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

Assignment 3
Association Rules

Instructions:

1. Assignment should be done individually; copies will be graded to zero.

2. Total grade is 5 marks.

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

4. No late submissions are allowed.

5. The submission will be on classroom.

Problem Description:

- You will be given a dataset for **Bank clients**, The Database contains examples with the structural information directly relates clients like (Age, Job, Marital, Education, etc.) to predict if the product (bank term deposit) would be ('yes') or not ('no') subscribed.
  - **Hint:** use the most related attributes that will affect the prediction result. (see dataset description)

- For this assignment you are being asked to apply the **Bayesian classifier** and **decision tree** classifier to correctly access if the product (bank term deposit) would be ('yes') or not ('no') subscribed. A term deposit based on a set of 5 features.

- The class label is y - has the client subscribed a term deposit? (binary: 'yes', 'no')

Data Description:

- The data is related with direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be ('yes') or not ('no') subscribed.

Feature	Values Range	Description
Feature 1	**Age:** (numeric)	Age
Feature 2	**Job :** (categorical: 'admin.', 'blue-collar', 'entrepreneur', 'housemaid', 'management', 'retired', 'selfemployed', 'services', 'student', 'technician', 'unemployed', 'unknown')	Type of job
Feature 3	**Marital :** (categorical: 'divorced', 'married', 'single', 'unknown'; note: 'divorced' means divorced or widowed)	Marital status
Feature 4	**Education:**(categorical:'basic.4y', 'basic.6y', 'basic.9y', 'high.school', 'illiterate', 'professional.course', 'university.degree', 'unknown')	Education level
Feature 5	**Housing:** (categorical: 'no','yes','unknown')	Has housing loan?

Class	**Output variable**: (desired target Y): Yes,No	Loan acceptability

Requirements:

    i.   The interface should enable user to select the percentage of the data needed to be read from the input file e.g. if the file contains 100 records, and the user needs to read 70% of the file then the classification should be done on 70 records only.

   ii.   Apply the Bayesian classifier and decision tree classifier, to build two (classifier models) from the first set "Training set".

  iii.   Apply the Bayesian classifier and decision tree classifier you built on the second set "Testing Set" to calculate the accuracy of the classifiers.

   iv.   Compare the results of 2 classifiers Bayesian and Decision tree.