Customer Lifetime Value Prediction Model

Introduction:

Customer Lifetime Value (LTV) is a useful metric for companies to forecast the total revenue a customer brings into a company over the time they are involved with a brand. Companies can segment their customers, maximize marketing efforts, and distribute resources optimally across channels through effective knowledge of LTV. For this project we will develop a customer LTV model based on their buying habits to enable brands to reach high-value customers and enhance their retention strategy.

Abstract:

The goal of this project is to forecast Customer Lifetime Value (LTV) using customer buying history and applying machine learning methods. By applying XGBoost and Random Forest algorithms, we preprocess and engineer features like frequency, recency, and average order value (AOV) to train the algorithm. We use MAE and RMSE metrics for model validation, and customer segments are formed on the basis of forecasted LTV values. The outputs are a trained model, visualizations, and a CSV file with the ultimate LTV predictions.

Tools Used:

- Python (Sklearn, XGBoost): For data processing, machine learning model building, and evaluation.

- Pandas: For data manipulation and feature engineering.

- Matplotlib: For visualizing the results, including LTV distribution.

- Joblib: For saving and loading the trained model.

- Excel: For analyzing data and making quick

Steps in Developing the Project:

Data Preprocessing & Collection

- Gather customer transaction data and join it with customer information based on customer ID.

- Replace missing values and outliers as needed.

- Join transaction data with information related to the customers to get an appropriate dataset ready.

- Determine how many times each customer has made a purchase using the feature engineering frequency.

Recency:

- Determine how many days have passed since the client's most recent purchase.

- Determine the average order cost, or AOV (Average Order Value).

- Incorporate these characteristics into the dataset used to train the model.

Model Training:

- Split the data into training and testing sets.

- Train machine learning models such as XGBoost or Random Forest on the engineered features to predict LTV.

- Evaluate model performance using metrics like MAE (Mean Absolute Error) and RMSE (Root Mean Squared Error).

Customer Segmentation:

- Use the predicted LTV values to segment customers into different value categories (e.g., High, Medium, Low).

- These segments can be used for targeted marketing strategies.

Visualization & Evaluation:

- Visualize the distribution of predicted LTV values and assess model performance.

- Present the segmentation of customers based on LTV categories.

Deliverables:

- Python notebook with all the steps from preprocessing to model evaluation.

- Trained model saved using Joblib.

- LTV predictions CSV file with customer ID, predicted LTV, and their segment.

Conclusion:

Conclusion Customer Lifetime Value (LTV), a useful indicator for organizations to identify high-value consumers, is effectively predicted by the study based on purchase history. More focused marketing tactics may be influenced by the model's predictions, which would eventually increase client retention and boost revenue. Customer segmentation according to estimated LTV enables resource allocation and tailored communication. Future developments may entail enhancing the model with additional sophisticated characteristics or outside data sources to increase forecast accuracy even more.

(KUNCHALA SAI ROHITH)

- adjustments (optional).