ANALYSIS AND FORECASTING SALES IN SUPER MARKET USING MACHINE LEARNING

Mr. K. TULASI KRISHNA KUMAR [*1], Ms. VELUDOOTI RUPA [*2]

[*1] Associate Professor, Department of Computer Science and Application,
Sanketika Vidhya Parishad Engineering College, P.M. Palem,
Visakhapatnam, Andhra Pradesh

[*2] MCA Student, Department of Computer Science and Application,
Sanketika Vidhya Parishad Engineering College, P.M. Palem,
Visakhapatnam, Andhra Pradesh

ABSTRACT

Sales forecasting is critical for businesses involved in retailing, shipping, manufacturing, marketing, and wholesaling. It enables businesses to deploy resources more efficiently, estimate sales revenue, and develop better strategies for the future. In this research, predicting product sales from a certain retailer is done in a method that outperforms any machine learning technique. This project makes use of Big Mart Sales data from 2013.Nowadays, shopping malls and supermarkets keep track of specific item sales data in order to forecast client demand in the future. It holds a vast amount of consumer information as well as item attributes. Furthermore, regular patterns are discovered through mining the data warehouse. The data can then be utilised to forecast future sales for organisations like Big Mart using various machine learning approaches (algorithms). In this research, we propose a model based on the Xgboost algorithm for predicting sales of enterprises such as Big Mart and show that it outperforms other existing models. This project compares this model against other models in terms of their performance measures. Big Mart is a low-cost internet marketplace where users may purchase, sell, or advertise their goods. The purpose of this article is to make Big Mart a shopping paradise for buyers as well as a marketing solution for merchants. The ultimate goal is complete customer pleasure. The project SUPERMARKET SALES PREDICTION creates a predictive model that determines the sales of each product at a specific store. This model is used by the Big Mart to underpin the properties of the products, which plays a significant part in raising sales. This can also be done on the basis of the hypothesis, which should be done prior to looking at the data.

KEY WORDS:

Big Mart, Supermarket Sales Prediction, Forecast, Wholesaling, Data Warehouse.

1. INTRODUCTION

Forecasting has traditionally relied on personnel's experience-based knowledge. However, as grocery shops become larger and contain a greater variety of products with varying features, knowledge-based forecasting becomes a more difficult task. It is possible to use data for forecasting with greater data collection capability. Statistical models are frequently used to calculate how sales have behaved in the past, and these data are then utilised in conjunction with experience to forecast future sales. With improved processing capacity, it may be possible to apply complex machine learning models and depend more heavily on data for projecting sales.

The essential ideas of sellers and buyers are supply and demand. The ability to effectively predict demand is crucial for organisations in order to establish plans. Big Mart is a low-cost online marketplace where we can buy, sell, or advertise your merchandise. Big Mart's main purpose is to become a shopping paradise for buyers and a marketing solution for vendors. The ultimate goal is to grow in number of customers. The project SUPERMARKET SALES PREDICTION seeks to create a predictive model and determine the sales of each product at a certain shop. This model can be used by the Big Mart to comprehend the qualities of the products, which is important in increasing the sales. This can also be done on the basis of the hypothesis, which should be done prior to looking at the data. In this study, we present a predictive model employing the XG Boost technique for estimating the sales of companies such as Big Mart, and we discovered that the model outperforms existing models. The main goal of this machine learning project is to create a predictive model as well as to find sales of each product at a specific retailer. The Supermarket sales forecast uses this machine learning model to try to identify the qualities of the products and stores, which play a vital role in improving product sales.

Using this model, we will attempt to comprehend the qualities of the products and retailers that play a significant influence in growing sales. Python is utilised as the programming language here, and Jupyter Notebook is used as the tool. Machine learning features such as Supervised Learning task and Regression task are used to construct this application. This is mostly done to forecast future sales of a company's stores. The following processes are used: data exploration and pre-processing, feature engineering, model creation, and evaluation. Supervised learning assists you in understanding the flow of data and knowing the sale prices, among other things. It also comprises data visualization, cleansing, and transformation tasks. Linear Regression, Multiple Linear Regression, Decision Tree Regression, XG Boost Regression, and Random Forest Regression are some of the algorithms employed. In this research, we created a predictive model using

the XG Boost approach and tested it on the Big Mart 2013 dataset to anticipate product sales from a certain outlet.

