Capstone project: Group 12

Prateek Tripathi, Ranjanipriya Dharmalingam, Rituraj Anand, Sai Venkat

Industry: Retail problem

Type: Classification

Problem: Customer purchase intent Prediction

Problem Statement:

In the competitive online retail market, understanding and predicting customer purchase intent is crucial for optimizing sales and marketing strategies. This project focuses on developing a classification model that predicts whether a customer will complete a purchase based on their online behaviours such as browsing patterns, cart additions, duration of visit, and demographics. By leveraging this model, retailers can tailor their promotions, streamline the checkout process, and engage customers more effectively, converting interest into sales and enhancing both customer satisfaction and business revenue.

Dataset Description:

This dataset contains information about customer behaviour and e-commerce activities on Amazon India website and this dataset contains 2 boolean, 20 numerical and 9 categorical data.

Customer Behaviour Metrics:

Page type, Page value and duration: Type of page (informational, administrative, etc) and amount of time and number of pages visit by users.

Based on activity like exit and bounce: Pages leaving per session and Single-page visits before leaving.

Date time: Date and time of user's activity and if the day is close to a special day, weekday, or weekend.

System information: Operating system and browser information

Cart activity: Number of items added, remove, checkout complete and order amount of a user and if cart abandoned.

Customer and Session Attributes:

Visitor Type and Traffic type: Distinguishes between new, returning, and other types of visitors and the traffic source.

Age, Location and Gender: Demographic information about the customer like city, region, etc.

Outcome Metrics:

Revenue: Indicates if the visit resulted in a successful transaction on amazon.in

Metadata:

Below data provides the information about each field:

• Administrative: Number of pages visited by the user about account management.

• Administrative_Duration: Total duration spent (in seconds) on account management related pages.

• Informational: Number of informational pages like customer care contact, communication and address information of amazon visited by the user.

• Informational_Duration: Total duration spent on informational pages.

• ProductRelated: Number of product-related pages visited by the user.

• ProductRelated_Duration: Total duration spent (in seconds) on product-related pages.

• BounceRates: Average percentage of visitors who view only one page before leaving the amazon website

• ExitRates: Average exit rate of visitors who leave the website from a specific page, after having visited any number of pages of the amazon website.

• PageValues: Average page value of the pages visited by visitor.

• SpecialDay: Proximity to special days (e.g., holidays, promotions) when the visit occurred.

• OperatingSystems: Operating system used by the visitor.

• Browser: Web browser used by the visitor.

• Region: Geographic region of the visitor.

• Location: Location of the visitor.

• TrafficType: Type of traffic source leading to the website.

• Traffic Category: Category of traffic source leading to the website.

• VisitorType: Type of visitor (e.g., new, returning).

• Weekend: Indicates whether the visit occurred on a weekend.

• Age: Age of the visitor.

• Gender: Gender of the visitor.

• Date Time: Date and time of the visit.

• No_Items_Added_InCart: Number of items added to the cart during the visit.

• No_Items_Removed_FromCart: Number of items removed from the cart during the visit.

• No_Checkout_Initiated: Number of items for which checkout processes initiated during the visit.

• cart_abandoned: Indicates whether the cart was abandoned during the visit.

• Revenue: Indicates whether a purchase was made during the visit.

• Order_Amount: Amount of the order placed during the visit.