RESEARCH ARTICLE OPEN ACCESS

Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms via Continuous and Binary Data

Mrs R Jhansi Rani MCA [1], C Nithin [2]

[1] Asst. Professor, Department Computer Applications

[2] Student, Department of Computer Applications

[1], [2] Chadalawada Ramanamma Engineering College (Autonomous)

ABSTRACT

The nature of stock market movement has always been ambiguous for investors because of various influential factors. This study aims to significantly reduce the risk of trend prediction with machine learning and deep learning algorithms. Four stock market groups, namely diversified financials, petroleum, non-metallic minerals and basic metals from Tehran stock exchange, are chosen for experimental evaluations. This study compares nine machine learning models (Decision Tree, Random Forest, Adaptive Boosting (Adaboost), eXtreme Gradient Boosting (XGBoost), Support Vector Classifier (SVC), Naïve Bayes, K-Nearest Neighbors (KNN), Logistic Regression and Artificial Neural Network (ANN)) and two powerful deep learning methods (Recurrent Neural Network (RNN) and Long short-term memory (LSTM). Ten technical indicators from ten years of historical data are our input values, and two ways are supposed for employing them. Firstly, calculating the indicators by stock trading values as continuous data, and secondly converting indicators to binary data before using. Each prediction model is evaluated by three metrics based on the input ways. The evaluation results indicate that for the continuous data, RNN and LSTM outperform other prediction models with a considerable difference. Also, results show that in the binary data evaluation, those deep learning methods are the best; however, the difference becomes less because of the noticeable improvement of models' performance in the second way.

Keywords: - The nature of stock market movement has always been ambiguous for investors because of various influential factors.

I. INTRODUCTION

Stock Market prediction and analysis is the act of trying to determine the future value of a company stock or other financial instrument traded on an exchange. Stock Market is the important part of the economy of the country and plays a vital role in the growth of the industry and commerce of the country that eventually affects the economy of the country. Both investors and industry are involved in the stock market and want to know whether some stock will rise or fall over a certain period of time. The stock market is the primary source for any company to raise funds for business expansions. It is based on the concept of demand and supply. If the demand for a company's stock is higher, then the company share price increases and if the demand for company's stock is low then the company share price decreases. Another motivation for research in this field is that it possesses many theoretical and experimental challenges. The most important of these is the Efficient Market Hypothesis (EMH), the hypothesis says that in an efficient market, stock market prices fully reflect available information about the market and its constituents and thus any opportunity of earning excess profit ceases to exist. One example of a big exchange is the New York Stock Exchange. The task of stock prediction has always been a challenging problem for statistics experts and finance. The main reason behind this prediction is buying stocks that are likely to increase in price and then selling stocks that are probably to fall. Generally, there are two ways for stock market prediction. Fundamental analysis is one of them and relies on a company's technique and fundamental information like market position, expenses and annual growth rates. The second one is the technical analysis method, which concentrates on previous stock prices and values. This analysis uses historical charts and patterns to predict future prices. Stock markets were normally predicted by financial experts in the past time. However, data scientists have started solving prediction problems with the progress of

learning techniques. Also, computer scientists have begun using machine learning methods to improve the performance of prediction models and enhance the accuracy of predictions. Employing deep learning was the next phase in improving prediction models with better performance. Stock market prediction is full of challenges, and data scientists usually confront some problems when they try to develop a predictive model. With the evolution of computer science, various new disciplines came into existence which provided better prediction models. One such discipline of computer science is Machine Learning. Over the years, machine learning has played a vital role in predictions. Predictions like workload management in cloud [1] [41-42], heart disease prediction [37], house rent price prediction [38], stock market price prediction [18] etc. were now possible with various techniques of machine learning. It helped in building new and improvised prediction models, which gave better results with lesser complexity. In context with stock market prediction, many researchers have been able to devise models for stock market prediction which uses various techniques of machine learning such as SVM (Support Vector Machine), Linear Regression, Random Forest, K-Nearest Neighbour (KNN), ANN, deep learning, LSTM, MLP, Boosted Decision Tree, Evolutionary algorithms and many more hybrid techniques which would be further discussed in this paper. This paper also discusses the challenges that are faced or can be faced by researchers while devising prediction models

