Customer Lifetime Value Prediction Model

Abstract

Customer Lifetime Value (CLTV) is a key metric that estimates the total revenue a business can expect from a customer throughout their relationship. Accurate prediction of CLTV helps businesses make informed decisions related to marketing strategies, resource allocation, and customer retention. In this project, a data-driven approach is used to predict CLTV using past purchase behavior. Features such as Frequency, Recency, and Average Order Value are extracted from transaction data. Machine learning models like XGBoost and Random Forest are trained to predict LTV. The results are evaluated using MAE and RMSE, and customers are segmented based on their predicted LTV for targeted marketing.

Introduction

In a competitive marketplace, understanding customer value is crucial for businesses to prioritize efforts and optimize profits. CLTV prediction provides insights into the future worth of customers, enabling better strategic planning. By analyzing historical transactions, we can identify patterns that indicate customer loyalty and purchasing behavior. This project focuses on building a machine learning model to estimate the lifetime value of customers, assisting marketing teams in identifying high-value segments and personalizing their approach accordingly.

Tools Used

- **Programming Language:** Python
- **Libraries:**
    - Pandas, Numpy (Data manipulation)
    - Seaborn, Matplotlib (Data visualization)
    - Scikit-learn (Model evaluation)
    - XGBoost, Random Forest (Model training)
- **Software:** Jupyter Notebook, Excel
- **File Format:** CSV (input and output)

Steps Involved in Building the Project

1. Data Collection & Loading

- Import customer transaction history and customer details.
- Combine transaction and customer data using customer ID as a key.

2. Data Preprocessing

- Handle missing values and remove outliers in purchase data.
- Convert date fields into datetime format for temporal calculations.

- Filter transactions to include only relevant and valid entries.

3. Feature Engineering

- **Recency**: Number of days since the customer's last purchase.
- **Frequency**: Total number of purchases made by the customer.
- **Monetary Value**: Total revenue generated by the customer.
- **AOV (Average Order Value)**: Monetary / Frequency
- Create customer-level aggregates to summarize purchase behavior.

4. Model Preparation

- Split the dataset into training and testing sets.
- Normalize features if required.
- Select features such as Recency, Frequency, and AOV as inputs.

5. Model Training

- Apply **Random Forest Regressor** and **XGBoost Regressor** to the training data.
- Perform hyperparameter tuning (optional) for better performance.

6. Model Evaluation

- Evaluate models using:
  - **MAE (Mean Absolute Error)**
  - **RMSE (Root Mean Squared Error)**
- Compare models based on accuracy and generalization capability.

7. Customer Segmentation

- Use predicted LTV to divide customers into segments:
  - High Value
  - Medium Value
  - Low Value
- Create visualizations like bar plots or pie charts for each segment.

8. Output Generation

- Export the final predicted LTV values to a CSV file.
- Include segment labels for each customer.
- Save trained model for future deployment or testing.

Conclusion

The Customer Lifetime Value Prediction Model provides an effective way to estimate the future revenue potential of individual customers using historical purchase data. By applying machine learning techniques such as XGBoost and Random Forest, we achieved accurate and interpretable predictions. Segmenting customers based on LTV enables targeted marketing and resource optimization. This model can be integrated into CRM systems for real-time decision-making and long-term business planning.