# Prediction and Clustering Techniques used in the Development of Stock Forecasting Model

**Shailaja K P, Manjunath M**

***Abstract*: With the advancement in science and technology more and more real time data is being accumulated in the digital repositories. One such highly accumulating data is stock market data. Prediction of stock market data and its analysis is a challenging task as it is a highly sophisticated time series data vulnerable to sudden changes. The data and its relevance in the real world has attracted the interest of many researchers. Research literature provides many contributions by eminent researchers for analyzing and developing models for stock data. In this research paper an effort has been made present two things. One, building a stock prediction model using artificial neural network using different learning functions and we found that the different learning functions produces different results. In our experiment we have achieved a highest accuracy of 94.55%. The second thing we are trying to do in this article is to provide a detailed information on various techniques used for stock data prediction. More than 50 articles have been studied to present the significant contribution of the researchers. The articles are categorized into two main sections namely prediction methodologies used for stock data analysis and clustering methodologies used for stock data analysis. The sections, prediction methodologies used for analysis and clustering methodologies used for analysis are further explored into eight and four sub categories respectively depending upon the methodologies used. After presenting the detailed analysis, we have also highlighted the research gaps existing in the methodologies discussed in this paper.***

***Keywords*: Data Mining, Stock Market, Prediction, Forecasting, Stock data.**

.

## I. INTRODUCTION

In this 21st century, each and every aspect of our life is getting digitalized. Every second an enormous amount of data is being generated from every aspects of our life due to more and more digitalization, every aspects of our life are generating a huge amount of data every second. This huge data can be converted to a useful and profitable only if one knows how to do it. In this regard data mining is playing a very vital role. By applying traditional and advanced data mining techniques, we can convert any raw data into gold. Data mining has been playing an important role for some time now in many fields either in solving many real world problems or providing a decision making support system to improvise our life. For instance, sensor networks' main problem is how to utilize the available limited energy to improve the life span of the network. In this regard data mining has provided many strategies for sensor networks such as identification of optimal routing path for data transmission [1], data prediction [2], to reduce the energy consumption. The major contribution of data mining is in the medical field where data mining techniques can be used for early detection and prediction of life threatening diseases such as cancer in particularly breast cancer [3], [4], [5]. Data mining applications are not just limited to science and technology but also include finance related areas. Traditionally, people related to finance were solely dependent on traditional statistical methods for analysis of stock data. But the digitalization of the world has forced the people to use data mining techniques in almost all fields for analysis of data and to provide a decision support system. And finance related area such as stock market is not exempted from it. Stock market is one place where gigantic amount of financial data is generated every second during the trading hours. This gigantic amount of data can be converted into useful knowledge using which one can make fortune. The analysis and prediction stock market data, identifying the reasons behind the movement of stock price, the factors influencing the stock values, all these are very complicated, difficult, and challenging task. However, the reward is also very attractive and profitable because of which many individuals and institutes are spending their time and effort to build the model to understand the stock market and make fortune out of it. Many researchers have proposed many techniques and these techniques can be categorized according to their nature such as analysis of stock data based prediction methodologies namely- artificial neural networks, convolutional neural network, decision support system, hidden markov models, recurrent neural network, support vector machines, support vector regression, and clustering methodologies namely- filtering, fuzzy, K-means. This article is one effort to provide the detailed analysis of different methodologies used for development of stock data prediction models.

## II. LITERATURE SURVEY

### A. Prediction methodologies used for analysis

#### 1. Analysis of stock data using Artificial Neural Network

Human brain has more capabilities than we think. And we are yet to find ways to use our brain more efficiently.

**Revised Manuscript Received on January 5, 2020**

***** Correspondence Author

**Shailaja K P ***, Assistant Professor, Department of Computer Science, shailaja.mca@bmsce.ac.in B.M.S College of Engineering, Bangalore. Email: shailaja.mca@bmsce.ac.in

**Manjunath M**, Assistant Professor, Department of Computer Science, R V College of Engineering, Bangalore. Email: manjunathm@rvce.edu.in

One of the ways is to simulate the brain activity using computers. This is where the artificial neural network (ANN) contribute significant role. ANN mimics the working style of a brain which also has the capability in identifying the patterns and learning it. This makes the ANN more efficient compared to traditional statistical methods.

Predicting an approximate price of a stock is very difficult but if one succeeds in it, it can turn out to be a fortune. ANN has been used in many ways in developing prediction models. One such model is Bayesian regularized ANN [6]. The authors used it in order to predict the closing price of a stock for the trading day using both stock price and technical indicator. The model was able to minimize the over fitting and over-training.

An adaptive model which is Computationally Efficient Functional Link ANN (CEFLANN) was proposed by [7] for the prediction of Indian stock exchange indices. For training the network, the model used least mean square error (LSM) and learning function. The excellence of the predicted data compared to the actual data is estimated using mean absolute percentage error (MAPE). Technical indicators and historical data were used in this model for prediction of stock data for the following day. The CEFLANN model also used differential evolution technique to optimize the weight used in the model to increase the overall accuracy.

Particle Swarm Optimization (PSO) was used in [8] in an integrated functional link interval type 2 fuzzy neural network for construction of prediction model for prediction of stock data. For testing purpose, the model was used to predict the various market indices such as BSE, S and P 500. Here, for evaluation purposes MAPE and root mean square error (RMSE) were used.

Feature selection is a very important stage as this acts as the foundation of the model. The models use the features to accomplish the task. To predict the stock data various factors can be considered but which one are important are decided by the feature selection methods but there are hundreds of factors which impact the stock data. Deciding which features to use is a very difficult task. In this regard [9] have used three dimensionality techniques such as the Principal Component Analysis (PCA), fuzzy robust PCA, kernel based PCA to decrease the dimensionality of the data. Once the features are selected, ANN was used to predict the stock data.

ANN was studied thoroughly in [10] to build stock prediction model. For experiment purpose, the authors used the model for prediction of stock data of an Iranian tractor company.

