

Stock Market Analysis using Supervised using Machine Learning Technique

Jayashree¹, A Sanchith Savio², Krishna Neelaraddi³, Ganesh R⁴, H Puneeth Sheety⁵ and Shyamashri Das⁶

¹⁻⁶Nitte Meenakshi Institute of Technology, Bangalore, India

Email: jayashree.cse@nmit.ac.in

Abstract—The stock market is extremely volatile and complex in nature, yet it is crucial to the financial element of the country's progress. Important political concerns, analyst calls, news stories, firm development and growth plans, and many more have an impact on it. Therefore, any investor would be interested in learning how the stock market has changed through time and how the aforementioned elements have an impact on the stock market's behavior. Millions of traders invest in the stock market each working day. While some of these investors do well, the majority do not. Loss or gain, however, is uneven when taking any trading day into account. There is a huge need for stock market analysis because it is so difficult to anticipate stock values. This project focuses on using statistical indicators to analyse a stock of any specific firm. These metrics come in both probabilistic and deterministic varieties. The goal of this project is to maximize profit while minimizing the risk of loss in every deal.

Index Terms— Data Science Algorithms, Supervised algorithm, Stock analysis, Stock Values, Stock prediction.

I. INTRODUCTION

The term "stock" has probably been used by all of us at some point. In particular, stock is connected to partners and businesses that have undergone commercialization and are establishing themselves in the market economy. Share, another term for stock that is often used in everyday speech, is another option. People even refer to it as an investment strategy since they believe it to be a long-term investment that will guarantee and offer a comfortable retirement income.

The greatest approach to discovering dependability is to be aware of the market environment and do research into it, which is why many agents have made a career out of it and are very successful. They provide predictions and offer advice, but their fees are greater and the stock assessment is never less than uniform. Even within a single day, the market might experience several highs and lows depending on the resources, timing, and actions of external and internal agents.

A significant component of contemporary industry and research is machine learning (ML). The use of algorithms [32] and neural network models enhance the performance of computer systems over time. Machine learning algorithms use sample data, also referred to as "training data", to automatically develop a mathematical model without being explicitly trained to make particular conclusions.

The project's only goal was to develop and implement the next method of computing. The first one deals with data extraction, which is accomplished by removing data and data chunks from the data set or database. The source training is completed and categorized in the second flow. The supervision is carried out at the same time,

and the creation of the yield—which offers the outcome after calculation of the same—comes last.

II. LITERATURE SURVEY

Software development requires legitimate materials that are readily available. This section aids in learning about the information that has been developed and identifying its application and execution in the present. The economics and the quality of the product are crucial to development. The support and resource flow must be tracked and calculated after the invention has completed the construction phase. The research phase, as it is sometimes called, is when all the research is integrated and carried out.[28][29]

A. Stock Market Analysis by Machine Learning

Kranthi Sai Reddy carried out work on the trading of stocks is one of the most important financial activities. The act of attempting to forecast the future price of a business or other asset type traded on a securities exchange is known as stock market prediction. The prediction of a stock using ML algorithms is described in this article. Most brokerages utilise technical, structural, or time series research when making stock projections. Machine learning is used to provide stock market forecasts using the programming language Python.[8]

B. Showing the Stock Market Values Using AI Methods

Ronald Lufuno Marwala et al. conducted study for a thesis that was presented to the University of the Witwatersrand Faculty and the Built Structures in Johannesburg in order to satisfy the criteria for the Master of Science in Engineering degree. The weak version of market efficiency states that it is difficult to predict a property's future price using information from its past pricing. In other words, the market operates on a chaotic basis, which makes forecasting challenging. Financial forecasting is a difficult task due to the financial system's inherent complexity.

C. Indian stock market prediction using artificial neural networks on tick data

The study was conducted by Dharmaraja Selvamuthu, Vineet Kumar, and Abhishek Mishra at the mathematics department of the Indian Institute of Technology Delhi in Hauz Khas, New Delhi, India (110016). A stock market is a venue where shares and derivatives of a corporation can be traded at a predetermined price. The basis of the stock market is the supply and demand for shares. The stock market is one of the global industries with the fastest growth rates. Many people nowadays are employed in this industry, either directly or indirectly. Trends in the market must be understood.\cite{b8}

D. Stock Market and its Investment opportunities

The study by Manh Ha Duong Boriss Siliverstovs examines the relationship between stock prices and ultimate investments in the Netherlands and the United Kingdom, two significant European countries. As European financial markets grow increasingly linked, it is projected that there would be an even stronger correlation between share prices in different European countries. This approach may also lead to convergence in the economic growth of European countries if stock market patterns have an effect on actual economic fundamentals like investment and consumption.

