**Name:** Chandrakant Dattatrey Thakare

**Roll no:** 282014  **PRN:** 22310303

**Class :** SY CSE AI**Batch:** B1

**Assignment 6: Temperature Prediction using Regression**

**Problem Statement:**

Download and analyze the **India Temperature Dataset** from the following link:  
<https://www.kaggle.com/venky73/temperaturesof-india?select=temperatures.csv>

The dataset provides **average month-wise temperatures** recorded across various places in India (in Celsius).  
The objective is to use **Linear Regression** to model and predict month-wise temperatures, assess the performance of the model, and visualize the results.

**Objective**

* Understand how to apply **Linear Regression** to real-world climate data.
* Gain experience in evaluating regression models using statistical metrics.
* Develop the ability to visualize regression lines and interpret results.

**Tools and Resources**

* **Software Used**: Google Colab / Jupyter Notebook
* **Libraries Used**: Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn

**Key Concepts and Functions Used**

* **Linear Regression** (LinearRegression from sklearn.linear\_model)
* **Evaluation Metrics**:
  + **Mean Squared Error (MSE)**
  + **Mean Absolute Error (MAE)**
  + **R-Squared Score (R²)**
* **Visualization**:
  + Line plots using Matplotlib and Seaborn for regression lines.

**Methodology**

**1. Data Loading and Preparation**

* Load the dataset using pandas.read\_csv()
* Explore and understand the structure of the dataset (months vs temperatures)
* Encode the **Month** column (if categorical) into numerical form
* Visualize the data trend using a line chart to understand temperature fluctuations across months

**2. Applying Linear Regression**

* Define input features (e.g., month numbers) and target variable (average temperature)
* Split the data into training and testing sets (optional in simple regression)
* Fit a **Linear Regression model** using Scikit-learn's LinearRegression()
* Use the trained model to predict month-wise temperatures

**3. Model Evaluation**

* Evaluate the performance using metrics like:
  + **MSE (Mean Squared Error)**
  + **MAE (Mean Absolute Error)**
  + **R² Score (Coefficient of Determination)**
* Interpret each metric to assess model accuracy

**4. Visualization**

* Plot the **actual vs predicted** temperatures using a line graph
* Visualize the **regression line** fitted on the data to understand the model's learning pattern

**Advantages of Linear Regression in Climate Analysis**

* Simple yet effective model for identifying linear trends in climate data
* Helps forecast future temperatures and seasonal patterns
* Easy to interpret, visualize, and deploy in real-world applications

**Challenges**

* Temperature data may not follow a strict linear trend; hence, the model might underfit in some cases
* Handling **seasonality** and **cyclic behavior** of temperature data may require advanced models
* Requires clean and well-formatted data for accurate results

**Conclusion**

This assignment helped develop a practical understanding of **regression modeling** using Python. Through this task, I:

* Applied **Linear Regression** to climate data for temperature prediction
* Evaluated the model using statistical performance metrics
* Visualized the model to interpret relationships and forecast values effectively

This assignment forms a foundation for more advanced regression techniques and time-series forecasting in future projects.