

BIGBAZAAR SALES PREDICTION

Swathi M (9922146017) Student, Department of M. Sc. Data Science, Kalasalingam Academy of Research and Education, Krishnan Koil, Srivilliputhur, Tamil Nadu 626126.

ABSTRACT

In the era of the internet, shopping malls like big bazaar keep the track of their sales data of each and every individual item 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. The dataset which I used is Big bazaar Sales Data can be used for predicting future sales volume with the help of different machine learning techniques. In this Project I have proposed a predictive model that is Linear Regression, Random Forest Regression and Xgboost Regressor for predicting the sales of a company and found that the model produces better performance as compared to existing models.

Keywords: Big bazaar Sales Prediction, Linear Regression, Random Forest Regression and Xgboost Regressor.

1.INTRODUCTION

Day by day competition among different shopping malls as well as D marts is getting more serious and aggressive only due to the rapid growth of the global malls and on-line shopping. Every mall or mart is trying to provide personalized and short-time offers for attracting more customers depending upon the day, such that the volume of sales for each item can be predicted for inventory management of the organization, logistics and transport service, etc. Present machine learning algorithm are very sophisticated and provide techniques to predict or forecast the future demand of sales for an organization, which also helps in overcoming the cheap availability of computing and storage systems. In this paper, we are addressing the problem of D mart sales prediction or forecasting of an item on customer's future demand in different D mart stores across various locations and products based on the previous record. Different machine learning algorithms like linear regression analysis, random forest, are used for prediction or forecasting of sales volume. As good sales are the life of every organization so the forecasting of sales plays an important role in any shopping complex. Always a better prediction is helpful, to develop as well as to enhance the strategies of business about the marketplace which is also helpful.

2.SYSTEM REQUIREMENTS

HARDWARE REQUIREMENTS:

- System : Pentium i3 Processor.
- Hard Disk : 500 GB.
- Monitor : 15" LED
- Input Devices : Keyboard , Mouse.

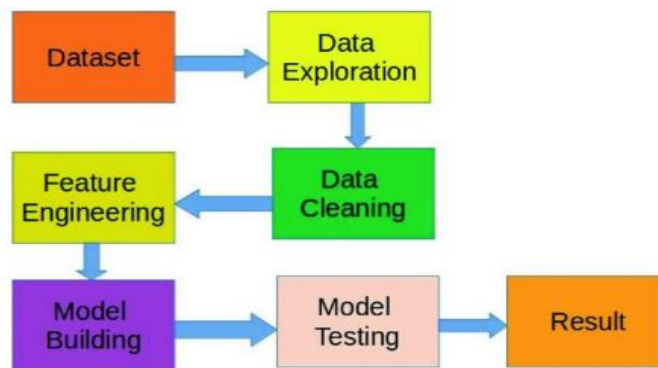
- Ram : 4 GB

SOFTWARE REQUIREMENTS:

- Operating system : Windows 10.
- Coding Language : Python
- Web Framework : Flask

3. METHODOLOGY

The aim is to build a predictive model and find out the sales of each product at a particular store. Using this model, Big bazaar will try to understand the properties of products and stores which play a key role in increasing sales.



Workflow of this project

MODULES DESCRIPTION:

- Data Collection
- Dataset
- Data Preparation
- Model Selection
- Analyze and Prediction
- Accuracy on test set
- Saving the Trained Model

Data Collection:

The dataset which I used in this project was Bigbazaar_sales_data which is available in Kaggle.

Link : <https://www.kaggle.com/datasets/atulmittal199174/bigbazaar-sales-data>

Dataset: The dataset consists of 8523 individual data. There are 12 columns in the dataset, which are described below.

1. Item Identifier ---- Unique product ID

- 2.ItemWeight ---- Weight of product
- 3.ItemFatContent ---- Whether the product is low fat or not
- 4.ItemVisibility ---- The % of the total display area of all products in a store allocated to the particular product
- 5.ItemType ---- The category to which the product belongs
- 6.ItemMRP ---- Maximum Retail Price (list price) of the product
- 7.OutletIdentifier ---- Unique store ID
- 8.OutletEstablishmentYear ---- The year in which the store was established
- 9.OutletSize ---- The size of the store in terms of ground area covered
- 10.OutletLocationType ---- The type of city in which the store is located
- 11.OutletType ---- Whether the outlet is just a grocery store or some sort of supermarket
12. ItemOutletSales ---- sales of the product in t particular store. This is the outcome variable to be predicted.

Data Preparation:

Clean that which may require it (remove duplicates, correct errors, deal with missing values, normalization, data type conversions, etc.)

Visualize data to help detect relevant relationships between variables or class imbalances (bias alert!), or perform other exploratory analysis and Split the data into training and evaluation sets

Model Selection:

I have used machine learning algorithm, I got a accuracy of 94.04% on train set so we implemented this algorithm.

In Linear Regression machine learning algorithm, I got a accuracy of 50.71% on train set.

In Xgboost Regressor machine learning algorithm, I got a accuracy of 87.62% on train set.

Analyze and Prediction:

In the actual dataset, I chose only 9 features:

- 1.ItemWeight ---- Weight of product
- 2.ItemFatContent ---- Whether the product is low fat or not
- 3.ItemVisibility ---- The % of the total display area of all products in a store allocated to the particular product
- 4.ItemType ---- The category to which the product belongs
- 5.ItemMRP ---- Maximum Retail Price (list price) of the product
- 6.OutletEstablishmentYear ---- The year in which the store was established

7.OutletSize ---- The size of the store in terms of ground area covered

8.OutletLocationType ---- The type of city in which the store is located

9.OutletType ---- Whether the outlet is just a grocery store or some sort of supermarket.

