Reflection

In this week, I learned about classification using the same dataset from previous week, fraud detection (0 and 1) but with a different model; Logistic Regression. In this model, I learn that the prediction process is similar to Linear Regression which use a linear formula: $y = w_1x_1 + w_2x_2 + b$

However, Linear Regression predicts continuous values, whereas Logistic Regression applies a sigmoid function to the result, which constrains the output to a range between 0 and 1. By using 0.5 as the cutoff point, predictions above 0.5 are classified into one class, and predictions below 0.5 into the other.

Through the coding, I explored more about the parameter for logistic regression and tried with different parameter and notice that the precision dropped and recall decrerase when adding the class_weight='balanced'. I found that the reason may be because of the trade-off between precision and recall in imbalanced datasets.