2.   LITERATURE SURVEY

This section will deal with all the previous information related to brain tumor and several methods for sales forecast prediction. Literature survey is the most important step in software development process. For any software or application development, this step plays a very crucial role by determining the several factors like time, money, effort, lines of code and company strength. Once all these several factors are satisfied, then we need to determine which operating system and language used for developing the application. Once the programmers start building the application, they will first observe what are the pre-defined inventions that are done on same concept and then they will try to design the task in some innovated manner.

MOTIVATION

1) Forecasting Supermarket Sales

Authors: Manjuprasad Shetty and Vishwasa Nawada

Most businesses like to forecast prospective trades. Superior forecasting can prevent them from overestimating or underestimating future trades, which can cause significant harm to the companies. Companies that use trustworthy trade calculation may be able to assign their properties more sensibly and get more income. However, forecasting sales is difficult due to a number of internal and external elements from the surrounding environment. As a result, in today's market, there is an urgent need for the development of a smart forecasting system that is rapid, adaptable, and capable of providing high accuracy. We intend to use several machine learning techniques to build and refine a sales forecasting model, as well as to estimate sales data.

2) Using Machine Learning to Forecast Daily Supermarket Sales
Authors: DANIEL FREDÉN and HAMPUS LARSSON

Improved sales estimates for individual products in retail establishments can benefit both the environment and the economy. Historically, projections were made using a combination of statistical observations and experience. However, with the increasing processing power accessible in modern computers, there has been a surge in interest in using machine learning to solve this challenge. The goal of this thesis was to determine whether machine learning approach could estimate sales the best using two years of sales data, yearly calendar events, and meteorological data. The approaches tested were XGBoost, ARIMAX, LSTM, and Facebook Prophet. Overall, the XGBoost and LSTM models fared the best, with lower mean absolute value and symmetric performance.

3) A Comparative Study of Big Mart Sales Prediction

AUTHORS: Gopal Behera, Neeta Nain

Nowadays shopping malls and Big Marts keep the track of their sales data of each and every individual items for predicting future demand of the customer and update the inventory management as well. These data stores basically contain a large number of customer data and individual item attributes in a data warehouse. Further anomalies and frequent patterns are detected by mining the data store from the data warehouse. The resultant data can be used for predicting future sales volume with the help of different machine learning techniques for the retailers like Big Mart. In this paper, we proposed a predictive model using XgBoost technique for predicting the sales of a company like Big Mart and we found our model produces better performance as compare to other model.

3. EXISTING METHODOLOGY

Various regression models are used in crime prediction, health prediction, home prediction, and sales prediction, among other things. Based on XG Boost, predict cardiovascular risk. Sales Forecast is used to forecast product sales in Big Mart Company's numerous outlets. Growing areas become more predictable as product volume increases, but anticipating them by hand becomes increasingly challenging. Python is employed as a programming language in this case, and Jupyter Notebook is used as a tool. Machine learning elements such as the supervised learning function and the regression function were used to create this software. This is done mostly to forecast future sales of the company's store merchandise.

LIMITATION OF EXISTING SYSTEM

1. Low Accuracy

2. Growing areas become more predictable as product volume increases, but anticipating them by hand becomes increasingly challenging.

4. PROPOSED SYSTEM & ITS ADVANTAGES

The proposed approach uses pre-processing and filling the missing values, we used ensemble classifier using Random Forest, Decision trees, Ridge regression, Xgboost and Linear regression..

ADVANTAGES OF PROPOSED SYSTEM:

The following are the benefits of the proposed system. They are:

1. High accuracy

2. Low complexity

3. Python is employed as a programming language in this case, and Jupyter Notebook is used as a tool. Machine learning elements such as the supervised learning function and the regression function were used to create this software.

