Reflections or Learning Outcomes

During this learning process, I gained a deeper understanding of how to use decision trees, random forests, and boosting algorithms to analyze e-commerce customer data. The following is a summary of my reflections in the learning environment:

What you learned:
1. The importance of data preprocessing: This time there were about 176 missing values in the reflow, and there were duplicates in some categories (e.g. age). I learned how to deal with missing data, encode categorical variables and standardize numerical data. This is to ensure The basic steps for the model to learn effectively.
2. Model selection and application: I understood why random forests and boosting methods were chosen to process complex customer behavior data, and learned the principles behind these algorithms.
3. Model evaluation: I learned to use different performance metrics to evaluate models and select the best model by comparing these metrics.

Challenges faced:
1. Software operation: Because I am using a Mac under the OS system, the software is not compatible, so I used a virtual machine to build it. However, because I am not familiar with the operating system, I don't understand why the Chinese version causes some trouble.
2. Model interpretation: How to reasonably explain the model after it is set up poses a challenge to provide easy-to-understand business insights.
3. Parameter adjustment: With limited computing resources, finding the optimal model parameter configuration is a challenge.

How to overcome these challenges:
1. Software operation: By viewing the past learning videos and the teacher's explanations in class, the SAS software can run smoothly. Because the version is incompatible, I chose the cloud service to complete the data preprocessing.
2. Model simplification and selection: When the model is too complex and difficult to interpret, I will try to simplify the model structure, or select a model with better interpretability, such as a decision tree.

In short, this process not only deepened my understanding of machine learning algorithms, but also improved my ability to solve practical problems. By working through each challenge, I improved my problem-solving skills and became better able to apply theoretical knowledge to real-world situations.