The stock prediction models are often complex in nature and use many inputs. But the idea is to have a prediction model with simple structure and minimum number of inputs and this is achieved in [11]. Here, in the first step the factor that has the most impact on the model is identified using stepwise regression analysis (SRA). In the second step, data is distributed among number of clusters using self-organizing map (SOM). In the final phase, the clusters formed in the second step are fed to the genetic fuzzy system (GFS) as an input. MAPE was used to calculate the performance of the model.

The fusion of machine learning technique for development of prediction model was proposed by [12]. In the first stage the proposed model used Support Vector Regression (SVR). Later, SVR is combined with Random Forest (RF), and ANN in the second stage. This resulted in development of three fusion models namely SVR-ANN, SVR-Random Forest, and SVR-SVR.

The impact and efficiency of the ANN in building a stock prediction model was thoroughly investigated in [13]. Another researcher [14] also investigated ANN contribution in building stock prediction model. In particularly, the researcher investigated the dynamic ANN, multilayer perceptron (MLP), generalized autoregressive conditional heteroscedasticity (GARCH).

Recently the impact of learning functions on artificial neural network was investigated in [15].

### 2. Analysis of stock data using Convolutional Neural Network

Intraday trading is where many day traders make money. In order to increase the profit margin one needs to make the right decision at a very quick time. To make the right decision one needs the accurate report of the events going on at that moment which could make an impact. But analyzing huge amount of data quickly is very difficult. In this regard a prediction model was built using Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN) and the model used the S and P 500 index data to test the accuracy of the model [16].

A structure including both CNN and Long Short Term Memory was used in forecasting the stock market in [17]. This model used the publicly available data for training and testing the model.

So far only one or two data source were used to analyze the information and use it for forecasting the stock data. However, data which could impact the stock could come from multiple sources and hence it becomes absolutely necessary to develop a model which can extract the relevant information from multiple sources and utilize the extracted information to forecasting the data. Such a technique was used in [18] which used RNN for development of the prediction model.

### 3. Analysis of stock data using Decision Support System

The stock data prediction is very complicated but profitable task. The predictions of stock data are basically done in two ways. One is to develop the prediction model and second way is to build a decision support system. Basically the decision support systems gives buy or sell signal to the user based on which user can take decision to maintain their stock portfolio. One such example of decision support system is [19] where the author used the oscillation box along with two support vector machine. The support vector machines were used in this intelligent model for finding the upper and lower bound of the price. Once the lower and upper bounds are estimated, a trading strategy are executed based on the pattern.

### 4. Analysis of stock data using Hidden Markov Model

Hidden Markov Model (HMM) is best suited for developing the model and analyzing the results related to time related phenomena.

One such application is analysis of stock market data which is a time dependent data such as time series data. Using Hidden Markov Model along with Principal Component Analysis, a prediction model was developed [20]. In this model, PCA is used to identify the most relevant technical factor which could be used for forecasting of Indian stock market data. Once the most relevant technical indicators are identified, Hidden Markov Model is utilized to develop the stock market prediction model. The experimental study showed that the prediction model built using Hidden Markov Model along with the Principal Component Analysis produced less than 2 percent in-terms of Mean Absolute Percentage Error.

The Hidden Markov Model's ability to deal with the time series data was used to develop a Posterior Hidden Markov prediction model in [21]. In this model, the stock price, highest and lowest price of the stock for the intraday is utilized to predict the price for the following day.

## 5. Analysis of stock data using Neural Network

The NN is one among the best ways to identify the hidden pattern and learn them. This is made possible only because the neural networks follow the same principle or methodology as the brain. Hence, one can say that neural networks are computer simulated versions of brain. Many researchers have used the neural network for development of stock prediction model and in this section we will describe few as follows.

A NN based model was utilized to identify the trends of the currently following stocks [22]. The article addresses a new mechanism of increasing the accuracy of the prediction. The model was designed to allow multiple hidden layers involving random connections and evolutionary algorithms were used to optimize the training weights. A sine function in place of sigmoid function was used as an activation function for training the NN. The efficiency of the model was validated by performing repeated experiments to find the variation between the predicted values and the actual values and based upon the predicted accuracy the over fitting issue was investigated. Finally a comparative study was conducted where the outcome of the proposed model was matched with other existing models. It was witnessed that the model gave higher result than its counter parts.

An innovative NN model for predicting stock market data was introduced by [23].The main aim was to demonstrate the impact of internet of things related to multimedia for stock analysis. In order to carry out real-time data analytics and visualization of prediction results dataset was collected from live stock market. Generally it was observed in the state of art that traditional neural networks may incorrectly predict the stock rate because of random initialization of weights and thus leading to incorrect predictions. In order to overcome this drawback two new models namely **stock vector** and **long short-term memory neural network model (LSTM)** were demonstrated. Stock vector model is a deep learning based model where **word vector** concept was used. In this model instead of taking a single stock index as an input, multiple indices from the historical data were considered. The prediction model was named as LSTM. For vectoring and forecasting the values, this model included the automatic encoder. The accuracy of the respective models for Shanghai A-shares composite index and for individual stocks was found to be 57.2% and 56.9%.

A new system called Wave Analysis Stock Prediction (WASP) was developed by [24] to forecast the future stock rate. Based on the Elliott wave theory (EWT), a neuro-fuzzy system was designed by the authors. The EWT has been used earlier in many applications and is found to be useful in solving forecasting based problems. The authors make use of neuro-fuzzy logic to forecast the stock price and yielded promising results.

