International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 02, 2020
ISSN: 1475-7192

Stock Market Prize Prediction Using Linear Regression and Spring XD Framework

[1]Mr.S.M Prabin ,[2] Dr.M.S Thanabal

ABSTRACT--*Huge data in continuous manner is required for prediction in stock market. This research work proposes a linear regression based prediction model for real time data analysis. The data pipelining technique is employed that provides the data to the regression model for prediction. This work focuses on the stock prediction from APPLE, AMAZON and GOOGLE sources and in future can be extended for multiple stocks. The spring XD framework was used for the pipeline of data and four different prediction models are employed in this work. The outcome of this work can be used for real time stock prize prediction.*

Key words*-- Data mining, prediction, stock prize, date pipelining, stock market*

I. INTRODUCTION

In today's scenario, market prize prediction is vital in all domains and stock market prediction is defined as the prediction of estimated stock value. The prediction using single data set is not beneficial and prediction from multiple sources is needed. The machine learning algorithms role is inevitable in all domains for the detailed analysis of data and bringing a fruitful outcome for a problem. A wide number of neural network and fuzzy logic algorithms are there for solving complex real time problems. The machine learning algorithms plays pivotal role in stock market prize prediction from the data. The outcome of the stock market prize prediction helps to invest shares in the appropriate companies that yield profit. The financial markets are highly unstable and the daily prizes have to be compared and prediction is needed to decide the future aspect.

A detailed study on different types of stock prediction techniques based on financial information has been discussed in [1]. The support vector regression was found to be efficient than support vector machine in stock market price prediction [2]. The hidden markov model was found to be robust in the stock market prize prediction and yields efficient results, when compared with the classical algorithms [3]. The radial basic function neural network was found to be proficient in the stock market price prediction, when compared with the back propagation neural network [4].The optimization algorithms plays vital role in machine learning techniques, PSO was utilized in the tuning of parameters in the least square SVM, when compared with the when compared with the least square SVM and BPN approaches [5] [6]. The wavelet based SVM was found to be efficient in the stock prize prediction, when compared with the classical approaches [7] [8]. The clustering techniques plays vital role in financial data analysis and novel algorithm based on decision tree were found to be efficient in the stock market price prediction [9] [10]. A novel type-2 fuzzy logic system was deployed for the stock market price prediction; the parameter of the fuzzy logic system was tuned by the genetic algorithm [11]. The neural and fuzzy logic

[1] *Assistant Professor /Computer science and Engineering,PSNA College of Engineering and Technology-Dindigul.*

[2] *Professor /Computer science and Engineering,PSNA College of Engineering and Technology-Dindigul.*

Received: 02 Dec 2019 | Revised: 21 Jan 2020 | Accepted: 05 Feb 2020

techniques yields significant results for stock market price prediction [12][13].The hybrid forecasting based on RST algorithm was used for the stock market prize prediction [14].Section 2 discusses the proposed methodology describing the proposed methodology for stock prize prediction and section 3 highlights the results and discussion and finally conclusion is drawn in section 4.

II. PROPOSED METHODOLOGY

In order to generate an accurate prediction model, it requires huge amount of historic and current data to be analysed in a continuous manner, it can be achieved through data pipeline. The scope of this work is to build both pipeline & prediction model. The building data pipeline itself is a complex task, Spring XD (eXtreme Data) is the current trend in building data pipeline. Sprint XD supports the pipeline jobs to be deployed in cluster environment, which ensures HA high availability of services. Once the pipeline jobs are built using Spring XD, it can be scheduled to run in batch mode or in streaming mode. In this work, the pipeline jobs will be running in batch mode to analyse the historic data. Later it can be scheduled to run in stream mode to analyse the real time data. The flexibility is the advantage of this project.In this work, pipeline process the stock and pushes in to KAFKA. The data from KAFKA can be easily moved to BIG DATA(Hadoop) , also KAFKA offers distributed messaging and high availability,that ensures there won't be any data loss.

