**Code: 1**

*- Students are allowed to use 1 A4 paper of material*

*- The exam has 2 pages*

**Question 1: (2.0 scores)** Select ONE of the following questions:

1. Present basic architecture of artificial neural network. Show advantages and disadvantages of artificial neural network.

2. Compare classification with clustering. Take examples to show the differences.

3. Give one example of data mining application in the field of **healthcare** or **human resource**. Based on your example, what kind of data, and data mining method you can use?

**Question 2: (6.0 scores)**

Suppose that a *Departure information for domestic flights* from Ho Chi Minh city as in the following table (Let *Result* be the decision attribute).

*Note: Students can use abbreviations (for example: R for **Result**) to present the examination.*

| 1. | Carrier Airline (CA) | To | Flight date [mm.yyyy] (FD) | Scheduled departure time [hh:mm] (SDT) | Result (R) |
|---|---|---|---|---|---|
| 1 | Vietjet | Da Lat | 6.2023 | 17:25 | Late |
| 2 | Vietnam | Phu Quoc | 5.2023 | 5:40 | On time |
| 3 | Vietjet | Ha Noi | 5.2023 | 9:30 | Late |
| 4 | Pacific | Da Nang | 4.2023 | 10:10 | On time |
| 5 | Vietnam | Da Lat | 6.2023 | 17:25 | On time |
| 6 | Vietjet | Phu Quoc | 6.2023 | 9:30 | Late |
| 7 | Vietnam | Ha Noi | 5.2013 | 10:10 | Late |
| 8 | Vietjet | Da Nang | 4.2023 | 9:30 | Late |
| 9 | Bamboo | Da Lat | 6.2023 | 5:40 | On time |
| 10 | Pacific | Da Nang | 5.2023 | 17:25 | On time |

1. *Let min_sup=25% and min_conf=75%.* Using Apriori, find all frequent itemsets. Select ONE maximal frequent set, then list all association rules. *(1.75 scores)*

2. Suppose B = {*To, Flight date*}, X={1, 3, 6, 7, 8} (*Result = "Late"*). Use rough set to compute: upper approximation, lower approximation, and quality coefficient *(1.0 score)*

3. Determine the root of Decision Tree using Gini Index. *(1.75 scores)*

4. Given a sample *X={Carrier Airline="Vietjet", To="Phu Quoc", Flight date="6.2023", Scheduled departure time="9:30"}*, what would a Naïve Bayesian classification using Laplacican correction of the **Result** for sample X be? *(1.5 scores)*

**Question 3: (2.0 scores)**

Suppose that 8 points as: x1={3, 8}, x2={2, 7.5}, x3={3, 7}, x4={4, 7}, x5={8, 3}, x6={7, 2.5}, x7={8, 2}. And the matrix U0 is:

| U0 | x1 | x2 | x3 | x4 | x5 | x6 | x7 | x8 |
|-----|----|----|----|----|----|----|----|----|
| C1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| C2 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| C3 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 |

Cluster the data to 3 clusters using K_means algorithm and Euclidean distance.

*Note: Only show 3 steps:*

- *Step 1: calculate center of each cluster*
- *Step 2: calculate distances*
- *Step 3: show matrix U1*

<center>END</center>

**Question and Course Learning Outcome table.**

| Question | Course Learning Outcome |
|----------|------------------------|
| 1 | G1 |
| 2 | G2, G4 |
| 3 | G2, G4 |

UNIVERSITY OF INFORMATION TECHNOLOGY

Faculty of Information Systems

Code: 2

FINAL EXAMINATION

Semester II

Academic year: 2022-2023

Subject: Data mining

During: 90 minutes

*- Students are allowed to use 1 A4 paper of material*

*- The exam has 2 pages*

**Question 1: (2.0 scores)** Select ONE of the following questions:

1. Present basic architecture of artificial neural network. Show advantages and disadvantages of artificial neural network.

2. Compare classification with clustering. Take examples to show the differences.

3. Give one example of data mining application in the field of **healthcare** or **human resource**. Based on your example, what kind of data, and data mining method you can use?

**Question 2: (6.0 scores)**

Suppose that a *Departure information for domestic flights* from Ho Chi Minh city as in the following table (Let *Result* be the decision attribute).

*Note: Students can use abbreviations (for example: R for **Result**) to present the examination.*

|  | **Carrier Airline (CA)** | **To** | **Flight date [mm.yyyy] (FD)** | **Scheduled departure time [hh:mm] (SDT)** | **Result (R)** |
|---|---|---|---|---|---|
| 1 | Vietnam | Phu Quoc | 5.2023 | 5:40 | On time |
| 2 | Vietjet | Ha Noi | 5.2023 | 9:30 | Late |
| 3 | Vietnam | Da Lat | 6.2023 | 17:25 | On time |
| 4 | Vietnam | Ha Noi | 5.2013 | 10:10 | Late |
| 5 | Pacific | Da Nang | 5.2023 | 17:25 | On time |
| 6 | Vietjet | Da Lat | 6.2023 | 17:25 | Late |
| 7 | Pacific | Da Nang | 4.2023 | 10:10 | On time |
| 8 | Vietjet | Da Nang | 4.2023 | 9:30 | Late |
| 9 | Vietjet | Phu Quoc | 6.2023 | 9:30 | Late |
| 10 | Bamboo | Da Lat | 6.2023 | 5:40 | On time |

1. *Let min_sup=25% and min_conf=75%.* Using Apriori, find all frequent itemsets. Select ONE maximal frequent set, then list all association rules. *(1.75 scores)*

2. Suppose B = {*To, Flight date*}, X={1, 3, 5, 7, 10} (*Result = "On time"*). Use rough set to compute: upper approximation, lower approximation, and quality coefficient *(1.0 score)*

3. Determine the root of Decision Tree using Gini Index. *(1.75 scores)*

4. Given a sample *X={Carrier Airline="Vietnam", To="Da Lat", Flight date="6.2023", Scheduled departure time="17:25"}*, what would a Naïve Bayesian classification using Laplacican correction of the **Result** for sample X be? *(1.5 scores)*

**Question 3: (2.0 scores)**

Suppose that 8 points as x1 ={3, 8}, x2 ={3, 7.5}, x3 ={3, 7}, x4 ={4, 7}, x5 ={8, 3}, x6 ={9, 2.5}, x7 ={8, 2}. And the matrix U0 is:

| U0 | x1 | x2 | x3 | x4 | x5 | x6 | x7 | x8 |
|----|----|----|----|----|----|----|----|----|
| C1 | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| C2 | 0  | 1  | 0  | 0  | 0  | 0  | 0  | 0  |
| C3 | 0  | 0  | 1  | 1  | 1  | 1  | 1  | 1  |

Cluster the data to 3 clusters using K_means algorithm and Euclidean distance.

*Note: Only show 3 steps:*

- *Step 1: calculate center of each cluster*
- *Step 2: calculate distances*
- *Step 3: show matrix U1*

END

*June 15th, 2023*

Faculty of Information Systems                    Lecturer

**Question and Course Learning Outcome table.**

| Question | Course Learning Outcome |
|----------|-------------------------|
| 1        | G1                      |
| 2        | G2, G4                  |
| 3        | G2, G4                  |