# United International University (UIU)

Dept. of Computer Science & Engineering (CSE)

**Assignment 1: Summer – 2024**

**Marks-10**

Course: CSE 4891 || Data Mining

---

1. A bookstore is analyzing the purchasing behavior of its customers to identify frequent item sets for bundling promotions. They have the following transactiondata from their sales:

| Transaction ID | Items Purchased |
|---|---|
| 1 | Book A, Book B, Book C |
| 2 | Book A, Book D, Book E |
| 3 | Book B, Book C, Book D, Book E |
| 4 | Book A, Book B, Book D |
| 5 | Book A, Book C, Book D, Book E |

**Table 1: Transactions**

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

(b) Calculate the confidence for the following association rules derived from the frequent itemsets: {Book A} → {Book B}, {Book A} → {Book D}.

*(c)* Is the following set both *Closed Pattern* and *Max Pattern:*
{Book A, Book C, Book D}
Explain the reason behind your answer.

# United International University (UIU)

Dept. of Computer Science & Engineering (CSE)

**Assignment 2: Summer – 2024**

**Marks-10**

Course: CSE 4891 || Data Mining

1.(a) Look at the data in Table 4. Build a decision tree to predict ***loan approval*** based on the given features and data by calculating ***Gini Impurity***.

| Application ID | Credit Score | Annual Income | Employment Status | Existing Debt | Loan Approved |
|---|---|---|---|---|---|
| 1 | High | High | Employed | Low | Yes |
| 2 | Medium | Medium | Employed | Medium | Yes |
| 3 | Low | Low | Unemployed | High | No |
| 4 | High | Low | Employed | High | No |
| 5 | Medium | High | Unemployed | Medium | No |

**Table 4: Data for Decision Tree Construction**

(b) A health clinic has developed a predictive model to diagnose whether patients have a certain disease based on various health indicators. The model classifies patients as either "Positive" (disease present) or "Negative" (disease absent). The clinic has tested the model on a set of 10 patients and obtained the following results:

| Patient ID | Actual Diagnosis | Predicted Diagnosis |
|---|---|---|
| 1 | Positive | Positive |
| 2 | Positive | Negative |
| 3 | Negative | Negative |
| 4 | Positive | Positive |
| 5 | Negative | Positive |
| 6 | Negative | Negative |
| 7 | Positive | Positive |
| 8 | Negative | Negative |
| 9 | Positive | Negative |
| 10 | Negative | Negative |

Answer the following question given the above scenario:

- Construct the confusion matrix based on the given data.
- Calculate Accuracy, Precision, Recall and F1-Score.