National University of Computer and Emerging Sciences



Lab Manual

Artificial Intelligence Lab

Department of Computer Science

FAST-NU, Lahore, Pakistan

# Lab Exercises (10 marks)

In this exercise, you will be working with the titanic dataset. The aim of this task is to test your skills in data manipulation and visualization using Python.

**Perform an Exploratory Data Analysis (EDA) on the provided Titanic dataset to gain insights into the passenger demographics and factors influencing survival rates.**

1. **Data Overview:**
   * What is the size of the dataset? How many columns and rows does it contain?
   * Are there any missing values in the dataset? If so, which columns have missing data and what percentage of data is missing?
2. **Passenger Demographics:**
   * What is the distribution of passengers across different passenger classes (Pclass)?
   * How is the age distribution of passengers? Are there any outliers?
   * What is the gender distribution among passengers?
3. **Survival Analysis:**
   * What is the overall survival rate of passengers?
   * Is there any relationship between passenger class and survival rate?
   * How does age correlate with survival? Are children more likely to survive?
   * Is there any significant difference in survival rates based on gender?
   * Does having family members (siblings/spouses or parents/children) aboard affect survival chances?
4. **Ticket Fare and Cabin:**
   * What is the distribution of ticket fares? Are there any outliers?
   * Is there a correlation between ticket fare and passenger class?
   * How does the availability of cabin data relate to survival rates?
5. **Embarkation Port:**
   * What is the distribution of passengers based on the port of embarkation?
   * Is there any correlation between the embarkation port and survival rates?
6. **Additional Insights:**
   * Are there any interesting patterns or insights that stand out in the data?
   * Can you identify any potential biases or limitations in the dataset?
7. **Visualization:**
   * Create all possible visualizations (e.g., histograms, box plots, bar plots, etc.) to better understand the relationships between different variables and survival outcomes.