A Mini Project Report on

Big Mart Sales Prediction

T.E. - I.T Engineering

Submitted By

Mayur Jain (20104024)

Aryan Amin (20104071)

Manashree Chavan(20104028)

Madhur Dukhande (20104014)

Under The Guidance Of

Ms.Rujata Chaudhari

DEPARTMENT OF INFORMATION TECHNOLOGY

A.P.SHAH INSTITUTE OF TECHNOLOGY

G.B. Road, Kasarvadavali, Thane (W), Mumbai-400615 UNIVERSITY OF MUMBAI

Academic year : 2022-23

CERTIFICATE

This to certify that the Mini Project report on Big Mart Sales Prediction has been submitted by Mayur Jain(20104024), Aryan Amin(20104071),
Manashree Chavan(20104028), Madhur Dukhande(20104014) who are Bonafede students of A. P. Shah Institute of Technology, Thane, Mumbai, as a partial fulfillment of the requirement for the degree in **Information Technology**, during the academic year **2022-23** in the satisfactory manner as per the curriculum laid down by University of Mumbai.

Ms.Rujata Chaudhari,
Guide

Dr. Kiran Deshpande
Head Department of Information Technology

Dr. Uttam D.Kolekar
Principal

External Examiner(s)
1.
2.

Place: A.P.Shah Institute of Technology, Thane
Date:

ACKNOWLEDGEMENT

This project would not have come to fruition without the invaluable help of our guide Rujata Chaudhari. Expressing gratitude towards our HoD, **Dr. Kiran Deshpande**, and the Department of Information Technology for providing us with the opportunity as well as the support required to pursue this project. We would also like to thank our teacher Ms. Mansi Choche who gave us her valuable suggestions and ideas when we were in need of them. We would also like to thank our peers for their helpful suggestions.

ABSTRACT

In the modern era of reaching new lengths of advancement, every company and enterprise are working on their customer demands as well as their inventory management. The models used by them help them predict future demands by understanding the pattern from old sales records. Lately, everyone is abandoning the traditional prediction models for sales forecasting as it takes a prolonged amount of time to get the expected results. Therefore now the retailers keep track of their sales record in the form of a data set, which comprises price tag, outlet types, outlet location, item visibility, item outlet sales etc.

In this project, exploratory machine learning approaches are used to forecast big-mart store sales. In general, sales forecasting is crucial for advertising, merchandising, warehousing, and production, and it is done in a variety of organizations. To modify the business strategy to predicted results, the sales estimate is based on Big Mart sales from different stores. Different machine learning approaches may then be applied to forecast possible sales volumes for stores like Big Mart.

TABLE OF CONTENTS

1. Introduction……...1

   1.1.Purpose..... 1

   1.2.Problem Statement……...2

   1.3.Objectives..... ..2

   1.4.Scope..... 3

2. Literature Review……....4

3. Proposed System…….....7

   3.1. Features and Functionality……7

4. Data Processing and Methodology.....9

5. Project Design.....10

   5.1.Use Case diagram……...10

   5.1. System Architecture……..11

6. Technical specification……...…….....13

8. Result and Discussion…… .17

9.Conclusion and Future Scope.....18

10. References……19

References

Chapter 1

Introduction

Customer Satisfaction and keeping up with the demand for products is very important for any store to survive in the market and to compete with other stores. And these two can only be achieved when you have a future demand figure for coming up with new plans for a flourishing business .Many businesses and enterprises keep track of the statistical data of their products so as to analyze their future demand in the market. The stored statistical data consists of the amount of items sold and its categories and other attribute details to their supply, to grow their business, and improvise their sales strategies and this might come in handy in the near future or seasonally in-order to create short-term discount offers to attract more customers towards their brand.

The aim of our project is to build a machine learning model which can predict the sales of a product and understand its patterns and trends which is an important part of a big mart's management.

To keep customer satisfaction and rising demands for products is very important for any store to survive in the market

This can be achieved when you have a future demand figure for coming up with new plans for a flourishing business.

- Problem Identified:
- Big Mart faced challenges in accurately predicting the sales of its products, which led to issues such as overstocking or understocking of products in their stores.
- Big Mart faced difficulties in setting optimal prices for its products due to the lack of accurate sales forecasting, which affected its competitiveness in the market.

- Solution Purposed:
- Accurate sales forecasting using machine learning, leveraging historical sales data and other relevant factors, leading to reduced wastage and lost revenue.
- Optimal pricing strategies based on demand, competition, and other factors, leading to increased competitiveness in the market.

