

EGCO 425 Data Mining (T2/2018)

SO 1 An ability to identify, formulate, and solve complex engineering problems by applying principles of engineering, science, and mathematics

PI 1.1 Identify and formulate engineering problems

PI 1.2 Solve engineering problems by applying mathematics and engineering knowledge

Passing criterion $\geq 70\%$

PI 1.1 Identify and formulate engineering problems

Assessment 1.1.1 Midterm exam : identify effects of parameter setups

Attainability = 18 out of 22 students (81.8%)

3.6 MinPts ในอัลกอริทึม DBSCAN มีบทบาทในการทำ clustering อย่างไร ถ้าพารามิเตอร์ตัวนี้มีค่าสูงเกินไปจะมีผลอย่างไร และถ้ามีค่าต่ำเกินไปจะมีผลอย่างไร (3 คะแนน)

MinPts = เกณฑ์ที่บอกว่าจุด p ใดๆ เป็น core point หรือไม่ (ภายในรัศมี r จากที่มี p เป็นจุดศูนย์กลาง ต้องมีสมาชิกอย่างน้อย MinPts) (1)

สูงเกินไป : p อาจไม่ผ่านเกณฑ์ กลายเป็น noise ทั้งที่ไม่ควรเป็น (1)

ต่ำเกินไป : มี noise ผ่านเกณฑ์ กลายเป็น core point (1)

ABET : course objective 1, SO 1 → pass at 1.5 points

Assessment 1.1.2 Final exam : identify problems in rule set

Attainability = 10 out of 22 students (45.5%)

2.2 เมื่อใช้ ruleset ในข้อ 2.1 ไปจำแนกข้อมูลในตารางที่ 2 พบว่า ruleset นี้ขาดคุณสมบัติ exhaustive ให้หาว่าเรคคอร์ดใดในตารางที่ 2 ทำให้เกิดปัญหาดังกล่าว และควรแก้ปัญหานี้ได้อย่างไร (3 คะแนน)

record 1, 4 (2)

เติม default rule โดยให้ทำนายตาม majority class : () → no (1)

ABET : course objective 1, SO 1 → pass at 1.5 points

Average attainability = $(81.8 + 45.5)/2 = 63.65\%$

Student performance

	Student ID	Assessment 1.1.1 Pass at 1.5 points	Assessment 1.1.2 Pass at 1.5 points
1	5713374	1.5	0.0
2	5813163	0.0	0.5
3	5813351	0.0	0.0
4	5813365	3.0	0.0
5	5913197	3.0	0.5
6	5913198	2.0	0.5
7	5913201	2.5	2.5
8	5913203	2.5	2.5
9	5913208	0.0	0.0
10	5913212	2.5	3.0
11	5913213	3.0	2.5
12	5913216	1.0	2.5
13	5913219	3.0	3.0
14	5913220	3.0	1.0
15	5913221	3.0	0.5
16	5913375	2.5	2.5
17	5913377	2.0	3.0
18	5913378	2.0	0.0
19	5913380	2.5	3.0
20	5913382	3.0	1.5
21	5913383	1.5	0.5
22	5913384	3.0	1.0

PI 1.2 Solve engineering problems by applying mathematics and engineering knowledge

Assessment 1.2.1 Midterm exam : calculate evaluation metrics from given scenario
Attainability = 20 out of 22 students (90.9%)

2.1 ทำค่า confidence และค่า lift ของกฎต่อไปนี้ (4 คะแนน)		
กฎ	Confidence	Lift
Coco ==> Logan	$= 0.5/0.7 = 0.7143$	$= 0.5/(0.7*0.8) = 0.8929$
Logan ==> Baywatch	$= 0.5/0.8 = 0.625$	$= 0.5/(0.8*0.5) = 1.25$

ABET : course objective 1, SO 1 → pass at 2 points

Average attainability = 90.9%

Student performance

	Student ID	Assessment 1.2.1 Pass at 2 points
1	5713374	2.0
2	5813163	2.0
3	5813351	0.0
4	5813365	4.0
5	5913197	4.0
6	5913198	0.0
7	5913201	4.0
8	5913203	2.0
9	5913208	4.0
10	5913212	4.0
11	5913213	2.0
12	5913216	4.0
13	5913219	3.0
14	5913220	4.0
15	5913221	4.0
16	5913375	4.0
17	5913377	4.0
18	5913378	3.0
19	5913380	4.0
20	5913382	2.0
21	5913383	4.0
22	5913384	4.0

SO 6 An ability to develop and conduct appropriate experimentation, analyze and interpret data, and use engineering judgment to draw conclusions

PI 6.1 Develop and conduct experimentation

PI 6.2 Analyze and interpret data, and use engineering judgment to draw conclusions

Passing criterion $\geq 70\%$

PI 6.1 Develop and conduct experimentation

Assessment 6.1.1 Group project 1. Apply association analysis to SPECT data (<https://bit.ly/2NdrEbK>) and write a report. This assessment focuses on parameter setup, discussion on results, and conclusion (marking criteria 3, 4, 5 – pass at 40 out of 75 points).

