JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Prediction System for Stock Market Indices Using Machine Learning

[1]Ashish Premji Nitnaware, [2]Dr. Sharvari Govilkar

[1]Student, [2]Project Guide

[1]Department of Information Technology,

[1]PCE, Mumbai, India

Abstract : The objective of modern finance is to discover effective methods for comprehending, processing, and interpreting stock market data into knowledge that may be used to make better investment decisions. Financial tasks usually have many different and complex structural characteristics which are influenced by a wide range of political and economic circumstances, and frequently have unexpected, dynamic, nonlinear, time-varying, and changeable structural qualities. Powerful software recently developed using machine learning techniques can extract useful data from a huge database. The whole people involved with this domain will not be expert. It is always be a requirement of system which can predict the share trends.in this study researcher developed a system using machine learning classification model namely k Nearest Neighbor (KNN), SVM algorithms, Adaboost, and decision trees. The developed system takes date as input and predicts categorical output for next date market trends.

I. INTRODUCTION

Due to the obvious significant financial benefits involved, stock market forecasting has long been an interesting activity for many experts worldwide. The capability of stock to acquire and respond to information that is quickly reflected in their pricing makes them an absolutely interesting investment option. Academicians and researchers have shown keen interest in studying the predictability of the stock prices, since it throws more light in understanding the behavior and dynamics of the stock market.

1. Introduction to Share market

The stock market is one of the pillars of an economic system. It enables businesses to raise capital through the sale of stock shares and corporate bonds, and it offers investors the ability to gain from the financial success of the company through capital gains and dividend payments.

Recently, significant sums of money are traded globally on the stock market. National economies are closely related and have a significant impact on how their stock markets operate. Consequently, they are linked to macroeconomic factors and have a more obvious effect on daily life. They therefore make up a system that has significant and immediate societal repercussions.

Participants in the stock market want to optimize their investment and produce a healthy return by buying or selling their assets at the ideal moment. The price of a stock security is heavily influenced by a number of factors, including corporate fundamentals, stock demand and supply, government and international policies, inflation, interest rates, etc. Predicting the stock market is difficult since there is constant ambiguity regarding these elements.

The use of computer-based information systems (CBIS) has the potential to enhance a person's capacity for problem solving beyond what is typically possible. Domain expertise and reasoning mechanisms are used in the design of knowledge-based systems to process the knowledge base and allow users to make informed decisions.

Numerous techniques have been used to predict stock market returns in the literature. These approaches can be divided into the following major categories: Techniques for Traditional Time Series Forecasting, Fundamental Analysis, and Machine Learning

1.2Machine Learning

An area of artificial intelligence (AI) and computer science called "machine learning" combines data and algorithms to replicate human learning while gradually increasing the accuracy of the results. Supervised, unsupervised, and reinforcement learning are the three main subcategories of machine learning. One of the most fundamental types of machine learning is supervised learning. In this case, labelled data is used to train the machine learning system. Although accurate data labeling is necessary for this method to work, supervised learning is quite effective when used in the right settings. Unsupervised machine learning has the advantage of using unlabeled data. The datasets may now be made machine-readable without any assistance from a person, allowing the application to work with much larger datasets. Using the reinforcement learning (RL) technique, an entity can learn through trial and error in an enhanced way using knowledge from its own actions and experiences. Machine learning's most transform is training. To find patterns and generate predictions, simply input the prepared data to any machine learning model during training. In order to execute the required task, the model learns from the data. Finding the best algorithm involves significant trial and error; The size, nature, and goal of the data are important considerations when choosing an algorithm.

Four different machine learning techniques, including K Nearest Neighbor (KNN), SVM algorithms, Adaboost, and decision trees, were used in this study to predict the share market. This algorithm was chosen based on evidence from a literature review and the desire to test a novel algorithm on stock market data. In this project, the performance of each algorithm is calculated using a performance matrix, and the results of all algorithms are compared.

1.3 Problem Statement

A Stock Market Prediction Using Machine Learning is a system of prediction of trends of stock (UP/DOWN). The work mainly focuses on investigating the predictability of the stock market indices using various Machine Learning models i.e., k-Nearest Neighbor (KNN), Decision tree, Adaboost and SVM. The developed system can be used as prediction tool for next day share market indices.

