International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 6 June 2021, pp: 38-48 www.ijaem.net ISSN: 2395-5252

"Machine Learning: A Gateway for Stock Market Predictions"

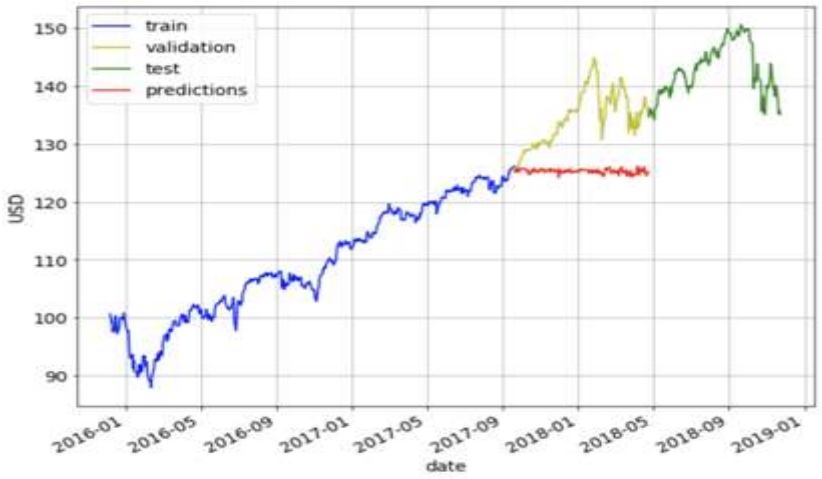
Lakshay Aggarwal
B.tech(information technology) final year student hmr institute of technology & management

Submitted: 25-05-2021 Revised: 31-05-2021 Accepted: 03-06-2021

ABSTRACT: Today it is globally known that there is no point in investing your money into a new venture or a start-up business. It is being observed that in Today's time, the market has turned out to be financially volatile due to aggressive competition between the vendors for the same products in the market. Because of the various financial and economic crises in the industry, today, every person smells it risky to put the money in any ongoing business. Thus, we need such ideas and techniques that can help the market and economy grow.

To secure the investments we need to move to a new Entrepreneur Gateway- "Stock Market". The current globalization and distribution of markets have meant that companies and organizations frequently need to adapt their internal structure and operating processes, to demonstrate efficiency and effectiveness for every customer and user, especially when considering aspects that are beyond logical definitions, such as emotions. To cope with that, this work proposes an alternative to model the distributed Stock Exchange Scenario with ontologies and their futuristic predictions. The proposed model considers that each investor can invest using information obtained by communication with different traders or investors. Each investor has its knowledge represented by ontologies, which is composed of technical knowledge together with internal training states of the data to present the graph. Our preliminary results show the possibility to use ontologies as knowledge representation mechanisms for domains that consider the human emotional dimension for decision-making processes.

Our objective is to identify the best possible algorithm for predicting future stock market performances. The successful prediction of the stock market will have a very positive impact on the stock market institutions and the investors also. In other words, we can say that Stock Market is the way out for every investor in Today's time to make their money by scoring profits in the market itself. Therefore, the correct identification of algorithms for the stock market prediction model is needed so that an investor can successfully raise profits.



International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 6 June 2021, pp: 38-48 www.ijaem.net ISSN: 2395-5252

I. INTRODUCTION

Stock Market consists of investors and sellers of the stock. Stock exchanges can be channelized by Financial Institutions which allow the exchange of various goods and services between the stockbroker components for monetary benefits. The scope of huge turnovers in this trading line grasps the attention of various people to invest the money in the market and score exponential profits. Profits in the stock market are directly proportional to the entry value of stock in the exchange market and the volume of its transactions.

(Fig-1. Predicted Global Stock Exchange Market 2020-2024) (Fig-2. Market Analysis 2025)

Stock market prediction means determining the futuristic behaviour of the stocks in the market. A technical system or software is essential to be built which will work with maximum possible accuracy and it should consider all important factors that could influence the results in the exchange market. Various researches have already been done to predict stock market prices. The researches have been done over business and computer science domains for the better synchronization of technology with the stock market. Sometimes, the stock market does well even when the economy is falling because there are various reasons for the profit or loss of a share like increase in prices of goods of that particular company in the market, investment by government policies, the initial performance of the stocks in the market, etc. Predicting the performance of a stock market is tough as it takes into account various factors such as Stock Entry Value, Volume of Transactions, Government Policies, and Stock Closing Value, etc. The main aim is to identify the sentiments of investors and based upon them predict the stock prices for the future with the highest accuracy. It is usually difficult as there must be a rigorous analysis of national and international events as well. An investor needs to know about the events related to the stock company & also the current price of the stock in which he wants to invest so that accordingly, system or software can update investor with future stock market prices with the highest accuracy of the predictions which in turn shall be highly profitable for the investor in the stock exchange market