II. LITERATURE SURVEY

A local and global event sentiment based efficient stock exchange forecasting using deep learning

Stock exchange forecasting is an important aspect of business investment plans. The customers prefer to invest in stocks rather than traditional investments due to high profitability. The high profit is often linked with high risk due to the nonlinear nature of data and complex economic rules. The stock markets are often volatile and change abruptly due to the economic conditions, political situation and major events for the country. Therefore, to investigate the effect of some major events more specifically global and local events for different top stock companies (country-wise) remains an open research area. In this study, consider four countries- US, Hong Kong, Turkey, and Pakistan from developed, emerging and underdeveloped economies' list. We have explored the effect of different major events occurred during 2012–2016 on stock markets

Predicting stock returns by classifier ensembles

The problem of predicting stock returns has been an important issue for many years. Advancement in computer technology has allowed many recent studies to utilize machine learning techniques such as neural networks and decision trees to predict stock returns. In the area of machine learning, classifier ensembles (i.e. combining multiple classifiers) have proven to be a method superior to single classifiers. In order to build a better model for predicting stock returns effectively and efficiently, this study aims at investigating the prediction performance that utilizes the classifier ensembles method to analyze stock returns. In particular, the hybrid methods of majority voting and bagging are considered. Moreover, performance using two types of classifier ensembles is compared with those using single baseline classifiers (i.e. neural networks, decision trees, and logistic regression). These two types of ensembles are 'homogeneous' classifier ensembles (e.g. an ensemble of neural networks) and 'heterogeneous' classifier ensembles (e.g. an ensemble of neural networks, decision trees and logistic regression). Average prediction accuracy, Type I and II errors, and return on investment of these models are also examined

A Morphological-Rank-Linear evolutionary method for stock market prediction

This work presents an evolutionary morphological-rank-linear approach in order to overcome the random walk dilemma for financial time series forecasting. The proposed Evolutionary Morphological-Rank-Linear Forecasting (EMRLF) method consists of an intelligent hybrid model composed of a Morphological-Rank-Linear (MRL) filter combined with a Modified Genetic Algorithm (MGA), which performs an evolutionary search for the minimum number of relevant time lags capable of a fine tuned characterization of the time series, as well as for the initial (sub-optimal) parameters of the MRL filter. Then, each individual of the MGA

population is improved using the Least Mean Squares (LMS) algorithm to further adjust the parameters of the MRL filter, supplied by the MGA. After built the prediction model, the proposed method performs a behavioral statistical test with a phase fix procedure to adjust time phase distortions that can appear in the modeling of financial time series.

III. EXISTING STOCK MARKET PREDICTION TECHNIQUES

Many models of prediction have been proposed till date to forecast the stock prices and stock market trends. Some of the machine learning techniques have been discussed in this paper. All the techniques have been classified into various subcategories like classification techniques, regression techniques, ensemble algorithms, evolutionary techniques, deep learning, hybrid models and some other additional techniques. A. Classification Techniques 1) Support Vector Machine (SVM): One of machine learning algorithms that possesses the desired features such as the decision function, usage of kernel methods and also the sparsity of the solution is known as the Support Vector Machine (SVM) technique. Random Forest Classifier Random forest classifier is a type of ensemble classifier and also a supervised algorithm. It basically creates a set of decision trees, that yields some result. The basic approach of random class classifier is to take the decision aggregate of random subset decision trees and yield a final class or result based on the votes of the random subset of decision trees. B. Random Forest Algorithm Random forest algorithm is being used for the stock market prediction. Since it has been termed as one of the easiest to use and flexible machine learning algorithms, it gives good accuracy in the prediction. This is usually used in the classification tasks. Because of the high volatility in the stock market, the task of predicting is quite challenging. In stock market prediction we are using a random forest classifier which has the same hyperparameters as the decision tree. The decision tool has a model similar to that of a tree. It takes the decision based on possible consequences, which includes variables like event outcome, resource cost, and utility. The random forest algorithm represents an algorithm where it randomly selects different observations and features to build several decision trees and then takes the aggregate of the several decision trees outcomes. The data is split into partitions based on the questions on a label or an attribute. The data set we used was from the previous year's stock markets collected from the public database available online, 80 % of data was used to train the machine and the rest 20 % to test the data. The basic approach of the supervised learning model is to learn the patterns and relationships in the data from the training set and then reproduce them for the test data.

