



SCHOOL OF SCIENCE & TECHNOLOGY

Semester 1: 2024

FINAL EXAMINATION PAPER

ITC355: BUSINESS INTELLIGENCE

Important Instructions to Students:

1. Read the questions, instructions carefully and answer the questions.
2. You are allowed an **extra 10 minutes reading time** (during this time, you are not allowed to write with pen) and **180 minutes (3 Hours) to complete this paper**.
3. Silent, non-programmable calculators are allowed, but not supplied.
4. Write all your answers in the answer booklet provided. If you need extra sheets, please ask from the supervisor. Write your name, course code and student identification number in the answer booklet and in the attendance sheet.
5. **This Final Examination Paper is worth 100 marks contributes 50% of the total course assessments.** Students must **score at least score 40 marks** in this final exam to pass the course.
6. This paper contains two (2) sections A and B. All questions are compulsory to answer **except** Section B has choices.
7. Write all your answers legibly. The marker has the right to award zero for answers that are not legible.
8. This is a **Closed Book Examination**. The University views plagiarism, cheating and dishonest practice as serious offences and if found, offenders are penalized.

Examination Paper Content (Format):

<u>Section</u>	<u>Question</u>	<u>Content</u>	<u>Marks</u>	<u>Suggested Time</u>
A	1 – 5	Short Answers	30	54 minutes
B	1 – 8	Discussion Questions (Choose any SEVEN out of Eight)	70	126 minutes
		TOTAL	100	180 Minutes

SECTION A – SHORT ANSWERS (30 MARKS)

All questions are compulsory to answer.

1. State a data mining method and an algorithm for each type of data mining task. (3m)

Table:

	<i>Method</i>	<i>Algorithm</i>
Classification		
Association		
Segmentation		

2. Study the weather.nominal.arff dataset in Fig.1 and answer questions in parts (a – m).

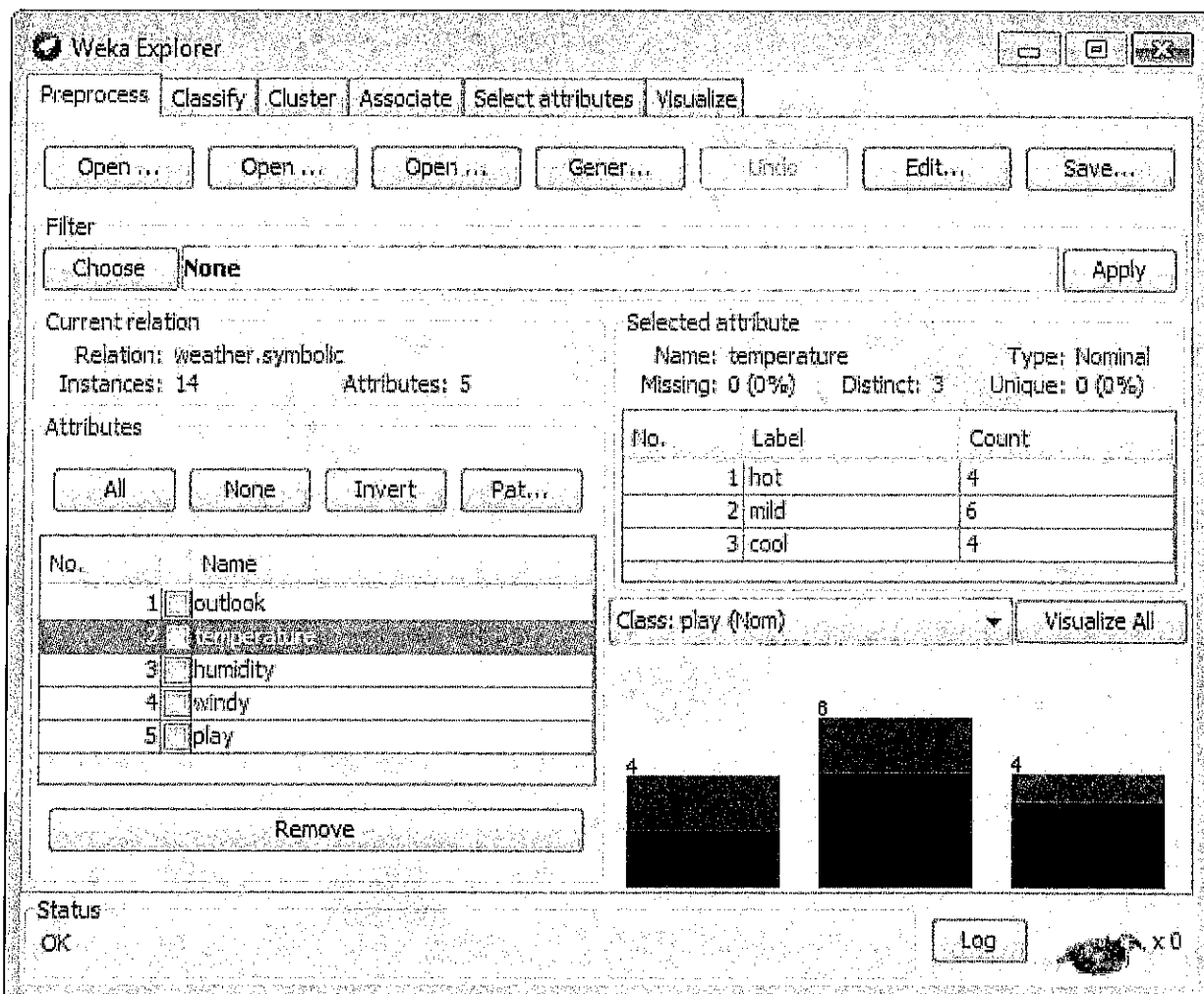
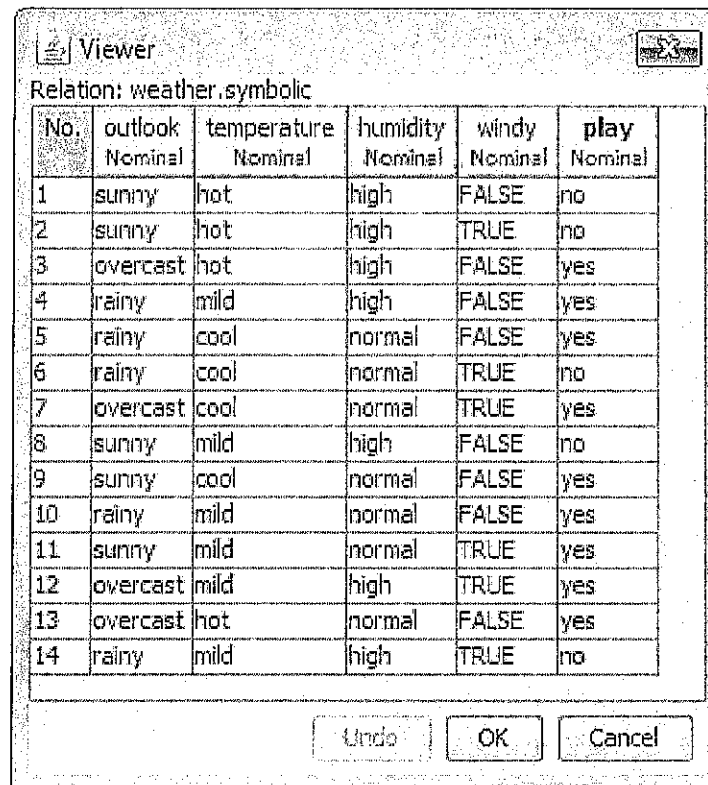


Fig. 1

- a. How many fields are there? (0.5m)
- b. How many records are there? (0.5m)
- c. What are the values that the *temperature* can have? (1m)
- d. What type of data does the temperature have? (0.5m)

e. Which button will you press to view the dataset as shown in Fig.2 below? (0.5m)



Viewer

Relation: weather.symbolic

No.	outlook Nominal	temperature Nominal	humidity Nominal	windy Nominal	play Nominal
1	sunny	hot	high	FALSE	no
2	sunny	hot	high	TRUE	no
3	overcast	hot	high	FALSE	yes
4	rainy	mild	high	FALSE	yes
5	rainy	cool	normal	FALSE	yes
6	rainy	cool	normal	TRUE	no
7	overcast	cool	normal	TRUE	yes
8	sunny	mild	high	FALSE	no
9	sunny	cool	normal	FALSE	yes
10	rainy	mild	normal	FALSE	yes
11	sunny	mild	normal	TRUE	yes
12	overcast	mild	high	TRUE	yes
13	overcast	hot	normal	FALSE	yes
14	rainy	mild	high	TRUE	no