II. LITERATURE SURVEY

In literature review it is evident that many of the researchers deployed different prediction techniques in share market prediction few of many are addressed below.

Oyelade Iyinoluwa* and Adewale Olumide Sunday (2019) [16] To help investors choose when to purchase, sell, or keep a company in order to make a profit, this research was driven by the need to foresee the stock market. Using technical indicators, historical (financial) data were first converted from the original stock market data. Using these technical indications, the datasets needed for analysis were created. The Frequent Pattern Growth technique was then used to build frequent patterns. The fuzzy C-means clustering method was employed to create the prediction model based on these common patterns. Finally, to predict stock market fluctuations, the K-Nearest Neighbor classifier classification method was used. The accuracy of the stock market trend forecast was evaluated by validating its results using the hit ratio evaluation metric and Comparative research was done.

Nemes, M. D., & Butoi, A. (2013) [17]. In this study, a series of neural networks created for stock exchange rate forecasting were applied to three Romanian stocks traded on the Bucharest Stock Exchange. An approach known as the multistep forward strategy was used to predict short-term price changes. Later, it is possible to integrate the findings of this study with a data mining and data stream processing model to create an intelligent multi-agent system that can help consumers decide whether to buy or sell stocks.

J. Montenegro, C., & Molina, M. (2020) [18]. The short-term step-by-step technique suggested by this study is to integrate two sources of information that investors can examine before making a choice. First, the index data is used as the training data for a deep learning neural network to represent and predict the value of stocks for the subsequent day. Secondly, using Feature Selection Analysis, this study determines the most representative businesses that are listed on Index and represent the Index behavior inclination. The procedure shows promising outcomes to influence the investor's decision, and the outputs are completed and verified.

Huang, C. Y., & Lin, P. K. (2014) [19]. This study presents a solution to the specific problem. The system incorporates multiple data mining methodologies and supports in stock trade decision-making. Top-down trading theory, Bayesian probability theory, technical analysis, dynamic time series theory, and theory of artificial neural networks are all included in the proposed system. To experimentally analyses the trading return of the suggested system, two scenarios are examined. That was the first time the 240-trading day Taiwan Semiconductor Manufacturing Company (TSMC) data set was used.

Rout, A., Bar, A. K., & Saha, S. P. (2022) [28]. The proposed solution is broad so it uses numerous feature engineering techniques, an RNN-based system for predicting stock market price trends, and pre-processing of the stock market information. Historical prices have been taught and have an accuracy of 84.0 percent in the yearly forecasting model. They have assessed several popular machine learning models and came to the conclusion that, primarily to the extensive feature engineering they developed, the proposed solution exceeds others. Its planning process, carefully considered prediction term lengths, extraction of features, and data pre-processing techniques will assist investors in making less risky stock investments by allowing them to regularly compare the stocks of other companies. Additionally, it will benefit the technical and financial aspects of the stock analysis research field.

13. Selvamuthu, D., Kumar, V., & Mishra, A. (2019) [30]. For stock market prediction using tick data as well as 15-min data of an Indian company, they apply neural networks based on three distinct learning algorithms in this study that are Levenberg-Marquardt, Scaled Conjugate Gradient, and Bayesian Regularization. The outcomes of study are compared.

Shaikh, A., & Shinde, S. (2022) [31]. This study makes stock price predictions for the next trading day and for a certain time. To improve the model's prediction, moving average approach is applied. The most widely used tool among all markers is the moving average. The goal of this study is to demonstrate whether or not examiners and financial backers are appraising the case without qualification. The goal of this tool is to present opportunities and determine whether enhancing future security is worthwhile. Sentiment analysis is the process of mining text for context while identifying and extrapolating subjective information from the source material. A user's opinion on a specific company can be determined via real-time sentiment analysis of stock prices.

Özgül ÖZER Nazlı ARDA(2022)[33] The purpose of this study is to develop several machine learning algorithms and evaluate how well they predict the accurate diagnostic parameters of celiac serological testing. With an accuracy rate of 86.7 percent, decision trees are efficient machine learning methods for predicting probable covariates.