Proposed Model

IV. PROPOSED SYSTEM

In the proposed system, the system concentrates on comparing prediction performance of nine machine learning models (Decision Tree, Random Forest, Adaboost, XGBoost, SVC, Naïve Bayes, KNN, Logistic Regression and ANN) and two deep learning methods (RNN and LSTM) to predict stock market movement.

Ten technical indicators are utilized as inputs to our models. The proposed study includes two different approaches for inputs, continuous data and binary data, to investigate the effect of preprocessing; the former uses stock trading data (open, close, high and low values) while the latter employs preprocessing step to convert continuous data to binary one. Each technical indicator has its specific possibility of up or down movement based on market inherent properties. The performance of the mentioned models is compared for the both approaches with three classification metrics, and the best tuning parameter for each model (except Naïve Bayes and Logistic Regression) is reported. All experimental tests are done with ten years of historical data of four stock market groups (petroleum, diversified financials, basic metals and non-metallic minerals), that are totally crucial for investors, from Tehran stock exchange. We believe that this study is a new research paper that incorporates multiple machine learning and deep learning methods to improve the prediction task of stock groups' trend and movement.

Advantages

- In the proposed system, each of the algorithms can effectively solve stock prediction problems.

- To The system is more effective due to presence of eXtreme Gradient Boosting (XGBoost), Support Vector Classier (SVC) techniques.

V. IMPLEMENTATION

There are Two modules involved in getting the information retrieval system up and running they are as follows

- Service provider
- Remote User

Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Upload & 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 trends Results, View Stock Market Profits Results, View Stock Market down trends Results,.

Remote User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like post stock market trends data sets, search on stock market trends data sets, and view your profile.