Attainability = 18 out of 22 students (81.8%)

Average attainability = 81.8%

Student performance

	Student ID	Assessment 6.1.1 Pass at 40 points
1	5713374	21.5
2	5813163	21.5
3	5813351	0.0
4	5813365	21.5
5	5913197	75.0
6	5913198	72.0
7	5913201	75.0
8	5913203	45.0
9	5913208	55.0
10	5913212	72.0
11	5913213	60.0
12	5913216	72.0
13	5913219	45.0
14	5913220	57.5
15	5913221	57.5
16	5913375	60.0
17	5913377	75.0
18	5913378	40.0
19	5913380	55.0
20	5913382	40.0
21	5913383	60.0
22	5913384	60.0

PI 6.2 Analyze and interpret data, and use engineering judgment to draw conclusions

Assessment 6.2.1 Final exam : calculate and compare performance of 2 classifiers
 Attainability = 15 out of 22 students (68.2%)

4.1 เมื่อทดลองใช้ classifier 2 ตัวในการจำแนกข้อมูลชุดหนึ่ง ได้ผลตามตาราง

Classifier 1

Confusion matrix

	Predict A	Predict B	Predict C
Actual A	18	0	2
Actual B	0	13	7
Actual C	1	4	15

Area under ROC of class A = 0.938
 Area under ROC of class B = 0.775
 Area under ROC of class C = 0.763

Classifier 2

Confusion matrix

	Predict A	Predict B	Predict C
Actual A	18	2	0
Actual B	1	14	5
Actual C	2	5	13

Area under ROC of class A = 0.908
 Area under ROC of class B = 0.763
 Area under ROC of class C = 0.819

หาค่า F-measure ในการจำแนกคลาส B ของ classifier 2 ตัวนี้ และจาก F-measure ที่ได้ ควรสรุปว่า classifier ตัวใดจำแนกคลาส B ได้ดีกว่ากัน (4 คะแนน)

ABET : course objective 1, SO 1 → pass at 2 points

Average attainability = 68.2%

Student performance

	Student ID	Assessment 6.2.1 Pass at 2 points
1	5713374	0.0
2	5813163	1.0
3	5813351	0.0
4	5813365	4.0
5	5913197	4.0
6	5913198	1.0
7	5913201	4.0
8	5913203	4.0
9	5913208	0.0
10	5913212	4.0
11	5913213	4.0
12	5913216	4.0
13	5913219	4.0
14	5913220	4.0
15	5913221	4.0
16	5913375	3.0
17	5913377	4.0
18	5913378	0.0
19	5913380	4.0
20	5913382	4.0
21	5913383	0.0
22	5913384	3.0

SO 7 An ability to acquire and apply new knowledge as needed, using appropriate learning strategies**PI 7.2 Use appropriate sources of knowledge****Passing criterion $\geq 70\%$**

PI 7.2 Use appropriate sources of knowledge

Assessment 7.2.1 Group project 1. Apply association analysis to SPECT data (<https://bit.ly/2NdrEbK>) and write a report. This assessment focuses on searching relevant information and background about heart disease & SPECT data (marking criteria 1,2 – pass at 13 out of 25 points).

Attainability = 18 out of 22 students (81.8%)

Average attainability = 81.2%**Student performance**

	Student ID	Assessment 7.2.1 Pass at 13 points
1	5713374	12.0
2	5813163	12.0
3	5813351	0.0
4	5813365	12.0
5	5913197	23.0
6	5913198	25.0
7	5913201	23.0
8	5913203	23.0
9	5913208	21.0
10	5913212	25.0
11	5913213	23.0
12	5913216	25.0
13	5913219	23.0
14	5913220	23.0
15	5913221	23.0
16	5913375	23.0
17	5913377	23.0
18	5913378	23.0
19	5913380	21.0
20	5913382	23.0
21	5913383	23.0
22	5913384	23.0

Summary

	PI	Attainable ($\geq 70\%$)	Reasons	Remedial Actions	Action Plan	Measurements
SO1	1.1	No	<ul style="list-style-type: none"> Some students cannot apply rule induction concepts to given scenarios 	<ul style="list-style-type: none"> More examples / case studies 	Next year	Next year
	1.2	Yes				
SO6	6.1	Yes				
	6.2	No	<ul style="list-style-type: none"> Some students are confused about TP, TN, FP, FN, and thus cannot calculate and compare performance metrics 	<ul style="list-style-type: none"> More focus on calculation and more examples on performance evaluation 	Next year	Next year
SO7	7.2	Yes				