

## Data Science Pathway Test

## Part V: Data Science Pathway Capstone

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

จงเลือกคำตอบที่ถูกต้องเพียง 1 ข้อ

- 1 จงดาวน์โหลดข้อมูล Ch3Thailand\_Capstone.csv เพื่อตอบคำถามข้อ 1-7  
Ch3Thailand\_Capstone.csv Link : <https://tinyurl.com/y897ey7d>

จากข้อมูลที่กำหนดให้ จำนวน Like ในทุก Post มีค่าเฉลี่ย (Mean) และ ส่วนเบี่ยงเบนมาตรฐาน (SD) เป็นเท่าไร

- ☐ a. Mean = 2053 , SD = 1197
- ☒ b. Mean = 2053 , SD = 7196
- ☐ c. Mean = 1423 , SD = 7196
- ☐ d. Mean = 1423 , SD = 1197

- 2 จากข้อมูลที่กำหนดให้ จำนวน Comment ในทุก Post มีค่ามัธยฐาน (Median) และ ฐานนิยม (Mode) เป็นเท่าไร

- ☐ a. Median = 254, Mode = 1
- ☐ b. Median = 12, Mode = 0
- ☐ c. Median = 1089 , Mode = 1
- ☒ d. Median = 4 , Mode = 0

- 3 จากข้อมูลที่กำหนดให้ Field ใดที่มี Missing Value

- ☐ a. like

- ☐ b. Id
- ☒ c. share
- ☐ d. time
- 

4 จากข้อ 3 ควรแทนที่ Missing Value ใน Field ดังกล่าว ด้วยข้อมูลใด

- ☐ a. 0
- ☒ b. ค่าเฉลี่ย (Mean)
- ☐ c. Max - Min
- ☐ d. ไม่มีข้อถูก
- 

5 จากข้อมูลที่กำหนดให้ หากต้องการแบ่งจำนวนยอด Like ในแต่ละโพสต์เป็น น้อย กลาง และ มาก โดยใช้สองค่า ได้แก่ Mean - (SD/5) และ Mean + (SD/5) มา

- น้อย คือ มีจำนวนยอด Like น้อยกว่า Mean - (SD/5)
- กลาง คือ มีจำนวนยอด Like ระหว่าง Mean - (SD/5) ถึง Mean + (SD/5)
- มาก คือ มีจำนวนยอด Like มากกว่า Mean + (SD/5)

ค่าทั้งสองที่นำมาใช้แบ่งดังกล่าวควรมีค่าเป็นเท่าใด

- ☐ a. 1333, 2772
- ☐ b. 1093, 3012
- ☐ c. 253, 3851
- ☒ d. 613, 3492
- 

6 จากข้อ 5 โพสต์ที่มีจำนวน Like น้อย มีจำนวนเท่าใด

- ☒ a. 893
- ☐ b. 186
- ☐ c. 1021
- ☐ d. 506
- 

7 จากข้อ 5 โพสต์ที่มีสัดส่วนจำนวน Like ปานกลางต่อจำนวน Like มาก มีค่าเท่าใด ?

- ☒ a. 1.03

▼

☐

**b. 4.11**

☐

**c. 7.29**

☐

**d. 2.72**

**8** จงดาวน์โหลดข้อมูล Train.csv เพื่อสร้าง model Decision Tree สำหรับทำนาย likeRank

Train.csv Link : <https://tinyurl.com/ycd04hky>

Module Decision Tree

criterion : gain\_ratio

maximal depth : 5

confidence : 0.25

minimal gain : 0.01

minimal leaf size : 2

minimal size for split : 4

number of prepruning ... : 3

จาก Model หากข้อมูลที่ต้องการทำนาย มี Attribute ต่างๆ ดังนี้

Data A : checkHashtags = 1 , Weekday = Tuesday , timeInterval = 12-15

Data B : checkHashtags = 1 , Weekday = Friday , timeInterval = 15-18

Data A และ Data B จะถูกทำนายให้ likeRank เป็นอะไรตามลำดับ

☐

**a. Data A = Low, Data B = Low**

☒

**b. Data A = Medium , Data B = Low**

☐

**c. Data A = Medium , Data B = Medium**

☐

**d. Data A = Low, Data B = Low**

**9** จากโมเดลในข้อ 8 หากต้องการ Post ให้มีผู้เข้ามากด Like อยู่ในระดับ High จะต้องโพสต์ที่ช่วงเวลาใดของวันใด

☐

**a. 12 - 15 , Thursday**

☐

**b. 9 - 12 , Monday**

☒

**c. 6 - 9 , Monday**

☐

**d. 15 - 18 , Thursday**



1.12.6.1 | [Privacy Policy](#) updated on Jun 30<sup>th</sup>, 2017

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