International Journal of Future Generation Communication and Networking
Vol. 13, No. 3s, (2020), pp. 1486–1496

Determining Stock Market Prediction Using Opinion Mining

Miss Asmita Kadam, Miss Kajal Warghude, Miss Pranjali Dudhal, Miss Taniya Nimbalkar

Department of Computer Engineering, Savitribai Phule Pune University

asmitakadamm5@gmail.com
kajalwarghde@gmail.com
dudhalpranjali15@gmail.com
nimbalkar.taniya1998@gmail.com

Abstract

The field of stocks or securities exchange has faced many transformations over the years. While the only thing constant in the stock market domain is its volatility, it still remains the most powerful investment strategy if you want to make money even when you sleep. In this paper, we propose an online website , that will make predictions on the performance of stocks in the future and help investors reach the most accurate decisions about investments. For this, we use a machine learning model, the base of which is a linear regression algorithm. Recent studies have suggested using a support vector machine (SVM) to make the predictions, which lead to a look-ahead bias, which was then tackled by coupling it with a rolling window approach to improve its accuracy. We, on the other hand, use linear regression to make predictions. The results of our experiment show that our method is more accurate as compared to SVM. Furthermore, any new investor who needs to brush up his knowledge about stocks will be redirected to articles or teaching videos about what the stock market is.

Keywords— *Stock market prediction, Linear Regression, Machine Learning, Classification, Data mining*

Introduction

The stock market is the most interesting and powerful domain with every working professional imagining to become rich by making investments in the stock market. A stock (also known as "securities exchange") is a kind of security that designates proportionate possession in the giving organization. The financial exchange works like a sale where speculators purchase and sell portions of stocks. Stock costs normally mirror the financial specialist's assessments of what the organization's income will be.

There are various factors that affect the prices of stocks which can be categorized in two ways. Fundamental factors and technical factors. When a company returns huge profits solely on the basis of production and sale of goods and services, it means the fundamental factors are deciding the prices of stock. In technical factors, the company's historical data is studied based on charts, behavioral factors of traders and investors, etc.

While the fundamental factors may not be in our hands, researches have dedicated their time and money in exploiting the technical factors to give accurate predictions on the performance of a stock in the future so as to decide whether to buy or sell a particular stock. There are too many parameters that can be manipulated to make predictions regarding stock market such as opening price, closing price, volumes of shares traded, etc. In this paper, we propose a website which will predict the opening price of a company the customer or user is interested in.

In the backend the foundation of our website uses a machine learning model to make predictions We are going to consider historical data of 5 companies to train our model. It will act as an input to our

ISSN: 2233-7857 IJFGCN

model. The algorithm used in our model is Linear Regression which helps us make predictions. In the front end, there are two modules - user module and admin module.

The remainder of the paper is sorted in the following manner. Section II gives the Motivation and Need for our project. Section III discusses the Literature Survey which demonstrates how the issue of "securities exchange" has been handled in earlier years. Section IV makes you acquainted with the existing framework and our proposed framework. Segment V elaborates on our proposed framework in detail covering algorithms and modules of our site. In section VI results of our experiment have been shown. And finally, we conclude our topic in section VII.

Motivation

We selected such a topic with the view of contributing to our country's economy. India is the second most populated country of the world after China having 1.19% growth rate as per 2016 census. And majority of the population lies between the range of 15-60 years as per 2018 study. If more people invest in stocks it will be beneficial to the overall economy of India as companies will get cash flow to clear-off their debts and launch new products and services which is just one of the many benefits of investing in stock.

Need

There are two types of investors who invest in stock. Either full time investors or people investing after regular intervals of time. The latter type of investors back-out of the process if they face any losses. To encourage them to keep investing regularly we are building this site.

Literature Survey

ASHISH SHARMA ,DINESH BHURIYA ,UPENDRA SINGH[1]

Stock market is nonlinear in nature and also the analysis on stock exchange is one of the vital problems in recent years. For predicting, the stock trade peoples search such alternatives and tools which can increase their profits, whereas minimize their risks. Prediction plays a vital role in available market business that is incredibly more difficult using ancient ways like elementary and technical analysis might not make sure the accuracy of the prediction. to do predictions multivariate analysis is employed largely. In this paper they did the survey of well-known regression approaches to predict the stock trade value. In future the results of multivariate analysis approach might be improved victimization additional variety of variables.