4. This is done mostly to forecast future sales of the company's store merchandise.

5. PROPOSED DATASET

In this proposed application we try to implement the dataset collected from KAGGLE website.

Kaggle DATASET:

In this phase useful information about the data has been extracted from the dataset. That is trying to identify the information from hypotheses vs available data. There are 1559 different products and also have 10 unique outlets which are present in the dataset. The Item type contains 16 unique values. Whereas two types of Item Fat Content are there but some of them are misspelled as regular instead of 'Regular' and low fat, LF instead of Low Fat.

The dataset is from Kaggle "Big Mart Sales Prediction Datasets" is:
https://www.kaggle.com/datasets/shivan118/big-mart-sales-prediction-datasets

6. IMPLEMENTATION PHASE

The step of implementation is when the theoretical design is translated into a programmatically-based approach. The application will be divided into a number of components at this point and then programmed for deployment. The front end of the application takes Google Collaboratory and as a Back-End Data base we took Kaggle /Big Mart Sales Prediction as input dataset. Python is being used in this instance to implement the present application. The following 3 modules make up the bulk of the application. They are listed below:

1) Data Exploration Module:

This phase extracts important information about the data from the dataset. This is an attempt to identify information from hypotheses vs available data. There are 1559 distinct products in the dataset, as well as 10 distinct outlets. The Item type has 16 distinct values. Although there are two sorts of Item Fat Content, some of them are misspell, such as regular instead of 'Regular' and low fat, LF instead of Low Fat.

2) Data Cleaning Module

The previous section revealed that the attributes Outlet_Size and Item_Weight have missing values. In our work, we substitute missing values in the Outlet Size property with the mode of that attribute, and missing values in the Item Weight field with their attribute value. The missing values are numerical, and substituting them with the mean and mode decreases the correlation between the imputed variables. For our model, we assume that there is no link between the measured and imputed attributes.

3) **Feature Engineering Module:**

During the data exploration phase, some peculiarities were discovered in the data-set. As a result, this phase is utilised to resolve all of the subtleties detected in the dataset and prepare them for developing an acceptable model. Because it was discovered that the value of Item visibility was zero, the mean value of that product's item visibility is utilised in place of the zero values attribute. All discrepancies in categorical attributes are resolved by altering all categorical attributes into a suitable one.
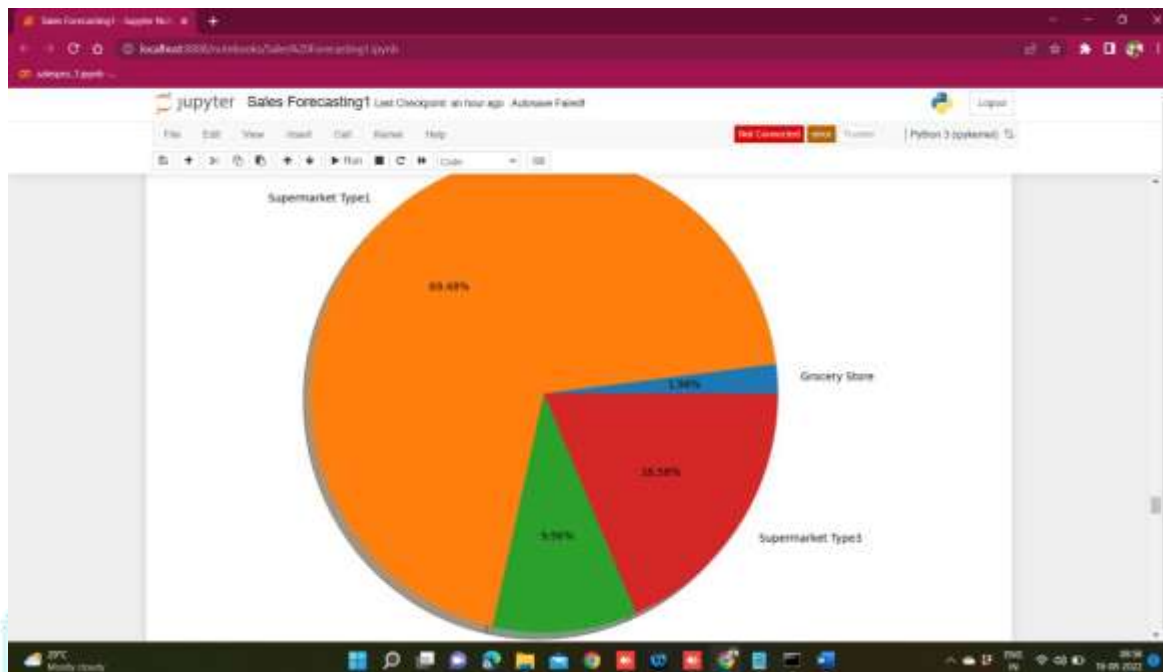
7. EXPERIMETAL REPORTS

