Product Analysis of a Reail Store Using Machine Learning Algorithms -Data Science Consideration.

[1]M. Kumaran, [2]G. Sundararaju, [3]J. Lin Eby Chandra, [4].V.Seedha Devi,
[5]S. Rajendran

1, 2, 3. Assistant Professor, Department of Computer Science and Engineering, Jaya Engineering College
Professor, Department of Mechanical Engineering, Jaya Engineering College Chennai 602024, INDIA.
4. Associate Professor Department of Computer Science Engineering, Jaya Engineering College
5. Professor, Department of Mechanical Engineering, Jaya Engineering College Chennai 602024, INDIA.
Email: kumaran.ma@gmail.com

Abstract
Many household products are sold by various subsidiaries of the retail store network which are geographically located at various locations. Supply chain inefficiencies will occur at different locations when the market potential will not evaluated by the retailers. Many times it is not easy for the retailers to understand the market condition at various geographical locations. The organization of retail store network has to understand the market conditions to intensify its goods to be bought and sold so that many number of customers get attracted in that direction. Business forecast helps retailers to visualize the big picture by forecasting the sales we get a general idea of coming years if any changes are needed then those changes are done in the retail store's objective so that success is achieved more profitably .In this paper we make an attempt by understanding the retail store business's driving factors by analyzing the sales data of Walmart store that is geographically located at various locations and the forecast of sales for coming 22 weeks is done. By sales forecasting the retail networks are supported so that the resources can be managed efficiently.

Key words: Business forcast, matplotlib, K-Neighbours, Random Forest

INTRODUCTION:

We all are very curious about our future! very excited to know what will happen with all of us the very next moment, tomorrow ,similarly the retailers also curious about their business, its progress, curtailing factors of it. By adopting certain steps the aspects that can cause damage or reduce the profit can be avoided. In this goal of forecasting the sales business the data from the various sectors is collected and the data analytics is done the efficient understanding of the observed data by common steps is not practically possible because the data is very huge the masses of data of organization is molded in such a way that its having meaning by

understanding deeply the suitable actions can be taken. Let us consider the retail stores network example Wallmart it is an example for giant stores etc.

The retail store sells the household products and obtains profit by that. There are different subsidiaries of the retail store network whose locations are variously located at various geographical locations most of the time retailers will not be successful in understanding the customer's needs because they will be able in the evaluation of market potential at that location, during special occasions the rate of sales or shopping is more sometimes this may cause inefficiency of the products ,the relationship between the customers and the stores is analyzed and the changes that need to obtain more profit is done.

In this paper we forecast the sales by using three modules they are hive, python programming and mathplotlib, storage in. MongoDB is provided by hive that provides interface to the data stored in HDP hive used for data processing and understanding in hive partitioning and bucketing is done. In Python programming statistical is done in Mathplotlib the interactive visualization of the data on the product that are focused on the business.

BACKGROUNDS:

Fundamentals:

In this implementation the main focus is on the analytics part that is been visualized by matplotlib and the machine learning techniques. The mongo db is mainly used for database storage and it has live interaction with the customers also it has the fluency of performing tasks using machine learning algorithms like random forest , Kneighbours, Skit Learn, Statistical Computing and various cloud computing algorithms can als be implemented and this is mainly been may for wallmart sales prediction for a particular product for 70 years by obtaining 22 graphs using the native libraries of python programming like data visualization libraries.

IMPLEMENTATION OF SALES FORCASTING:

### Sales Forecasting System Architecture

Architecture is the most important aspect of any system represents the system architecture.

The Forecasting Process:

The process of forecasting is a group of methods to predict the sales. It is initiated after determining the objective. It may include the sales amount in dollars, the number of employees to be appointed .In Fig.2 selection of dependent and independent variables are done. The forecasting results like sales data or the number of employees to be appointed in the upcoming year. The market factor include the factors like products existence in the store, its quality and the demand of the item .Market index is a market factor that is expressed as the amount of percentage relatively with some base content. When the market index is increased then the industry sales is increased. The index consists of many market factors like price, population of the area, personal income that is disposable. Then in the forecasting process the procedures of forecast and methods that are useful for data analysis are

