

Company's Stock Price Predictor using Machine Learning

Prof. Vikas Palekar¹, Piyush Zade², Bhumika Vaidya³, Rachana Mehra⁴, Nikhil Pradhan⁵,
Ashwini Lambat⁶

¹Assistant Professor, Dept. of Computer Science & Engineering, Datta Meghe Institute of Engineering Technology & Research, Wardha, Maharashtra, India

^{2,3,4,5,6} Student, Dept. of Computer Science & Engineering, Datta Meghe Institute of Engineering Technology & Research, Wardha, Maharashtra, India

Abstract - Due to the unsettled, composite and time-to-time changing of prices, stock market prediction has attracted a lot of interest from investors and researchers for quite a time. It is very difficult to make dependable predictions. The stock market prices are predicted using the machine learning algorithms like k-Nearest Neighbors, in this system. The model that has been highly triumphant for classification and regression is the k-Nearest Neighbors model. KNN algorithm is a very popular classification algorithm demonstrating good performance characteristics and also a short period of training time. We must specify such sets of neighborhoods that might approve the close neighbors, as stated by K nearest neighbor's distribution. We use Support vector machine for the purpose of classification and it is a machine learning model. This model is also used for classification for most of the times. Based on historical data, this system provides the information whether the price of stock will go up or down using these techniques. This also provides deep knowledge of the models that are used in this process.

Key Words: stock market, economic growth, dynamic nature, investments, prediction, data mining, machine learning, k-Nearest Neighbours.

1. INTRODUCTION

In the growth of developing countries like India, the stock market plays a very vital role. Hence the growth of many developing nations depend on their countries stock market. If stock market is doing well and going high, then the countries growth would be good but if the stock markets are not doing well, then it may affect the countries growth negatively. [1][2] We can conclude that any countries growth and its stock market are very closely connected. In a country like India with its huge population there are only 10% of people who indulge themselves to trade in stock market. This is only because of the nature of stock market as it is highly volatile. [2]. Some people have wrong conceptions about the stock market and they consider it as a place where people do gambling. As countries development depends on stock market, so this misconception of people should be cleared and awareness about stock markets should be brought. These

misconceptions can be cleared by using the prediction systems in stock market. The more accurate predictions can make the positive beliefs of people on stock market. The mindset of people can be changed by providing an accurate prediction system. To predict the future trends and the prices of stocks in future, Data Mining plays a very important role. They help the companies to take knowledgeable decisions based on analysis. [3][4] Data Mining is one of the fragment of knowledge discovery but in some cases knowledge discovery and data mining are used in similar way. [3][4][5].

1.1 RELATED WORK

In this paper the prediction of stock market prices is suggested by using learning models like Support Vector Machine and Random Forest model. For classification and regression, the Random Forest model has been proved very successful. Classification is usually done by the Support Vector Machine. [1] This literature consists of various distinct machine learning algorithms such as Artificial Neural Networks (ANN). [2] In situations where statistical and traditional methods are not giving accurate results, to increase the accuracy KDD is used. This KDD can find the hidden patterns thus increasing the level of accuracy. [4]

1.2 PROBLEM STATEMENT

The financial organizations that are responsible for the conveyance of various different goods between the people or investor and stock brokers are the stock exchanges. The yield of stock market is in billions of dollars so this makes the people very anxious to make profits in such a market. The value of a share decreases if it is transacted in less volume. The profits and losses in stock market completely depends on the ability to predict the prices in future. Hence, the problem for investors is that at what time they should enter or buy the stock and at what time they should exit or sell the stock of any company. This is such an area which has

charmed the likes of many researchers for predicting the future prices of various goods.

The main problem is to get the real time data from the internet. Fetching the data automatically from the internet and working on real time data is not available in existing system.

As the real-time data is not available, the correct and accurate analysis of data is not possible due to which accurate predictions of stock price is not possible in this conditions. This creates a huge problem for the traders as they are unable to get the correctly predicted price for any stock.

