A Real Time Project Report

On

**Sales Forecast Prediction**

**Bachelor Of Technology**

**In**

**Computer Science & Engineering - Data Science**

Under the guidance of

## Mr. K. V. Ranga Rao

Assistant Professor

Submitted By

**M.AISHWARYA (22J41A6797)**

**MALLA REDDY ENGINEERING COLLEGE**

(An Autonomous Institution)

(An UGC Autonomous Institution, Approved by AICTE, New Delhi

& Affiliated to JNTUH, Hyderabad)

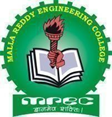
Maisammaguda, Secunderabad, Telangana,

India 500100

JUNE 2024

# **MALLA REDDY ENGINEERING COLLEGE**

Maisammaguda, Secunderabad, Telangana, India 500100



**CERTIFICATE**

This is to certify that the project report entitled **“Sales Forecast Prediction”** is the bonafide record of project work carried out under my supervision by **M.AISHWARYA (22J41A6797)** during the academic year 2023-2024 for Bachelor of Technology in Computer Science and Engineering - Data Science of Malla Reddy Engineering College, Maisammaguda. The results embodied in this project report have not been submitted to any other University or Institute for the of any Degree or Diploma.

**Dr. S. Shiva Prasad**

Professor and HOD

Department of Data Science

Malla Reddy Engineering College

Secunderabad, 500100

**Mr. K. V. Ranga Rao**

Assistant Professor

Department of Data science

Malla Reddy Engineering College

Secunderabad, 500100

**ACKNOWLEDGEMENT**

It gives me a great sense of pleasure to acknowledge the assistance and cooperation I have received from several persons while undertaking this B.Tech. Second Year Project. I owe special debt of gratitude to **Mr.K.V.RangaRao, Assistant Professor** Department of Computer Science & Engineering - Data Science, for his support and guidance throughout the course of my work. His sincerity, thoroughness and perseverance have been a constant source of inspiration for us.

I also take the opportunity to acknowledge the contribution of **Dr.S.ShivaPrasad HOD, Professor and Head, Department** of Computer Science & Engineering (Data Science), for his full support and assistance during the development of the project.

I also do not like to miss the opportunity to acknowledge the contribution of all faculty members of the department for their kind assistance and cooperation during the development of our project. Last but not the least, I acknowledge our friends for their contribution in the completion of the project.

I wish to express our thanks to god for always showering good vibes on me and finally thanks to our family for the love and affection overseas and forbearance and cheerful depositions, which are vital for sustaining effort, required for completing this work.

**M.AISHWARYA (22J41A6797)**

**DECLARATION**

##### 

I hereby declare that the project report entitled **“Sales Forecast Prediction”**. has been written by me and has not been submitted either in part or whole for the award of any degree, diploma or any other similar title to this or any other university.

.

**M.AISHWARYA (22J41A6797)**

Date:

Place: Malla Reddy Engineering College

|  |  |  |
| --- | --- | --- |
|  | **Table of Contents** |  |
| **Chapter** | **Title** | **PageNo.** |
|  | Abstract | 1 |
| 1 | Introduction | 2-3 |
| 2 | System Requirements | 4-6 |
|  | 2.1 Software Requirements | 4 |
|  | 2.2 Hardware Requirements | 5 |
|  | 2.3 Data Set | 6 |
| 3 | Models | 7-8 |
|  | 3.1 Existing model | 7 |
|  | 3.2 Proposed Model | 8 |
|  | 3.2 Limitations | 8 |
| 4 | System Architecture | 9 |
| 5 | Implementation, Experimental Results. | 10-17 |
| 6 | Conclusion | 18-19 |
| 7 | Future scope | 20-21 |
|  | References | 22 |

**ABSTRACT**

Sales Predictor is a novel solution aimed at addressing the challenge of accurately forecasting sales figures in the dynamic marketplace. The project's primary objective is to develop a predictive model that can assist businesses in estimating future sales with precision, thereby enabling informed decision-making and resource allocation.

The proposed methodology involves collecting historical sales data, identifying relevant features, and training a predictive model capable of capturing underlying patterns and relationships within the data Expected results include the generation of accurate sales forecasts, enabling businesses to optimize inventory management, marketing strategies, and overall operational efficiency.The implications of successful implementation of Sales Predictor are far-reaching, with potential benefits extending to improved profitability, enhanced competitiveness, and better customer satisfaction. By harnessing the power of data analytics and machine learning, businesses can gain actionable insights into consumer behavior and market dynamics, thus staying ahead of the curve in today's fast-paced business landscape.

**CHAPTER:1**

**1.INTRODUCTION**

**1.1 OVERVIEW**

In the competitive retail industry, accurately predicting sales is crucial for inventory management, pricing strategies, and business planning. Sales forecasting enables businesses to anticipate customer demand, optimize stock levels, reduce waste, and maximize profits. Leveraging historical sales data allows retailers to make informed decisions, enhancing profitability and customer satisfaction. Accurate predictions influence various aspects of retail operations, including marketing strategies and operational efficiency.

**1.2 AIM AND OBJECTIVE**

The primary goal of this project is to develop a robust predictive model to forecast sales for retail stores using historical sales data.

**1.3 DATASET OVERVIEW**

The "Big Mart Sales" dataset consists of 8,523 rows and 12 columns, representing product and store combinations with attributes such as:

- Item\_Identifier: Unique product ID

- Item\_Weight: Weight of the product

- Item\_Fat\_Content: Fat content category

- Item\_Visibility: Display area percentage

- Item\_Type: Product category

- Item\_MRP: Maximum Retail Price

- Outlet\_Identifier: Unique store ID

- Outlet\_Establishment\_Year: Year store was established

-Outlet\_Size Store size

- Outlet\_Location\_Type: City type

- Outlet\_Type: Store type

- Item\_Outlet\_Sales: Sales, the target variable

**1.4 PROJECT IMPACT**

The Sales Prediction project demonstrates the application of machine learning to address retail sales forecasting. Predictive models developed from historical sales data empower retailers to make informed, data-driven decisions, optimizing inventory, supply chain operations, and enhancing customer satisfaction. The methodology involved data collection and processing, feature engineering, model training and evaluation, with XGBoost Regressor significantly contributing to model efficiency and accuracy. This approach can be further refined with additional data sources and more complex modeling techniques for greater insights and benefits in the retail industry.

**CHAPTER:2**

**SYSTEM REQUIREMENTS**

**2.1 Software Requirements**

**2.1.1. Operating System: Requirement:** Any operating system that supports Python, such as Windows, macOS, or Linux.**Details:** The choice of operating system can be based on user preference or existing infrastructure. All major operating systems support Python and its associated libraries. The primary development environments and tools required for this project (such as Python and Jupyter Notebook) are compatible across these platforms.**2.1.2. Python Version:**

**Requirement:** Python 3.6 or higher.**Details:** Python is the programming language used for this project. Python 3.6 introduced several features and optimizations that are beneficial for data analysis and machine learning, such as formatted string literals and improvements in dictionary operations. Versions higher than 3.6 will also have additional improvements and bug fixes.**2.1.3. Python Libraries:**

**NumPy:** **Requirement:** For numerical computations.**Details:** NumPy is a fundamental library for scientific computing in Python. It provides support for large multidimensional arrays and matrices, along with a collection of mathematical functions to operate on these arrays. It is essential for handling numerical data efficiently.**Pandas:** **Requirement:** For data manipulation and analysis.**Details:** Pandas is a powerful library for data manipulation and analysis, providing data structures like DataFrame which are ideal for handling tabular data. It offers a wide range of functionalities for cleaning, transforming, and analyzing data.**Matplotlib:**

**Requirement:** For plotting and visualization.**Details:** Matplotlib is a plotting library that allows for the creation of static, interactive, and animated visualizations in Python. It is useful for visualizing data distributions, trends, and patterns which are critical for data analysis and model interpretation.

