# Indian Institute of Information Technology, Allahabad

**C1- Examination (Feb 2022)**

**Paper: Data Mining**

**B.Tech. (IT), VIth Semester** **Max. Marks : 40** **Duration: 02 Hours**

**Course Instructor: Prof. O.P. Vyas, Dr. Manish Kumar & Dr. Muneendra Ojha**

---

**Ques 1:**Define the importance of following entities in respective Data Mining algorithm along with their mathematical formulation (Define each one with their corresponding algorithm):

**(05 Marks)**

a). Support
b). Confidence
c). Entropy
d). Information Gain

**OR**

Highlight the problem of the Data Analyst while applying mining techniques to any domain in terms of **(05 Marks)**

a) Data pre-processing
b) Choice of Algorithms
c) Choice of Software
d) Final pattern

**Ques 2:** Assume that minimum support threshold (s = 33.33%) and minimum confidence threshold (c = 60%), Find the frequent itemsets and generate association rules on this.

**(10 Marks)**

| Transaction ID | Items |
|---|---|
| Trans 1 | Burger, Pizza, Ketchup |
| Trans 2 | Burger, Pizza |
| Trans 3 | Burger, Beer, Biscuit |
| Trans 4 | Biscuit,Beer |
| Trans 5 | Biscuit, Ketchup |
| Trans 6 | Burger, Beer, Biscuit |

**Ques 3:** Consider the dataset shown in below Table **(10 Marks)**

| Customer ID | Transaction ID | Items Bought |
|---|---|---|
| 1 | 0001 | $\{a, d, e\}$ |
| 1 | 0024 | $\{a, b, c, e\}$ |
| 2 | 0012 | $\{a, b, d, e\}$ |
| 2 | 0031 | $\{a, c, d, e\}$ |
| 3 | 0015 | $\{b, c, e\}$ |
| 3 | 0022 | $\{b, d, e\}$ |
| 4 | 0029 | $\{c, d\}$ |
| 4 | 0040 | $\{a, b, c\}$ |
| 5 | 0033 | $\{a, d, e\}$ |
| 5 | 0038 | $\{a, b, e\}$ |

(a) Compute the support for itemsets $\{e\}$ , $\{b, d\}$ and $\{b, d, e\}$ by treating each transaction ID as a market basket.

(b)Use the results in part (a) to compute the confidence for the association rules $\{b, d\} \rightarrow \{e\}$ and $\{e\} \rightarrow \{b, d\}$. Is confidence a symmetric measure?

(c) Repeat part(a) by treating each CustomerID as a market basket. Each item should be treated as a binary variable (1 if an item appears in at least one transaction bought by the customer, and 0 otherwise).

(d)Use the results in part(c) to compute the confidence for the association rules $\{b, d\} \rightarrow \{e\}$ and $\{e\} \rightarrow \{b, d\}$.

**Ques 4:** The table provided below lists a small dataset for soybean crop. If we had to compare between two features namely "precip" and "temp", then which feature would be a better choice for split w.r.t. the three flavors of decision trees discussed in class i.e. ID3, C4.5 and CART.Demonstrate all the calculations clearly and in sequential manner. **[Marks: 15]**

| date | plant-stand | Precip | temp | hail | crop-hist | area-damaged | severity | seed-tmt | Germination | Class |
|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 0 | 2 | 1 | 0 | 2 | 0 | 2 | 1 | 1 | D1 |
| 3 | 0 | 2 | 1 | 0 | 2 | 1 | 1 | 0 | 1 | D1 |
| 3 | 0 | 2 | 1 | 0 | 1 | 0 | 2 | 1 | 2 | D1 |
| 4 | 0 | 2 | 1 | 1 | 1 | 0 | 1 | 0 | 2 | D1 |
| 4 | 0 | 2 | 1 | 0 | 3 | 0 | 2 | 0 | 2 | D1 |
| 5 | 0 | 2 | 1 | 0 | 3 | 1 | 1 | 1 | 2 | D1 |
| 5 | 0 | 2 | 1 | 0 | 2 | 0 | 1 | 1 | 0 | D1 |
| 6 | 0 | 2 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | D1 |
| 6 | 0 | 2 | 1 | 0 | 3 | 0 | 1 | 1 | 1 | D1 |
| 6 | 0 | 2 | 1 | 0 | 1 | 0 | 1 | 0 | 2 | D1 |
| 3 | 0 | 0 | 2 | 1 | 0 | 2 | 1 | 0 | 1 | D2 |
| 3 | 0 | 0 | 1 | 0 | 1 | 2 | 1 | 0 | 0 | D2 |
| 4 | 0 | 0 | 1 | 0 | 2 | 3 | 1 | 1 | 1 | D2 |
| 4 | 0 | 0 | 1 | 1 | 1 | 3 | 1 | 1 | 1 | D2 |
| 5 | 0 | 0 | 2 | 0 | 3 | 2 | 1 | 0 | 2 | D2 |
| 5 | 0 | 0 | 2 | 1 | 2 | 2 | 1 | 0 | 2 | D2 |
| 5 | 0 | 0 | 2 | 1 | 3 | 3 | 1 | 1 | 2 | D2 |
| 6 | 0 | 0 | 2 | 1 | 0 | 2 | 1 | 0 | 0 | D2 |
| 6 | 0 | 0 | 1 | 1 | 3 | 3 | 1 | 1 | 0 | D2 |
| 6 | 0 | 0 | 2 | 0 | 1 | 3 | 1 | 1 | 0 | D2 |
| 0 | 1 | 2 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | D3 |
| 0 | 1 | 2 | 0 | 0 | 0 | 1 | 1 | 1 | 2 | D3 |
| 0 | 1 | 2 | 0 | 0 | 2 | 1 | 1 | 1 | 1 | D3 |
| 0 | 1 | 2 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | D3 |
| 0 | 1 | 2 | 0 | 0 | 1 | 1 | 2 | 1 | 2 | D3 |
| 2 | 1 | 2 | 0 | 0 | 3 | 1 | 2 | 0 | 1 | D3 |
| 2 | 1 | 2 | 0 | 0 | 2 | 1 | 1 | 0 | 2 | D3 |
| 2 | 1 | 2 | 0 | 0 | 3 | 1 | 2 | 0 | 2 | D3 |
| 3 | 0 | 2 | 0 | 1 | 3 | 1 | 2 | 0 | 1 | D3 |
| 4 | 0 | 2 | 0 | 1 | 0 | 1 | 2 | 0 | 2 | D3 |