The Prediction model was built using Linear regression. Basically, the model will analyse the stock date and stock price and build the regression line. This was done through supervised learning techniques so that an accurate prediction model was constructed. Whenever it requires we can train the model with more data. The Java OOP supports reusability of existing code. Since we are using JAVA for building the Linear regression model, it can be easily extended to use multiple Linear regression or different machine learning algorithms.

A reliable data pipeline was built to bring all the stock data into kafka distributed storage, since it is fault tolerant and then finally load the data into the linear regression model. The Spring XD was used to build the data pipeline which is user interface and it provides all necessary details to deploy the job. The four different models that will predict 4 different graphs for comparison which includes the stock data parameters like close price,average price,high price and low price are proposed in this work.There is a user interface using Spring XD local host records and shows all the details of the deployed jobs and pipeline modules. It shows the time and date when a pipeline is being executed and launched. It also helps us to shutdown our module. It records the execution count of pipeline jobs as well as the status of it whether it is completed or not, it provides options to deploy or un-deploy or destroy.

2.1 Stock Market Price Prediction System Design

The system design was done using UML(Unified *ModellingLanguage*). The use case diagram is a graphical representation of the components of a system. It is a technique used in system to identify, clarify, and organize system requirements. The use case diagram of the proposed system is depicted in figure 1. The use case diagram of the end user is depicted in figure 2. The activity diagrams are similar to flow charts and depict the operations of the system. The activity diagrams for the proposed stock market prediction system are as follows.

