

DAYANANDA SAGAR COLLEGE OF ENGINEERING

(An Autonomous Institute affiliated to VTU, Belagavi, Approved by AICTE & ISO 9001:2008 Certified)

Accredited by National Assessment & Accreditation Council (NAAC) with 'A' grade,

Shavige Malleshwara Hills, Kumaraswamy Layout, Bengaluru-111



Mini Project Report

on

“BIG-MART SALES PREDICTION”

Submitted by

Ramachandra (1DS19CS124)

Siddesh Pattanshetty (1DS19CS157)

Pavan Patel G S (1DS19CS111)

Sachin C B (1DS19CS137)

Sixth Semester B.E (CSE)

Mini Project

19CS6DCMIP

Under the guidance of

Abhinav R B

Assistant Professor

Dept. of CSE

DSCE, Bangalore

Department of Computer Science and Engineering

Dayananda Sagar College of Engineering

Bangalore-111

VISVESVARAYA TECHNOLOGICAL UNIVERSITY

DAYANANDA SAGAR COLLEGE OF ENGINEERING

Shavige Malleshwara Hills, Kumaraswamy Layout, Bangalore - 560078

Department of Computer Science & Engineering



CERTIFICATE

This is to certify that the project entitled **Big Mart Sales Prediction** is a bonafide work carried out by **Ramachandra[1DS19CS124]**, **Pavan Patel G S[1DS19CS111]**, **Siddesh Pattanshetty [1DS19CS157]**, **Sachin C B[1DS19CS137]** in partial fulfilment of 6th semester, Bachelor of Engineering in Computer Science and Engineering under Visvesvaraya Technological University, Belgaum during the year 2021-22.

Abhinav R B

(Internal Guide)

Asst. Professor

Department of CSE,

DSCE

Dr. Ramesh Babu

Vice principal & Head

Department of CSE,

DSCE

Dr.CPS Prakash

Principal,

DSCE

Signature.....

Signature.....

Signature.....

Name of the Examiners:

Signature with date:

1.

.....

2.

.....

ACKNOWLEDGEMENT

We are pleased to have successfully completed the Mini project “**Big Mart Sales Prediction**”. We thoroughly enjoyed the process of working on this project and gained a lot of knowledge doing so.

We would like to take this opportunity to express our gratitude to **Dr. C P S Prakash**, Principal of DSCE, for permitting us to utilize all the necessary facilities of the institution.

We also thank our respected Vice Principal, HOD of Computer Science & Engineering, DSCE, Bangalore, **Dr. Ramesh Babu D R**, for his support and encouragement throughout the process.

We are immensely grateful to our respected and learned guide, **Prof. Abhinav R B**, Asst. Professor, CSE, DSCE for his valuable help and guidance. We are indebted to them for their invaluable guidance throughout the process and their useful inputs at all stages of the process.

We also thank all the faculty and support staff of Department of Computer Science, DSCE. Without their support over the years, this work would not have been possible.

Lastly, we would like to express our deep appreciation towards our classmates and our family for providing us with constant moral support and encouragement. They have stood by us in the most difficult of times.

PAVAN PATEL G S-	1DS19CS111
RAMACHANDRA-	1DS19CS124
SIDDESH PATTANSHETTY -	1DS19CS157
SACHIN C B-	1DS19CS137

CONTENTS

SL. NO	CONTENT	Page No.
1.	Abstract	4
2.	Introduction	5-6
3.	Literature Survey	7-8
4.	System design & Methodology	9-10
5.	Snapshots and Results	11-12
6.	Conclusion & Future Enhancements	13
7.	Appendix	14
8.	References	15

ABSTRACT

Machine Learning is transforming in every walk of life and has become a major contributor in real world scenarios. The revolutionary applications of Machine Learning can be seen in every field including education, healthcare, engineering, **sales**, entertainment, transport and several more; the list is never ending. The traditional approach of sales and marketing goals no longer help the companies, to cope up with the pace of competitive market, as they are carried out with no insights to customers' purchasing patterns. Major transformations can be seen in the domain of sales and marketing as a result of Machine Learning advancements.

Nowadays shopping malls and Big Marts retain 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. 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 like Big Mart. In this paper, we are making some models using for predicting the sales of a company like Big Mart and found which model produces better performance as compared to other models. A comparative analysis of the model with others in terms of performance metrics is also explained in detail.

