

A Major Project Synopsis on

**BIGMART SALES PRICE PREDICTION**

Submitted to Manipal University, Jaipur

Towards the partial fulfillment for the Award of the Degree of

**MASTER OF COMPUTER APPLICATIONS**

2023-2025

by

PRIYANKA CHOUDHARY

23FS20MCA00043

Under the guidance of

DR. REKHA JAIN

**Department of Computer Applications**

**School of AIML, IoT&IS, CCE, DS and Computer Applications**

**Faculty of Science, Technology and Architecture**

**Manipal University Jaipur**

**Jaipur, Rajasthan**

**2025**

|  |  |  |
| --- | --- | --- |
| **S.NO** | **TOPICS** | **PAGE NO.** |
| 1. | INTRODUCTION | **2-3** |
| 2. | MOTIVATION | **3** |
| 3. | PROBLEM STATEMENT | **4-5** |
| 4. | METHODOLOGY/PLANNING OF WORK | **5-6** |
| 5 | REQUIREMENTS FOR PROPOSED WORK | **6-7** |
| 6. | BIBLIOGRAPHY/REFERENCES | **7-8** |
| 7. | CONCLUSION | **8** |

1. **INTRODUCTION**

BigMart Sales Price Prediction is a data-driven approach to forecasting sales for retail products across multiple stores. This project leverages machine learning techniques to analyze historical sales data and predict future sales trends.

The data includes product attributes such as weight, visibility, type, and outlet information such as location, size, and type. The target variable is the sales price of each product.

It provides insights of how Machine Learning can help to predict sales prices for BigMart products. The target variable is the sales price of each product.

The data can be used to build predictive models to predict sales for future time periods. This information can be used by Big Mart to make a variety of business decisions, such as:

• Pricing: Big Mart can use the models to set optimal prices for its products in each store. This can help to maximize profits and increase sales.

• Inventory management: Big Mart can use the models to forecast sales and ensure that it has the right amount of inventory in stock to meet demand. This can help to reduce costs and improve customer satisfaction.

• Marketing: Big Mart can use the models to identify which products are selling well and which products are not. This information can be used to develop targeted marketing campaigns to promote specific products.

The data can also be used to gain insights into customer behaviour and shopping trends. For example, the data can be used to identify which products are most popular in different stores and which products are often bought together. This information can be used to improve the store layout and product placement. Overall, the Big Mart sales prediction project data is a valuable resource that can be used to improve business decisions and increase sales.

1. **MOTIVATION**

The motivation for this project is to develop a model to predict sales at Big Mart, a large retail chain in India. The system is designed to address key challenges in sales forecasting for retail businesses. It aims to:

1. Sales prediction allows businesses to make informed decisions about pricing and optimize stock levels and prevent overstocking or understocking.
2. The proposed project will use machine learning to develop a model to predict sales at Big Mart. The model will be trained on historical sales data and will take into account a variety of factors, such as product attributes, outlet information, and past sales data. The goal of the model is to predict sales for future time periods with high accuracy.
3. Machine learning is a powerful tool that can be used to develop predictive models from data. Machine learning models can learn from historical data to identify patterns and relationships. This knowledge can then be used to predict future values, such as sales.
4. **PROBLEM STATEMENT**

* **For Store Owners**
  1. Difficulty in predicting sales trends accurately.
  2. Inefficient inventory management leads to losses.
  3. Lack of data-driven decision-making tools.
* **For Customers**
  1. Frequent stock outs of popular products.
  2. Inconsistent pricing and discounts.

It is crucial to understand the impact of various product attributes and store-specific factors on sales. The primary problem we aim to tackle is as follows: "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, our project seeks to construct a predictive model that can provide insights into the key factors influencing sales for every store. The model will also offer recommendations on potential changes to product characteristics or store attributes that can enhance sales performance. The primary questions we aim to answer through this project are:

* + - 1. What are the key attributes, such as weight, visibility, and type, that significantly influence sales prices for Big Mart products?
      2. How do store-specific attributes, including location and size, impact sales performance?
      3. Can a predictive model be developed to forecast sales prices accurately, and how can it be leveraged to optimize pricing strategies and boost revenue for Big Mart? This problem statement forms the foundation of our project, driving us to explore the relationships within the provided data, apply machine learning techniques, and ultimately provide actionable insights to enhance Big Mart's sales strategies and profitability in the highly competitive retail market.

