Proposal: Customer Segmentation on E-commerce Data

Author: Kausik Chattapadhyay

Date: 04/08/2024

DSC-680 Applied Data Science

Project Proposal: Customer Segmentation for E-commerce Personalization

Topic: Implementing Customer Segmentation for Personalized Marketing Strategies in E-commerce.

Business Problem: The project aims to solve the challenge of ineffective generic marketing strategies by developing a customer segmentation model that can accurately categorize customers based on their behavior, preferences, and purchase history. This will enable the E-commerce platform to tailor its marketing efforts, promotions, and product recommendations to specific customer segments, ultimately improving customer satisfaction, retention, and sales.

This project aims at analyzing the content of an E-commerce database that lists purchases made by ~~4000 customers over a period of one year (from 2010/12/01 to 2011/12/09). Based on this analysis, I develop a model that allows to predict the purchases that will be made by a new customer, during the following year and this, from its first purchase.

Datasets: The data for this project will be sourced from the E-commerce platform's database, which has details of purchases made by customers over a period. The dataset includes information such as customer demographics, purchase history, product categories, transaction amounts, and timestamps.

https://www.kaggle.com/datasets/carrie1/ecommerce-data

This data holds 8 variables that correspond to:

InvoiceNo: Invoice number. Nominal, a 6-digit integral number uniquely assigned to each transaction. If this code starts with the letter 'c', it shows a cancellation.

StockCode: Product (item) code. Nominal, a 5-digit integral number uniquely assigned to each distinct product.

Description: Product (item) name. Nominal.

Quantity: The quantities of each product (item) per transaction. Numeric.

InvoiceDate: Invoice Date and time. Numeric, the day and time when each transaction was generated.

UnitPrice: Unit price. Numeric, Product price per unit in sterling.

CustomerID: Customer number. Nominal, a 5-digit integral number uniquely assigned to each customer.

Country: Country name. Nominally, the name of the country where each customer lives.

Methods:
The project will use data preprocessing techniques to clean and prepare the dataset for analysis. Feature engineering will be performed to extract relevant features such as customer lifetime value, purchase frequency, average order value, and product category preferences. Unsupervised learning algorithms such as clustering (e.g., K-means, hierarchical clustering) will be used for customer segmentation based on these features. Additionally, supervised learning techniques like classification and regression may be employed to predict customer behavior and preferences.

Advanced methods in customer segmentation for E-commerce include deep learning models, feature selection/engineering, ensemble learning, sequential modeling, graph-based methods, lifelong learning, unsupervised representation learning, hybrid models, transfer learning, and reinforcement learning. These techniques enhance segmentation accuracy, capture complex patterns, adapt over time, and optimize segmentation strategies. Implementation requires expertise in machine learning, careful experimentation, and validation for effectiveness in E-commerce contexts.

Ethical Considerations:
Ethical considerations include ensuring data privacy and security, obtaining consent for data usage, and avoiding biases in the segmentation model. It's crucial to handle sensitive customer information responsibly and transparently, adhering to legal regulations such as GDPR or CCPA.

Ethical considerations in customer segmentation for E-commerce encompass transparency, informed consent, data security, non-discrimination, fairness, data minimization, accountability, respecting customer preferences, ethical marketing practices, customer empowerment, and continuous monitoring and evaluation. Addressing these aspects requires proactive measures, collaboration, adherence to regulations, ethical guidelines, and a commitment to ethical decision-making throughout the segmentation process.

Challenges/Issues:
Challenges may include dealing with missing or incomplete data, selecting proper features for segmentation, avoiding overfitting or underfitting in the models,

and interpreting and confirming the results effectively. Addressing these challenges will require careful data analysis, feature selection, model evaluation, and validation techniques.

Customer segmentation for E-commerce faces several challenges, including ensuring data quality and consistency, scalability of the segmentation model, adapting to dynamic customer behavior, keeping interpretability, following data privacy regulations, handling segment overlapping, managing long-term model maintenance, ensuring fairness and avoiding biases, integrating with business processes, and dealing with limited labelled data for training. These challenges require a comprehensive approach involving technical ability, domain knowledge, stakeholder collaboration, and continuous evaluation and refinement of the segmentation strategies.

References: To confirm the results and support the project topic, academic research papers, industry case studies, relevant books, and online resources on customer segmentation, machine learning techniques, and E-commerce personalization will be referenced. Additionally, consultation with domain experts and collaboration with stakeholders from the E-commerce platform will provide valuable insights and validation for the project.

https://www.kaggle.com/datasets/carrie1/ecommerce-data