Customer Churn Prediction and Offer Recommendation in Banking

Akshata Gurayya Annigeri
ANA 540 – Machine Learning

Abstract:

This project focuses on understanding customer churn behavior and building a targeted recommendation system using enriched banking transaction data. Key goals include segmenting customers based on transactional patterns, identifying churn risk, and offering personalized financial products to improve retention. Unsupervised learning (KMeans) is applied for behavioral segmentation, while supervised models (Logistic Regression, Decision Tree, Random Forest) are used to predict churn. Additionally, the project evaluates existing bank-provided recommendations and finds them largely ineffective due to lack of personalization. By leveraging behavioral insights and churn risk, the proposed recommendation strategy outperforms the original system and offers more relevant, data-driven suggestions tailored to each customer segment.

Introduction:

Customer retention is critical for financial institutions, as acquiring new customers is significantly more expensive than retaining existing ones. This project aims to analyze customer behavior using transactional data, predict the likelihood of churn, and generate personalized product recommendations that align with customer needs and risk profiles.

The analysis begins with customer segmentation based on behavioral features such as transaction frequency, product usage, tenure, and payment behavior. Churn prediction models are then trained to identify at-risk customers with high accuracy.

A key part of the investigation involves assessing the effectiveness of the bank's original recommendation system. Findings reveal that the same financial products are often assigned across customer segments without considering churn risk, credit behavior, or recent activity resulting in low relevance and limited impact. In contrast, the improved system introduced in this project generates offers based on a combination of recent behavior, customer score, and predicted churn, leading to more actionable and personalized targeting.

By combining machine learning with behavioral segmentation, the solution aims to enhance customer engagement and retention while demonstrating measurable improvements over existing strategies.

Methodology:

Data Preparation:

- The dataset includes 20,000 banking transactions with features such as TransactionDate, Amount, ProductCategory, Channel, CreditCardFees, InsuranceFees, LatePaymentAmount, CustomerScore, MonthlyIncome, and RecommendedOffer.
- Additional features such as transaction_count, product_count, tenure_days, late_payment, days_since_last_txn, and txn_last_90 were engineered.

Segmentation:

- KMeans clustering was applied using features like tenure_days, late_payment, and logic_score.
- Two segments emerged: "Loyal" and "Risky" customers.

Churn Prediction:

- Churn labels were defined based on behavioral cutoffs such as inactivity and negative payment patterns.
- Logistic Regression, Decision Tree, and Random Forest classifiers were trained.
- Evaluation metrics included Accuracy, Precision, Recall, and AUC-ROC.

Recommendation System:

- The bank's original recommendations were compared with segment-specific offers.
- New recommendations were generated based on recent behavior, churn status, and customer profile.
- Most recent offer per customer was retained to ensure contextual relevance.

Results:

Model Performance and Churn Insights:

To identify customers at risk of churn, three classification models were implemented: Logistic Regression, Decision Tree, and Random Forest. Among them, Random Forest delivered the highest AUC-ROC score, indicating strong predictive performance and the ability to effectively distinguish between churned and retained customers.

Behavioral segmentation using KMeans revealed two key customer clusters: Loyal and Risky. As expected, risky customers exhibited lower tenure, reduced transaction volume, and higher churn probability. However, a surprising insight emerged from the loyal segment - over 65% of customers labeled as Loyal had actually churned.

This revealed a major blind spot: behavioral metrics like tenure and transaction count alone were not sufficient indicators of loyalty or retention. It highlighted the importance of not only identifying loyalty based on past behavior but also monitoring recent engagement signals such as declining activity or gaps in transactions.

Evaluation of Bank's Existing Recommendation System:

An in-depth analysis of the bank's existing recommendation strategy showed limited personalization:

- The same 2- 3 product offers were repeatedly assigned across all customer segments.
- The system ignored churn risk altogether, recommending complex investment services even to high-risk and disengaged customers.
- Loyal customers who churned were often assigned generic financial products, failing to address their disengagement signals.

Visualizations and cross-tabulations confirmed that the majority of churned customers, regardless of their segment, were offered irrelevant products, reducing the effectiveness of the recommendation strategy.

Redesigned Recommendation System: Behavior-Aware and Churn-Sensitive

To improve relevance and retention, a new recommendation framework was built using:

- Behavioral features like transaction frequency in the last 90 days, days since last activity, and customer segment (Loyal/Risky).
- Churn prediction results to avoid recommending high-risk offers to disengaged users.

This system prioritized simple, trust-building offers (e.g., Financial Literacy Programs, Cashback Offers) for churn-prone and risky users. For truly loyal and recently active customers, it recommended high-value offers such as Premium Investment Services or Gold Cards with Travel Benefits.

This behavior-aware strategy ensured that recommendations were both personalized and retention-oriented.

Visual Evidence and Business Alignment

Supporting visualizations made the contrast clear:

- Crosstab heatmaps showed how the original system assigned the same product to both churned and retained customers.
- Charts comparing the original vs. new recommendations demonstrated that the new system better matched customer profiles and aligned with risk levels.
- The new system reduced mismatches like offering investment products to customers showing signs of disengagement.

Conclusion:

This project demonstrated how combining machine learning with behavioral segmentation can significantly improve customer retention strategies in the banking domain. By analyzing transactional patterns, applying unsupervised clustering, and developing supervised models to predict churn, deeper insights into customer behavior were uncovered. Notably, 65% of customers identified as “Loyal” had churned, revealing gaps in existing loyalty definitions.

The evaluation of the bank’s current recommendation system showed poor personalization, with high-risk customers receiving complex offers and churn-prone segments being neglected. The redesigned recommendation framework addressed this issue by incorporating recent engagement metrics and churn probabilities. As a result, personalized offers were better aligned with customer needs i.e. low-risk, trust-building options for disengaged users, and high-value services for active, profitable segments.

These findings highlight the importance of integrating behavioral analytics, predictive modeling, and customer segmentation into recommendation systems to enhance both customer satisfaction and business performance.

References:

- Freitas, F. H. de. (2018). Predicting article retweets and likes based on the title using Machine Learning.
- Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow (2nd ed.). O'Reilly Media.
- Course lectures and machine learning notes.
- Fp20 dataset.

Appendix:

Please find my code here and insights here FinalProject_MachineLearning_AkshataAnnigeri.ipynb