1. **METHODOLOGY/ PLANNING OF WORK**
2. **Data Collection and Preprocessing**
   * Gather historical sales data from BigMart.
   * Clean and preprocess the data to remove inconsistencies.
   * Handle missing values and perform feature engineering.
3. **Exploratory Data Analysis (EDA)**
   * Identify patterns and relationships between sales, store types, and product categories.
   * Visualize trends using graphs and statistical tools.
4. **Model Selection and Training**
   * Use regression-based models such as:
     + Linear Regression
     + Random Forest Regressor
     + XGBoost
   * Train models using historical sales data.
   * Optimize hyperparameters for better accuracy.
5. **Model Evaluation**
   * Test models using validation datasets.
   * Evaluate performance using metrics such as Mean Squared Error (MSE) and R-Squared.
6. **Deployment and User Interface**
   * Develop a web-based dashboard using React.js.
   * Integrate backend services using Flask/Django.
   * Allow users to input store and product details for sales predictions.
7. **REQUIREMENTS FOR PROPOSED WORK**

**Software Requirements**:

* Python: Python is a popular programming language for machine learning due to its rich ecosystem of libraries and frameworks. EnsurePython installed on your system.
* Integrated Development Environment (IDE): You can use IDEs like Jupyter Notebook, Jupyter Lab, Google Collab or Visual Studio Code for data exploration, analysis, and model development.
* Data Manipulation and Analysis Libraries: You will need libraries like NumPy and Pandas for data handling, manipulation, and exploratory data analysis.
* Data Visualization Libraries: Libraries such as Matplotlib and Seaborn are useful for creating visualizations to understand the data better.
* Machine Learning Libraries: The primary libraries for machine learning in Python are scikit-learn and TensorFlow (for deep learning tasks).
* Model Evaluation Libraries: Libraries like scikit-learn provide functions to evaluate the performance of machine learning models.

**Hardware Requirements:**

* **Central Processing Unit (CPU):** A multi-core processor is essential for handling the data preprocessing, feature engineering, and model training tasks efficiently.
* **Random Access Memory (RAM):** Sufficient RAM is crucial, especially when working with large datasets and complex machine learning models. At least 8GB or more is recommended.
* **Storage:** Adequate storage is required to store the dataset, model files, and any intermediate results. SSDs are recommended for faster data access. Minimum 20GB free space
* **Graphics Processing Unit (GPU) (Optional):** Using a GPU can significantly speed up model training, especially for Machine learning models. However, it is not strictly necessary for all machine learning tasks.

1. **REFERENCES /BIBLIOGRAPHY**

 **Pandas Documentation.** (2021). *Pandas: Python Data Analysis Library*. Retrieved from <https://pandas.pydata.org>

 **Matplotlib Documentation.** (2021). *Matplotlib: Python Plotting Library*. Retrieved from <https://matplotlib.org>

 **Scikit-learn Documentation.** (2021). *Linear Regression and Random Forest Regressor*. Retrieved from <https://scikit-learn.org>

 **Chai, T., & Draxler, R. R. (2014).** *Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) for Evaluating Model Performance*. Journal of Geophysical Research, 119(7), 1303-1317. DOI: 10.1002/2013JD021043.

 **Hastie, T., Tibshirani, R., & Friedman, J. (2009).** *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.

 **React Documentation.** (2021). *React: A JavaScript Library for Building User Interfaces*. Retrieved from <https://reactjs.org>

 **Flask Documentation.** (2021). *Flask: Web Development, One Drop at a Time*. Retrieved from https://flask.palletsprojects.com

 **Django Documentation.** (2021). *Django: The Web Framework for Perfectionists with Deadlines*. Retrieved from <https://www.djangoproject.com>

1. **CONCLUSION**

In Conclusion, Our Logical Analysis Has Centered around The Critical Data Preparation Steps For The Bigmart Sales Prediction Project. We've Successfully Addressed Missing Data By Replacing It With The Mean Or Mode Values, Ensuring That Our Dataset Is Complete And Ready For Further Analysis. We've Gained A Deep Understanding Of The Data By Identifying The Most Correlated Columns And Have Introduced New Attributes Through Feature Engineering, Which Improves The Performance Of The Machine Learning Models. Moreover, The Exploratory Data Analysis (EDA) Has Allowed Us To Uncover Essential Insights Into The Dataset's Characteristics.

Execution Working –

* With The Essential Preprocessing Steps , We were Well-prepared To Progress To The Next Phase Of Our Analysis. Further we applied One-hot Encoding To Categorical Variables, Transforming Them Into A Suitable Format For Machine Learning Algorithms.
* Our Careful Groundwork In Data Preparation Sets the Stage For The Successful Implementation Of Predictive Models Aimed At Accurately Forecasting Bigmart Sales Price Prediction.