1.1 Purpose :

1) To provide trends and patterns to the organization regarding their supply.
2) To improvise their sales strategies.
3) To help the retailers to set the stock quantity more accurately.
4) Easy to use.
5) To find the factors that attract customers towards Big-mart Shops.

1.2 Problem Statement :

With an increased population, the number of stores and shopping malls is also increasing creating competition between different enterprises for bigger sales and popularity. Along with grocery shops and stores, even enterprises need an analysis to check about the patterns and predict future sales. In order to maintain the efficacy of marketing organizations, all suppliers must use an efficient and optimum forecasting method. Manual material handling of this work may result in significant mistakes, leading to poor organization management, and, more significantly, that would be time consuming, which is not desired in today's fast-paced environment.

The main aim of business sectors is to attract the target audience. As a result, it's critical that the firm has already been capable of reaching this goal through the use of a prediction model. Big Mart’s trends are extremely important, using a computer to predict Big Mart sales allows data scientists to explore different patterns by shop and product to get the best results.

1.2 Objectives :

1) To predict future sales from a given dataset
2) To understand the key features that are responsible for the sale of a particular product
3) To Find the best algorithm that will predict sales with the greatest accuracy
4) To pick the foremost suitable demand prediction and with the aid of which Big Mart will prepare its marketing campaigns

1.4 Scope :

1) To develop an accurate model that can forecast sales for each product at each store.
2) To Collect & analyze historical sales data for each product and store to help identify trends, patterns, and seasonality in sales
3) Financial Goals: To predict sales that help Big Mart to set financial goals and plan its revenue growth.
4) Work force management: To predict sales that help Big Mart to adjust the number of staff.

Chapter 2

Literature Review

Sr.no	Title	Author(s)	Year	Algorithms	Limitations	Result
1.	Sales Prediction using machine learning	K.Saraswati, P.Naveen	2021	-Simple Linear Regression -Multile Linear Regression	Linear Regression is very famous for prediction but it gives less accuracy.	The data is processed to select the features and extract those features. Accurate projections make it easier for the shop to boost demand growth and a higher degree of sales generation.
2.	Big Mart Sales prediction using Machine learning	Kopilaka Rajesh, V.Prabhakar, Bhuvya Klayan	2022	XG Boost algorithm	System accuracy is not being improved over time.	When the user submits Details of a particular item, the system will predict sales generated by that item. Hence, leads to an increase in in sales.
3.	Big Mart Sales using machine learning with data analysis	Asha Jyoti Kalluri, Arun Kumar, Aishwarya Poojari	2020	XGBoost Decision Tree	insufficient data to accurately predict sales, the models performance may be compramised	The Machine Learning Methods here, should provide an effective method for data shaping and decision-making. It will help to select the most suitable demand prediction algorithm and with the aid of which BigMart will prepare its marketing campaigns.

Chapter 3

PROPOSED SYSTEM

3.1 Features and Functionality

1. Data preprocessing: cleaning, transforming, and normalizing the dataset.
2. Feature engineering: identifying and selecting relevant features for the model.
3. Exploratory data analysis: visualizing and understanding the patterns and trends in the data.

Chapter 4

4.1 DATA PROCESSING AND METHODOLOGY

A. Data Collection: We have collected the data securely in accordance with an agreed methodology. The procedure for the collected data may differ from client to client and is dependent on the type, quantity, availability and need of data. Data Cleaning and Preprocessing

B. Data Cleaning and Preprocessing : The collected data is passed through a 'cleaning' process, so as to make sure that the data is segregated properly and identified gaps in the data are filled with the appropriate information, making data compatible and also fixing errors in storage systems which can cause data redundancy.

C. Data Modeling: This is primarily a process in which the given dataset and the objects in it are analyzed to get a clear view of the requirements that may help us support our business model. Based on the analysis on patterns present in the data, models are then created on the established flow of the project.

D.Data Prediction: Machine Learning prediction models are trained in this process and then later on evaluated using the data. This will then be applied to the preprocessed dataset. Some of the Models to be used for the prediction are:

Linear Regression

Random Forest

Decision tree

XG Boost Regressor

E. Data Visualization

Data Analyzed is then further picturized for customers and the admin to reach out conclusions and take effective decisions.

