A PROGNOSIS FOR STOCK MARKET PREDICTION BASED ON TERM STREAK INFORMATION

Pugazhenthi.P[1], Ruban Raj.K[2], Saadh.A[3] , PRIYANKA.S.V[4]

B.E, (Computer Science and Engineering, T.J.S Engineering College, Tamilnadu, India.

ABSTRACT

The Stock market process is full of uncertainty and is affected by many factors. Hence the Stock market prediction is one of the important exertions in finance and business. In this system technical analysis are considered. Technical analysis is done using historical data of stock prices by applying machine learning. The learned model can then be used to make future predictions about stock values. The system is trained by using machine learning algorithm. Then the correlation between the stock values is analyzed. The learned model can then be used to make future predictions about stock values. It can be shown that this method is able to predict the stock performance. In this system we applied prediction techniques approach in order to predict stock prices for a sample companies. In stock predictions, a set of pure technical data, fundamental data, and derived data are used in prediction of future values of stocks. The pure technical data is based on previous stock data while the fundamental data represents the companies' activity and the situation of market. All the data will be classified and clustering using the data mining techniques.

Keyword: - *BIG Data, Networking, Internet Protocol, JAVA, J2EE, JAVA Servlets, My Sql, Modules, Testing, Use Case Diagrams, HDFS, Map Reduce, Database, Clustering, Net Beans, etc…*

1. INTRODUCTION

Now a day's Academia and Industry people are working on large amount of data, in petabyte, and they are using technique of Map Reduce for data analysis. The input for such framework is very large and main requirement for theses inputs are that all the files cannot be kept on single node. After putting all data on single machine, we have to process it parallel. Hadoop is a framework which enables applications to work on large amounts of data on clusters with thousands of nodes. A distributed file system (HDFS) stores the data on these nodes, enabling a high bandwidth across the cluster. Hadoop also implements a parallel computational algorithm, Map Reduce, which divides the main task into small chunks and these work in parallel known as mapping, and all the results are combined into a final output, the reduce stage. This paper is based on Hadoop Based Stock forecasting using neural networks. Stock Market has high profit and high risk features that's why its prediction must be in the parallel of accuracy, the main issue about such data are, and these are very complex nonlinear function and can only be learnt by a data mining method. We have tried to utilize distributing capability of Hadoop ecosystem which is parallel too. Map-Reduce for managing training of large datasets on the neural network. Our experimental results basically show the speedup achieved by increasing number of processors to the hadoop cluster for an artificial neural network. . To analyze the large volume of data and to process it, is difficult and challenging and there are different methods. Hadoop is a very fast way for massively parallel processing. Hadoop analyze the scattered data and predict the future trends and business intelligence solutions which would benefit the enterprise and client all together. In this system, we develop a broadly applicable parallel programming method. We adapt Google's map-reduce paradigm to demonstrate this parallel speed up technique on feed neural network. Big Data is a collection of large datasets that cannot be processed using traditional computing technique.

Map Reduce: Map reduce is the high level programming system. This helps in doing the computation of the problem in parallel using all the connected machines so that the output, results are obtained in efficient manner. DFS also

provides data replication up to three times to avoid data loss in case of media failures. The Master node stores the huge data HDFS and runs parallel computations on all the data i.e. Map Reduce.

1. The NameNode coordinates and monitors the data storage function (HDFS), while the Job Tracker coordinates the parallel processing of data using Map Reduce.
2. Slave node does the actual work of storing the data and running the computations. Master nodes give the instructions of their Slave node. Each slave runs both a Data node and a Task Tracker daemon that communicate with their respective Master nodes.
3. The Data node is a slave to the Name node.
4. The Task Tracker is a slave to the Job Tracker.

