**BigMart Sales Prediction Documentation**

**Project Title:**

**BigMart Sales Prediction using Machine Learning**

**Objective:**

**The goal of this project is to predict sales for different products in various BigMart outlets using machine learning models. We use XGBoost Regression with feature engineering and hyperparameter tuning to improve accuracy.**

**📌 Steps Followed in the Project:**

**1️ Data Preprocessing & Cleaning**

* **Fixed categorical inconsistencies in Item\_Fat\_Content by replacing:  
  ✅ "lf" → "Low Fat"  
  ✅ "reg" → "Regular"**
* **Handled missing values using KNNImputer to fill gaps based on similar data points.**

**2️ Feature Engineering**

* **Created new features:  
  ✅ Outlet\_Age = 2025 - Outlet\_Establishment\_Year (Captures store age effect on sales).  
  ✅ Item\_Weight\_MRP = Item\_Weight \* Item\_MRP (New interaction feature).  
  ✅ Applied Log Transformation to Item\_Weight\_MRP for better data distribution.**

**3️ Encoding & Scaling**

* **Used Label Encoding for categorical variables.**
* **Standardized numerical features using StandardScaler to improve model performance.**

**4️ Model Training & Hyperparameter Tuning**

* **Split the data into training and validation sets (80%-20%).**
* **Used XGBoost Regressor (XGBRegressor) for prediction.**
* **Performed GridSearchCV to tune hyperparameters for the best performance.**

**5️ Model Evaluation**

* **Evaluated the model using Root Mean Squared Error (RMSE).**
* **Ensured no negative sales predictions by applying np.maximum(0, predictions).**

**6️ Saving Predictions**

* **Created a submission file: submission\_optimized\_File1.csv containing predicted sales for each product.**

**🔹 Technologies & Tools Used:**

**✔ Python (pandas, numpy, sklearn, xgboost)  
✔ Machine Learning (XGBoost, GridSearchCV, Feature Engineering)  
✔ Data Preprocessing (Handling Missing Values, Encoding, Standardization)  
✔ Data Visualization & Analysis (Exploring Feature Impact on Sales)**

**📈 Results & Impact**

**📌 The model successfully predicts BigMart sales with optimized accuracy.  
📌 Hyperparameter tuning improved RMSE performance.  
📌 Feature engineering added valuable insights into sales trends.**

**🚀 Next Steps: Further optimization using RandomizedSearchCV and more advanced feature engineering.**