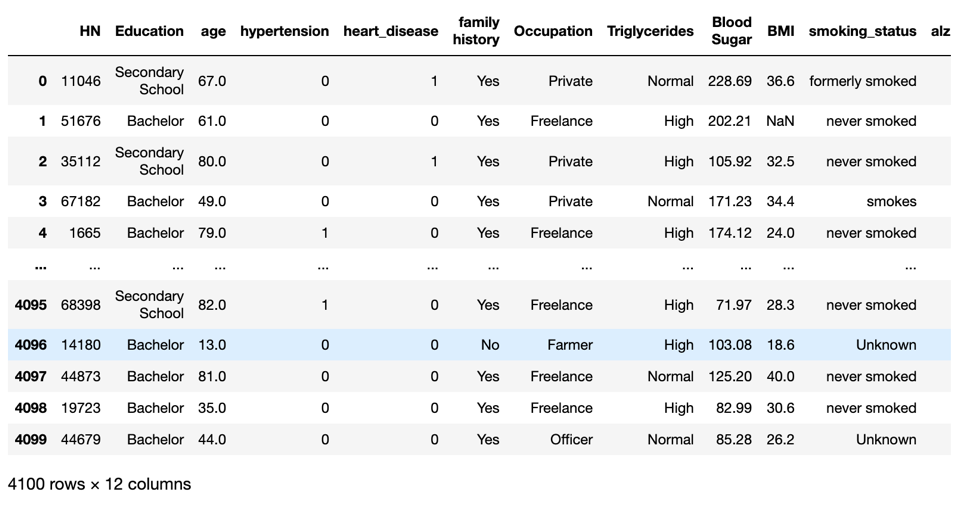
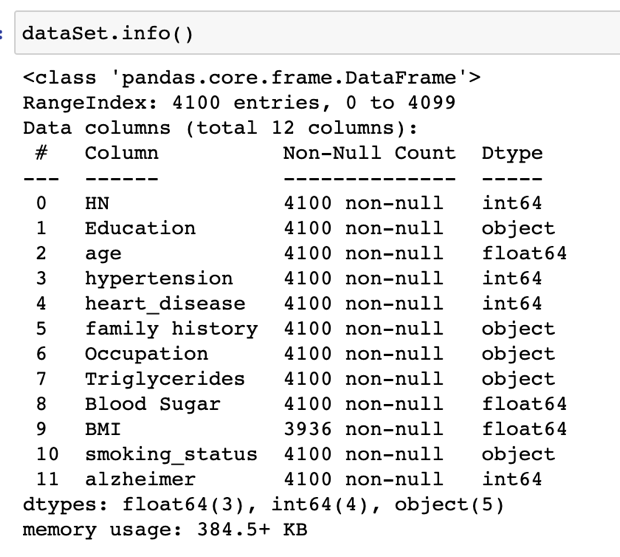
**6614450042**

**ปัญญวัตร สุวรรณทัต**

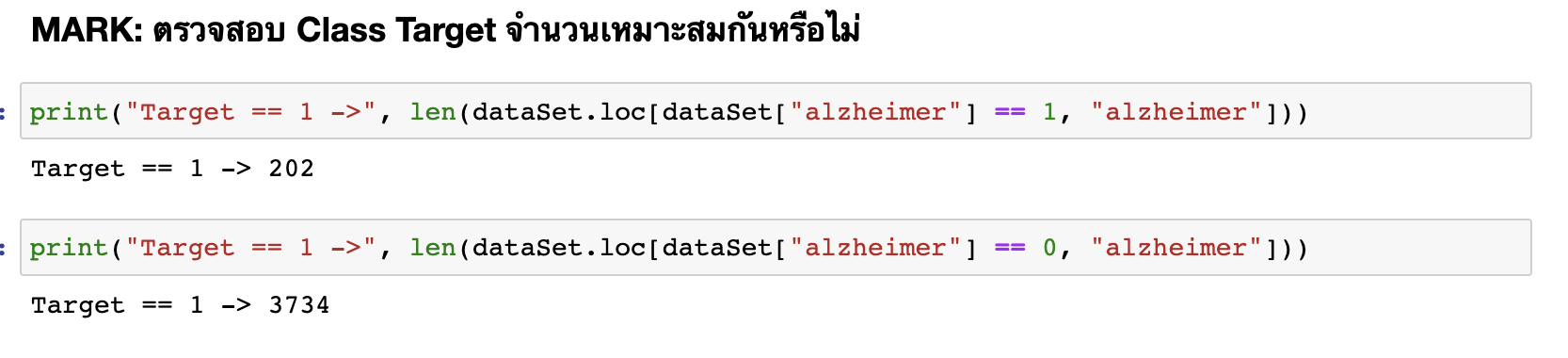
1. ตรวจสอบ Data Set Alzheimer.csv มีลักษณะข้อมูลเป็นอย่างไร type ข้อมูลแบบใด



