

Customer Lifetime Value (LTV)

Prediction using Machine Learning

Introduction

Customer Lifetime Value (LTV) is a metric that estimates the total revenue a business can expect from a single customer over the entire relationship. Accurate LTV prediction allows businesses to make data-driven decisions in marketing, budgeting, and customer retention strategies. In this project, we built a machine learning model to predict LTV based on customer shopping behavior.

Abstract

The primary goal of this project is to predict the future value a customer will bring to the business using historical transactional data. We used a dataset containing customer shopping activity including invoice details, purchase dates, and amounts. By calculating features such as recency, frequency, and average order value (AOV), we trained a Random Forest Regressor to estimate each customer's LTV. The model demonstrated strong predictive performance and can help prioritize high-value customers.

Tools Used

- Python 3
- Pandas, NumPy
- Matplotlib, Seaborn
- Scikit-learn
- Joblib
- VS Code
- XGBoost

Steps Involved in Building the Project

1. Data Loading: Loaded customer shopping data from a CSV file.
2. Preprocessing:
 - Converted date fields.
 - Filtered recent data (post-2021).
 - Removed duplicates.
3. Feature Engineering:

- Recency: Days since last purchase.
- Frequency: Number of unique invoices.
- AOV: Average Order Value = Total spent / Number of invoices.
- Calculated LTV = AOV × Frequency.

4. Model Building:

- Split the data into training and testing sets.
- Trained a Random Forest Regressor.

5. Evaluation:

- Evaluated model performance using MAE, RMSE, and R^2 metrics.
- Visualized predictions and feature importance.

6. Saving Model:

- Exported the model using joblib.
- Saved customer LTV predictions and features to CSV.

7. Segmentation:

- Classified customers into High, Medium, and Low LTV segments.
- Plotted LTV distribution for analysis.

Conclusion

The model accurately predicted Customer Lifetime Value with an R^2 score close to 1.0, indicating strong predictive power. This system can help businesses identify and target high-value customers, optimize marketing spend, and enhance customer retention. With further refinement, the model could be extended to real-time or multi-channel data for broader impact.