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

PREDICTING STOCK MARKET TRENDS USING MACHINE LEARNING AND DEEP LEARNING ALGORITHM

Nannam. Swetha
MTech Department of computer science
Andhra University College of Engineering, Visakhapatnam, A.P, India

ABSTRACT: **Financial market predictions are challenging in the best of times and especially when markets experience economic distress or rapid flux. This study aims to find better prediction models using AI and significant learning computations. It looks at four trade areas for testing appraisals: widened financials, oil, non-metallic minerals, and key metals from the Tehran stock exchange. This study investigates nine AI models (Decision Tree, Random Forest, adaptive boosting (Adaboost), eXtreme gradient boosting (XGBoost), support vector classifiers (SVC), Naïve Bayes, K-nearest neighbors (KNN), logistic regression, and artificial neural network (ANN) along with two significant learning procedures, recurrent neural network (RNN) and long short-term memory (LSTM). We look at ten specific markers from ten years of data and two distinct ways of assessing them. Each assumption model is surveyed by three estimations subject to the data. Our results show that for the perpetual data, RNN and LSTM are superior to other prediction models with noteworthy differentiation. Results further show that in the combined data evaluation, those significant learning techniques work exceedingly well.**

***KEYWORDS*: *Stock market, trends prediction, classification, machine learning, deep learning*.**

I. INTRODUCTION

Historical Background:

Gauging the movement of the stock market has always been a troublesome issue for subject matter experts. If the logical strategy is to buy stocks that are likely to increase in value and sell stocks that are going to lose value, then the challenge has been how best to make the predictions that guide those choices. Overall, there are two distinct methods for protections trade assumption. Fundamental assessment is one, and it relies upon an association's strategy and head information like imperfection ket position, expenses, and yearly advancement rates. The other method is the specific examination procedure, which centers around past stock expenses and characteristics. This assessment uses certain graphs and guides to anticipate future expenses [1], [2]. In the past, stock trades were assessed by financial professionals, but today, data scientists have begun using AI methods to deal with the introduction of assumption models and work on the precision of conjectures.
Significant learning helps create assumption models with better execution [3], [4]. Market prediction is a process riddled with hardships, and data analysts face certain issues when they try to construct a farsighted model. Multifaceted design and nonlinearity are two challenges contributing to the instability of monetary trade [5]. There are also unpredictable variables, like the public images of associations or political conditions of countries, which can negatively impact predictability. However, if the data obtained from stock characteristics are viably preprocessed and sensible computations are used, stock characteristics can be reliably predicted, and historical trends recorded.
In monetary trade assumption structures, AI and significant learning approaches can help investors and dealers make decisions. These procedures can see and learn plans even when dealing with colossal amounts of information. The estimations can be effectively self-learning and can predict worth well enough to create trading frameworks [6].

II. THEORY

Existing System:
Predicting markets and trade can be impacted by external elements like public end and political events. The goal of this assessment was to find whether public and political situations on a given day can impact trade among individual companies or the overall market. Inclination and situation features were used in an AI model to find the effect of public and political conditions on the assumption precision of computations for seven days in the future. We also considered interdependencies among associations and protections trades. For experimentation, credible trade data were downloaded from Yahoo! Cash, and public events

were culled from Twitter. Critical political events data of Pakistan were crawled from Wikipedia.
The rough substance data were then pre-arranged, and the inclination and situation features were created to make the last instructive assortments. Ten AI estimations were applied to the last educational assortments to predict future trades. The test outcomes showed that the assessment improves assumption accuracy of AI computations by 0–3%, and political situations impact the conjecture precision of estimations up to about 20%. Also, the supposition trademark was best on day seven, while the political situation quality was best on day five. SMO estimation was found to show the best display, while ASC and bagging give a dull appearance. The interdependency results showed that protections trades in a comparative industry have a medium positive relationship with each other.