determined .If the procedures were not used in prior then the firm may want to test the procedures. Then gathering and analyzing of data is done. Certain assumptions are made about the forecasted sales. Then the sales forecast is finalized as the time passes and the results are evaluated.
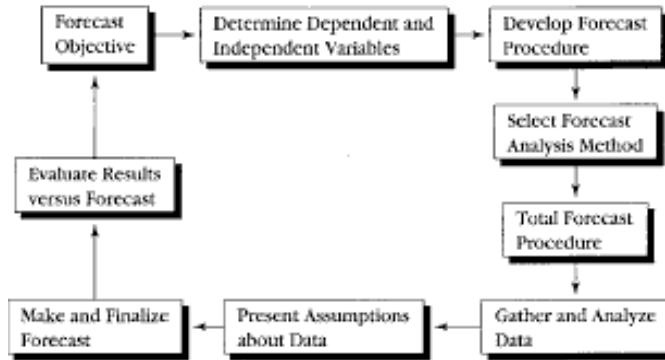
SYSTEM ARCHITECTURE
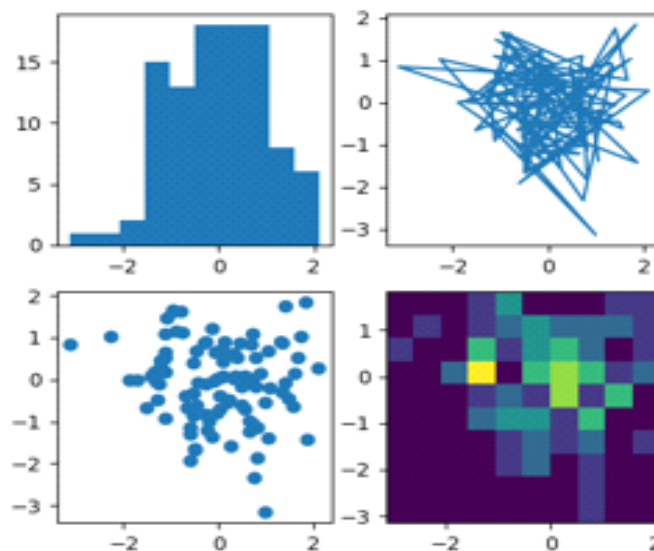
MATPLOTLIB ARCHITECTURE

METHODOLOGY:

Strategic Research:

The data after processed by map reduced though is in understandable format, but is too large hence it is difficult to draw conclusions out of that. Processing using hive is done by loading the data sets supplied then dynamic partition and bucketing is applied. The monthly sales of 45 stores and 99 departments are calculated. The average feature across them is calculated.

Machine Learning:

Machine learning algorithm which provides efficient results are applied so that huge data can be modeled R is a programming language and software environment for statistic computing and graphs. The R programming [12] is used by the statisticians and data miners widely to develop the statistical software and data analysis. The Holt winters [16] algorithm is used to predict the sales .The seasonality, trend and randomness is observed in the algorithm .The algorithm is used for train data sets and then the sales prediction is done.

Graphical Representation:

The processed data by the machine learning algorithm which provides efficient results can be understood but this obtained information is too large and hence conclusions are difficult. We all know the picture representation is worth than many words.

RESULTS AND DISCUSSION:

Results and discussions are very important so that readers can understand the information of the paper, results are the descriptive content information .Below are the few results obtained by observing these results the sales is forecasted for the next year. represents the different combinations of the graphs. represents the sales and stores, x axis represents the stores and y axis represents the weekly sales. b the store, department and sales it's a bubble chart, weekly sales and the department, x axis represents the weekly sales and y axis represents the departments, suppose if we want the sales of department in store 1 then we can select the required department and the sales is represented. represents the weekly sales and the year, the weekly sales according to year is represented ,x axis represents the year and y axis represents the weekly sales. represents the sales and the stores x axis represents the different number of the stores and the y axis represents the sales in each department. represents the sales in stores by using Matplotlib, it is the bubble chart and are the graphs of the data collected from 3 years of walmart store, now we have to predict the sales for the next 22 weeks and should be graphically represented.
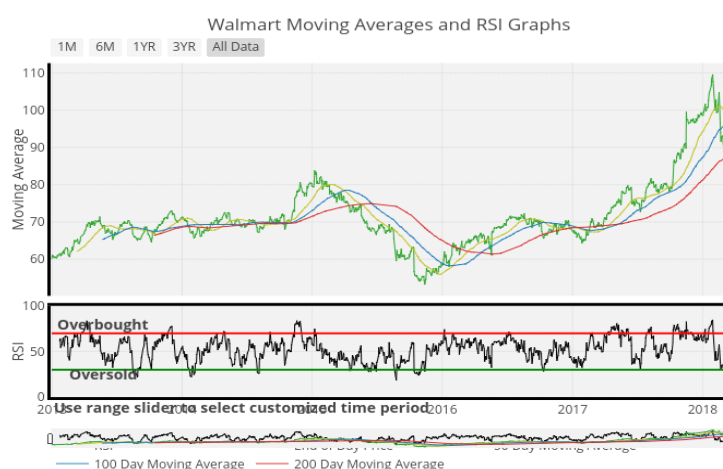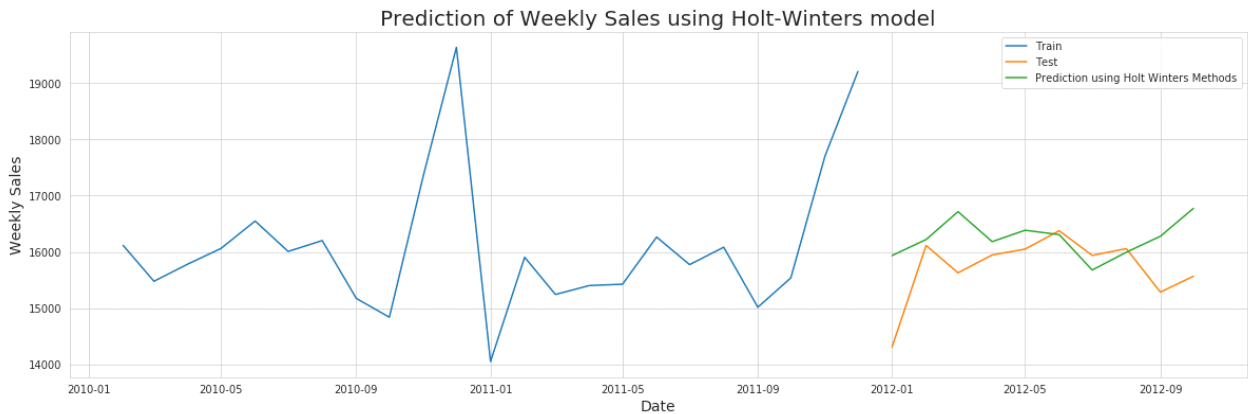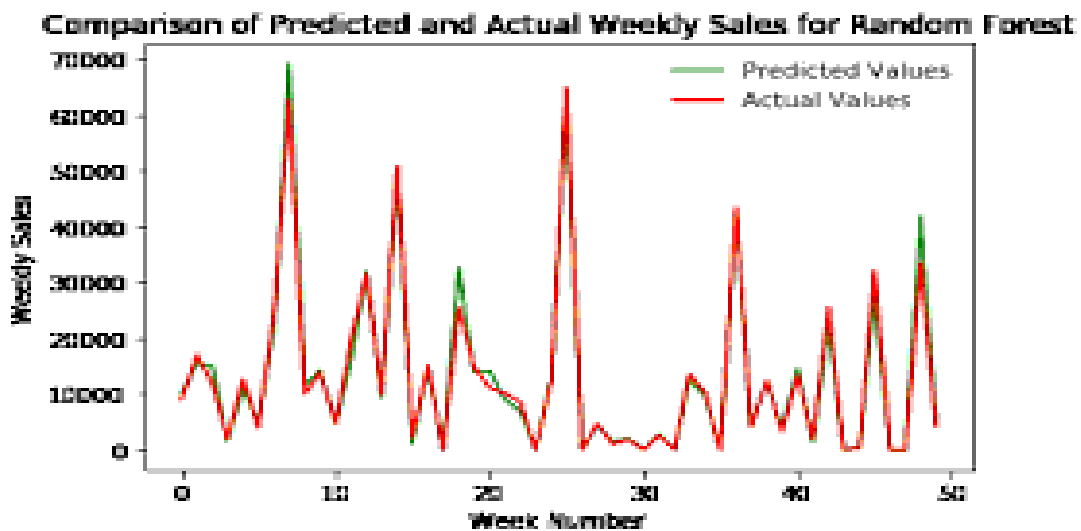