E. KNN recommender systems

The KNN or K nearest neighbor algorithms produce suggestions by classifying people and things based on similarity metrics like Cosine similarity and Pearson similarity. User-based KNNs group K users who have scored related items into one cluster in order to predict the score of an item by the target user. Item-based KNN locates the top k-similar things that the user has already scored in order to estimate the score of an item. An extremely effective and user-friendly tool for creating suggestions is the KNN algorithm. However, these recommend-er systems are prone to cold start issues, data sparsity issues, and exclusively proposing well-liked goods. By integrating using a context-sensitive KNN recommendation system, the cold start problem can be resolved. When there are many users and things in a system, only a tiny fraction of the data are connected, the issue of information sparsity occurs. As a result, similarity metrics are less reliable. One might consider the amount of convergence, the average of all the items, the time interval of scoring, and the interval of item scoring while addressing the data sparsity problem.}

Recurrent neural networks are what it is. Three "gate" structures make up the unique network structure known as LSTM. An LSTM unit has three gates: an input gate, a forgetting gate, and an output gate. Information may be chosen by rules when it enters the LSTM network. Information that does not comply with the algorithm will be erased by the forgetting gate, leaving only the data that does. The historical data that were collected from the

Internet and used as experimental data in this study. The experiments made use of three data sets. It is necessary to find an optimization method with a quicker convergence rate and less resource requirements.[24]

- Used the LSTM neural network with automated encoder and the embedded layer of LSTM (long short-term memory).
- To prevent gradients from bursting and disappearing, LSTM is employed in place of RNN.
- In this project, the model is trained in Python, and the input dimensions are reduced using MATLAB. Data is stored and retrieved using MySQL as a data-set.
- The table of historical stock data includes data on the starting price, maximum price, lowest price, closing price, transaction date, volume, and other metrics.
- LSTM model's accuracy in this project is 57%.

F. Deep Learning

It is possible to extract features using convolution neural networks (CNN). They may be used to the semantic analysis procedure, which may be useful for interpreting text data provided by users in the form of reviews. They can learn about product attributes and consumer behavior from text reviews. CNN can perform better when combined with an attention network while learning features from text input. CNN's attention-based algorithms may be used to extract the most eye-catching textual traits or hashtags for use in recommendation systems. Concise assessments from real reviews may be produced using attention-based deep learning networks. In order to generate ratings, CNN may utilize these condensed evaluations to anticipate user preferences and commercial success criteria on its own. \cite{b31}

Recurrent neural networks (RNN) may process sequential or time series data. RNNs have a design that allows them to do computations using the results of earlier rounds. However, they simply pay attention to the outcome of a single iteration and do not take the long-term results into account. To create suggestions, RNNs can be organised hierarchically to contain details about the order and spacing of the user's ratings. The two levels of the hierarchical structure are a layer for short-term events and a layer for long-term events. \cite{b22}

Deep learning networks and matrix factorization (MF) techniques may also be used to predict ratings from both linear and non-linear viewpoints. Here, the deep learning model learns nonlinear interactions between latent features while the MF learns linear interactions using the user-item interaction data. In order to predict ratings, deep learning networks may be utilised to extract the most crucial user preference data and item attributes. The decision to make a certain recommendation to a specific user can therefore be supported by these characteristics. By combining a range of user-interest representations, including user embedding (latent factors), item-level representations, neighbor-assisted representations, and category-based representations, deep learning algorithms may also be used to learn about users' interests. The variability of user preferences is thus better accounted for by blending these representations using either early fusion or late fusion. Deep learning techniques may also be used to simulate a user's dynamic interest in the next item recommendation by dynamically combining the user's general preferences with sequential patterns.[22]-[27]

G. Support Vector Machine (SVM)

SVM is one of the most used supervised learning algorithms. It is one of the most popular training sets that may be used to classify data and address regression-related issues. It is one of the most frequent training sets, which can be used for categorising data and regression concerns. Also, basically, it is utilised for categorising challenges in Machine Learning. Building the optimum regression line or line that can divide an n-dimensional space into classes in order to quickly and easily place a data frame in the appropriate category is the aim of the SVM technique. The term "hyperplane" refers to this ideal choice limit. The local maxima that help create the hyperplane are identified via SVM. The method is known as SVM because these extreme situations are known as vertices.[25]