Received: 02 Dec 2019 | Revised: 21 Jan 2020 | Accepted: 05 Feb 2020

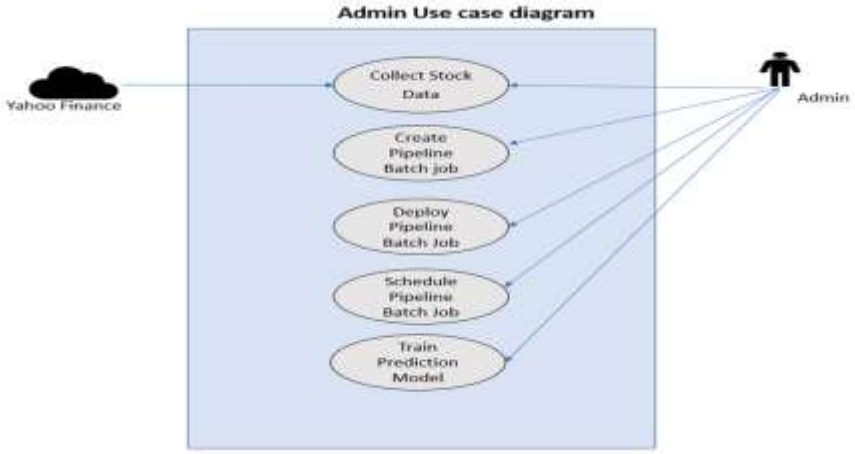


Figure 1: Use case diagram of admin for the stock prediction model

Figure 2: Use case diagram of end user for the stock prediction model

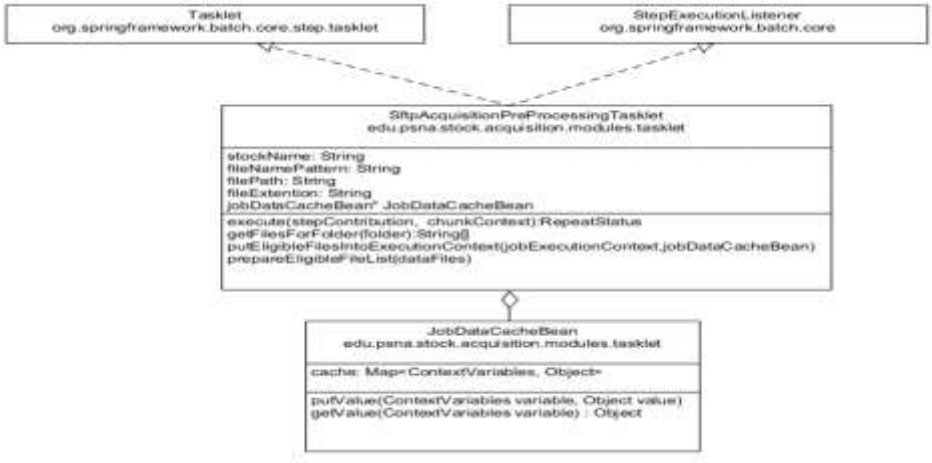


Figure 3: Class diagram for pipeline pre-processing module

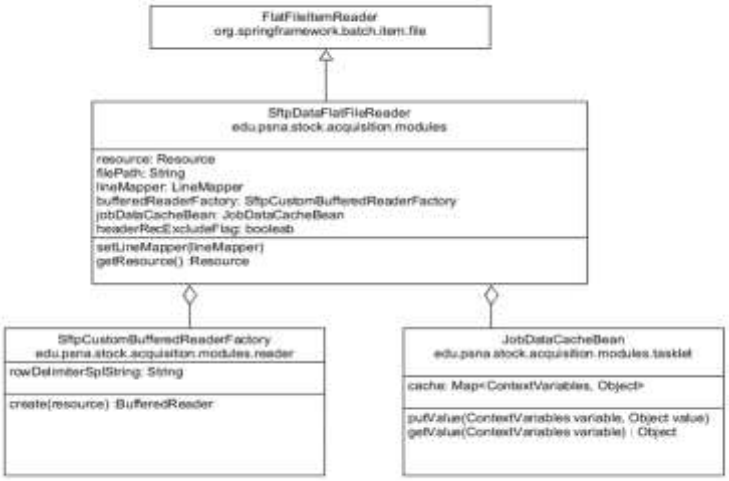


Figure 4: Class diagram for Pipeline Reader Module

The class diagram in UML describes the system sub classes and its features. In the class diagram, the top compartment comprises of the class details and the middle compartment comprises of the attributes of the class. The class diagram of pipeline preprocessing module, reader module and writer module are depicted in figure 3, 4 and 5.

