Task-6: Time Series Analysis

Objective

Analyze sales data over time and forecast future sales using statistical models.

Project Steps

1. Dataset Selection

- The dataset should include at least two columns:
    - **Date:** The timestamp of the recorded sales.
    - **Sales:** The number of units sold or revenue on that date.

2. Tasks to Perform

1. **Visualize Sales Trends Over Time**
    - Plot sales data to identify patterns such as seasonality, trends, or irregularities.
    - Use line charts or moving averages to observe fluctuations over time.
2. **Use ARIMA for Forecasting**
    - Apply **AutoRegressive Integrated Moving Average (ARIMA)** to model the sales data.
    - Train the model and generate sales forecasts for future periods.
    - Validate the accuracy of the model using metrics like RMSE or MAPE.

3. Deliverables

- **Forecasted Sales Values:** A table or dataset with predicted future sales.
- **Plots:**
    - A **trend analysis plot** showing historical sales.
    - A **forecasting plot** displaying the predicted sales alongside actual sales for comparison.

Project Description: Predicting Heart Disease Using Logistic Regression

Objective

Predict whether a patient has heart disease based on key medical parameters.

Project Steps

1. Dataset Selection

- **Dataset Name:** `heart_disease.csv`
- **Columns:**
  - **Age** – Patient's age
  - **Gender** – Male or Female
  - **Cholesterol** – Cholesterol levels
  - **Blood Pressure** – Systolic/Diastolic pressure
  - **Heart Disease** – Target variable (Yes/No or 1/0)

2. Tasks to Perform

1. **Load the Dataset**
   - Check for missing or inconsistent data.
   - Clean the dataset by handling null values and duplicates.
2. **Feature Engineering**
   - Normalize or scale numerical features like **Age, Cholesterol, and Blood Pressure** to improve model performance.
3. **Model Training**
   - Train a **Logistic Regression model** to classify patients as having heart disease or not.
4. **Model Evaluation**
   - Assess model accuracy using a **confusion matrix** and classification metrics like:
     - Precision
     - Recall
     - F1-score

3. Deliverables

- **Logistic Regression Model:** Trained model to predict heart disease.
- **Evaluation Report:**
  - Confusion matrix
  - Accuracy, precision, recall, and F1-score insights

Would you like recommendations on how to improve model performance? 😊

Deadline Compliance

- **Restriction**: **Submit the project within 7 days** from the start date.
- **Reason**: Meeting deadlines is crucial in the real-world software development environment. This restriction helps students practice **time management** and **task prioritization**. In professional settings, tight deadlines are often the norm, and learning to meet them without compromising quality is an essential skill.
- **Learning Outcome**: Students will learn to manage their time effectively, complete projects under pressure, and **deliver results on time**, which are all important skills in the workplace.