HDFS: HDFS is a distributed file system that provides a limited interface for managing the file system to allow it to scale and provide high throughput. HDFS creates multiple replicas of each data block and distributed them on computers throughout a cluster to enable reliable and rapid access. When a file is loaded into HDFS, it is replicated and fragmented into "blocks" of data, which are stored across the cluster nodes; the cluster nodes are also called the Data Node. The Name Node is responsible for storage and management of metadata, so that when Map Reduce or another execution framework calls for the data, the Name Node informs it where the data is needed resides.
Variety - The next aspect of Big Data is its variety. BigData can be generated from various fields. This means that the category to which Big Data belongs to is also a very essential fact that needs to be known by the data analysts.
Velocity - The term _velocity 'in the context refers to the speed of generation or processing of data. Since the speed of generation of this data is very high it increases the level of difficulty to process it.
Complexity - Data management is a very complex process, especially when large volumes of data come from different sources. These data need to be linked, connected and correlated so that it can grasp the information that is supposed to be conveyed by these data. And hence the complexity of Big-Data

1.1 Existing System

- Supervised learning approach is used to build dictionary which is time consuming, because of initial level of manual work. They are used index tracking algorithm.
- Setting some threshold values words are added to either positive dictionary of negative dictionary. This approach is not suitable for real time analytics until the dictionary is complete.
- Stock market data used for stock market prediction, data set which is used for train the model is very less.
- In existing system they are find the sentiment of the user comments and predict stock market status.

1.2 Objective

The stock market shows the variation of the market economy, and receives millions of investors 'attention from the time of opening development each day. The stock market is characterized by high-risk, high-yield, hence investors are concerned about the analysis of the stock market and trying to forecast the trend of the stock market. However, stock market is affected by various factors like the politics, economy , along with the complexity of its internal law, such as price changes in the non-linear, and so on therefore the traditional mathematical statistical methods to predict the stock market has not yielded suitable results. Thus, it is very suitable for the analysis of stock data.

1.3 Contribution

Analysis of stocks using data mining will be useful for new investors to invest in stock market based on the various factors considered by the software. Demand and Supply of shares of a company is a major reason price change in stocks. When Demand Increase and Supply is less, price rises.

2. LITERATURE SURVEY

Large-scale oil refineries are equipped with mission-critical heavy machinery (boilers, engines, turbines, etc.) and are continuously monitored by thousands of sensors for process efficiency, environmental safety, and predictive maintenance purposes. However, sensors themselves are also prone to errors and failure. The quality of

data received from these sensors should be verified before being used in system modeling. There is a need for reliable methods and systems that can provide data validation and reconciliation in real-time with high accuracy. In this paper, we develop a novel method for real-time data validation, gross error detection (GED) and classification (GEC) over multivariate sensor data streams. The validated and high- quality data obtained from these processes is used for pattern analysis and modeling of industrial plants. We obtain sensor data from the power and petrochemical plants of an oil refinery and analyze them using various time-series modeling and data mining techniques that we integrate into a complex event processing (CEP) engine. Next, we study the computational performance implications of the proposed methods and uncover regimes where they are sustainable over fast streams of sensor data. Finally, we detect shifts among steady-states of data, which represent systems' multiple operating modes and identify the time when a model reconstruction is required using DBSCAN clustering algorithm. This paper proposes a novel technique to approximate functions for time series and signal processing using the special type of neural network, called Critical Support Vector Machine (CSVM). CSVM is a combination of the Support Vector Machine, the Nearest Neighbor Algorithm and the Perceptron. The CSVM has been shown to be an effective method for classification problems. In this work, we generalize CSVM so that it can be used for tbe application of time series prediction. The experiment on the chaotic Mackey- Glass time series significantly verifies the performance of our algorithm. Data representation and similarity measurement are two basic aspects of similarity detection in time series data mining. In this paper, we present two novel approaches to perform similarity detection efficiently and effectively. One is composed of a new time series representation model and a corresponding similarity measure, which is called fragment alignment distance (FAD); the other applies dynamic time warping to the representation model of FAD and is called FAD_DTW. The new data representation model is based on the trend information of time series, which can provide a concise yet feature-rich representation of time series. FAD is able to align the segments of time series in linear time, which greatly accelerates the similarity detection process. We extensively compare FAD and FAD_DTW with state-of-the- art time series representation models and similarity measures in classification and clustering frameworks. Experimental results from efficiency and effectiveness validations on various data sets demonstrate that FAD and FAD_DTW can achieve fast and accurate similarity detection. In particular, FAD is much faster than the other methods.

