PROJECT REPORT

Customer Lifetime Value (LTV) Prediction Model

Abstract

Customer Lifetime Value (LTV) is a critical business metric that represents the total revenue a company can expect from a customer throughout their relationship. Accurately predicting LTV helps organizations optimize marketing strategies, improve customer retention, and allocate resources effectively.

This project focuses on building a machine learning-based regression model to predict customer LTV using historical purchase data. By applying feature engineering techniques such as Recency, Frequency, and Average Order Value (AOV), and training an XGBoost regression model, customers are segmented into value-based categories to support data-driven decision-making.

Introduction

In today's competitive business environment, understanding customer behavior is essential for long-term success. Traditional marketing approaches treat all customers equally, which often leads to inefficient spending and reduced profitability.

Customer Lifetime Value (LTV) provides a data-driven way to identify high-value customers and tailor marketing efforts accordingly. This project aims to predict LTV using past transaction data and machine learning techniques, enabling businesses to focus on customer retention, loyalty programs, and targeted promotions.

Tools Used

The following tools and technologies were used to build this project:

- **Platform:** Google Colab
- **Programming Language:** Python
- **Libraries:**
  - pandas (data manipulation)
  - numpy (numerical computations)
  - scikit-learn (model evaluation & preprocessing)

    - xgboost (regression model)
    - matplotlib & seaborn (data visualization)
- **Dataset Source:** Kaggle (Online Retail II – UCI dataset)

Steps Involved in Building the Project

1. Data Collection and Preprocessing

Transaction-level customer purchase data was loaded into Python. Data cleaning steps included handling missing customer IDs, removing invalid or negative transaction values, and converting date fields into datetime format.

2. Feature Engineering

Customer-level features were created using purchase history:

- **Recency:** Number of days since the last purchase
- **Frequency:** Total number of purchases
- **Average Order Value (AOV):** Average spend per transaction
- **Total Spend:** Cumulative customer expenditure

These features capture customer behavior effectively and serve as inputs for the prediction model.

3. Target Variable Definition

Historical total spend per customer was used as a proxy for Customer Lifetime Value (LTV).

4. Model Training

The dataset was split into training and testing sets. An **XGBoost Regression model** was trained to learn the relationship between customer behavior features and LTV.

5. Model Evaluation

Model performance was evaluated using:

- Mean Absolute Error (MAE)

- Root Mean Squared Error (RMSE)

These metrics ensured prediction accuracy and reliability.

6. LTV Prediction and Segmentation

The trained model was used to predict LTV for all customers. Based on predicted values, customers were segmented into:

- Low Value
- Medium Value
- High Value
- VIP Customers

This segmentation enables targeted marketing strategies.

Conclusion

This project successfully demonstrates how machine learning can be applied to predict Customer Lifetime Value using transactional data. By leveraging RFM-based features and an XGBoost regression model, customers were accurately classified into value-based segments.
The insights generated from this model can help businesses optimize marketing campaigns, improve customer retention, and maximize revenue. The project can be further enhanced by predicting future LTV over a fixed time horizon and integrating results into business intelligence dashboards.