Undo OK Cancel

Fig. 2

- f. What is the function of the first column (No.) in the Viewer window? (1m)
- g. How many numeric attributes are in this dataset? (0.5m)

- h. What will be the result of applying the chosen filter shown in Fig.3 on *weather.nominal* dataset in Fig.1? (2m)

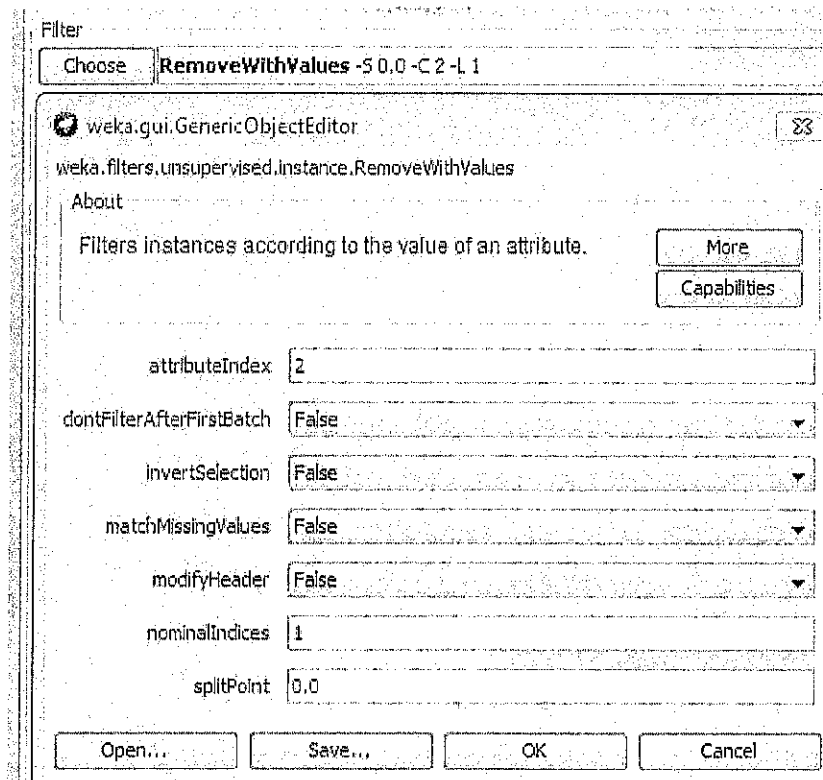


Fig. 3

- i. What button will you press in Fig. 1 to cancel the change to the dataset that you just performed in Fig.3, and verify that the data has reverted to its original state? (0.5)
- j. What is another name of *J48* algorithm? (0.5m)

- k. Fig. 4 shows the output after building and testing the $J48 - C 0.25 - M 2$ classifier. Interpret the output. (2m)

```

Classifier output

=== Classifier model (full training set) ===

J48 pruned tree
-----

outlook = sunny
|  humidity = high: no (3.0)
|  humidity = normal: yes (2.0)
outlook = overcast: yes (4.0)
outlook = rainy
|  windy = TRUE: no (2.0)
|  windy = FALSE: yes (3.0)

Number of Leaves   :    5

Size of the tree   :    8

=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances      14           100 %
Incorrectly Classified Instances     0             0 %
Kappa statistic                     1
Mean absolute error                  0
Root mean squared error              0
Relative absolute error              0 %
Root relative squared error          0 %
Total Number of Instances          14

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
                1       0       1         1         1         1       yes
                1       0       1         1         1         1       no
Weighted Avg.   1       0       1         1         1         1

=== Confusion Matrix ===

a b  <-- classified as
9 0 | a = yes
0 5 | b = no

```

Fig. 4

- l. Draw a *Tree View* that would be produced by the *Visualize tree* Command for the *J48* pruned tree in Fig.4. (5m)
- m. Explain the visualization shown in Fig.6 below. (2m)