2. ARCHITECTURE

2.1 DATA FLOW DIAGRAM

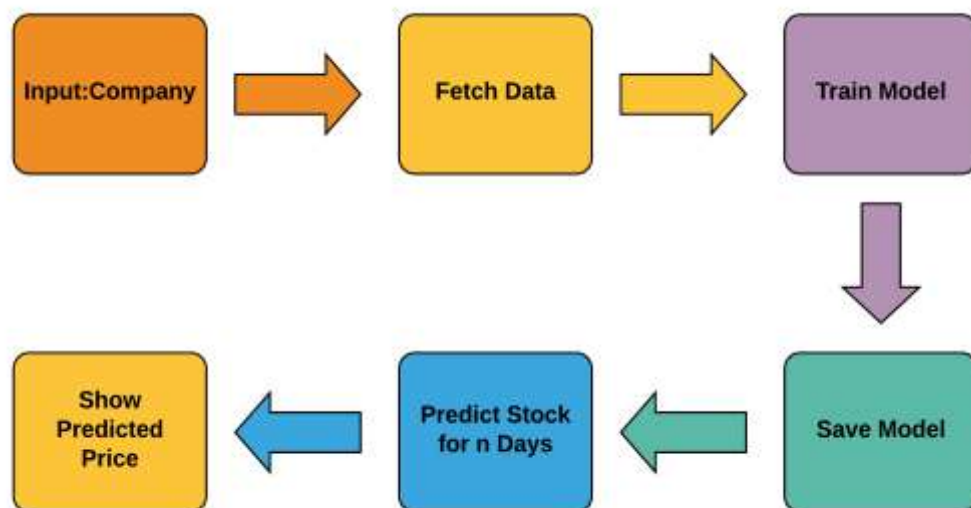


Fig. 2.1 Data Flow Diagram

- The User will visit the application and ask for the company’s stock data.
- Previously saved model is loaded as well as new data will be fetched from website.
- Previously saved model is trained with new data, i.e., newly fetched data will be stored in database.
- Stock data will be shown to the user.
- User requests for prediction to be made.
- The Stock Market Prediction System will predict the price with the help of saved model.
- Result is displayed and predicted price is shown.

