**Name: Prachi Shedge**

**Roll No: 281050**

**Batch: A2**

**Assignment 6**

**Statement**

In this assignment, we aim to:
a) Implement Linear Regression using a suitable library function to predict month-wise temperatures.
b) Evaluate the regression model's performance using MSE, MAE, and R-Square metrics.
c) Visualize the regression model to analyze predictions.

**Objective**

1. Understand how to analyze structured temperature data using Python.

2. Apply regression techniques to forecast temperature trends.

3. Assess model effectiveness using statistical evaluation metrics.

4. Develop visualization techniques for performance analysis.

**Resources Used**

- **Software:** Visual Studio Code

- **Libraries:** Pandas, NumPy, Matplotlib, Scikit-learn

**Introduction to Regression Analysis**

Regression analysis is a fundamental machine learning technique used to predict continuous values. In this assignment, we apply **Linear Regression** to model and forecast month-wise temperature trends in India.

**Key Concepts:**

- **Linear Regression:** Establishes a linear relationship between independent and dependent variables.

- **Model Evaluation:** Uses MAE, MSE, RMSE, and R-squared to assess model accuracy.

- **Data Visualization:** Uses scatter plots and regression lines to analyze predictions.

**Basic Functions Used**

1. pd.read_csv() – Loads data into a Pandas DataFrame.

2. drop(columns=[]) – Removes unnecessary dataset columns.

3. train_test_split() – Splits data into training and testing sets.

4. LinearRegression() – Implements a linear regression model.

5. fit() – Trains the model on training data.

6. predict() – Predicts target values using input features.

7. mean_absolute_error() – Computes MAE to evaluate prediction accuracy.

8. mean_squared_error() – Calculates MSE and RMSE.

9. r2_score() – Measures how well the model explains variance in data.

10. scatter() & plot() – Visualizes regression results.

**Methodology**

**1. Data Collection and Exploration**

- The dataset consists of month-wise temperature records for India.

- The first step is loading and inspecting the dataset.

**2. Feature Selection & Preprocessing**

- Unnecessary columns (e.g., YEAR) are removed.

- Independent variables (X) and the target variable (y) are defined.

**3. Model Training**

- The dataset is split into **80% training** and **20% testing** data.

- A **Linear Regression model** is trained on the training set.

**4. Model Evaluation**

- Performance metrics **MAE, MSE, RMSE, and R-squared** are computed.

- Predicted values are compared with actual values.

**5. Visualization**

- Scatter plots illustrate regression performance.

- A regression line is plotted for reference.

**Results & Observations**

- The model successfully predicted month-wise temperature trends.

- Performance metrics indicated a reasonable level of accuracy with some error.

- Visualization techniques helped analyze model effectiveness.

**Advantages of Regression Modeling**

1. **Predictability:** Enables effective temperature forecasting.

2. **Simplicity:** Easy to implement and interpret.

3. **Data Insights:** Helps understand temperature variations.

**Disadvantages**

1. **Assumptions:** Assumes a linear relationship, which may not always hold.

2. **Accuracy Limitations:** Depends on data quality and relevant feature selection.

**Conclusion**

This assignment demonstrated the **application of Linear Regression** in predicting month-wise temperatures. The process involved **data preprocessing, model training, evaluation, and visualization**. The model provided useful insights into temperature trends, showcasing the significance of regression in data science applications.