Implementation of a Machine Learning Framework for Predicting Customer Lifetime Value (CLV) in a Bank

NABIGH Mohamed

(*Received 5 August 2024*)

Abstract

This report presents a detailed plan to implement a sophisticated model for predicting Customer Lifetime Value (CLV) in our bank. Inspired by the techniques discussed in the article "A novel approach to predicting customer lifetime value in B2B SaaS companies," this implementation aims to enhance our ability to identify and nurture high-value customers, optimize marketing strategies, and improve customer retention.

1 Executive Summary

In this report, I present a detailed plan to implement a sophisticated model for predicting Customer Lifetime Value (CLV) in our bank. This model is inspired by the advanced techniques discussed in the article "A novel approach to predicting customer lifetime value in B2B SaaS companies." By adapting these techniques to our context, we aim to enhance our ability to identify and nurture high-value customers, optimize marketing strategies, and improve customer retention.

2 Understanding the Article's Approach

The article proposes a flexible machine learning framework specifically designed to predict CLV in Business-to-Business (B2B) Software-as-a-Service (SaaS) companies. Key elements of this approach include:

1. **Hierarchical Model Structure**:
   - **Short-Term Prediction (T')**: Uses recent data to predict CLV over a shorter period.
   - **Long-Term Prediction (T)**: Extends short-term predictions to cover the full customer lifetime.
2. **Ensemble Learning**:
   - Combines multiple machine learning models to improve prediction accuracy.

  - Uses different models for different customer segments based on specific characteristics.

3. **Feature Engineering**:
   - Creates meaningful features from raw data, such as aggregate metrics, trend indicators, lag features, and interaction terms.
   - Emphasizes recent data to ensure predictions are based on the most relevant information.
4. **Model Evaluation**:
   - Uses metrics like RMSE (Root Mean Squared Error), MAE (Mean Absolute Error), and SMAPE (Symmetric Mean Absolute Percentage Error) to evaluate model performance.
   - Performs residual analysis to refine segmentation and model selection.

3 Implementing the Approach in Our Case Study

3.1 Project Goals

Our goal is to predict the cumulative CLV of clients at our bank. By implementing the techniques from the article, we aim to:

- Identify high-value customers early.
- Optimize marketing campaigns to target these high-value customers.
- Improve customer retention by understanding their long-term value.

3.2 Project Steps

3.2.1 Data Collection

- Collect data from various sources within the bank, including transaction history, customer demographics, product holdings, and engagement metrics.

3.2.2 Data Preprocessing

- Clean and transform the data, handling missing values and normalizing numerical features.
- Encode categorical variables into numerical values.

3.2.3 Feature Engineering

- **Aggregate Metrics**: Calculate total transactions, average transaction value, and total balance over time.
- **Trend Indicators**: Derive growth rates and percentage changes in balances and transaction volumes.
- **Lag Features**: Create features representing past values, such as the previous month's balance or last quarter's transactions.
- **Interaction Terms**: Generate new features that capture the combined effect of multiple variables, such as balance and transaction count.

3.2.4 Model Training

- Split data into training and testing sets.
- Use machine learning models like Random Forest and XGBoost to train on the engineered features.
- Perform hyperparameter tuning to optimize model performance.

3.2.5 Model Evaluation

- Evaluate models using RMSE, MAE, and SMAPE.
- Analyze residuals to identify patterns and improve segmentation.

3.2.6 Model Deployment

- Integrate the final model into the bank's IT infrastructure.
- Automate data collection, preprocessing, feature engineering, and model retraining.

3.2.7 Business Application

- **Marketing Optimization**: Use CLV predictions to allocate marketing spend effectively and target high-value customers with personalized offers.
- **Customer Retention**: Identify at-risk customers and develop targeted retention strategies.
- **Product Development**: Tailor product offerings based on predicted customer value and behavior.
- **ROI Analysis**: Conduct ROI analysis to guide strategic resource allocation and optimize campaign performance.

4 Conclusion

By implementing the machine learning techniques from the article, we can significantly enhance our ability to predict and leverage Customer Lifetime Value. This will enable us to make more informed decisions, optimize marketing efforts, and improve overall customer satisfaction and retention. The project will involve careful data collection, robust feature engineering, and advanced modeling techniques to achieve these goals.