H. Random Forest

Random forest and other supervised machine learning algorithms are commonly used in classification and regression problems. Random forest is a directed Algorithm that's also employed often in Machine learning challenges. It creates decision trees on general populations and employs appropriate simple majority for classification and mean into regression issues. One of the really essential properties of the Decision Trees is that this could handle the information source comprising outcome variable as the average case and predicted values as in the classification step. It gives superior outcomes for categorization issues.[8]

I. Decision Tree

Decision tree classifier learning approach which might be utilised in both classification concerns, but normally it is indicated for handling Classification problems. It is a plant predictor, where core nodes indicate the attributes of a data-set, branches reflect underlying formulas and that each grid cell reflects the result. In a Regression model, there will also be 4 parts, it's these Transformation To convert a set and Leaf Node.

III. ALGORITHM DESIGN AND METHODOLOGY

A. Comparative Analysis

Choosing of the algorithms to be used in the application. Conducted a comparative analysis on various algorithms, as a result, the Algorithms chosen are KNN, SVM-Linear, RBF, Linear Regression, Random Forest, Decision Tress, LSTM. We split the data-set into two parts for training and testing them as per the algorithms. Trained data-set using the Algorithm. After the above we calculate the Mean Squared Error(MSE) of the algorithms, to find their deviation or error from the actual data.

$$MSE = \left(\frac{1}{n} \star \Sigma(actual - forecast)^2\right) \quad (1)$$

Where P n is number of items,
 Σ is summation notation,
actual = original or observed y-value,
forecast = y-value from regression.

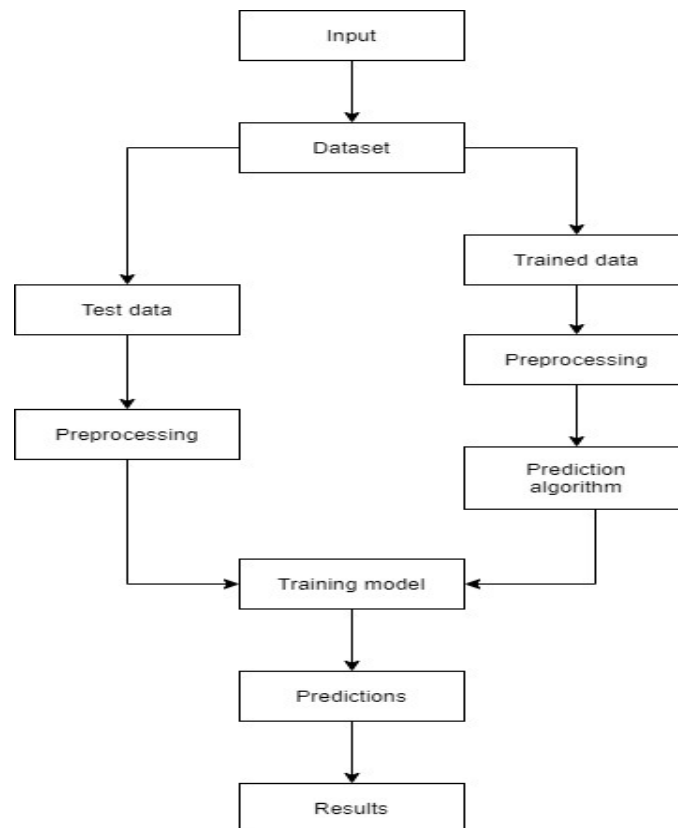


Figure 1. Block Diagram of Comparative Analysis

B. Web Application Development

Developed a Web Application which runs on Flask server, for the visualization and plotting of the stock values. The values are retrieved from algorithms chosen during the comparative analysis. Thus creating an Application for Stock Market Analysis. After development this application can be supplied to any Stock Market enthusiast, who want to analyze the variations and trends in stock.

After development this application can be supplied to any Stock Market enthusiast, who want to analyze the variations and trends in stock.

