



**UNIVERSITI
MALAYA**

WQD7005 DATA MINING

2023/2024 S1

Learning Outcomes

Matric Number: S2164046

Name: HU LIANGLIANG

Learned

- I learned how to effectively clean and preprocess complex data using different tools (SAS EM, Talend), especially in dealing with missing values, outliers and data transformations. Understand the importance of feature engineering, especially in selecting and constructing features that have a significant impact on model predictions.
- I mastered the application of various models (such as decision trees, random forests, gradient boosted trees) and understood their performance differences on different types of data sets through the interpretation of output results and images. Learned how to adjust strategies based on model results and optimize parameters to improve the prediction accuracy of the model.
- I mastered extracting valuable insights from complex data. Through iterative analysis and continuous experimentation, I learned to deal with complex data problems more effectively. It also deepens the understanding of the business background and can better combine data analysis results with actual business needs.

Challenges encountered:

Using different tools was a bit complicated for me at the beginning, but after carefully reading the courseware, I became very familiar with it.

Interpreting the output of different model results will be a bit complicated, especially at the beginning when the decision tree results were not satisfactory. Fortunately, other models were built later and finally the ideal model was obtained and a high-quality analysis was output.

Conclusion

Through this project, I improved my technical skills in data processing and model building, mastered the application of different tools, and gained valuable experience in solving practical business problems. The challenges in the project prompted me to continue learning, gave me a deeper understanding of the complexity and dynamics of data science, and strengthened my problem-solving strategies and methods. Overall, it

not only improved my data analysis skills, but also deepened our understanding of the relationship between business needs and data, allowing me to better know how to apply data to real-life scenarios.