Zhao, Lei, Wang, Lin[2]

In this paper they have used unique information mining approach to predict long term of stock trend. Traditional techniques on stock trend forecast have shown their disadvantages. In their analysis, a unique outlier mining algorithmic rule is used to discover anomalies on the idea of volume arrangement of high recurrence tick by tick data. Such anomaly trades continuously logical thinking with the stock value within the stock exchange.Results show that their approach makes profits on the Chinese stock exchange.

Ryota Kato and Tomoharu Nagao [3]

In this paper, we have a tendency to propose a securities market prediction methodology supported statistic information. There are plenty of securities market prediction models, there are few models that predict stock by considering alternative statistic information. So we have a tendency to

ISSN: 2233-7857 IJFGCN

concentrate on extracting interrelationships between the expected stock and varied statistic information, like alternative stocks, world securities market indices, foreign exchanges and oil costs.

Mustain Billah ,Sajjad Waheed ,Abu Hanifa[4]

Giving closing stock price accurately is associate degree difficult job. Laptop assisted systems are proved to be useful tool for stock prediction like Artificial Neural Network, accommodative Neuro Fuzzy abstract thought System , etc. Latest analysis works prove that accommodative Neuro Fuzzy abstract thought system shows higher results than Neural Network for stock prediction. During this paper, associate degree improved Levenberg Marquardt coaching formula of artificial neural network has been planned. Improved Levenberg Marquardt formula of neural network will predict the attainable day- end closing stock worth with less memory and time required, provided previous traditional exchange knowledge of Dacca stock market like gap worth, highest worth, lowest worth, total share listed. Moreover , improved Levenberg Marquardt algorithm can predict day-end stock price with fifty three less error than ANFIS and ancient lumen formula. It conjointly needs half-hour less time, fifty four less memory than ancient lumen and forty seventh less time, fifty nine less memory than ANFIS

Tejas Mankar, Tushar Hotchandani, Manish Madhwani, Akshay Chidrawar, Lifna C.S [5]

Machine learning and computing techniques are getting used alongwith data processing to resolve excess of world issues. To add to their annual financial gain, nowadays, individuals have started observing stock investments as a remunerative possibility. With professional steering and intelligent coming up with, we will nearly double our annual revenue through stock returns. The concern of losses additionally acts as a deterrent to several. These facts propelled USA to harness the ability of ML to predict the movement of stocks. Victimisation sentiment analysis on the tweets collected victimisation the Twitter API and additionally the closing values of assorted stocks, we have a tendency to get to make a system that forecasts the stock worth movement of assorted firms. Such a prediction would greatly facilitate possible stock capitalist in taking enlightened selections which might directly contribute to profit.

Carol Hargreaves ,Yi Hao[6]

With the easy access to share information and data, many investors worldwide are interested in predicting stock prices. Our two goals are to validate our stock selection methodology and to determine whether our trading strategy allows us to outperform the Australian market. Simulation results show that our selected stock portfolios outperform the Australian All-Ordinaries Index.

AbdullahAl-Luhaib, KhaledAl-Ghoneim, Yousef-Al-Ohali[7]

Features from the Saudi stock exchange (SSM) are examined to aim to predict the direction ofdailyprice changes. Backpropagation neural network has been applied to predict the direction of worth changes for the listed stocks in SSM. the value modification in SSM ranges between -10% and 100 percent. The target encompasses a illustration ofthree categories one, -1 and zero that severally represent the rise, decrease or insignificant modification within the stock costs. The dynamic target could be a novel improvement to the normal objective perform mean-squared-error (MSE) for higher classification. Our preliminary results showthat the classifier's performance improved victimisation dynamic targets in terms of quantitative performance and qualitative performance. additionally, experiments were conducted to work out the most effective hardening perform for objective targets.

ISSN: 2233-7857 IJFGCN

Yaoqing Wang ,Yaojun Wang [8]

Price prediction available market is taken into account to be one amongst the foremost tough tasks, attributable to the worth dynamic. Previous study found that stock value volatility during a short term is closely associated with the market sentiment; particularly for capitalization stocks. This paper used the social media mining technology to quantitative analysis market section, and together with alternative factors to predict the stock value trend . Experiment results show that by exploiting social media mining combined with alternative info, the stock costs prediction model will forecast additional correct

Ze Zhang, Yongjun Shen, Guidong Zhang, Yongqiang Song, Yan Zhu [9]

Stock worth is one among Byzantine non-linear dynamic system. Typically, Elman neural network may be a native perennial neural network, having one context layer that memorizes the past states, that is kind of suitable partitioning statistic problems. This paper verifies that model by some stock cost and compares with BP network and Elman network, therefore on draw the result that shows the exactitude and stability of this declaration model each are superior to the neural network. This paper predicts the opening Price of the Stock Market. precision and stability of this declaration model measure superior to the normal neural network. This paper predicts the opening Price of the Stock Market.