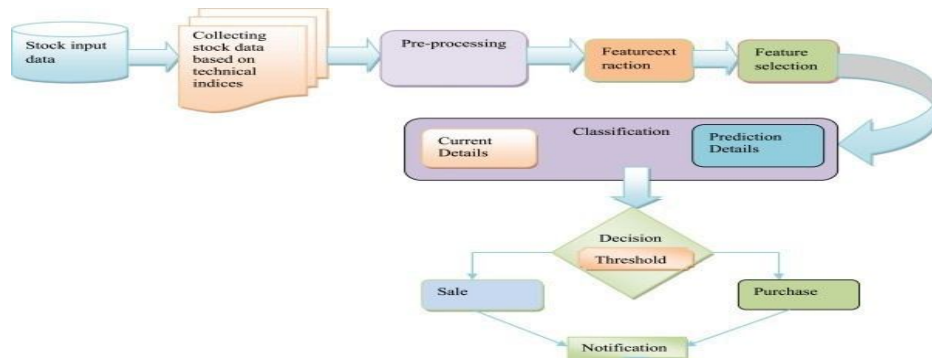


Figure 2. Methodology of Stock Market Analysis

C. Description of Process

It is taken the Yahoo Stock Data-set website for the retrieval of Stock values of various companies. The format of data is like this:

1. The stock values of companies on at a particular date.
2. Opening and Closing values of the stock.
3. High and low value of the stock.

Fig 2 Methodology of Stock Market Analysis. Choosing of the algorithms to be used in the application. Conducted a comparative analysis on various algorithms, as a result, the Algorithms chosen are:

- KNN
- SVM- Linear, RBF
- Linear Regression
- Random Forest
- Decision Tress
- LSTM

We split the data-set into two parts for training and testing them as per the algorithms. Trained data-set using the Algorithm. Developed a Web Application which runs on Flask server, for the visualization and plotting of the stock values. Thus creating an Application for Stock Market Analysis. [30]

D. Pseudo Code

1. Read all stock files in directory of Stocks csv file
2. Convert an array of values into a data-set matrix.
3. For the data set selected, create data-set from the dataframe, reshape, train the selected model.
4. Calculating, inverting the results.
5. Calculate the RMS value.

IV. EXPERIMENT AND RESULT

Experiment includes following steps as-

- Data is taken from website which is in the form of csv file.
- Feature extraction depends upon the algorithms used here.
- Similarly divide the data-set for testing and training.
- Visualize the results and Analyze.

Technologies used for Application Development -HTML,CSS, JavaScript, Bootstrap, Flask

A. HTML

HTML is used for building web addresses and web programs. Below offered is the example of HTML.

B. CSS

CSS (Cascading Style Sheets) allows you to design stunning web sites, working of it is. We can observe This article teaches what CSS utilised a basic syntax demonstrations and also covers numerous crucial characteristics of the technology. CSS is a framework for getting to know how documents are showed to users.

C. JavaScript

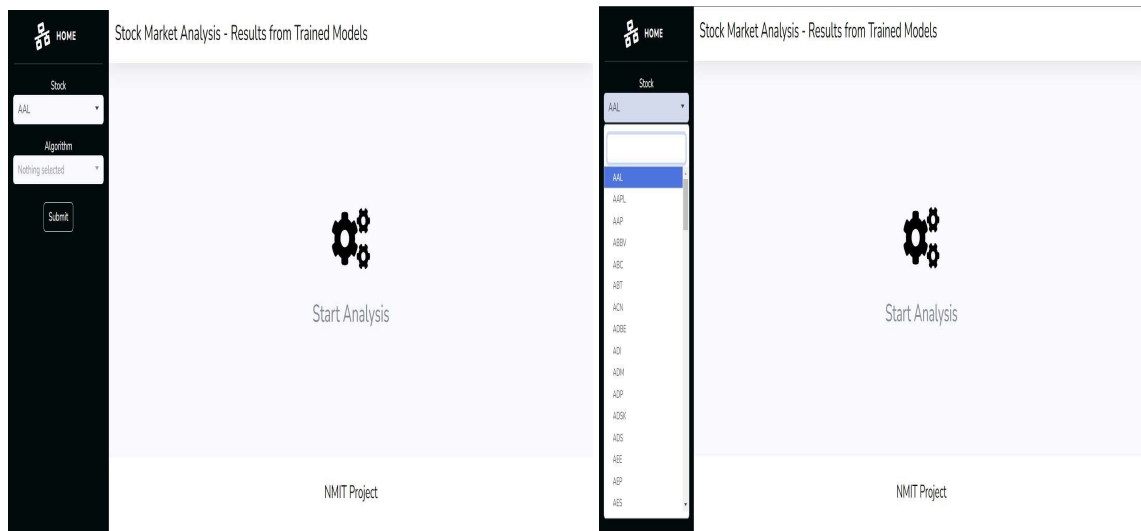
JavaScript is interface that enables users to enact way of dealing on webpages — so each time a web application seems to be doing more than persists there and demonstrate critical data for you all to look at — showing timely manner better content, user behavior, hyper stylized letterier graphics, browsing clip music boxes, etc. — you can genuinely think that JavaScript is likely involved. It is the third component of the layer cake of common network skills, both under which (HTML and CSS).

