

# **Project Report**

## **Customer Lifetime Value Prediction Model using Machine Learning**

### **1. Project Objective**

The primary objective is to model Customer Lifetime Value (CLV) using transactional data from an online retail business. This enables customer segmentation, targeted marketing, and better resource allocation by identifying high-value customers.

### **2. Dataset Description**

- **Source:** Kaggle - Online Retail Dataset
- **Attributes:** InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice, CustomerID, Country
- **Time Period:** December 2010 to December 2011
- **Size:** ~500k rows

### **3. Methodologies Applied**

#### **Data Preprocessing**

- Removed nulls and canceled transactions.
- Filtered only positive quantities and prices.
- Converted InvoiceDate to datetime.
- Created training and testing splits based on a cut-off date (2011-08-31).

#### **Feature Engineering**

- **latetime:** Days since last transaction
- **earlytime:** Days since first transaction
- **freq:** Total transactions
- **freq\_3m:** Transactions in recent 3 months
- **target:** Future transactions in test period

## 4. Modeling Approaches

### Probabilistic Model: BG/NBD + Gamma-Gamma

- Fitted the model to historical transaction data to predict CLV.
- Assumptions:
  - Purchase follows a Poisson-Gamma process.
  - Dropout follows a Geometric-Beta process.

#### Model Performance:

- **RMSE:** 1.6298
- **R<sup>2</sup>:** 0.5610

### Machine Learning Model: XGBoost Regressor

- Used engineered features to train a predictive model.



#### Model Performance:

- **RMSE:** 0.9432
- **R<sup>2</sup>:** 0.8530

**Conclusion:** XGBoost significantly outperforms the probabilistic model in accuracy but lacks interpretability and simplicity of assumptions.

## 5. Visualizations

- **Poisson / Gamma / Beta / Geometric Distributions** used in probabilistic modeling.
- **RFM Segmentation** heatmap for strategic targeting.

## 6. Recommendations

Segment	Strategy
<b>Champions</b>	Reward, promote referrals, early access to launches
<b>Loyal Customers</b>	Encourage reviews, loyalty programs
<b>At Risk</b>	Win-back campaigns, personalized offers
<b>Hibernating</b>	Reactivation emails, feedback surveys
<b>Potential Loyalists</b>	Promote upsell/cross-sell, early incentives
<b>Can't Lose Them</b>	Urgent retention via discounts, premium engagement

## **Summary**

- **XGBoost** > **Probabilistic models** in predictive power.
- However, **BG/NBD** + **Gamma-Gamma** provides strong interpretability.
- Combining both helps balance accuracy and business insight.

~ **AKASH KUMAR RAJAK**