Process Report in Assignment 2

Group 77- Yulu Wang(2727657) Yunshan Wang(2721621) Siqi Wei(2724831)

Vrije Universiteit, Amsterdam, Netherlands

1 Time Schedule

Table 1. Project Timeline

Period	Work Content
01/05-03/05	Explore the dataset & Find methods for memory optimization
04/05-05/05	Basic data cleaning
06/05-08/05	Looking for suitable approaches from materials
09/05-10/05	Finish data preprocessing
11/05	Discuss feature selection
12/05	Feature Engineering-establishing features
13/05-14/05	Feature Engineering-merge and generate final dataset
15/05	Second EDA
16/05	Use LightGBM for basic model training and prediction in label
17/05	Use XGBoost for basic model training and predictio in label
18/05-19/05	Try to use labelall in both models
20/05-23/05	Do analysis in feature significance in both models
16/05-27/05	Repeated tuning & Submit predictions to Kaggle for testing
18/05-27/05	Complete Reports

2 Overall cooperation reflection

All three members of Group 77 were very active in group meetings and discussions. In order to ensure the smoothness of communication, we have established a small group. No matter who asked a question or has a new idea, all team members would reply as soon as possible. When encountering more complex content, we would set up a zoom meeting or have a face-to-face discussion. All study materials, including related projects and references, were shared in our group for reading comprehension.

Our team members are very friendly, everyone was full of responsibility, and all decisions were made after a thorough discussion among three people. Everyone actively wanted to contribute to the group. Our groups meet on campus with problems, asked each other for advice, and exchanged ideas on projects together.

In addition, we have asked questions from experienced programmers to help us solve the difficulties encountered during the project. We also exchanged opinions on model selection, feature engineering, and parameter tuning strategies

through some social platforms or face-to-face with other teams and students, and obtained many meaningful suggestions.

We are very happy to have such a positive and open atmosphere, and everyone believes that everyone shares the knowledge contribution of the project.

3 Contribution

Table 2. Contribution

Work Content	Name
Explore the dataset(First EDA)	Yunshan Wang
Basic data cleaning	Yunshan Wang
Find methods for memory optimization	Yulu Wang
Looking for suitable approaches from materials	Siqi Wei
Data preprocessing before feature engineering	Siqi Wei
Feature engineering process	Yunshan Wang&Yulu Wang
Second EDA	Siqi Wei
LightGBM model constraction	Yunshan Wang
XGBoost model constraction	Siqi Wei
Model Tuning	Yunshan Wang&Yulu Wang&Siqi Wei
Labelall in training and prediction	Yulu Wang
Analysis in feature significance	Yunshan Wang&Yulu Wang&Siqi Wei
Reports	Yunshan Wang&Yulu Wang&Siqi Wei