D. Bootstrap

Bootstrap is the most popular HTML, CSS, and JavaScript framework for building adaptable and mobile-friendly websites. It is a front-end framework intended to speed up and simplify web development. [26] Web framework Flask is a Python package that makes it simple to create web applications.

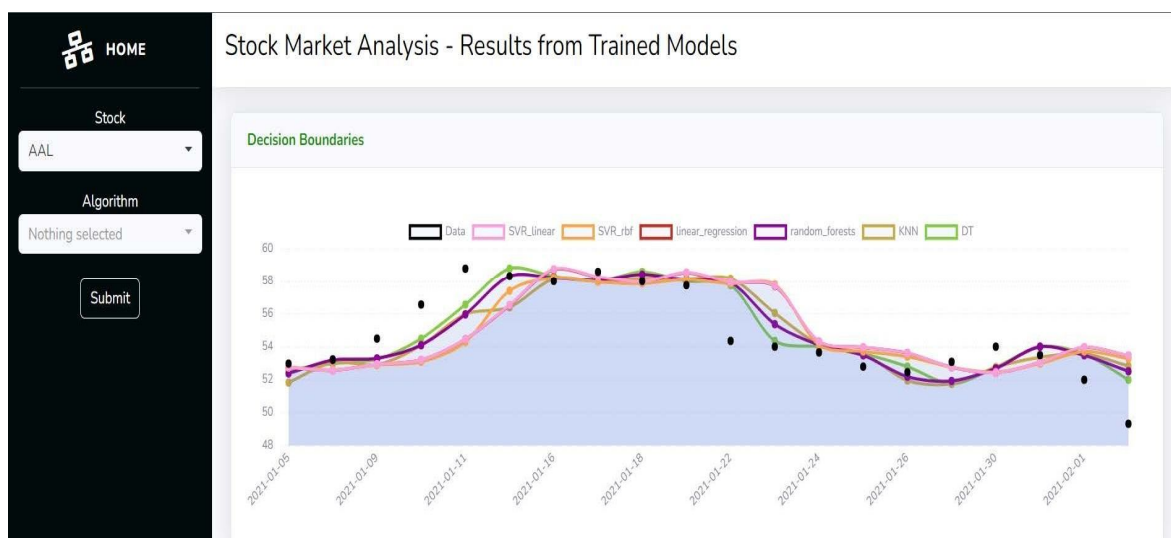
E. Results

The results of the Web Application which were built with description. Screen shot of home page of the Application as shown in the figure 3a from it has to select name of the company and algorithm used for analysis as shown in figure 3b and 3d respectively. Figure3c shows Visualization of stock values using different algorithms and figure3e shows the the comparison of Actual and Predicted Data.