2.2 SYSTEM ARCHITECTURE

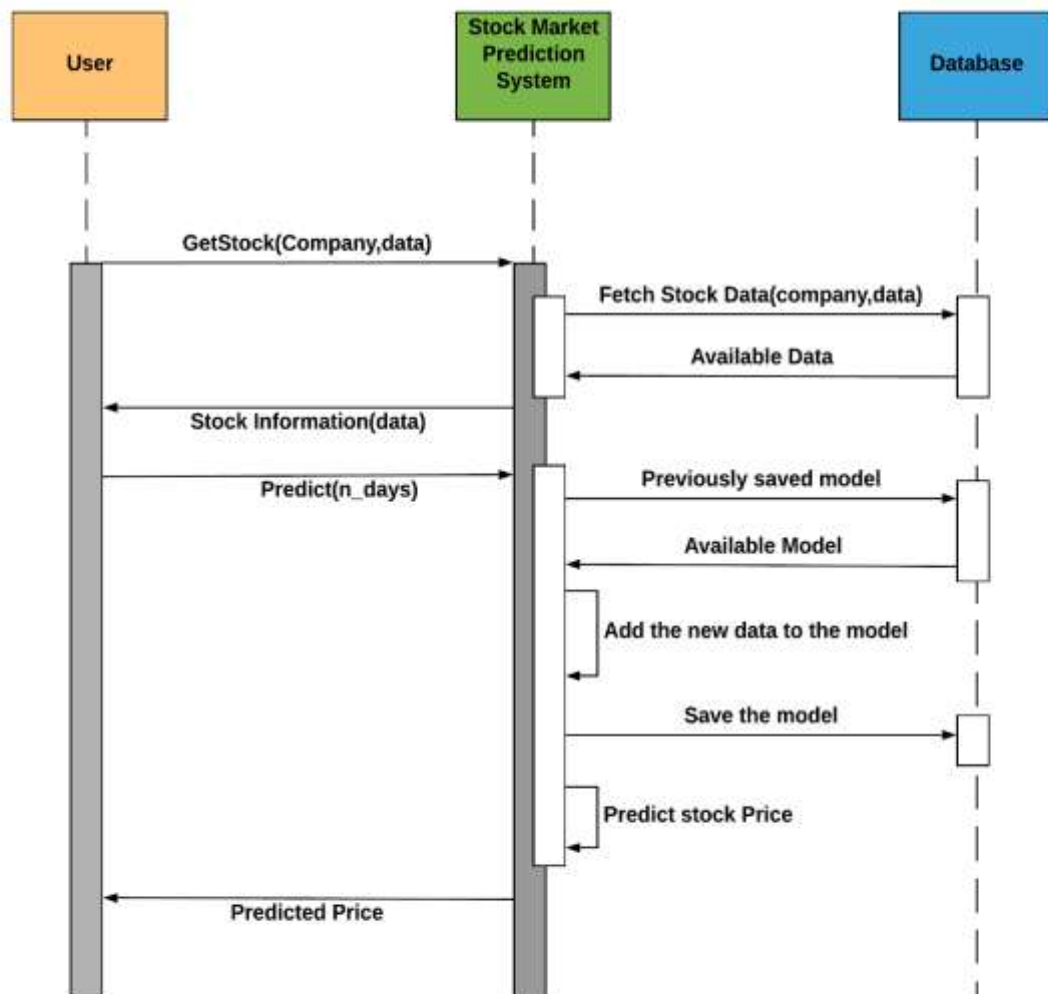


Fig. 2.2 System Architecture

3. METHODOLOGY

3.1 MODULE DESCRIPTION

Module 1: Data Fetching

- Web Crawling:** The main use of web crawler is that it is used for web indexing. It indexes each and every page one at a time of a website until every page is done. Gathering the data about any website and all the links associated to them is done with the help of web crawlers. They help in affirmation of the HTML hyperlinks and code. Web crawler is often referred to as an automatic indexer or web spider. [7]