Chapter 1

INTRODUCTION

Day by day competition among different shopping malls as well as big 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.

Big Mart is a Grocery Super Market Brand. Big Mart Brand has started out its journey with free home delivery offerings of food and grocery. Big Mart lets in you to walk far away from the drudgery of grocery shopping and welcome a clean comfortable way of browsing and shopping for groceries. Discover new merchandise and shop for all of your food and grocery desires from the comfort of your private home or workplace. No greater getting stuck in traffic jams, procuring parking, standing in long queues and wearing heavy bags – get everything you want ,when you want, right at the doorstep.

A standard sales prediction study can help in deeply analyzing the situations or the conditions previously occurred and then, the inference can be applied about customer acquisition, funds inadequacy and strengths before setting a budget and marketing plans for the upcoming year. In other words, sales prediction is based on the available resources from the past. In depth knowledge of past is required for enhancing and improving the likelihood of marketplace irrespective of any circumstances especially the external circumstance, which allows to prepare the upcoming needs for the business.

1.1 Machine Learning

The data available is increasing day by day and such a huge amount of unprocessed data is needed to be analysed precisely, as it can give very informative and finely pure gradient results as per current standard requirements. It is not wrong to say as with the evolution of Artificial Intelligence (AI) over the past two decades, Machine Learning (ML) is also on a fast pace for its evolution. In machine learning, one deals with both supervised and unsupervised types of tasks and generally a classification type problem accounts as a resource for knowledge discovery. It generates resources and employs regression to make precise predictions about future, the main emphasis being laid on making a system self-efficient, to be able to do computations and analysis to generate much accurate and precise results. By using statistic and probabilistic tools, data can be converted into knowledge. The statistical inferencing uses sampling distributions as a conceptual key.

1.3 Dataset Description of Big Mart:

In our work we have used 2013 Sales data of Big Mart as the dataset. Where the dataset consists of 12 attributes like Item Fat, Item Type, Item MRP, Outlet Type, Item Visibility, Item Weight, Outlet Identifier, Outlet Size, Outlet Establishment Year, Outlet Location Type, Item Identifier and Item Outlet Sales. Out of these attributes response variable is the Item Outlet Sales attribute and remaining attributes are used as the predictor variables. The data-set consists of 8523 products across different cities and locations. The data-set is also based on hypotheses of store level and product level. Where store level involves attributes like: city, population density, store capacity, location, etc and the product level.

1.2 Problem statement :

To find out what role certain properties of an item play and how they affect their sales by understanding Big Mart sales.” In order to help Big Mart achieve this goal, a predictive model can be built to find out for every store, the key factors that can increase their sales and what changes could be made to the product or store’s characteristics.

Chapter 2

LITERATURE SURVEY

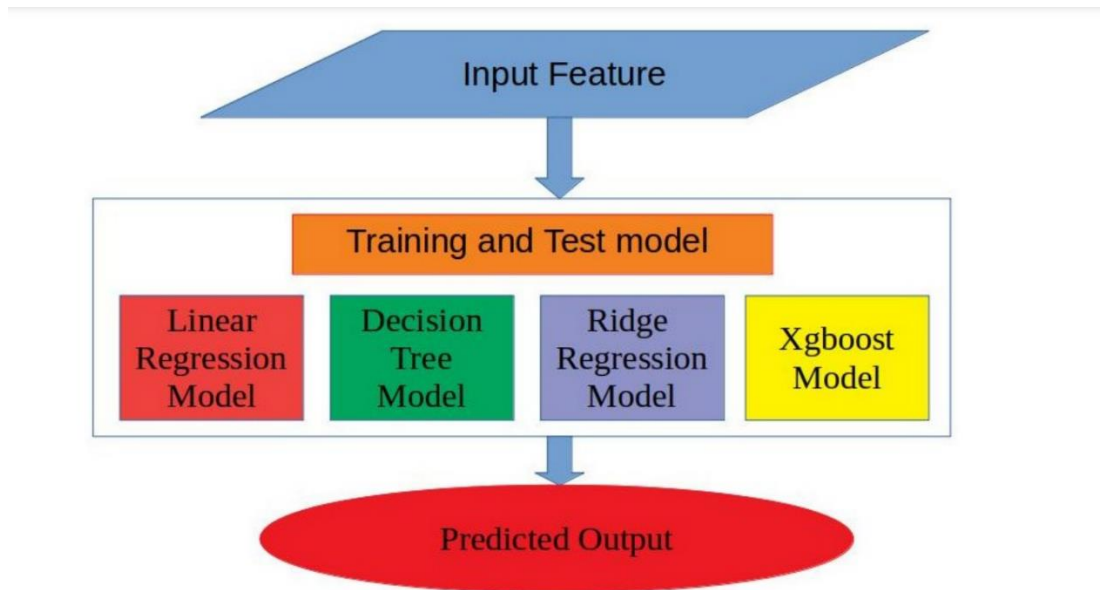
- For the project, Big mart sales prediction, the following research papers were referred before the implementation process. These research papers are studied and tabulated as follows having the title, algorithm/technique used, as well as their performance and the dataset used in the implementation.