Prediction of Stock Market by Principal Component Analysis " Muhammad Waqar[10]

The categorization of high dimensional knowledge gift a desirable challenge to machine learning models as frequent range of extremely related to dimensions or attributes will have an effect on the accuracy of classification model. during this paper, the matter of high spatial property of stock market is investigated to predict the market trends by applying the principal element analysis (PCA) with regression toward the mean. PCA will facilitate to enhance the predictive performance of machine learning strategies whereas reducing the redundancy among the information. Experiments are administered on a high dimensional spectral of three stock exchanges such as: the big apple stock market, London stock market and Karachi stock market. The accuracy of regression toward the mean classification model is compared before and when applying PCA. The experiments show that PCA will improve the performance of machine learning generally if and provided that relative correlation among input options is investigated and careful choice is completed whereas selecting principal parts. Root mean sq. error (RMSE) is employed as associate analysis metric to guage the classification modelSystem architecture.

System Architecture

1. Existing System

The Stock Market data is obtained through web crawling. This Stock Market data is basically of two types:-

-News Data.

News data is data about stock which is available on daily basis through news on websites. Based on this daily data we apply a sentence-based sentiment analysis approach which is used to process the textual data during a specific period. Sentiment is a opinion or feeling that we have about something.

ISSN: 2233-7857 IJFGCN

We divide the document into sentences first as deciding the sentiment of sentence is easier as compared to seperate words[11]. After that we segment the sentences into separate words and project them on sentiment space and then count the number of positive and negative words and then decide the polarity of each sentence. The value of sentiment ranges from -1 to +1, where 0 means people hold neutral value, +1 means positive view and -1 means negative view. Further we consider the day-of-the-week effect, which states that average investment on monday is much less than that on other days of the week because if an investor suffers loss he will need atleast two or three days to overcome his thoughts and invest in a new stock After that we get the modified sentiment index which becomes the first input for SVM.

-Market Data

In market data, there is price and trade related data such as stock exchange data. Market data allows traders and investors to know the latest price and see historical trends for currencies or stocks. On market data we apply attributes such as high, low, open, close and volume. *Open* and *Close* represent the starting and final price at which the stock is traded on a particular day. *High*, *Low* and *Last* represent the maximum, minimum, and last price of the share for the day[12]. Also we give labels to indicate 1 for sale and +1 for buying. This acts as the second input to SVM model. To increase the accuracy of SVM model, we combine it with rolling-window approach. In addition, a stop-loss order strategy is applied to limit the potential losses, and it accomplishes a much better performance.

Disadvantages:

-It was not giving accurate results.
-Provided less stability.

2) Proposed System

In Proposed system we are making use of regression analysis as the comparative study has revealed that linear regression gives more accuracy than SVM. In Proposed system, we intend to abuse financial specialist conclusion to conjecture securities exchange development heading by underscoring the job of speculators. Financial specialist brain science drives the securities exchange and it makes a difference for our exploration. As needs be, client created content on the Internet gives a valuable source to reflect financial specialist brain science. Conclusion examination is utilized to change over unstructured printed records into day by day assessment files.

Advantages:

-The system is user friendly.
-System is flexible, so user is more satisfied.
-Stock data classification result accuracy is efficient**.**

