Introduction to Machine Learning

Lab 2: Mortality Prediction using Decision Tree and Random Forest

Student ID: 111060013   Name: 劉祐廷

1. Attributes Setting

ATTRIBUTE	MAX DEPTH	MIN SAMPLE SPLIT	TREES	FEATURES	SAMPLE SIZE
VALUE	7	10	100	5	0.7

For better stability, I used **100 trees** to build the random forest. And I choose **5 features** randomly because some articles say that the square root of the number of instances is usually the best number of the features. I didn't split the data for validation, therefore **the number of instances** I used to build each tree is 8000 * 0.7 = **5600**.

2. Difficulty

At first, I wasn't sure how to tune the parameters, and I thought that increasing the number of trees and features would automatically improve the model. As a result, I spent a lot of time waiting for the program to finish running, but the results were often disappointing. After adjusting some parameters, I had to wait again for the program to finish executing, which took a lot of time. Additionally, I'm not very familiar with the operations in pandas, so I spent quite some time learning how to use it.

3. How I Solve the Difficulty and My Reflections

I'm grateful to live in an era where ChatGPT-4o is available, which saves me from spending excessive time searching for how to use pandas. Now, I can simply ask AI how to implement the functions I need, avoiding the irrelevant information that often came up in past searches. As for parameter tuning, I started by asking ChatGPT and used its suggestions as a foundation to guide my further research. Gradually, I found better ways to adjust the parameters, allowing my model to achieve stable and solid performance.