Financial management alert system using Machine Learning

The banking industry's increasing reliance on debit card transactions has generated a wealth of valuable data for understanding consumer behavior. This study aims to develop a machine learning model to predict a customer's next expenditure and the corresponding merchant category using 50 customers' debit card transaction data for 11 years. Unlike existing research focused on bankrupt users and fraud detection, this study addresses the next expenditure prediction with merchant categories. For the bank, predicting a customer's next expenditure and merchant category enables targeted marketing efforts. The bank can send alert messages with discount offers specifically to each customer's spending habits, reducing marketing costs by only targeting relevant customers for relevant merchant types. Additionally, customers benefit from early reminders, allowing them to manage their finances effectively. For instance, a customer can receive a reminder about an upcoming insurance payment and allocate funds accordingly, avoiding unnecessary expenses. This proactive approach can help reduce the number of bankrupt customers and long-term customer relationships. Challenges in this study include obtaining a dataset that is not readily available on the internet. The dataset was provided by the People's Bank Head Office Digital Banking Department while ensuring data privacy. Data preprocessing involved removing null values, and unnecessary columns and creating customer ID instead of account numbers. Then identified 36 customers who consistently used debit cards and categorized merchant names into 11 groups. The dataset was split into training and testing sets using a specific date. Four machine learning algorithms Gradient Boosting Regressor, Random Forest Regressor, Random Forest Classifier, and Long Short-Term Memory were employed. The first method used a Gradient Boosting Regressor to predict expenditures and merchant categories after encoding the categories using one-hot encoding. The second method used two separate algorithms, Random Forest Regressor for expenditure prediction and Random Forest Classifier for merchant category prediction. Ordinal Encoding was used to convert categories into numerical values and as the third method, LSTM-based neural networks was used to predict expenditure and merchant category. Model performance was optimized through hyperparameter (learning rate, number of trees, maximum depth of each decision tree, minimum number of samples required to split an internal node, minimum number of samples required to be at a leaf node, and fixed random seed for reproducibility) tuning using grid search, evaluating various combinations of hyperparameters through cross-validation. Models run through each customer's unique dataset since expanding patterns are different from each other. The results showed that the LSTM method achieved higher accuracy compared to the other two methods. This was evident from R2 scores (0.6822, 0.9866, and 0.9947) for expenditure prediction and R2 values (0.6393, 0.8605, and 0.9895) for category prediction. In the future, this study could be extended to predict the exact time and date of transactions using techniques like Long Short-Term Memory (LSTM) with a larger dataset like 1000 customers.

Index Terms—Expenditure prediction, Gradient Boosting Regressor, Long Short-Term Memory, Random Forest Classifier, Random Forest Regressor