

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 an online retail website.

Customer Behaviour Metrics:

"Administrative", "Administrative Duration", "Informational", "Informational Duration", "Product Related" and "Product Related Duration": represent the number of different types of pages visited by the visitor in that session and total time spent in each of these page categories.

The values of these features are derived from the URL information of the pages visited by the user and updated in real time when a user takes an action, e.g. moving from one page to another.

The **"Bounce Rate", "Exit Rate" and "Page Value"** features represent the metrics measured by "Google Analytics" for each page in the e-commerce site.

The value of **"Bounce Rate"** feature for a web page refers to the percentage of visitors who enter the site from that page and then leave ("bounce") without triggering any other requests to the analytics server during that session.

The value of **"Exit Rate"** feature for a specific web page is calculated as for all pageviews to the page, the percentage that were the last in the session.

The **"Page Value"** feature represents the average value for a web page that a user visited before completing an e-commerce transaction.

The "**Special Day**" feature indicates the closeness of the site visiting time to a specific special day (e.g. Mother's Day, Valentine's Day) in which the sessions are more likely to be finalized with transaction. The value of this attribute is determined by considering the dynamics of e-commerce such as the duration between the order date and delivery date. For example, for Valentine's day, this value takes a nonzero value between February 2 and February 12, zero before and after this date unless it is close to another special day, and its maximum value of 1 on February 8.

The key elements of the dataset except above attributes are:

Bounce Rates and Exit Rates: Measures the percentage of visitors who leave the site after a single page view or after completing a session.

Shopping Cart and Checkout Activities: Items Added/Removed from Cart: Captures the number of items added to and removed from the shopping cart.

Cart Views and Checkout Confirmations: Tracks the number of times the cart was viewed and the number of completed checkouts.

Cart Abandonment: Indicates if the customer abandoned the cart without completing the purchase.

Customer and Session Attributes:

Visitor Type: Distinguishes between new, returning, and other types of visitors.

Age and Gender: Demographic information about the customer.

Outcome Metrics:

Revenue: Indicates if the visit resulted in a successful e-commerce transaction.

Metadata:

Below data provides the information about each fields.

1. Customer_ID: Unique identifier for each customer.
2. Administrative: Number of administrative pages visited by the customer.
3. Administrative_Duration: Total time spent on administrative pages.
4. Informational: Number of informational pages visited.
5. Informational_Duration: Total time spent on informational pages.

6. ProductRelated: Number of product-related pages visited.
7. ProductRelated_Duration: Total time spent on product-related pages.
8. BounceRates: Percentage of visitors who enter the site and then leave without interacting further.
9. ExitRates: Percentage of pageviews that were the last in the session.
10. PageValues: Average value for a web page that a user visited before completing an e-commerce transaction.
11. SpecialDay: Proximity of the site visit to a special day (e.g., Mother's Day, Valentine's Day).
12. Month: Month of the year when the visit occurred.
13. OperatingSystems: Operating system used by the visitor.
14. Browser: Web browser used by the visitor.
15. Region: Geographic region of the visitor.
16. TrafficType: Type of traffic source.
17. VisitorType: Type of visitor (returning, new, etc.).
18. Weekend: Indicator if the visit occurred on a weekend.
19. Revenue: Indicator if the visit resulted in revenue.
20. ID: Unique identifier for each record.
21. Is_Product_Details_viewed: Indicator if product details were viewed.
22. Session_Activity_Count: Number of activities in a session.
23. No_Items_Added_InCart: Number of items added to the cart.
24. No_Items_Removed_FromCart: Number of items removed from the cart.
25. No_Cart_Viewed: Number of times the cart was viewed.
26. No_Checkout_Confirmed: Number of confirmed checkouts.
27. No_Cart_Items_Viewed: Number of items viewed in the cart.
28. No_Customer_Login: Number of customer logins.
29. No_Page_Viewed: Total number of pages viewed.
30. Customer_Segment_Type: Type of customer segment.
31. Cart_Abandoned: Indicator if the cart was abandoned.
32. No_Checkout_Initiated: Number of initiated checkouts.

- 33. Age: Age of the customer.
- 34. Gender: Gender of the customer.
- 35. Date_Time: Date and time of the visit.
- 36. Location: Location of the visit.
- 37. Date: Date of the visit.
- 38. Time: Time of the visit.