|  |  |  |
| --- | --- | --- |
| https://www.universitiesegypt.com/ImageHandler.ashx?Id=12743&SS=45d51611eb2f67a4fbfc98c93ac6c147 | **Data Mining** **Information Systems Department**  **Faculty of Computers and Artificial Intelligence** **Cairo University** | **C:\Users\DELL\AppData\Local\Microsoft\Windows\Clipboard\HistoryData\{9ECED2FC-CAE1-4516-AA65-10689F539445}\{79AB2326-BB10-46F2-95AB-064C792A48D8}\ResourceMap\{19483E05-F456-4AD4-BE6F-4F1B169CFA85}** |

**Assignment 1**

**Association Rules**

Instructions:

1. Assignment should be done individually; copies or any other method of cheating will be graded to - 5.
2. Each student can solve one problem or both, and in the second case, we will consider the higher mark.
3. The total grade is 3 marks.
4. No late submissions are allowed.
5. Discussion will be during the office hours.
6. The deadline will be on 5/4/2025.
7. Your program should include a graphical user interface.
8. The interface should enable the 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 analysis should be done on 70 records only.
9. Using the programming language you prefer, write a program with the following specifications:

a.      **Inputs**:

                                                    i.     Excel, text, or CSV file

                                                  ii.     Min support count

                                                iii.     Min confidence (percentage value)

**b.     Outputs**

                                                    i.     The frequent item sets

                                                  ii.     The strong association rules

**Problem 1**

**Description:**

    Consider the sets of transactions for groceries in the attached file "Groceries.csv".

o   columns (Member\_number, Date) represent transactions that a set of items defined with their names at a specific time.

o   Each transaction has many items.

       We need to know the association between items.

**Requirements:**

    Write a program in any programming language that implements one of the association algorithms (Apriori, FP-Growth, or vertical data format) on this set of transactions.

    Minimum support & minimum confidence should be variable as per user input during runtime.

    Then generate allassociation rules that can be mined from the transactions.

    The final output of your program should show the frequent item sets and association rules with their confidence.

**Problem 2**

**Description:**

* Consider the attached file "categories.txt".
* The provided input file consists of the category lists of 77,185 places in the US. Each line corresponds to the category list of one place, where the list consists of a number of category instances (e.g., hotels, restaurants, etc.) that are separated by semicolons.
  + An example line is provided below:

Local Services; IT Services & Computer Repair

In the example above, the corresponding place has two category instances: “Local Services” and “IT Services & Computer Repair”.

       Each line corresponds to the category list of one place.

**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 that can be mined from the transactions and meet 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.