**Retail Sales Prediction using Machine Learning**

**1. Project Title**

**Retail Sales Prediction using Machine Learning**

**2. Project Description**

This project aims to predict the **Total Sales Amount** from **retail transaction data** based on product quantity and unit price. We built multiple regression models, compared their performance, and finalized the best one for deployment-ready prediction.

**3. Dataset Details**

* **Filename**: online\_retail.csv downloaded form UCI site. (https://archive.ics.uci.edu/dataset/352/online+retail)
* **Records**: 541,909
* **Features**:
  + InvoiceNo
  + StockCode
  + Description
  + Quantity
  + InvoiceDate
  + UnitPrice
  + CustomerID
  + Country

**4. Project Pipeline**

| **Stage** | **Description** |
| --- | --- |
| **Data Loading** | Loaded using load\_data() function from data\_loader.py |
| **EDA** | Explored dataset structure, types, nulls, duplicates |
| **Data Preprocessing** | Removed missing/duplicate entries, feature engineering (TotalSales, DayOfWeek, Month) |
| **Modeling** | Built Linear Regression, Decision Tree, Random Forest models |
| **Evaluation** | Compared models based on MAE, MSE, RMSE, R² Score |
| **Prediction** | Predicted on random samples and plotted actual vs predicted |

**5. Feature Engineering**

* Created **TotalSales** = Quantity × UnitPrice
* Extracted **DayOfWeek** and **Month** from **InvoiceDate**

**6. Models Trained**

| **Model** | **MAE** | **MSE** | **RMSE** | **R² Score** |
| --- | --- | --- | --- | --- |
| Linear Regression | 10.70 | 2848.05 | 53.37 | 0.58 |
| Decision Tree Regressor | 0.16 | 127.68 | 11.30 | 0.98 |
| Random Forest Regressor | **0.14** | **72.24** | **8.50** | **0.99** |
|  |  |  |  |  |

**7. Final Model Selection**

* **Best Model**: Random Forest Regressor
* **Reason**: Achieved the lowest MAE, MSE, RMSE and the highest R² score (0.99)

**8. Prediction Results**

After training, the model was tested on unseen random samples.

| **Metric** | **Value** |
| --- | --- |
| MAE | 0.24 |
| MSE | 2.12 |
| RMSE | 1.45 |
| R² Score | 1.00 |

* The graph plotted showed the prediction line (red) almost perfectly passing through the actual data points (blue).

**9. Tools and Technologies**

* Python 3.12
* Pandas
* NumPy
* Matplotlib
* Scikit-learn (sklearn)

**10. Project Folder Structure**

Retail\_Sales\_Prediction/

│

├── src/

│ ├── data\_loader.py

│ ├── eda.py

│ ├── modeling.py

│ ├── predict.py

│ └── online\_retail.csv

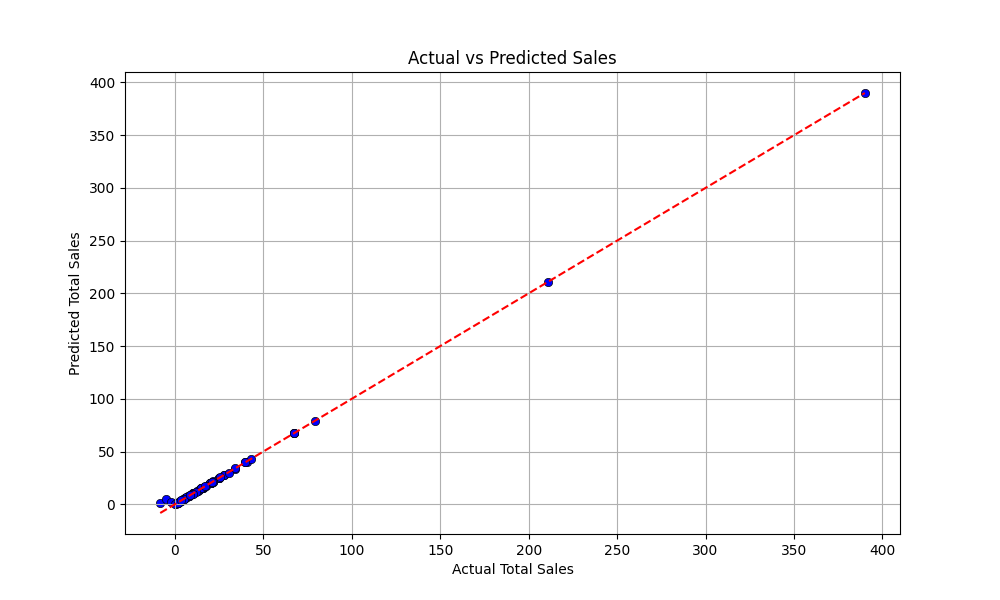
│

├── venv/ # Virtual environment (ignored)

│

└── README.md, .gitignore, etc.

11. PLOT



**📜 End of Documentation**