

Data Science Pathway Test

Part III: c. Theorem in Lectures

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

จงเลือกคำตอบที่ถูกต้องเพียง 1 ข้อ

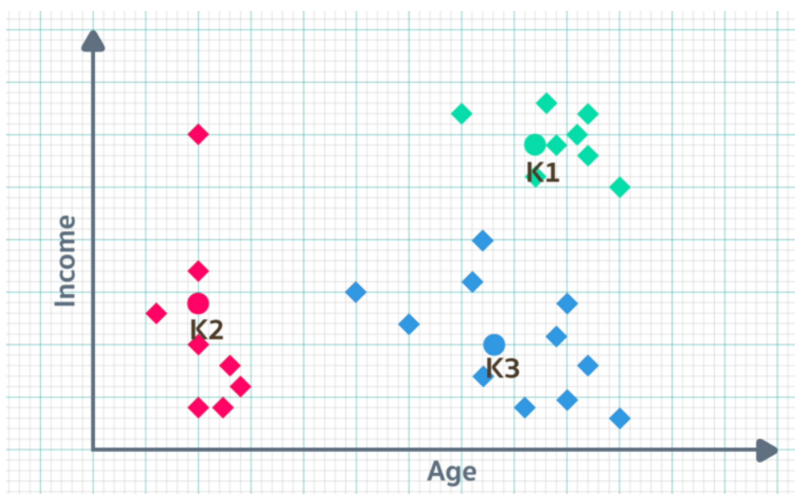
1 ข้อใดต่อไปนี้เป็นชื่อเทคนิคในการจัดกลุ่มข้อมูล (Clustering)

- ☐ a. Regression
- ☐ b. Neural Networks
- ☐ c. Decision Tree
- ☒ d. DBSCAN

2 ข้อใดต่อไปนี้เป็นสิ่งที่ผู้ใช้อ้างอิงใน K-means Clustering

- ☐ a. จำนวนกลุ่ม
- ☐ b. จำนวนรอบสูงสุดที่ทำงาน
- ☐ c. ระยะห่างขั้นต่ำระหว่างข้อมูล และตัวแทนกลุ่ม
- ☒ d. ทั้งข้อ 1 และข้อ 2

3 จากรูปผลลัพธ์การจัดกลุ่มของลูกค้าที่เข้าพักในโรงแรม ข้อใดควรเป็นคำอธิบายของกลุ่มที่ 1 (K1)



กดที่รูปเพื่อดูรูปขนาดใหญ่นะ

- ☐ a. กลุ่มที่อายุน้อย รายได้สูง
- ☐ b. กลุ่มที่อายุเยอะ รายได้ต่ำ
- ☐ c. กลุ่มที่อายุน้อย รายได้ต่ำ
- ☒ d. กลุ่มที่อายุเยอะ รายได้สูง

4 จากรูปผลลัพธ์การจัดกลุ่มในข้อที่ผ่านมา หากทางโรงแรมต้องการนำเสนอรายการนำเที่ยวแบบผาดโผน เช่น ส่องแท่ง ปีนเขา เป็นต้น ทางโรงแรมควรนำเสนอให้กับกลุ่มไหน

- ☐ a. K1
- ☒ b. K2
- ☐ c. K3
- ☐ d. ไม่มีกลุ่มใดเหมาะสม

5 ข้อใดต่อไปนี้เป็นเป้าหมายในการจัดกลุ่มข้อมูล

- ☒ a. เพื่อให้เข้าใจพฤติกรรมของลูกค้าในภาพรวม
- ☐ b. เพื่อให้ทำนายกลุ่มลูกค้าเป้าหมายได้ถูกต้อง
- ☐ c. เพื่อลดจำนวนตัวแปร
- ☐ d. เพื่อนำผลลัพธ์ของการจัดกลุ่ม ไปสร้างตัวแปรใหม่ที่เกี่ยวข้องกับกลุ่ม

6 เทคนิคในการทำนายข้อมูลใดต่อไปนี้ไม่เหมาะกับการทำ Classification

- ☐ a. Decision Tree
- ☐ b. Linear Regression

- ☒ c. Neural Networks
☐ d. k-NN

7 ผลการทำงานของ Decision Tree เกิดขึ้นที่ใด

- ☐ a. Root Node
☐ b. Internal Node
☒ c. Leaf Nodes
☐ d. โหนดใด ๆ ก็ได้

8 ข้อใดเป็นข้อจำกัด (assumptions) ของสมการ Linear Regression

- ☐ a. ตัวแปรต้น ทุกคู่ตัวแปรต้นมีความสัมพันธ์เชิงเส้นตรง
☒ b. ตัวแปรต้น และตัวแปรเป้าหมายต้องมีความสัมพันธ์เชิงเส้นตรง
☐ c. ตัวแปรต้น และตัวแปรเป้าหมายต้องมีความสัมพันธ์เชิงเส้นตรง ที่เป็นแบบ positive เท่านั้น
☐ d. ตัวแปรต้น และตัวแปรเป้าหมายต้องมีความสัมพันธ์เชิงเส้นตรง ที่เป็นแบบ negative เท่านั้น

