

Answer to the question – 01(A)

Applying the Apriori algorithm to the transaction data to find all frequent itemsets with a minimum relative support threshold of 50%.

To find the frequent itemsets, we first calculate the support for each item and then the combinations of items.

Minimum support threshold = 50% → 0.50 \* 5 (total transactions) = 2.5 ~ 3 transactions.

Step 1: Count individual items:

* Book A: 4/5 = 80% (frequent)
* Book B: 3/5 = 60% (frequent)
* Book C: 3/5 = 60% (frequent)
* Book D: 4/5 = 80% (frequent)
* Book E: 3/5 = 60% (frequent)

Step 2: Generate 2-itemsets and calculate support:

* {Book A, Book B}: 2/5 = 40% (not frequent)
* {Book A, Book C}: 2/5 = 40% (not frequent)
* {Book A, Book D}: 3/5 = 60% (frequent)
* {Book A, Book E}: 2/5 = 40% (not frequent)
* {Book B, Book C}: 2/5 = 40% (not frequent)
* {Book B, Book D}: 2/5 = 40% (not frequent)
* {Book B, Book E}: 1/5 = 20% (not frequent)
* {Book C, Book D}: 2/5 = 40% (not frequent)
* {Book C, Book E}: 2/5 = 40% (not frequent)
* {Book D, Book E}: 2/5 = 40% (not frequent)

Step 3: Generate 3-itemsets and calculate support:

* Only frequent items are {Book A, Book D}.

Answer to the question – 01(B)

Calculating the confidence for the following association rules derived from the frequent itemsets:

1. Confidence for {Book A} → {Book B} = Support({Book A, Book B}) / Support({Book A}) = 2/4 = 50%.
2. Confidence for {Book A} → {Book D} = Support({Book A, Book D}) / Support({Book A}) = 3/4 = 75%.

Answer to the question – 01(C)

1. Closed Pattern: As we can see that there are no super-set of this itemset has the same support so we can say this is a closed pattern.

2. Max Pattern: As it is a subset of {Book A, Book D} which has higher support so we cannot consider this as max pattern.

1. Closed Pattern: YES
2. Max Pattern: No

Answer to the question – 02(A)

**Step 1:**

Calculate Gini Impurity for the full dataset

There are 5 records:

* 2 "Yes" (Loan Approved)
* 3 "No" (Loan Not Approved)

Gini impurity for the dataset is calculated as follows:

Gini\_total = 1 - (P(Yes)^2 + P(No)^2)

Gini\_total = 1 - ((2/5)^2 + (3/5)^2)

Gini\_total = 1 - (0.16 + 0.36) = 1 - 0.52 = 0.48

**Step 2:**

Calculate Gini Impurity for each feature

We will calculate the Gini impurity for each feature and determine the best split.

1. Split by Credit Score:

* High (2 records): 1 Yes, 1 No
* Gini\_High = 1 – ((1/2)^2 – (1/2)^2 = 1 - 0.5 = 0.5
* Medium (2 records): 1 Yes, 1 No
* Gini\_Medium = 1 – ((1/2)^2 – (1/2)^2 = 1 - 0.5 = 0.5
* Low (1 record): 0 Yes, 1 No
* Gini\_Low = 1 – ((0/1)^2 – (1/1)^2 = 1 - 0.5 = 0
* Weighted Gini for Credit Score:
* Gini\_CreditScore = 2/5\*0.5+2/5\*0.5+1/5\*0=0.4

2. Split by Annual Income:

* High (2 records): 1 Yes, 1 No
* Gini\_High = 0.5
* Medium (2 records): 1 Yes, 1 No
* Gini\_Medium = 0.5
* Low (1 record): 0 Yes, 1 No
* Gini\_Low = 0
* Weighted Gini for Annual Income:
* Gini\_AnnualIncome = 2/5\*0.5 + 2/5\*0.5 + 1/5\*0 = 0.4

3. Split by Employment Status:

* Employed (3 records): 2 Yes, 1 No
* Gini\_Employed = 1 – ((2/3)^2 – (1/3)^2 = 1 - 0.44 - 0.11 = 0.44
* Unemployed (2 records): 0 Yes, 2 No
* Gini\_Unemployed = 1 – ((0/2)^2 – (2/2)^2 = 0
* Weighted Gini for Employment Status:
* Gini\_EmploymentStatus = 3/5\*0.44 + 2/5\*0 = 0.264

4. Split by Existing Debt:

* Low (1 record): 1 Yes, 0 No
* Gini\_Low = 0
* Medium (2 records): 1 Yes, 1 No
* Gini\_Medium = 0.5
* High (2 records): 0 Yes, 2 No
* Gini\_High = 0
* Weighted Gini for Existing Debt:
* Gini\_ExistingDebt = 1/5\*0 + 2/5\*0.5 + 2/5\*0 = 0.2

**Step 3:**

Choose the best feature for the first split

Here are the Gini values:

* Credit Score Gini: 0.4
* Annual Income Gini: 0.4
* Employment Status Gini: 0.264
* Existing Debt Gini: 0.2

The best feature to split on is “Existing Debt” because it has the lowest Gini impurity (0.2).

**Step 4:**

Split on Existing Debt

New branches:

* Low Debt (1 record): Loan Approved = Yes
* High Debt (2 records): Loan Approved = No
* Medium Debt (2 records): We need to further split this group.

**Step 5:**

Further split on Medium Debt

For records with Medium Debt (Applications 2 and 5):

* Credit Score: Medium, Medium
* Annual Income: Medium, High
* Employment Status: Employed, Unemployed
* Loan Approved: Yes, No

Split by Employment Status:

* Employed (1 record): Loan Approved = Yes
* Unemployed (1 record): Loan Approved = No

Gini impurity for this split is 0 (perfect classification).

Final Decision Tree:

* Existing Debt
  + Low Debt → Yes
  + High Debt → No
  + Medium Debt
    - Employed → Yes
    - Unemployed → No

Interpretation:

* If the applicant has low debt, their loan is approved.
* If the applicant has high debt, their loan is not approved.
* If the applicant has medium debt, we look at their employment status:
* If employed, the loan is approved.
* If unemployed, the loan is not approved.

Answer to the question – 02(B)

**Confusion Matrix**

|  |  |  |
| --- | --- | --- |
|  | Predicted Positive | Predicted Negative |
| Actual Positive | 4 | 2 |
| Actual Negative | 2 | 2 |

* Accuracy = (TP + TN) / Total = (4 + 2) / 10 = 0.60 or 60%.
* Precision = TP / (TP + FP) = 4 / (4 + 2) = 0.67 or 67%.
* Recall = TP / (TP + FN) = 4 / (4 + 2) = 0.67 or 67%.
* F1-Score = 2 \* (Precision \* Recall) / (Precision + Recall) = 2 \* (0.67 \* 0.67) / (0.67 + 0.67) = 0.67 or 67%.