Assignment Date: 23.11.2021 11:30PM

Due: 25.11.2021 23:59PM

**Question 1 (50 pts):** Kazakh Naval Forces invited you again to develop a specific algorithm for shipwreck investigations. The new algorithm is called CaspianML, and it will use ensemble learning. You are supposed to determine if a cruise passenger is alive or not using voting approach. You will implement decision tree, logistic regression and SVM algorithms. Your final decision will depend on highest vote. Your model score will be the average score of algorithms. Use test\_size = 0.2 for all experiments.

For decision tree criterion: gini, test\_size=0.2

For SVM C=10, gamma=5

You are given the dataset **Sank.csv :**

1. What is the model score of CaspianML ? (25 pts)
2. Express what result will CaspianML predict for the following rows. Is it any different than other algorithms ? Elaborate. **(25 pts)**

Class = 1, Male, Age: 28, Fare: 20.5, survived: 1

Class = 2, Male, Age: 70, Fare: 7.5, survived: 0

Class = 3, Female, Age: 25, Fare: 6.76, survived: 1

Class = 2, Female, Age: 43, Fare: 12.88, survived: 0

**Note:** **Class** column stands for passenger class, **Alive** is either a passenger was survived or not (**survived:1, not survived:0**), and **Fare** is the price in USD paid for cruise trip in the dataset.

**Question 2 (50 pts):** *k* nearest neighbour (called *X*) has been used to estimate the price of the houses based on three different variables which are the number of rooms, size of house in m2, and age of houses. The training set is given below which is used to train the algorithm. Use the test dataset to estimate the price of the houses based on the *k* nearest neighbour algorithm (for *k*=1 and *k*=2).

*Training dataset*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sample No. | y: Price of the house | x1: Number of rooms | x2: Size of house (m2) | Age of House (year) |
| 1 | 500.000 $ | 2 | 45 | 25 |
| 2 | 800.000 $ | 3 | 65 | 30 |
| 3 | 1.000.000 $ | 6 | 100 | 40 |
| 4 | 350.000 $ | 2 | 30 | 20 |
| 5 | 100.000 $ | 2 | 25 | 20 |

*Testing dataset*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sample No. | y: Price of the house | x1: Number of rooms | x2: Size of house (m2) | Age of House (year) |
| 1 | ? | 4 | 100 | 25 |
| 2 | ? | 1 | 60 | 20 |