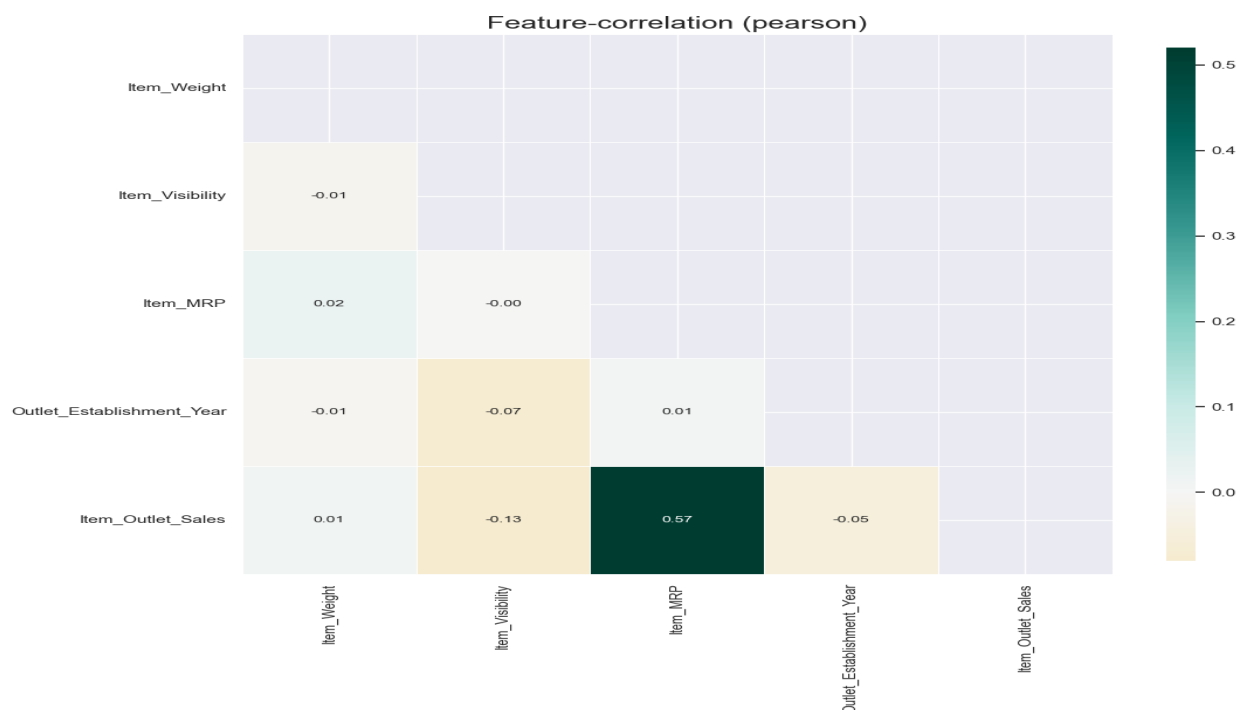
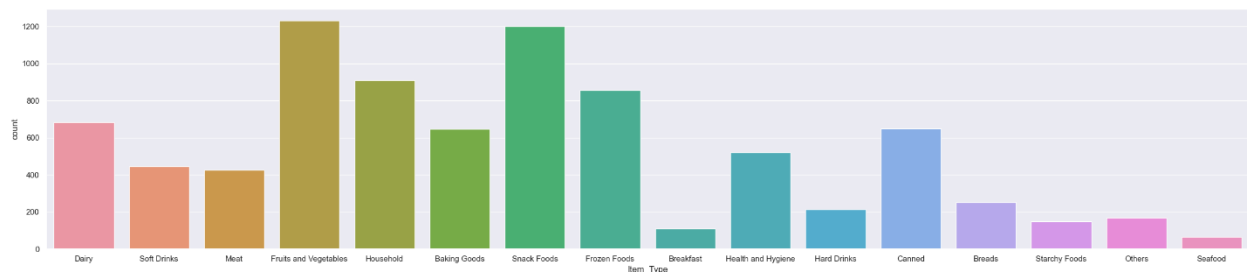
Accuracy on test set:

I got an accuracy of 55.62% on test set by the selected model Random Forest Regressor.

Saving the Trained Model:

save our trained model using joblib, you can use the “dump” function, which takes two arguments: the trained model object and the filename for saving the model. Make sure you have joblib installed in your environment.

4. EXPERIMENTAL RESULTS



Bigbazaar Sales Prediction

5.920000

Enter Item Weight

Low Fat

0.016047

Enter Item Visibility

Soft Drinks

48.269199

Enter Item MRP

2009

Outlet Establishment Year (YYYY)

Medium

Tier 2

Supermarket Type2

Activate Windows

Go to Settings to activate Windows.

5.CONCLUSION

In this documentation, basics of machine learning and the associated data processing and modeling algorithms have been described, followed by their application for the task of sales prediction in Big bazaar shopping centers at different locations. On implementation, the prediction results show the correlation among different attributes considered and how a particular location of medium size recorded the highest sales, suggesting that other shopping locations should follow similar patterns for improved sales. Accuracy, which plays a key role in prediction-based systems, can be significantly increased as the number of parameters used are increased. best performance-algorithm, here propose software to using regression approach for predicting the sales centered on sales data from the past the accuracy of linear regression prediction can be enhanced with this method, and Random Forest Regression can be determined. So, we can conclude Random Forest Regression gives the better prediction with respect to Accuracy.