PA 2: Classification - Nearest Neighbors

Student Details (1 Point)

First Student Name and ID: ABC 1001XXXXXX

Second Student Name and ID: XYZ 1001XXXXXX

Notes: When submitting, fill your name and ID in this cell. [1 point]

Do not to forget to cite any external sources used by you.

Programming Assignment Submission Instructions (2 Points)

Step 1: Rename this submission file as 'yourLastName_Last4digitsofyourID_NN.ipynb' [1 point]

Step 2: Place this file inside the folder 'PA#2_Classification_yourLastName' [1 point]

Do not upload the database file [-20 points]

Programming Assignment Details (27 Points)

Titanic Dataset Details:

- survival - Survival (0 = No; 1 = Yes) [Target Variable]
- class - Passenger Class (1 = 1st; 2 = 2nd; 3 = 3rd)
- name - Name
- sex - Sex
- age - Age
- sibsp - Number of Siblings/Spouses Aboard
- parch - Number of Parents/Children Aboard
- ticket - Ticket Number
- fare - Passenger Fare
- cabin - Cabin
- embarked - Port of Embarkation (C = Cherbourg; Q = Queenstown; S = Southampton)

For this assignment use Jupyter notebook, Panda, and scikit.

- Load titanic dataset attached with this file in csv format. [1 points]
- Select best 3 attributes for training and testing your model. [4 points]
- Find the best K. [3 points]
- Split your dataset 70% for training, and 30% for testing the classifier. [2 points]
- Use Euclidean distance.[3 points]
- Test the classifier with three diferent numbers for neighbors and record the results.[3 points]
- Use comments to explain your code and variable names.[2 points]
- Calculate and print the confusion matrix, and the classification Report (includes:precision, recall, f1-score, andsupport) for all three different numbers.[6 points]

- Plot the classifier in a 2D projection.[3 points]

Report (20 Points)

For each classification task you need to submit a report (Microsoft Word, or PDF) that you have to:

- Describe the Nearest Neighbors method. [5 points]
- Explain what was your criteria for selecting the three attributes. [5 points]
- Visualizations of the classifier in a 2D projection, and write your observations. [5 points]
- Interprete and compare the results.[5 points]

Do not to forget to cite your sources!

Please consult the TA before using any other packages apart from sklearn,numpy,pandas, matplotlib and seaborn