A significant research contribution utilizing daily stock data, bond data, and currency data specific to nearly around 39 nations was proposed by the authors in [25] to predict the market crash. The works showcases the transmission mechanisms of crash events across the stock market and its effect from bond and currency markets. The authors make use of two approaches namely Boosting and Deep learning. In Boosting of bag of various machine learning algorithms like SVM, neural networks, random forests, etc were used. In deep learning based approach a sophisticated deep neural network was built to carry out the prediction task. The main aim was to come up with a mechanism capable of predicting the probability of any market crash. The main variables considered for the forecasting of crash event were selected based upon the returns and volatility of the market. Machine learning algorithms were used to select the relevant features and bootstrap sampling was used to ensure that the data is properly fitted. The deep neural networks yielded in higher prediction accuracy than any other neural networks. The experimental results prove that there exists a strong relation between stocks, bond and currency of a nation. The results also reveal that stock market crises incline to show the persistence. The authors claim that even though there are many applications of boosting and deep learning it is for the first time these two mechanisms are used to forecast the crash events in various time frames at a larger scale involving 39 economies and their allied economic data. The research work produced by [25] offers a systematic tool to the central banks that warns whenever there is a risk of sensitive market so that they can adjust their monetary policies to stabilize the financial state.

A hybrid model for the forecasting of Shanghai stock exchange data was proposed by [26]. In this article the authors design an optimized NN model to forecast the movement of stock indices. The authors make use of a Radial Basis Function NN (RBFNN) optimized with a nature inspired algorithm called artificial fish swarm algorithm (AFSA) embedded with in K-means clustering technique is utilized to train the stock data. A comparative study was carried out and the proposed hybrid algorithm was compared along with popular ARIMA, back propagation (BP), SVM, nature inspired genetic algorithms (GA) and particle swarm optimization (PSO). Out of all the algorithms used for comparison the proposed algorithm yielded highest accuracy and least error. Another hybrid prediction model involving nature inspired algorithm was proposed by [27].

The authors claim to build an intelligent model capable of handling the complicated associations between inputs and outputs and predicting the financial data with high accuracy. A combination of genetic algorithm (GA) and Levenberg–Marquardt (LM) technique and a simple FNNs algorithm is used for stock data prediction. In this article data transformation and feature selection are carried out as the initial pre-processing tasks. The weights are optimized by using genetic algorithm and then the LM algorithm is used to make the FNN learn the patterns useful for predicting the stock data. In order to evaluate the performance of the proposed model, it was tested on various stock exchange indices available in the literature and it was found that the model resulted in significant accuracy. The main advantage of this model is it is capable of understanding the data fluctuations and mould accordingly yielding better results.

Current state of art provides many articles related to prediction of stock data using technical indicators. However in [28] the author discusses the impact of hybrid indicators - fundamental, technical and experts' advice in the prediction of stock data. Various features extracted from the mentioned hybrid indicators were fed into a neuro-fuzzy prediction model trained by sigmoid function. The empirical results reveal that the accuracy obtained from the proposed hybrid indicator neuro-fuzzy model was comparatively better than the individual indicator models which entirely rely on the historical price data.

The current literature provides many research work relate to single classifiers and multi classifiers. In [29] the authors attempt to check the impact of bagging and boosting on multiple classifiers. The authors check the working of multiple homogeneous and multiple heterogeneous classifiers in an ensembled approach for prediction of stock returns. The models were validated against average forecasting accuracy, Type I error, Type II errors, and profit gained on investments. The results indicate that the predictive performance of multiple classifiers was found to be higher than the individual classifiers. They also outperformed in predicting the profit gained on investment. In addition to that, heterogeneous classifier ensembles outperformed homogeneous counterparts. However, it was also noted that the difference between majority voting and bagging mechanism had no significance. The authors conclude that the prediction of stock returns with neural network based homogeneous multiple classifiers with voting mechanism show the best performance.

## 6. Analysis of stock data using Recurrent Neural Network

The authors in [30] came up with a hybrid approach to build a prediction model. The authors have combined RNN and artificial bee colony algorithm (ABC) to predict the stock data in order to maximize the profit. The hybrid model was simulated in three steps. The first step involved data transformation using Haar wavelet. The wavelet transform eliminated the noise in the time series data. Second step involved the creation of feature matrix using RNN. This includes utilization of RNN with numerous technical indicators and fundamental indicators to build a feature matrix using RNN architecture. The features were selected based on Stepwise Regression-Correlation Selection (SRCS). Third and final step involved the optimization of weights and biases using artificial bees colony algorithm under a parameter space design. The RNN was trained with sigmoid function resulting in higher predictive accuracy. The simulation was carried out on various stock indexes including Tokyo, Taiwan and London stock data. The simulation results confirm that the proposed hybrid approach returned promising results and can be used to implement real-time stock trade.

The authors of [31] proposed a research work where they used precision and average profit rate as the evaluation criteria instead of usual accuracy for studying the performance of recurrent neural network (RNN) on time series data. The RNN was applied on China Shanghai Shenzhen 300 Index stock data, specifically to understand the pattern of intraday trade. The experimental results on this time series data shows the RNN performed well with larger testing data providing useful information to the stake holders of the stock market.

A combined model including RNN and sentiment analysis (SA) was utilized to predict the volatility of the stock and mood of the investors who invested their money in Chinese stock market [32]. The model which is an amalgamation of Recurrent Neural Networks (RNN) with Gated Recurrent Units (GRU) predicted the presence of volatility in the Chinese stock data. Also a sentiment analysis base approach was followed to analyses the sentiments of the people extracted from the posts available at social platform called Sina Weibo. The experimental result reveal that an influential feature(s) can be derived by analyzing the opinions of the stake holders and could be used to improvise the prediction performance.

## 7. Analysis of stock data using Support Vector Machine

A comparative study to check the performance of three different prediction models was carried out by [33]. The three prediction models namely adaptive neuro-fuzzy inference systems (ANFIS), ANN and Support Vector Machine (SVM) was examined on the basis of sensitivity, and accuracy criteria. To check the performance of the models eight year long Borsa Istanbul BIST 100 Index stock data ranging from 2007 to 2014 was used. A random sampling and ten-fold cross validation mechanism was used for dataset division and performance testing respectively. Experimental results reveal that all the three models performed well in prediction of the down movement as compared to prediction of upward movement of stock. However it was found that in all cases SVM over cross ANFIS and ANN in terms of accuracy.

