stage process of this project. we are passing NSE data set as a input the system will detect the features.

Level 2:

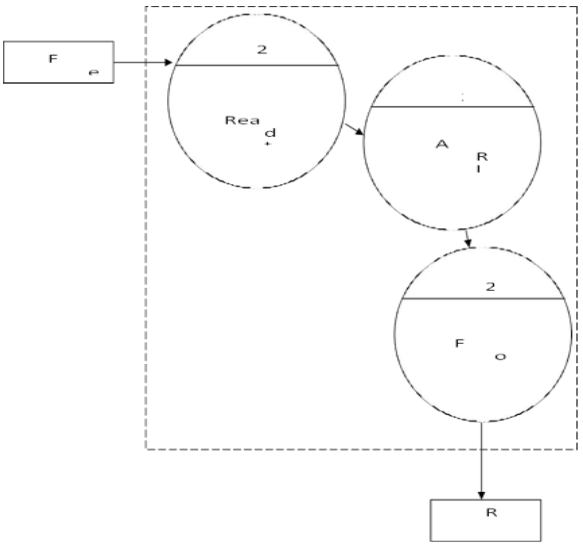
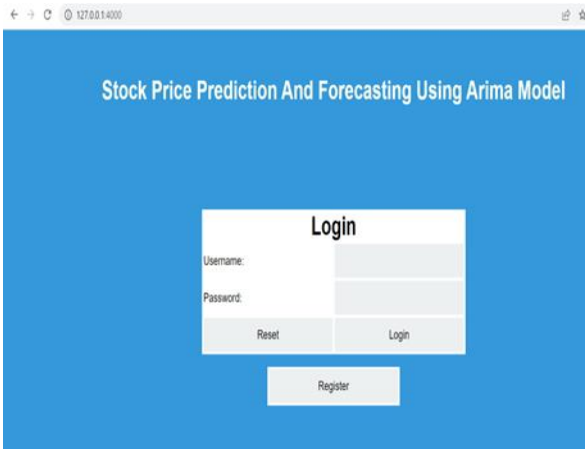
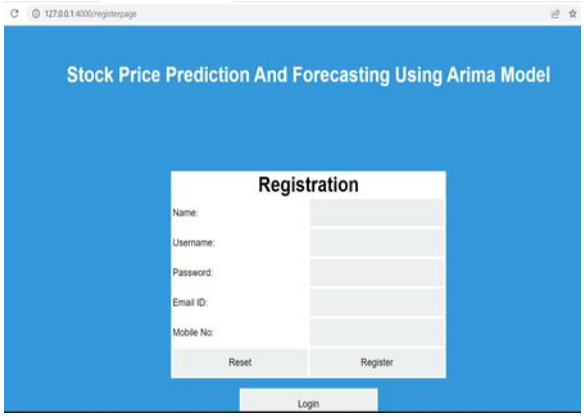


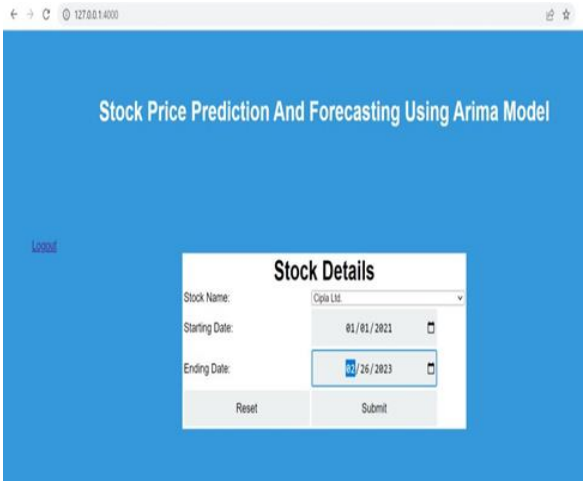

Fig 7.3level 2

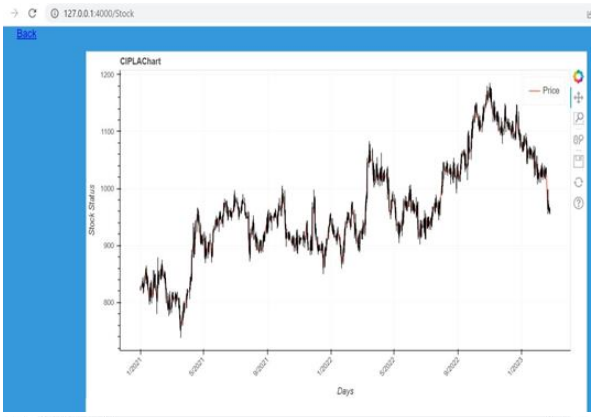
Level 2 Describes the final stage process of this project. we are passing extracted features from level by applying ARIMA model system will predict the stock valuesand forecast the values.