In this proposed application, we try to use google collab or Jupitor as working platform and try to show the performance of our proposed application.

1) **MAIN PAGE**

In the below main page we can see application is loaded on Jupitor notebook and we tried to load the dataset with several attributes.
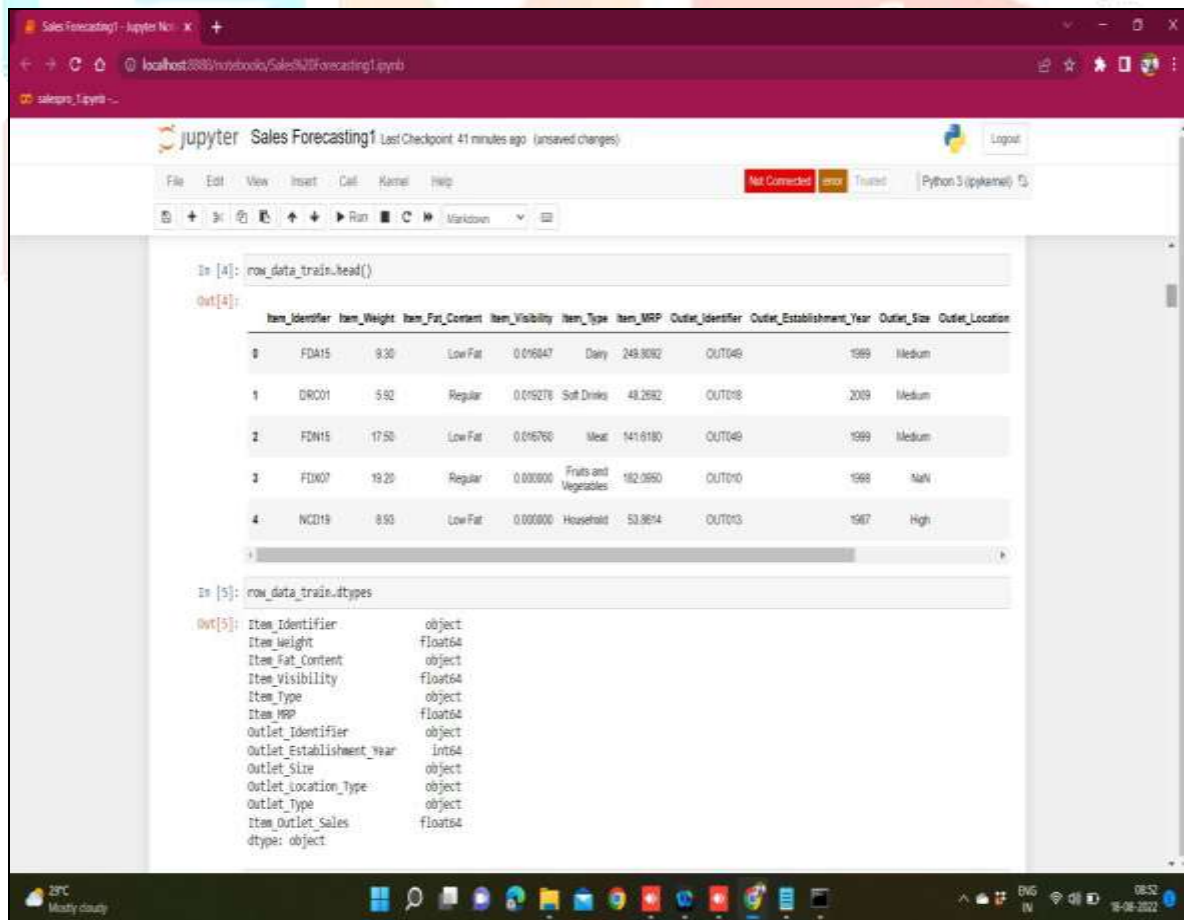
2)   ACCURACY OF SUPERMARKET PREDICTION

