Customer Lifetime Value (CLV) Prediction Model

Introduction:

The Customer Lifetime Value Prediction project focuses on building machine learning models to estimate the lifetime value of customers based on their purchasing behavior. Accurate CLV predictions enable businesses to allocate marketing resources more effectively, enhance customer retention, and drive profitability.

Abstract:

Using the Online Retail II dataset, key RFM (Recency, Frequency, Monetary) metrics were derived. The models built include Random Forest Regressor and XGBoost Regressor. Both models were evaluated based on MAE, RMSE, and R^2 scores to determine prediction performance. Feature importance analysis revealed that Frequency and AOV are significant predictors of CLV.

Tools Used:

- Python (Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, XGBoost)
- Jupyter Notebook

Steps Involved in Building the Project:

1. Data Cleaning: Removed missing Customer IDs, canceled invoices, negative quantities/prices.
2. Feature Engineering: Calculated Recency, Frequency, Monetary value, and Average Order Value (AOV) for each customer.
3. Data Splitting: Divided into training and testing sets.
4. Feature Scaling: Applied StandardScaler to normalize feature values.
5. Model Building: Trained Random Forest and XGBoost Regressors.
6. Evaluation: Achieved R^2 score of 0.87 using XGBoost, indicating good predictive power.
7. Visualization: Displayed feature importance using Random Forest.
8. Output: Saved predicted CLV values in 'Predicted_CLV.csv'.

Conclusion:

The CLV prediction models demonstrated robust performance with significant insights derived from RFM features. The analysis will aid in customer segmentation and strategic marketing initiatives.

Customer Lifetime Value (CLV) Prediction Model

Author:

Created by Susan Chandra Bonkuri