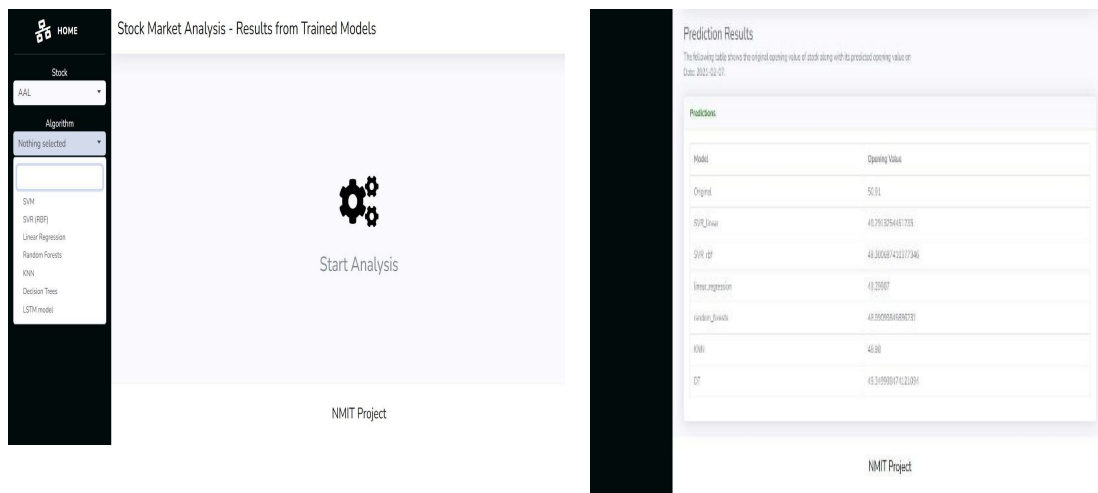


(a) Home page of the application

(b) Choose the Company of which the Stock value has to be analysed



(c) Visualization of the Stock Value



(d) Choose the Algorithm, through which the visualization and analyzing of the Stock.

(e) The comparison of Actual and Predicted Data

Fig. 3: Snapshots of web application results

V. CONCLUSIONS

We have provided a quick overview of some of the most prominent recommend-er system strategies in this post. Additionally, we tried to summarise the most recent advancements in CNN and deep learning. We've made an attempt to emphasise how different strategies might be used to overcome the challenges of creating a recommend-er system. Additionally, we have included a brief explanation of the assessment criteria and procedures that may be utilised to test your recommender systems. We have also made an effort to give a brief outline of some of the most typical issues encountered while developing recommend-er systems. At the conclusion, we want to emphasise that not all sophisticated algorithms provide superior outcomes. As these designs are complicated and may be computationally expensive, one should pick an architecture based on his or her needs rather than on what is cutting edge.

ACKNOWLEDGMENT

This material is based upon work supported by the Google Cloud Research Credits program with the award GCP19980904.

REFERENCES

- [1] R. Ahuja, H. Rastogi, A. Choudhuri and B. Garg, "Stock market forecast using sentiment analysis," 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom), 2015.
- [2] I. Parmar et al., "Stock Market Prediction Using Machine Learning," 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), 2018.
- [3] I. Parmar et al., "Stock Market Prediction Using Machine Learning," 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), 2018.
- [4] A. Sharma, D. Bhuriya and U. Singh, "Survey of stock market prediction using machine learning approach," 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA), 2017.
- [5] M. K. Sophan, M. N. Fitriantama and S. Wahyuni, "Collaborative Filtering Item Recommendation Methods based on Matrix Factorization and Clustering Approaches," 2020 10th Electrical Power, Electronics, Communications, Controls and Informatics Seminar (EECCIS), 2020.
- [6] Patel, Yagnesh G. and Vishal P. Patel. "A Survey on Various Techniques of Recommendation System in Web Mining." (2015).
- [7] N. Sharma and A. Juneja, "Combining of random forest estimates using LSboost for stock market index prediction," 2017 2nd International Conference for Convergence in Technology (I2CT), 2017.
- [8] Sachin Gavhane, Kartik Sharma, Akilesh Rao, Rushabh Shah, Sujay Sail, Stock Market Analysis, International Journal for Research in Applied Science and Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887 Volume 6 Issue V, May 2018.