By studying the advancement of social media and linguistic technologies researchers came up with new modes of predicting stock data. One such mode of prediction was proposed by [34]. In this paper the authors address the mechanism of identifying the predominant stock indicators by analyzing the moods of the users using opinion mining. The authors adapt a lexicon based approach to analyze the emotions of 755 million tweets acquired using twitter API.

The tweets were subjected to classification to analyze the moods of the users on P500 and DJIA and S indicators. The analysis reveal that round 80% of tweets had no significance (i.e. were neutral) the tweets from celebrities influenced the mindset of the common people. If the tweets were positive it had positive impact on common man, if the tweets were negative it had negative impact.

Feature selection is salient step in prediction task. Stronger the feature set higher the predictive accuracy. There are many feature selection techniques available in literature which claims high performance. One such research work based on feature selection is projected by [35]. Here a hybrid feature selection technique involving fractals and SVM is used to select strong features from the stock data. Fractal method is most suitable for finding solutions to non-linear problem and thus fits the domain of stock analysis. The performance of Fractal-SVM method was compared with some of the widely used feature selection techniques to confirm that the performance of prediction using features selected by fractal-SVM was higher than the counter parts.

**8. Analysis of stock data using Support Vector Regression**

Stock data are very complex in nature and also they are non-stationary because of which the traditional stock prediction technique such as ARIMA are not enough to build a reliable prediction model. To predict the non-linear and continuously changing stock data, a three stage prediction model was proposed by [36]. In the first stage, delay co-ordinates were used to recreate the unknown unseen phase. In the second stage, Support Vector Regression parameters were optimized using nature inspired Firefly Algorithm (FA). Once the support vector regression parameters are optimized, they are used to predict the stock data. In this system, chaotic mapping, nature inspired firefly algorithm, and SVR technique were combined to develop a hybrid stock prediction model. The proposed model was compared with prediction models based on various techniques such as genetic algorithm, ANN and ANFIS, and it was found that the proposed model out performed all other techniques.

The feature extraction is one of the most significant phase in building any prediction model because the relevance of the feature directly impact on the accuracy of the model. Hence, one needs to be very careful and confident in selecting the features which are going to be used in the model and one such feature extraction technique is Non-linear Independent Component Analysis (NLICA). Using the NLICA as the feature extraction technique in the pre-processing stage, a stock prediction model was proposed by [37]. Here, once the features are identified, features are used in SVR as input to develop the stock prediction model. The proposed model was tested on two major Asian indices namely Shanghai Stock Exchange of China and Nikkei stock exchange of Japan. The experimental study exhibited that the NLICA improved the accuracy of the prediction model and outperformed the Principal Component Analysis based prediction model.

Another variant of SVR used for prediction of stock data is multi-output SVR. Further, a hybrid FA multi-output SVR model for prediction of stock data was proposed [38]. The firefly algorithm was used for finding the optimal parameters for multi-output support vector regression.

**B. Clustering methodologies used for analysis**

**1. Analysis of stock data using Filtering**

In stock market, the main objective is to increase the profit margin. However, the price of the stock is vulnerable to many factors due to which stock price may either increase or decrease without any sign or warning. In such situations making sure the minimum profit margin is the absolute goal of a trader. For such situations traders can use the stop loss window which will be triggered when the stock price hit a certain price ensuring the minimum profit or loss. Such a mechanism of using stop loss was introduced in [39]. Setting a stop loss is itself is a tricky thing and usually the traders make use of many technical indicators for analysis the stock data. One such technical indicator is exponential moving average (EMA) which is used in this model. The EMA follows the market trend and gives an idea about where the stock price may head. The EMA indicator is estimated for fifteen minute and one day time frame which helps the traders to look into the short term and medium term investment strategies. The proposed model was tested on intraday readings of DJIA index and the stop loss was kept for 100 points. The proposed model with the stop loss can be used to buy or sell the stock and ensure at least the minimum profit or damage.

One of the most popular statistical techniques utilized for building stock forecasting is ARIMA and many researchers have used it and one such example is described in [40]. Here, ARIMA was used to construct the prediction model and it was tested on couple of real world stock exchanges namely New York and Nigeria Stock Exchanges. The results showed that the simple but effective ARIMA models are very competitive when compared with other available prediction models.

The stock market data is a very complicated data which is vulnerable to many factors due to which it becomes very volatile. To build an effective prediction model one should consider the volatility factors. In this regard [41] have proposed a forecasting model for volatility in particularly conditional variance and the model was tested on Indian stock market index SENSEX data over a decade from 1996 to 2010. In the proposed model, Generalized Autoregressive Conditional Heteroskedasticity (GARCH), Threshold GARCH, and Exponential GARCH were considered for the study and the results displayed that the GARCH models were performing superior in forecasting the conditional variance in Indian stock exchange [42].

**2. Analysis of stock data using Fuzzy based clustering**

Stock prediction models may show some accuracy in predicting the stock price based on which one may wish to invest the money but one should remember that buying selling comes with a cost such as commission. If the prediction model fails to ensure that the profit earned is after all the expenses including the commission cost then there is no use of such models. To deal with such thing [43] proposed a fuzzy rule based prediction model. In this model fuzzy C-means algorithm was used to recognize the relevant rules which could be used for making the prediction. The parameters used in this model were optimized using ANFIS and the model produced a very high accuracy when it was tested on a real world Tehran stock exchange.

Stock predictions are done to make profit in three ways by either predicting the individual stock or the stock index value or both. To do this, a three staged forecasting model was developed by [44]. In this proposed model, Multiple Regression Analysis (MRA) was used in the first stage to identify the correlation among the variable impacting the stock price both in terms of economic and financial variables. In the second stage a stock prediction model was developed using the Differential Evolution (DE) based Fuzzy Clustering (FC) technique. In the final stage, Fuzzy Neural Network is used for prediction of stock. In this fashion, a hybrid prediction model was created using MRA, DEFC and fuzzy NN.

