Chapter 8
Predicting Customer Transactions Using Machine Learning

Ansh Patel
https://orcid.org/0009-0001-3639-8722
VIT Bhopal University, India

Kalp Prajapati
https://orcid.org/0009-0004-2564-3886
VIT Bhopal University, India

D. Lakshmi
https://orcid.org/0000-0003-4018-1208
VIT Bhopal University, India

ABSTRACT

In the contemporary financial landscape, predicting customer transactions plays a crucial role in enhancing customer service, personalizing marketing strategies, and improving operational efficiency. This research paper delves into the prediction of customer transactions using machine learning. Various machine learning techniques have been employed in previous research to predict customer transactions. Utilizing the anonymized Customer Transaction Prediction dataset, this study undertakes a comprehensive data analysis, rigorous feature engineering, and model training. The primary aim is to predict the likelihood of a customer making a specific transaction in the future. The methodology encompasses various data visualization techniques, statistical analyses, and model evaluation metrics to ensure robust and accurate predictions. Our findings demonstrate the effectiveness of the LightGBM model in handling large-scale datasets with numerous features, achieving a competitive

DOI: 10.4018/979-8-3693-6386-7.ch008

AUC score.

1. INTRODUCTION

1.1 Background

In the current financial scenario, predicting customer transactions is a very critical activity for banks and other financial institutions. The accurate forecasting of customer behavior, especially in terms of transactions, is integral to several business strategies, such as personalized marketing, risk management, fraud detection, and customer retention. Big data and advanced machine learning algorithms have made it possible to mine large chunks of transactional data to discover patterns that may predict with great accuracy what a customer might do in the future. Bank accounts have the capability to gather humongous amounts of data on a daily basis; with the right technique in handling these data points, there could be greater possibilities for understanding customer behaviors deeply. This data would likely be high-dimensional in most cases, anonymized and highly imbalanced by conventional predictive modeling techniques; furthermore, feature anonymization presents difficulties in direct feature-related applications due to domain knowledge constraints that can be addressed via deep latent discovery without features requiring detailed descriptions.

Over the past years, different machine learning models have been used for solving this problem of prediction of customers' transactions. These vary from simple models of logistic regression to sophisticated methods, such as a decision tree, random forest(Joshi, Gupte, & Saravanan, 2018) and even neural networks. However, with the growing availability of high-performance computing resources and more advanced algorithms, gradient boosting machines (GBMs) have emerged as one of the most powerful tools for this task. Among these, Light Gradient Boosting Machine (LightGBM) has gained popularity due to its efficiency and superior performance, especially when dealing with large datasets and complex feature interactions.

1.2 Related Work

Predictive modeling in finance is one of the highly studied issues in literature. Machine learning models were used successfully for other tasks like forecasting the likelihood of a client being retained by retail services (Schaeffer & Sanchez, 2020), techniques comparing for customer retention (Sabbeh, 2018),predict consumer behavior in digital markets (Orogun & Onyekwelu, 2019). Several avenues of improvement in accuracy and reliability, especially for predicting customer transaction types, have been pursued. Among the oldest and easiest models for binary