Data Mining
Information Systems Department
Faculty of Computers and Artificial Intelligence
Cairo University

Assignment 3
Classification

Instructions:

1. Assignment should be done individually; copies will be graded -5.

2. Total grades are 5 marks.

3. The assignment discussion will be held in the week starting on 14 May 2022.

4. No late submissions are allowed.

5. Each student should choose and implement only one of the two problems described below.

Problem description:

- Consider the car evaluation dataset, attached in the file "car_data.csv". The dataset contains examples with six structural features directly related to cars (described in the table below) to evaluate and determine the car acceptability level.
- We need to build a classifier that evaluate a given car based on its structural information
- The class label for the classifier is the car acceptability which is the last column in the provided file.

Data description:

Feature	Description	Values
Feature 1	Buying price	vhigh, high, med, low.
Feature 2	Maintenance Price	vhigh, high, med, low.
Feature 3	Number of doors	2, 3, 4, 5more.
Feature 4	Capacity in terms of persons to carry	2, 4, more.
Feature 5	the size of luggage boot	small, med, big.
Feature 6	Estimated safety of the car	low, med, high.
Class label	car acceptability	unacc, acc, good, vgood

Requirements:

For this assignment you are being asked to apply the ***Naïve Bayesian classifier*** to classify the acceptability level of a car based on the 6 features given in the dataset, and then compare its accuracy with the K-nearest classifier.
You should do the following:

1. Divide the dataset into 2 subsets, 1st one will be **75%** of the data and call it *"Training Set"*, 2nd set will be **25%** of the data, and call it *"Testing set"*
2. Build your Bayesian classifier on the first set *"Training set"*.
3. Apply the Bayesian classifier you built in step-2 on the second set *"Testing Set"* to calculate the accuracy of the classifier.

4. Use the *k-nearest algorithm* to get the 4-nearest classes for each record in the "Testing Set", use majority voting to classify such records and calculate the accuracy for this classifier.
5. Compare the 2 classifiers, Naïve Bayesian and K-Nearest, and comment on them.