**Assignment 6**

**Name: Ashutosh Shinde**

**Roll no: 282006**

**Batch: B1**

**Problem Statement**

Predict month-wise temperatures in India using Linear Regression and evaluate the model using:

1. **Mean Squared Error (MSE)**
2. **Mean Absolute Error (MAE)**
3. **R-Squared (R²) Score**

**Objective**

1. To preprocess and analyze temperature data.
2. To implement Linear Regression for temperature prediction.
3. To evaluate and visualize the model's performance.

**Resources Used**

* **Dataset:** [Temperatures of India](https://www.kaggle.com/venky73/temperaturesof-india) (temperatures.csv).
* **Libraries:**
  + Pandas, NumPy (Data Manipulation)
  + Scikit-learn (Regression Modeling)
  + Matplotlib (Visualization)

**Methodology**

**1. Data Loading & Exploration**

* **Features:** YEAR (independent variable), ANNUAL temperature (dependent variable).
* **Dataset Shape:** 117 rows × 18 columns (1901–2017).
* **Key Statistics:**
  + Mean Annual Temperature: **29.18°C**
  + Minimum: **28.11°C** (1917), Maximum: **31.63°C** (2016).

**2. Data Preprocessing**

* **Train-Test Split:** 75% training, 25% testing.
* **Feature Scaling:** Not required for simple linear regression with one feature.

**3. Model Training**

* **Algorithm:** LinearRegression() from Scikit-learn.
* **Training Data:** x\_train (Years), y\_train (Temperatures).

**4. Model Evaluation**

* **Metrics:**
  + **R² Score:** 0.659 (65.9% variance explained).
  + **MSE:** 0.165 (Low error).
  + **MAE:** 0.271 (Average prediction error in °C).

**5. Visualization**

* Scatter plot of actual vs. predicted temperatures with regression line.

**Interpretation of Results**

1. **R² Score (0.659):**
   * The model explains **65.9%** of the variance in temperature data.
   * Indicates a moderate fit; higher values (closer to 1) are better.
2. **MSE (0.165) & MAE (0.271):**
   * Low error values suggest the model's predictions are close to actual temperatures.
   * MAE of **0.271°C** means predictions are off by ~0.27°C on average.
3. **Trend Visualization:**
   * The regression line shows a gradual increase in temperatures over years, aligning with global warming trends.

**Conclusion**

1. The Linear Regression model **effectively predicts** annual temperatures with reasonable accuracy.
2. **Improvement Suggestions:**
   * Include more features (e.g., monthly temperatures, CO₂ levels).
   * Try polynomial regression for non-linear trends.
3. **Business Impact:**
   * Useful for climate studies and agricultural planning.