Customer Lifetime Value (LTV) Prediction Using Machine Learning

Introduction

Customer Lifetime Value (LTV) represents the total revenue a business can expect from a customer over the duration of their relationship. Predicting LTV helps organizations identify valuable customers and design targeted marketing and retention strategies.

Abstract

This project focuses on predicting Customer Lifetime Value using historical transaction data. Customer behavior features such as recency, frequency, monetary value, and customer age are engineered and used to train a machine learning regression model. The predicted LTV values are further used to segment customers into value-based groups.

Dataset Description

The dataset consists of online retail transactions containing invoice details, purchase quantity, unit price, transaction date, and customer identifiers. Invalid records such as missing customer IDs and cancelled transactions were removed during preprocessing.

Tools and Technologies Used

Python, Pandas, NumPy, Scikit-learn, XGBoost, Google Colab

Methodology

1. Data cleaning and preprocessing
2. Feature engineering using Recency, Frequency, Monetary value, and Customer Age
3. LTV calculation
4. Model training using XGBoost regression
5. Model evaluation using MAE, RMSE, and R2 score
6. Customer segmentation based on predicted LTV

Conclusion

The developed model effectively predicts customer lifetime value and segments customers into low, medium, and high value groups. This enables businesses to focus retention efforts on high-value customers and improve overall profitability.