- [9] B. Panwar, G. Dhuriya, P. Johri, S. Singh Yadav and N. Gaur, "Stock Market Prediction Using Linear Regression and SVM," 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 2021.
- [10] Q. Yunneng, "A new stock price prediction model based on improved KNN" 2020 7th International Conference on Information Science and Control Engineering(ICISCE), 2020.
- [11] Ruslan Salakhutdinov, Andriy Mnih, and Geoffrey Hinton. 2007. Restricted Boltzmann machines for collaborative filtering.
- [12] H. N. Shah, "Prediction of Stock Market Using Artificial Intelligence," 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), 2019.
- [13] R. Jindal, N. Bansal, N. Chawla and S. Singhal, "Improving Traditional Stock Market Prediction Algorithms using Covid-19 Analysis," 2021 International Conference on Emerging Smart Computing and Informatics (ESCI), 2021.
- [14] Z. Liu, Z. Dang and J. Yu, "Stock Price Prediction Model Based on RBFSVM Algorithm," 2020 International Conference on Computer Engineering and Intelligent Control (ICCEIC), 2020.
- [15] Z. Sharifi, M. Rezghi and M. Nasiri, "New algorithm for recommender systems based on singular value decomposition method," ICCKE 2013.
- [16] C. Shi, B. Hu, W. X. Zhao and P. S. Yu, "Heterogeneous Information Network Embedding for Recommendation," in IEEE Transactions on Knowledge and Data Engineering, vol. 31, no. 2, pp. 357-370, 1 Feb. 2019.
- [17] H. Zhang, I. Ganchev, N. S. Nikolov, Z. Ji and M. O'Droma, "FeatureMF: An Item Feature Enriched Matrix Factorization Model for Item Recommendation," in IEEE Access, vol. 9, pp. 65266-65276, 2021.
- [18] K. Senthamarai Kannan, P. Sailapathi Sekar, M.Mohamed Sathik and P. Arumugam, "Financial stock market forecasting using data mining Techniques", 2010.
- [19] Tiffany Hui-Kuang yu and Kun-Huang Huarng, "A Neural network-based fuzzy time series model to improve forecasting", Elsevier, 2010.
- [20] Md. Rafiul Hassan and Baikunth Nath, "Stock Market forecasting using Hidden Markov Model: A New Approach", Proceeding of the 2005 5th International conference on intelligent Systems Design and Application 0-7695-2286-06/05, IEEE 2005.
- [21] Bonde, Ganesh, and Rasheed Khaled. "Extracting the best features for predicting stock prices using machine learning." Proceedings on the International Conference on Artificial Intelligence. The Steering Committee of The World Congress in Computer Science, (WorldComp), 2012.
- [22] Raj, J.S., Palanisamy, R., Perikos, I. and Shi, Y., 2022. Intelligent Sustainable Systems: Proceedings of ICISS 2021. Springer.
- [23] S. Vijaya Shetty, Khush Dassani, G. P. Harish Gowda, H. Sarojadevi, P. Hariprasad Reddy, Sehaj Jot Singh. "Chapter 35 A Review of the Techniques and Evaluation Parameters for Recommendation Systems", Springer Science and Business Media LLC, 2022
- [24] Wang, G., Wei, W., Jiang, J., Ning, C., Chen, H., Huang, J., Liang, B., Zang, N., Liao, Y., Chen, R. and Lai, J., 2019. Application of a long short-term memory neural network: a burgeoning method of deep learning in forecasting HIV incidence in Guangxi, China. *Epidemiology and Infection*, 147.
- [25] Huang, Wei, Yoshiteru Nakamori, and Shou-Yang Wang. "Forecasting stock market movement direction with support vector machine." *Computers and operations research* 32, no. 10 (2005): 2513-2522.
- [26] Gaikwad, Suraj Shahu, and P. R. A. T. I. B. H. A. Adkar. "A review paper on bootstrap framework." *IRE Journals* 2, no. 10 (2019): 349-351.
- [27] S. Vijaya Shetty, Khush Dassani, G. P. Harish Gowda, H. Sarojadevi, P. Hariprasad Reddy, Sehaj Jot Singh. "Chapter 35 A Review of the Techniques and Evaluation Parameters for Recommendation Systems", Springer Science and Business Media LLC, 2022
- [28] Vinaydeep Kaur, Akriti Tyagi, M Kritika, Pratima Kumari, Sanket Salvi. "Crowd-sourcing based android application for structural health monitoring and data analytics of roads using cloud computing", 2017 International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), 2017
- [29] Ajinkya Yelne, Dipti Theng. "Stock Prediction and analysis Using Supervised Machine Learning Algorithms", 2021 International Conference on Computational Intelligence and Computing Applications (ICCICA), 2021
- [30] Bhanu Prakash Gopularam, C. B. Yogeesh. "Mechanism for on demand Tag-Based software testing in virtualized environments", 2012 Fourth International Conference on Advanced Computing (ICoAC), 2012
- [31] Sanket Salvi, Pramod Jain Sa, Sumeet Kumar S, Nithin D Jacob, Suhas M V, Shubham S, Harshitha T K. "PI-aaS: An Experimental Study of Computer Vision based Platooning as a Service", 2021 IEEE 2nd International Conference on Technology, Engineering, Management for Societal impact using Marketing, Entrepreneurship and Talent (TEMSMET), 2021
- [32] Jayashree, Shivaprakash T. "Optimal Value for Number of Clusters in a Dataset for Clustering Algorithm", *Artificial Intelligence and Sustainable Computing. Algorithms for Intelligent Systems*. Springer, Singapore. https://doi.org/10.1007/978-981-19-1653-3_48.