

School of Computer Science Engineering and Technology
Assignment-05

Course- B.Tech
Code-23CS301PC206
Year- 2024-2025
Date- 27-08-2024

Type- Core
Course Name- Artificial Intelligence & Machine Learning
Semester- Even, **Instructor:** Prof. E.L.N. Kiran
Batch- AIML-A,B

1 Implement Logistic Regression Model Using Titanic Ship Dataset

Part 1 – Import the required Python, Pandas, Matplotlib, Seaborn packages. [CO2]

1. Load the Titanic Ship Passengers dataset into a dataframe using `pandas`
2. Check the data types of each feature(column) in the dataset.
3. Generate a summary of the dataset for `min`, `max`, `stddev`, `quartile` values for 25%,50%,75%,90%,
4. List the names of columns/features in the dataset
5. Perform Exploratory analysis - plot numeric features, check relative size of survived/unsurvived, check if any pattern on gender. passenger class, class-wise survival rate, siblings, overall age distribution, class-wise age distribution - apply bar plot, histogram, box plots to visualize.
6. Perform data wrangling - impute age data, drop unnecessary features, recode the categorical features to a class

Part 2 – Model training and Fit the data to Model. [CO3]

1. Split the data generated from list created as `X`, `Y` is distributed using `train_test_split` function as `X_train`, `Y_train`, `X_test`, `Y_test`
2. Apply the logistic regression model of `sklearn` package
3. Fit the data to the Logistic Model using `fit`
4. Compute the F1 score as function of penalty , F1 as function of test data. Plot both outcomes using scatter plots.