Stock market forecasting has long been a fascinating effort for many researchers all over the world due to the huge financial benefits involved. Studies of the literature have shown that the outcomes of predictions differ between financial markets. An interesting finding is that machine learning technologies consistently beat other statistical techniques. The study predicts that machine learning tools will be more precise than existing methods at predicting changes in stock market values. The most evident benefit of data mining technology is that it is superior to and between 5 and 20% more accurate than traditional statistical methods.

III. PROPOSED SYSTEM

This work is predicting the movement and price of selected Indian stock indices using machine learning algorithms namely k-Nearest Neighbor (KNN), Decision tree, AdaBoost and SVM. Classification model is built for each machine learning algorithm using training data set further classification model is used with respect to user choices for next day trends of share indices.

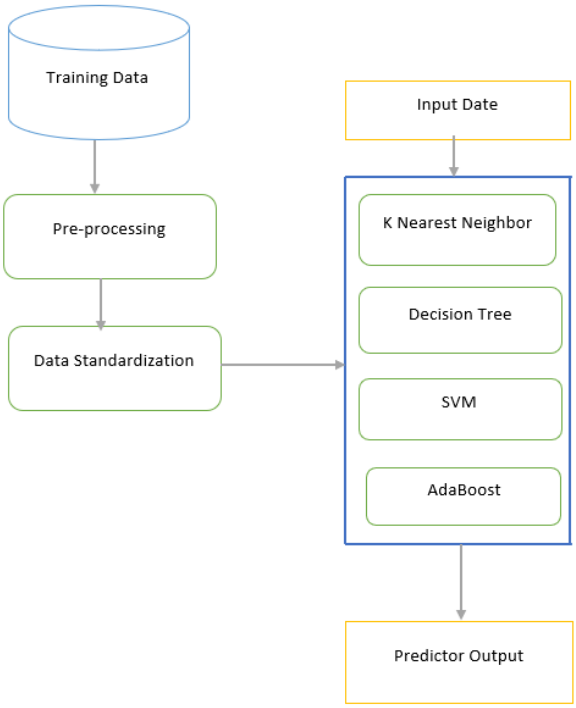

Figure1: Architecture of proposed system

Above figure 1 is the architecture of the proposed system is mainly divided into two sub parts. First is the training part and second is the prediction of trends.

Stock Market Dataset: In this study, the closing values of NSE-NIFTY and BSE stock data are examined in relation to the opening, highest, lowest, and closing prices as predictors. The NSE-NIFTY and BSE stock index data are retrieved individually from the NSE and BSE websites for the years 1990 to 2022. The data is divided into two 75:25 sub-tests. The system can operate on any suitable share market data set.

Preprocessing: This process involves data validation. Selected data set scanned for appropriate data. Developed system gives error message if data is not in appropriate format.

Data Standardization: The predictors are very important in prediction system because it play important role in the accuracy of prediction model. In view of above facts ten commonly used technical indicators, including the stochastic percent K, stochastic percent D, Momentum indicator, Price rate of change, William's percent R, Moving Average Convergence and Divergence (MACD), Price minus Moving Average, A/D oscillator, Price oscillator, and Relative Strength index, are used in this study. Additionally, intraday index movements, such as the low price, open price, high price, and close price, are used as predictors.

Classification model using K-Nearest Neighbor (KNN): The k-NN algorithm ranks among the simplest machine learning techniques. It is based on the theory that items closer to one another will also have same attributes. In this way, it is simple to forecast an object's immediate neighbors given one of its defining characteristics are known.

An enhancement over the nearest neighbour method is the k-NN method, which bases predictions on votes from "k" nearest neighbours rather than just one. The idea behind it is that each new instance can be defined by a simple majority vote among its k neighbours, where k is a positive integer and is often a small number. With respect to the day for which a prediction is to be made, the k-NN method identifies the k nearest neighbours in the training dataset. Once k-NNs have been found, the prediction for that day is calculated as the weighted average of the neighbours average closing prices for the following day. The k-NN method experiments with different values of k in the training dataset to get the ideal value of k that yields the best prediction outcome. The test dataset is then used to apply this predictive model, with the optimal value of k, in order to forecast the closing price for the following day. The output of the predictive model and the actual values of the test dataset are used to generate performance metrics, which are used to measure the effectiveness of the predictive model.