ISSN: 2233-7857 IJFGCN

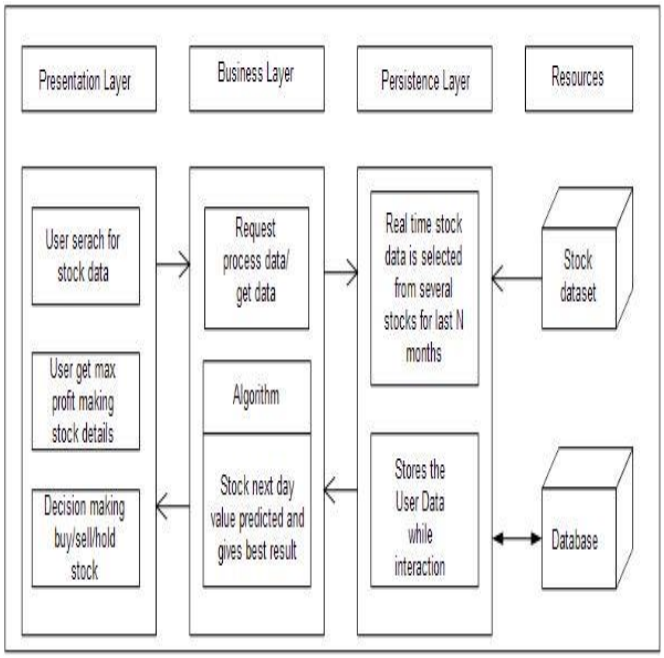

Fig 1 Proposed System Architecture

Algorithm used In Proposed System

Linear regression is one of the most well known algorithm in machine learning. Linear regression is a part of a special area of statistics called as Bivariate statistics. Bivariate means two variables (one dependent and one independent). The value of one variable is a function of the other. The value of y is a function of x; y=f(x). The value of dependent variable is a function of the independent variable. Dependent variable are continuous in nature. In our case dependent variable is stock and independent variable is time.

Linear Regression could be methodology accustomed model a relationship between a variable(y), associated and experimental variable(x) with straight forward statistical regression, there'll solely be one experimental variable. There will be several free lance variables which might compromise the class of multiple statistical regression during this circumstance, we tend to solely have one experimental variable that is that the date. The date are described by associate number beginning at one of the primary date intensifying to the length of the vector of dated which may vary looking on the statistical knowledge. Our variable, of course , are the value of the stock so as to know statistical regression, you need to perceive a reasonably elementary equation you most likely learned to soon at school.

$y = a + bx$
Where:

Y = anticipated worth or variable quantity b = slope
x = constant or variable quantity
a =y-intercept

ISSN: 2233-7857 IJFGCN

Modules

There are 2 types of modules:

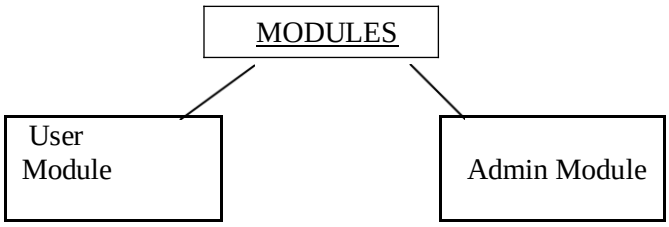

Fig 2 Module

Let us discuss them in detail:

1) User Module

There are sub modules in the user module. They are:

- New To Stock Market

If the user is new to stock market , that means the user doesn't know anything related to stocks and stock market then the user will be provided with four videos. By seeing those videos he will be able to understand what stock market is, how to invest, what are the advantages, what profit the user will get and lot many things will be cleared.

If the user doesn't wants to see the video and he wants something theoretical then the user needs to click on more about stock market field and then the user will be redirected to a google page from where he can read all the related information about stocks.The user can also see the current stocks rates which will be more helpful for him so that he can invest properly or buy the stocks accordingly.

- Stock Details

This sub module will help the user to see the stock details of a particular company. The user will be provided with a drop down list from where he can select the desired company he wants and can see the stock details of that company.
This will be helpful for him because he can compare the stock details of other companies and then he can take his decision.

2) Admin Module

In admin module, the admin can upload the dataset of various companies which will be helpful for the user. As the number of companies will increase it will be good for the user to compare and then invest.