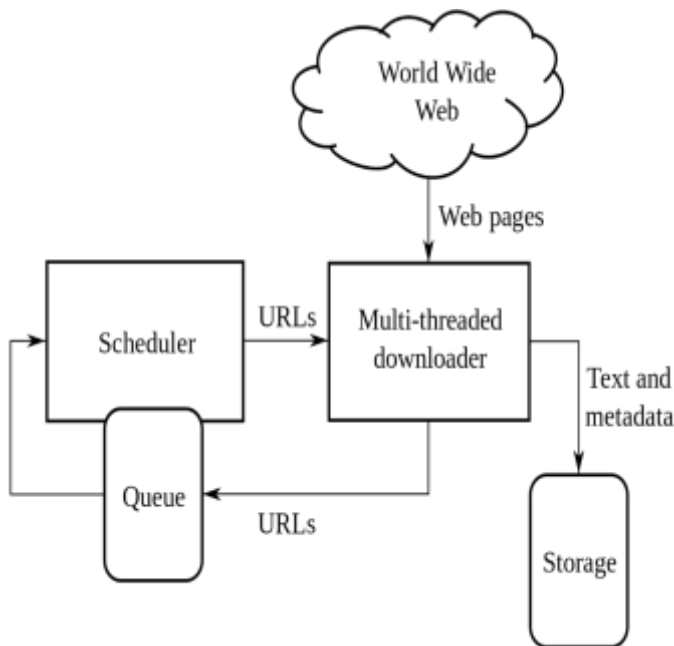


Fig. 3.1 Architecture of a Web crawler

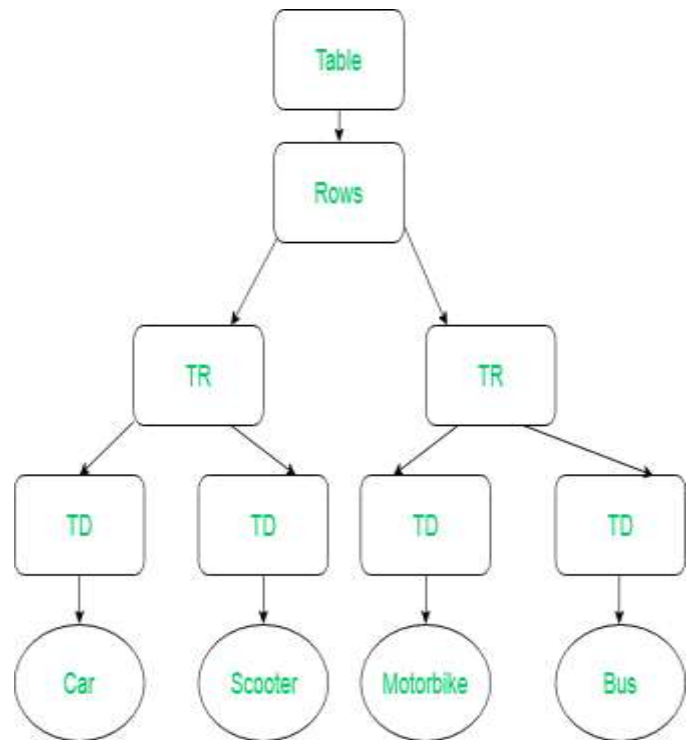


Fig. 3.2 Structure of Dom Tree for example

- **Applying Dom Tree Algorithm:** The Document Object Model(DOM) is a programming port. It is used for XML and HTML files. It shows a way in which a file is manipulated and accessed and also the logical structure of file.
- We can consider DOM as a tree or a forest. Tree-like portrayal of a file can be described by the structure model. We can have two DOM application to represent a same file. This nature is called as structural isomorphism. [6]

Module 2: Data Storing

- We are storing the data fetched from the website to the database.
- The data is to be stored by making several tables in the database to access the data of any given day easily.

Module 3: Analysis & Prediction

- **Analysis:** Analyzing the data collected using backpropagation of neural network algorithm.

Backpropagation of neural network algorithm:

Backpropagation algorithms are used to accurately train the artificial neural networks which follows a gradient-based optimization algorithm due to which the chain rule breaks. Efficient, iterative and recursive are the three key properties of backpropagation which helps in computing the weight which in turn is useful and improves the network until it performs the task accurately for which it has been trained.

- **Prediction:** We are predicting the price by using KNN Algorithm
- The KNN algorithm is used for classification or regression. It is a non-parametric algorithm which means that it does not make any prediction regarding the basic information or its dispersion.
- For each data point, the algorithm finds the k closest observations, and then classifies the data point to the majority. [8]

KNN Algorithm:

1. Fetch and fill the data.
2. For each selected number of neighbors, initialize K.
3. For each every in the given data
 - 3.1 Compute the distance among the present example and the query example from the information.
 - 3.2 Take sum of the index and distance calculated to an ordered collection.
4. The group of indices and distances should be sorted in ascending order as per the distances.
5. From the sorted collection, draw the first K entries.
6. Obtain the labels for the K entries which were selected.
7. If it comes to be regression, then mean of K labels should be returned.
8. If it comes to be classification, then mode of K labels should be returned. [8]

K-Median Algorithm:

K-Medians clustering algorithm is an algorithm that is referred as cluster analysis algorithm. It is slightly different from k-means clustering as in the k-means we calculate the mean for each cluster while in k-median, median is calculated.

In the Manhattan- distance formulation the distance is calculated in every single dimension therefore each individual attribute will come from dataset. Thus the algorithm becomes very reliable specially for the even binary or discrete datasets.[9]

We are going to calculate or compute the prediction of stock prices over a week or over a month. For example, if we want to calculate the price of stock of any company on any given Monday, then we will be able to calculate the price on the basis of the prices on Monday's in previous weeks. Same goes with the monthly predictions. For example, if we want to know the trend of the price of stock of any company on any given month say May, then we are able to do it from the data of previous years May months' data.

We are going to calculate and make this prediction by using K-Median algorithms. In this method, we will take the median of the prices of previous days and use that value with our algorithms to calculate the predicted price of the stock. In above example for Monday, we can calculate its price by taking the median and applying algorithms on the data of previous Mondays that is being collected in the database.

