

Lab 6

School of Computer Science Engineering and Technology

| | | | |
|-------------|------------|-------------|--|
| Course | B. Tech. | Type | Core |
| Course Code | CSET301 | Course Name | Artificial Intelligence and Machine Learning |
| Year | 2025 | Semester | Odd |
| Date | 14/08/2025 | Batch | 2023–2027 |

CO-Mapping

| | CO1 | CO2 | CO3 | CO4 | CO5 | CO6 |
|----|-----|-----|-----|-----|-----|-----|
| Q1 | | | √ | | | |

AI/ML Lab –Logistic Regression with scikit-learn

Objective:

Total Marks: 1.0

This lab aims to make the students understand and implement classification model using Logistic Regression algorithm. Perform data preprocessing including handling missing values and encoding. Visualize relationships and model performance. Evaluate the model using standard classification metrics.

Problem Statement:

Use Titanic dataset identify the survival probability. You can download the dataset from the given link: <https://www.kaggle.com/datasets/yasserh/titanic-dataset>

Instructions:

Perform the following tasks:

1. To load the data and print first 5 rows.
2. Explore and visualize the dataset using different columns such as total number of people survived, survival rate based on gender, survival rate based on age (Hint: for age column create bins and divide age data into different categories).
3. Pre-process the data by dropping irrelevant columns, filling missing values, encoding categorical columns, etc.
4. Define X matrix (independent features) and y vector (target feature).
5. Split the dataset into training and testing subsets.
6. Train Logistic Regression Model (from sklearn.linear_model import LogisticRegression)
7. Print the classification metrics for train and test subsets.