3. PROPOSED SYSTEM

- This paper proposes a major information display for securities exchange forecast machine learning calculations.
- In rehearse; the machine will be prepared utilizing the profound learning technique.
- The information gathered from the online sources utilizing money markets programming interface. That information will be unstructured arrangement. That will clean utilizing the preprocess method.
- The cleaning information will be jumped into various parts and assembled into comparable data's. That every one of the information dealt with by the Hadoop system.
- The gathering information going to the calculation and foresee the element stock promoting report.

3.1 Advantages of Proposed System

- The put away information broke down utilizing mapreduce calculation and SVM calculation used to characterize and grouping process.
- The mining method to foresee the stock promoting status in view of utilizing the authentic information like value, low, high, open and close everything utilizing the verifiable information.

4. RESULT

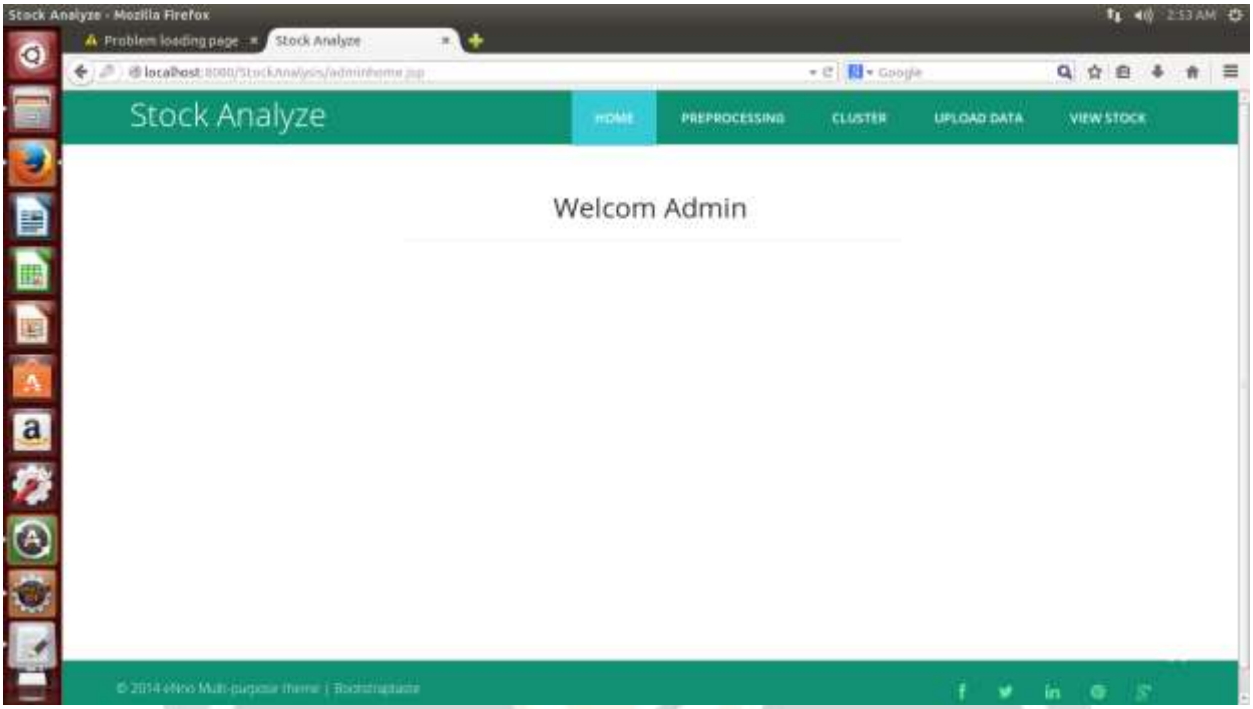

Fig.No.1 Screenshot of the Project

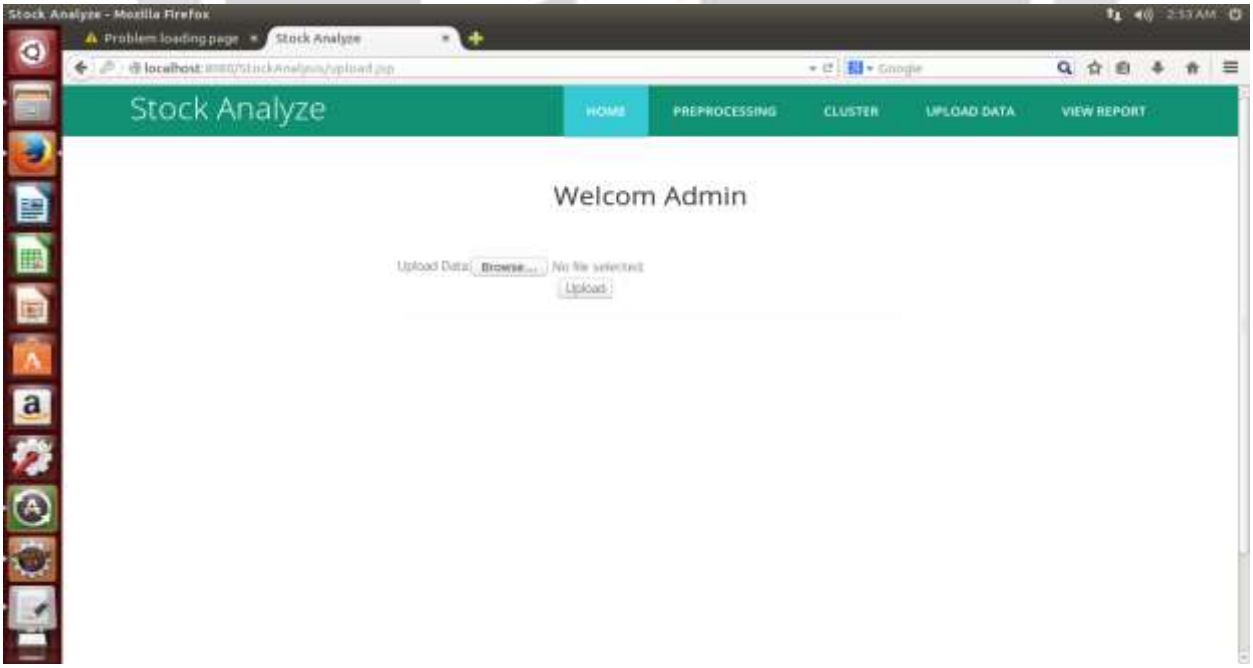

Fig.No.2 Screenshot of the Project

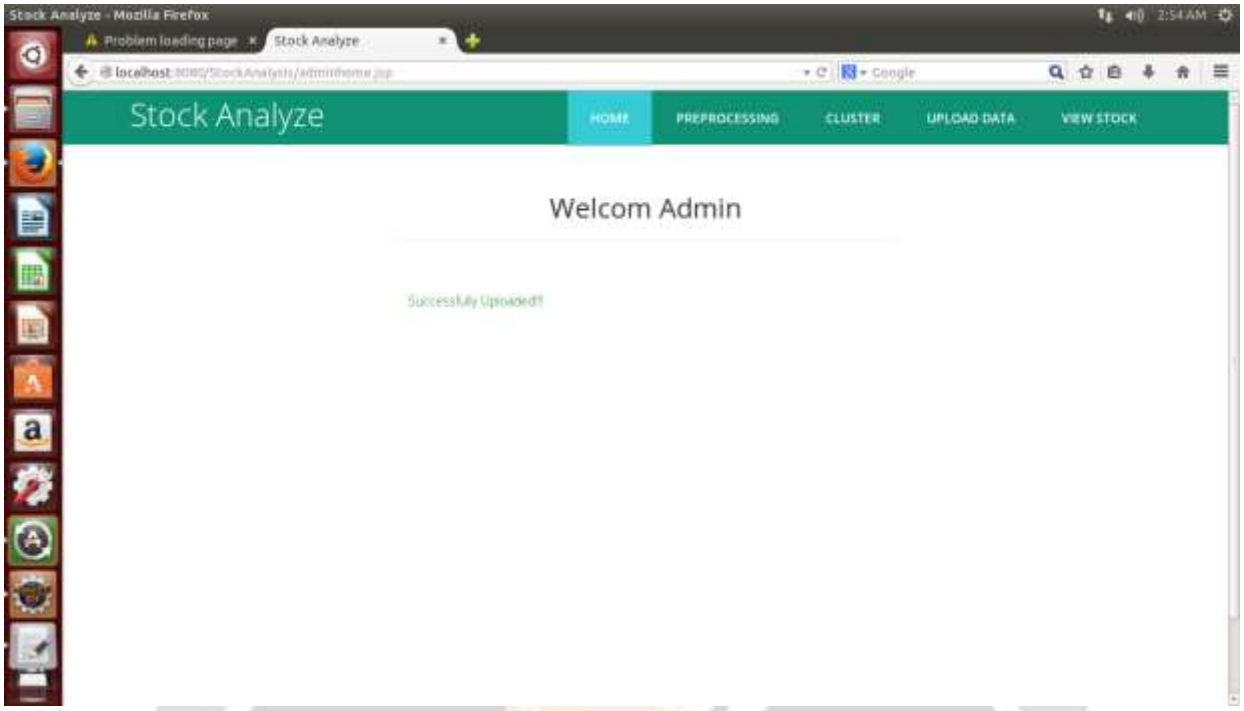

Fig.No.3 Screenshot of the Project

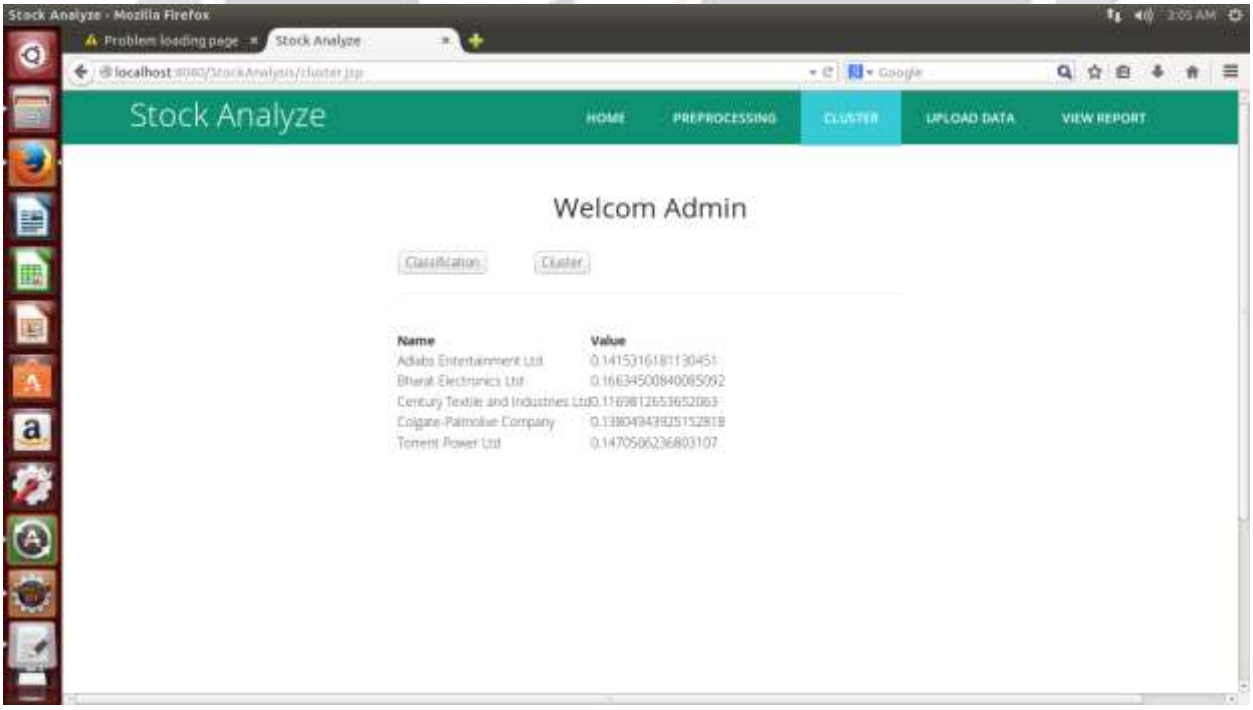

Fig.No.4 Screenshot of the Project

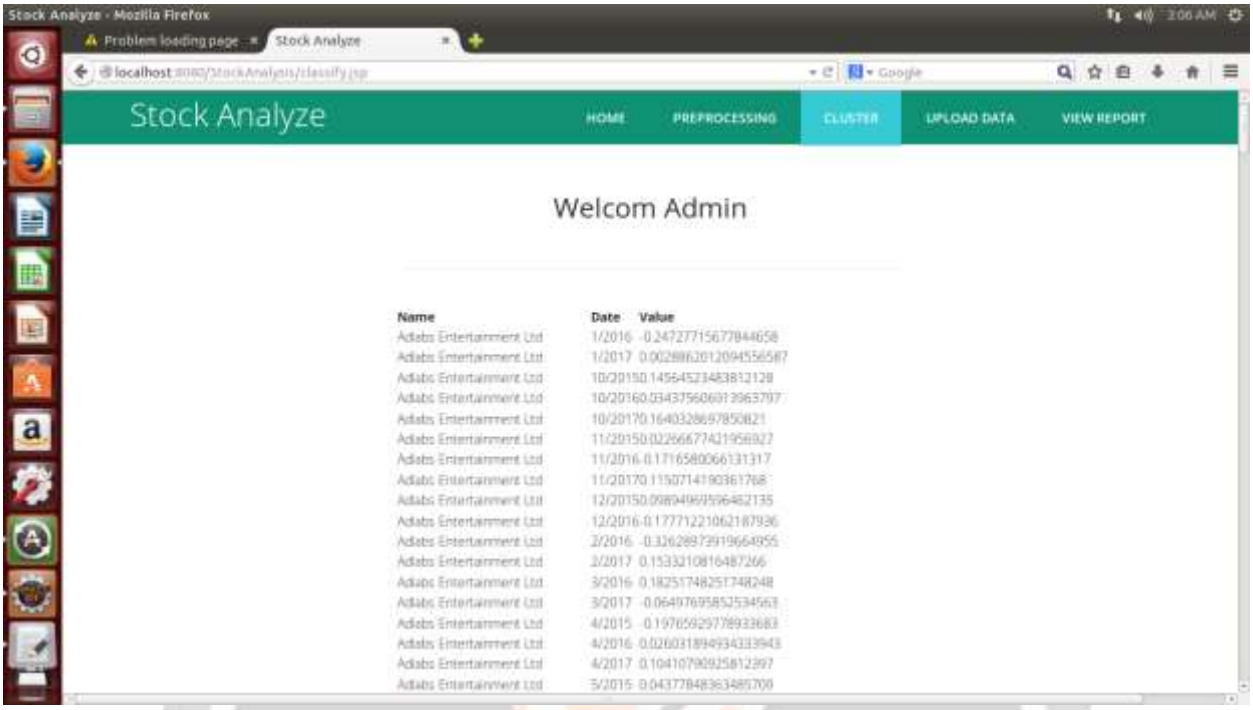

Fig.No.5 Screenshot of the Project