Disadvantages of the Existing System

In the existing system, the framework for stock market prediction is riddled with difficulties, and information researchers typically encounter problems when they attempt to foster a prescient model. It is a lower performing system in with consistently unusual factors like the public picture of organizations or political circumstance of nations, which influence financial exchange patterns

Proposed System:

The proposed system centers around the execution of nine AI models (Decision Tree, Random Forest, Adaboost, XGBoost, SVC, Naïve Bayes, KNN, Logistic Regression and ANN) and two significant learning strategies (RNN and LSTM) to make improved trade predictions. Ten specific data markers were utilized by our models. The proposed study consolidates two particular procedures for inputs, steady data and combined data, to explore the effect of preprocessing. The former utilizes stock trading data (open, close, high, and inferior qualities) while the latter uses preprocessing steps to change persevering data to twofold. Each marker has a specific possibility of greater or lesser improvement reliant upon market intrinsic properties.
The display of the two models was differentiated for the two procedures and three gathering estimations, and the best tuning limit for each model (beside Naïve Bayes and Logistic Regression) was represented. All exploratory tests were done with ten years of chronicled data of four monetary trade get-togethers (oil, widened financials, crucial metals, and non-metallic minerals) that were vital for investors in the Tehran stock exchange. We acknowledge that this assessment is another investigation paper that joins various AI and significant learning methods to further improve stock market predictions.

Advantages of the Proposed System

In the proposed system, each of the algorithms can effectively solve stock prediction problems. The system is more effective due to the presence of XGBoost and SVC techniques.

III. LITERATURE REVIEW

In recent years, various methods have improved prediction of monetary trade designs. Hassan et al. [7] proposed the execution of a mixed model with Genetic Algorithms (GA), Artificial Neural Networks, and a Hidden Markov Model (HMM). They changed the step-by-step stock characteristics to free social occasions of expenses as commitments to HMM. Huang et al. [8] examined the consistency of financial examples using a support vector machine (SVM) model to evaluate the step-by-step NIKKEI 225 record. Their goal was to assess advantages and disadvantages of SVM, the linear discriminant method, and their results showed that SVM was the best classifier technique.
Sun et al. [9] proposed a new financial gauge estimation subject to the SVM outfit and a system for picking the SVM outfit's base classifiers by thinking about both assortment examination and individual assumption. Their results showed that for planning, SVM gathering was fundamentally more effective than individual SVM. Ou et al. [10] used ten data mining strategies to analyze data from the Hang record from the Hong Kong market. The methodologies included tree-based portrayal, K-nearest neighbor, Bayesian gathering, SVM, and neural association. Their results showed that the SVM outmaneuvered other insightful models. Liu et al. [11] utilized Legendre neural association to analyze data on the characteristics of investor positions and their decisions. They dissected a discretionary limit (time strength) in the estimate model. Araújo et al. [12] proposed the morphological position straight gauging approach to manage contrast in its results. Including a formative decision-making methodology and multi-layer perceptron networks, their method was able to suitably handle stock figure issues. Their conjecture results were affected not only by the depiction of the data but also depended upon the assumption strategy. Using prominent features and remembering them as data can effectively cultivate the precision of estimation models.
Another kind of assessment uses tree-based assembling strategies and significant learning estimations for predicting stock and monetary trade. Tsai et al. [13] used two different kinds of

company classifiers, pressing and vote strategies, as heterogeneous and homogeneous systems. They similarly consider macroeconomic features and money-related outcomes from the Taiwan monetary trade to review prediction models. Their results showed that with respect to the theory returns and conjecture precision, company classifiers were superior to single classifiers. Ballings et al. [14] dissected the introduction of Ada Boost, Random Forest, and bit fabricating plant versus single models including SVM, KNN, logistic regression, and ANN. Their results showed that Random Forest beat all the other models, and they were able to successfully predict the European association's expenses for a year in advance.
Basak et al. [15] used XGBoost and Random Forest to calculate the stock value based on past behavior, and their results showed that their method outperformed current ones. To dissect macroeconomic pointers to correctly predict monetary trade for one month ahead, Weng et al. [16] created four company models: boosting regressor, stowing regressor, neural association bunch regressor, and sporadic forest regressor. They also used a cross-variety strategy for long short-term memory (LSTM) to show that macroeconomic features are the best markers for protections trade. Progressing forward using significant learning computations, long et al. [17] investigated a significant neural association estimation using trade records and public market data to assess stock worth examples. Their results showed that their specific method of utilizing bidirectional LSTM could predict the market exceedingly well.
Rekha et al. [18] broke down recurrent (RNN) and convolutional neural network (CNN) estimations to differentiate between their accuracy and that of certified characteristics of monetary trade. Pang et al. [19] utilized