Motivation & Requirements for the Proposed System

The real motivation for this application is to make the easy process of all the investors, sellers, and brokers in the stock exchange market processes like Stock Entry Value in the market, Volume of Transactions of the Stock Quantities, Stock Closing Value, Stock Related Events and Policy Announcements, Futuristic Stock Price Predictions, etc. We have often seen that to find out the performance of particular stocks within the market, the investor has to go through various Broker Tips and Market analyses which results in wastage of time. By implementation of this system, it will become easy to manage all such processes thus saving time and earning high profits. So now by taking the motivation of this scenario which was

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 6 June 2021, pp: 38-48 www.ijaem.net ISSN: 2395-5252

regularly done in the stock market we are designing this system which can be benefitted for the investors, stockbrokers & sellers. This system would make it possible for all the people to trade easily in the market by showing fast, effective and accurate predictions about the stocks.
The major requirements of this system are as follows-:

- To implement this application we need to understand the basic terminologies associated with the Stock Exchange Market like Stock Entry Value, Stock Closing Value, Volume of Transactions of the Stock, Total Traded Quantity of Stock, etc.
- There must be a data set to implement the system. The description of the dataset is given in Table1.

Description of dataset Table 1.

Feature	Description
Symbol	Symbol of the listed company
Series	Series of the equity(EQ, BE, BL, BT, GC, IL)
Open	Starting price at which a stock is traded in a day
High	Highest price of equity symbol in a day
Low	Lowest price of share in a day
Close	Final price at which a stock is traded in a day
Last	Last traded price of the equity symbol in a day
Previousclose	The previous day closing price of equity symbol in a day
TOTTRDQTY	Total traded quantity of equity symbol on the date
TOTTRDVAL	Total traded volume of equity symbol on the date

3. The dataset needs to be converted into a format that can be analyzed. Therefore some steps are performed before building the model:

- Handling missing data
- One Hot Encoding: It converts categorical data to quantitative variables as any data in the form of string or object does not help in analysing data. The first step is to convert the columns to category data type. The second step is to apply label encoding to convert it into numerical values which will be valuable for analysis. The third step is to convert the column into binary value (either 0 or 1).
- Data Normalization: It is often possible that if data is not normalized, the column with high values will be given more importance in prediction. To tackle that, we scale the data.

II. RESEARCH METHODOLOGY

The study is qualitative and descriptive in nature and most of the information is predicated on secondary sources of survey data. Such an approach is adopted within the study because the area of research is extremely broad and sources of information are also spread across multiple locations. To attain a conclusive idea of the larger picture on Stock Exchange Market Predictions information systems, analysing the prevailing survey data and certain successful case studies of the Stock Market system could provide a good end in finding the predictions with utmost accuracy to the research question framed.

III. PROPOSED APPROACH

To carve out a comprehensive Prediction information systems model for the Stock Exchange Market, some of the various technical analysis tasks are discussed and summarized in the context of the current study.
The most important tasks in information systems can be summarized as follows:
1. Statistical method: Statistical methods were widely used before the advent of machine learning. The popular techniques are ARIMA, ESN, and Regression. The main features of the statistical approach are linearity and stationary. An analysis of statistical approaches like Linear Discriminant Analysis (LDA), regression algorithms, and Quadratic Discriminant Analysis (QDA) is done in Pattern Recognition. An approach to using time series as input variables is Auto-Regressive Moving Average (ARMA). ARMA model combines Auto-Regressive models. ARIMA can reduce non-stationary series to a stationary series and is also an extension to ARMA models.
2. Pattern Recognition: This method focuses on pattern detection. It studies data rigorously and identifies a pattern. Traders find buy and sell signals in Open-High- Low-Close Candlestick charts. A study is done on the pattern of stock prices that can help in predicting the future of a stock. An analysis of patterns is done, by studying charts to develop predictions of the stock market. A comparison of market price and its history to chart

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 6 June 2021, pp: 38-48 www.ijaem.net ISSN: 2395-5252

patterns for predicting future stock prediction is done.

3. Machine learning: Machine learning is used in many sectors. One of the most popular applications is being stocked market prediction itself. Machine learning algorithms are either supervised or unsupervised by nature. In Supervised learning models, labelled input data is trained and the algorithm is applied. Classification and regression are types of supervised learning. It has a higher controlled environment. Unsupervised learning models have unlabelled data but have a lower controlled environment. It analyses pattern, correlation, or cluster.