VI. RESULT

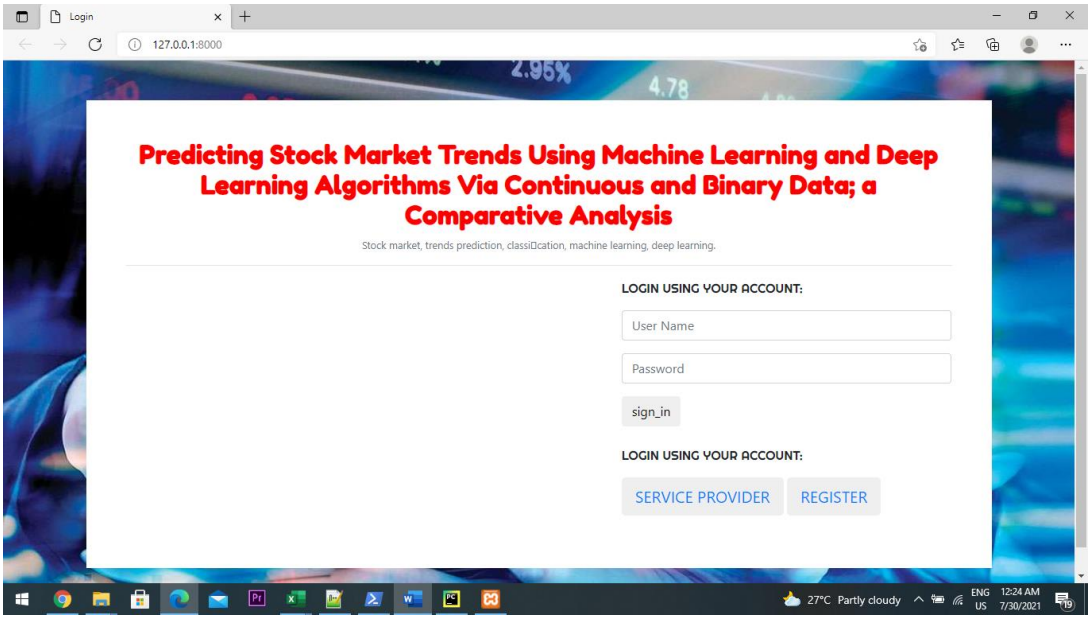


Fig 1. Home Page

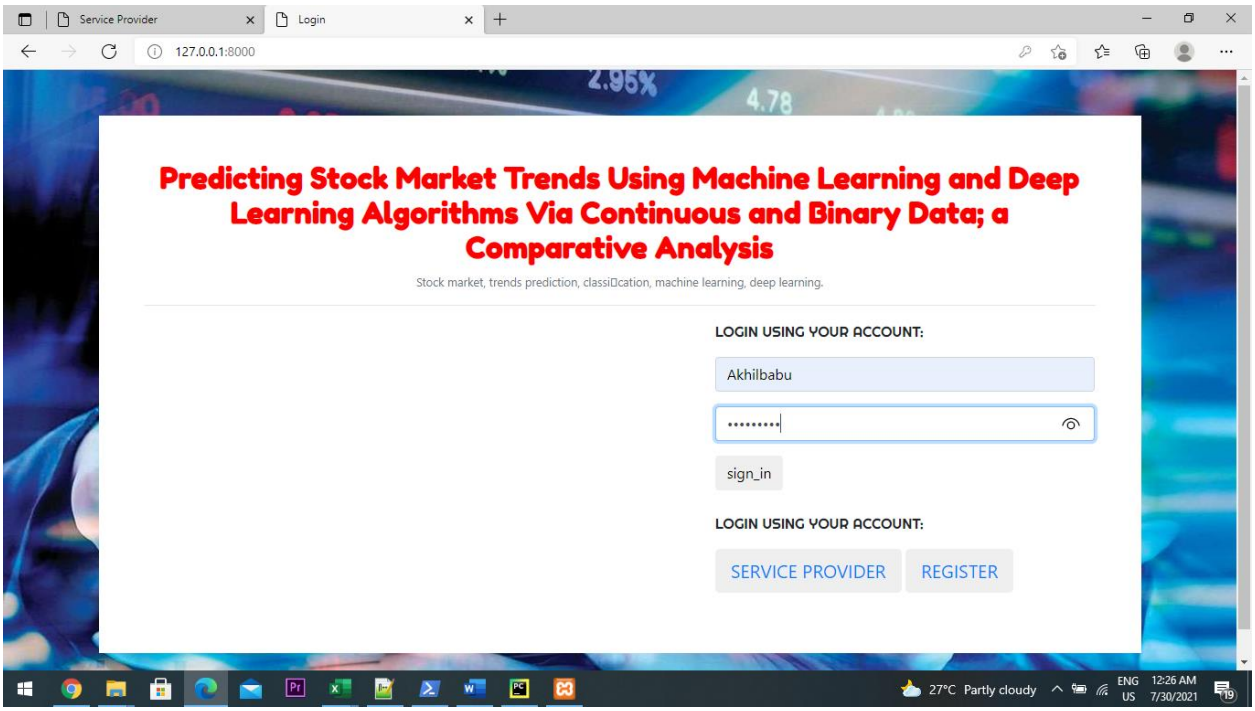