Classification model using Decision tree: Decision tree is a popular decision-making tool and it is a widely used classifier among the data miners. It is based on constructing hierarchical trees from the historical dataset available and the fully grown trees and pruned trees are then used for forecasting new instances. An example of decision tree depicting conventional trading decisions. One of the major attractions of the decision tree lies in its ability to generate rules for decision making [4]. To forecast the class of a given dataset, a decision tree's procedure begins at the root node and goes upward. Based on a comparison between the values of the record (actual dataset) attribute and the values of the root attribute, this method follows the branch and leaps to the next node. After comparing the attribute value to those of the other sub-nodes, the procedure advances to the next node. It continues until it reaches the tree's leaf node.

Classification model using AdaBoost: Machine learning uses the Boosting technique known as the AdaBoost algorithm, sometimes referred to as Adaptive Boosting. The practice of reassigning weights to each instance, with higher weights going to cases that were incorrectly classified, is known as "adaptive boosting." Boosting is used in supervised learning to reduce bias and variance. The sequential learning concept serves as its foundation. Every succeeding learner, with the exception of the first, is created from learners who have already been generated. In other words, a weak learner becomes a strong learner. With one little exception, the AdaBoost algorithm runs according to the same principles as Boost.

Classification model using SVM: The SVM method operates as follows, assuming that the dataset comprises two tags (green/Up and blue/Down), as well as two features (x1 and x2) (In our project it is 14). The pair of coordinates (x1, x2) needs to be split into green/Up and blue/Down categories by a classifier. We can readily discriminate between these two classes using only a straight line because it is a two-dimensional space. A hyperplane is a zone or boundary that is best for a choice that has been determined using

the SVM method. These groupings could be divided along a number of lines, though. We need a classifier that can divide the pair of coordinates (x1, x2) into green/Up and blue/Down categories. Since it is a two-dimensional space, we may easily distinguish between these two classes by utilizing merely a straight line. The SVM algorithm assists in identifying the best line or decision boundary; this best border or region is known as a hyperplane. However, there may be several lines that can split these groups.

IV. CREATING THE ENVIRONMENT FOR IMPLEMENTATION

The proposed method is implemented using Python. The suggested approach is completely implemented using a variety of libraries and technologies. One of them is Matplotlib, a Python plotting library, which is an illustration of a machine learning library. TensorFlow and Sklearn are a couple of the other software packages used in the implementation.

Visually presented information is more rapidly understood than words. Because they don't require any prior knowledge of computing methods, GUIs are easy for non-programmers to use. They don't care about writing or fixing code. Users find GUI to be easy to understand as a result.

An effective graphical user interface (GUI) is created to facilitate effective human–computer interaction which is shown below Figure2 &Figure3.

Figure2: User interface.

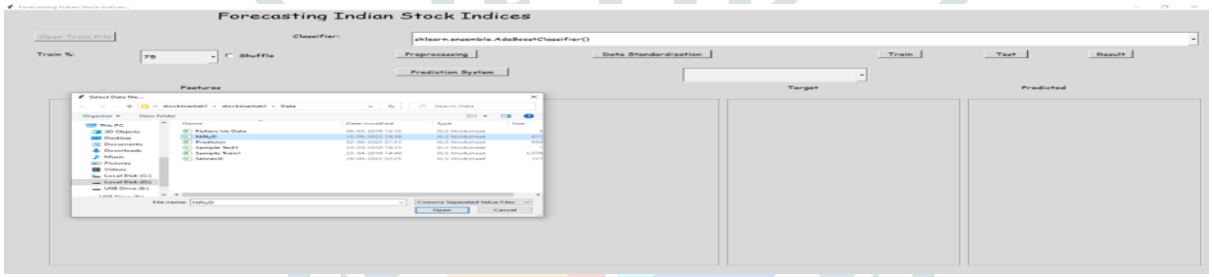

Figure3: Share Indices Data Selection

Data preprocessing is the process of validating data, which involves determining if the data is present in an appropriate format or not. If not, the data must be manually converted into an appropriate format, as shown in figure4.