AVERAGE DAY MOVING OF PRODUCTS
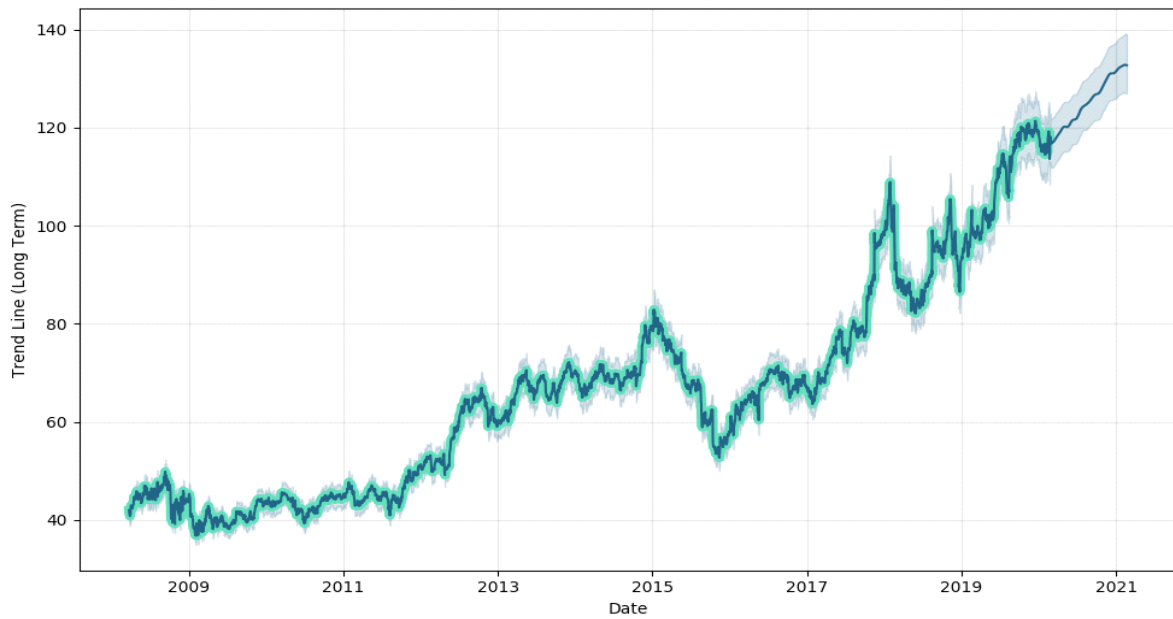
SALES PREDICTION OF PRODUCTS
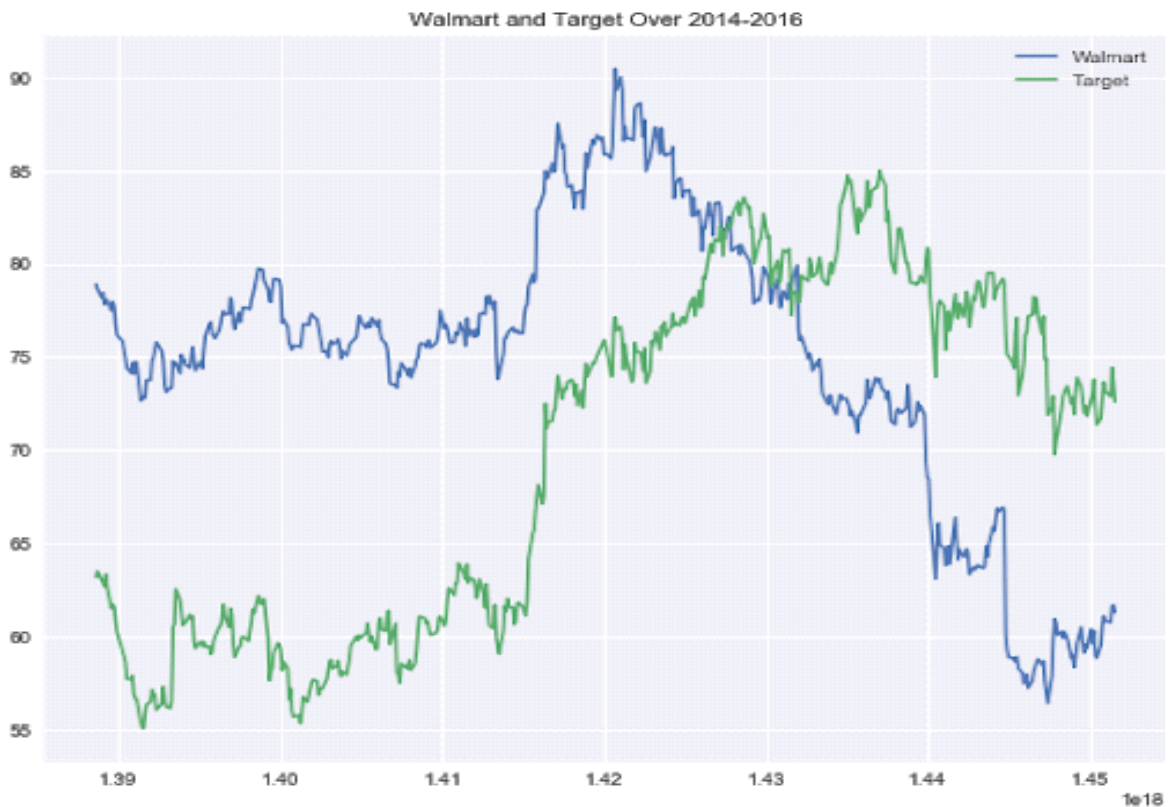
K-NEIGHOBOURS ALGORITHM

