Project1 Review

Summary

First, the LightGBM model is used to fit and predict the data, calculate the accuracy and extract the importance of each feature. The report then picked out features that were more important and used the model to fit the prediction again, and the result was slightly better than before.

Strengths

In the face of a large number of data and missing values, the report skillfully uses the LightGBM model to process and fit the data, simplifying the operation steps without reducing the accuracy of prediction.

In addition, there is no over-fitting problem common to LightGBM model in the process of model fitting, and the ROC predicted by the model is good.

Weaknesses

One-hot coding is a general approach to handling category features, however it may not necessarily be a good approach in tree models. When there are many class values, this encoding method will result in less data for each class, which will cause an imbalance in sharding, which means that the sharding gain will also be small.

Score

Evaluation on Clarity and quality of writing (1-5): 4
Evaluation on Technical Quality (1-5): 4
Overall rating: 4
Confidence on your assessment (1-3): 3