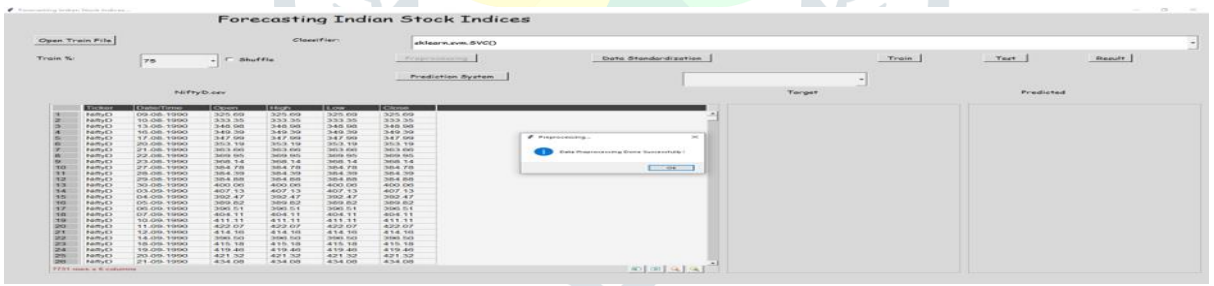

Figure 4: Preprocessing of Data

Data standardization, which transforms data into 10 separate indicators and then serves as a parameter or predictor for the prediction system, is a crucial stage in this approach. The process of data standardization is shown in the figure5.

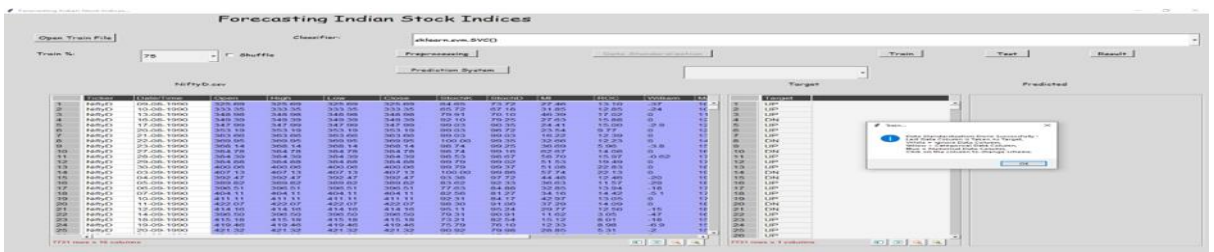

Figure5: Data Standardization.

After data standardization, the next step is to create a classification model. Each classification model needs to be trained in this view, the classification model is trained using 75% of the total data set, as shown in figure 6 below. The model is then ready for use after a number of iterations.

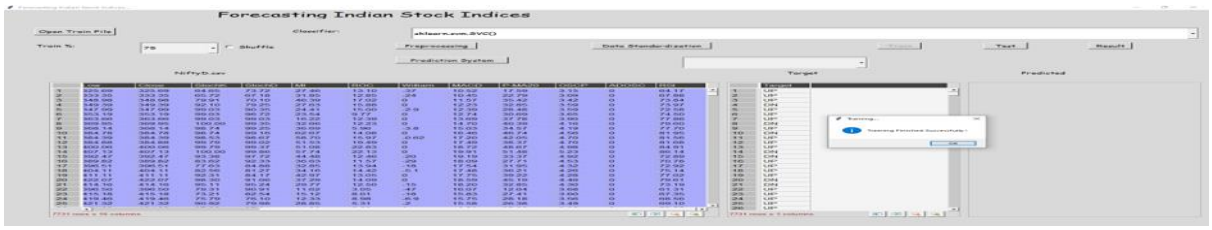

Figure6: Training Data set.

IV. RESULT ANALYSIS

In machine learning, whenever a model is constructed for any classification task, the accuracy of the model is typically the first parameter to be assessed. This is also referred to as classification accuracy, and it describes how accurate the model's predictions are. Classification accuracy can be calculated by dividing the total number of predictions made by the model by the number of correct predictions. Then after achieving significant accuracy model are deployed for prediction of share market indices.