4. OUTPUT/RESULT

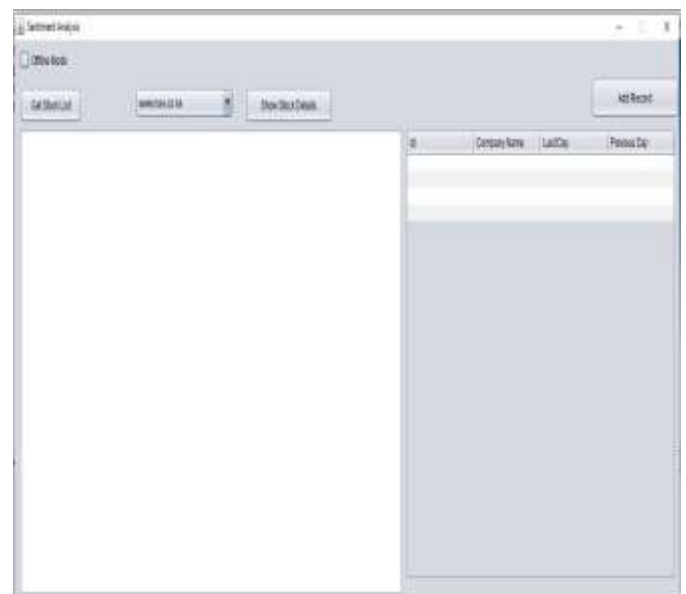


Fig 4.1: Screenshot1

In **Screenshot 1**, we are representing the module 1 which is data fetching from the url specified. AWT is used to form the front end of the application. On this page we are fetching the data from the website - <https://www.nse.co.ke/market-statistics/equity-statistics.html>.

We are providing an option for offline mode, in case the internet connection is not available so the previously loaded data can be used.

Once the user clicks on Get Stock List button, the data is fetched from the website.