Title	Algorithms/Technique	Description	Author
1. A Forecast for Big Mart Sales Based on Random Forests and Multiple Linear Regression (2018)	Random Forests and Multiple Linear Regression	A Forecast for Big Mart Sales Based on Random Forests and Multiple Linear Regression used Random Forest and Linear Regression for prediction analysis which gives less accuracy. To overcome this we can use XG boost Algorithm which will give more accuracy and will be more efficient.	Kadam, H., Shevade, R., Ketkar, P. and Rajguru
2. Prediction of retail sales of footwear using feed forward and recurrent Neural Networks (2018)	feed forward and recurrent neural networks , XG Boost	Prediction of retail sales of footwear using feed forward and recurrent neural networks used neural networks for prediction of sales. Using neural network for predicting of weekly retail sales, which is not efficient, So XG boost can work efficiently.	Das, P., Chaudhury
3. Comparison of Different Machine Learning Algorithms for Multiple Regression on Black Friday Sales Data (2018)	Multiple Linear Regression and Neural networks	To overcome this Complex models like neural networks is used for comparison between different algorithms which is not efficient so we can use more simpler algorithm for prediction.	C. M. Wu, P. Patil and S. Gunaseelan

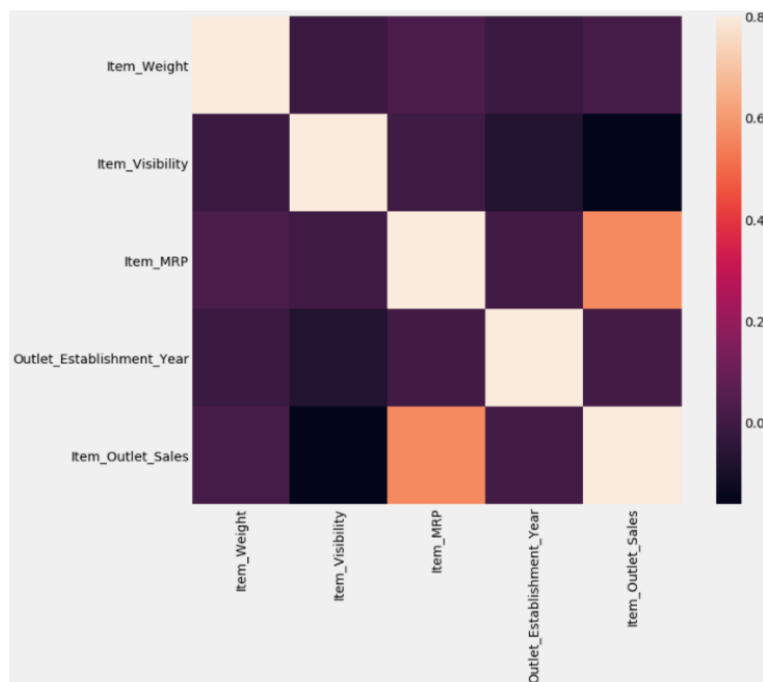
Big-Mart Sales Prediction

4. Big Mart Sales Prediction Using Machine Learning	Multiple Linear Regression and Neural networks	Machine Learning is a category of algorithms that allows software applications to become more accurate in predicting outcomes without being explicitly programmed.	Nimit Jain