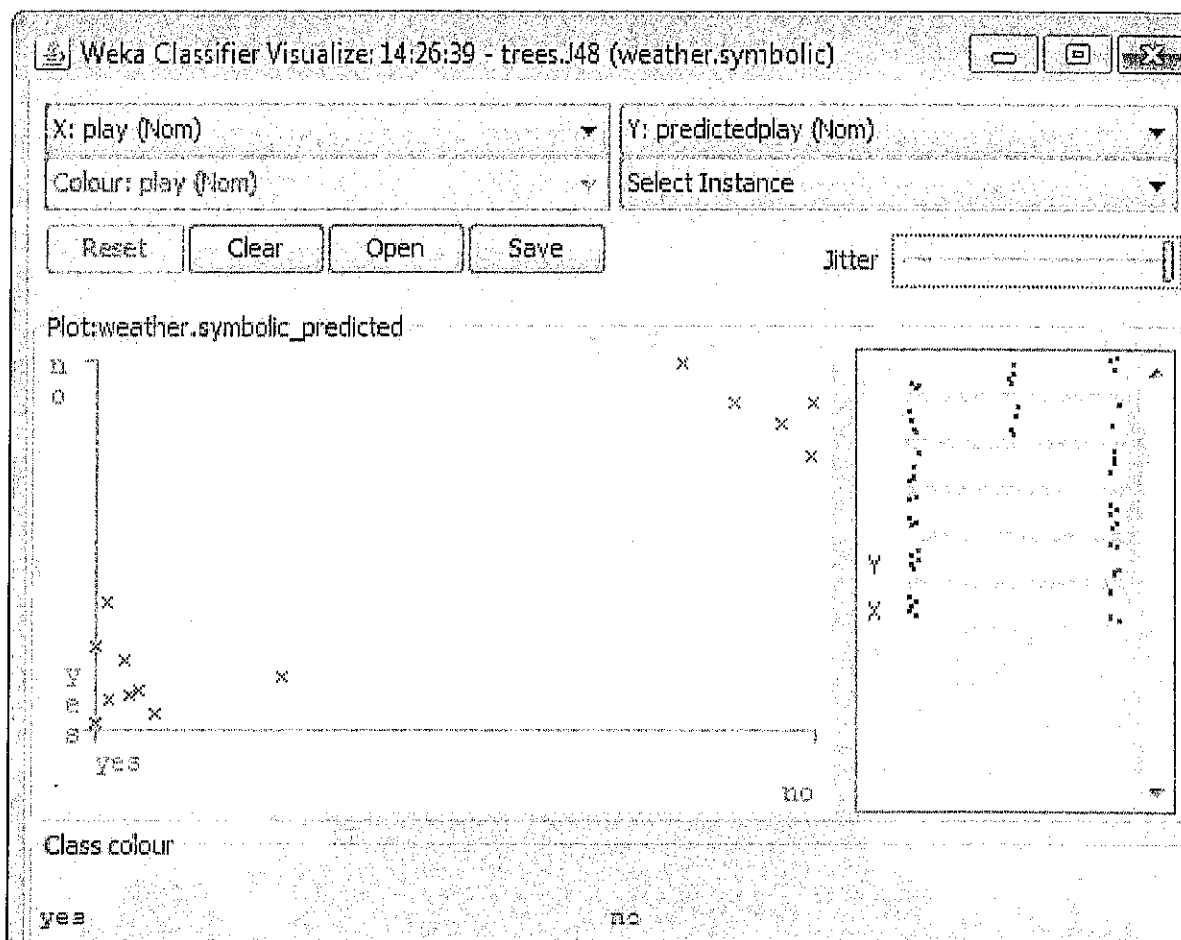


Fig. 6

3. Study the Fig.7 below on the application of association rule mining on weather.nominal.arff data and answer parts a – e.