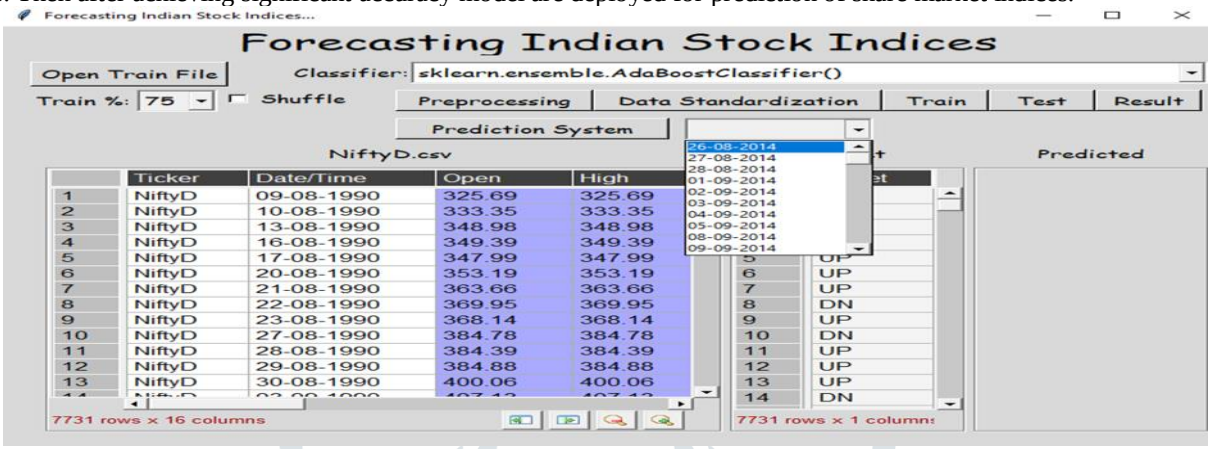

Figure7: Prediction System input.

Above figure no 7 shows the input that is in terms of date given to the system for which the share market indices have to be predict. After gating valid input date user has to select any developed and trained classification model with user choice. Prediction system after processing gives prediction of share market indices for selected date. Which is shown figure8.

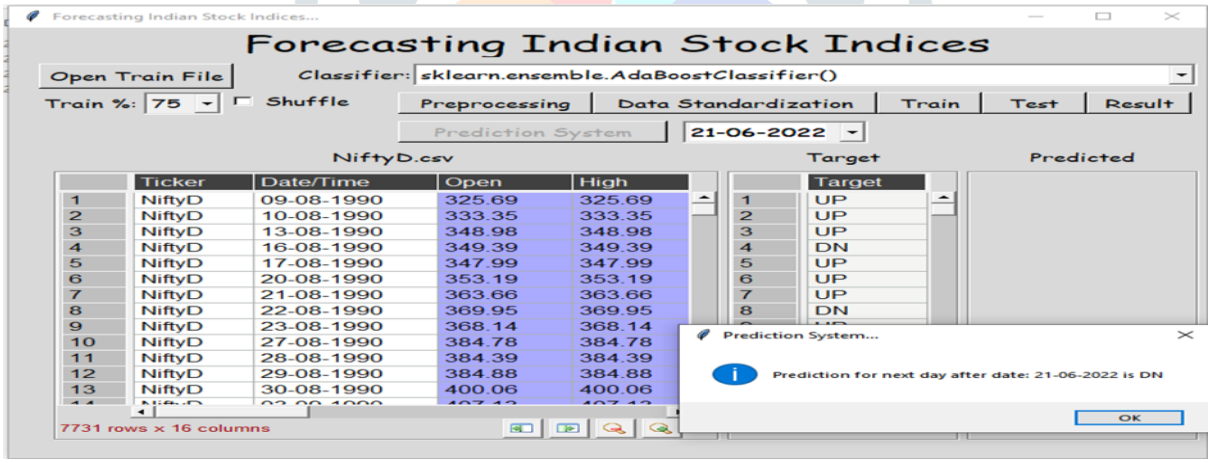

Figure8: Prediction System Output.

V. CONCLUSION

In this study, machine learning techniques were used to forecast future stock indices. In order to predict future stock indices, a KNN, Decision Tree, SVM, and AdaBoost have first been considered. The accuracy, precision, recall, and F1 Score for each prediction class for each algorithm are checked while developing this project. It was found that nearly all classification models performed well on the share market data set. Comparatively SVM and AdaBoost outperform the remaining prediction models.

In developed system technical indicator are used as a predictor which is the major part of this project which can help investor this software as a tool to forecast the market trained. User is having choice as well as facility to use multiple models simultaneously and that to very easy manner. This project can be subsystem for large financial project.