4.2 Technical Feasibility:

Technical Feasibility assessment is focused on the present technical resource available in the organization. It studies if the technical resources including the technical team are capable of converting the ideas into working system. Software Requirements are:

Frontend: HTML, CSS

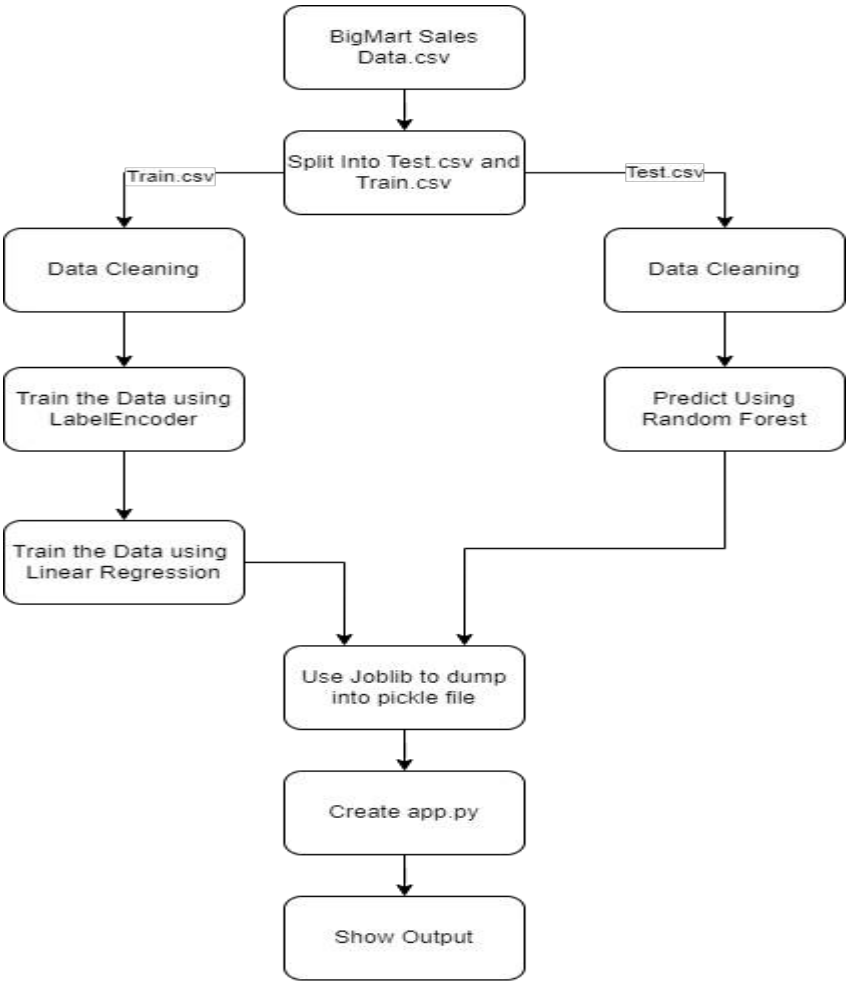
Backend: Python

Database: MySQL

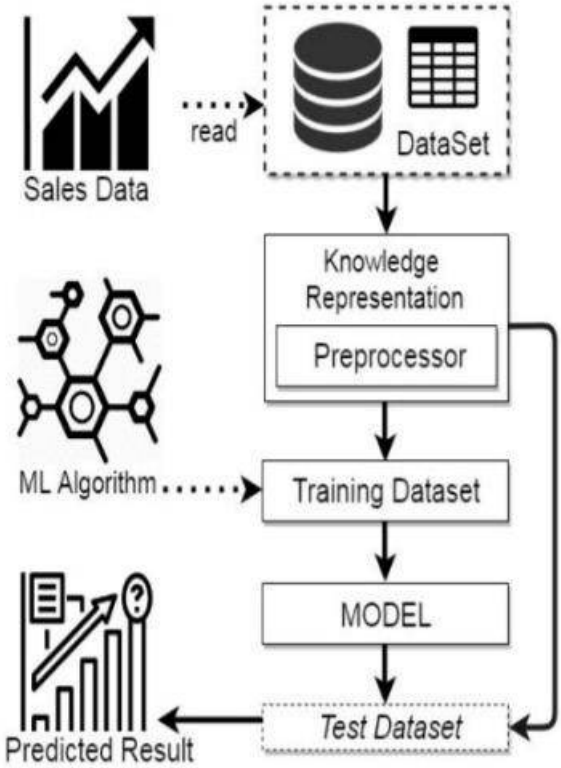
Development Environment: Microsoft Visual Studio

Chapter 5

5.1 Block Diagram:



5.2 System Architecture:



Chapter 6

Technical Specification

Frontend:

HTML:

HTML stands for HyperText Markup Language. It is used to design web pages using the markup language. HTML is the combination of Hypertext and Markup language. Hypertext defines the link between the web pages and markup language defines the text document within the tag that define the structure of web pages.

