**CA-1**

**INT234**

**Predictive Analytics**

**Even**

**Instructions: Each question is of 10 marks**

Q1. Using the titanic dataset from the titanic package in R or openly available, how can you apply a K-Nearest Neighbors (KNN) algorithm to predict whether a passenger survived based on features such as age, sex, and class?

Q2.Perform all the steps of data preprocessing on the customer dataset

Q3. Consider the CSV file crimesonwomen and apply the following:

1. Write an sqldf query to find the state with the highest total number of crimes.
2. Write an sqldf query to calculate the total number of crimes for each state across all years.
3. Write an sqldf query to count the number of records for each type of crime in the dataset.
4. Write an sqldf query to select records from the dataset for the state of Maharashtra in the year 2020.

**CA-1**

**INT234**

**Predictive Analytics**

**Odd**

**Instructions: Each question is of 10 marks**

Q1. In the wine dataset which is available through the rattle package in R, how would you use K-Nearest Neighbors (KNN) to classify different types of wine based on attributes like alcohol content, malic acid, and color intensity?

Q2.Perform all the steps of data preprocessing on the product dataset.

Q3. Using the Titanic dataset “titanic\_data\_odd”, answer the following questions with the help of sqldf:

1. Retrieve the names and ages of the passengers who survived the disaster.
2. Find the number of passengers who embarked from each port (C, Q, S) and survived.
3. Calculate the average age of passengers who did not survive, grouped by passenger class (Pclass).
4. List the top 5 youngest passengers who survived, showing their names, ages, and classes.