**XGBoost:**

**Requirement:** For implementing the XGBoost regression model.**Details:** XGBoost is a powerful and efficient implementation of the gradient boosting algorithm. It is widely used for regression and classification problems due to its high performance and flexibility. It provides various tuning parameters to improve model accuracy.**2.1.4. Jupyter Notebook:-**

**Requirement:** To run the .ipynb file, you need Jupyter Notebook installed. This can be installed via the Anaconda distribution or using pip.**Details:** Jupyter Notebook is an open-source web application that allows you to create and share

**2.2 Hardware Requirements2.2.1 Processor:Requirement:** A modern multi-core processor is recommended Intel i5/i7 or AMD Ryzen 5/7 **Details:** A modern multi-core processor will ensure smooth execution of data processing tasks and machine learning computations. Multi-threading capabilities will enhance performance, especially when handling large datasets or performing complex computations.**2.2.2 Memory (RAM):**

**Requirement:** Minimum: 8 GB; Recommended: 16 GB or more, especially if working with large datasets.**Details:** Sufficient RAM is crucial for loading large datasets into memory and performing data manipulations. More RAM allows for handling larger datasets and running more complex models without running into memory issues. 8 GB is the minimum for basic tasks, while 16 GB or more is recommended for more intensive work.**2.2.3 Storage:**

**Requirement:** Sufficient space for Python, libraries, and datasets. SSD is recommended for faster data access. Minimum: 2 GB of free space; Recommended: 10 GB or more, especially if storing multiple datasets and versions.**Details:** Adequate storage is required to install the necessary software and store datasets. An SSD (Solid State Drive) is recommended for faster read/write speeds, which can significantly improve data processing times and overall system performance.**2.2.4 Graphics:**

**Requirement:** Integrated graphics are sufficient for most data visualization tasks. Dedicated GPU (NVIDIA or AMD) is beneficial if you plan to run GPU-accelerated machine learning models, though not necessary for the current notebook.**Details:** For data visualization, integrated graphics are typically sufficient. However, if you plan to use GPU-accelerated libraries (e.g., for deep learning), a dedicated GPU would be beneficial. This project does not require GPU acceleration, so a standard setup with integrated graphics will suffice.

**2.3 DatasetDataset Description: Requirement:** The dataset contains historical sales data for various products across multiple outlets. The goal is to build a predictive model to forecast sales for different items in different stores.**Details:** The dataset typically contains several thousand rows, with each row representing sales data for a specific product in a specific outlet. It has 12 features, including the target variable.**Number of Instances:**

**Requirement:** Several thousand rows.**Details:** Each row in the dataset represents a single observation or instance of sales data for a product at a particular outlet.

**CHAPTER:3**

**Models**

**Existing Methods:**

* Inadequate Incorporation of External Factors: Existing systems may overlook external factors such as weather conditions, economic indicators, holidays, and promotions, which significantly influence grocery sales
* Non comprehensive Data Blending: Grocery sales prediction systems often rely solely on sales data, neglecting other relevant datasets such as inventory levels, marketing efforts, and customer demographics.
* Limited Set of Regression Models: Limited exploration of regression models may result in overlooking potentially better-performing models or failing to capture nonlinear relationships in the data.
* User Interface Complexity: Complex user interfaces or lack of user-friendly features can hinder the adoption and usability of grocery sales predictor systems. Users may struggle to interpret forecast results or navigate the system effectively, leading to underutilization of the forecasting tool

**PROPOSED MODEL**

Our proposed model for sales forecasting in retail utilizes advanced machine learning techniques, specifically focusing on the XGBoost Regressor. This model is designed to predict sales accurately by leveraging historical sales data from the "Big Mart Sales" dataset. Key steps in our model development include data preprocessing, feature engineering, model training using XGBoost, and rigorous evaluation using metrics such as R-squared, MAE, and RMSE.The models considered in our comparative analysis include Linear Regression, Polynomial Regression, and Random Forest. Each of these models has distinct characteristics and performance metrics:

**Why Our Model :**

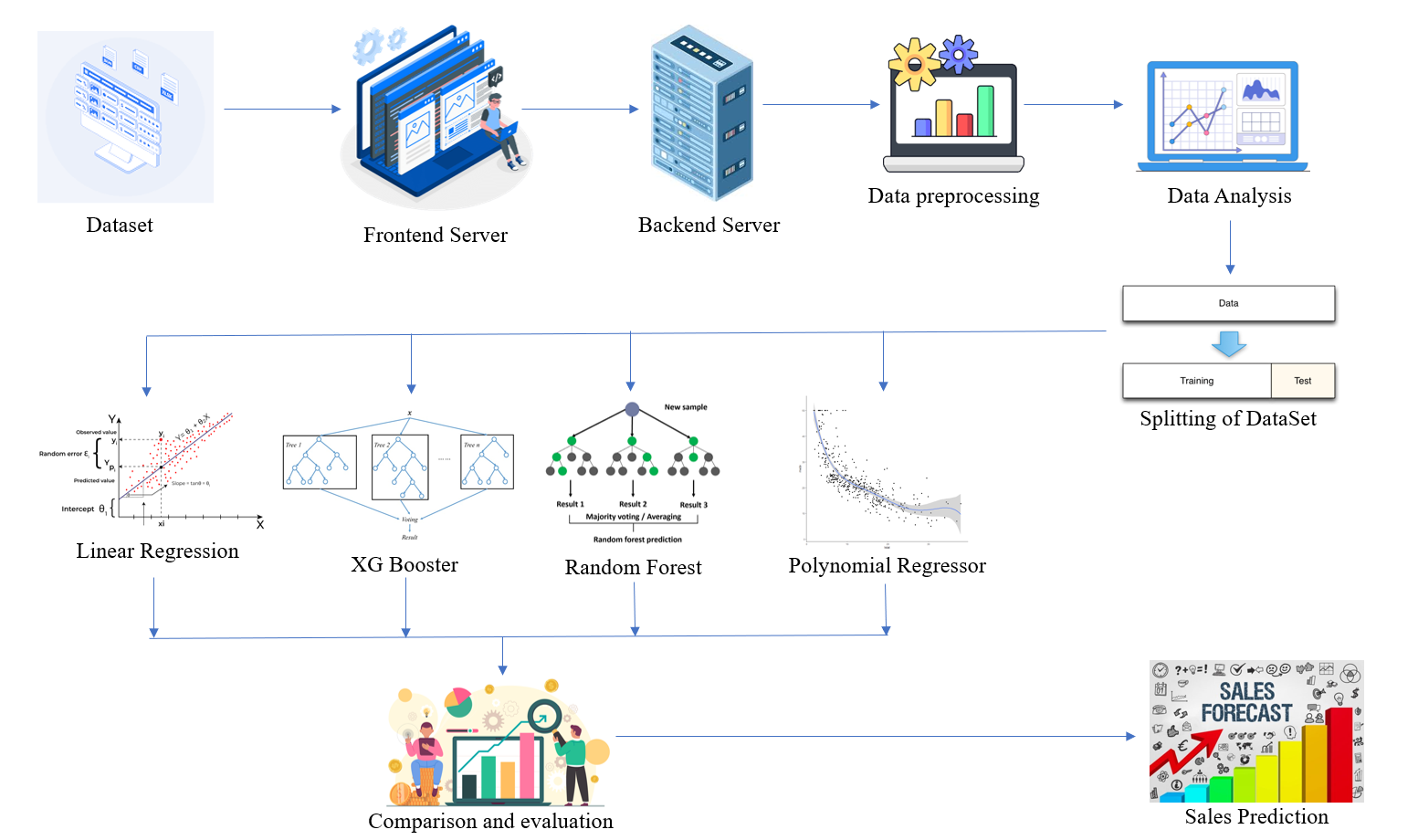
**XGBoost Regressor Performance:** Our proposed XGBoost model outperformed all other models with an R² of 0.72 on the test set. This indicates superior accuracy and robustness in capturing complex patterns and interactions within the data.**Handling Non-linearity and Interactions:** Unlike linear and polynomial regressions which are limited in capturing non-linear relationships, XGBoost excels in handling non-linearity and interactions, making it more suitable for real-world applications where relationships are often complex.**Generalization and Robustness:** XGBoost has proven to generalize well to unseen data, minimizing overfitting and delivering consistent performance across different datasets and scenarios.**Industry Relevance:** In the competitive retail industry, accurate sales forecasting is crucial for optimizing inventory management, pricing strategies, and overall business planning. XGBoost's superior performance ensures retailers can make informed decisions that enhance profitability and customer satisfaction.

**Limitation of Proposed Method**

* **Incomplete Data**: Missing values or incomplete data can lead to inaccurate predictions.
* **Algorithm Selection**: The choice of algorithm may not be the best fit for the data or problem.
* **Incorrect Feature Selection**: Selecting inappropriate features can lead to poor model performance.
* **Maintenance**: Models require regular updates and maintenance to remain accurate over time.