HTML is used to create the structure of web pages that are displayed on the World Wide Web (www). It contains Tags and Attributes that are used to design the web pages. Also, we can link multiple pages using Hyperlinks.

The basic structure of an HTML page is laid out below. It contains the essential buildingblock elements (i.e. doctype declaration, HTML, head, title, and body elements) upon which all web pages are created. The HyperText Markup Language or HTML is the standard markup language for documents designed to be displayed in a web browser. It can be assisted by technologies such as Cascading Style Sheets (CSS) and scripting languages such as JavaScript. Web browsers receive HTML documents from a web server or from local storage and render the documents into multimedia web pages. HTML describes the structure of a web page semantically and originally included cues for the appearance of the document.

CSS:

CSS(CSS stands for Cascading Style Sheets)

CSS describes how HTML elements are to be displayed on screen, paper, or in other media, CSS saves a lot of work. It can control the layout of multiple web pages all at once External stylesheets are stored in CSS files Cascading Style Sheets (CSS) is a style sheet language used for describing the presentation of a document written in a markup language such as HTML or XML (including XML dialects such as SVG, MathML or XHTML). CSS is a cornerstone technology of the World Wide Web, alongside HTML and JavaScript. CSS is designed to enable the separation of presentation and content, including layout, colors, and fonts. This separation can improve content accessibility; provide more flexibility and control in the specification of presentation characteristics; enable multiple web pages to share formatting by specifying the relevant CSS in a separate css file, which reduces complexity and repetition in the structural content; and enable the

css file to be cached to improve the page load speed between the pages that share the file and its formatting. Separation of formatting and content also makes it feasible to present the same markup page in different styles for different rendering methods, such as on-screen, in print, by voice (via speech-based browser or screen reader), and on Braille-based tactile devices. CSS also has rules for alternate formatting if the content is accessed on a mobile device.

Backend:

Python:

Python is a high-level, general-purpose programming language. Its design philosophy emphasizes code readability with the use of significant indentation. Python is dynamically typed and garbage-collected. It supports multiple programming paradigms, including structured, object-oriented and functional programming.

Database:

MySQL:

MySQL is an open-source relational database management system (RDBMS). Its name is a combination of "My", the name of co-founder Michael Widenius's daughter My, and "SQL", the abbreviation for Structured Query Language. A relational database organizes data into one or more data tables in which data may be related to each other; these relations help structure the data. SQL is a language programmers use to create, modify and extract data from the relational database, as well as control user access to the database. In addition to relational databases and SQL, an RDBMS like MySQL works with an operating system to implement a relational database in a computer's storage system, manages users, allows for network access and facilitates testing database integrity and creation of backups.

