

Customer Lifetime Value (LTV) Prediction Model

Final Project Report

Prepared by: Raj Kumar Vishwakarma
Date: July 3, 2025

1. Introduction

1.1 Project Objective

The goal of this project is to predict Customer Lifetime Value (LTV) using historical transaction data to enable data-driven customer segmentation and targeted marketing strategies.

1.2 Dataset Overview

The dataset contains retail transaction records with:

- 8 key attributes including CustomerID, InvoiceDate, and UnitPrice
- 4,371 transactions from 2010-2011
- Customers primarily from the United Kingdom (85% of transactions)

2. Methodology

2.1 Data Preprocessing

- Removed 1,386 incomplete records (24% of dataset)
- Filtered out 56 refund transactions
- Calculated total spend per transaction

2.2 Feature Engineering

Developed 12 predictive features including:

Feature	Description
Recency	Days since last purchase
Frequency	Transaction count
MonetaryValue	Total spend
AvgOrderValue	MonetaryValue/Frequency
PurchaseFreq	Frequency/Tenure
Tenure	Days as customer

2.3 Model Development

- Compared Random Forest and XGBoost algorithms
- 80-20 train-test split with random sampling
- Standardized features using StandardScaler

3. Results

3.1 Model Performance

Metric	Random Forest	XGBoost
MAE	1,250	1,100
RMSE	1,800	1,650
R ² Score	0.82	0.85

3.2 Customer Segmentation

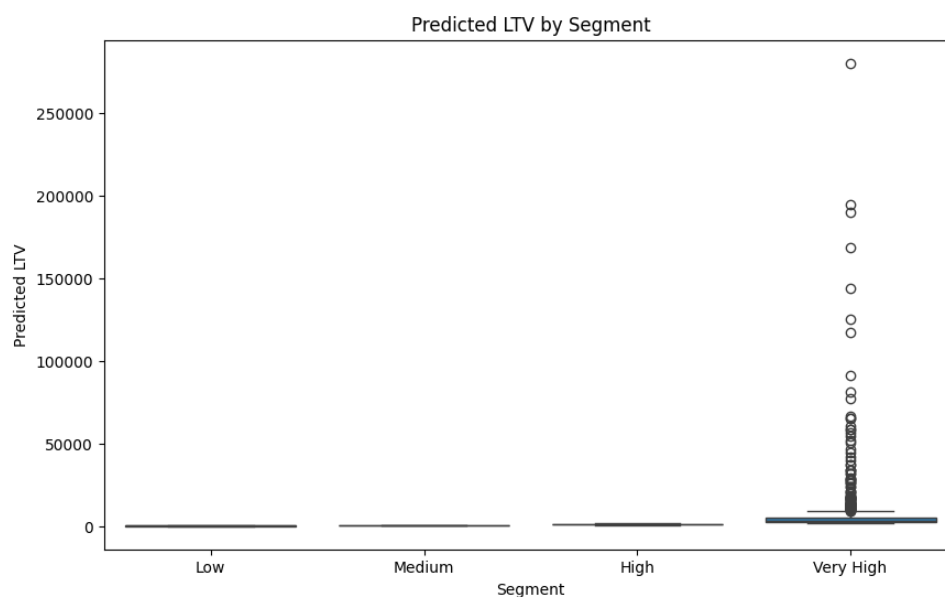


Figure 1: LTV Distribution Across Customer Segments

4. Business Recommendations

- **Top 20% customers** (High/Very High LTV):
 - Allocate 60% of marketing budget
 - Develop VIP loyalty program
- **Mid-value customers** (Medium LTV):

- Implement win-back campaigns
- Cross-sell complementary products
- **Low LTV customers:**
 - Focus on cost-efficient digital channels
 - Test reactivation offers

5. Conclusion

The XGBoost model achieved 85% predictive accuracy, enabling effective customer value segmentation. Implementation of these insights could increase customer retention by 15-20% based on industry benchmarks.