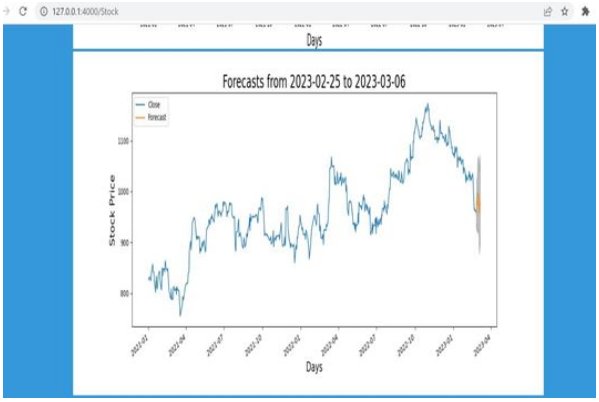
SNAPSHOTS




CONCLUSION

The purpose of this study was the prediction task of stock market movement by machine learning and deep learning algorithms. Four stock market groups, namely diversified financials, petroleum, non-metallic minerals and basicmetals, from Tehran stock exchange were chosen, and the dataset was based on ten years of historical records with ten technical features. Also, nine machine learning models (DecisionTree, RandomForest, Adaboost, XGBoost, SVC, Naïve

Bayes, KNN, Logistic Regression and ANN) and two deep learning methods (RNN and LSTM) wereemployed as predictors. We supposed two approaches,for input values to models continious data and binary data and we employed

REFERENCE


[1] J. J. Murphy, *Technical Analysis of the Financial Markets: A Comprehen- sive Guide to Trading Methods and Applications*. Penguin, 1999.

[2] T. Turner, *A Beginner's Guide To Day Trading Online*, 2nd ed. New York, NY, USA: Simon and Schuster, 2007.

[3] H. Maqsood, I. Mehmood, M. Maqsood, M. Yasir, S. Afzal, F. Aadil,
M. M. Selim, and K. Muhammad, ''A local and global event sentiment based efficient stock exchange forecasting using deep learning,'' *Int. J. Inf.Manage.*, vol. 50, pp. 432–451, Feb. 2020.

[4] W. Long, Z. Lu, and L. Cui, ''Deep learning-based feature engineer- ing for stock price movement prediction,'' *Knowl.-Based Syst.*, vol. 164, pp. 163–173, Jan. 2019.

[5] J. B. Duarte Duarte, L. H. Talero Sarmiento, and K. J. Sierra Juárez, ''Evaluation of the effect of investor psychology on an artificial stock market through its degree of efficiency,'' *Contaduría y Administración*, vol. 62, no. 4, pp. 1361–1376, Oct. 2017.

[6] Lu, Ning, *A Machine Learning Approach to Automated Trading*. Boston, MA, USA: Boston College Computer Science Senior, 2016.

[7] M. R. Hassan, B. Nath, and M. Kirley, ''A fusion model of HMM, ANN and GA for stock market forecasting,'' *Expert Syst. Appl.*, vol. 33, no. 1, pp. 171–180, Jul. 2007.

[8] W. Huang, Y. Nakamori, and S.-Y. Wang, ''Forecasting stock market movement direction with support vector machine,'' *Comput. Oper. Res.*, vol. 32, no. 10, pp. 2513–2522, Oct. 2005.

[9] J. Sun and H. Li, ''Financial distress prediction using support vector machines: Ensemble vs. Individual,'' *Appl. Soft Comput.*, vol. 12, no. 8, pp. 2254–2265, Aug. 2012.

[10] P. Ou and H. Wang, ''Prediction of stock market index movement by ten data mining techniques,'' *Modern Appl. Sci.*, vol. 3, no. 12, pp. 28–42, Nov. 2009.

[11] F. Liu and J. Wang, ''Fluctuation prediction of stock market index by legen- dre neural network with random time strength function,'' *Neurocomputing*, vol. 83, pp. 12–21, Apr. 2012.

[12] C.-F. Tsai, Y.-C. Lin, D. C. Yen, and Y.-M. Chen, ''Predicting stock returns by classifier ensembles,'' *Appl. Soft Comput.*, vol. 11, no. 2,pp. 2452–2459, Mar. 2011.