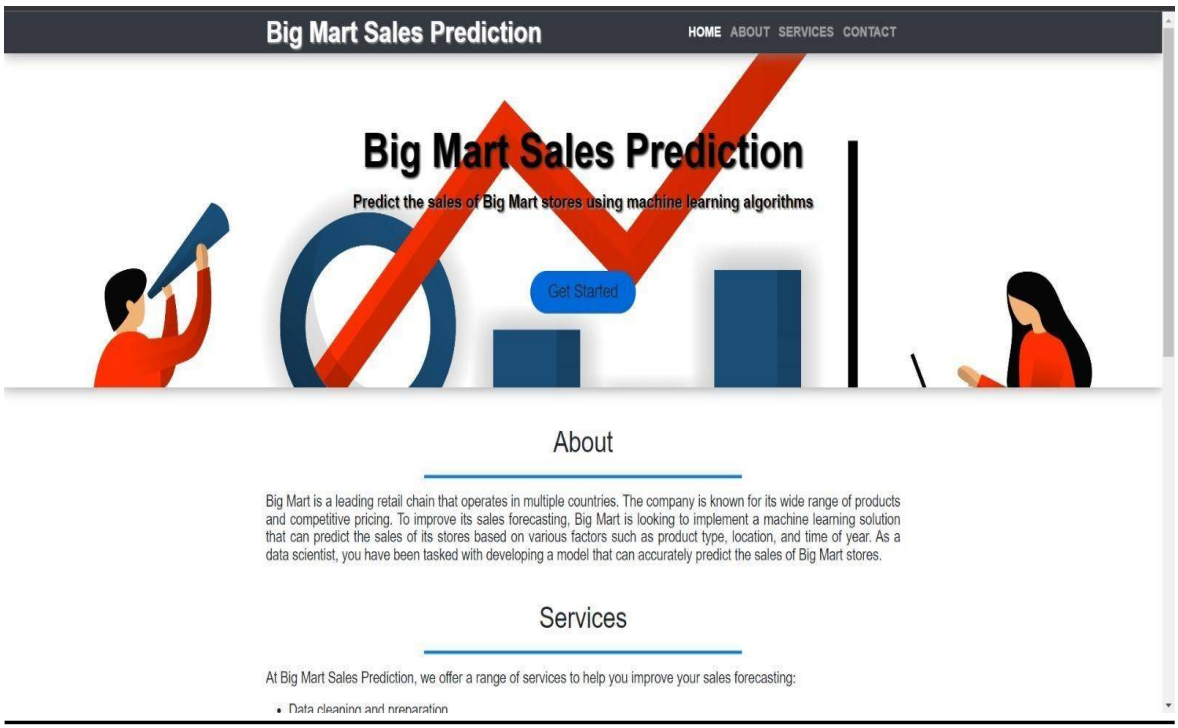
Chapter 7

Implementation:

Sr. No	Group Member	Time duration	Work to be done
1	Aryan Amin Manashree Chavan, Madhur Dukande, Mayur Jain	1st week of January	Implementing 1st module/ functionality)
		2nd week of January	Testing 1st module
2	Aryan Amin Manashree Chavan, Madhur Dukande, Mayur Jain	3rd week of January	Implementing 2nd module/ functionality)
3	Aryan Amin Manashree Chavan, Madhur Dukande, Mayur Jain	By the end of February month	Implementing 3rd module/ functionality)

Chapter 8

Implementation:



Sales Prediction Input

Enter Item Weight

Enter Item Visibility

Enter Item MRP

Outlet Establishment Year

Item Fat Content

Select an option

Item Type

Item Type

Outlet Size

Select an option

Outlet Location Type

Select an option

Outlet Type

Grocery Store

Submit Reset

Back to Home Page

Chapter 9

Result and Discussion

Result:

This project proposed a framework that predicts mart's sales using a machine learning model and different techniques. This project uses the data of various marts and then combines and analyses the data so that any mart can check the product's demand and sales overall. This forecasting helps the retailers to set the stock quantity more accurately. Next, we use many models to study the scores outcomes. Considering the outcomes, it is suitable for present data, but extensive data might be unsuitable or change the model selection. For more accuracy, we need a massive amount of data with minimum outliers. Therefore, the authors seek to check the individual product demand in a particular area in future work.

Chapter 10

Conclusion and Future Scope

Conclusion:

This project proposed a framework that predicts mart's sales using a machine learning model and different techniques. This project uses the data of various marts and then combines and analyses the data so that any mart can check the product's demand and sales overall. This forecasting helps the retailers to set the stock quantity more accurately. Next, we use many models to study the scores outcomes. Considering the outcomes, it is suitable for present data, but extensive data might be unsuitable or change the model selection. For more accuracy, we need a massive amount of data with minimum outliers. Therefore, the authors seek to check the individual product demand in a particular area in future work.

Future scope:

Further, in the future, a retailer checks the score of a specific product by entering product attributes and its store's information, like location, and culture. Also, we consider an online App for the costumer's review regarding the stores and specific products for future work. Customers rank the stores by giving feedback, this helps the other customers to move on towards the stores. Sort of forum allows checking the demand for a specific store's specific product. Also, this kind of portal helps the retailer compare the stocks and scores of the public leader board.

REFERENCES

1. Rohit Sav, Pratiksha Shinde, Saurabh Gaikwad (2021, June). Big Mart Sales Prediction using Machine Learning.2021 International Journal of Research Thoughts (IJCRT).

2. Heramb Kadam, Rahul Shevade, Prof. Deven Ketkar , Mr. Sufiyan Rajguru (2018). A Forecast for Big Mart Sales Based on Random Forests and Multiple Linear Regression. (IJEDR).

3. Gopal Behere, Neeta Nain (2019). Grid Search Optimization (GSO) Based Future Sales Prediction for Big Mart. 2019 International Conference on Signal-Image Technology & Internet-Based Systems (SITIS).

4. Archisha Chandel, Akanksha Dubey, Saurabh Dhawale, Madhuri Ghuge (2019, April). Sales Prediction System using Machine Learning. International Journal of Scientific Research and Engineering Development.