The ANFIS model is used worldwide for developing the stock prediction and hence, one needs to investigate the effectiveness of ANFIS model. In this regard [45] invested their time and resource to investigate ANFIS model of prediction. They implemented the prediction model using ANFIS and tested the model on Istanbul stock exchange and they found that the ANFIS model is quite effective in forecasting the stock returns.

Fuzzy Time Series has been used in the literature for developing a prediction model and this has motivated the [46] to develop a multilayer prediction model using the fuzzy time series. Each layer is used to deal with certain specific problem. The model was tested on Taiwan stock exchange, NASDAQ, DJI, and S and P datasets. However, each layer failed to understand the working style of other layers as a result it has created a research opportunity for others to work on this inefficiency to improve the accuracy of prediction models using fuzzy time series.

While dealing with the fuzzy time series, the distribution of dataset play an important role and hence it needs to be addressed in order to create a prediction mode. In this regard, [47] proposed to use discretization based on entropy and partition based on binning along with granular computing. The proposed model was compared with SVM based models, GARCH, and fuzzy models. However, this model fails to take other factors which could impact the stock such as volume of shares trading, technical indicators, and financial reports of the concerned company.

Technical analysis is used to build prediction models in the past [48] have several drawbacks such as the models were complex and produced unintelligent rules, mathematical models used were complex and investors failed to understand them. To handle it effectively [49] proposed an ANFIS which used multi-technical indicator for forecasting the stock. This hybrid model used technical indicators which converted into linguistic values using clustering method and finally linguistic rules were used by fuzzy inference system to extract the rules. The model produced better results when it was compared with [50].

The estimation of trend can be done by developing an advanced fuzzy set [51]. In order to do it, first the past historical data needs to be fuzzified into fuzzy sets and fuzzy relationships are estimated. The entire model is trained using the imperialist competitive algorithm. Once training is done, defuzzification takes place to predict the stock data.

### 3. Analysis of stock data using K-means

Clustering is among the major techniques in data mining and it has been widely used in various applications. Even in building stock prediction model, clustering technique such as K-means is used in the past. In [52], classifications of stocks were first performed. Once the classification is done, then the clusters of stocks based on investment criteria were formed. For this, K-means clustering technique was used. As a result of this model, investors can divide their investment in different sectors which eventually decreases the loss or risk.

### 4. Analysis of stock data using Optimization

Building a prediction model for stocks is very difficult but it is even more difficult to find the appropriate parameters of the model which can yield better, optimal, and desired results. Many researchers have used different techniques to optimize the model and in this section we are showing few among them.

The forecasting models built using ARIMA involve mathematical equations which are complicated and not easily understood by traders. The rules generated by neural networks are also not easily understandable. Hence, a simplified and effective hybrid model was proposed by [48] which consist of four procedures. First, technical indicators which are more relevant are selected. Second, the technical indicators are converted into linguistic datasets. Third, rough set theory algorithm is used to extract the linguistic rules. Fourth, genetic algorithms are used to optimize and find the optimal rules which are used to predict the stock data.

A hybrid Morphological Rank Linear (MRL) along with genetic algorithm was proposed by [53] to build a stock prediction model. Here, the most relevant time lags were identified using genetic algorithm which are also capable of fine tuning the characteristics of time series. Genetic algorithms are also used in this model to find the initial parameters of Morphological Rank Linear module. Finally least mean square algorithm was used to predict the stock data.

A multiple kernel learning algorithm was designed in two stages [54] to predict the stock data. The model adopted the gradient projection technique and sequential optimization method. The key benefit of this method is, it has the capability to combine the different hyper-parameter settings.

## III. EXPERIMENT AND RESULT

We have proposed to use ANN along with four different learning functions to build a prediction model. Whichever learning function produces the highest accuracy will be used in the final prediction model. The four learning functions are mentioned in the table 1.

**Table 1 Learning Functions used in our model**

| Sl No. | Function | Formula |
|---|---|---|
| 1 | F1 | $\frac{1}{(1+exp^{-a*x})}$ |
| 2 | F2 | $\frac{2}{(1+exp^{-a*x})}-1$ |
| 3 | F3 | $a*tanh(b*x)$ |
| 4 | F4 | $\frac{1}{\sqrt{2\pi}\sigma}*exp^{\frac{-(x-\mu)^2}{2*\sigma^2}}$ |

Before we start the experiment, Nifty index data was normalized and the entire data was divided for training and testing purpose in the ration 60:40. The parameters of the neural network such as number of hidden layer was set to 2, maximum epoch was set to 30, number of previous observation taken for prediction was set to 5. Experiment was conducted for all four learning function for all possible different values of the variables of their respective formula of the learning function and the obtained results are shown in table 2 and 3.

**Table 2 Prediction accuracy for different learning functions F1, F2, and F4.**

| Sl. No | Alpha or Sigma value | Accuracy of F1 in % | Accuracy of F2 in % | Accuracy of F4 in % |
|---|---|---|---|---|
| 1. | 0.1 | 75.77 | 83.17 | 23.55 |
| 2. | 0.2 | 88.89 | 89.74 | 23.31 |
| 3. | 0.3 | 93.78 | **90.24** | 59.85 |
| 4. | 0.4 | 91.85 | 90.09 | 58.22 |
| 5. | **0.5** | **94.55** | 80.60 | 80.35 |
| 6. | 0.6 | 94.19 | 69.51 | 78.04 |
| 7. | 0.7 | 94.29 | 44.89 | 83.87 |
| 8. | 0.8 | 88.35 | 82.03 | 84.51 |
| 9. | 0.9 | 94.06 | 82.55 | **86.43** |
| 10. | 1.0 | 87.77 | 48.45 | 75.57 |

