

Customer Lifetime Value Prediction Model

By- Aashish Kumar Saini

Laxmi Devi Institute of Engineering and Technology,

Alwar, Rajasthan

Abstract

This project aims to predict Customer Lifetime Value (LTV) using transaction data and behavioral features to support targeted marketing strategies. By engineering RFM (Recency, Frequency, Monetary) and Average Order Value (AOV) features, an XGBoost regression model was developed to forecast LTV. The model output was used to segment customers into actionable groups. The result empowers businesses to focus on high-value customers and personalize retention efforts.

Introduction

Customer Lifetime Value is a key business metric that estimates the total revenue a customer will generate throughout their relationship with a company. Accurate LTV prediction enables companies to prioritize marketing budgets, design loyalty programs, and reduce churn. This project uses historical transaction data to predict LTV using a machine learning approach and classify customers into strategic segments.

Tools Used

- **Python:** Core programming language
- **Pandas, NumPy:** Data cleaning and feature engineering
- **XGBoost:** For training the regression model
- **Scikit-learn:** Model evaluation and preprocessing
- **Matplotlib, Seaborn:** Data visualization
- **Excel:** Initial data review

Steps Involved in Building the Project

1 Data Preprocessing

- Merged transaction and customer data using Customer ID.
- Defined reference date as one day after the last transaction.

2 Feature Engineering

- **Recency:** Days since the last purchase.
- **Frequency:** Unique invoice count per customer.
- **Monetary:** Total revenue from a customer.
- **AOV:** Monetary / Frequency.
- Used Monetary as proxy LTV for training the model.

3 Model Building

- Features used: Recency, Frequency, AOV.
- Trained an XGBoost Regressor to predict LTV.
- Evaluate Using:

- 1 MAE: 377.79
- 2 RMSE: 4112.67

- Extracted feature importance: Frequency and AOV contributed the most.

4 Customer Segmentation

- Predicted LTV using the trained model.
- Segmented customers into Low, Medium, High, and VIP using quartiles.
- Calculated average Recency, Frequency, and AOV per segment.

Conclusion

The project demonstrates that machine learning can effectively forecast Customer Lifetime Value using simple behavioral metrics. The segmentation provides actionable insights—VIP customers can be rewarded to increase loyalty, while Low-value customers can be targeted with reactivation offers. The project deliverables include the Python notebook, trained XGBoost model, visualizations, and final CSV with predicted LTVs and segments.

Future Scope

- Integrate **churn prediction** for dynamic LTV estimation.
- Deploy the model as a **web-based dashboard** for real-time business use.
- Expand features to include **channel of purchase, product type, or seasonality**.