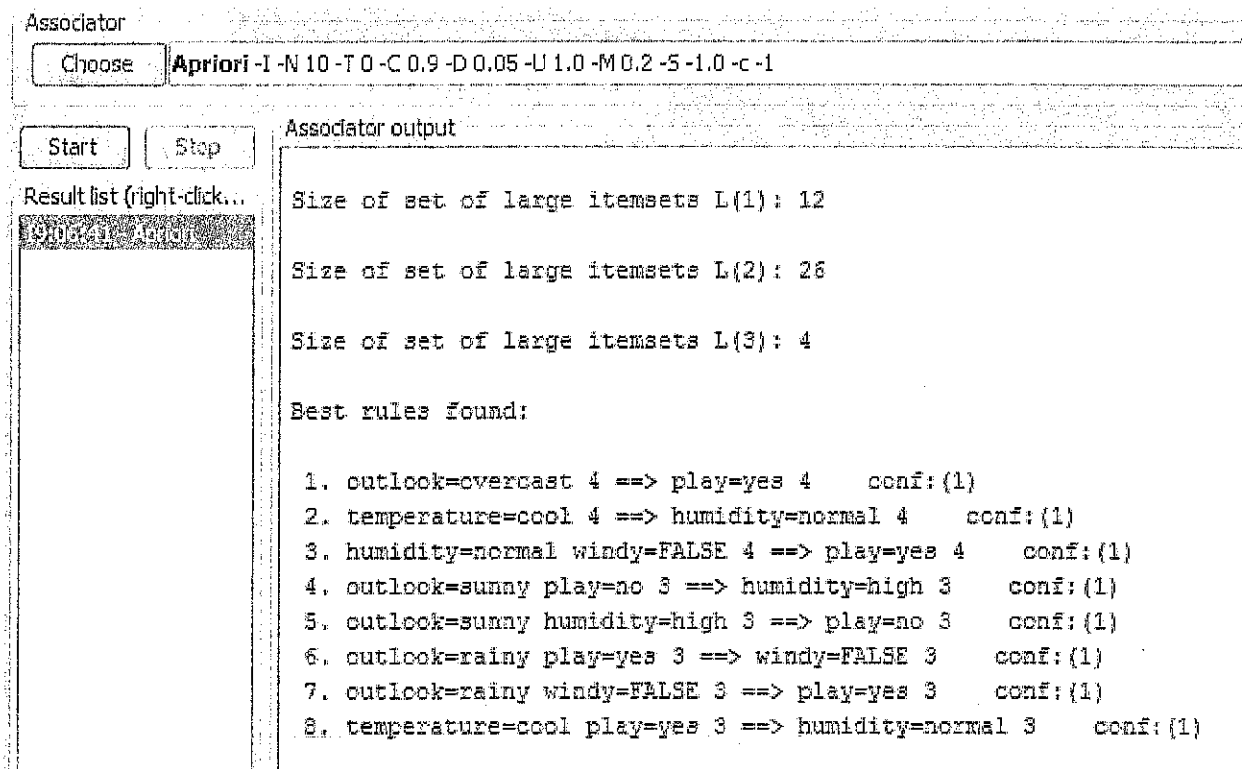


Fig. 7

What values have been used for each of the following parameters? (2.5m)

- upperBoundMinSupport
- significanceLevel
- numRules
- minMetric <confidence>
- Minimum Support

4. In which of the following classifiers in Fig.8, will you find the algorithms mentioned in parts a – j? (3m)

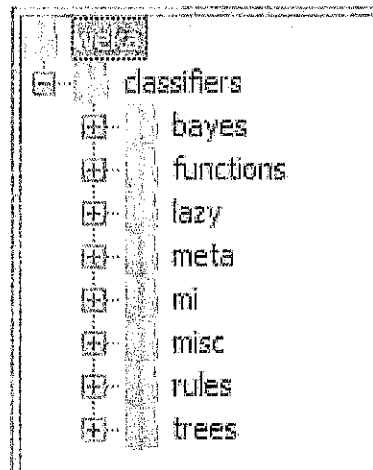


Fig. 8

- a. IBK
- b. UserClassifier
- c. OneR
- d. Naïve Bayes
- e. FilteredClassifier
- f. Logistic