**Table 3 Prediction accuracy of the learning functions F3.**

| Sl. No | Value of variable a | Value of variable b | Accuracy of F3 in % |
|---|---|---|---|
| 1. | 0.1 | 0.1 | 45.31 |
| 2. | 0.2 | 0.2 | 53.33 |
| 3. | 0.3 | 0.3 | 67.68 |
| 4. | 0.4 | 0.4 | 74.05 |
| 5. | 0.5 | 0.5 | 80.07 |
| 6. | 0.6 | 0.6 | 78.58 |
| 7. | 0.7 | 0.7 | 79.54 |
| 8. | 0.8 | 0.8 | 61.15 |
| 9. | 0.9 | 0.9 | 52.27 |
| 10. | 1.0 | 1.0 | 43.88 |
| 11. | 0.1 | 1.0 | 52.09 |
| 12. | 0.2 | 0.9 | 60.80 |
| 13. | 0.3 | 0.8 | 68.63 |
| 14. | 0.4 | 0.7 | 75.01 |
| 15. | 0.5 | 0.6 | 80.52 |
| 16. | 0.6 | 0.5 | 83.07 |
| 17. | 0.7 | 0.4 | 85.34 |
| 18. | 0.8 | 0.3 | 88.79 |
| 19. | **0.9** | **0.2** | **90.01** |
| 20. | 1.0 | 0.1 | 89.90 |

From table 2 and 3 it is clear that the our ANN model with different learning functions produced an accuracy of **94.55% for F1**, 90.24% for F2, 90.01% for F3, and 86.43% for F4. Even though, the ANN and its parameters were same, different learning functions produced different results. Hence, it becomes very necessary to select the appropriate learning function to get better results. In our case the learning function F1 is best suited for stock prediction purposes as it is producing the highest accuracy among other learning functions.

## IV. RESEARCH GAP

Despite many researchers have proposed different techniques, still there are many drawbacks and limitations exist. In this section we are trying to mention few.

Neural networks used in building the prediction model contain huge number of neurons in their hidden layer which fail to deal with the high computations [7].

If the training and testing rates are kept very low, it will affect the overall performance of the model. And neural networks can also be trapped in local minima [8].

Failure to avoid misclassification and optimization of the parameters of the NN will result in low accuracy of the model [23].

The convolutional neural networks along with deep learning are not suitable for highly extensive applications such as stock prediction and hence when compared to other techniques, it is found that CNN produces low accuracy [55].

While designing a decision support system for stock prediction model, techniques and knowledge is required [19].

Artificial neural network spends a lot of time for training and testing, and it also fail to explain the obtained result [33].

Fuzzy time series are time consuming, complex in nature and the results are affected by the methods used for splitting [47].

Support Vector Machines are very effective in extracting the features. But SVM fails to specify the minimum number of features essential to get the best outcomes out of the prediction model [35].

## V. CONCLUSION

The objective of this research article is twofold. One is to use the ANN for building the stock prediction model. From our experiment we found that the learning functions has produced an accuracy of **94.55% for F1**, 90.24% for F2, 90.01% for F3, and 86.43% for F4.

From the results we conclude that the ANN with learning function F1 is best suited for prediction purposes when compared with F2, F3, and F4.

The second objective of this research article is to deliver a detailed analysis on the methodologies used to develop stock market data prediction model by eminent researchers. The detailed analysis presented in this paper is divided into two categories. One, based on the prediction methodologies used for analysis. Two, based on the clustering methodologies used for analysis.

The prediction methodologies used for analysis is further explored into eight sub categories namely,

1. Analysis of stock data using Artificial Neural Network.
2. Analysis of stock data using Convolutional Neural Network.
3. Analysis of stock data using Decision Support System.
4. Analysis of stock data using Hidden Markov Model.
5. Analysis of stock data using Neural Network.
6. Analysis of stock data using Recurrent Neural Network.
7. Analysis of stock data using Support Vector Machine.
8. Analysis of stock data using Support Vector Regression.

The clustering methodologies used for analysis is also further explored into four sub categories namely,

1. Analysis of stock data using Filtering
2. Analysis of stock data using Fuzzy based clustering
3. Analysis of stock data using K-means
4. Analysis of stock data using Optimization

After presenting a detailed analysis available in the literature, we have also highlighted the research gaps existing in the methodologies we mentioned in this paper.