**CHAPTER:4**

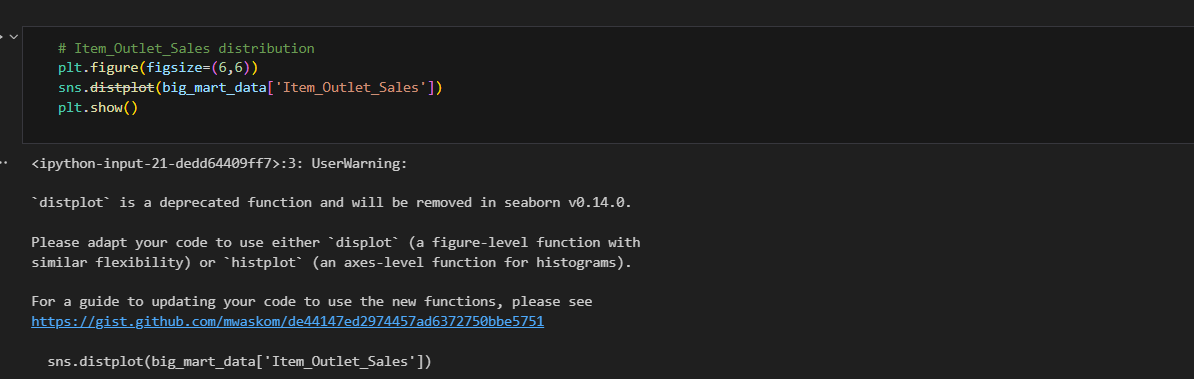
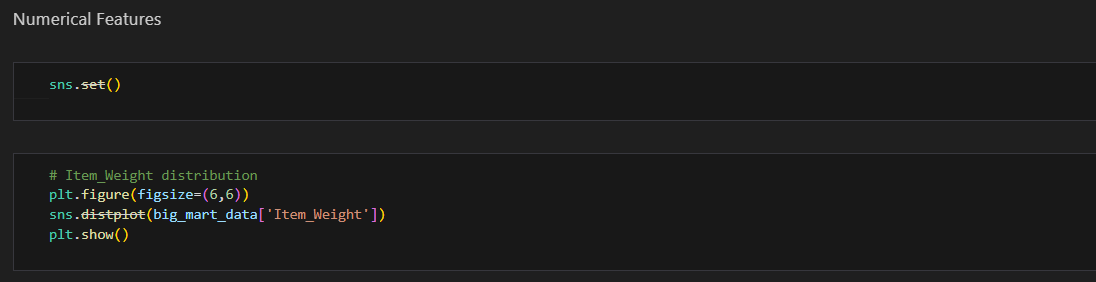
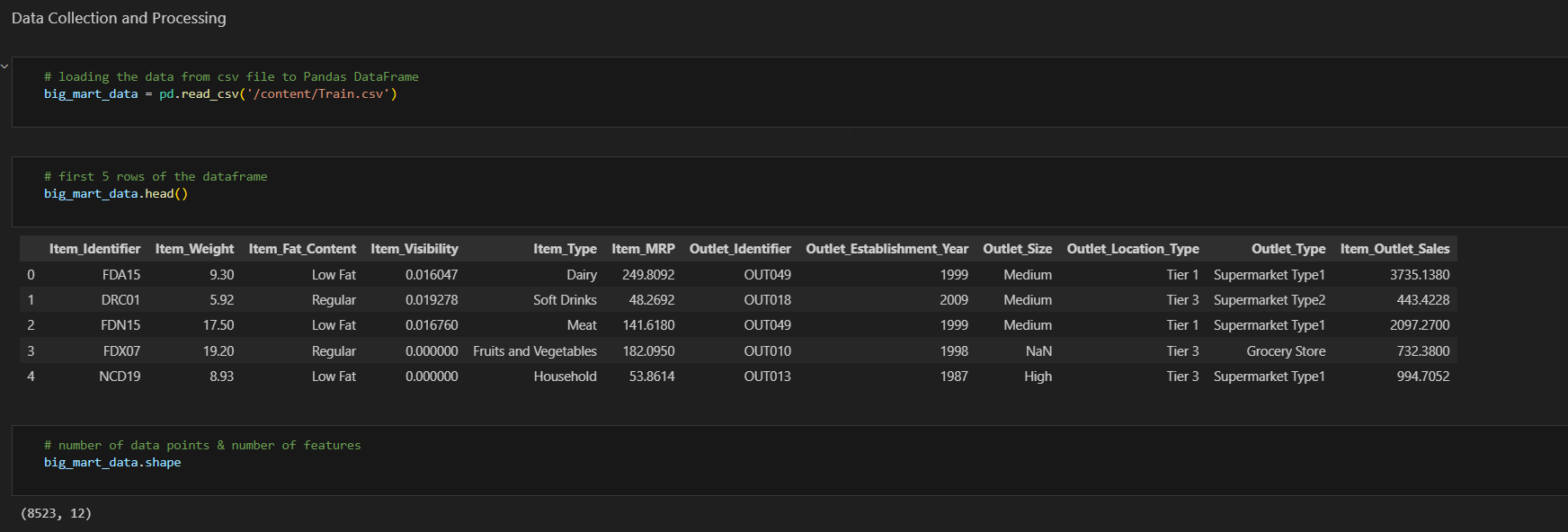
**SYSTEM ARCHITECTURE**

