

Part III: b. Clustering Lab

- This test/quiz can be taken ONLY ONCE.
- Correct answers will NOT be revealed after submission.

จงเลือกคำตอบที่ถูกต้องเพียง 1 ข้อ

1 แลปที่ 2: Clustering Lab (ชุดข้อมูลการประเมินรถ)

ประกอบด้วยข้อมูล 6 ตัวแปรดังต่อไปนี้

1. ตัวแปร buying เช่น vhigh, high, med, low
2. ตัวแปร maint เช่น vhigh, high, med, low
3. ตัวแปร doors เช่น 2, 3, 4, 5 เป็นต้น
4. ตัวแปร persons เช่น 2, 4 เป็นต้น
5. ตัวแปร lug_boot เช่น small, med, big
6. ตัวแปร safety เช่น low, med, high

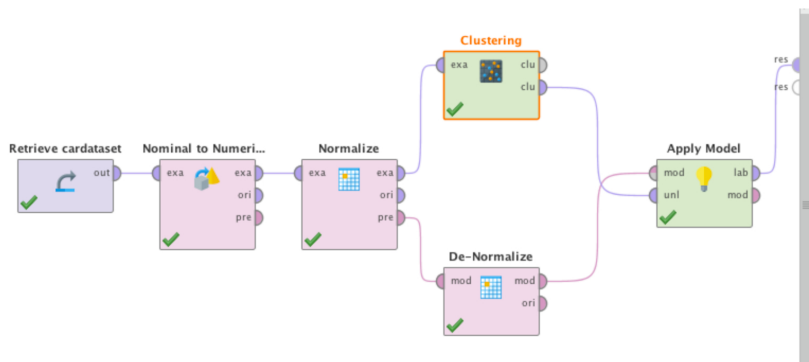
** สำหรับแลปนี้ให้ทำการตัดตัวแปร Class ออก เนื่องจากไม่สามารถนำมาใช้ในการจัดกลุ่มได้ เนื่องจากเป็นตัวแปรผลเฉลย

** คำอธิบายของข้อมูลแต่ละตัวแปรสามารถดูเพิ่มเติมได้จาก <https://archive.ics.uci.edu/ml/datasets/car+evaluation>

จากนั้น จงใช้ Rapid Miner จัดกลุ่มของการประเมินรถ ดังขั้นตอนดังต่อไปนี้

1. นำข้อมูล car dataset เข้า RapidMiner โดย Exclude column ของตัวแปรเป้าหมาย Class ให้ออกจากชุดข้อมูลนี้ ส่วนชนิดของตัวแปรนำเข้าทุกตัวให้ใช้ค่า Default
2. ทำการแปลงข้อมูลที่เป็นชนิด Nominal ให้กลายเป็นข้อมูลชนิด Numerical โดยต้องทำการตั้งค่า Parameters ตามรายการด้านล่าง
 - Attribute filter type ตั้งค่าให้เป็น all
 - Coding type ตั้งค่าให้เป็น unique integers
3. ปรับช่วงข้อมูลด้วยโหมด Normalize โดยให้เลือกใช้ z-transformation (** ปรับ Attribute filter type เป็น all)
4. จากนั้นจึงจัดกลุ่มข้อมูลด้วยโหมด Clustering (k-Means) ซึ่งมีการกำหนดค่า K=4, max runs=500, measure types=NumericalMeasures โดยกำหนดให้ numerical measures เป็น EuclideanDistance นอกจากนี้ขอให้อัตโนมัติ random seed=1992 รวมทั้งต้องเลือก determine good start values และเลือก "add as label"
5. ปรับช่วงของค่าให้กลับเป็นช่วงข้อมูลเดิม โดยการใช้โหมด De-normalize พร้อมทั้งใช้งานโดยการให้โหมด Apply Model

ภาพผลลัพธ์การต่อ Process ทั้งหมดของแลปที่ 2 จะเป็นดังภาพด้านล่าง



กดที่รูปเพื่อดูรูปขนาดใหญ่นะ

ค่า Fraction ของตัวแปร doors ที่มีค่า 3 เป็นเท่าไร

- ☐ a. 0.15
- ☒ b. 0.25
- ☐ c. 0.35
- ☐ d. 0.50

2 ค่า Absolute Count ของตัวแปร Persons ที่มีค่า 4 เป็นเท่าไร

- ☐ a. 476
- ☒ b. 576
- ☐ c. 623
- ☐ d. 723

3 ผลลัพธ์จากการทำ Z-Transformation ให้ค่า variance ของตัวแปร safety เป็นเท่าไร

- ☐ a. 0.251
- ☐ b. 0.017
- ☒ c. 0.667
- ☐ d. 0.722

4 จากตาราง Centroid Table ค่า Centroid ของตัวแปร buying ของ cluster 2 เป็นเท่าไร

- ☒ a. 0.348
- ☐ b. -0.348
- ☐ c. 0.944

☐ d. -0.944

5 จากตาราง Plot ของการจัดกลุ่มทั้ง 4 กลุ่ม กลุ่มรหัสหมายเลขใด น่าจะมีแนวโน้มในการให้ความปลอดภัยที่สูงแต่สามารถบรรจุคนได้ค่อนข้างน้อย (กลุ่มนักแข่งรถ)

- ☐ a. Cluster_1
☐ b. Cluster_2
☐ c. Cluster_3
☒ d. Cluster_0

6 จากตาราง Plot ของการจัดกลุ่มทั้ง 4 กลุ่ม กลุ่มรหัสหมายเลขใด น่าจะมีแนวโน้มในการมีกำลังซื้อรถที่ค่อนข้างสูงและเลือกรถที่สามารถบรรจุคนได้เยอะด้วย (กลุ่มคนฐานะดี ที่มีครอบครัวใหญ่)

- ☐ a. Cluster_1
☒ b. Cluster_2
☐ c. Cluster_3
☐ d. Cluster_0

7 ตัวอย่างที่ 109 ถูกจัดอยู่ในกลุ่มของ Cluster หมายเลขใดต่อไปนี้

- ☐ a. Cluster_1
☐ b. Cluster_2
☒ c. Cluster_3
☐ d. Cluster_0

The latest submission was made on 28-Feb-2020 @ 01:43:26



112.61 | [Privacy Policy](#) updated on Jun 30th, 2017

The development of myCourseVille is partially supported by [Learning Innovation Center](#), Chulalongkorn University