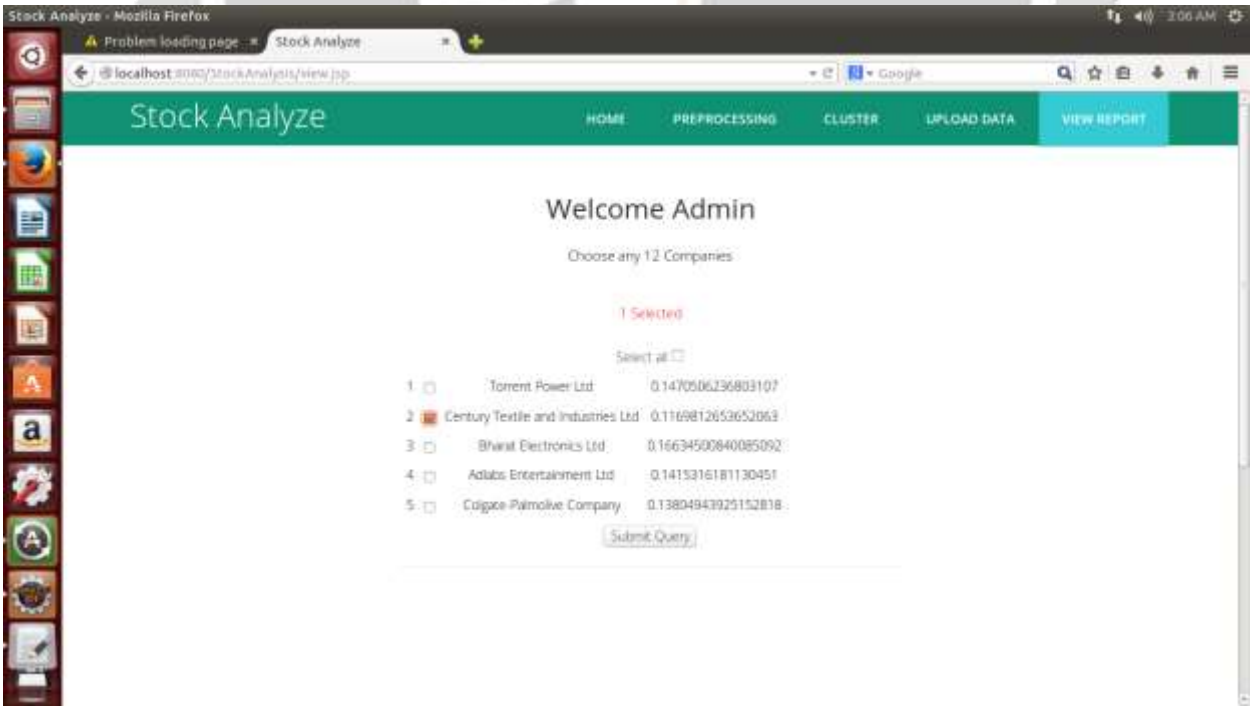

Fig.No.6 Screenshot of the Project

Fig.No.7 Screenshot of the Project

5. CONCLUSIONS

The stock advertising information is expanding day by day with the information will be produced in various associations. The information will gathered and stacked into the hdfs utilizing the Hadoop system. The put away information broke down utilizing mapreduce calculation and SVM calculation used to characterize and grouping process. The mining method to foresee the stock promoting status in view of utilizing the authentic information like value, low, high, open and close everything utilizing the verifiable information.

6. REFERENCES