5. Big Mart Sales Prediction and Analysis	Machine Learning, Sales Prediction, Big Mart, Random Forest, Linear Regression	Machine Learning is a technology that allows machines to become more accurate in predicting outcomes without being explicitly programmed for it. The basic premise of machine learning is to build models and deploy algorithms that can receive input data and use statistical analysis to predict an output while modifying outputs as the new data becomes available.	Shantanu Choudhary, Utkarsh Singh, Nikhil Saxena, Sameer Jain
6. A Two-Level Statistical Model for Big Mart Sales Prediction	feed forward and recurrent neural networks , XG Boost	Sales forecasting is an important aspect of different companies engaged in retailing, logistics, manufacturing, marketing and wholesaling. It allows companies to efficiently allocate resources, to estimate achievable sales revenue and to plan a better strategy for future growth of the company.	<u>Kumari Punam</u> ; <u>Rajendra Pamula</u> ; <u>Praphula Kumar Jain</u>

SYSTEM DESIGN & METHODOLOGY

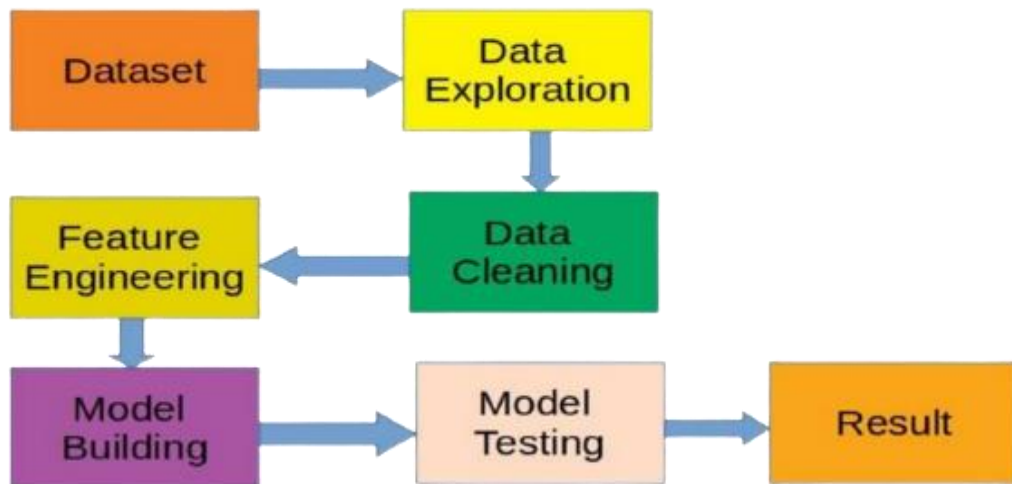


Framework of proposed model. Model received the input features and split it into training and test set. The trained model is used to predict the future sales.

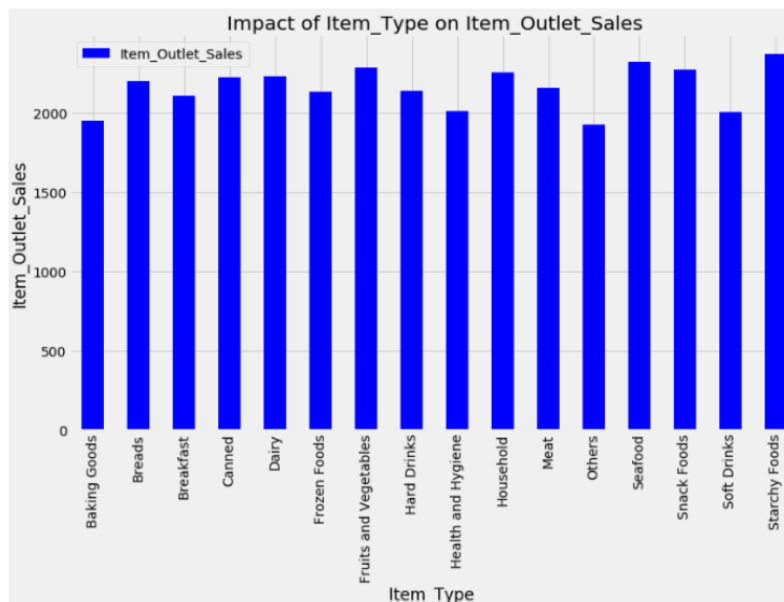


Correlation among features of a dataset. Brown squares are highly correlated whereas black square represents bad correlation among attributes.

