**BigMart Sales Prediction - Approach Note**

**Objective:**

The goal of this hackathon project was to develop a predictive model to estimate the sales of different products across multiple BigMart outlets using historical sales data. The prediction of Item\_Outlet\_Sales enables better decision-making in inventory management, demand forecasting, and pricing strategies.

**Data Understanding & Preprocessing:**

The dataset contained 8523 training instances and 5681 test instances, each with multiple product and store-related attributes. The key preprocessing steps undertaken included:

1. **Handling Missing Values:**
   * Item\_Weight: Missing values were imputed using the mean strategy.
   * Outlet\_Size: Missing values were filled using the most frequent category (mode).
2. **Feature Engineering:**
   * Item\_Identifier and Outlet\_Identifier were preserved for final submission but not used for model training.
   * Categorical variables (Item\_Fat\_Content, Item\_Type, Outlet\_Location\_Type, Outlet\_Type, Outlet\_Size, Outlet\_Identifier) were one-hot encoded using OneHotEncoder.
3. **Data Transformation:**
   * One-hot encoding was applied to categorical variables.
   * Ensured all feature names were converted to strings to avoid compatibility issues in model training.

**Model Building & Experimentation:**

Two models were trained and evaluated for performance:

1. **Linear Regression:**
   * Baseline model to understand the linear relationship between features and sales.
   * RMSE on validation data was calculated for performance evaluation.
2. **Random Forest Regressor:**
   * A more robust ensemble model that captures complex relationships between features.
   * RMSE was compared against Linear Regression, showing an improvement in accuracy.

**Final Prediction & Submission:**

* The trained Random Forest model was used for final test dataset predictions.
* The final output was formatted to include Item\_Identifier, Outlet\_Identifier, and Item\_Outlet\_Sales.
* The submission file was saved as submission.csv.