Customer Lifetime Value (CLV) Prediction Model

Internship Project – Elevate Labs Ministry of MSME

Table of Contents

- Introduction
- Abstract
- Tools
- Steps
- Conclusion
- Insights

1.Introduction

In today’s competitive business landscape, understanding customer value is crucial for long-term growth. This project focuses on predicting Customer Lifetime Value (CLV) using historical purchase behaviour. By identifying high-value customers, businesses can optimize marketing spend, improve retention strategies, and drive profitability

2.Abstract

The objective of this project was to build a machine learning model that predicts the lifetime value of customers based on their purchase history, tenure, and total spending. Using a Random Forest Regressor, the model achieved strong performance with a Mean Absolute Error (MAE) of ₹354.46 and Root Mean Squared Error (RMSE) of ₹590.36. Customers were segmented into Low, Medium, and High value tiers based on predicted CLV, enabling targeted business strategies.

3.Tools

- Python: Data cleaning, feature engineering, model training
- Pandas, Seaborn, Matplotlib: Data manipulation and visualization
- Scikit-learn: Machine learning model and evaluation
- Jupyter Notebook: Development environment

4.Methodology

Step 1: Data Cleaning

- Removed duplicates and null values
- Verified data types and summary statistics

Step 2: Feature Engineering

- Selected key features: purchase_history, tenure, total_spent
- Defined CLV as the target variable

Step 3: Model Training

- Used RandomForestRegressor with 100 estimators
- Split data into training and test sets (80/20)

Step 4: Evaluation

- MAE: ₹354.46
- RMSE: ₹590.36
- Model showed stable and consistent predictions

Step 5: Prediction & Segmentation

- Predicted CLV for all customers
- Segmented into Low, Medium, High using quantiles
- Exported final CSV with customer ID, predicted CLV, and segment

Step 6: Visualization

- Plotted CLV distribution using Seaborn
- Observed right-skewed pattern, indicating a few high-value customers drive most revenue

5.Conclusion

This project successfully demonstrates how machine learning can be used to predict customer lifetime value and segment customers for strategic decision-making. The model's performance was strong, and the segmentation revealed actionable insights. Future improvements could include adding behavioral or demographic features, testing XGBoost for comparison, and building a dashboard for business users.

6. Insights

- **Majority of customers fall into the Low CLV segment**, indicating that most users contribute modest revenue. This aligns with the right-skewed distribution observed in the CLV histogram.
- **High CLV customers are fewer but significantly more valuable**, suggesting that targeted retention strategies for this group could yield strong ROI.
- **Tenure and purchase history show strong correlation with CLV**, confirming that long-term engagement and frequent purchases are key drivers of customer value.
- **Model predictions are stable and consistent**, with low error margins (MAE: ₹354.46, RMSE: ₹590.36), making the segmentation reliable for business decisions.
- **Quantile-based segmentation using pd.qcut provides balanced customer tiers**, enabling marketing teams to personalize campaigns for each segment.
- **Visualization confirms business intuition**: a small group of loyal customers drive most of the revenue, while the majority are casual or low-engagement users.
- **Opportunities for future enhancement** include adding behavioral features (e.g., time between purchases), testing alternative models like XGBoost, and building dashboards for real-time CLV monitoring.

Thank You,

I would like to express my sincere gratitude to **Elevate Labs** and the **Ministry of MSME** for the opportunity to work on this project. This experience has helped me strengthen my skills in data analysis, machine learning, and business storytelling.

Special thanks to my mentors and reviewers for their guidance and support throughout the journey. Their feedback and encouragement played a key role in shaping this project.

This project marks a meaningful step in my transition toward a career in data analytics, and I look forward to applying these skills to real-world business challenges.

— Umasankar G
Aspiring Data Analyst

giriumasankar855@gmail.com