Big-Mart Sales Prediction

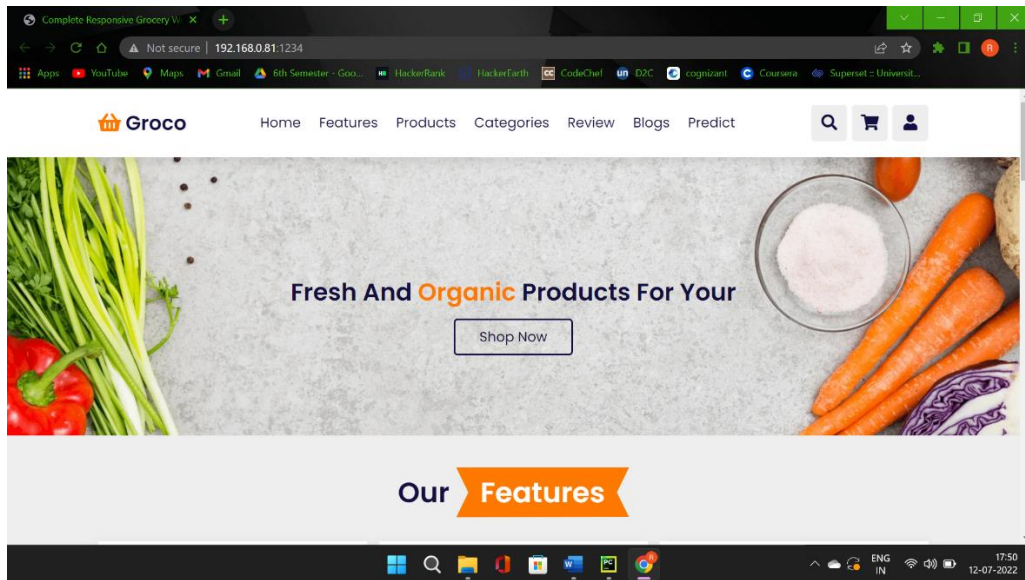


1. **Data Exploration** – looking at categorical and continuous feature summaries and making inferences about the data.
2. **Data Cleaning** – imputing missing values in the data and checking for outliers
3. **Feature Engineering** – modifying existing variables and creating new ones for analysis
4. **Model Building** – making predictive models on the data

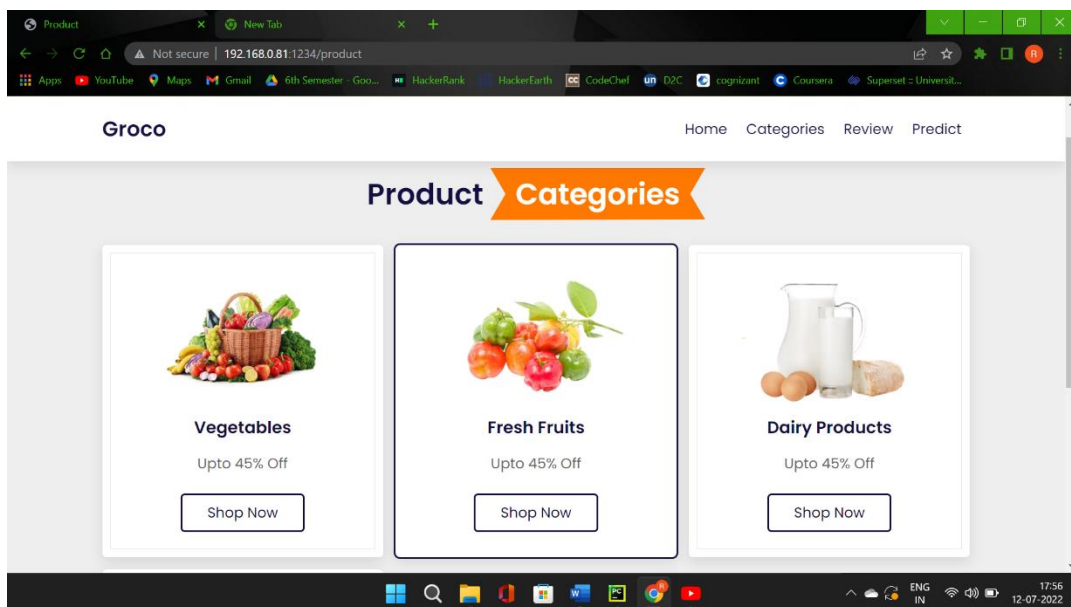


Impact of item type on target variable item outlet sale

SNAPSHOTS AND RESULTS



This is the Home page of the Big-Mart website. Which consist of different types of button in the header to navigate into the other pages like categories, blogs and predict etc.