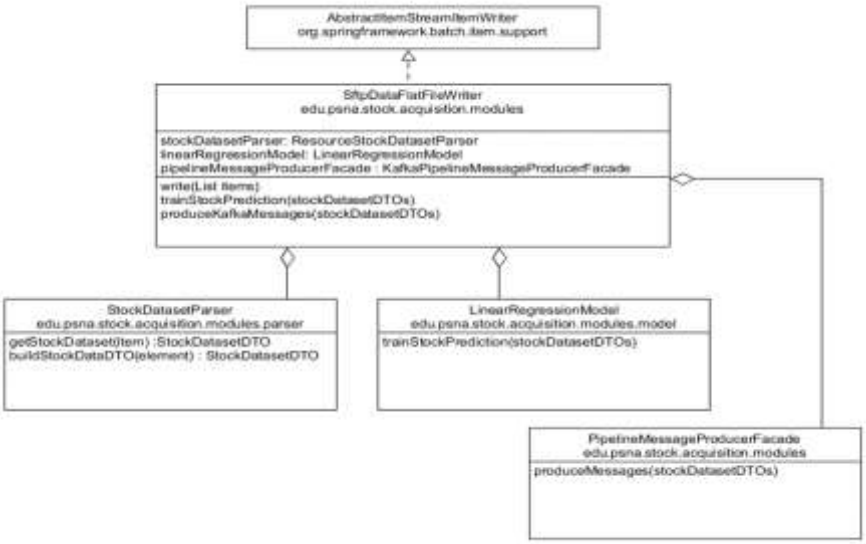


Figure 5: Class diagram for pipeline writer module

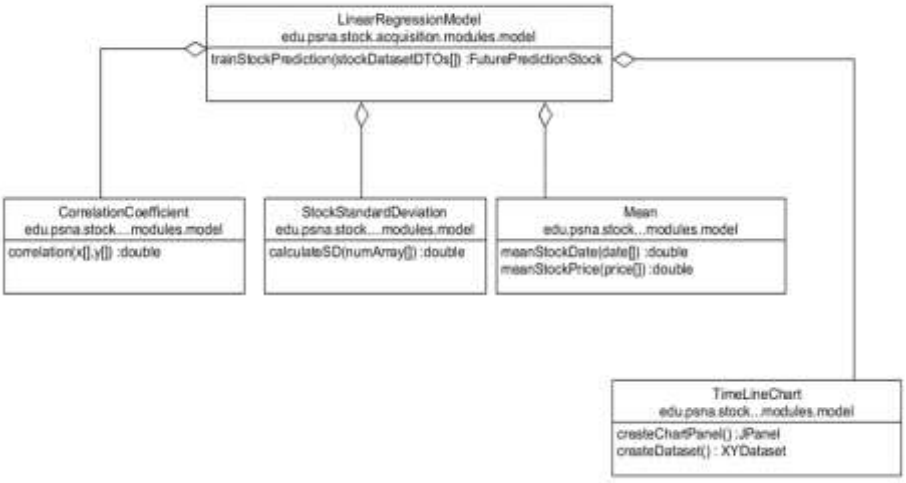