LSTM with a customized encoder and LSTM with an introduced layer to improve protections trade appraisals. They found that LSTM with an introduced layer predicted the Shanghai composite record with 57.2% precision. Kelotra and Pandey [20] used the significant convolutional LSTM estimation to efficiently learn monetary trade improvements. They employed a Rider-based, ruler-rich improvement strategy to achieve RMSE and MSE of 2.6923 and 7.2487, respectively

Baek and Kim [21] prescribed a deciding LSTM model and an over fitting expectation LSTM module to predict protections trade. They showed that using the over fitting neutralization module generated more accurate results. Chung and Shin [22] used a hybrid method for LSTM and GA to cultivate another protections trade estimate methodology. Their procedure beat the benchmark model, which centered around macroeconomic or specific features with late AI methods that may not have adequately considered appropriate preprocessing methodologies.

The Tehran market has some outstanding features that are unique to other country's monetary trades; for example, there is an overseeing esteem limitation that is 5% of the opening expense for each record in each trading day. This matter hampers the dissemination of market shocks and inconsistent market incitations, policy driven issues, etc. In any case, the effect of accessible head limits is huge, and the assumption undertaking of future advancements is not straightforward [23]. This study used monetary trade get-togethers (highly important for sellers) to investigate the task of expecting future examples.

Despite groundbreaking progress in the Tehran monetary trade in the new decade, there have not been enough studies AI estimations of its stock values and behaviors. However, Nabipour et al. [23] used tree-based models and significant learning estimations to evaluate future stock expenses as a backslide issue over 30 days. They showed that LSTM (as the unmatched model) could adequately predict values (from the Tehran Stock Exchange) with only minor uncertainty.

IV. METHODOLOGY:

Methodology of designing the software system proposed in this study, from practical and economic feasibility through approval.

Feasibility Study

In this step, the project's feasibility was assessed, and a business proposal was presented, complete with a broad project plan and cost estimates. The feasibility assessment assures that the proposed system will not be a financial burden to the company. A basic grasp of the system's primary requirements was required for feasibility study. Three key considerations involved in the feasibility analysis were economic feasibility, technical feasibility, and social feasibility.

Economic Feasibility

This study was carried out to check the economic impact that the system will have on the organization. The amount of funds that the organization can pour into the research and development of the system is limited, so expenditures must be justified. The system developed here was well within the budget, and this was achieved because most of the technologies used were free. Only the customised items have to be bought.

Technical Feasibility

This study was carried out to check technical feasibility, i.e., the technical requirements of the system. Any system developed must not place a high demand on available technical resources or on clients. The system developed here will present only modest demand on our technical resources, as only minimal or null changes are required for its implementation.

Social Feasibility

The aspect of study checked the level of acceptance of the system by users. This includes the process of training users to use the system efficiently. Users must not feel threatened by the system and instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make them familiar with it. Their level of confidence must be raised so that they can offer constructive criticism, which is welcomed, as they are the final users of the system.

Request Clarification

After approving the project guide and the initial process, the organization requested more detail about the precise requirements of developing the system. This project is basically meant for users within the organization whose systems are interconnected by the Local Area Network (LAN). In response to the need for an efficient and fast system, this study clarified its request by introducing the creation of a portal.

Request Approval

Not all proposed projects are desirable or feasible. Some organizations can only approve a small percentage of the many proposals they receive. However, projects that are both feasible and desirable should be scheduled. After a project request is approved, its cost, priority, completion time, and personnel requirement are estimated and used to determine its scheduling.

V. INPUT DESIGN

Input/data design deals withs the associations between the information structure and the customer. It incorporates decision making based on data availability, and those methods are essential in making trade data usable. These data can be cultivated by surveying the PC to examine data from a created or printed file, or it can occur by having people enter the data directly into the structure.

The arrangement of data revolves around controlling the proportion of data required, controlling the errors, avoiding delay, and keeping the collaboration fundamental.

The data is arranged in such a way that it outfits security and convenience.