4. Sentiment analysis: Sentiment analysis is an approach that is used concerning the latest trends 8. It observes the trends by analysing news and social trends like tweet activity. A study is done on using segment signals from text to improve the efficiency of models to analyse trends in the stock market.

IV. RELATED WORK

There have been several research works on implementing machine learning algorithms for predicting the stock market. A study is done by implementing machine learning algorithms on Karachi Stock Exchange (KSE). It compared Single Layer Perceptron (SLP), Multi-Layer Perceptron (MLP), Radial Basis Function (RBF), and Support Vector Machine (SVM). MLP performs best as compared to others. A comparison of four techniques Artificial Neural Network (ANN), Support Vector Machine (SVM), random forest, and Naive-Bayes is done. A study used unsupervised learning as a precursor for supervised tasks. A study compared various machine learning techniques like Random Forest, AdaBoost, Kernel Factory, Neural Networks, Logistic Regression, Support Vector Machine, KNN, on a dataset of European companies. An application of various machine learning algorithms (SVM, Naïve Bayes, Random Forest) was done and it was found that random forest gives the highest F- score. The research applied RNN, LSTM, and Gated Recurrent Unit (GRU) on the stock dataset and found that LSTM outperforms other algorithms. An application of LSTM to predict Nifty prices is done. It aims to predict stock price and also an interval of growth rate. They found that this method is better than some existing methods in terms of accuracy.

V. DATASET

The dataset is downloaded from Kaggle. The description of dataset is given in Table1.

Table 1. Description of dataset

Feature	Description
Symbol	Symbol of the listed company
Series	Series of the equity(EQ, BE, BL, BT, GC, IL)
Open	Starting price at which a stock is traded in a day
High	Highest price of equity symbol in a day
Low	Lowest price of share in a day
Close	Final price at which a stock is traded in a day
Last	Last traded price of the equity symbol in a day
Previous close	The previous day closing price of equity symbol in a day

VI. DATA PRE-PROCESSING

The dataset is in raw format. The dataset needs to be converted into a format that can be analysed. Therefore some steps are performed before building the model:

1. Handling missing data

2. One Hot Encoding: It converts categorical data to a quantitative variable as any data in the form of string or object does not help in analysing data. The first step is to convert the columns to category data type. The second step is to apply label encoding to convert it into numerical values which will be valuable for analysis. The third step is to convert the column into binary value (either 0 or 1).

3. Data Normalization: It is often possible that if data is not normalized, the column with high values will be given more importance in prediction. To tackle that, we scale the data.

VII. CLASSIFIERS

Classifiers are given training data, it constructs a model. Then it is supplied testing data and the accuracy of the model is calculated. The classifiers used in this paper are:

❖ Random Forest Classifier:

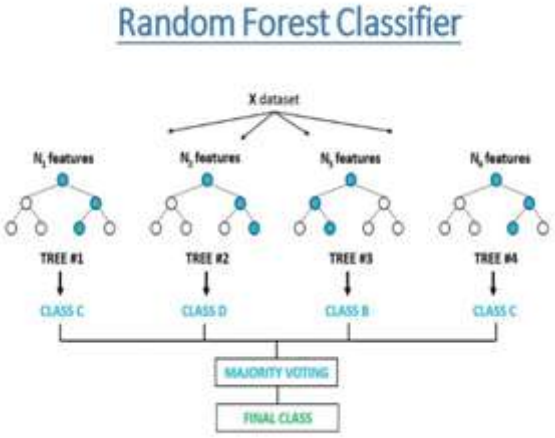
It is a supervised algorithm and a type of ensemble learning program. It is a very versatile algorithm capable of performing regression as well as classification. It is built on decision trees. It

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 6 June 2021, pp: 38-48 www.ijaem.net ISSN: 2395-5252

builds multiple decision trees and merges them for producing the result. In this algorithm, only a subset of features is taken into consideration. It has the same hyper parameters as a decision tree. The advantages of Random Forest are that it works very effectively on a large dataset. It can work for both regression and classification problems. It adds more randomness to the model which makes it a better model. The disadvantage of this model is that it makes use of a large number of trees that make it slow.

Algorithm:

1. Randomly select m features.
2. For a node, find the best split.
3. Split the node using best split.
4. Repeat the first 3 steps and build the forest by repeating last 4 steps.