REFERENCES

[1] Le, Thang & Garcia, Rafael & Casari, Paolo & Östberg, P-O. (2019). Machine Learning Methods for Reliable Resource Provisioning in Edge-Cloud Computing: A Survey. ACM Computing Surveys. 52. 1-39. 10.1145/3341145.

[2]K-Nearest Neighbor (KNN) Algorithm for Machine Learning - Javatpoint. (2021, April 2). Www.Javatpoint.Com. Retrieved May 8, 2022, from https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning

[3] Support Vector Machine (SVM) Algorithm - Javatpoint. (2019, February 3). Www.Javatpoint.Com. Retrieved April 6, 2022, from https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm

[4] Machine Learning Decision Tree Classification Algorithm - Javatpoint. (2019, April 6). Www.Javatpoint.Com. Retrieved April 5, 2022, from https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm

[5] Artificial Neural Network Tutorial - Javatpoint. (2021, September 9). Www.Javatpoint.Com. Retrieved April 7, 2022, from https://www.javatpoint.com/artificial-neural-network
[6] Fumo, D. (2018, June 19). Linear Regression — Intro To Machine Learning #6 - Simple AI. Medium. Retrieved April 5, 2022, from https://medium.com/simple-ai/linear-regression-intro-to-machine-learning-6-6e320dbdaf06
[7] Alsammak, I. L. H., Mohammed, A. H., & Nasir, N. S. (2022). E-learning and COVID-19: Predicting Student Academic Performance Using Data Mining Algorithms. Webology, 19(1), 3419–3432. https://doi.org/10.14704/web/v19i1/web19225
[8] Buche, A., & Chandak, M. (2019). Stock Market Forecasting Techniques: A Survey. Journal of Engineering and Applied Sciences, 14(4), 1649–1655. https://doi.org/10.36478/jeasci.2019.1649.1655
[9] Das, D., & Shorif Uddin, M. (2013). Data Mining and Neural Network Techniques in Stock Market Prediction: A Methodological Review. International Journal of Artificial Intelligence & Applications, 4(1), 117–127. https://doi.org/10.5121/ijaia.2013.4109
[10] D. S, S., V, S., M.L.S, V., & Garcia Jacob, S. (2015). Improved Turnover Prediction of Shares Using Hybrid Feature Selection. International Journal of Data Mining & Knowledge Management Process, 5(6), 45–52. https://doi.org/10.5121/ijdkp.2015.5604
[11] Educational Data Classification and prediction using Data Mining Algorithms. (2019). International Journal of Recent Technology and Engineering, 8(3), 8674–8678. https://doi.org/10.35940/ijrte.c6457.098319
[12] Emami, S. S. (2018). Predicting Trend of Stock Prices by Developing Data Mining Techniques with the Aim of Gaining Profit. Journal of Accounting & Marketing, 07(04). https://doi.org/10.4172/2168-9601.1000304
[13] Garg*, P., & Vishwakarma, S. K. (2019a). An Efficient Prediction of Share Price using Data Mining Techniques. International Journal of Engineering and Advanced Technology, 8(6), 3110–3115. https://doi.org/10.35940/ijeat.f9085.088619
[14]Suganthi, R., & Kamalakannan, P. (2015). Analyzing Stock Market Data Using Clustering Algorithm. International Journal of Future Computer and Communication, 4(2), 108–111. https://doi.org/10.7763/ijfcc.2015.v4.366
[15] Gupta, A., Bhatia, P., Dave, K., & Jain, P. (2019a). Stock Market Prediction Using Data Mining Techniques. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3370789
[16] Iyinoluwa, O. (2019). Stock Market Trend Prediction Model Using Data Mining echniques. Current Trends in Computer Sciences & Applications, 1(5). tps://doi.org/10.32474/ctcsa.2019.01.000122
[17] Nemes, M. D., & Butoi, A. (2013). Data Mining on Romanian Stock Market Using Neural Networks for Price Prediction. Informatica Economica, 17(3/2013), 125–136. https://doi.org/10.12948/issn14531305/17.3.2013.11
[18] Montenegro, C., & Molina, M. (2020). Improving the Criteria of the Investment on Stock Market Using Data Mining Techniques: The Case of S&P500 Index. International Journal of Machine Learning and Computing, 10(2), 309–315. https://doi.org/10.18178/ijmlc.2020.10.2.936