Our design considered the following:

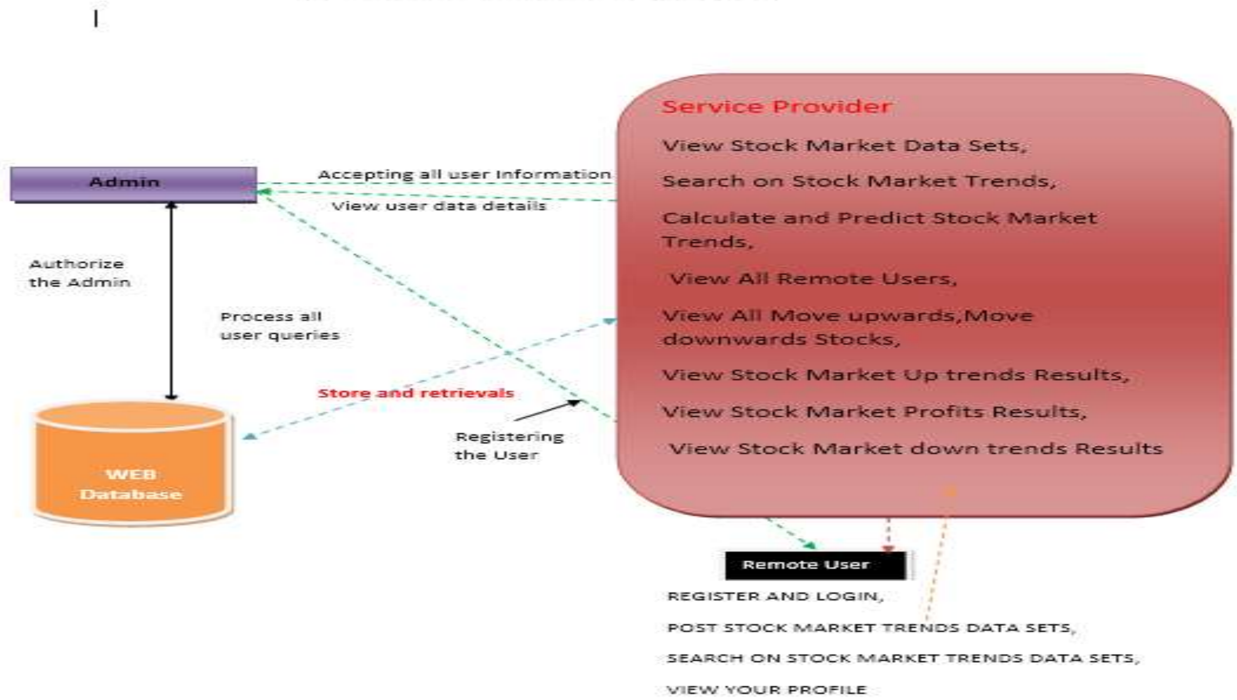
Which data should be offered as input, how should the data be formatted or coded, the dialogue to aid the operational staff in providing input, and techniques for preparing input validations and steps to take when errors occur were all factors in our design.

Objectives

The process of transforming a user-oriented description of an input into a computer-based system is known as input architectural design. This design is important to avoid errors in the data input process and show management the correct direction for getting accurate information from the computerized system.

It is accomplished by designing user-friendly data entry panels that can manage a vast amount of data. The purpose of input design is to make data entry easier and error-free. The data entry screen is designed in such a way that all data manipulation can be performed. It also provides record viewing facilities.
When the data is entered it is checked for its validity. Screens can be used to enter information. Appropriate messages are provided as needed so that the user does not get lost in a maze of data and fields.

Architecture Diagram



VI. OUTPUT DESIGN

A good output is one that satisfies the end user's needs and delivers the information clearly. The outputs of any system transmit the outcomes of processing to users and other systems. Output design determines how the information is to be displaced for immediate need and the hard copy output. It is the user's primary and most direct source of information. The system's relationship with the user is improved via efficient and intelligent output design.

Computer output should be built in an orderly, well-thought-out manner; the proper output must be developed while ensuring that each output part is designed in such a way that people will find the system easy to use. When programmers create computer output, they should consider the following factors.

- identify the specific output that is needed to meet the requirements:
- select the method for presenting information, and
- create documents, reports, or other formats that contain information produced by the system.

