Vishwakarma Institute of Information Technology, Pune

**(An Autonomous Institute Maharashtra)**

**A Report**

**on**

**A blue and red logo

Description automatically generated**

**CSE (AI) Department**

**Vishwakarma Institute of Information Technology**

**Academic Year: 2023-24**

**“Assignment 6: Regression Analysis”**

**Submitted by**

|  |  |  |
| --- | --- | --- |
| **Sr No** | **Name** | **Email** |
| **1** | **Prajwal Dilip Gaikwad** | [prajwal.22211012@viit.ac.in](mailto:prajwal.22211012@viit.ac.in) |

**Under Guidance of**

Dr. Anuradha Yenkikar

# Problem Statement: -

# 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 place’s 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.

# Packages / Libraries used: -

* **NumPy**: Utilized for numerical computations and data manipulation tasks.
* **Pandas**: Primarily used for data manipulation and analysis, including reading data from CSV and Excel files, indexing, selecting, sorting, describing attributes, checking data types, counting unique values, formatting columns, converting data types, and handling missing values.

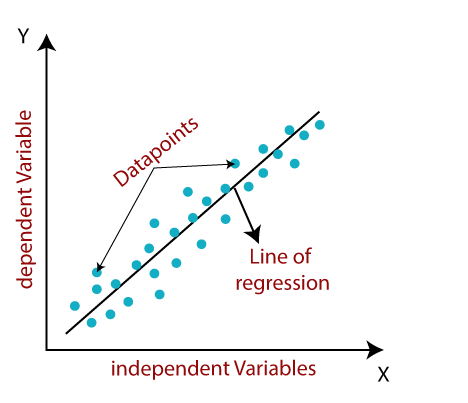
# Theory: -

The regression technique involves applying linear regression to predict month-wise temperatures using temperature data from India, recorded in Celsius. By leveraging a suitable library function, such as the Linear Regression module from scikit-learn, the objective is to build a regression model that accurately predicts temperatures based on historical data. The performance of the regression model is assessed using key metrics including Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-Squared (R2). Additionally, the task entails visualizing the simple regression model to gain insights into the relationship between independent variables (months) and the target variable (temperatures). This endeavor aims to demonstrate the application of regression analysis in understanding and predicting temperature patterns, contributing to broader insights in climate science and related fields.

## Linear Regression: -

Linear regression is one of the easiest and most popular Machine Learning algorithms. It is a statistical method that is used for predictive analysis. Linear regression makes predictions for continuous/real or numeric variables such as sales, salary, age, product price, etc.

Linear regression algorithm shows a linear relationship between a dependent (y) and one or more independent (y) variables, hence called as linear regression. Since linear regression shows the linear relationship, which means it finds how the value of the dependent variable is changing according to the value of the independent variable.



## Types of Linear Regression

* **Simple Linear Regression:**  
  If a single independent variable is used to predict the value of a numerical dependent variable, then such a Linear Regression algorithm is called Simple Linear Regression.
* **Multiple Linear regression:**  
  If more than one independent variable is used to predict the value of a numerical dependent variable, then such a Linear Regression algorithm is called Multiple Linear Regression.

## Methodology: -

1. Data Preparation:
   * Load the temperature data from the provided CSV file into a DataFrame.
   * Separate the independent variables (months) and the dependent variable (temperatures) from the dataset.
2. Train-Test Split:
   * Split the dataset into training and testing sets using the train\_test\_split function from scikit-learn. This allows for evaluating the model's performance on unseen data.
3. Linear Regression Model Training:
   * Instantiate a LinearRegression object from scikit-learn.
   * Fit the model to the training data using the fit method. This process involves finding the optimal coefficients that minimize the residual sum of squares between the observed and predicted temperatures.
4. Prediction:
   * Use the trained model to make predictions on the test data using the predict method.
5. Evaluation:
   * Calculate key performance metrics, including Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE), as well as R-squared (R2) score. These metrics provide insights into the model's accuracy and predictive power.
6. Interpretation:
   * Interpret the model coefficients obtained from the trained model to understand the relationship between the independent variables (months) and the dependent variable (temperatures).
7. Visualization:
   * Visualize the simple regression model by plotting the predicted temperatures against the actual temperatures on a scatter plot. Additionally, the coefficients of the regression line can be visualized to understand their impact on temperature prediction.

## Application: -

# **Weather Forecasting**:

# Regression models can be used to predict future temperatures based on historical data. This is valuable for weather forecasting agencies, allowing them to provide accurate temperature predictions for different regions.

# **Climate Analysis**:

# Regression analysis helps in studying long-term temperature trends and patterns, which are crucial for understanding climate change and its impacts. By analyzing historical temperature data, researchers can identify trends, anomalies, and potential climate shifts.

# **Agriculture**:

# Farmers can use temperature predictions to make informed decisions about crop planting, irrigation schedules, and pest management. By knowing the expected temperature patterns, farmers can optimize crop yields and minimize risks associated with extreme temperatures.

# **Energy Consumption**:

# Regression models can predict energy consumption based on temperature forecasts. Utility companies can use these predictions to plan energy production, manage resources efficiently, and optimize energy distribution networks.

# **Urban Planning**:

# Urban planners use temperature predictions to design sustainable cities and infrastructure. By considering future temperature trends, planners can implement strategies to mitigate the urban heat island effect, improve air quality, and enhance overall livability.

# **Healthcare**:

# Healthcare professionals can use temperature forecasts to prepare for potential health impacts associated with extreme temperatures, such as heatwaves or cold snaps. Hospitals and public health agencies can implement measures to protect vulnerable populations and reduce the burden on healthcare systems.

# **Tourism and Recreation**:

# Temperature predictions are essential for tourism and recreation industries. Travel agencies, event organizers, and outdoor recreation businesses rely on accurate weather forecasts to plan activities, attract visitors, and ensure the safety of participants.

# **Environmental Monitoring**:

# Environmental agencies use temperature data to monitor ecosystem health, track changes in biodiversity, and assess the impact of climate change on natural habitats. By analyzing temperature trends, scientists can identify areas at risk and develop conservation strategies.

# Conclusion: -

In conclusion, the regression analysis conducted on temperature data offers valuable insights into temperature patterns and trends, with practical applications across various sectors. By applying linear regression techniques, we can accurately predict month-wise temperatures based on historical data, providing valuable information for weather forecasting, climate analysis, agriculture, energy management, urban planning, healthcare, tourism, and environmental monitoring. The evaluation of regression model performance using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R2) score allows us to assess the model's accuracy and predictive power. Furthermore, visualization of the regression model provides a clear understanding of the relationship between independent variables (months) and the dependent variable (temperatures), enhancing interpretation and decision-making. Overall, the regression technique serves as a powerful tool for understanding temperature dynamics and informing strategic initiatives across diverse domains, contributing to more informed and resilient decision-making processes.