

Customer Lifetime Value (CLTV) Prediction Project

Thirupathi Yuvaraju

1. Objective

The objective of this project is to predict Customer Lifetime Value (CLTV) using transaction data. By analyzing customer behavior — specifically recency, frequency, and average order value — we aim to predict the total monetary value a customer is expected to bring over their relationship with the business. This enables better segmentation and targeting in marketing strategies.

2. Tools and Libraries Used

- Python
- Pandas, NumPy
- Scikit-learn
- XGBoost
- Matplotlib, Seaborn
- Jupyter Notebook

3. Project Steps

Step 1: Data Collection and Loading

We started by loading a customer transaction dataset, which includes details such as InvoiceDate, Quantity, UnitPrice, and CustomerID. The data was filtered to include only transactions from the United Kingdom.

Step 2: Data Preprocessing

The data was cleaned by removing null CustomerIDs and converting InvoiceDate to datetime format. A new column for total transaction amount was computed using $\text{Quantity} \times \text{UnitPrice}$.

Step 3: Feature Engineering

Three main features were engineered:

- Recency: Days since the last purchase by the customer.
- Frequency: Number of unique invoices per customer.
- Monetary Value and AOV (Average Order Value): Total spending and average value per order.

Step 4: Model Building

We used XGBoost Regressor to train a regression model using the Recency, Frequency, and AOV features. The target variable was the total monetary value (used as a proxy for LTV). The dataset was split into training and testing sets for validation.

Step 5: Model Evaluation

The model was evaluated using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). The results were:

- MAE: 129.36
- RMSE: 902.58

Step 6: Prediction Visualization

A scatter plot was created to visualize the predicted vs actual CLTV values, demonstrating how well the model was able to estimate customer value.

4. Actual vs Predicted Plot