## REFERENCES


1. Srinivas Narasegouda, M Umme Salma, and Anuradha N Patil. "Nature inspired algorithm approach for the development of an energy". Computational Intelligence in Sensor Networks, 776:55, 2018.
2. M Umme Salma, Srinivas Narasegouda, and Anuradha N Patil. "A metaheuristic based hybrid predictive model for sensor network". Computational Intelligence in Sensor Networks, 776:167, 2018.
3. M Umme Salma, and Doreswamy. "Bat-elm: a bio inspired model for prediction of breast cancer data". In Applied and Theoretical Computing and Communication Technology (iCATccT), 2015 International Conference on, pages 501–506. IEEE, 2015.
4. M Umme Salma, and Doreswamy. "Fast modular artificial neural network for the classification of breast cancer data". In Proceedings of the Third International Symposium on Women in Computing and Informatics, pages 66–72. ACM, 2015.
5. M Umme Salma "et al. Pso based fast k-means algorithm for feature selection from high dimensional medical data set". In Intelligent Systems and Control (ISCO), 2016 10th International Conference on, pages 1–6. IEEE, 2016.
6. Jonathan L Ticknor. A" bayesian regularized artificial neural network for stock market forecasting". Expert Systems with Applications, 40(14):5501–5506, 2013.
7. Ajit Kumar Rout, Birendra Biswal, and Pradipta Kishore Dash. "A hybrid flann and adaptive differential evolution model for forecasting of stock market indices". *International Journal of Knowledge-based and Intelligent Engineering Systems*, 18(1):23–41, 2014.
8. Sreejit Chakravarty and Pradipta K Dash. "A pso based integrated functional link net and interval type-2 fuzzy logic system for predicting stock market indices". Applied Soft Computing, 12(2):931–941, 2012.
9. Xiao Zhong and David Enke. "Forecasting daily stock market return using dimensionality reduction". Expert Systems with Applications, 67:126–139, 2017.
10. Amir Omidi, Esmaeil Nourani, and Mehdi Jalili. "Forecasting stock prices using financial data mining and neural network". In 2011 3rd International Conference on Computer Research and Development, volume 3, pages 242–246. IEEE, 2011.
11. Esmaeil Hadavandi, Hassan Shavandi, and Arash Ghanbari. "Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting". Knowledge-Based Systems, 23(8):800–808, 2010.
12. Jigar Patel, Sahil Shah, Priyank Thakkar, and Ketan Kotecha. "Predicting stock market index using fusion of machine learning techniques". Expert Systems with Applications, 42(4):2162–2172, 2015.
13. Amin Hedayati Moghaddam, Moein Hedayati Moghaddam, and Morteza Esfandyari. "Stock market index prediction using artificial neural network". *Journal of Economics, Finance and Administrative Science*, 21(41):89–93, 2016.
14. Erkam Guresen, Gulgun Kayakutlu, and Tugrul U Daim. "Using artificial neural network models in stock market index prediction". Expert Systems with Applications, 38(8):10389–10397, 2011.
15. Shailaja. K P, Manjunath. M, and M Umme Salma. "Impact of learning functions on prediction of stock data in neural network". In 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS), pages 82–86, Dec 2018.
16. Manuel R Vargas, Beatriz SLP De Lima, and Alexandre G Evsukoff. "Deep learning for stock market prediction from financial news articles". In 2017 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA), pages 60–65. IEEE, 2017.
17. Xingyu Zhou, Zhisong Pan, Guyu Hu, Siqi Tang, and Cheng Zhao. "Stock market prediction on high-frequency data using generative adversarial nets". Mathematical Problems in Engineering, 2018, 2018.
18. Xi Zhang, Siyu Qu, Jieyun Huang, Binxing Fang, and Philip Yu. "Stock market prediction via multi-source multiple instance learning". IEEE Access, 6:50720–50728, 2018.
19. Qinghua Wen, Zehong Yang, Yixu Song, and Peifa Jia. "Automatic stock decision support system based on box theory and svm algorithm". Expert Systems with Applications, 37(2):1015–1022, 2010.
20. Jyoti Badge et al. "Forecasting of Indian stock market by effective macro-economic factors and stochastic model". *Journal of Statistical and Econometric Methods*, 1(2):39–51, 2012.
21. Aditya Gupta and Bhuwan Dhingra. "Stock market prediction using hidden markov models". In 2012 Students Conference on Engineering and Systems, pages 1–4. IEEE, 2012.
22. Pei-Chann Chang et al. "A novel model by evolving partially connected neural network for stock price trend forecasting". Expert Systems with Applications, 39(1):611–620, 2012.
23. Xiongwen Pang, Yanqiang Zhou, Pan Wang, Weiwei Lin, and Victor Chang. "An innovative neural network approach for stock market prediction". The Journal of Supercomputing, pages 1–21, 2018.
24. George S Atsalakis, Emmanouil M Dimitrakakis, and Constantinos D Zopounidis. "Elliott wave theory and neuro-fuzzy systems, in stock market prediction: The wasp system". Expert Systems with Applications, 38(8):9196–9206, 2011.
25. Sotirios P. Chatzis, Vassilis Siakoulis, Anastasios Petropoulos, Evangelos Stavroulakis, and Nikos Vlachogiannakis. "Forecasting stock market crisis events using deep and statistical machine learning techniques". Expert Systems with Applications, 112:353 – 371, 2018.
26. Wei Shen, Xiaopen Guo, Chao Wu, and Desheng Wu. "Forecasting stock indices using radial basis function neural networks optimized by artificial fish swarm algorithm". Knowledge-Based Systems, 24(3):378–385, 2011.
27. Shahrokh Asadi, Esmaeil Hadavandi, Farhad Mehmanpazir, and Mohammad Masoud Nakhostin. "Hybridization of evolutionary levenberg–marquardt neural networks and data pre-processing for stock market prediction". Knowledge-Based Systems, 35:245–258, 2012.
28. AA Adebiyi, CK Ayo, and SO Otokiti. "Fuzzy-neural model with hybrid market indicators for stock forecasting". *International Journal of Electronic Finance*, 5(3):286–297, 2011.
29. Chih-Fong Tsai, Yuah-Chiao Lin, David C Yen, and Yan-Min Chen. "redicting stock returns by classifier ensembles". Applied Soft Computing, 11(2):2452–2459, 2011.