Figure 6: Class diagram for linear regression module

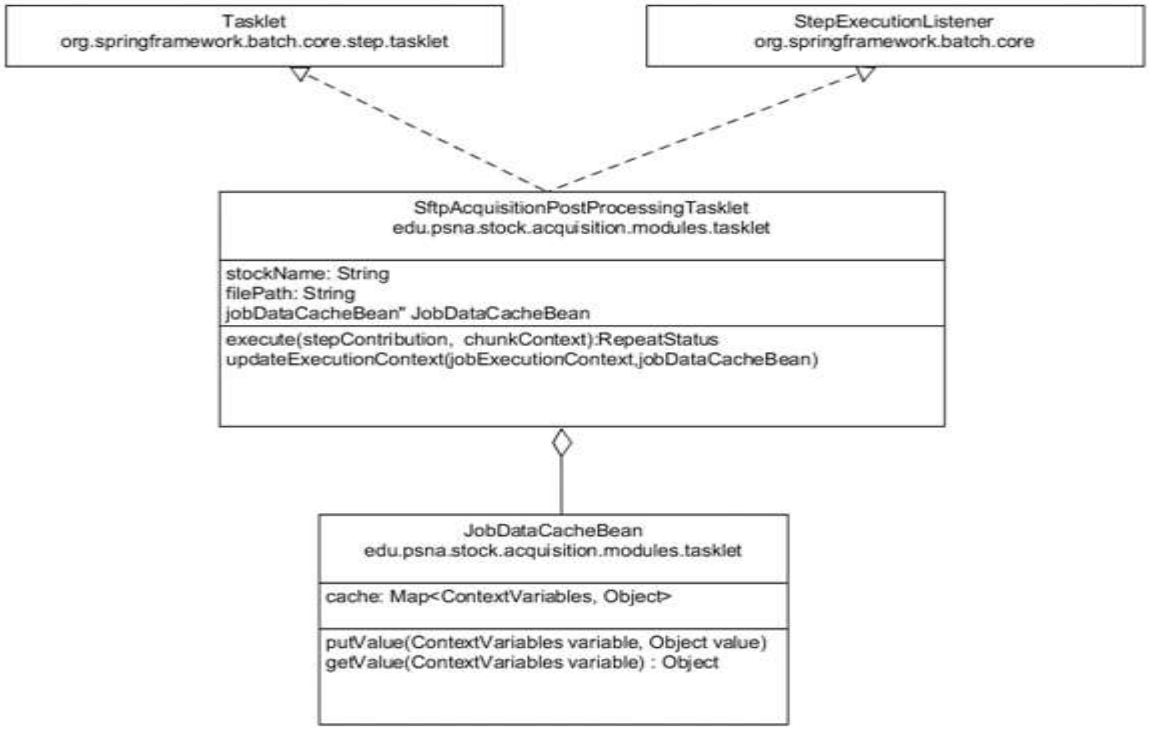
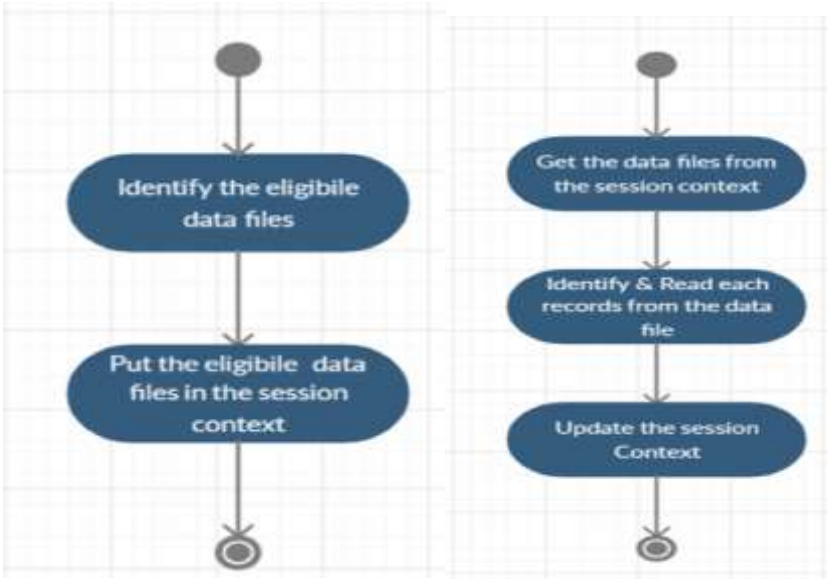
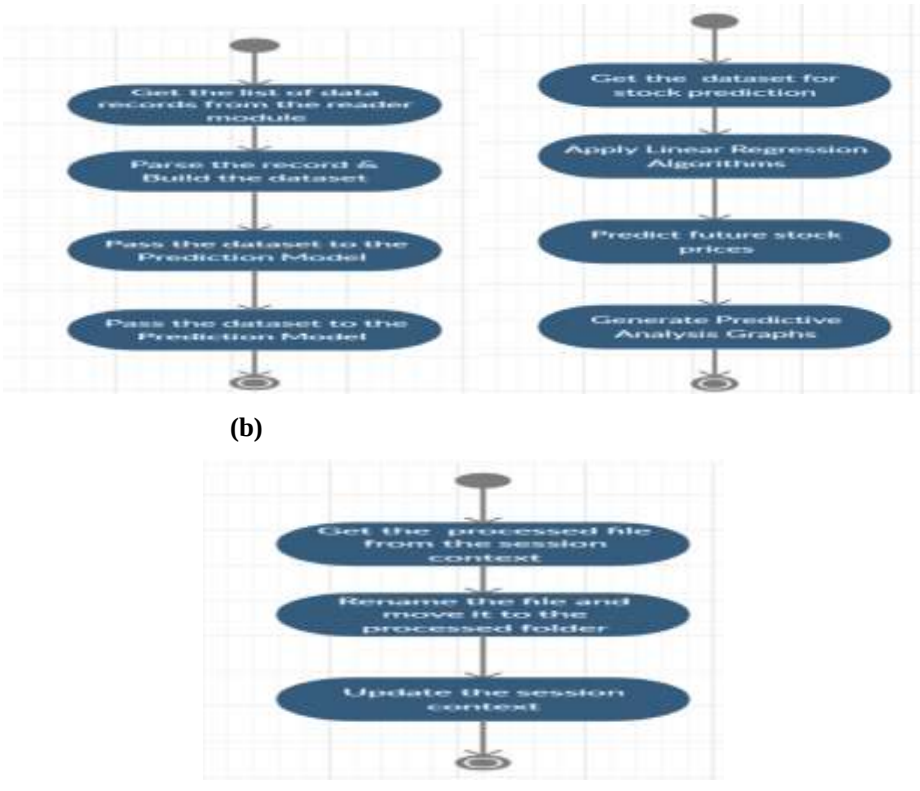