Fig 2. Remote User Login

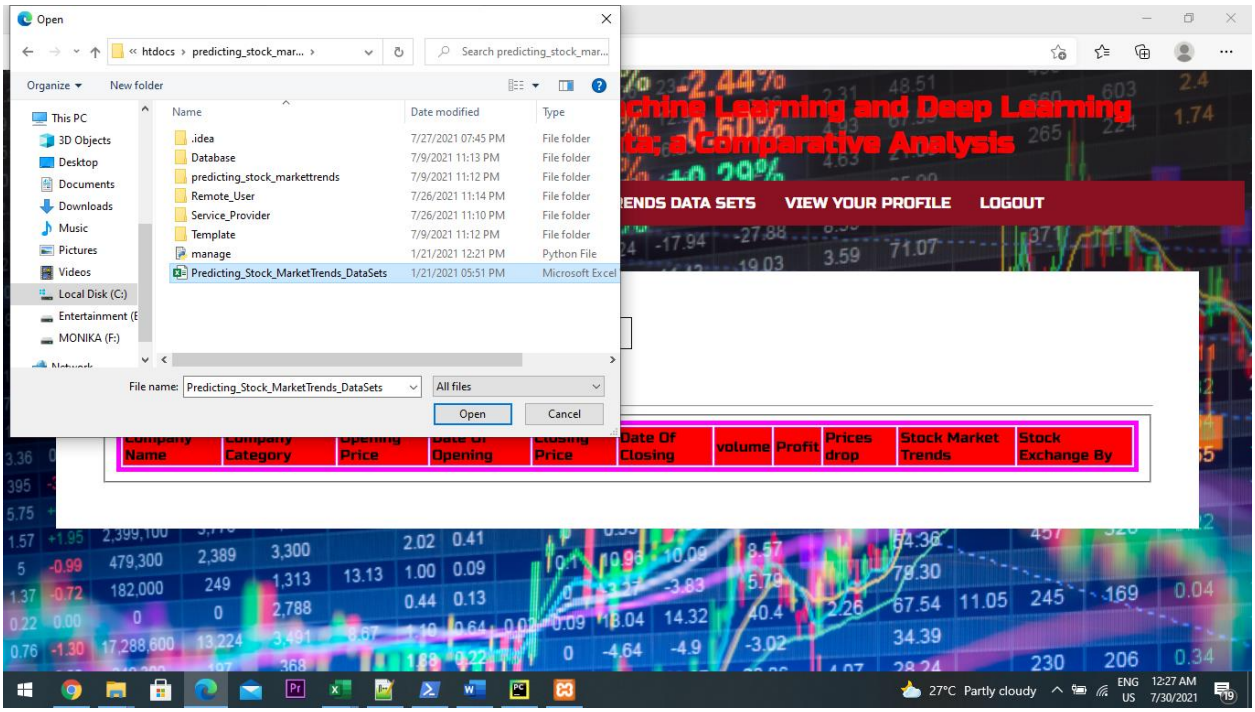
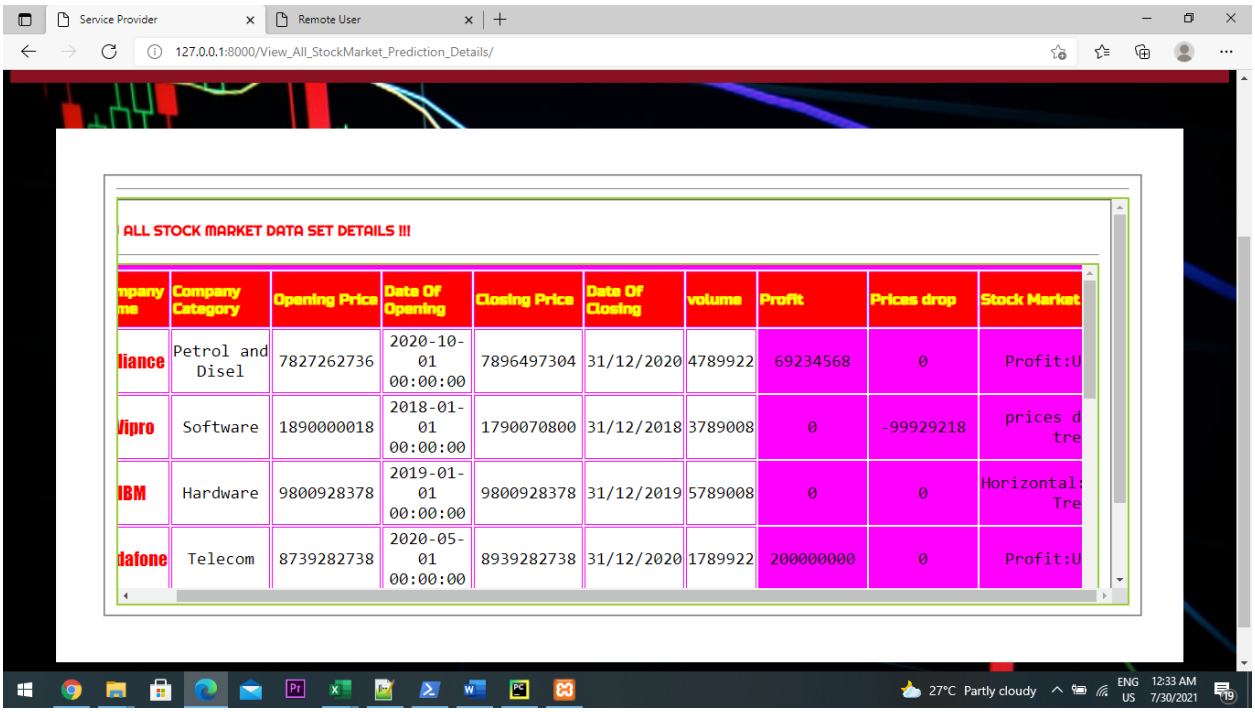