30. Tsung-Jung Hsieh, Hsiao-Fen Hsiao, and Wei-Chang Yeh. "Forecasting stock markets using wavelet transforms and recurrent neural networks: An integrated system based on artificial bee colony algorithm". Applied soft computing, 11(2):2510–2525, 2011.
31. Xin-Kai Xie and Hong Wang. "Recurrent neural network for forecasting stock market trend". In Computer Science, Technology and Application: Proceedings of the 2016 International Conference on Computer Science, Technology and Application (CSTA2016), pages 397–402. World Scientific, 2017.
32. W. Chen, Y. Zhang, C. K. Yeo, C. T. Lau, and B. S. Lee. "Stock market prediction using neural network through news on online social networks". In 2017 International Smart Cities Conference (ISC2), pages 1–6, Sep. 2017.
33. Asil Oztekin, Recep Kizilaslan, Steven Freund, and Ali Iseri. "A data analytic approach to forecasting daily stock returns in an emerging market". European Journal of Operational Research, 253(3):697–710, 2016.
34. Alexander Porshnev, Ilya Redkin, and Alexey Shevchenko. "Machine learning in prediction of stock market indicators based on historical data and data from twitter sentiment analysis". In 2013 IEEE 13th International Conference on Data Mining Workshops, pages 440–444. IEEE, 2013.
35. Li-Ping Ni, Zhi-Wei Ni, and Ya-Zhuo Gao. "Stock trend prediction based on fractal feature selection and support vector machine". Expert Systems with Applications, 38(5):5569–5576, 2011.
36. Ahmad Kazem, Ebrahim Sharifi, Farookh Khadeer Hussain, Morteza Saberi, and Omar Khadeer Hussain. "Support vector regression with chaos-based firefly algorithm for stock market price forecasting". Applied soft computing, 13(2):947–958, 2013.
37. Ling-Jing Kao, Chih-Chou Chiu, Chi-Jie Lu, and Jung-Li Yang. "Integration of nonlinear independent component analysis and support vector regression for stock price forecasting". Neurocomputing, 99:534–542, 2013.
38. Tao Xiong, Yukun Bao, and Zhongyi Hu. "Multiple-output support vector regression with a firefly algorithm for interval-valued stock price index forecasting". Knowledge-Based Systems, 55:87–100, 2014.
39. Rub´en Ar´evalo, Jorge Garc´ıa, Francisco Guijarro, and Alfred Peris. "A dynamic trading rule based on filtered flag pattern recognition for stock market price forecasting". Expert Systems with Applications, 81:177– 192, 2017.
40. Adebiyi A Ariyo, Adewumi O Adewumi, and Charles K Ayo. "Stock price prediction using the arima model" In 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, pages 106–112. IEEE, 2014.
41. P Srinivasan and P Ibrahim." Forecasting stock market volatility of bse-30 index using garch models". Asia Pacific Business Review, 6(3):47–60, 2010.
42. Saikat Sovan Deb, Srivyal Vuyyuri, and Bijan Roy. "Modeling stock market volatility in india: A comparison of univariate deterministic models". ICFAI Journal of Applied Finance, pages 19–33, 2003.
43. Akbar Esfahanipour and Werya Aghamiri. "Adapted neuro-fuzzy inference system on indirect approach tsk fuzzy rule base for stock market analysis". Expert Systems with Applications, 37(7):4742–4748, 2010.
44. David Enke, Manfred Grauer, and Nijat Mehdiyev. "Stock market prediction with multiple regression, fuzzy type-2 clustering and neural networks". Procedia Computer Science, 6:201–206, 2011.
45. Melek Acar Boyacioglu and Derya Avci. "An adaptive network-based fuzzy inference system (anfis) for the prediction of stock market return: the case of the istanbul stock exchange". Expert Systems with Applications, 37(12):7908–7912, 2010.
46. Hossein Javedani Sadaei and Muhammad Hisyam Lee. "Multilayer stock forecasting model using fuzzy time series". *The Scientific World Journal, 2014.*
47. Mu-Yen Chen and Bo-Tsuen Chen. "A hybrid fuzzy time series model based on granular computing for stock price forecasting". Information Sciences, 294:227–241, 2015.
48. Ching-Hsue Cheng, Tai-Liang Chen, and Liang-Ying Wei. "A hybrid model based on rough sets theory and genetic algorithms for stock price forecasting". Information Sciences, 180(9):1610–1629, 2010.
49. Liang-Ying Wei, Tai-Liang Chen, and Tien-Hwa Ho. "A hybrid model based on adaptive-network-based fuzzy inference system to forecast taiwan stock market". Expert Systems with Applications, 38(11):13625–13631, 2011.
50. Sheng-Chai Chi, Wei-Ling Peng, Pei-Tsang Wu, and Ming-Wei Yu. "The study on the relationship among technical indicators and the development of stock index prediction system". In 22nd International Conference of the North American Fuzzy Information Processing Society, NAFIPS 2003, pages 291–296. IEEE, 2003.
51. Hossein Javedani Sadaei, Rasul Enayatifar, Muhammad Hisyam Lee, and Maqsood Mahmud. "A hybrid model based on differential fuzzy logic relationships and imperialist competitive algorithm for stock market forecasting". Applied Soft Computing, 40:132–149, 2016.
52. SR Nanda, Biswajit Mahanty, and MK Tiwari. "Clustering indian stock market data for portfolio management". Expert Systems with Applications, 37(12):8793–8798, 2010.
53. Ricardo De A Ara´uJo and Tiago AE Ferreira. "A morphological-rank linear evolutionary method for stock market prediction". Information Sciences, 237:3–17, 2013.
54. Chi-Yuan Yeh, Chi-Wei Huang, and Shie-Jue Lee. "A multiple-kernel support vector regression approach for stock market price forecasting". Expert Systems with Applications, 38(3):2177–2186, 2011.
55. Bo Xu, Dongyu Zhang, Shaowu Zhang, Hengchao Li, and Hongfei Lin. "Stock market trend prediction using recurrent convolutional neural networks". In CCF International Conference on Natural Language Processing and Chinese Computing, pages 166–177. Springer, 2018.

## AUTHORS PROFILE

**Shailaja K.P,** received her B.Sc degree from Bangalore University in 1995, MCA degree from Bangalore University in 1999 and M.Phil degree from BharatiDasan University in 2006 and currently working as Assistant Professor in Department of Computer Application, B.M.S College of Engineering, Bangalore. At present she is pursuing her PhD under the guidance of Dr. Manjunath M in the Department of Master of Computer Applications, RVCE, Bangalore from VTU, Belgaum. Her areas of interest are Data Mining and Analytics and Machine Learning.

**Dr. Manjunath M,** received his B.Sc degree from Bangalore University in 2007, MCA degree from VTU, Belgaum in 2011 and PhD from Mangalore University, Mangalore in 2016. At present he is working as Assistant Professor in Department of Master of Computer Applications, RV College of Engineering, Bangalore. His areas of expertise include Wireless Communication and Networking, Data Mining and Analytics and Web Technologies.