

## 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

- Apply Linear Regression using a suitable library function and predict the Month-wise temperature.
- Assess the performance of regression models using MSE, MAE and R-Square metrics
- Visualize a simple regression model.

### Software Library Package:

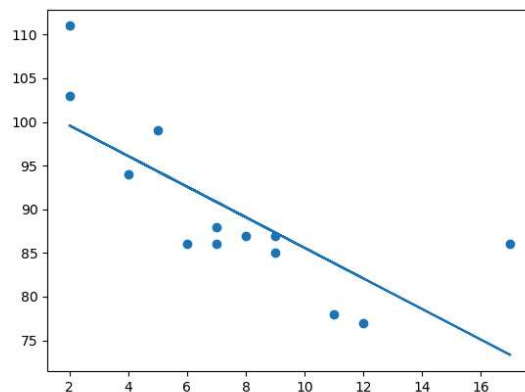
For this task, we'll use the following Python libraries:

- `pandas` for data manipulation.
- `scikit-learn` for linear regression modeling.
- `matplotlib` for data visualization.
- `numpy` for mathematical operations.

### Theory:

i) Methodology:

- **Linear Regression:** Linear regression is a statistical method used to model the relationship between a dependent variable (target) and one or more independent variables (features). It assumes a linear relationship between the variables and estimates the coefficients of the linear equation that best fits the data.



**Fig 1: Linear Regression**

- Advantages: Simple and interpretable model, easy to implement and understand, provides insights into the relationship between variables.
- Applications: Used in various fields such as finance, economics, healthcare, etc., for predicting outcomes based on input variables.
- Limitations/Example: Assumes linearity between variables, sensitive to outliers, may not capture complex relationships in the data.

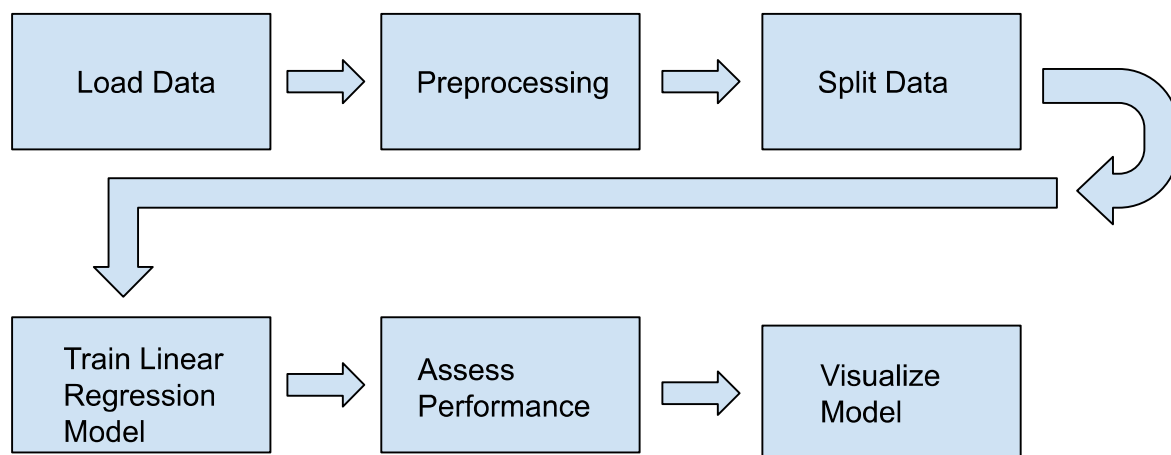
ii) Advantages and Applications and Limitations/Example:

- Advantages: Linear regression is straightforward to implement, easy to interpret, and computationally efficient. It can provide insights into the relationship between the independent and dependent variables.
- Applications: Used in various domains such as finance, economics, healthcare, etc., for predicting outcomes based on input variables.
- Limitations/Example: Linear regression assumes a linear relationship between variables, which may not always hold true. It may not capture complex nonlinear relationships in the data. Additionally, it is sensitive to outliers and multicollinearity.

## Working/Algorithm:

- Load the temperature data into Python using `pandas`.
- Preprocess the data if necessary, such as handling missing values or encoding categorical variables.
- Split the data into training and testing sets.
- Fit a linear regression model to the training data using `scikit-learn`.
- Assess the performance of the model using metrics such as mean squared error (MSE), mean absolute error (MAE), and R-squared.
- Visualize the regression model using `matplotlib`.

## Diagram:



**Fig 2: Workflow Diagram**

## **Conclusion:**

Linear regression is a simple yet powerful technique for modeling the relationship between variables. By fitting a linear equation to the data, it allows us to make predictions and understand the influence of independent variables on the dependent variable. However, it's important to assess the model's performance using appropriate metrics and visualize the results to ensure the model's validity and interpretability.