Figure 7: Class diagram for pipeline post processing module



(a) **(b)**

Figure 8: Activity diagram for pipeline pre-processing module and pipeline reader module



(a) **(b)**

(c)

Figure 9: Activity diagram for pipeline writer module, linear regression model and post processing module

III. RESULTS AND DISCUSSION

The pipeline was build using spring-xd to bring the historical stock data that is taken from yahoo. The stock data's will be stored in the .csv file extension.The spring XD framework was used to build out pipeline and it works 24x7 without much human intervention. A dedicated pipeline batch job was created for each STOCK and the job will be scheduled to run on every day. The batch pipeline job ingests the data into Kafka & the Data Model.The codes for the entire module will be stored in a .jar file. The jar file was uploaded as the custom module in spring-xd.As the pipeline jobs gets deployed and launched, four models will predict four different graphs for comparison to find the future closing stock price of the stock. Each includes parameters stock price in y axis and date in the x axis which is default. The graphs and the prediction of future stock prices for the stock data will be stored in a separate file.

At end of every day, the stock data for different companies (eg: Apple, Amazon, Google etc) will be downloaded at C:\stockData. The scheduled pipeline job extract the right data file from C:\stockData and it transforms the dataset and load into Kafka & Data models.The Data model does the prediction using Linear regression, it generates the prediction graph at C:\stockPrediction\graph. There are 4 prediction models using various parameters of the stock data; prediction using stock close price , prediction using stock High price, prediction using stock Low price and prediction using stock average Price.

The model will convert the entire date format into number format for prediction.The Data model predicts the future stock price for another 7 days. The comparison between different prediction models will be saved at **C:\stockPrediction\prediction.** The blue line indicates the future close price and the red line indicates the historical stock prices stored in stock data.As soon as the pipeline job is launched, four different models will be predicting and displaying four different graphs for comparison to find the future closing price of the stock. Each includes parameters stock price in y axis and date in the x axis which is default.The blue line indicates the future close price and the red line indicates the historical stock prices stored in stock data.