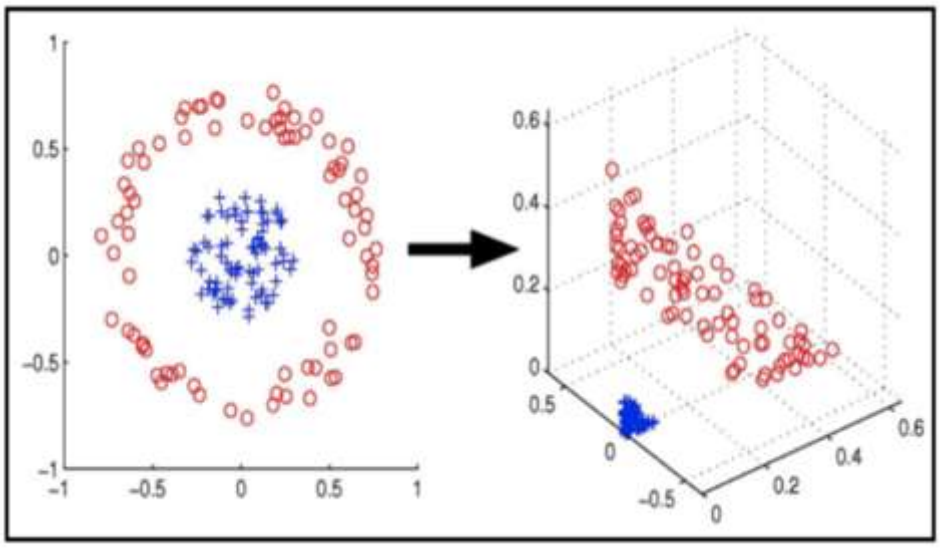


(Fig-3. Workflow diagram of RFC)

❖ SVM (Support Vector Machine):

It is a supervised learning algorithm that classifies cases by a separator. It works by mapping data to high dimensional feature space and then finds a separator. It finds n-space that categorizes data points. This algorithm finds the best plane. This plane must have a maximum margin. The boundary that classifies data points is called hyper planes. The data points are classified based on a position concerning hyper planes. Kernel parameter, gamma parameter, and regularization parameter are tuning parameters of SVM. Linear kernel predicts new input by the dot product between input and support vector. Mapping data to a higher dimensional space is called kernelling. Kernel function can be linear, polynomial, RBF, and Sigmoid. The regularization parameter is the C parameter with a default value of 10. Less regularization means the wrong classification. A small value of gamma means not able to find the region of data. One can improve the model by increasing the importance of the classification of each data. The advantages of SVM are that it is a good algorithm for estimation in high dimensional space and it is very memory efficient. Disadvantages of SVM are that it can suffer from over-fitting and that it works very well on small datasets.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 6 June 2021, pp: 38-48 www.ijaem.net ISSN: 2395-5252



(Fig-4. Workflow diagram of SVM)

❖ KNN (K-nearest neighbour):

It is an algorithm for classifying similar cases. It produces results only when they are requested. Therefore, it is called a lazy learner because there is no learning phase. The advantages of KNN are that it is one of the simplest algorithms as it has to compute the value of k and the Euclidean distance only. It is sometimes faster than other algorithms because of its lazy learning feature. It works well for the multiclass problem. The disadvantages of KNN are that the algorithm may not generalize well as it does not go through the learning phase. It is slower for a large dataset as it will have to calculate sort all the distances from the unknown item. Data normalization is necessary for the KNN algorithm to get the best result.

Algorithm for KNN-

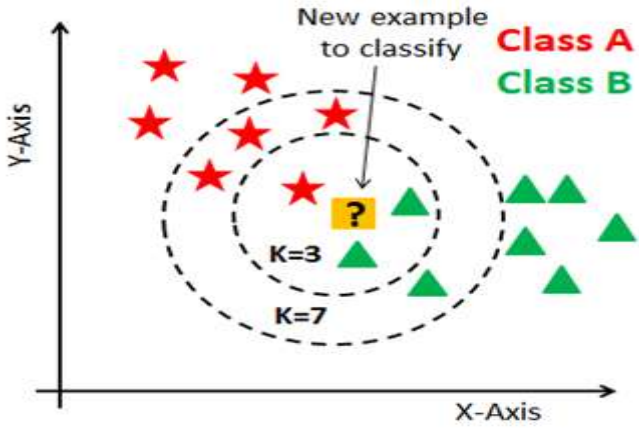
1. Choose k.
2. Calculate the Euclidean distance of all cases from the unknown case.

The Euclidean distance (also called the least distance) between sample x and y is:

Where,

Xi is the i^{th}element of the instance x, Yi is the i^{th} element of the instance y, n is the total number of features in the data set.

3. K number of data points are chosen near unknown data.
4. The unknown data will belong to the majority of cases in chosen k neighbours.