ตรวจสอบข้อมูล เช่น จำนวน row, column

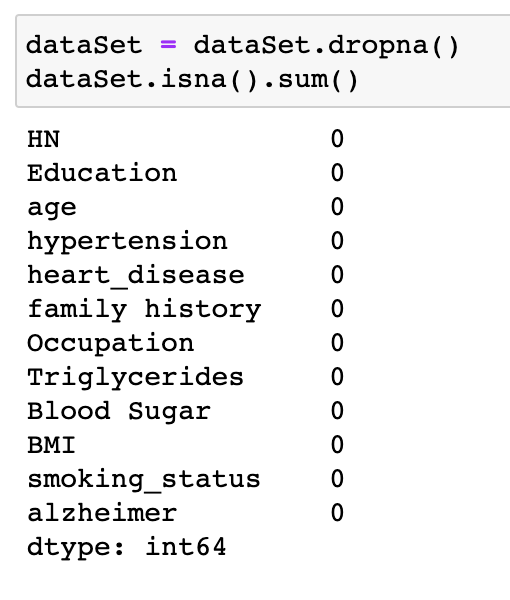


ตรวจสอบ type ข้อมูล



ตรวจสอบความ Balance ของ Class target

1. ทำ Data Processing



Drop NA ของ Column BMI ทิ้ง เพราะไม่มีส่วนสูงและน้ำหนักให้คำนวณ รวมไปถึง ค่า NA ใน Column BMI นั้น มีไม่เยอะมาก ไม่น่าส่งผลกับ model เท่าที่ควร



เปลี่ยนข้อมูล Nominal ให้อยู่ในรูป Labelencode

1. ทำการ SMOTE ข้อมูล เพื่อให้จำนวน Class Target เหมาะสมกัน
2. ทําการแบ่งชุดข้อมูลฝึกฝนและชุดข้อมูลทดสอบโดยกําหนดชุด ข้อมูลฝึกฝนมี 70% และชุดข้อมูลทดสอบมี 30%
3. ทำการฝึกฝน Model โดยเลือก 3 model ดังนี้
   1. Decision tree
   2. K nearest neighbors
   3. Neural Network

โดยทำการทดฃอง model ละ 2 ครั้ง โดยครั้งแรกเลือก feature ทั้งหมด และครั้งที่ 2 เลือก มาเฉพาะ Feature importance 4 อันดับแรก

1. ผลการทดลอง

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Feature** | **Parameter** | **accuracy** | **precision** | **recall** | **F1** |
| Decision tree | All | Default | 0.8951 | 0.9 | 0.9 | 0.9 |
| Top 4 | Default | 0.8911 | 0.89 | 0.89 | 0.89 |
| Top 4 | max\_depth = 9 | 0.8380 | 0.84 | 0.84 | 0.84 |
| K nearest neighbors | All | Default  K = 1 | 0.8920 | 0.9 | 0.89 | 0.89 |
| Top 4 | Default  K = 1 | 0.8741 | 0.88 | 0.87 | 0.87 |
| Neural Network | All | Default  hidden\_layer\_sizes(64,32) | 0.8995 | 0.9 | 0.9 | 0.9 |
| Top 4 | Default  hidden\_layer\_sizes(64,32) | 0.8500 | 0.86 | 0.85 | 0.85 |

จากผลการทดลองจะเห็นได้ว่า Neural network แบบเลือก Feature ทั้งหมด Parameter แบบ Default ให้ความแม่นยำอยู่ที่ 89.95% ลำดับที่สองคือ Decision tree แบบเลือก Feature ทั้งหมด Parameter แบบ Default ให้ค่าความแม่นยำ อยู่ที่ 89.51% และสุดท้าย KNN เลือก Feature ทั้งหมด Parameter แบบ Default ให้ค่าความแม่นยำอยู่ที่ 89.20%