5. Explain the effect of applying the following filters (a – e) on Iris dataset. (5m)

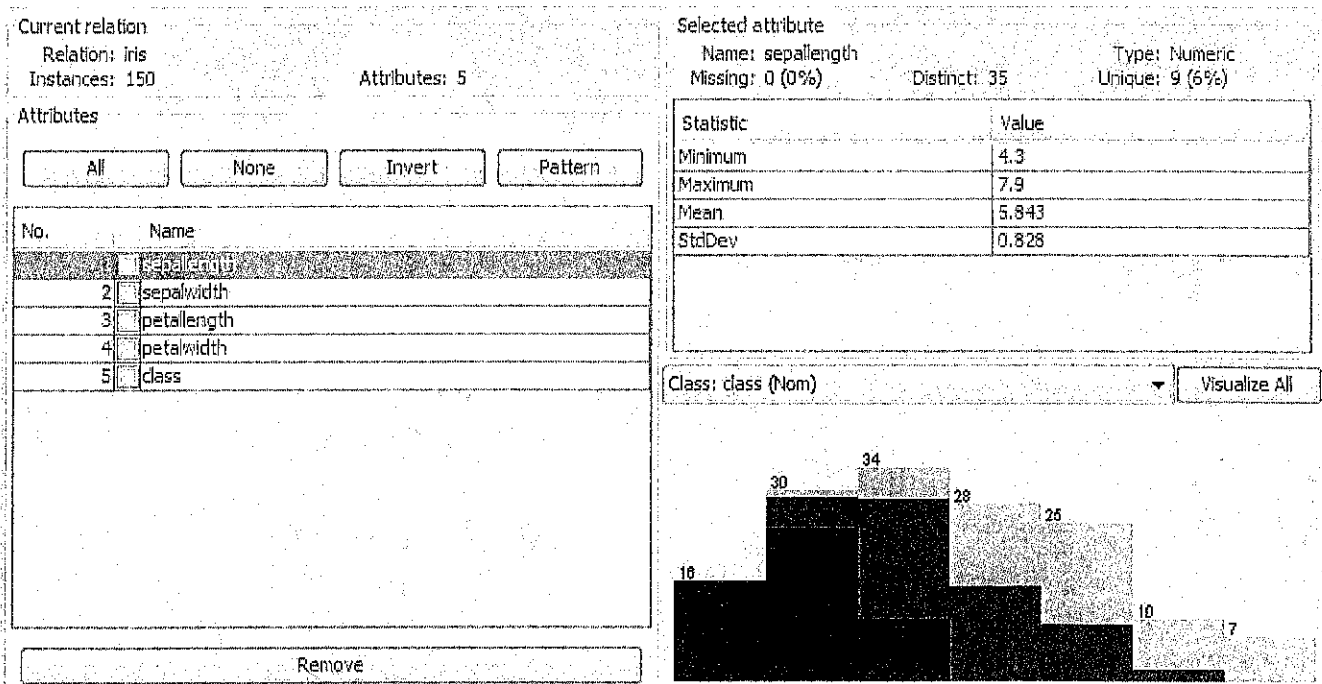


Fig. 9

- Iris-weka.filters.unsupervised.attribute.Remove -R 1
- Iris-weka.filters.unsupervised.attribute.Copy -R 3
- Iris-weka.filters.unsupervised.instance.Resample -S 1 -Z 10.0
- Iris-weka.filters.unsupervised.attribute.Discretize -B 10 -M -1.0 -R first-last
- Iris-weka.filters.unsupervised.attribute.Discretize -D -R first-last

SECTION B - DISCUSSION QUESTIONS (70 MARKS)

Each question is worth 10 marks.

Answer any SEVEN (7) out of Eight questions.

Use a new page for each question.

1. Explain the **three** types of business analytics – **Descriptive**, **predictive**, and **prescriptive** by comparing and contrasting (a) the questions they answer, (b) the enabling technologies they use, and (c) the outcomes they achieve.
2. There are many **types of charts/graphs**. Explain how you should determine right chart/graph for the job?
3. In data analytics, explain types of **structured data**, with examples.
4. Compare **OLTP** and **OLAP** based on the criteria – *purpose, data source, reporting, resource management, and execution speed*.
5. Explain the **CRISP-DM** data mining process with the help of a diagram.
6. **Simple Split** and **k-fold cross validation** are two of the most popular estimation methodologies used for classification type data mining models. Explain each model with the help of diagrams.
7. Explain the **three** main areas of **web mining**, data sources they use, and the business problems and opportunities they address.
8. Explain the general **process of simulation**.

END