I. INTRODUCTION

In modern life, credit card consumption and online shopping have become more and more frequent, and a large amount of daily transaction data has also been generated. For the banking industry, proper management and rational use of this data is one of the keys to prevent financial risks and improve customer service quality. Although the potential of data is huge, it is not easy to extract useful information accurately from the huge and chaotic transaction data and use it as the basis for business improvement.

Lloyd's Banking Group (LBG), one of the largest financial services organisations in the UK, generates a large amount of customer transaction data every day. These data contain extremely rich transaction details, if it can be properly analyzed and used, it will not only greatly improve customer satisfaction, but also gain more benefits for the bank. Based on this need, this research project was born to demonstrate how machine learning techniques can be used to conduct in-depth analysis of simulated transaction data provided by LBG through a comprehensive data analysis framework. Through this project, we will conduct data preprocessing, visual analysis, transaction amount prediction, and build a personalized recommendation system based on user consumption behavior.

The first step of this research is to conduct accurate pre-processing of the transaction data. Since the data source contains some missing values, which is also a possible problem in the actual transaction or data preservation, we converted the time stamp and explored the method of data repair, so as to provide more accurate and complete data for the subsequent analysis. Next, the project uses data visualization techniques to detail time series analysis of transaction activity to help banks identify and understand patterns and trends in consumer behavior. In terms of transaction amount prediction, this project compares a variety of machine learning models, such as support vector machines, decision trees, random forests, linear regression and multi-layer perceptrons, in order to select the optimal model to predict the future transaction amount, which supports the transaction risk prediction of banks. In addition, the classification based recommendation system developed in this research can recommend potential users for merchants by analyzing users' consumption behaviors, and recommend merchants that cardholders may be interested in, so as to improve customers' personalized experience and satisfaction.

在现代生活中，刷卡消费、网络购物变得越来越频繁，大量的日常交易数据也随之产生。对于银行业来说，妥善管理并合理利用这些数据是预防金融风险、提升客户服务质量的关键之一。虽然数据的潜力巨大，但是要从庞大且混乱的交易数据中准确地提取有用的信息，并把它们作为业务提升的基础，并非易事。

洛伊兹银行集团（LBG）作为英国最大的金融服务机构之一，每天都会产生大量的客户交易数据。这些数据包含着极为丰富的交易细节，如果能被妥善分析和利用，不仅会极大地提升客户满意度，而且可以为银行收获更多的利益。基于这一需求，本研究项目应运而生，旨在通过一个综合的数据分析框架，展示如何利用机器学习技术对LBG提供的模拟交易数据进行深入分析。通过此项目，我们将进行数据预处理、可视化分析、交易金额预测，并建立一个基于用户消费行为的个性化推荐系统。

本研究的第一步是对交易数据进行精确的预处理。由于数据源包含一些缺失值，这也是实际交易或数据保存中可能出现的问题，所以我们对时间戳进行转换，并且探索了数据修复的方法，为后续的分析工作提供了较为准确和完成的数据。接下来，项目利用数据可视化技术，详细展示了交易活动的时间序列分析，帮助银行识别和理解消费行为的模式和趋势。在交易金额预测方面，本项目比较了多种机器学习模型，如支持向量机、决策树、随机森林、线性回归和多层感知机，目的是选择最优模型以预测未来的交易金额，这对银行的交易风险预判起到支持作用。此外，本研究开发的基于分类的推荐系统，通过分析用户的消费行为，为商家推荐潜在用户，同时为持卡者推荐他们可能感兴趣的商家，以此提高客户的个性化体验和满意度。