An information system's output form should achieve one or more of the following goals:
provide information regarding previous activities, current situations, or future projections provide information regarding previous activities, current situations, or future projections

VII. MODULES

Service Provider

In this module, the Service Provider has to login by using valid username and password. Figure 4 shows a flowchart for actions that can be performed after login: *View Stock Market Data Sets, Search on Stock Market Trends, Calculate and Predict Stock Market Trends, View All Remote Users, View All Move Upwards/Move downwards Stocks, View Stock Market Up Trend Results, View Stock Market Profit Results, and View Stock Market Down Trends Results.*

View and Authorize Users

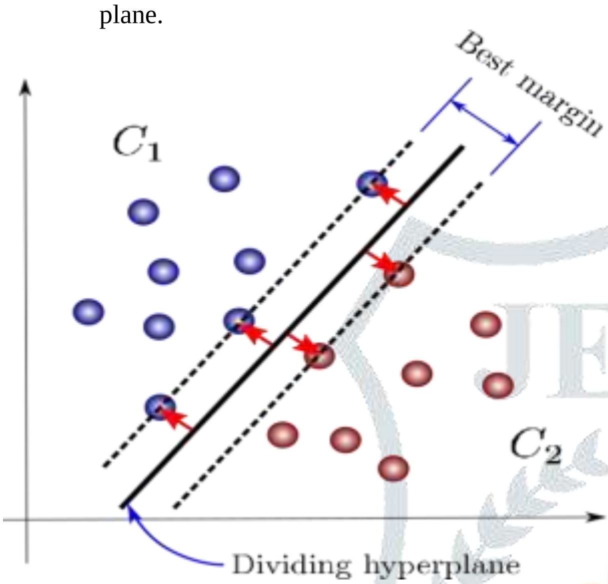
In this module, the admin can view the list of registered users and details such as username, email, and address.
Together, the remote user, service provider, and server form front end development. Figure 5 details the formation and features of all users and their relationships in the system. It also shows the flow from web server register and login, post stock trends to remote server, and from remote server to service provider.

Remote User
There are a total of n users in this module. Users should register before doing any operations. Once a user registers,

their details will be stored to the database. After successful registration, they must login with an authorized username and password. on successful login, the user can perform operations like
post stock market trends data sets, search on stock market trends data sets, and view your profile.

SVM – Support Vector Machine

- SVM is a frontier which best segregates two classes via hyper plane
- SVM solves classification and regression problems
- SVM classifies a set of points through hyper plane.



- Significance of margin: for creating a generalized model to get better accuracy.
- This hyper plane is giving a cushion in dividing the points as positive or negative in a better way
- But the hyper plane is not the end apart from that plane it creates two margin hyper planes parallel to hyper plane , separating a distance.
- Creating two hyper planes must make sure that they pass through one of the nearest data points.
- The distance between the two created hyper planes is called margin.

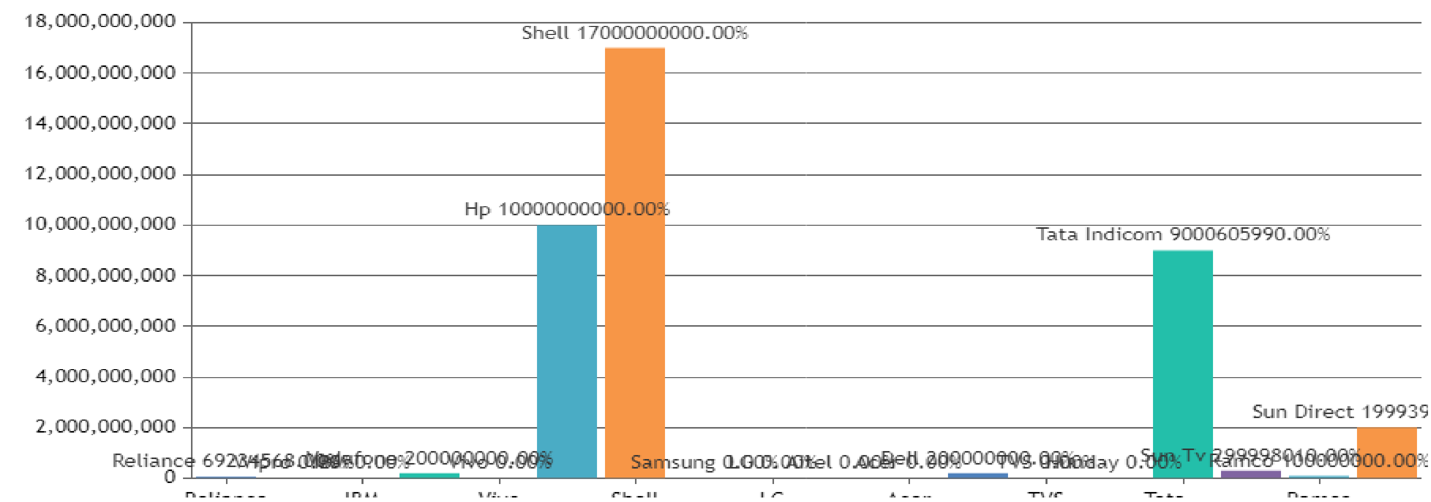
$$f(x) = \mathrm{sgn}(\sum_{i=1}^{n} \alpha_i y_i \cdot K(x, x_i) + b)$$

and becomes the main difference between logistic and SVM.

