

## LAB\_FINAL (20 คะแนน)

จงเขียนโปรแกรมภาษา Python ตามโจทย์ดังกล่าวเป็น Jupyter Notebook โดยให้ 1 ข้อเป็นหนึ่ง Block

สิ่งที่ต้องส่ง

- ส่งไฟล์ Jupyter Notebook (.ipynb) โดยตั้งชื่อเป็นรหัสนักศึกษา\_LAB\_FINAL (Ex 61605052\_LAB\_FINAL.ipynb) เป็นคำตอบ
- ส่งไฟล์ที่ได้จากข้อ 3 โดยตั้งชื่อเป็น รหัสนักศึกษา.csv เช่น (61605052.csv)

1. จงเขียนโปรแกรมรับค่าเข้าเป็นชื่อ - สกุล และรหัสนักศึกษาของนักศึกษา เพื่อแสดงค่าผลลัพธ์เป็นชื่อ - สกุล ตนเอง, รหัสนักศึกษา, ผลลัพธ์ของเลขโดดของรหัสนักศึกษารวมกัน, ผลรวมของเลขโดดข้างต้นหารเอาเศษด้วย 5

เช่น

### Input

Nattadet Vijaranakul

61605052

### Output

Name : Nattadet Vijaranakul

Student ID : 61605052

Sum : 25

Sum%5 : 0

2. สร้าง Numpy Array ขนาด (x,2) จากนั้นสุ่มค่าจำนวนเต็มระหว่างเลข (0 - 100) ใส่ลงไปยัง Array แต่ละช่อง เมื่อ x คือผลรวมของเลขโดดจากข้อ 1. และแสดงค่าทั้งหมดที่อยู่ใน Array  
2.1. ใช้ np.random.rand(array\_size) เพื่อ random ค่า 0 - 1(ค่าเป็น 0.xxxx) ใส่ Array

2.2. เขียนโปรแกรมอีก 2 บรรทัดจะทำให้ค่าใน Array เป็น 0 – 100 และแสดงค่าที่อยู่ใน Array

3. ให้นำ Numpy Array ที่ได้มาสร้างเป็น Pandas DataFrame โดยให้ Column แรก ชื่อ x และ Column ที่สอง ชื่อ y จากนั้นทำการ Export DataFrame ที่ได้ออกเป็นไฟล์ชื่อ รหัสนักศึกษา.csv (เช่น 61605052.csv) โดยไม่ต้อง Save เลข index ออกมาด้วย

4. ให้อ่านไฟล์ดังนี้แล้วโหลดขึ้นมาเป็น DataFrame และแสดงผลที่ได้

4.1. คนที่ค่าผลรวมเลขโดดจากข้อ 1. เป็นเลขคู่ให้อ่านไฟล์ Classification\_Data.xlsx ใน Sheet ที่ชื่อว่า Sport ส่วนคนที่ผลรวมจากข้อ 1. เป็นเลขคี่ ให้อ่านไฟล์ Classification\_Data.xlsx ใน Sheet ที่ชื่อว่า Fever

5. ให้โหลดข้อมูลที่ Export ไว้ในข้อ 3 ขึ้นมาเป็น Pandas DataFrame จากนั้นให้แปลง DataFrame ดังกล่าว เป็น Array ด้วย function to\_numpy() จากนั้นให้ทำการ นำ Array ที่ได้มา Plot Graph Scatter ด้วย matplotlib

6. ให้นำข้อมูลที่ได้จากข้อ 4.1. มา Plot ดังนี้

คนที่ได้ค่า Sum%5 จากข้อ 1. เป็นเลขคู่ให้ plot ข้อมูลแต่ละ Feature ออกเป็น Pie Chart

คนที่ได้ค่า Sum%5 จากข้อ 1. เป็นเลขคี่ให้ plot ข้อมูลแต่ละ Feature เป็น Bar Chart ด้วย matplotlib

