# SNS COLLEGE OF ENGINEERING

Kurumbapalayam(Po), Coimbatore – 641 107
Accredited by NAAC-UGC with 'A' Grade
Approved by AICTE, Recognized by UGC & Affiliated to Anna University, Chennai

## Department of Information Technology

**Course Name – 23ADT202 Fundamental of Data science and Analytics**

**II Year / IV Semester**

**Unit 2 – Descriptive Analytics**

**Regression**

# Introduction to Regression

- Regression is a statistical technique to study the relationship between variables.
- It predicts the dependent variable (output) based on independent variables (predictors).
- Widely used in forecasting and understanding variable relationships in various fields.
- Essential in business, healthcare, and engineering for decision-making.

# Types of Regression

- **Linear Regression**: Predicts a dependent variable based on one independent variable.
- **Multiple Linear Regression**: Uses multiple independent variables for predictions.
- **Logistic Regression**: Used for binary outcomes like "yes/no" or "pass/fail" predictions.
- Each type has specific use cases and applications in different industries.

# Linear Regression Overview

- Linear regression examines the relationship between two variables.
- The formula for linear regression is: $Y = \beta_0 + \beta_1 X + \epsilon$.
- Y is the dependent variable, X is the independent variable, $\beta_0$ is the intercept, and $\beta_1$ is the slope.
- The error term ($\epsilon$) accounts for deviations from the line.

# Steps in Linear Regression

- **Collect Data**: Gather independent and dependent variable data.
- **Visualize Data**: Plot data to check for linear trends.
- **Fit the Model**: Estimate the coefficients ($\beta_0$ and $\beta_1$)
- **Evaluate the Model**: Assess the model using metrics like

$R^2$ and MSE.

# Example of Linear Regression

- Example: Predicting sales based on advertising spend.
- Dataset: X = [10, 20, 30, 40, 50], Y = [15, 25, 35, 45, 55].
- Perform linear regression to estimate the coefficients.
- Use Python to implement and analyze the results.

# Model Evaluation - $R^2$ and MSE

- **$R^2$ (Coefficient of Determination)**: Measures how well the independent variable explains variability in the dependent variable.
- $R^2$ ranges from 0 (no fit) to 1 (perfect fit).
- **MSE (Mean Squared Error)**: Measures the average squared difference between predicted and actual values.
- Lower MSE indicates better model accuracy.

# Applications of Regression

- **Business**: Predict sales and market trends.
- **Healthcare**: Forecast patient recovery times based on treatments.
- **Education**: Analyze how study hours impact exam performance.
- Regression can improve decision-making and optimize operations in various sectors.

# Exercise 1 - Linear Regression

- Given dataset: X = [5, 10, 15, 20], Y = [10, 20, 25, 30].
- Task: Calculate the intercept ($\beta_o$) and slope ($\beta_1$) of the regression line.
- This exercise demonstrates the basic process of fitting a linear regression model.
- Understanding these values helps in predicting the dependent variable.

# Exercise 2 - Plotting a Regression Line

- Dataset: X = [1, 2, 3, 4, 5], Y = [3, 6, 9, 12, 15].
- Task: Plot the regression line for the given data.
- This exercise visualizes how the regression line fits the data points.
- Helps in understanding the relationship between X and Y and evaluating the model's accuracy.

# THANK YOU