NUMPY ANALYTICS

RANDOM FOREST ALGORITHM

CONCLUSION:

Most of the shopping malls / shopping centers plan to attract the customers to the store and make profit to the maximum extent by them. Once the customers enter the stores they are attracted then definitely they shop more by the special offers and obtain the desired items which are available in the favorable cost and satisfy them. If the products as per the needs of the customers then it can make maximum profit the retailers can also make the changes in the operations, objectives of the store that cause loss and efficient methods can be applied to gain more profit by observing the history of data the existing stores a clear idea of sales can be known like seasonality trend and randomness. The advantage of forecasting is to know the number of employees should be appointed to meet the production level. Sales drop is bad thing forecasting sales helps to analyze it and it can overcome through the sales drop to remain in the competition forecast plays a vital role.

REFERENCES

1. Food and agro-product quality evaluation based on spectroscopy and deep learning: A review, X Zhang, J Yang, T Lin, Y Ying - Trends in Food Science & Technology, 2021 – Elsevier

2. Product decision-making information systems, real-time sensor networks, and artificial intelligence-driven big data analytics in sustainable Industry 4.0
   A Novak, D Bennett, T Kliestik - Economics, Management and Financial …, 2021

3. Sentiment analysis on product reviews based on weighted word embeddings and deep neural networks
   A Onan - Concurrency and Computation: Practice 2021

4. Predicting customer value per product: From RFM to RFM/P
   R Heldt, CS Silveira, FB Luce - Journal of Business Research, 2021 - Elsevier

5. Intelligent machine vision model for defective product inspection based on machine learning
   T Benbarrad, M Salhaoui, SB Kenitar… - Journal of Sensor and …, 2021

6. Data Analysis in Context-Based Statistical Modeling in Predictive Analytics
   C Selvan, SR Balasundaram - Handbook of Research on …, 2021

WEB REFERENCES :

1. https://numpy.org/ (Numpy documentation for analytics).

2. https://www.scipy.org/ (Scipy documentation for analytics).

3. https://machinelearningmastery.com/linear-regression-for-machine-learning/

4. https://towardsdatascience.com/understanding-random-forest-58381e0602d2

5. https://towardsdatascience.com/machine-learning/home

6. http://scikit-learn.github.io/stable

7. https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761

8. https://developers.google.com/machine-learning/clustering/clustering-algorithms.