###### **Data Mining**

**Information Systems Department**

**Faculty of Computers and Artificial Intelligence**

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

**Assignment 1**

**Association Rules**

**Instructions:**

1. Assignment should be done individually; copies will be graded -5.
2. Total grades are 5 marks.
3. The assignment should be submitted before 14 April 2022 at 11pm.
4. No late submissions are allowed.
5. Each student should choose and implement only one of the two problems described below.

**Problem 1**

**Description:**

* Consider the transactions data of a retail market attached in the file “retail\_dataset.csv”:
  + The file contains 315 transactions.
  + Each row represents a single transaction.
  + Each transaction contains a set of purchased items; each one in a separate column.
* We need to know if there are any associations between purchased items in a single transaction.

**Requirements:**

* Write a program, using any programming language, that implements one of the association algorithms (Apriori, FP-Growth or vertical data format) on this dataset.
* Minimum support & minimum confidence should be variable as per user input during runtime.
* Then, generate all association rules which can be mined from the transactions and met the minimum support and confidence.
* The final output of your program should display:
  1. The frequent item sets.
  2. The association rules with their confidence.

**Problem 2**

**Description:**

* Consider the attached file “Scores.xlsx”.
* It contains the number of points collected by school members of 10000 students in Tennis, Basketball and Swimming in a specific competition.
* Each row represents the scores of a one student.
* The scores that should be considered are superior, top level and ranking. The range of scores are distributed as follows:
* Ranking: 65-74
* Top level: 75-84
* Superior: 85-100
* We need to know the association between scores in one sport with reference to the two other sports.
  + For example, superior in swimming derives to top level in basketball and ranking in tennis.

**Requirements:**

* Write a program, using any programming language, that implements one of the association algorithms (Apriori, FP-Growth or vertical data format) on this dataset.
* Minimum support & minimum confidence should be variable as per user input during runtime.
* Then, generate all association rules which can be mined from the transactions and met the minimum support and confidence.
* The final output of your program should display:
  1. The frequent item sets.
  2. The association rules with their confidence.