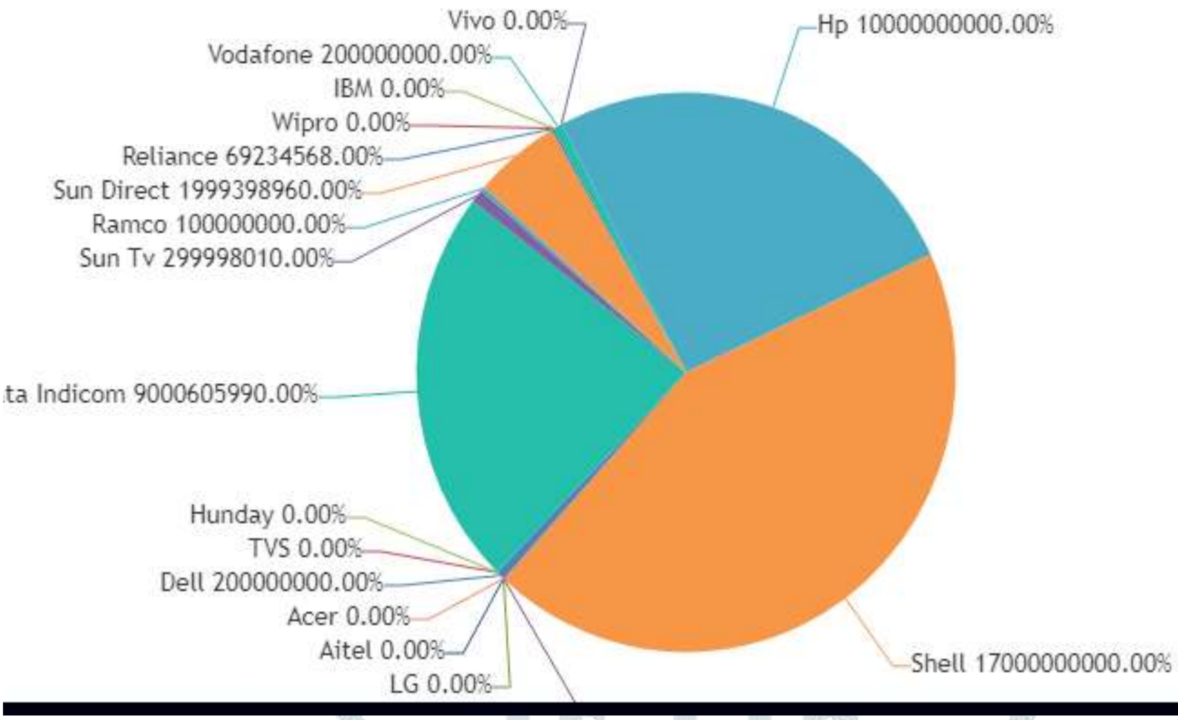
- whenever hyper planes are created, we should choose the hyper planes where the marginal distance is maximum to get a more generalized model for a new data set.
- Support vectors are the points that pass through the marginal hyper plane.