9 ข้อใดต่อไปนี้ไม่จริงเกี่ยวกับ Neural Networks

- ☐ a. เหมาะสำหรับการสัมพันธ์แบบ non-linear
☐ b. ประกอบหลัก 3 layers หลักคือ input, hidden, และ output layers
☐ c. การเพิ่มจำนวน hidden units เป็นการเพิ่มความเป็น non-linear
☒ d. เป็นเทคนิคที่พัฒนามาต่อมาจาก Decision Tree

10 ข้อใดต่อไปนี้ไม่จริงเกี่ยวกับการทำ Prediction Model

- ☐ a. ควรทดลองพัฒนาด้วยโมเดลหลากหลายวิธี แล้วเลือกวิธีการที่ได้ประสิทธิภาพดีที่สุด
☒ b. วิธีการ Neural Networks จะได้ผลลัพธ์ดีที่สุดเสมอ
☐ c. วิธีการ Decision Tree จะเหมาะสมสำหรับการให้เหตุผลในการทำนาย
☐ d. วิธีการ Regression จะรองรับเฉพาะตัวแปรแบบตัวเลขเท่านั้น

11 จงอธิบายความหมายของค่า support=40% ของกฎความสัมพันธ์ (Pizza, Pasta) => {Soda}

- ☐ a. จากรายการซื้อสินค้าทั้งหมด มีรายการที่ตรงกับกฎข้อนี้ 40%
☐ b. จากรายการซื้อสินค้าที่มี Soda จะมี 40% ที่ซื้อทั้ง Pizza และ Pasta
☒ c. จากรายการซื้อสินค้าที่มีทั้ง Pizza และ Pasta จะมี 40% ที่ซื้อ Soda ด้วย
☐ d. จากรายการซื้อสินค้าที่มีทั้ง Pizza หรือ Pasta จะมี 40% ที่ซื้อ Soda ด้วย

12 หากต้องการส่งเสริมยอดขายของ Donut กฎข้อใดต่อไปนี้ควรถูกพิจารณาเป็นลำดับแรก

- ☐ a. {Pizza, Donut} => {Soda} with confidence=0.9
☒ b. {Donut} => {Soda, Pizza} with confidence=0.7
☐ c. {Pizza, Soda} => {Donut} with confidence=0.9
☐ d. {Pizza} => {Donut} with confidence=0.8

13 หากท่านทำงานให้กับบริษัทขายของออนไลน์ และมีสมมติฐานเรื่องการสำรวจโฆษณาในหน้าแรกของเว็บ จะสามารถช่วยเพิ่มจำนวนลูกค้าที่ซื้อสินค้าได้มากขึ้น ในกรณีนี้ท่านควรจะทดสอบสมมติฐานของท่านอย่างไรด้วยหลักการของวิทยาศาสตร์ข้อมูล

- ☒ a. ทดลองสุ่มให้ลูกค้าแต่ละคน เห็นหน้าเว็บที่ไม่เหมือนกันทั้งที่มี และไม่มีวิดีโอโฆษณา แล้วนำจำนวนลูกค้าของ 2 กลุ่มนี้มาเปรียบเทียบกับหลักการ Hypothesis Test
☐ b. สอบถามหัวหน้างาน และเพื่อนร่วมงานว่าชอบแบบใดมากกว่ากัน
☐ c. เขียนโปรแกรมจำลองสถานการณ์เพื่อประเมินสมมติฐาน
☐ d. ทำแบบสอบถาม แจกให้กับลูกค้าว่าชอบแบบใดมากกว่ากัน

14 ข้อใดต่อไปนี้ ไม่ใช่ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลของบริษัทที่ดูแลโปรแกรมสั่ง-ส่งอาหาร

- ☐ a. วิเคราะห์ปริมาณความต้องการในการสั่งอาหาร เพื่อเตรียมพนักงานให้เหมาะสมกับแต่ละช่วงเวลา
☐ b. วิเคราะห์ร้านอาหารที่ยอดนิยม แยกตามพื้นที่ เพื่อจัดทำรายการส่งเสริมการขายที่เหมาะสม
☒ c. สรุปจำนวนลูกค้าทั้งหมดที่สั่งซื้อสินค้าในแต่ละวัน เพื่อคำนวณรายได้ของบริษัท
☐ d. วิเคราะห์พฤติกรรมความชอบของลูกค้าแต่ละประเภท เพื่อจัดทำระบบแนะนำร้านอาหารที่เหมาะสม

15 ข้อใดต่อไปนี้ไม่เป็นประเภทของ Data Mining

- ☐ a. Supervised Learning
☐ b. Unsupervised Learning
☐ c. Inductive Logic Programming
☒ d. จริงทั้งข้อ 1 และ 2