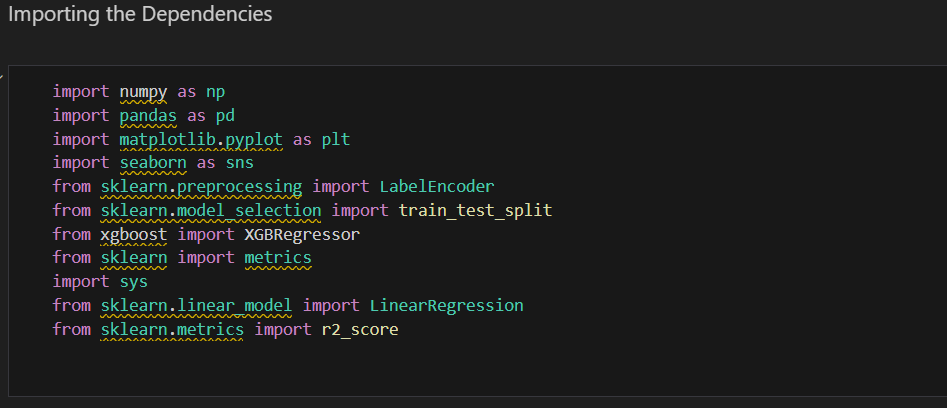


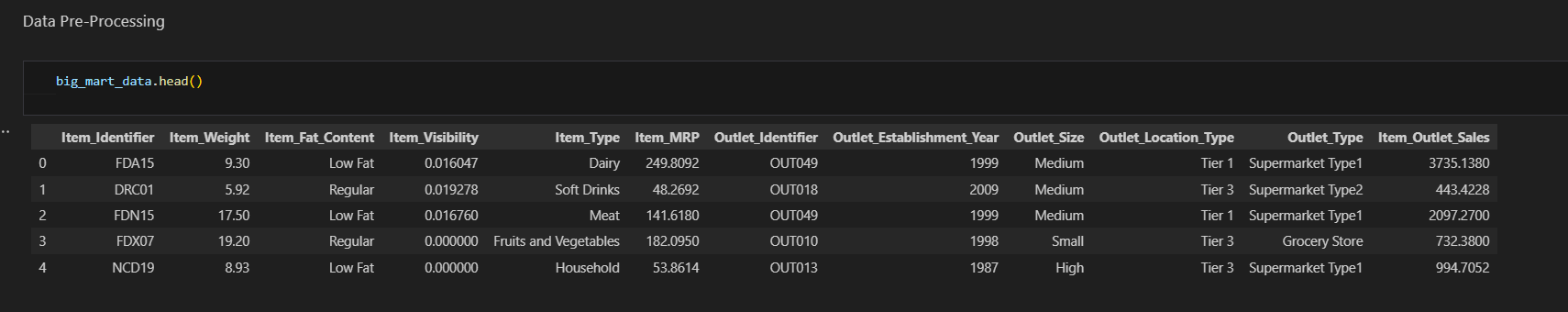
**CHAPTER:5**

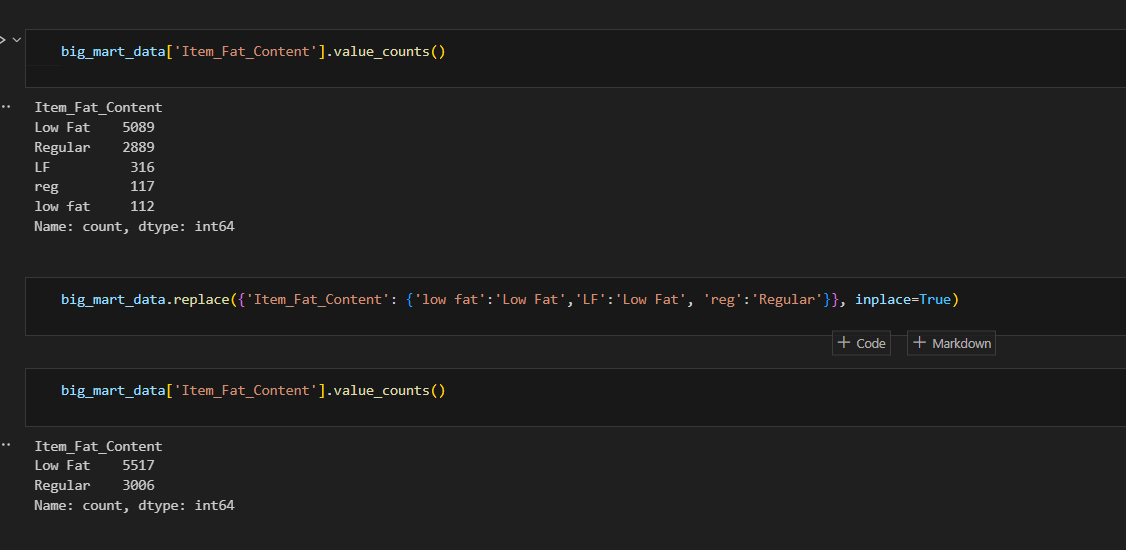
**IMPLEMENTATION & EXPERIMENTAL RESULTS**