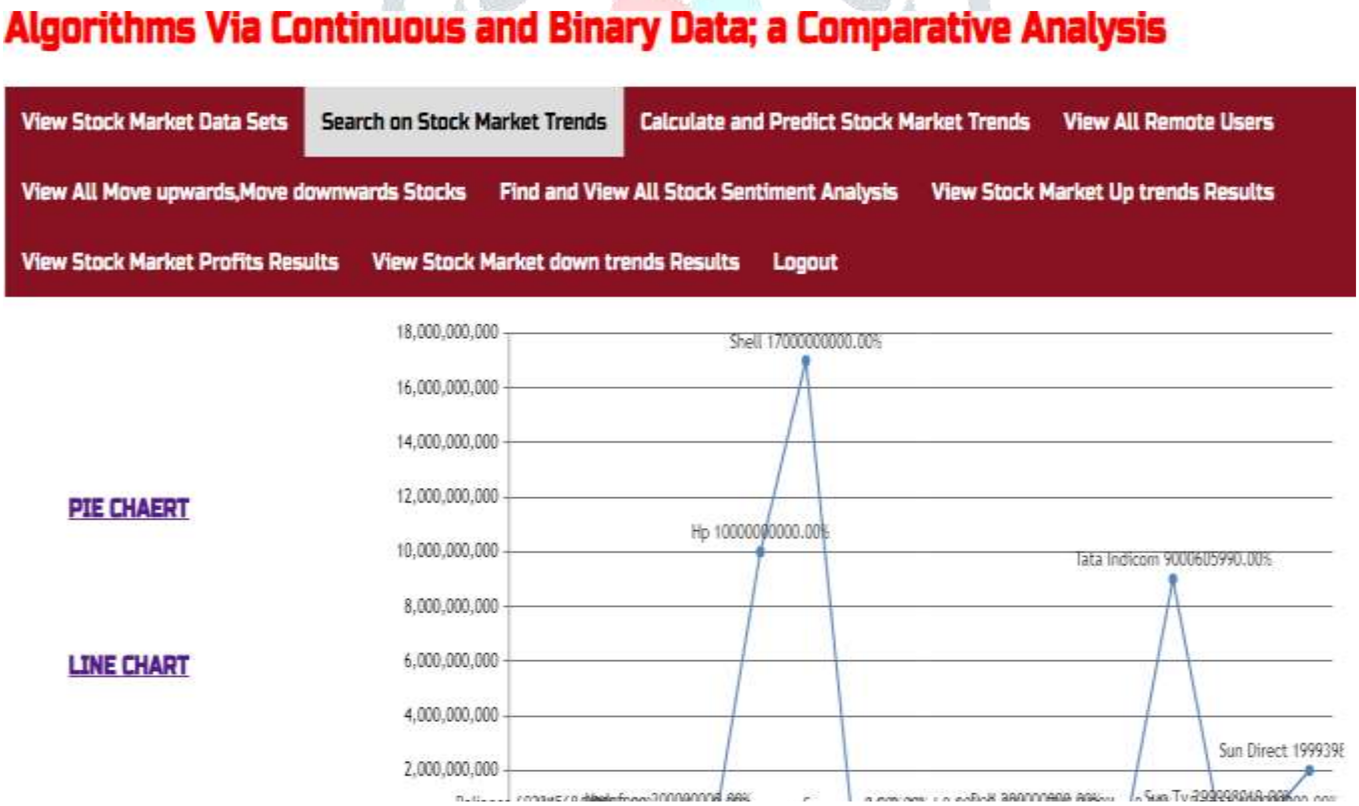
I. EXPERIMENTAL RESULTS:

PROFIT ANALYSI



UPWARD TRENDS PIE CHART





II. CONCLUSION

The inspiration driving this study was to improve stock market predictions using AI and significant learning computations. We looked at four trade areas for testing appraisals: widened financials, oil, non-metallic minerals, and key metals from the Tehran stock exchange. This study investigated nine AI models (Decision Tree, Random Forest, adaptive boosting (Adaboost), eXtreme gradient boosting (XGBoost), support vector classifiers (SVC), Naïve Bayes, K-nearest neighbors (KNN), logistic regression, and artificial neural network (ANN) along with two significant learning procedures, recurrent neural network (RNN) and long short-term memory (LSTM). We looked at ten specific markers from ten years of data and two distinct ways of assessing them. Results of our tests showed that there was a significant improvement in the models' output when they used twofold data.

References

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

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

[3] H. Maqsood, I. Mehmood, M. Maqsood, M. Yasir, S. Afzal, F. Aadil, M. M. Selim, and K. Muhammad, ``A local and global event sentiment based efcient stock exchange forecasting using deep learning," Int. J. Inf. Manage., vol. 50, pp. 432451, Feb. 2020.

[4] W. Long, Z. Lu, and L. Cui, ``Deep learning-based feature engineering for stock price movement prediction," Knowl.-Based Syst., vol. 164, pp. 163173, Jan. 2019.

[5] J. B. Duarte Duarte, L. H. Talero Sarmiento, and K. J. Sierra Juárez. “Evaluation of the effect of investor psychology on an articial stock market through its degree of efciency," Contaduría y Administración, vol. 62, no. 4, pp. 13611376, Oct. 2017.

