**Assignment No:-6**

**Problem Statement:-**

Assignment on Regression technique.

Download temperature data from the link below.

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 a 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 a simple regression model.

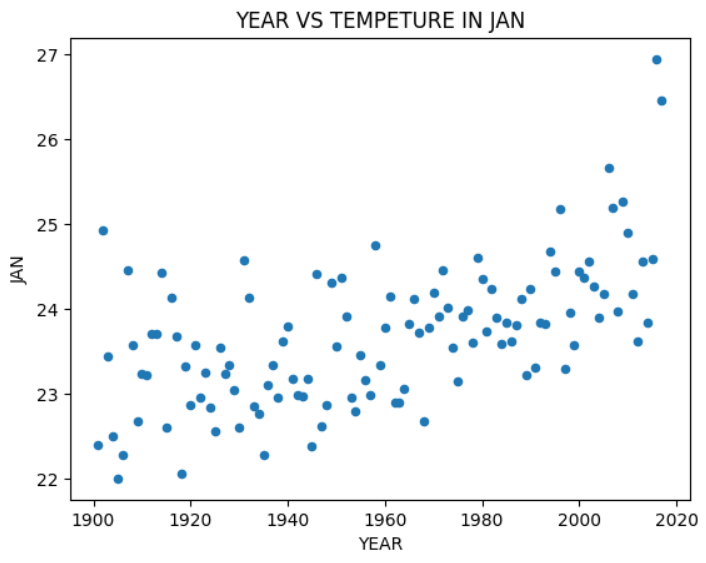
**Theory**:- **Methodology:**

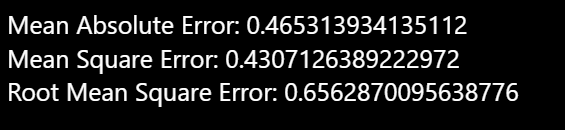
1. **Loading the Data:**
   * **Download the temperature data from the provided link and load it into a DataFrame using a suitable library like pandas in Python or read.csv in R.**
2. **Data Preprocessing:**
   * **Check for any missing values or outliers in the dataset and handle them appropriately. Convert the categorical variables like month into numerical format if needed.**
3. **Splitting the Data:**
   * **Split the data into features (X) and target variable (y), where X represents the independent variables (e.g., month) and y represents the dependent variable (temperature).**
4. **Applying Linear Regression:**
   * **Use a suitable library function (e.g., sklearn.linear\_model.LinearRegression in Python or lm() function in R) to apply linear regression on the dataset.**
5. **Predicting Month-wise Temperature:**
   * **Fit the linear regression model on the training data and use it to predict the month-wise temperature for the test data.**
6. **Assessing Model Performance:**
   * **Calculate metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-Squared to assess the performance of the regression model.**
7. **Visualizing the Regression Model:**
   * **Plot the actual vs. predicted temperatures to visualize how well the regression model fits the data.**

**Advantages and Disadvantages & Limitation/Example:**

1. **Advantages:**
   * **Interpretability: Linear regression models are easy to interpret, making them suitable for explaining relationships between variables.**
   * **Simple and Fast: Linear regression is computationally efficient and easy to implement.**
   * **Versatility: Can be applied to both small and large datasets and can handle multiple input variables.**
2. **Disadvantages & Limitations/Example:**
   * **Assumption of Linearity: Linear regression assumes a linear relationship between the independent and dependent variables, which may not always hold true.**
   * **Sensitivity to Outliers: Linear regression models are sensitive to outliers in the data, which can skew the results.**
   * **Limited Complexity: Linear regression models may not capture complex relationships between variables compared to more advanced regression techniques.**

**Diagram**

****

****

**Conclusion**  
  
In summary, the application of linear regression on the temperature dataset from India facilitated the prediction of month-wise temperatures. Evaluation of regression model performance through metrics such as MSE, MAE, and R-Square provided insights into the accuracy and effectiveness of the predictions. Additionally, visualization of the regression model enhanced the understanding of the relationship between independent and dependent variables, aiding in interpretation and decision-making processes.