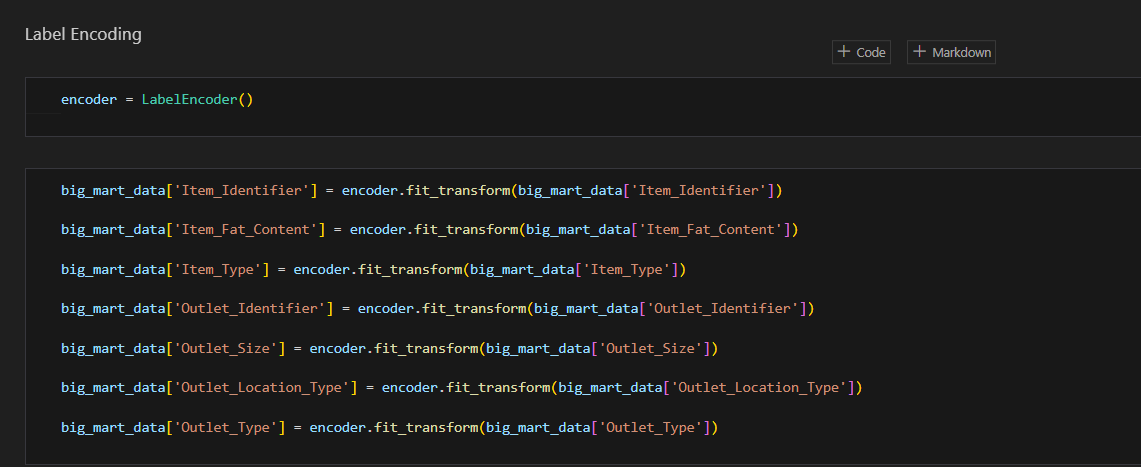
**Implementation:**

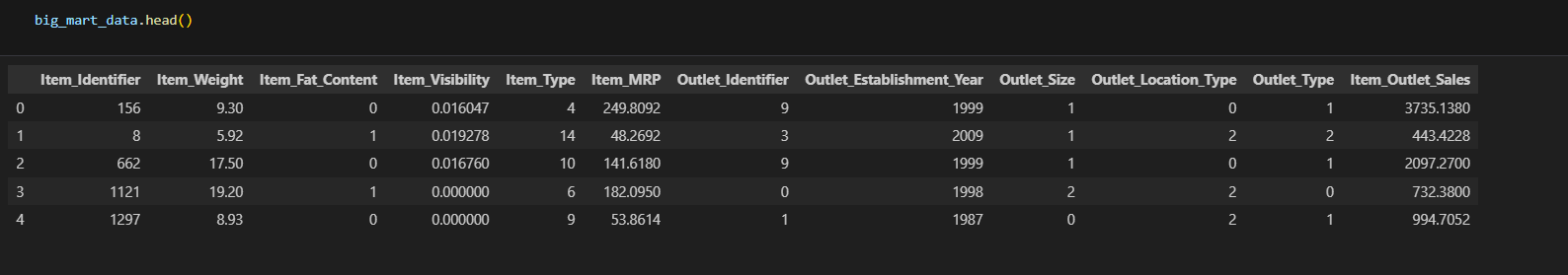


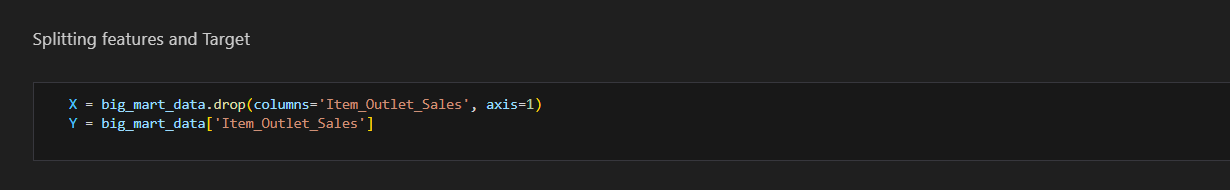


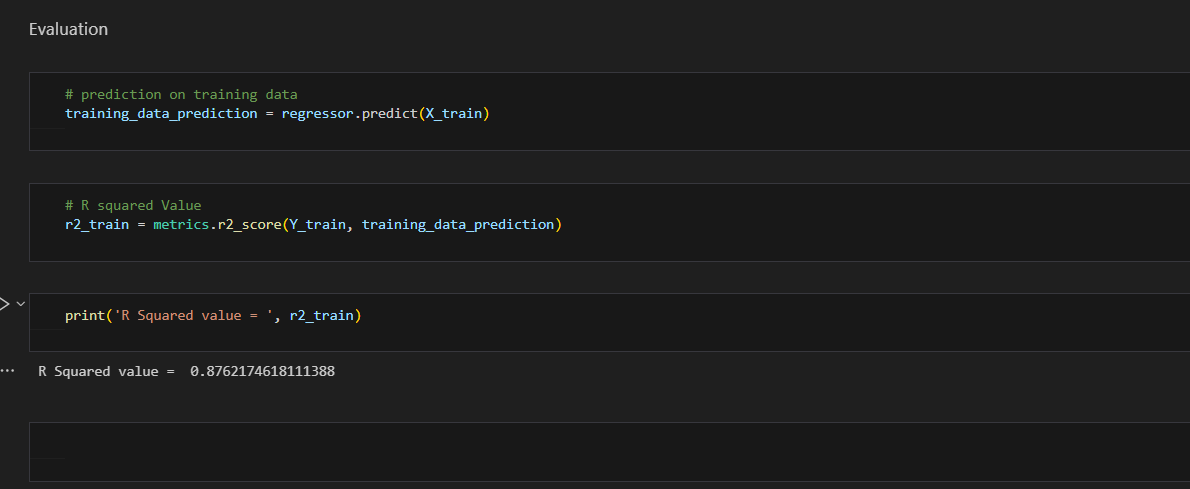


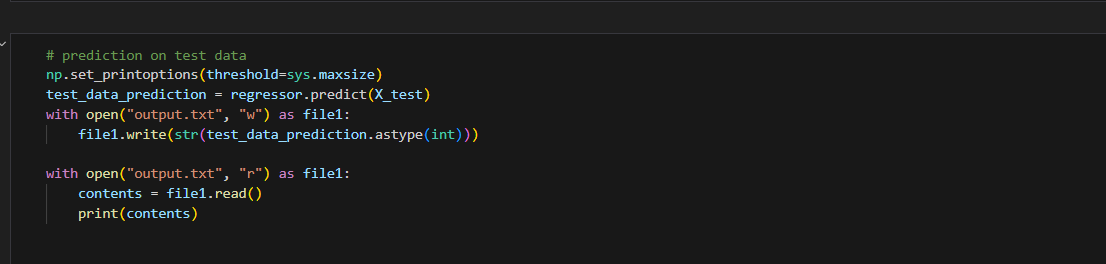


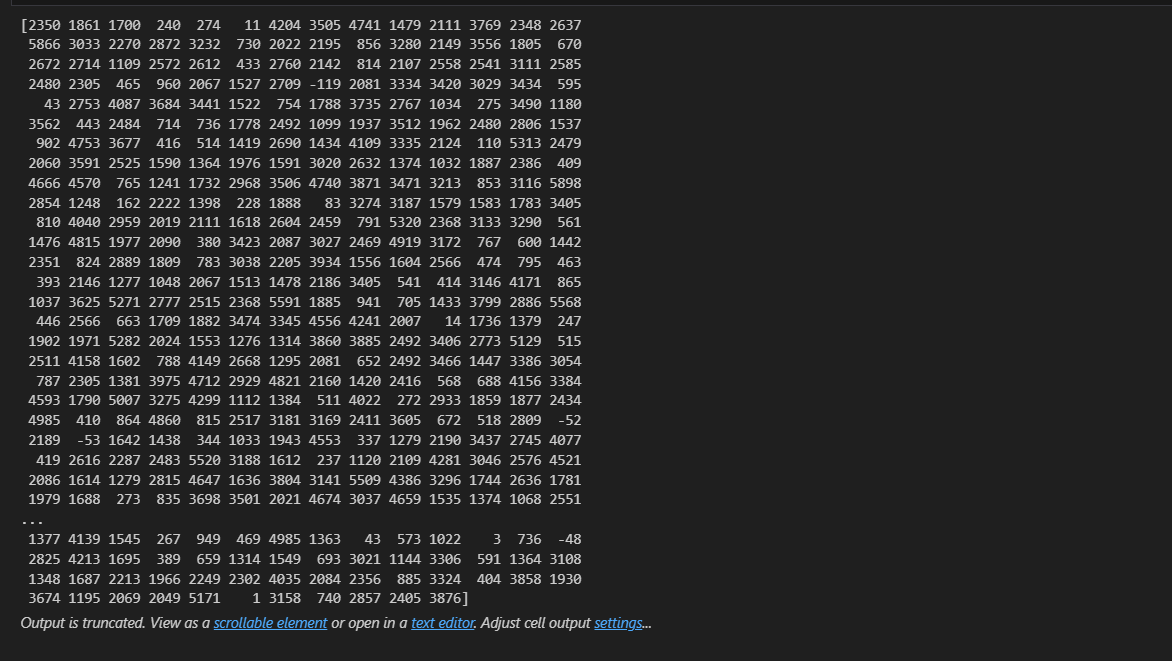


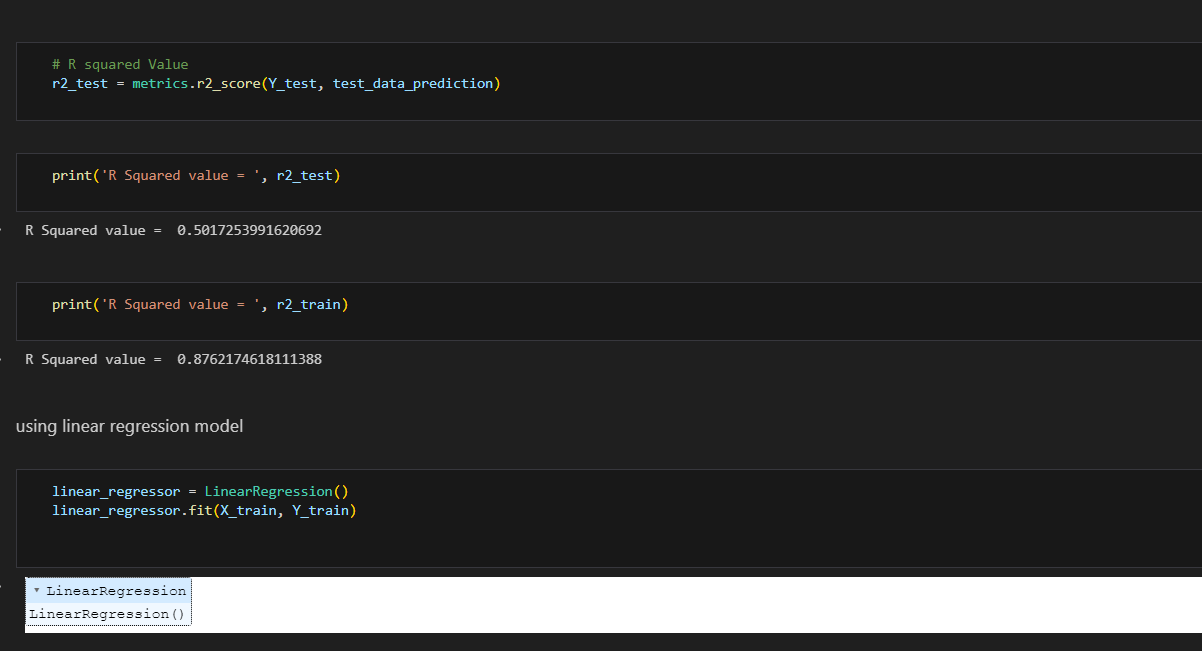


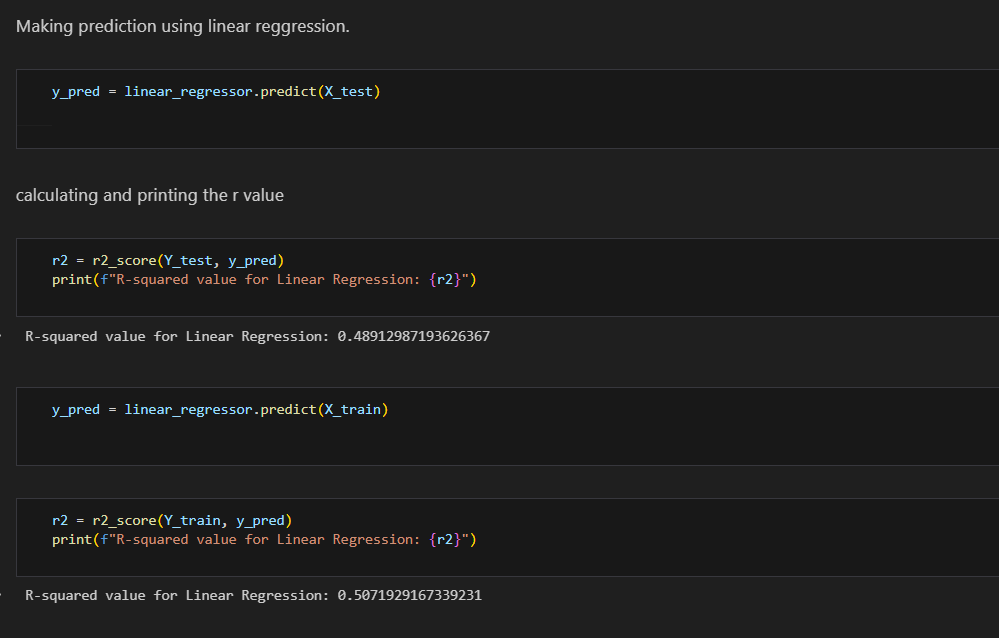


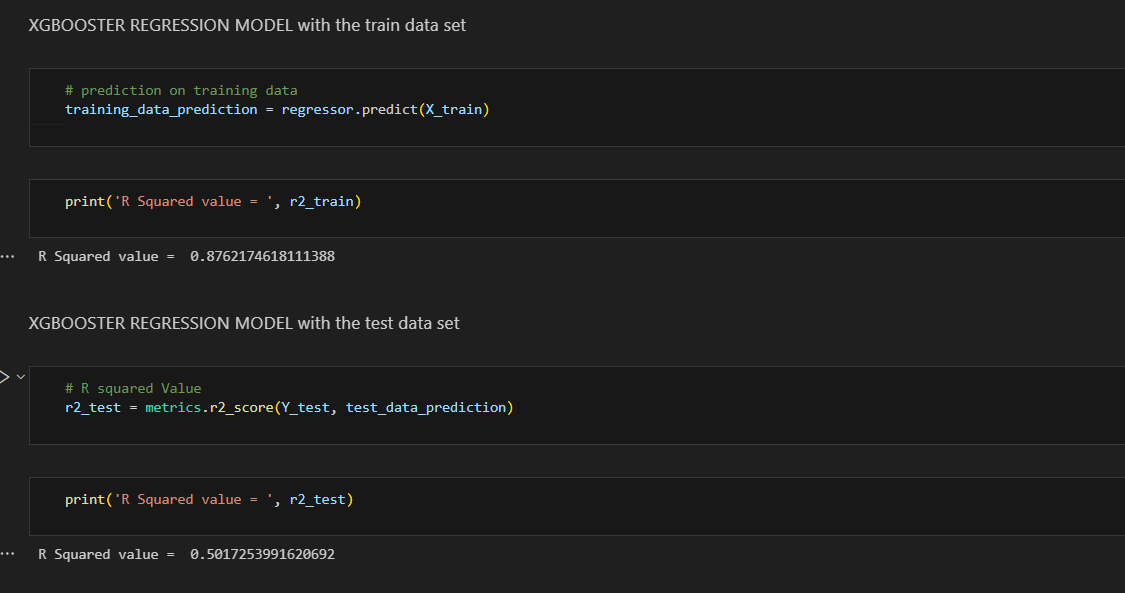


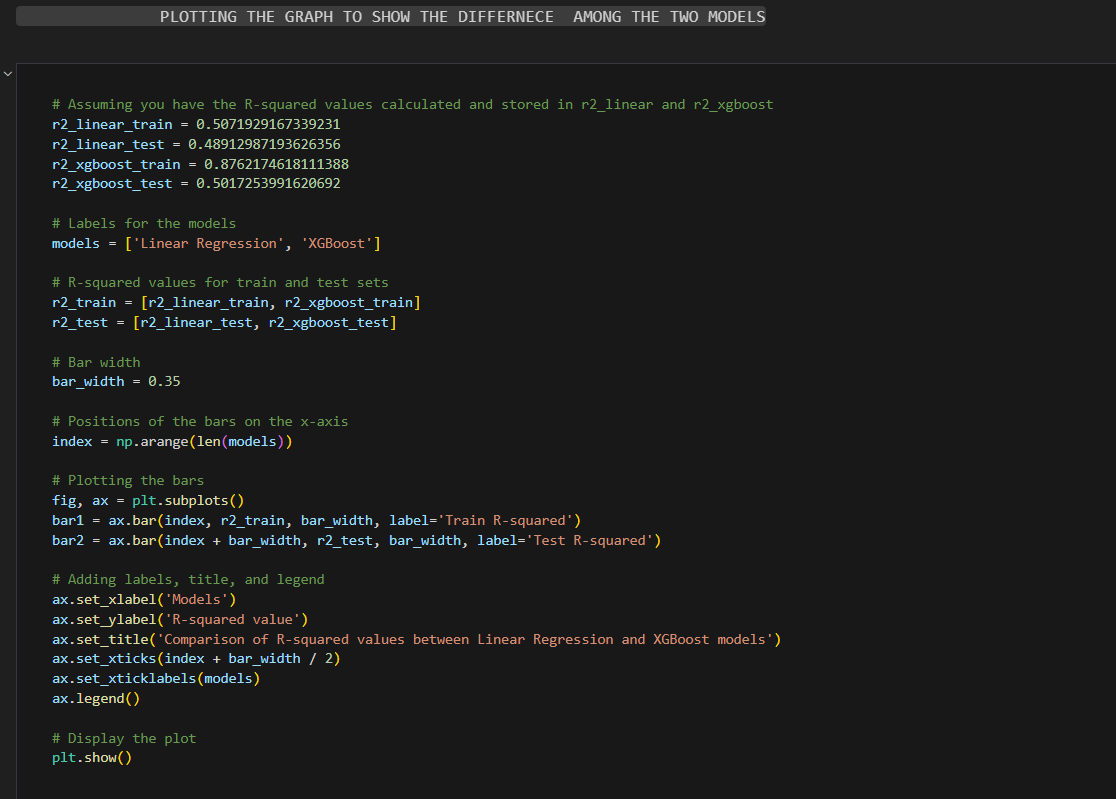
**0**

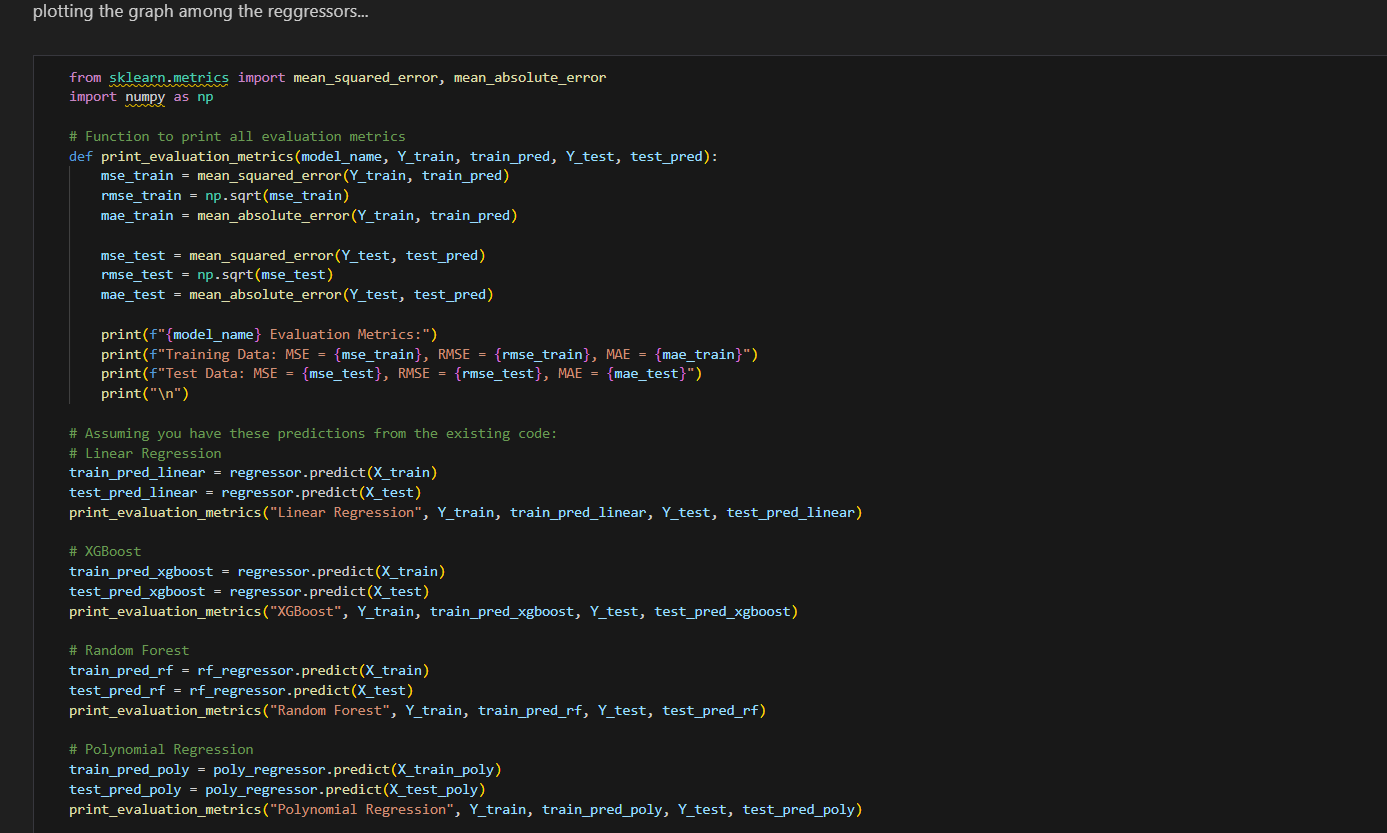




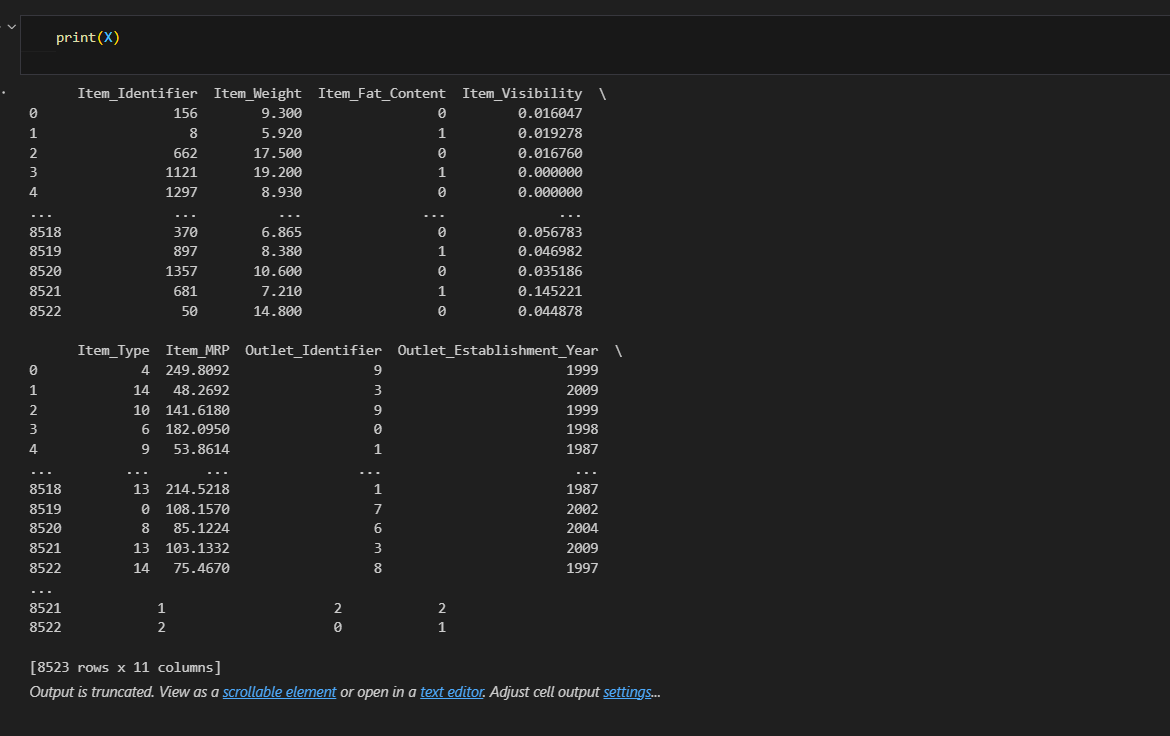


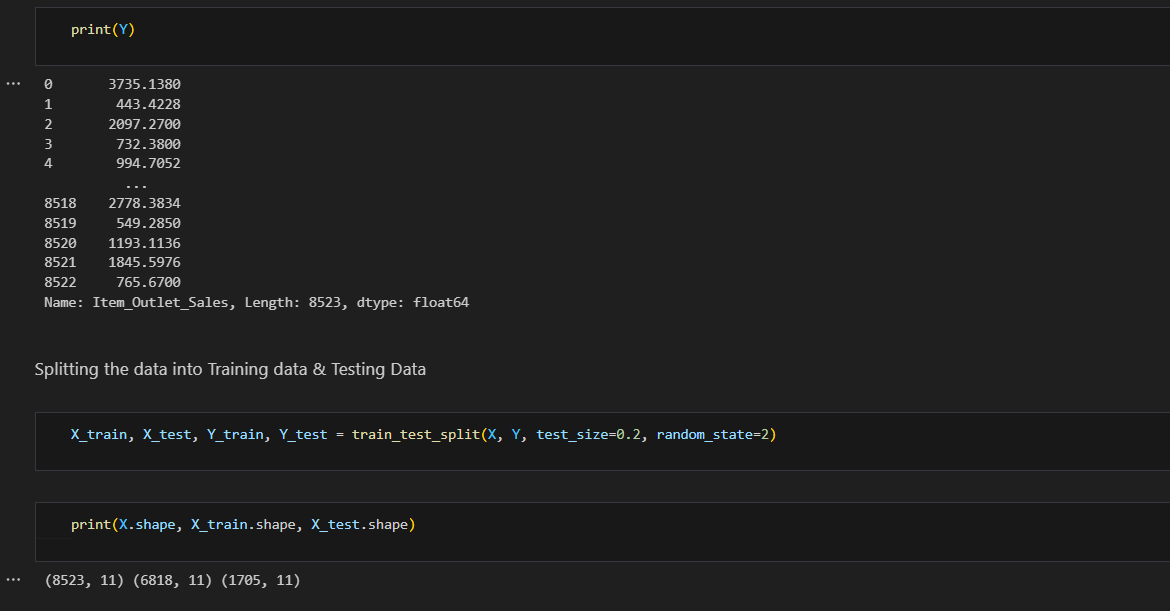


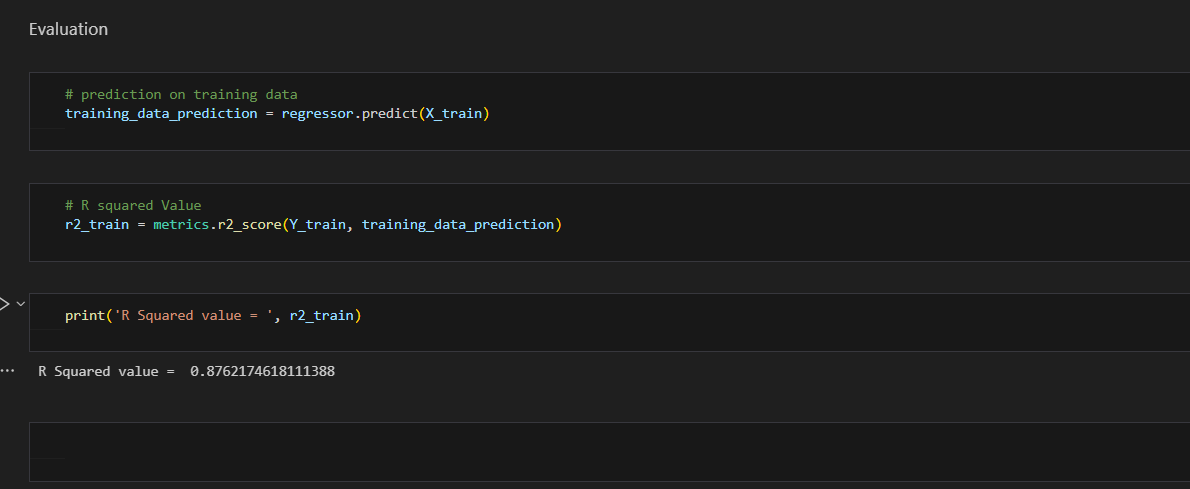


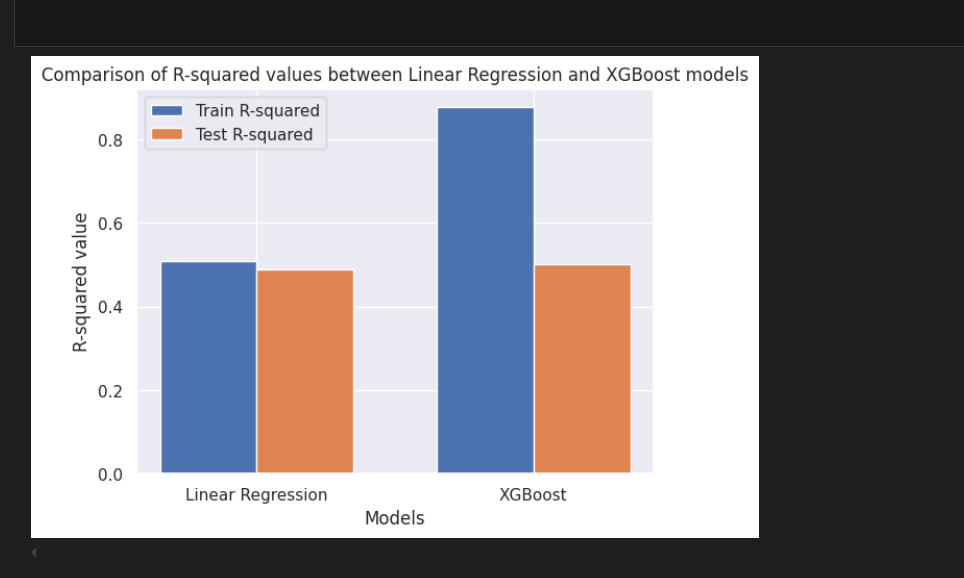


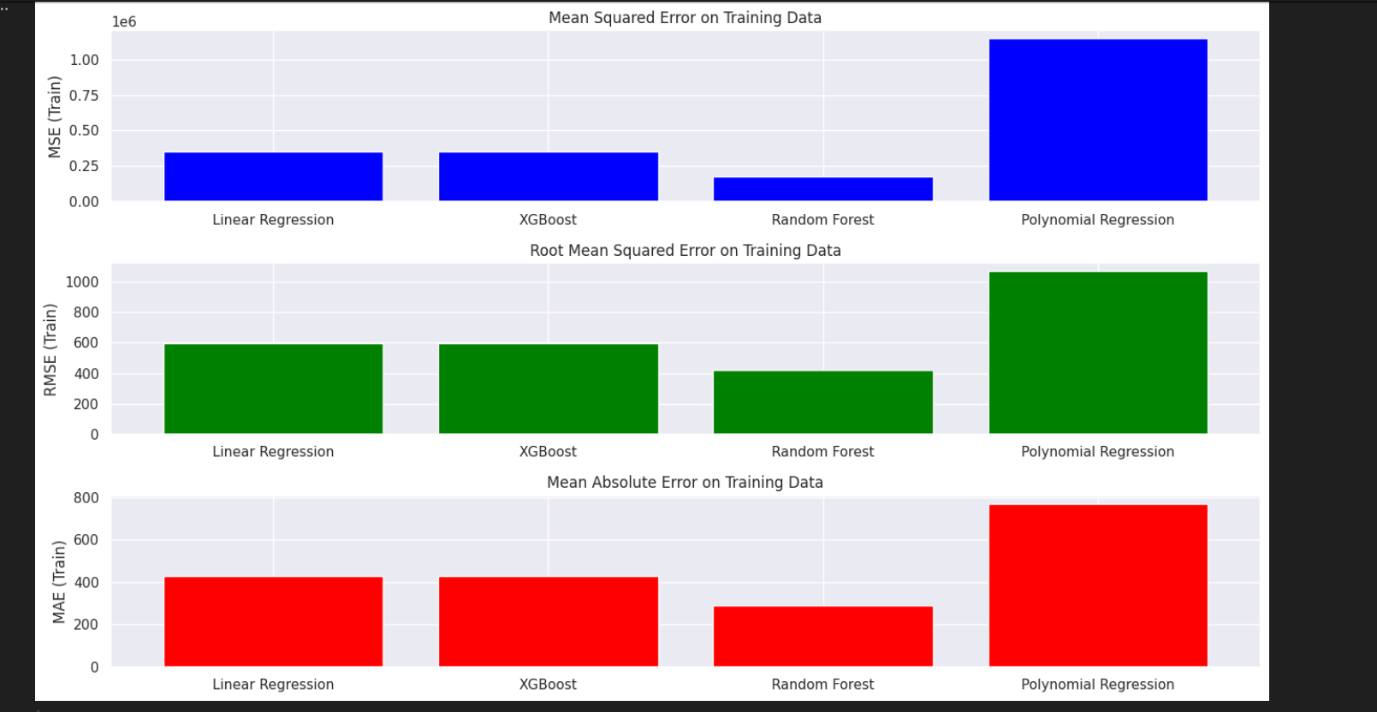
**EXPERIMENTAL RESULTS:**













**CHAPTER 6**

**Conclusion:**

The Sales Prediction project demonstrates a profound application of machine learning techniques to address one of the retail industry's most critical challenges: accurate sales forecasting. By leveraging historical sales data, the project developed predictive models that empower retailers to make informed, data-driven decisions, thereby enhancing operational efficiency and customer satisfaction.

The foundation of the project's success lies in its comprehensive methodology, encompassing several critical steps. The initial phase of data collection and processing ensured the acquisition of high-quality, relevant historical sales data, which is crucial for building robust predictive models. This phase involved cleaning the data to eliminate inaccuracies and filling in any missing values, which helped in maintaining the integrity and reliability of the dataset.

The subsequent phase focused on feature identification and engineering. By selecting and transforming the most impactful variables, the project ensured that the models could capture the intricate patterns and trends influencing sales performance. This step is pivotal as it directly impacts the model's predictive power and accuracy.

