Lloyds Banking Group: Analysis of retail-banking customer data

Ali Almalki
Faculty of Engineering
University of Bristol
Bristol, United Kingdom
ei20779@bristol.ac.uk

Ghadah Almuqrin
Faculty of Engineering
University of Bristol
Bristol, United Kingdom
cq20763@bristol.ac.uk

Haoge Liu
Faculty of Engineering
University of Bristol
Bristol, United Kingdom
py20346@bristol.ac.uk

***Abstract*— The project deals with two synthetic datasets provided by Lloyds Banking Group (LBG) containing bank transactional data that is very similar to real-world data. Performing an exploratory data analysis on these datasets can help answer several compelling questions that could reveal useful insights for decision-makers, such as: understanding customer behaviour, average customer spending, identifying peak spending months, which shops have the highest spending and whether customers that spend more tend to transfer money to other people, who also spend more. Our main contribution is analysing two synthetic datasets, consisting of 13234863 million transactional records and 174601 transactional records, respectively. This dataset was combined with insights from the analysis, which included a dashboard, representing a detailed customer transaction for money payed into the customer's account and money paid out of the account, along with details of the counter-party to the transaction. We also applied unsupervised and supervised learning techniques, including data clustering and building a decision tree classification model, to classify customers into four different groups with similar characteristics, based on the customer's spending density and transfer frequency. We found that the customers with the highest spending frequency in terms of transactions had been spending their money in bars or pubs. Moreover, we found that customer spending increased at the start of the year (January) and at the end of the year (December). We concluded that banks should classify customers according to their profits and loyalty to the bank when recommending suitable financial products to them with corresponding service quality. To create this recommendation engine, we have built a classification model that divides all customers into four groups: the ordinary customer, loyal customer, potential valuable customer and priority customer.**

Keywords — Exploratory Data Analysis, Data Mining, Classification, Clustering, Banking sector, central banks, finance.

I. INTRODUCTION

A vast amount of transactional bank data is generated daily and, essentially, refers to everything related to customer activity; their behaviour and personality is shown in their transactions. Therefore, conducting a proper analysis on this data is required in order to gain useful insights to help make better-informed decisions.

Most financial products and services are standard, making it difficult for banks to meet the unique needs of their customers. By understanding client behavioural trends and preferences, banks would be able to formulate an affective marketing strategy. The important question is, how could Lloyds Banking Group (LBG) take advantage of these huge amounts of data to help them increase their profits, as well as customer loyalty? Detecting fraud is another critical aspect to consider in the banking sector and this can be done by clustering customers and monitoring outliers and suspicious behaviours.

Exploratory Data Analysis (EDA) aims at analysing data sets encompass a broader overview of their main characteristics. This is done by using statistical graphics and other data visualisation methods. Although the available datasets do not have a customer class label, unsupervised learning addresses this problem by using unlabelled data and adding a class label for each customer, to learn classification models.

In this research paper, we employ two well-known supervised and unsupervised learning approaches, k-means clustering and decision tree classification, to cluster customers into different groups, sharing similar characteristics, before classifying them.

The rest of the paper is organised as follows: related work is presented in Section 2. In Section 3, we describe our dataset and the preprocessing techniques used to render the data ready for analysis. The methods used to derive our results are described in Section 4. Our results are described in Section 5, future work and conclusions in Section 6.

II. RELATED WORK

Our related is derived from different areas: central banking studies, fraud detection, classification models, customer retention in the banking sector and data mining in the banking sector. Wei Wei, Jinjiu Li, Longbing Cao, Yuming Ou and Jiahang Chen [1] propose a novel algorithm called 'ContrastMiner' that effectively detects and distinguishes fraudulent activity from genuine customer behaviour. It does this by building a contrast vector for each customer transaction, based on the customer's historical behavioural sequence and outlines the differentiating rate of each current transaction against the customer's behavioural preference. The results of the 'ContrastMiner' algorithm on real online banking data show high fraud detection accuracy and it works better than existing fraud detection methods and systems in both efficiency and accuracy. In addition, the algorithm can be merged with the existing banking fraud detection system.

Bholat [2], explains why 'Big Data' is likely to become of paramount importance to central banks in changing both internal operations and external economic and financial systems. 'Big Data' tools and techniques have had less of an impact on financial services than they have had on other sectors of the economy. Nevertheless, the situation appears to be changing rapidly. Having standardised granular data is

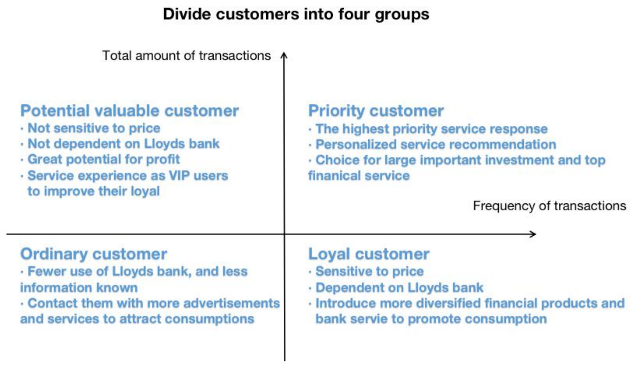

Figure 4.2: Description for each kind of customers

For different groups, we can adopt different sales strategies and recommend different financial services to them.