Fig 3. Browsing For Stock market Trends Dataset

Company Name	Company Category	Opening Price	Date Of Opening	Closing Price	Date Of Closing	volume	Profit	Prices drop	Stock Market Trends	Stock Exchange By
Reliance	Petrol and Disel	7827262736	2020-10-01 00:00:00	7896497304	31/12/2020	4789922	No	No	No	NSE
Wipro	Software	1890000018	2018-01-01 00:00:00	1790070800	31/12/2018	3789008	No	No	No	NSE
IBM	Hardware	9800928378	2019-01-01 00:00:00	9800928378	31/12/2019	5789008	No	No	No	NSE
Vodafone	Telecom	8739282738	2020-05-01 00:00:00	8939282738	31/12/2020	1789922	No	No	No	NSE
Vivo	Mobiles	19872635632	2018-11-01 00:00:00	19872635632	31/12/2018	2789008	No	No	No	NSE
Hp	Software	2800928378	2019-05-01 00:00:00	12800928378	31/12/2019	5789008	No	No	No	NSE
Shell	Petrol and Disel	1928374837	2020-10-01 00:00:00	18928374837	31/12/2020	3789922	No	No	No	NSE
Samsung	Mobiles	2890000018	2018-07-01 00:00:00	2790070800	31/12/2018	4789008	No	No	No	NSE
LG	Home Appliances	8273835617	2019-07-01 00:00:00	7273835617	31/12/2019	5789008	No	No	No	NSE
Aitel	Telecom	7625372837	2020-08-01 00:00:00	7625372837	31/12/2020	4780000	No	No	No	NSE
Acer	Hardware	8920283746	2018-12-01 00:00:00	6920283746	31/12/2018	3789008	No	No	No	NSE
Dell	Software	6718272363	14/1/2019	6918272363	31/12/2019	5789008	No	No	No	NSE
TVS	Motor	4892827383	2020-09-01	3992827383	31/12/2020	4789922	No	No	No	NSE

Fig 4. Stock Market Dataset Details



npany me	Company Category	Opening Price	Date Of Opening	Closing Price	Date Of Closing	volume	Profit	Prices drop	Stock Market
liance	Petrol and Disel	7827262736	2020-10-01 00:00:00	7896497304	31/12/2020	4789922	69234568	0	Profit:U
Vipro	Software	1890000018	2018-01-01 00:00:00	1790070800	31/12/2018	3789008	0	-99929218	prices d tre
IBM	Hardware	9800928378	2019-01-01 00:00:00	9800928378	31/12/2019	5789008	0	0	Horizontal: Tre
dafone	Telecom	8739282738	2020-05-01 00:00:00	8939282738	31/12/2020	1789922	200000000	0	Profit:U

