Customer Lifetime Value (LTV) Prediction Using RFM and XGBoost

Introduction:

Customer Lifetime Value (LTV) is a crucial metric for understanding how much revenue a business can expect from a customer throughout their relationship. Accurate LTV prediction allows businesses to prioritize customer retention, segment users, and personalize marketing strategies. This project aims to build a machine learning model that predicts the LTV of customers using historical transactional data.

Abstract:

In this project, we used a real-world online-retail dataset to identify high, medium, and low-value customers based on their predicted LTV. The approach involves extracting RFM features (Recency, Frequency, Monetary), deriving Average Order Value (AOV), and training an XGBoost regression model to predict customer lifetime value. The final model was evaluated using standard metrics such as MAE and RMSE and deployed into a user-friendly Power BI dashboard for visualization.

Tools Used:

- Programming Language: Python (Jupyter Notebook)

- Libraries: Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, XGBoost

- Model Saving: Pickle

- Dashboarding: Power BI Desktop

- Data Source: UCI Online Retail Dataset (or alternative CSV file)

Steps Involved in Building the Project:

Step 1: Data Preprocessing

- Loaded transactional data from CSV

- Cleaned missing values (CustomerID, InvoiceDate)

- Calculated TotalAmount = Quantity × UnitPrice

- Filtered out negative and zero-value transactions

Step 2: Feature Engineering (RFM)

- Recency: Days since last purchase

- Frequency: Number of unique invoices

- Monetary: Total amount spent

- AOV: Average Order Value = Monetary / Frequency

Step 3: Model Building

- Split data into training and testing sets (80/20)

- Trained an XGBoost regressor to predict Monetary (LTV proxy)

- Evaluated performance using MAE and RMSE

Step 4: Model Deployment

- Saved model using pickle

- Predicted LTV for all customers and exported to CSV

- Loaded results into Power BI for visualization

Step 5: Power BI Dashboard

- Bar Chart: Top customers by LTV

- Histogram: LTV distribution by bins

- KPI Cards: Average Recency, Frequency, AOV, LTV

- Slicer: Filter visuals by customer segment (High/Medium/Low)

Conclusion:

This project successfully demonstrates how to predict customer LTV using transactional data and machine learning. The RFM-based features provided a meaningful representation of customer behavior, and the XGBoost model showed good predictive power. The Power BI dashboard enables interactive business insights and supports strategic decisions in customer targeting and retention.