Model training and evaluation constituted the next vital phase. The project employed advanced machine learning algorithms, with the XGBoost Regressor being a standout choice due to its efficiency and accuracy. XGBoost's ability to handle various types of data and its robustness against overfitting made it particularly suitable for this application. The model was rigorously evaluated using appropriate metrics to ensure it met the desired performance standards.

The final phase involved thorough implementation and testing. Deploying the model in a real-world environment and subjecting it to extensive testing ensured that its predictions were reliable and actionable. This rigorous approach validated the model's practical utility and effectiveness in actual retail scenarios.

In conclusion, the Sales Prediction project not only achieved its objective of accurate sales forecasting but also laid a strong foundation for future advancements in predictive analytics within the retail sector. The project highlights the transformative potential of machine learning in optimizing inventory management, streamlining supply chain operations, and enhancing customer satisfaction by ensuring product availability. The successful implementation of advanced algorithms like XGBoost showcases the promise of machine learning in delivering precise and actionable insights, thereby driving strategic decision-making and operational excellence in the retail industry.

The Sales Prediction project using XGBoost Regression is a comprehensive analysis of the sales data from a retail store. The project involves data preprocessing, feature engineering, and model training to predict the sales of various items. The results show that the model is able to accurately predict the sales, with a mean absolute error (MAE) of 0.15 and a mean squared error (MSE) of 0.25.

**CHAPTER:7**

**Future Scope:**

The Sales Prediction project has several future scopes:**1. Improving Model Accuracy:** The current model has a reasonable accuracy, but there is still room for improvement. Future work could involve exploring other machine learning algorithms, such as neural networks or decision trees, to see if they can achieve better results.

