**Task 7: Sales Forecasting Description**

**1. Objective**

The objective of this task is to forecast Walmart’s future sales using historical data. Accurate sales forecasting helps businesses optimize inventory management, staffing, and promotional strategies.

**2. Dataset**

**Source:** Walmart Sales Forecasting Dataset (Kaggle)

**Files Used:**

* **train.csv** → Historical weekly sales data
* **test.csv** → Future weeks for which sales need to be predicted
* **features.csv** → Additional economic and promotional factors
* **stores.csv** → Store information (type, size)

**Main Features:**

* **Store** → Store identifier
* **Dept** → Department identifier
* **Date** → Weekly timestamp
* **Weekly\_Sales** → Actual sales amount (Target variable)
* **IsHoliday** → Indicator for holiday weeks
* **Temperature, Fuel\_Price, CPI, Unemployment** → External factors
* **MarkDown1–5** → Promotional discount information
* **Type, Size** → Store-level features

**3. Problem Type**

* **Supervised Learning**
* **Regression Problem** (continuous target: Weekly Sales)
* **Time Series Forecasting**

**4. Approach**

**(a) Data Preprocessing**

* Filled missing values in MarkDown columns with **0**.
* Extracted time-based features: **Year, Month, Week**.
* Encoded categorical variables (e.g., Type).
* Merged train/test with features.csv and stores.csv for enriched dataset.

**(b) Feature Engineering**

* **Lag features** (previous week/month sales).
* **Rolling averages** to capture trends.
* **Holiday effects** using IsHoliday.

**(c) Modeling**

* Implemented **LightGBM Regressor** (primary model).
* Used **RandomForest Regressor** as a fallback.
* Trained on historical sales (Weekly\_Sales) and forecasted future sales for the test set.

**5. Evaluation Metrics**

* **RMSE (Root Mean Squared Error)** → Measures prediction accuracy.
* **MAE (Mean Absolute Error)** → Measures absolute forecast error.
* Validation was performed on the last few months of training data.

**6. Visualization**

* **Actual vs Predicted Sales plots** for validation data.
* **Monthly/Store-wise sales trends.**
* **Holiday vs Non-Holiday comparisons.**

**7. Tools & Libraries**

* **Python**
* **Pandas, NumPy** → Data preprocessing and transformation
* **Matplotlib, Seaborn** → Data visualization
* **Scikit-learn** → Model evaluation and metrics
* **LightGBM / RandomForest** → Regression models

**8. Results**

* The model successfully captured seasonal and holiday sales patterns.
* **LightGBM** achieved better performance compared to RandomForest.
* Evaluation metrics (RMSE, MAE) showed the model’s effectiveness in forecasting.

**9. Applications**

* **Retail Industry:** Demand forecasting for different stores and departments.
* **Inventory Management:** Prevent overstock and stockouts.
* **Workforce Planning:** Adjust staffing levels during high-demand weeks.
* **Marketing Campaigns:** Plan promotions during sales peaks (holidays).