[1] Khodabakhsh, A., Arí, I., Bakír, M. and Ercan, A.O., 2018. Multivariate Sensor Data Analysis for Oil Refineries and Multi-mode Identification of System Behavior in Real-time. IEEE Access, 6, pp.64389- 64405

[2] Raicharoen, T., Lursinsap, C., & Sanguanbhokai, P. (2003, May). Application of critical support vector machine to time series prediction. In Circuits and Systems, 2003. ISCAS'03. Proceedings of the 2003 International Symposium on (Vol. 5, pp. V-V). IEEE.

[3] Zhang, M. and Pi, D., 2017. A New Time Series Representation Model and Corresponding Similarity Measure for Fast and Accurate Similarity Detection. IEEE Access, 5, pp.24503-24519

[4] Pati, J., Kumar, B., Manjhi, D. and Shukla, K.K., 2017. A Comparison Among ARIMA, BP-NN, and MOGA-NN for Software Clone Evolution Prediction. IEEE Access, 5, pp.11841-11851

[5] Zhang, Q., Li, F., Long, F. and Ling, Q., 2018. Vehicle Emission Forecasting Based on Wavelet Transform and Long Short-Term Memory Network. IEEE Access, 6, pp.56984-56994.

[6] Conejo, Antonio J., Miguel A. Plazas, Rosa Espinola, and Ana B. Molina. "Day-ahead electricity price forecasting using the wavelet transform and ARIMA models." IEEE transactions on power systems 20, no. 2 (2005): 1035-1042.

[7] Anaghi, M.F. and Norouzi, Y., 2012, December. A model for stock price forecasting based on ARMA systems. In Advances in Computational Tools for Engineering Applications (ACTEA), 2012 2nd International Conference on (pp. 265-268). IEEE.