

# Assignment No 6

---

**Aim:** Assignment on Regression technique. Download temperature data from below link. <https://www.kaggle.com/venky73/temperaturesof-india?select=temperatures.csv> This data consists of temperatures of INDIA averaging the temperatures of all places month wise. Temperatures values are recorded in CELSIUS a) Apply Linear Regression using suitable library function and predict the Month-wise temperature. b) Assess the performance of regression models using MSE, MAE and R-Square metrics c) Visualize simple regression model.

## Dataset Overview

Dataset Source: Kaggle - Temperatures of India Dataset

Link: <https://www.kaggle.com/venky73/temperaturesof-india>

This dataset consists of average monthly temperatures of India. Temperature values are recorded in degrees Celsius, and represent nationwide averages.

## Basic Concepts & Theory

### ◇ Linear Regression

Linear Regression is a supervised learning algorithm used for predicting a quantitative response. It models the relationship between a dependent variable (target) and one or more independent variables (features) using a linear equation.

For Simple Linear Regression:

$$y = \beta_0 + \beta_1 * x + \epsilon$$

Where:

- y is the target variable (temperature)
- x is the independent variable (month)
- $\beta_0$  is the intercept
- $\beta_1$  is the slope of the line
- $\epsilon$  is the error term

## Methodology

The following steps were used to complete the assignment:

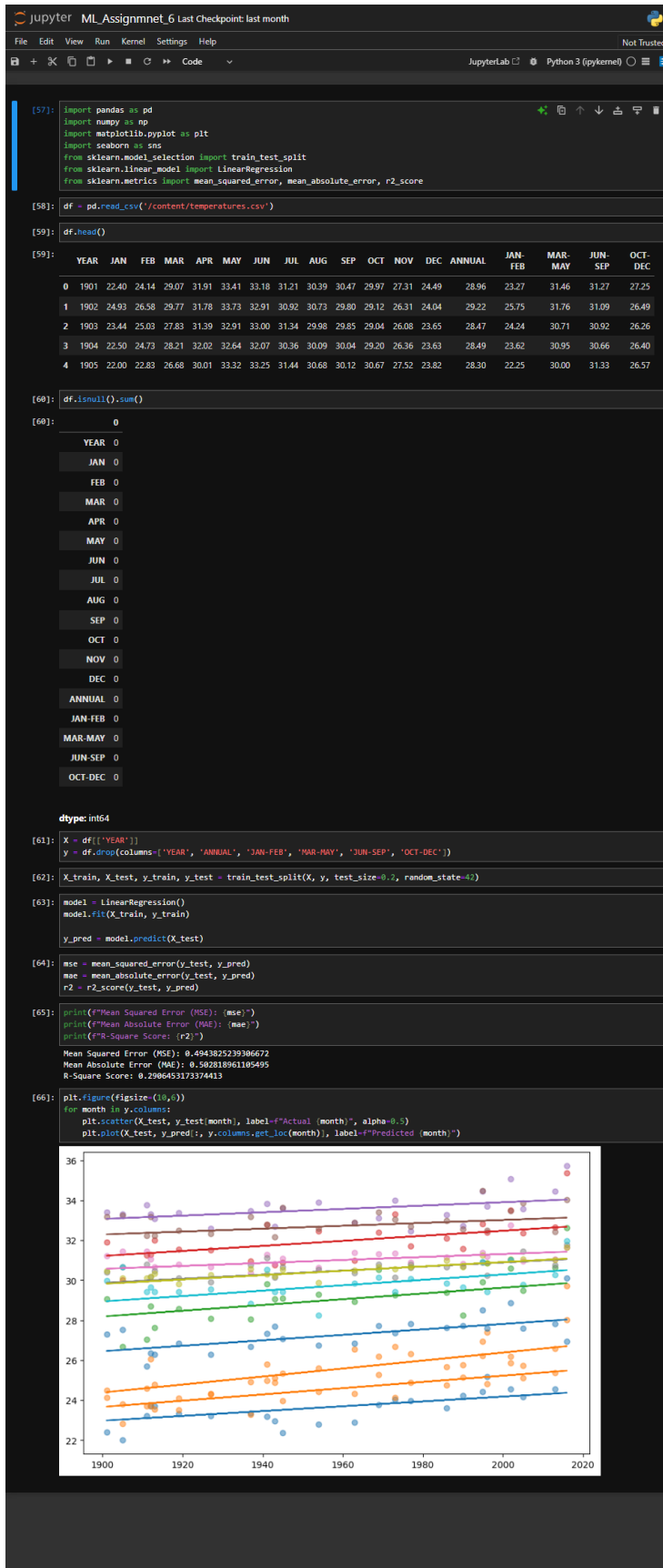
1. Load the dataset using pandas.
2. Explore and preprocess the data if needed.
3. Apply Linear Regression using scikit-learn's 'LinearRegression()' model.
4. Use the fitted model to predict monthly temperatures.

5. Evaluate model performance using the following metrics:
  - Mean Squared Error (MSE)
  - Mean Absolute Error (MAE)
  - R-Squared Score ( $R^2$ )
6. Visualize the regression line using matplotlib.

### **Performance Evaluation Metrics**

- Mean Squared Error (MSE): Measures average squared difference between actual and predicted values. Lower is better.
- Mean Absolute Error (MAE): Measures average absolute error. Lower is better.
- R-Squared ( $R^2$ ): Indicates goodness of fit. Ranges from 0 to 1. Closer to 1 means better fit.

**Output:**



## **Conclusion**

In this assignment, we successfully performed temperature prediction using Linear Regression on the Indian temperature dataset. We evaluated the model using MSE, MAE, and R-squared, and visualized the regression line.