·Ordinary customer:

'Ordinary customers' do not transfer funds frequently and the total amount of transfers is not considered substantial. They may not have a high demand for use, or they may use multiple banks at the same time. Information for ‘ordinary customers’ is restricted and they may be potential loyal customers or valuable customers. This type of customer is the key target group for activity promotion or preferential activities and the bank can reach these customers through telephone sales or gifting, to try to improve their loyalty.

The bank may introduce these customers to select, basic financial products, such as like credit loans, mortgages or loan securitisation. However, the loan limit is restricted, because the actual economic capabilities of such users is unknown. Thus the bank may only offer them the opportunity to experience more diversified financial services.

·Loyal customer:

'Local customers’ often use Lloyds Bank to make transfers, but the total amount of transfers is considered to be very low, which explains they are loyal to the bank, even though they do not have enough financial strength to make large investments. Banks can use such users as information dissemination sources, prompting them to recommend bank activities and products to people around them, and give them some benefits or discounts as rewards for attracting new customers.

Due to the high loyalty of such users, the bank can open all loan services for them with a medium loan amount. The bank may also recommend a variety of low-risk financial products to them, such as monetary fund, national bonds or a time deposit.

·Potential valuable customer:

'Potential valuable customers' have made high-value transfers through the bank, indicating that they have high asset strength and are not price sensitive. However, the Lloyds bank may not be a dependable choice for them. They have huge potential value, since they are very likely to become priority customers for the bank, and banks need to give them the best experiential service to increase their loyalty to the bank.

Lloyds banks may use professional sales staff to recommend this customer type with the most suitable financial services and investment consulting. The bank may also guide in becoming ‘super customers’ for the bank, which may encourage them to invest more money into the bank. However, for loan services, the banks can only give them a medium loan amount, since their credit for the bank may still be uncertain.

·Priority customer:

Such customers should be the top priority for the banks and they typically possess a high financial capacity, as well as loyalty. These customers determine the development status of the bank. Their needs should be responded to first and they should be prioritised for key investment projects by bank.

The bank should establish personalised investment strategies for these customers, providing them with the most suitable asset allocation schemes and financial planning. Therefore, the bank should understand the needs of these customers, their consumption preferences, credit limit, risk tolerance, asset allocation preferences, risk preferences and other information in order to achieve accurate product recommendations.

B. Second Dataset

For the second dataset, we used ‘Tableau' software, which helped us render our visualisation for the data analysis more dynamic and interactive. Hence, the benefit from that we can take an insight across bank-level ‘all costumers’ or using the filter to select particular customer or shops. The dashboard will present five different graphs. The first graph depicts 'monthly deposits to customers accounts’, the second presents the ‘monthly transaction per customer’, the third graph shows the 'places customers spend their money the most’, the fourth graph displays ‘hourly deposit transactions to customer accounts’ and the fifth graph shows ‘hourly withdrawal transactions per customer’. See figure 4.3 .

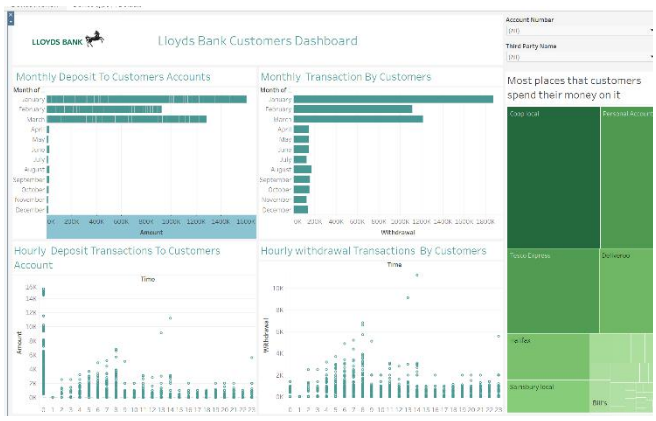

Figure 4.3: Lloyds bank customers dashboard

We have attempted to apply the same semi-supervised learning method to train the classification model. However, the second data set is not large enough its distribution is uneven. Therefore, we decide to use the first data set to train the model, before using the model to make predictions based on the customers of the second dataset.

V. Results

We started an exploratory data analysis for the bank level before expanding into the details of the customer level.