Figure 10: AAPL stock prediction using close price

Figure 11: AAPL stock prediction using average price

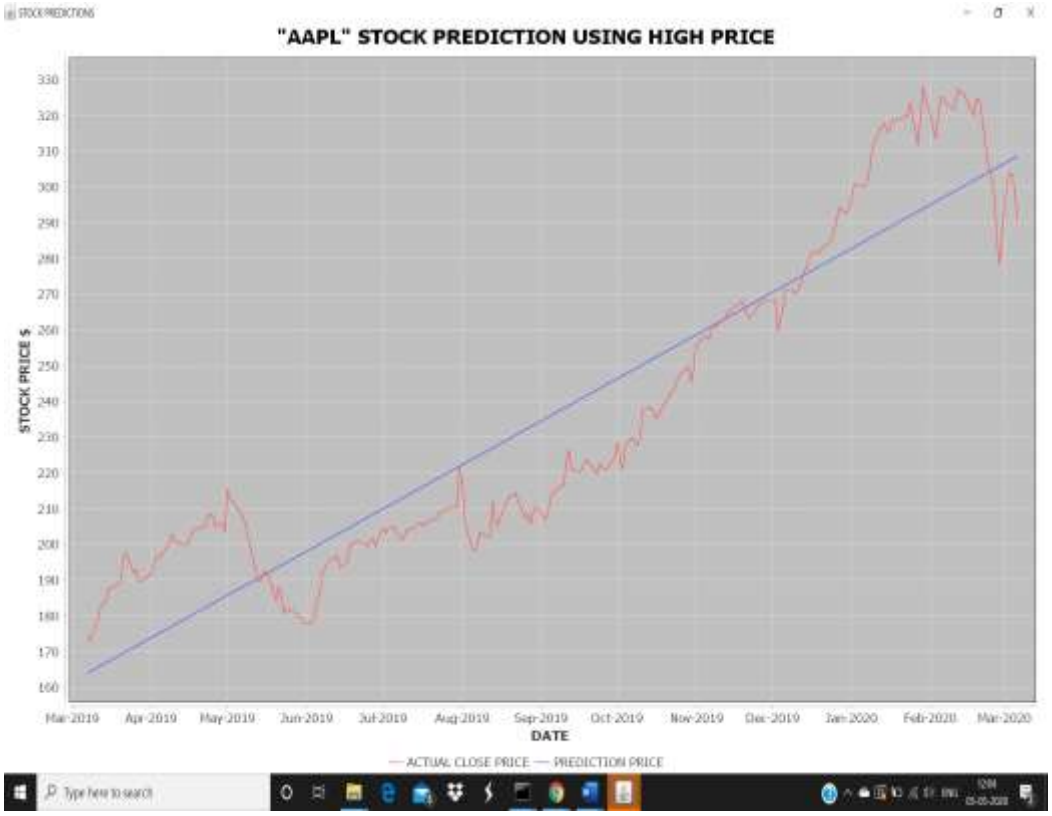


Figure 12: AAPL stock prediction using high price

Figure 13: AAPL stock prediction using low price

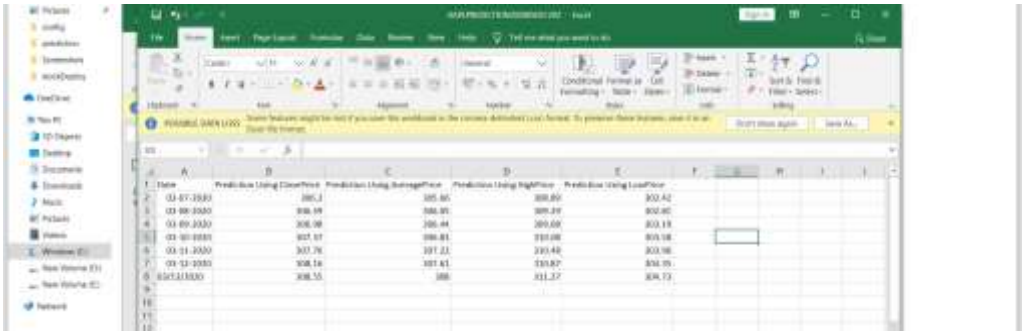


Figure 14: Processed AAPL file

It is uploaded as a jar file in the stock deploy. It is named as stock prediction.The Stock prices stored in a csv format is been created and deployed using xd server and xdshell.xd shell is used to communicate with the server.So now the file will be processed and deployed in the processed file with the date followed by it.

