Customer Lifetime Value Prediction Using Machine Learning

1. Introduction

In today's competitive market, understanding customer behavior and maximizing customer retention are critical to business success. One of the most valuable metrics that businesses can leverage is **Customer Lifetime Value (CLV)** — the projected revenue that a customer is expected to generate over their entire relationship with a company.

This project aims to build a machine learning model that predicts the CLV of customers based on historical transactional data. The model's output is visualized through an interactive Streamlit dashboard, which provides clear insights into customer behavior and value segmentation. This helps businesses target high-value customers more effectively and optimize their marketing strategies.

2. Abstract

The objective of this project is to predict the Customer Lifetime Value (CLV) using key behavioral indicators such as Recency, Frequency, and Average Order Value (AOV). By applying regression models like **XGBoost**, we estimate future customer value based on historical purchase data.

The project uses real-world transaction data, performs comprehensive feature engineering, trains a regression model, and visualizes the results using a fully interactive dashboard. The visual interface enables users to explore customer segments, understand patterns, and gain insights from multiple visualizations including histograms, time series trends, segment breakdowns, and correlation heatmaps.

3. Tools and Technologies Used

Tool/Library	Purpose
Python	Core programming language
Pandas	Data cleaning and manipulation
Seaborn & Matplotlib	Data visualization
Scikit-learn	Preprocessing and evaluation
XGBoost	Regression model for CLV prediction
Streamlit	Interactive dashboard development
Jupyter Notebook	Development and experimentation
Joblib	Model serialization (.pkl file)

4. Steps Involved in Building the Project

Step 1: Data Collection

- Collected customer transaction data including CustomerID, InvoiceDate, Quantity, and UnitPrice.

Step 2: Data Preprocessing

- Converted InvoiceDate to datetime.

- Calculated TotalAmount = Quantity × UnitPrice.
- Removed missing or invalid entries.

Step 3: Feature Engineering

- **Recency**: Days since last purchase
- **Frequency**: Number of purchases per customer
- **AOV (Average Order Value)**: TotalAmount / Number of transactions
- **Last Purchase Date**: Used for time series visualization

Step 4: CLV Calculation

- Aggregated total spending per customer as a proxy for lifetime value.
- Created segments: Low, Medium, High, Very High based on quartiles.

Step 5: Model Building

- Used XGBoost Regressor to predict CLV using Recency, Frequency, AOV.
- Split data into train and test sets.
- Evaluated using MAE and RMSE.

Step 6: Model Serialization

- Saved trained model using joblib as model.pkl.

Step 7: Dashboard Creation (Streamlit)

- Visualized:
    - Summary KPIs (Total customers, Avg CLV)
    - Histograms (Recency, Frequency, AOV, CLV)
    - Segment bar chart and boxplot
    - Time series chart (CLV by month and segment)
    - Correlation heatmap
    - Scatter plot (AOV vs Frequency)
    - Filtered data table

5. Conclusion

This project demonstrates how businesses can use machine learning to estimate Customer Lifetime Value based on transactional behavior. The predictive model allows for data-driven customer segmentation, enabling smarter marketing investment and retention strategies.

The dashboard provides a dynamic, visual interface to explore customer patterns, segment insights, and CLV trends — ultimately helping decision-makers identify high-value customers and align business strategies accordingly.

By combining data analytics, regression modeling, and interactive dashboards, this project serves as a practical example of how AI can enhance customer-centric decision-making in real-world business scenarios.