[19] Huang, C. Y., & Lin, P. K. (2014). Application of integrated data mining techniques in stock market forecasting. Cogent Economics & Finance, 2(1), 929505. https://doi.org/10.1080/23322039.2014.929505
[20] Humayun, A., & Waqar, A. (2017). A Comparative Study on Usage of Data Mining Techniques in the Healthcare Sector. International Journal of Computer Applications, 162(6), 13–15. https://doi.org/10.5120/ijca2017913331
[21] Iyinoluwa, O. (2019). Stock Market Trend Prediction Model Using Data Mining Techniques. Current Trends in Computer Sciences & Applications, 1(5). https://doi.org/10.32474/ctcsa.2019.01.000122
[22] Kaur, J. (2021). Crime Analysis and Prediction Using Data Mining and Machine Learning Techniques. International Journal for Research in Applied Science and Engineering Technology, 9(10), 1259–1270. https://doi.org/10.22214/ijraset.2021.38558
[23] Kaur, N. (2017). Prediction of Stock Market Price using Neural Network. IJARCCE, 6(1), 308–311. https://doi.org/10.17148/ijarcce.2017.6159
[24] Khedr, A. E., S.E.Salama, & Yaseen, N. (2017). Predicting Stock Market Behavior using Data Mining Technique and News Sentiment Analysis. International Journal of Intelligent Systems and Applications, 9(7), 22–30. https://doi.org/10.5815/ijisa.2017.07.03
[25] Kumar, D. S. (2015). A Comparative Study of Various Data Transformation Techniques in Data Mining. International Journal of Scientific Engineering and Technology, 4(3), 146–148. https://doi.org/10.17950/ijset/v4s3/305
[26] Lakshmi, K. (2013). Utilization of Data Mining Techniques for Prediction and Diagnosis of Tuberculosis Disease Survivability. International Journal of Modern Education and Computer Science, 5(8), 8–17. https://doi.org/10.5815/ijmecs.2013.08.02
[27] Preetha, R., & Vinila Jinny, S. (2021). Breast-cancer prediction strategies and experimental processing using DEFS algorithm. Materials Today: Proceedings, 47, 207–213. https://doi.org/10.1016/j.matpr.2021.04.097
[28] Rout, A., Bar, A. K., & Saha, S. P. (2022). Stock Market Prediction using Machine Learning Algorithm. IJARCCE, 11(3). https://doi.org/10.17148/ijarcce.2022.11339
[29] S., G., Dudhwala, N., Jadhav, K., Gabda, P., & Kishor, B. (2016). Prediction of Stock Market using Data Mining and Artificial Intelligence. International Journal of Computer Applications, 134(12), 9–11. https://doi.org/10.5120/ijca2016907635
[30] Selvamuthu, D., Kumar, V., & Mishra, A. (2019). Indian stock market prediction using artificial neural networks on tick data. Financial Innovation, 5(1). https://doi.org/10.1186/s40854-019-0131-7
[31] Shaikh, A., & Shinde, S. (2022). Prediction of Stock Market Prices using Prediction algorithm and Sentiment Analysis. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4111874
[32] DataCamp. (2020, May 2). AdaBoost Classifier Algorithms using Python Sklearn Tutorial. Https://Www.Datacamp.Com. Retrieved July 18, 2022, from https://www.datacamp.com/tutorial/adaboost-classifier-python
[33] ÖZER, Z., & ARDA, N. (2022). Comparision of Different Machine Learning Algorithms to Predict the Diagnostic Accuracy Parameters of Celiac Serological Tests. Sakarya University Journal of Computer and Information Sciences. https://doi.org/10.35377/saucis. . .1094043
[34] Fundamental analysis. (2019, April 5). Http://Capitalveda.in/Wp-Content/Uploads/2018/09/FA.Png. Retrieved May 5, 2022, from http://capitalveda.in/wp-content/uploads/2018/09/FA.png
[35] Technical analysis. (2020, April 6).
Https://School.Stockcharts.Com/Lib/Exe/Fetch.Php?Media=overview:Technical_analysis:Ta1-Ibm-Trendingtrading.Png. Retrieved March 8, 2022, from https://school.stockcharts.com/lib/exe/fetch.php?media=overview:technical_analysis:ta1-ibm-trendingtrading.png