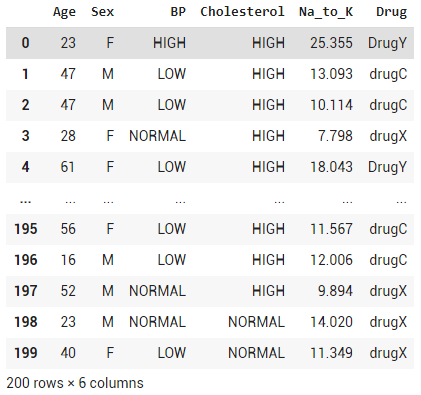
นายฟาริซ หะยีมะสาและ รหัส 61050258

**งานวิชา Machine Learning (แทนข้อสอบ Final)**

****

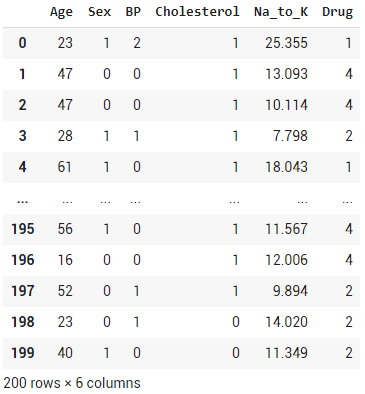
**อธิบาย dataset**

Dataset ที่เลือกใช้เกี่ยวกับการแยกประเภทยาของผู้ป่วยแต่ละคนโดย จะแยกประเภทของยาออกเป็น 5 ประเภท คือ DrugY , drugX , drugA , drugC , drugB และมีคอลัมที่ใช้ประกอบการแยกประเภท 5 คอลัม คือ Age(อายุของผู้ป่วย) , Sex(เพศของผู้ป่วย) , BP(ระดับความดันโลหิต) โดยความดันโลหิตจะแบ่งเป็น ความดันต่ำ , ความดันปกติ และความดันสูง , Cholesterol(ระดับคอเลสเตอรอล) โดยระดับคอเลสเตอรอลแบ่งเป็น คอเลสเตอรอลปกติ และคอเลสเตอรอลสูง , Na\_to\_K(อัตราส่วนโซเดียมต่อโพแทสเซียมในเลือด)

หมายเหตุ : ตัว dataset ไม่ได้ระบุว่าประเภทของยาแต่ละประเภทหมายถึงยาอะไร

**อธิบายการ dummy ข้อมูล จาก String เป็น ตัวเลข**

Colume Sex(เพศของผู้ป่วย) แทนค่า M เป็น 0 และ F เป็น 1  
Colume BP(ระดับความดันโลหิต) แทนค่า LOW เป็น 0 , NORMAL เป็น 1 และ HIGH เป็น 2  
Colume Cholesterol(ระดับคอเลสเตอรอล) แทนค่า NORMAL เป็น 0 และ HIGH เป็น 1  
Colume Drug(ประเภทของยา) แทนค่า DrugY เป็น 1 , drugX เป็น 2 , drugA เป็น 3 , drugC เป็น 4 และ drugB เป็น 5



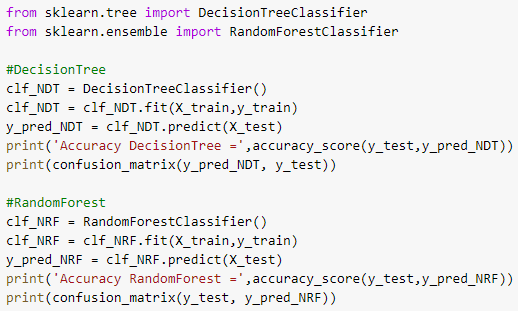
หลังจาก dummy ข้อมูล

**อธิบายการทดลอง**

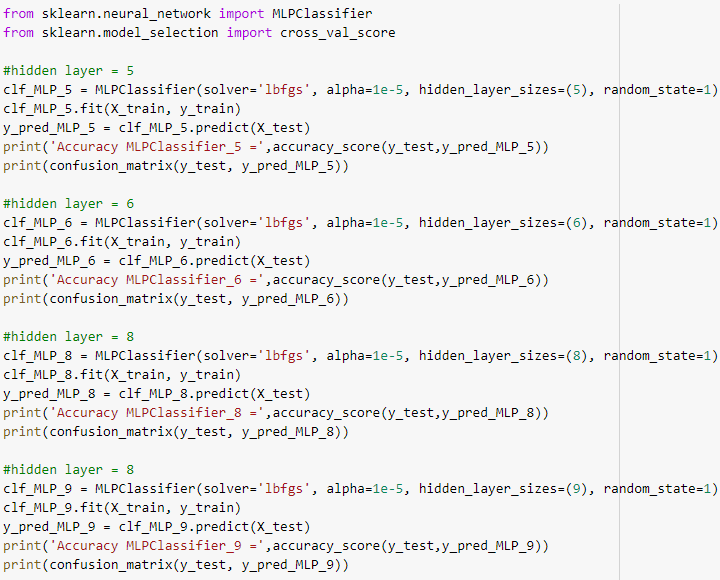
Direct Classification .ใช้โมเดล K-nearest-neighbors โดยทดลองทำโดยการหา N ที่เหมาะสมกับ dataset ที่สุด ทดลองใช้ N เป็น 3, 5, 10 และดูผลลัพธ์