ISSN: 2233-7857 IJFGCN

Result

- Snapshot

Fig 3 New To Stock Market

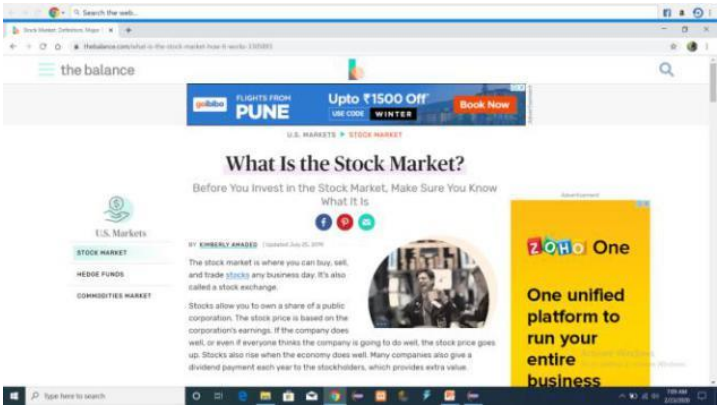

Fig 4 More About Stock Market

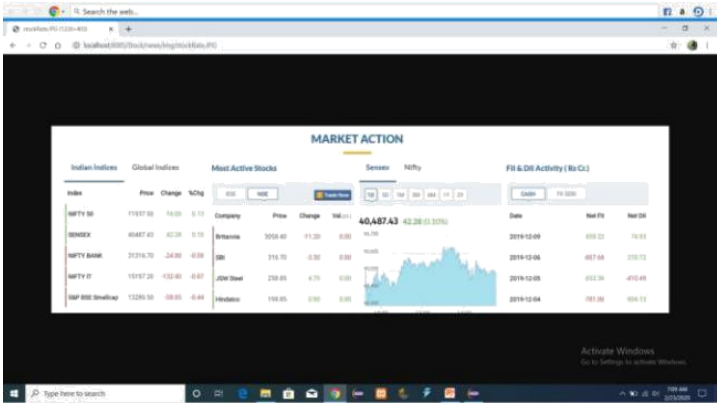

Fig 5 Todays Stock rate

ISSN: 2233-7857 IJFGCN

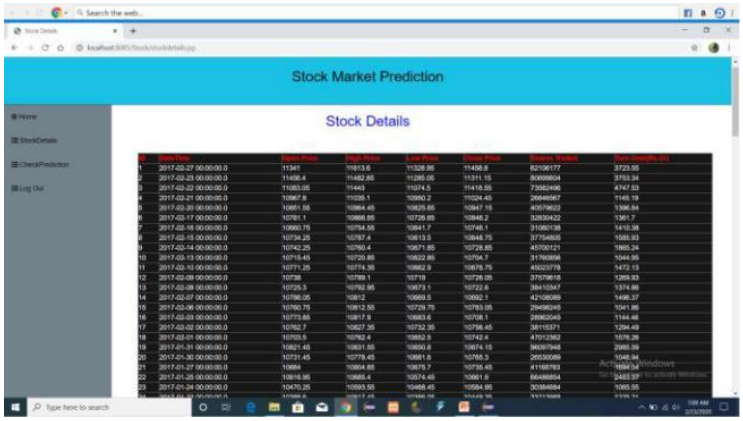

Fig 6 Stock Details

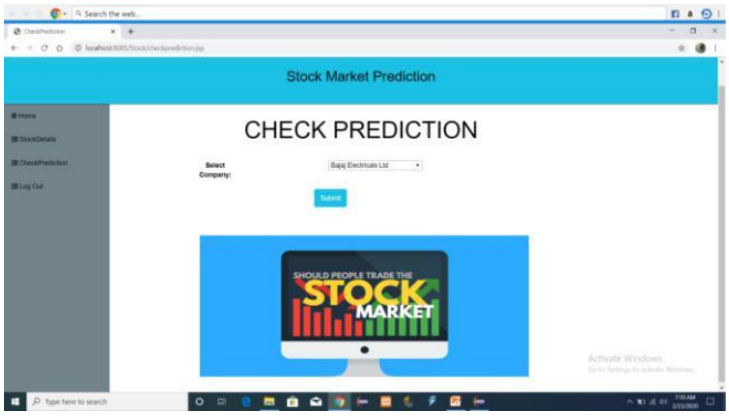

Fig 7 Check Prediction

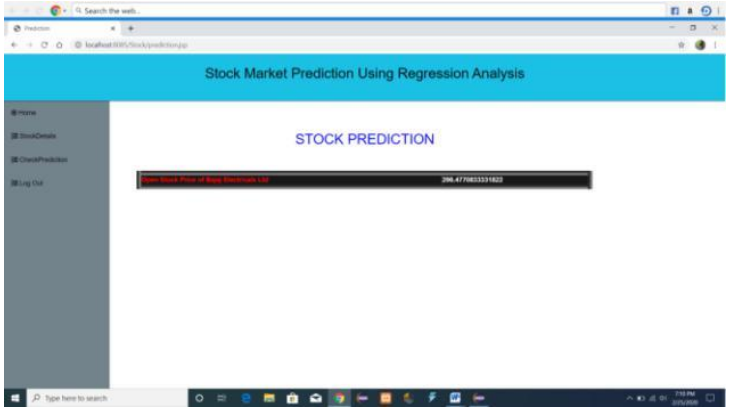

Fig 8 Prediction Result

ISSN: 2233-7857 IJFGCN

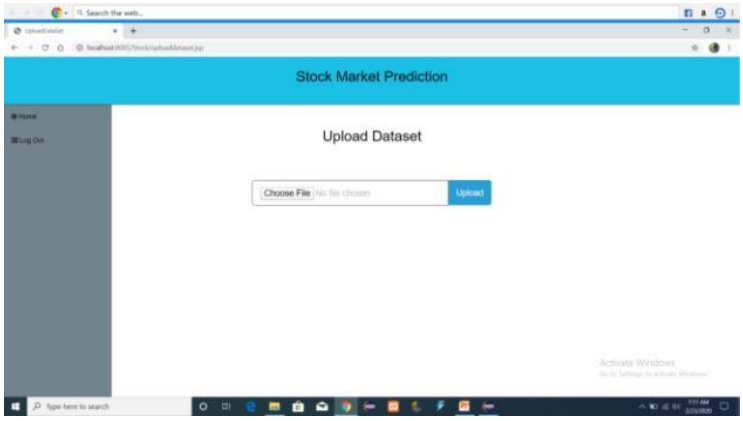

Fig 9 Upload Dataset

Future Work And Conclusion

- Future Work

As we have developed this project at college level we have kept a limited time span of 2 years and a limited number of stocks i.e. 149. In future if there is need of increasing the span it can be increased to 5years or 10 years. Our system is feasible for the increased time span. The number of stocks can also be increased without affecting the performance level

- Conclusion

In this paper, we proposed a stock market prediction using linear regression. For this purpose, we built a website that will predict the opening price of a particular company. We then compared the results shown by our model with the one using SVM as the base algorithm. We have used only 5 companies to train our model, for future scope increasing the quantity of dataset to train the model can be done. Also, if any other algorithm is better than Linear Regression can be found out.

References

[1] “Survey of Stock Market Prediction Using Machine Learning Approach”.ASHISH SHARMA HOD Govt. Women’s Polytechnic, Indore.DINESH BHURIYA Govt.Women’s Polytechnic, Indore.UPENDRA SINGH”,2017

[2] “Price Trend Prediction of Stock Market Using Outlier Data Mining Algorithm”.Zhao, Lei Baylor University.Wang, Lin Japan Advanced Institute of Science and Technology”,2015

[3] “Stock Market Prediction Based on Interrelated Time Series Data”. Ryota Kato and Tomoharu Nagao Graduate School of Environment and Information Sciences, Yokohama National University Kanagawa, Japan”,20l2