(Fig-5. Workflow diagram of k-NN)

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 6 June 2021, pp: 38-48 www.ijaem.net ISSN: 2395-5252

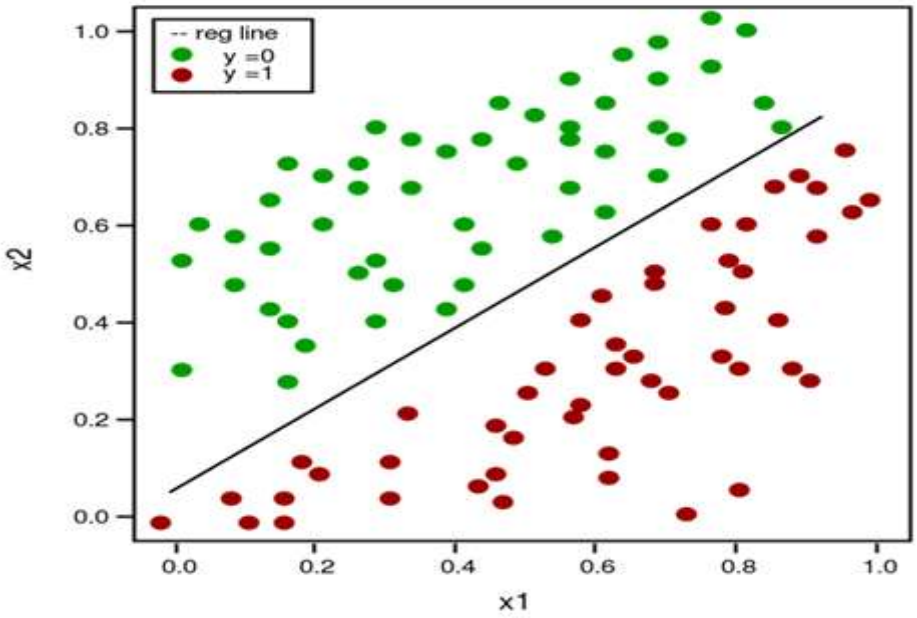
❖ Logistic Regression

This algorithm is used when the response is binary (either 1 or 0). It is used for both binary and multiclass classification. Logistic Regression provides the most accurate results among all but requires finding the best possible feature to fit. In this model, the relationship between Z and the probability of an event is given in 24 as,

Where, pi is the probability that ith case occurs Zi is unobserved continuous variable for ith case Z value is an odd ratio expressed in 24 as:

Zij is the j^{th} predictor of ith case Bj is the j^{th} coefficient

P is the number of predictors



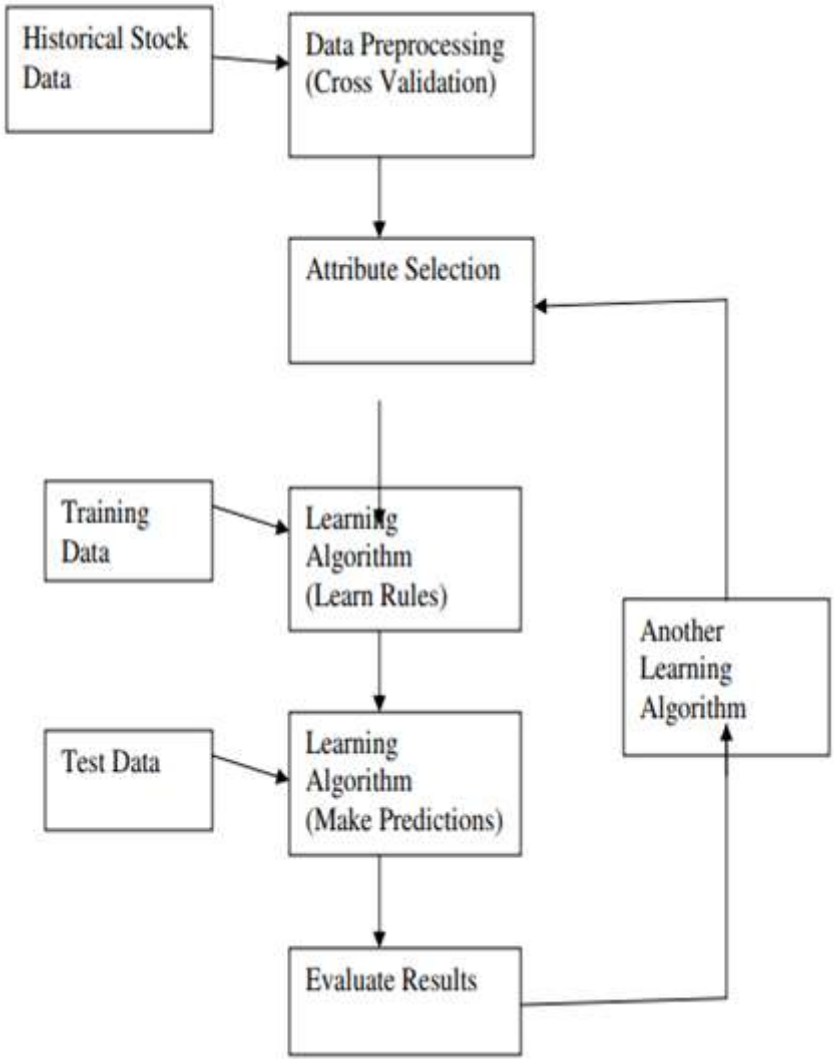
(Fig-6. Workflow diagram of Logistic Regression)