Traditional Classification .ใช้โมเดล Decision Tree และ Random Forest ในการทดลอง



Deep Learning ใช้โมเดล แบบ Multi-layer Perceptron ซึ่งเป็น Model Neural network ประเภทหนึ่ง และได้ทำการทดลองด้วยการปรับค่า hidden layer เท่ากับ 5, 6, 8, 9 ตามลำดับและดูผลลัพธ์



**ผลการทดลองของแต่ละโมเดล**

Direct Classification ได้ผลการทดลอง N ที่เหมาะสมกับ dataset ที่สุด คือ N=5 ได้ Accuracy เท่ากับ 0.76667







Traditional Classification ได้ผลการทดลองของทั้ง Decision Tree และ Random Forest เท่ากัน ได้ค่า Accuracy เท่ากับ 1 มีความแม่นยำในการทำนาย 100%





Deep Learning ได้ผลการทดลอง ของโมเดลแบบ Multi-layer Perceptron ได้ค่า Accuracy มากที่สุดเท่ากับ 0.9333 ในการใส่พารามิเตอร์ hidden layer เท่ากับ 8









**สรุปผลและอภิปรายผลการทดลอง**

สรุป Direst Classification

หลังจากนำ dataset ไปทำ Direct Classification พบว่ายิ่งใช้จำนวน N เพิ่มมากขึ้นมาโอกาศทำให้การทำนายผิดพลาดได้ เพราะจำนวนข้อมูลที่นำเข้ามาแต่ละ class มีไม่เท่ากันจึงอาจไปเจอ class ที่มีตัวนวนตัวอยู่ใกล้มากกว่าและทำให้การทำนายผิดพลายได้ และการที่จำนวน N น้อยเกินไปก็อาจทำให้ไม่ความแม่นยำลดลงได้เช่นเดียวกัน เพราะตัวที่อยู่ใกล้อาจมีจำนวนข้อมูลของแต่ละ class เท่ากัน จึงต้องเลือกใช้จำนวน N ที่เหมาะสมกับแต่ละ dataset โดยใน Dataset นี้ คือ N=5 จะให้ Accuracy เท่ากับ 0.76666667

## สรุป Traditional Classification

ผลการนำ dataset ไปเข้า model Decision Tree และ Random Forest ได้ค่า Accuracy และ confusion matrix เท่ากัน ค่า Accuracy เท่ากับ 1 ซึ่งหมายความว่า model มีการทำนายที่แม่นยำ ไม่มีข้อผิดพลาดเลย ซึ่ง Decision Tree และ Random Forest เป็นการนำ dataset ไปเข้า model เพื่อสร้างต้นไม้ออกมา เพื่อแยกประเภทของแต่ละclass ซึ่งแสดงว่า dataset นี้เหมาะกับการทำ Traditional Classification แบบ Decision Tree และ Random Forest

## สรุป Deep Learning

Deep Learning ได้ Accuracy มากที่สุดเท่ากับ 0.9333 ในการใส่พารามิเตอร์ hidden layer เท่ากับ 8 โดยจากการทดลองทำให้เห็นว่า dataset ชุดนี้เหมาะกับการใช้ hidden layer เท่ากับ 8 เพราะเมื่อเพิ่มจำนวน hidden layer ไปเป็น 9 ทำให้ค่า Accuracy ลดลง

สรุปผลการทดลองทั้งหมด

จากการทดลองนำ dataset ไปเข้าโมเดลทั้งแบบ Direst Classification , Traditional Classification และ Deep Learning ได้ข้อสรุปว่า dataset ชุดนี้เหมาะกับการทำ Traditional Classification ที่สุด เพราะได้ค่า Accuracy ดีที่สุด คือ 1 ซึ่งสามารถทำนายข้อมูล test ถูกต้อง 100%