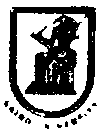
****Data Mining**

**Information Systems Department**

**Faculty of Computers and Artificial Intelligence**

###### **Cairo University**

Assignment 3

Classification

**Instructions:**

* This assignment should be performed individually, copies will be graded -5.
* The assignment total grade is 5.
* The assignment should be submitted before 10/1/2021 at 11pm on blackboard.
* Submission after 10/1/2020 and until 17/1/2020 will be considered a late submission and 50% of the grade will be lost.
* **Don’t use any built-in library. You should implement the code by yourself.**
* Write your code in any programming language.

**Problem**

Description:

* Consider the car evaluation dataset in the attached file, it contains examples with the structural information related to cars like (buying price, maintenance price, number of doors, capacity in terms of persons to carry, the size of luggage boot, estimated safety of the car) to evaluate the car level.

<car.data.csv>

* Apply the any classifier that correctly recognizes evaluation level of a car based on the 6 features.
* The class label is car acceptability which is the last column in the provided comma separated file.

Requirements:

1. Divide the data set 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. Apply the chosen classifier, to build a classifier (model) from the first set “Training set”.
3. Apply the chosen classifier you built in step-1 on the second set “Testing Set” to calculate the accuracy of the classifier.
4. Use the k-nearest algorithm to get the 5-nearest classes for each record in the “Testing Set”, use majority voting to classify such records and calculate the accuracy of this classifier.

Data Description:

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