[4] “Stock Market Prediction Using an Improved Training Algorithm of Neural Network” Mustain Billah Department of Information And CommunicationTechnology Mawlana Bhashani Science And Technology University Tangail, Bangladesh”,2016

[5] “Stock Market Prediction based on Social Sentiments using Machine Learning TejasMankar, Tushar Hotchandani”,2018

[6] “Does the use of Technical & Fundamental Analysis improve Stock Choice? : Data Mining Approach applied to the Australian Stock Market.Carol Hargreaves Institute of Systems Science

ISSN: 2233-7857 IJFGCN

National University of Singapore. Yi Hao School of Business Economics Wilfrid Laurier University Waterloo, Canada",2012

[7]"DYNAMIC TARGETS FOR STOCK MARKET PREDICTION AbdullahAl-Luhaib KhaledAl-Ghoneim YousefAl-Ohali",2017

[8] "Using Social Media Mining Technology to Assist in Price Prediction of Stock Market".Yaojun Wang Institute of Computing Technology, - Yaoqing Wang Zhejiang University of Science and Technology School of Economics and Management.",2016

[9] "Short-term Prediction for Opening Price of Stock Market Based on Self-adapting Variant PSO-Elman Neural Network". Ze Zhang, Yongjun Shen, Guidong Zhang, Yongqiang Song, Yan Zhu School of Information Science& Engineering Lanzhou University",20l7

[10]"Prediction of Stock Market by Principal Component Analysis.MuhammadWaqar, Hassan Dawood,Muhammad Bilal Shahnawaz, Mustansar Ali Ghazanfar.Software Engineering Department University of Engineering and Technology Taxila" , Pakistan,2017.

[11] Udayan Birajdar, Sanket Gadhave, Shreyas Chikodikar, Shubham Dadhich, Shwetambari Chiwhane, "Detection and Classification of Diabetic Retinopathy Using AlexNet Architecture of Convolutional Neural Networks", Proceeding of International Conference on Computational Science and Application, online 05 January 2020, pp 245-253.

[12] Dr. C. Nalini, Shwetambari Kharabe, Sangeetha S," Efficient Notes Generation through Information Extraction", International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8 Issue-6S2, August 2019.

ISSN: 2233-7857 IJFGCN