7. ให้นำข้อมูลที่ได้จากข้อ 5. มา plot Graph Line Plot ด้วย matplotlib

8. Regression Model โดย Model ที่สร้างให้ใช้ข้อมูลที่ได้จากข้อที่ 5 โดยให้ x เป็น train และ y เป็น validation

คนที่ได้ค่า Sum%5 = 0,1 ให้สร้างเป็น Linear Regression

คนที่ได้ค่า Sum%5 = 2,3 ให้สร้างเป็น Decision Tree Regression

คนที่ได้ค่า Sum%5 = 4 ให้สร้างเป็น KNNRegression (K = 2)

จากนั้นให้ Predict ค่าดังนี้ `model.predict(a)` และ `model.predict(b)` เมื่อ a เป็นผลรวมของเลขโดด 4 ตัวแรก และ b ผลรวมของเลขโดด 4 ตัวหลัง ของรหัสนักศึกษา เช่น 61605052 → a = 13, b = 12 เป็นต้น

จากนั้นให้แสดงค่า Coefficient of Determination ( $R^2$ ) ของ Model ที่ได้

9. Classification Model ให้ใช้ข้อมูลที่ได้จากข้อ 4 โดยจะต้องทำการ Label ค่าที่เป็น String ใหม่ให้เป็นค่า Integer โดย Label อย่างไรให้ Comment ไว้ด้วย (Label ตอนเป็น Data Frame ง่ายกว่า แล้วค่อยแปลงเป็น Array)

คนที่ได้ค่า  $\text{Sum} \% 5 = 0$  ให้สร้างเป็น Decision Tree Classification

คนที่ได้ค่า  $\text{Sum} \% 5 = 1$  ให้สร้างเป็น Naïve Bayes Classification

คนที่ได้ค่า  $\text{Sum} \% 5 = 2$  ให้สร้างเป็น KNN Classification ( $K = 2$ )

คนที่ได้ค่า  $\text{Sum} \% 5 = 3$  ให้สร้างเป็น Random Forest Classification (criterion = “gini”)

คนที่ได้ค่า  $\text{Sum} \% 5 = 4$  ให้สร้างเป็น Random Forest Classification (criterion = “entropy”)

จากนั้นให้ทำนายค่าดังนี้

จากข้อ 4 หากใครได้ Load Sheet ที่ชื่อว่า Sport ให้ predict ค่า (Sunny, 5 to 10, Indoor)

จากข้อ 4 หากใครได้ Load Sheet ที่ชื่อว่า Fever ให้ predict ค่า (37, Yes, No)

จากนั้นให้แสดง Probability ของแต่ละ class เมื่อทำนายข้อมูลจากข้อมูลข้างต้น

10. Unsupervised Learning ให้นำข้อมูลจากข้อที่โหลดจากข้อที่ 5 มาสร้าง Model ดังนี้

คนที่ได้ค่า  $\text{Sum} \% 5 = 0$  ให้สร้างเป็น Hierarchical Clustering โดยให้จำนวน Cluster = 4

คนที่ได้ค่า  $\text{Sum} \% 5 = 1$  ให้สร้างเป็น Hierarchical Clustering โดยให้จำนวน Cluster = 5

คนที่ได้ค่า  $\text{Sum} \% 5 = 2$  ให้สร้างเป็น KMean Clustering โดยให้จำนวน Cluster = 5

คนที่ได้ค่า  $\text{Sum} \% 5 = 3$  ให้สร้างเป็น KMean Clustering โดยให้จำนวน Cluster = 4

คนที่ได้ค่า  $\text{Sum} \% 5 = 4$  ให้สร้างเป็น DBSCAN โดยให้  $\text{eps} = 10$ ,  $\text{min\_sample} = 2$

จากนั้นให้แสดง Cluster Label และทำนายผล `model.fit_predict([[a,b]])` หรือ `model.predict([[a,b]])` เมื่อ a เป็นผลรวมของเลขโดด 4 ตัวแรก และ b ผลรวมของเลขโดด 4 ตัวหลัง ของรหัสนักศึกษา เช่น 61605052  $\rightarrow$  a = 13, b = 12 เป็นต้น