



امتحان الفصل
للعام الجامعي 2025/2024

المادة: إستخراج البيانات وتخزينها	المرحلة:
المدة:	السنة المنهجية: الثالثة
الدورة: الأولى	الاستاذ: عباس رمال

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%.

- Find all frequent itemsets using
 - A priori
 - FP-Growth.

Transaction ID	Items
T1	Apples, Milk, Bread, Cheese
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:

Weather	Day	Temperature	Play
Sunny	Weekday	30	Yes
Sunny	Weekend	32	Yes
Rainy	Weekday	20	No
Sunny	Weekend	28	Yes
Rainy	Weekday	18	No
Rainy	Weekend	22	No
Sunny	Weekday	26	Yes
Sunny	Weekend	35	No
Rainy	Weekday	17	No

Take the optimal split for the "Temperature" attribute in your decision tree: ≤ 27 and >27 . 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.

Problem 3: Data preprocessing (10 points)

You are working as a data analyst for a retail company. You are provided the following dataset about customer demographics, behavior, and platform usage. Some data is missing or noisy.

Customer ID	Age	Annual Income (\$)	Total Purchases
101	23	54000	4
102	35	76000	9
103	36	72000	8
104	45	82000	12
105	38	69000	6
106	50	115000	16
107	29	58000	5
108	41	87000	10
109	31	60000	6

1. Normalize the "Total Purchases" using Min-Max normalization.
 2. Apply smoothing by bin mean on the "Total Purchases" column using a bin depth of 3.
 3. Detect and handle outliers in the "Annual Income" attribute using the IQR method.
-