Fig 5. Predict Stock Market Trends

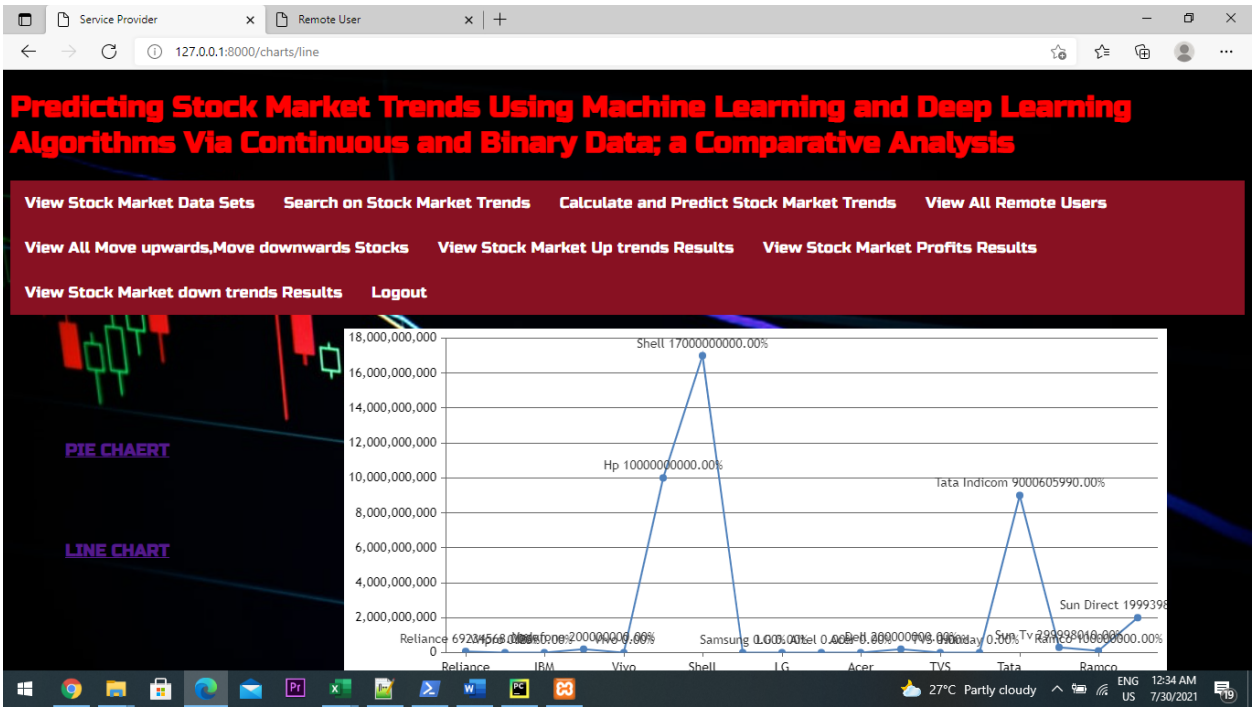


Fig 6 Stock Market Uptrends In Line Chart

VII. CONCLUSIONS

The purpose of this study was the prediction task of stock market movement by machine learning and deep learning algorithms. Four stock market groups, namely diversified financials, petroleum, non-metallic minerals and basic metals, from Tehran stock exchange were chosen, and the dataset was based on ten years of historical records with ten technical features. Also, nine machine learning models (Decision Tree, Random Forest, Adaboost, XGBoost, SVC, Naïve Bayes, KNN, Logistic Regression and ANN) and two deep learning methods (RNN and LSTM) were employed as predictors. We supposed two approaches for input values to models, continuous data and binary data, and we employed three classification metrics for evaluations. Our experimental works showed that there was a significant improvement in the performance of models when they used binary data instead of continuous one. Indeed, deep learning algorithms (RNN and LSTM) were our superior models in both approaches.

REFERENCE

[1] Murphy, John J. Technical analysis of the financial markets: A comprehensive guide to trading methods and applications. Penguin, 1999

[2] Turner, Toni. A Beginner's Guide to Day Trading Online 2nd Edition. Simon and Schuster, 2007.

[3] Chen, Yingjun, and Yongtao Hao. "A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction." Expert Systems with Applications 80 (2017): 340-355.

[4] Basak, Suryoday, et al. "Predicting the direction of stock market prices using tree-based classifiers." The North American Journal of Economics and Finance 47 (2019): 552-567

[5] Araújo, Ricardo De A., and Tiago AE Ferreira. "A morphological-rank-linear evolutionary method for stock market prediction." Information Sciences 237 (2013): 3-17.

[6] Weng, Bin, et al. "Macroeconomic indicators alone can predict the monthly closing price of major US indices: Insights from artificial intelligence, time-series analysis and hybrid models." Applied Soft Computing 71 (2018):685- 697.

[7] Baek, Yujin, and Ha Young Kim. "ModAugNet: A new forecasting framework for stock market index value with an overfitting prevention LSTM module

and a prediction LSTM module." Expert Systems with Applications 113 (2018): 457-480.
[8] Patel, Jigar, et al. "Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques." Expert systems with applications 42.1 (2015): 259-268
[9] Majhi, Ritanjali, et al. "Efficient prediction of stock market indices using adaptive bacterial foraging optimization (ABFO) and BFO based techniques." Expert Systems with Applications 36.6 (2009): 10097-10104.
[10]Sun, Jie, and Hui Li. "Financial distress prediction using support vector machines: Ensemble vs. individual." Applied Soft Computing 12.8 (2012): 2254- 2265. [11]Patel, Jigar, et al. "Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques." Expert systems with applications 42.1 (2015): 259-268