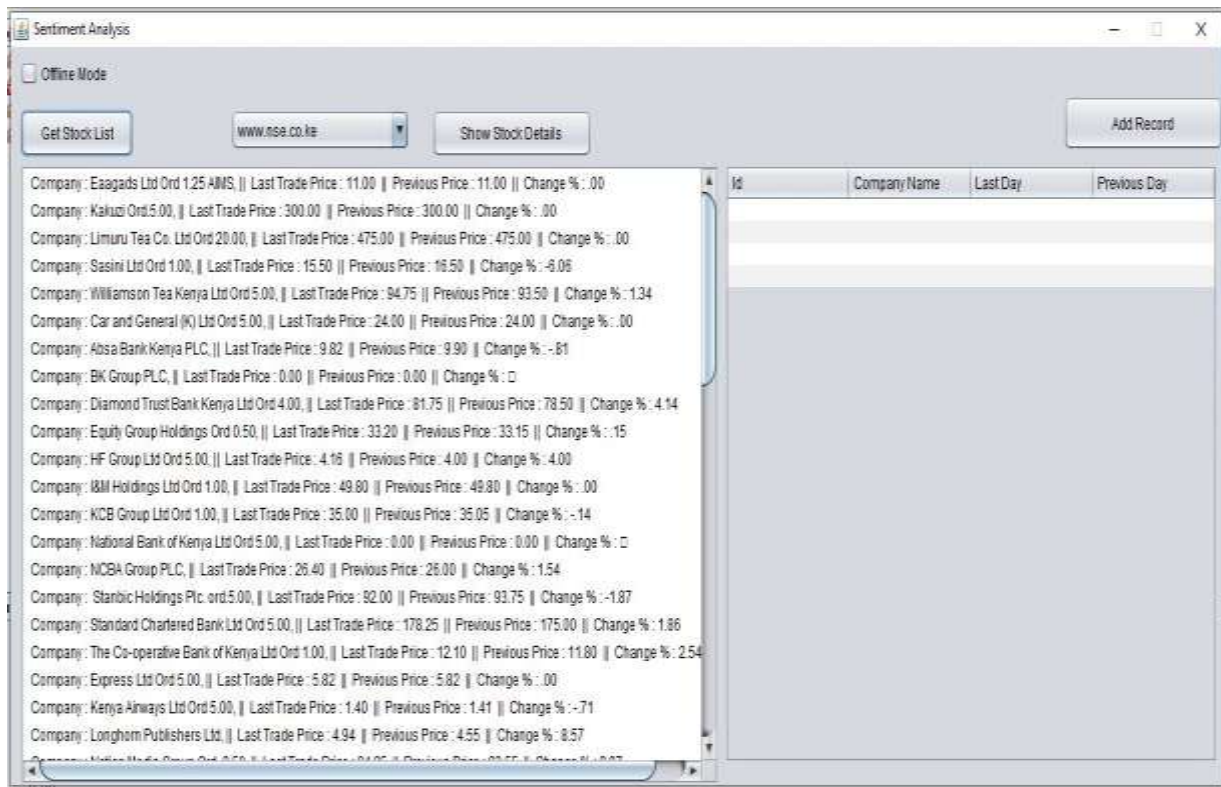


Fig 4.2: Screenshot2

In **Screenshot 2**, we have fetched the data from the provided website. This data is displayed in the form of list in the box.

This data can only be fetched when the system is connected to the internet otherwise the system will stop giving an error.

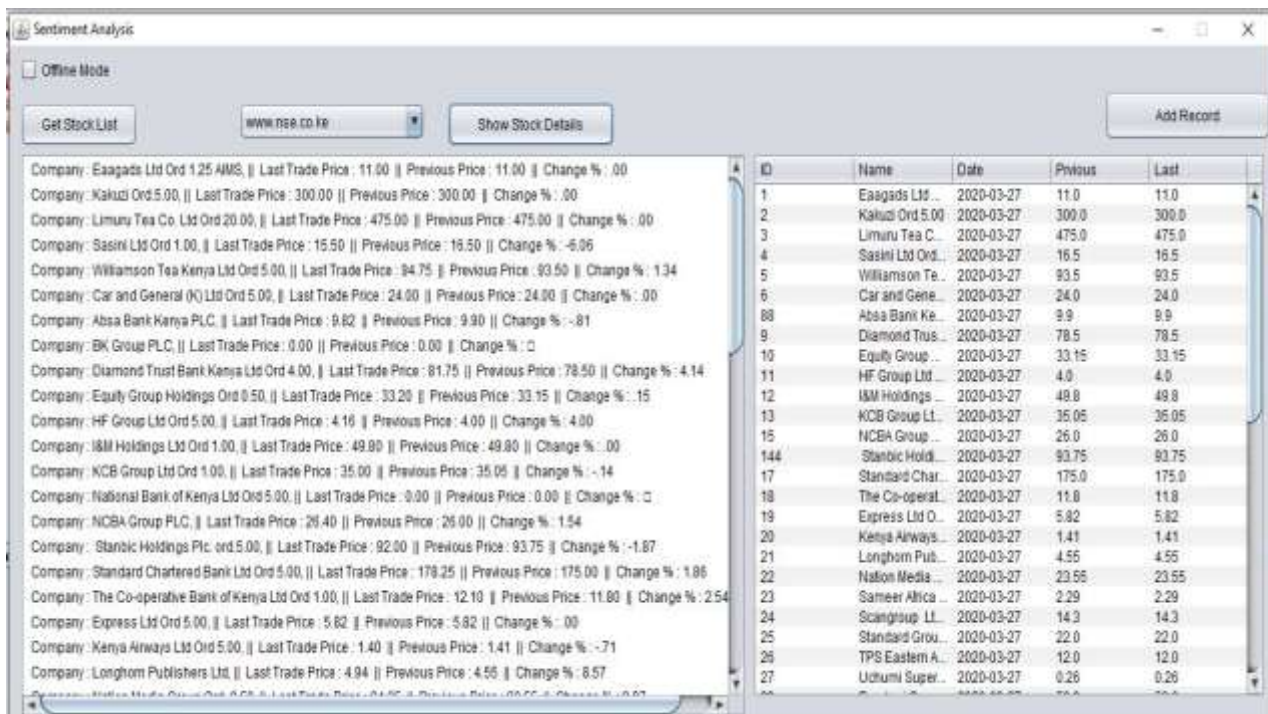


Fig 4.3: Screenshot3

In **Screenshot 3**, as we have fetched the data from the website, we are saving that fetched data in our database. Once the data is stored in our database we can use it for further operations. The stored data can be displayed in the form of list on the right (fig. 4.3).