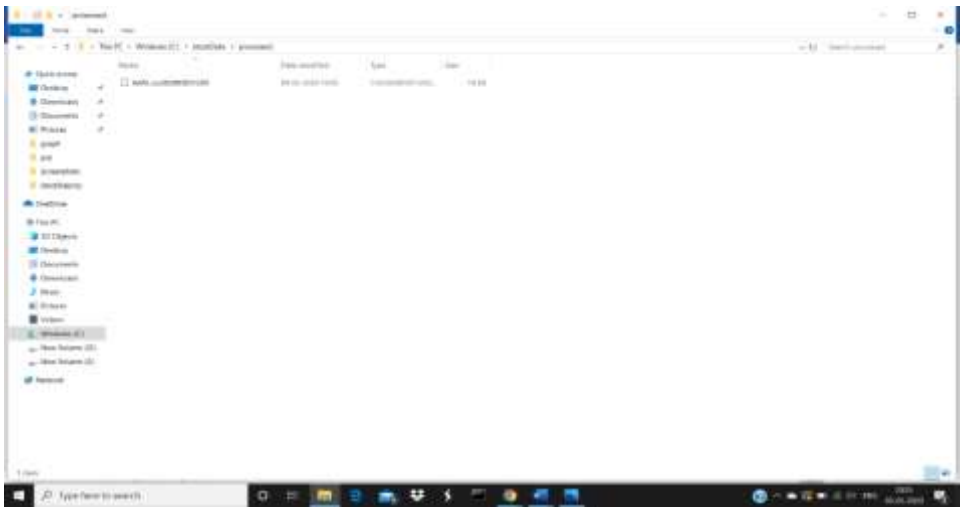


Figure 15: Processed AAPL job

The server will be recording and displaying all the details of the model once the model is launched. It calculates and shows standard deviation and mean of stock price, correlation coefficient, slope, intercept and all predicted dates and respective predictive prices.

IV. CONCLUSION

This research work proposessupervised Learning using Linear Regression Algorithms to build the Stock Prediction model. The algorithms are developed using Java Spring and to achieve the goal of distributed processing & load balancing cloud environment is used. Spring XD is used to build the data pipeline and it is an open source project, cost free, unified, distributed, and extensible service for data ingestion, real time analytics, batch processing, and data export. The data pipeline and Stock predictions module can be deployed in single server environment and as well as cloud cluster environment.

REFERENCES

1. Schumaker RP, Chen H. A discrete stock price prediction engine based on financial news. Computer. 2010 Jan 26;43(1):51-6.
2. Fulks R. Applying Our Measurement Expertise to Investing [Recalibration]. IEEE instrumentation& measurement magazine. 2010 Dec 17;13(6):40-4.
3. Economopoulos AA, Kouikoglou VS, Grigoroudis E. The base stock/base backlog control policy for a make-to-stock system with impatient customers. IEEE transactions on automation science and engineering. 2010 Jul 19;8(1):243-9.
4. Zhang L, Liu N, Yu P. A novel instantaneous frequency algorithm and its application in stock index movement prediction. IEEE Journal of Selected Topics in Signal Processing. 2012 May 11;6(4):311-8.
5. Chun C, Qinghua M, Shuqiang L.: *"Research on Support Vector Regression in the Stock Market Forecasting"* ©Springer, Advances in Intelligent and Soft Computing Volume 148, , pp 607-612, 2012.
6. Guo Z., Wang H., Liu Q. : "Financial time series forecasting using LPP and SVM optimized by PSO" © Springer, Soft Computing Methodologies and Applications , December 2012.
7. Huang C.; Huang L.; Han T.: *"Financial time series forecasting based on wavelet kernel support vector machine"* IEEE, Eighth International Conference on Natural Computation (ICNC), 2012.
8. Hegazy O, Soliman OS, Salam MA. A machine learning model for stock market prediction.arXiv preprint arXiv:1402.7351. 2014 Feb 28.
9. Kou G, Peng Y, Wang G. Evaluation of clustering algorithms for financial risk analysis using MCDM methods. Information Sciences. 2014 Aug 10;275:1-2.
10. Wu DD, Zheng L, Olson DL. A decision support approach for online stock forum sentiment analysis. IEEE transactions on systems, man, and cybernetics: systems. 2014 Jan 9;44(8):1077-87.
11. Zarandi MF, Rezaee B, Turksen IB, Neshat E. A type-2 fuzzy rule-based expert system model for stock price analysis.Expert Systems with Applications. 2009 Jan 1;36(1):139-54.

12. Yu TH, Huarng KH. A bivariate fuzzy time series model to forecast the TAIEX. Expert Systems with Applications. 2008 May 1;34(4):2945-52.
13. M.S Thanabal, L. Ganesan " A Contrary Recurrent Item Set Hierarchy in PPDDM" National Academy science Letters,Springer vol39 pages283-286.
14. Patel J, Shah S, Thakkar P, Kotecha K. Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. Expert systems with applications. 2015 Jan 1;42(1):259-68.
15. Ching-Hsue Cheng, Tai-Liang Chen, Liang-Ying Wei, " A hybrid model based on rough set theory and genetic algorithms for stock price forecasting", Pages. 1610-1629, 2010.