IV. 数据描述与准备

Data description and preparation

A. 数据描述

该项目中使用的数据集来自布里斯托尔劳埃德银行集团首席数据和分析办公室机器学习工程师 Marie Anderson 博士。 原始交易数据提供客户交互的记录，包括日期、时间、金额、模式和相关位置。 出于保密原因，原始交易数据经过基于代理的模拟，以生成个人级别的交易数据，组成了我们在本项目中使用的数据集。

模拟交易数据集中，共有230,596行交易数据和7个变量：交易日期（Date）、交易时间（Timestamp）、账号（Account No）、账户余额（Balance）、交易金额（Amount）、第三方 帐号（第三方帐号）和第三方名称（第三方名称）。

The dataset used in this project comes from Dr Marie Anderson, Machine Learning Engineer, Chief Data and Analytics Office, Lloyds Banking Group, Bristol. Raw transaction data provides a record of customer interactions, including date, time, amount, pattern and relative location. For confidentiality reasons, the raw transaction data were subjected to agent-based simulations to generate individual-level transaction data, which make up the dataset we use in this project.

In the simulated transaction dataset, there are 230,596 rows of transaction data and seven variables: transaction date (Date), transaction time (Timestamp), account number (Account No), account balance (Balance), transaction amount (Amount), third party account number (Third Party Account No), and third party name (Third Party Name).