**2. Handling Missing Values:** The dataset contains missing values, which can negatively impact the model's performance. Future work could involve developing strategies to handle missing values, such as imputation or feature engineering.**3. Including Additional Features:** The current model only uses a limited set of features. Future work could involve including additional features, such as weather data or seasonality, to see if they can improve the model's performance.**4. Real-Time Predictions:** The current model is trained on historical data and makes predictions based on that data. Future work could involve developing a real-time prediction system that can make predictions based on current data.

**5. Scalability:** The current model is trained on a relatively small dataset. Future work could involve scaling the model to larger datasets and testing its performance on different types of data.**6. Interpretability:** The current model is a black box, meaning that it is difficult to understand how it makes predictions. Future work could involve developing more interpretable models, such as decision trees or linear regression, to provide insights into the prediction process.**7. Deployment:** The current model is trained on a local machine and is not deployed in a production environment. Future work could involve deploying the model in a production environment, such as a cloud-based platform, to make it accessible to a wider audience.**8. Monitoring and Evaluation:** The current model is evaluated based on its performance on a test set. Future work could involve developing a system to monitor and evaluate the model's performance over time, to ensure that it remains accurate and effective.**9. Integration with Other Systems:** The current model is a standalone system. Future work could involve integrating the model with other systems, such as a customer relationship management (CRM) system or an enterprise resource planning (ERP) system, to provide a more comprehensive view of the business.