FINAL APPROACH

At present traditional Stock Market Prediction system in the market is a time-consuming and long-lasting process. Investors, sellers & brokers go virtually to the exchange market and they need to undergo have various processes for the trading of stocks. Apart from this these people rely upon various insider tips, Policy Announcements and then they need to wait till the stock attains its Final/Closing Value in the Market, and this process is very time-consuming and lengthy also. To overcome such pitfalls we have to design a system that can give Future Predictions about any stock with the highest possible accuracy. The advancements in technology and internet speed have made services like Stock Predictions, a dream come true for today‘s investor, seller, and broker agencies. The evolution of this technology in the Business world has made a boom in the economic growth of the industry. Stock Market Prediction can be cited as the provision of scoring high profits by easy trading of financial quantities. According to American Stock Market reports, top quality of knowledge storage, data speed, data exchange, profound knowledge about Machine Learning Algorithms, and networking for Stock Marketing Prediction systems (HIS) is compulsory for the efficient performance of the stocks in the market. Especially information storage requirements of stocks are very stimulating. Also, the developing technology and varied solutions in the predicting Stock Market domain necessitated the development of common protocols and standards at the global level.

The methodology of the project contains four major approaches Random Forest Classifier, Support Vector Machine (SVM), k-Nearest Neighbour (kNN), Logistic Regression. Initially, the dataset needs to be entered into the system, they need to get trained. According to their problem, algorithms of Machine Learning will be applied. The algorithm will process the trained data and accordingly, give predictions.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 6 June 2021, pp: 38-48 www.ijaem.net ISSN: 2395-5252

