**1. Project introduction**

**(1) Objectives**

The primary objective of our graduation project is to develop a robust and efficient fraud detection model for identifying fraudulent customer transactions in the e-commerce domain.

To accomplish this goal, the project will employ machine learning and deep learning techniques to analyze Vesta's real-world e-commerce transaction data, which contains a diverse array of features. By harnessing the power of these advanced algorithms, the project seeks to create an accurate and reliable model capable of identifying suspicious activities and minimizing the occurrence of false positives.

Another important objective of our project is to explore the possibility of generating new features that may enhance the model's performance. This will involve a comprehensive analysis of the available data and the creative application of domain knowledge to uncover patterns and relationships that may be indicative of fraudulent transactions.

Ultimately, we hope that our model will help businesses reduce fraud-related losses, increase revenue, and bolster customer trust in reality.

**(2) Scope**

**Dataset**: IEEE-CIS Fraud Detection dataset provided by the world’s leading payment service company Vesta.

**Models:** Machine learning and deep learning models.

**Application area:** customer credit analysis in banks

We plan to use machine learning and deep learning models to identify customer frauds in credit card lending based on the Kaggle IEEE dataset. Providing tools for scientific decision-making for banks to help them minimize the risk of losing money while banks lending to customers.

**(3) Schedule**

|  |  |  |  |
| --- | --- | --- | --- |
| **Deadline** | **Event** | **Details** | **Estimated hours** |
| 2023/4/20 | Hand in detailed Proposal | Read documents, learn research methods | 200 |
| Complete data preprocessing part | Data preprocessing | 90 |
| Construction of the basic webpage framework | Learn the methods of constructing webpages | 100 |
| - | Writing detailed proposal | 50 |
| 2023/6/1 | Hand in interim report | Writing interim report | 70 |
| Complete the report slides | Making interim report slides | 50 |
| Add data preprocessing part and proposal to the webpage | Read documents to prepare for model choosing | 100 |
| - | Webpage modifying | 50 |
| 2023/7/1 | Complete model construction and evaluation part | Model construction and evaluation | 100 |
| Add model construction and evaluation part to the webpage | Webpage modifying | 50 |
| 2023/7/15 | Complete final report | Writing final report | 80 |
| Complete webpage | Webpage modifying | 80 |
| Complete final report slides | Making final report slides | 70 |
| 2023/8/1 | Get all the materials modified and submitted | Modifying materials according to supervisor's comments | 60 |
| 2023/8/15 | Get all the materials revised and submitted | Modifying materials according to examiner's comments | 50 |
| - | - | **Total** | **1200** |

***Q1 do we need to use react to construct webpage? Can we use github ?***

***Q2 How and how often do we present the project progress to the supervisor? Zoom meeting or webpage? Webpage could take longer time.***

**(4) Data**

**Kaggle website:**

<https://www.kaggle.com/competitions/ieee-fraud-detection/overview>

**(5) Models**

**Initially choose some models, need to be determined by reading more cutting-edge documents**

**1) ML**

Logistic regression

Adaptive boosting

**2)** **DL**

Gradient boosting

Neural Networks

**2. Work Distribution**

**(1) Code**

**Data Processing**

Li Jiayao, Xia Linlong

**Models & Evaluation**

Each person is responsible for one model

**(2) Report**

**Proposal**

Li Jiayao, Xia Linlong

**Webpage**

Li Dongheng, Tang Yutian

**Interim report**

Each person responsible for their part

**Interim pre slides**

Each person responsible for their part

**Final Report**

Each person responsible for their part

**Final pre slides**

Each person responsible for their part