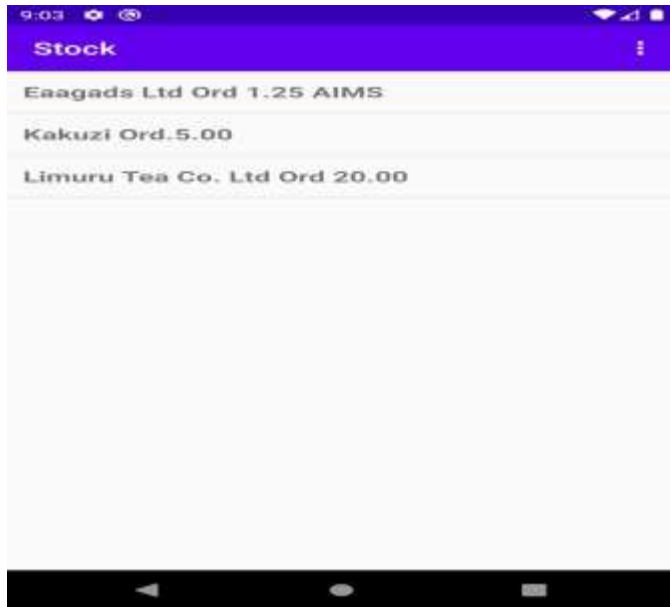


Fig 4.4: Screenshot4

In **Screenshot 4**, we have displayed the list of stocks of various such companies for which we want to see the prediction.



Fig 4.5: Screenshot5

In **Screenshot 5**, the prediction of stock prices of a particular company over the coming days of a week is given. Also the prediction for such stock prices of any

given company can be predicted over a period of few months. A graph is shown (Fig. 4.5), which signifies the prediction made for the upcoming days of a week.

5. CONCLUSION

- The proposed system is capable of fetching the data automatically from the website by using Web crawling & Dom tree algorithm. We can use Dom when we are working with any normal data structure and the present XML schema is not required, then we can find more than one object-oriented standards. In our case, we do not require the complete XML code but only some attributes from it so we have used Dom tree algorithm. We have used Dom because it has following advantages:
 - Relatively easy to change the data structure and extract data.
 - XML DOM is dynamic in nature and can be edited. It allows the developers to add, update, move or delete the nodes in the tree.
 - We get a platform and independence in languages due to XML DOM.
- We are able to analyze the data by using backpropagation of neural networks and predict the price using the KNN algorithm.
- We are using KNN algorithm for classification and prediction of stock prices. We are using KNN because it has following advantages:
 - Implementation is very simple.
 - We can update the classifier at a very little cost as instances of known classes are being presented.
 - The parameters to tune are firstly, distance metric and secondly, k.
- In this case we have used K-Median algorithms to predict and calculate the price of stocks. We have calculated the median of the prices from the data that is available with us in the database and then we have applied algorithms on this median values. After applying this algorithms, we found that we are able to predict the price of stock and it is quite correct method and it does gives quite accurate and suitable results in our case.

REFERENCES

- [1] Sahaj Singh Maini, K. Govinda Stock market prediction using data mining techniques “2017 International Conference on Intelligent Sustainable Systems (ICISS)”.

- [2] Radu Iacomin, Stock market prediction “2015 19th International Conference on System Theory, Control and Computing (ICSTCC)”.
- [3] Ko Ichinose Stock market prediction from news on the Web and a new evaluation approach in trading “2016 5th IIAI International Congress on Advanced Applied Informatics”.
- [4] Ashish Sharma, Dinesh Bhuriya, Upendra Singh Survey of stock market prediction using machine learning approach “2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)”.
- [5] Tao Xing, Yuan Sun, Qian Wang, Guo Yu The Analysis and Prediction of Stock Price “2013 IEEE International Conference on Granular Computing (GrC)”
- [6] GeeksforGeeks, “DOM (Document Object Model)”, <https://www.geeksforgeeks.org/dom-document-object-model/>
- [7] Wikipedia, “Web crawler” https://en.m.wikipedia.org/wiki/Web_crawler
- [8] Towards Data Science, “Machine Learning Basics with the K-Nearest Neighbors Algorithm” <https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761>
- [9] Wikipedia, “k-medians clustering” https://en.wikipedia.org/wiki/K-medians_clustering