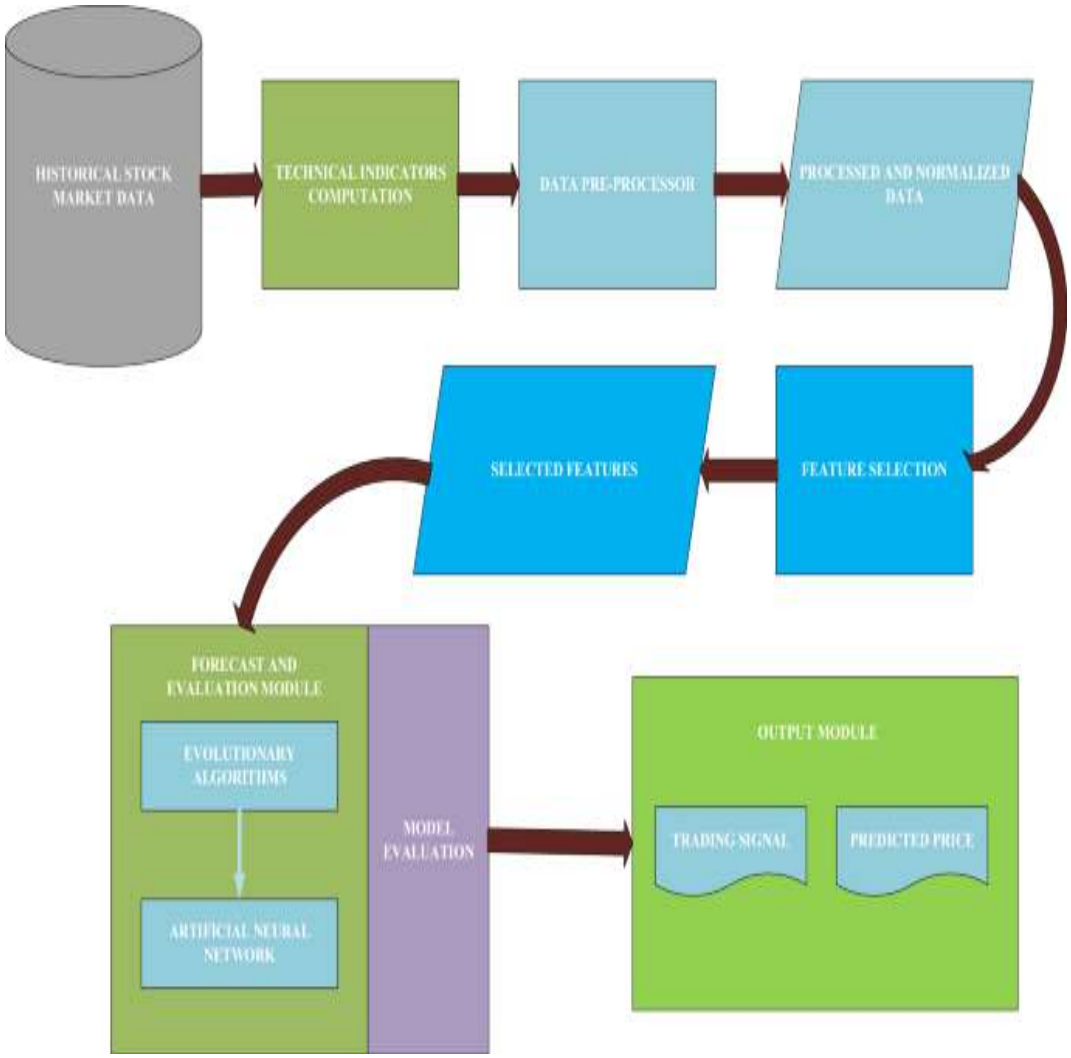


(Fig-7. Workflow diagram of Stock Market Prediction System)

SYSTEM ARCHITECTURE DESIGN

Towards this scope, two traditional deep learning architectures are designed in comparison: a long short-memory network and a temporal convolutional neural model. Based on their predictions, a trading strategy, whose decision to buy or sell depends on two different thresholds, is proposed.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 6 June 2021, pp: 38-48 www.ijaem.net ISSN: 2395-5252



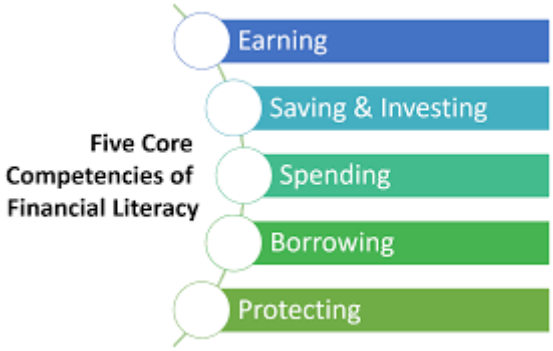
(Fig-8. Architecture Design of Stock Market Prediction System)

CONSUMER E-FINANCIAL LITERACY

Financial literacy is the ability to understand and make use of a variety of financial skills, including personal financial management, budgeting, and investing. It also means comprehending certain financial principles and concepts, such as the time value of money, compound interest, managing debt, and financial planning.

Achieving financial literacy can help individuals avoid making poor financial decisions to help them become self-sufficient and achieve financial stability. Key steps to attaining financial literacy include learning how to create a budget, track spending, pay off debt, and plan for retirement. Educating yourself on these topics also involves learning how money works, setting and achieving financial goals, becoming aware of unethical/discriminatory financial practices, and managing financial challenges that life throws your way.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 6 June 2021, pp: 38-48 www.ijaem.net ISSN: 2395-5252



(Fig-9. Financial Literacy for Stock Market Prediction System)

FUTURE ENHANCEMENT

The future scope of this project will involve adding more parameters and factors like financial ratios, multiple instances, etc. The more the parameters are taken into account more will be the accuracy. The algorithms can also be applied for analysing the contents of public comments and thus determine patterns/relationships between the customer and the corporate employee. The use of traditional algorithms and data mining techniques can also help predict the corporation's performance structure as a whole.

DISADVANTAGES OF THE SYSTEM

- The existing system fails when there are rare outcomes or predictors, as the algorithm is based on sampling.
- The previous results indicate that the stock price is unpredictable when the traditional classifier is used.
- The existence system reported highly predictive values, by selecting an appropriate period for their experiment to obtain highly predictive scores.
- The existing system does not perform well when there is a change in the operating environment.
- It doesn't focus on external events in the environment, like news events or social media.
- It exploits only one data source, thus highly biased.
- The existing system needs some form of input interpretation, thus need of scaling.
- It doesn't exploit data pre-processing techniques to remove inconsistency and incompleteness of the data.

VIII. CONCLUSION

We have successfully built a Stock Prediction Model for effective policing of the investors, broker agencies, and stock sellers. The successful implementation of Machine Learning Algorithms on the trained dataset has given away to the Exchange Market Experts & Traders to invest their money in the market and earn relatively high profits on traded stocks without any dependence on factors like Insider Tips, Brokerage Analysis, Market Researches, etc.