Here we can clearly see the accuracy of Supermarket prediction is observed and now we can see how much accuracy is available for each and every super market.

8. CONCLUSION

The fundamentals of machine learning, as well as the accompanying data processing and modelling methods, are presented in this project, as is their application in predicting sales at various Big Mart shopping stores. The implementation demonstrates the relationship between the various criteria studied and how a certain medium-sized location recorded the best sales, implying that additional shopping places should follow similar patterns for increased sales. Multiple instance parameters and various other factors can also be employed to more innovatively and successfully anticipate sales. Accuracy is important in prediction systems and can be considerably boosted by increasing the parameters employed. The way the sub-models work can also lead to an increase in system productivity.

9. REFERENCES

1)   Makridakis, S., Wheelwright, S.C., Hyndman, R.J.: Forecasting methods and applications. John wiley & sons (2008).

2)   Kadam, H., Shevade, R., Ketkar, P. and Rajguru.: ―A Forecast for Big Mart Sales Based on Random Forests and Multiple Linear Regression.‖ (2018).

3)   C. M. Wu, P. Patil and S. Gunaseelan: Comparison of Different Machine Learning Algorithms for Multiple Regression on Black Friday Sales Data (2018).

4) Das, P., Chaudhury: Prediction of retail sales of footwear using feed forward and recurrent neural networks (2018) 5. Das, P., Chaudhury,

5) Comparison of Different Machine Learning Algorithms for Multiple Regression on Black Friday Sales Data (2007).

6) Smola, A., & Vishwanathan, S. V. N. (2008). Introduction to machine learning. Cambridge University, UK, 32, 34.

7) Saltz, J. S., & Stanton, J. M. (2017). An introduction to data science. Sage Publications.

8) Shashua, A. (2009). Introduction to machine learning: Class notes 67577. arXiv preprint arXiv:0904.3664.

9) MacKay, D. J., & Mac Kay, D. J. (2003). Information theory, inference and learning algorithms. Cambridge university press.

10) Daumé III, H. (2012). A course in machine learning. Publisher, ciml. info, 5, 69.

11) Quinlan, J. R. (2014). C4. 5: programs for machine learning. Elsevier.

12) Cerrada, M., & Aguilar, J. (2008). Reinforcement learning in system identification. In Reinforcement Learning. IntechOpen.

13) Welling, M. (2011). A first encounter with Machine Learning. Irvine, CA.: University of California, 12.

10. ABOUT THE AUTHORS

Mr.K. TULASI KRISHNA KUMAR is currently working as Associate Professor in Department of Computer Science and Engineering at Sanketika Vidhya Parishad Engineering College, P.M. Palem, Visakhapatnam, Andhra Pradesh. He has more than 13 years of teaching experience. His research interest includes AI, Java and Python.

Ms. VELUDOOTI RUPA is currently pursuing her 2 years MCA in Department of Computer Science and Applications at Sanketika Vidhya Parishad Engineering College, P.M. Palem, Visakhapatnam, Andhra Pradesh. Her area of interest includes Python, Java, C, and C++.