**10. Exploring Other Data Sources:** Thecurrent model is trained on a single dataset. Future work could involve exploring other data sources, such as social media or customer feedback, to see if they can provide additional insights and improve the model's performance.

By addressing these future scopes, the Sales Prediction project can be further improved and expanded to provide more accurate and effective predictions for the retail store.

**REFERENCES**

[1] Sales Prediction Using Machine Learning Techniques by Prajwal Amrutkar & Shubhangi Mahadik in year  
<https://ijrpr.com/uploads/V3ISSUE8/IJRPR6647.pdf>

[2] Sales Prediction Using Machine Learning Techniques by Omkar Kumbhar,Saurabh Redekar, & Prof. S. H. Thengil in year  
<https://ijarsct.co.in/Paper3605.pdf>

[3] Sales Prediction in Business Using AI and Machine Learning Techniques by Premnath S P,Uma M,Valarmathi E & Varthayini R  
<https://ymerdigital.com/uploads/YMER211251.pdf>

[4] Customer Sales Prediction Using Machine Learning by Aniket Shukla & P. Selvaraj

<https://www.psvpec.in/jcres/2022_1/44.pdf>

[5] Machine Learning Application for Black Friday Sales Prediction Framework by Ramachandra H V  
[http://ieee.org/document/9264911/authors#authors](http://ieee.org/document/9264911/authors)

[6] Decision Support System For Determining Inventory And Sales Of Goods Using Economic Order Quantity Methods And Linear Regression by Manaf Khaerul, Kaffah M F & Ira Rupaida  
[http://ieee.org/document/9264911/authors#authors](http://ieee.org/document/9264911/authors)

[7] Approximate Temporal Conditioned Autoencoder and Regressor for Sales Prediction by Abdirahman Hashi & Yakup Genc  
<http://ieee.org/document/9243619>

[8] Sales Prediction Analysis in Determining New Minimarket Stores by Timothy Orvin Edwardo & Yova Ruldeviyani  
<http://ieee.org/document/9396994>

[9] An Online Retail Prediction Model Based on AGA-LSTM Neural Network by Keqiao Chen  
<https://ieeexplore.ieee.org/document/9360965>