AI PROJECT REPORT – MODULE E

STUDENT & PROJECT DETAILS

Student Name: Upasana Das
Mentor Name: Dr. Niranjan Deshpande
Project Title: AI-Based Demand Trend Analysis of Retail Products Using Machine Learning

---

1. PROBLEM STATEMENT

Retail businesses rely heavily on accurate demand forecasting to manage inventory and optimize sales. However, predicting product demand is challenging due to variations in customer behaviour, pricing strategies, discounts, and seasonal trends. Inaccurate demand estimation can result in excess inventory, increased holding costs, or loss of sales due to stock shortages.

This project aims to analyse historical retail sales data to understand demand patterns and predict future demand using machine learning techniques. The system focuses on demand analysis and prediction to support data-driven decision-making in retail operations.

The objective of the system is prediction and analysis.
Key assumptions include availability of historical sales data and consistent data quality. The scope of the project is limited to structured retail sales data and does not include external factors such as economic conditions or marketing campaigns.

2. APPROACH

- **System Overview**
  The system follows a structured workflow that includes data loading, preprocessing, exploratory data analysis, feature selection, machine learning model development, and evaluation. The goal is to predict product demand based on historical sales-related attributes.

- **Data Strategy**
  The dataset used in this project consists of historical retail sales data stored in an Excel file (superstore_sales.xlsx). The data includes attributes such as sales value, quantity sold, discount, profit, shipping cost, and time-related information. Data preprocessing involves handling missing values, selecting relevant numerical features, and preparing the dataset for analysis and modeling.

- **AI / Model Design**
  A supervised machine learning approach is used for demand prediction. Linear Regression is implemented as a baseline predictive model to estimate sales demand based on selected features followed by a Random Forest Regressor as prototype for performance comparison. The dataset is split into training and testing sets to evaluate

model performance. Model training is performed using historical data, and predictions are generated on unseen test data.

- **Tools & Technologies**
  The project is implemented using Python. Pandas and NumPy are used for data manipulation and numerical computations. Matplotlib and Seaborn are used for data visualization and exploratory analysis. Scikit-learn is used for model implementation, data splitting, and performance evaluation. Jupyter Notebook is used as the development and experimentation environment.

- **Design Decisions**
  Linear Regression was chosen as an interpretable baseline model to understand relationships between sales and influencing factors. Feature selection was guided by correlation analysis and business relevance to improve model simplicity and interpretability. Random Forest Regressor is used as prototype model.

3. KEY RESULTS

The system successfully analyses historical sales data and identifies relationships between key variables such as sales, quantity, discount, and profit. Exploratory analysis highlights demand trends and feature correlations.

The Linear Regression model is able to predict sales demand with reasonable accuracy. Model performance is evaluated using Root Mean Squared Error (RMSE), which provides a quantitative measure of prediction error. The results demonstrate that machine learning can be effectively applied for basic demand forecasting in retail scenarios.

The system is designed as a baseline demand prediction model using historical sales data. While it provides useful insights, its performance is influenced by data quality and the selected feature set. More advanced models and additional data sources could further enhance prediction accuracy.

4. LEARNINGS

- **Technical Learnings**
  This project provided hands-on experience with data preprocessing, exploratory data analysis, correlation analysis, and supervised machine learning using Linear Regression and Random Forest Regressor. It also strengthened understanding of model evaluation techniques such as RMSE.

- **System & Design Learnings**
  Designing an end-to-end data analysis and prediction pipeline highlighted the importance of feature selection, data quality, and model interpretability in real-world applications.

- **Challenges Faced**
  Key challenges included handling real-world retail data and identifying relevant features for prediction. These challenges were addressed through exploratory analysis and correlation-based feature selection.

- **Future Improvements**
  Future enhancements could include using advanced machine learning models, incorporating time-series forecasting techniques, adding external factors such as promotions or seasonality, and deploying the model as a web-based application.

5. REFERENCES & AI USAGE DISCLOSURE

- **Dataset Used**
  Retail sales dataset (Superstore sales data)

- **Tools and Libraries Referenced**
  Python, Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, Jupyter Notebook

- **AI Tools Used**
  No generative AI tools were used for model training. AI assistance tools were used for guidance and documentation support.