Different categories of products in our mart like vegetables, fruits and dairy products.

Big-Mart Sales Prediction

The screenshot shows a web browser window with the URL `127.0.0.1:8000/predict`. The page is titled "Groco" and has a navigation bar with links: Home, Products, Review, and Predict. The main content area is titled "Sales Prediction" and features a background image of a grocery store aisle. Overlaid on this are four input fields: "Enter Item Weight" (a text box), "Item Fat Content" (a dropdown menu), "Enter Item Visibility" (a text box), and "Item Type" (a dropdown menu). Below these is a label "Enter Item MRP" followed by another text box. The browser's taskbar at the bottom shows the time as 14:19 on 27-06-2022.

This is the Prediction page we can predict the price of the required item or a categories of the item

The screenshot shows the same web browser window, but now displaying the "Predicted" page. The URL is `192.168.0.81:1234/Predicted`. The navigation bar now includes "Categories" instead of "Products". Below the navigation bar is a table with the following data:

Item Weight	Fat Content	Type Of Product	MRP	Establishment Year	Future Sale Of Product
1	Low Fat	Baking Foods	10.0	2000	[-2224.51419266]

Below the table is a background image of a grocery store aisle. The browser's taskbar at the bottom shows the time as 18:03 on 12-07-2022.

Final output of our project which consist of the future of the particular product.

Chapter 5

CONCLUSION

FUTURE ENHANCEMENTS

We are predicting the accuracy for Linear regression algorithm. Our predictions help big marts to refine their methodologies and strategies which in turn helps to increase the profit. The results predicted will be very useful for the executives of the company to know about their sales and profits. This will also give them the idea for their new locations or Centre's of Big-mart.

In present era of digitally connected world every shopping mall desires to know the customer demands beforehand to avoid the shortfall of sale items in all seasons. Day to day the companies or the malls are predicting more accurately the demand of product sales or user demands. Extensive research in this area at enterprise level is happening for accurate sales prediction. As the profit made by a company is directly proportional to the accurate predictions of sales, the Big marts are desiring more accurate prediction algorithm so that the company will not suffer any losses. In this research work, we have designed a predictive model by modifying Gradient boosting machines as Xg-boost technique and experimented it on the 2013 Big Mart dataset for predicting sales of the product from a particular outlet. Experiments support that our technique produce more accurate prediction compared to than other available techniques like decision trees, ridge regression etc. Finally a comparison of different models is summarized in Table 2. From Table 2 it is also concluded that our model with lowest MAE and RMSE performs better compared to existing models.

Future enhancement: Hyper-parameter Tuning and Gradient Boosting.

Chapter 6

APPENDIX

Machine learning : Machine learning (ML) is a type of artificial intelligence (AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so.

Data warehouse :A data warehouse is a central repository of information that can be analyzed to make more informed decisions.

Artificial intelligence :Artificial intelligence is the simulation of human intelligence processes by machines, especially computer systems. Specific applications of AI include expert systems, natural language processing, speech recognition and machine vision.

Supervised learning :Supervised learning is an approach to creating artificial intelligence (AI), where a computer algorithm is trained on input data that has been labeled for a particular output.

Unsupervised learning :It refers to the use of artificial intelligence (AI) algorithms to identify patterns in data sets containing data points that are neither classified nor labeled.

Statistical inference : It is the process of using data analysis to infer properties of an underlying distribution of probability.

Chapter 7

REFERENCES

- Big Mart Sales Prediction and Analysis
By Shantanu Choudhary, Utkarsh Singh, Nikhil Saxena, Sameer Jain
<https://www.ijirt.org/Article?manuscript=155006>
- SALES PREDICTION MODEL FOR BIG MART
By Nikita Malik and Karan Singh
https://www.researchgate.net/publication/344099746_SALES_PREDICTION_MODEL_FOR_BIG_MART
- Dataset link -- <https://www.kaggle.com/code/hiralmshah/bigmart-sales-prediction>

PROJECT SOURCE CODE LINK:

<https://drive.google.com/drive/u/0/folders/1LGADwk-b8b0fl0yCL35uhQCoxJN05DZz>