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

[7] M. R. Hassan, B. Nath, and M. Kirley, ``A fusion model of HMM, ANN and GA for stock market forecasting," Expert Syst. Appl., vol. 33, no. 1, pp. 171180, Jul. 2007.

[8] W. Huang, Y. Nakamori, and S.-Y. Wang, ``Forecasting stock market movement direction with support vector machine," Comput. Oper. Res., vol. 32, no. 10, pp. 25132522, Oct. 2005.

[9] J. Sun and H. Li, ``Financial distress prediction using support vector machines: Ensemble vs. Individual," Appl. Soft Comput., vol. 12, no. 8, pp. 22542265, Aug. 2012.

[10] P. Ou and H. Wang, ``Prediction of stock market index movement by ten data mining techniques," Modern Appl. Sci., vol. 3, no. 12, pp. 2842, Nov. 2009.

[11] F. Liu and J.Wang, ``Fluctuation prediction of stock market index by legendre neural network with random time strength function," Neurocomputing, vol. 83, pp. 1221, Apr. 2012.

[12] C.-F. Tsai, Y.-C. Lin, D. C. Yen, and Y.-M. Chen, ``Predicting stock returns by classier ensembles," Appl. Soft Comput., vol. 11, no. 2, pp. 24522459, Mar. 2011.

[13] R. D. A. Araäjo and T. A. E. Ferreira, ``A Morphological-Rank-Linear evolutionary method for stock market prediction," Inf. Sci., vol. 237, pp. 317, Jul. 2013.

[14] M. Ballings, D. Van den Poel, N. Hespeels, and R. Gryp, ``Evaluating multiple classiers for stock price direction prediction," Expert Syst. Appl., vol. 42, no. 20, pp. 70467056, Nov. 2015.

[15] S. Basak, S. Kar, S. Saha, L. Khaidem, and S. R. Dey, ``Predicting the direction of stock market prices using tree-based classiers," North Amer. J. Econ. Finance, vol. 47, pp. 552567, Jan. 2019.

[16] B. Weng, W. Martinez, Y.-T. Tsai, C. Li, L. Lu, J. R. Barth, and F. M. Megahed. “Macroeconomic indicators alone can predict the monthly closing price of major U.S. indices: Insights from articial intelligence,time-series analysis and hybrid models," Appl. Soft Comput., vol. 71, pp. 685697, Oct. 2018.

[17] J. Long, Z. Chen, W. He, T. Wu, and J. Ren, ``An integrated framework of deep learning and knowledge graph for prediction of stock price tren: An application in chinese stock exchange market," Appl. Soft Comput., vol91, Jun. 2020, Art. no. 106205.

[18] G. Rekha, D. Bhanu Sravanthi, S. Ramasubbareddy, and K. Govinda, ``Prediction of stock market using neural network strategies," J. Comput. Theor. Nanoscience, vol. 16, no. 5, pp. 23332336, May 2019.

[19] X. Pang, Y. Zhou, P. Wang, W. Lin, and V. Chang, ``An innovativneural network approach for stock market prediction," J. Supercomput., vol. 76,no. 3, pp. 20982118, Mar. 2020.

[20] Kelotra, A. and P. Pandey, ``Stock market prediction using optimized deep convLSTM model," Big Data, vol. 8, no. 1, pp. 524, 2020.

[21] Y. Baek and H. Y. Kim, ``ModAugNet: A new forecasting framework for stock market index value with an overtting prevention LSTM module and a prediction LSTM module," Expert Syst. Appl., vol. 113, pp. 457480, Dec. 2018.

[22] H. Chung and K.-S. Shin, ``Genetic algorithm-optimized long short-term memory network for stock market prediction," Sustainability, vol. 10, no. 10, p. 3765, 2018.

[23] M. Nabipour, P. Nayyeri, H. Jabani, A. Mosavi, and E. Salwana, ``Deep learning for Stock Market Prediction," Entropy, vol. 22, no. 8, p. 840, Aug. 2020.

[24] Y. Kara, M. Acar Boyacioglu, and Ö. K. Baykan, ``Predicting direction of stock price index movement using articial neural networks and support vector machines: The sample of the Istanbul stock exchange," Expert Syst. Appl., vol. 38, no. 5, pp. 53115319, May 2011.

[25] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, ``Predicting stock market index using fusion of machine learning tech.”