REFERENCES

[1]. Shah D, Isah H, Zulkernine F. Stock Market Analysis: A Review and Taxonomy of Prediction Techniques. Int. J. Financial Stud., 2019, 7(2), pp. 1-22.

[2]. Zhong X, Enke D. Forecasting daily stock market return using dimensionality reduction. Expert Systems with Applications, 2017, vol. 67, pp. 126139.

[3]. Hiransha M, Gopalakrishnan E A, Menon V K, Soman K P. NSE stock market prediction using deep-learning models. Procedia Computer Science, 2018, vol. 132, pp. 13511362.

[4]. Velay M, Fabrice D. Stock Chart Pattern recognition with Deep Learning. ArXiv, 2018.

[5]. Parracho P, Neves R, Horta N. Trading in Financial Markets Using Pattern Recognition Optimized by Genetic Algorithms. 12th Annual Conference Companion on Genetic and Evolutionary Computation, 2010, pp. 2105-2106.

[6]. Nesbitt K V, Barrass S. Finding trading patterns in stock market data. IEEE Computer Graphics and Applications, 2004, 24(5), pp. 4555.

[7]. Leigh W, Modani N, Purvis R, Roberts T. Stock market trading rule discovery using technical charting heuristics. Expert Systems with Applications, 2002, 23(2), pp. 155159.

[8]. Bollen J, Mao H, Zeng X. Twitter Mood Predicts the Stock Market. Journal of Computational Science, 2011, 2(1), pp. 18.

[9]. Seng J L, Yang H F. The association between stock price volatility and financial news sentiment analysis approach. Kybernetes, 2017, 46(8), pp. 13411365.
[10]. Usmani M, Adil S H, Raza K, Ali S S A. Stock market prediction using machine learning techniques. 3rd International
[11]. Conference on Computer and Information Sciences (ICCOINS), 2016, pp. 322-327.
[12]. Patel J, Shah S, Thakkar P, Kotecha K. Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques. Expert Systems with Applications, 2015, 42(1), pp. 259-268.
[13]. Bhardwaj A, Narayan Y, Vanraj, Pawan, Maitreyee D. Sentiment analysis for Indian stock market prediction using Sensex and nifty. Procedia Computer Science, 2015, 70, pp. 8591.
[14]. Ballings M, Poel D V D, Hespeels N, Gryp R. Evaluating multiple classifiers for stock price direction prediction. Expert Systems with Applications, 2015, 42(20), pp. 704656.
[15]. Milosevic N. Equity Forecast: Predicting Long Term Stock Price Movement Using Machine Learning. ArXiv, 2016.
[16]. Luca D P, Honchar O. Recurrent Neural Networks Approach to the Financial Forecast of Google Assets. International Journal of Mathematics and Computers in simulation, 2017, vol. 11, pp. 713.
[17]. Roondiwala M, Patel H, Varma S. Predicting Stock Prices Using Lstm. International Journal of Science and Research (IJSR), 2017, vol. 6, pp. 17541756.
[18]. Yang B, Gong Z J, Yang W. Stock Market Index Prediction Using Deep Neural Network Ensemble. 36th Chinese Control Conference (CCC), 2017, pp. 2628.
[19]. Zhang J, Cui S, Xu Y, Li Q, Li T. A novel data-driven stock price trend prediction system. Expert Systems with Applications, 2018, 97(1), pp. 6069.
[20]. Hossain M A, Karim R, Thulasiram R K, Bruce N D B, Wang
[21]. Y. Hybrid Deep Learning Model for Stock Price Prediction. IEEE Symposium Series on Computational Intelligence (SSCI), 2018, pp. 1821.
[22]. Powell N, Foo S Y, Weatherspoon M. Supervised and Unsupervised Methods for Stock Trend Forecasting. Paper presented at the 40th South-eastern Symposium on System Theory (SSST), 2008, pp. 203-205.
[23]. Babu M S, Geethanjali N, Satyanarayana B. Clustering Approach to Stock Market Prediction. International Journal of Advanced Networking and Applications, 2012, vol. 3, pp. 1281-1291.
[24]. Wu K P, Wu Y P, Lee H M. Stock Trend Prediction by Using K-Means and Aprioriall Algorithm for Sequential Chart Pattern Mining. Journal of Information Science and Engineering, 2014, vol. 30, pp. 669686.
[25]. Peachavanish R. Stock selection and trading based on cluster analysis of trend and momentum indicators. International MultiConference of Engineers and Computer Scientists, 2016, vol. 1, pp. 1618.
[26]. Zaidi M, Amirat A. Forecasting Stock Market Trends by Logistic Regression and Neural Networks Evidence from KSA Stock Market. Euro-Asian Journal of Economics and Finance, 2016, 4(2), pp. 50-58.