**Lebanese University**
**Faculty of information**
**Section 1**

**Final Exam: Data Mining**

**Time: 90 minutes**
**Dr. Abbas RAMMAL**

---

**Problem 1: Association Mining Rules (10 points)**

We provide the following database of store and customer transactions. Create all association rules with a minimum support of 60% and a minimum confidence of 70%.

1. Find all frequent itemsets using
   a) A priori
   b) FP-Growth.
2. We have the following rule: {Milk, Bread} => {Yogurt}. Calculate the Lift, Leverage and conviction coefficients of this rule. Interpret.

| Transaction ID | Items |
|---|---|
| T1 | Apples, Milk, Bread, Cheese, Yogurt |
| T2 | Bananas, Milk, Bread, Yogurt |
| T3 | Apples, Bread, Eggs, Cheese |
| T4 | Bananas, Milk, Eggs, Yogurt |
| T5 | Apples, Milk, Bread, Yogurt |

---

**Problem 2: Decision Tree Classifier (10 points)**

We consider a real dataset with three categorical attributes (Weather, Temperature, Day) and a binary target variable (Play: Yes or No). Here are 10 records in the dataset:

| Color | Size | Prize | Label |
|---|---|---|---|
| Red | Small | 10 | Class A |
| Blue | Large | 15 | Class B |
| Green | Medium | 20 | Class A |
| Red | Large | 25 | Class B |
| Blue | Medium | 30 | Class A |
| Green | Small | 35 | Class B |
| Red | Medium | 40 | Class A |
| Blue | Small | 45 | Class B |
| Green | Large | 50 | Class A |
| Red | Large | 55 | Class B |

a) Take the optimal split for the "Prize" attribute in your decision tree: $< 35$ and $\geq 35$. Construct a decision tree based on this training data. For splitting, use information gain as measure for impurity. Build a separate branch for each attribute.

b) Split the Prize attribute in two way using Gini Index.

c) Create the contingency table for Color and Size attributes. Test whether the two variables (color and size) are related to each other. Take the critical value of chi-square is 9.488

---