

Report highlighting design decisions and thought process behind selecting Algorithm.



Online Customers Purchasing Intention

The dataset consists of feature vectors belonging to 12,330 sessions .By looking into attribute , we can say it is related to ecommerce website which is related to purchasing

In this project, I will analyse the Customer's Intentions based on the transactions, duration and rates made online in a year. The increasing popularity of online shopping has led to the emergence of new economic activities. To succeed in the highly competitive e-commerce environment, it is vital to understand Customer intention. Understanding what motivates consumer intention is critical because such intention is key to survival in this fast-paced and hypercompetitive environment.

By looking on to the Analysis we have done:

	Importance
PageValues	0.698107
ExitRates	0.084048
ProductRelated_Duration	0.059141
ProductRelated	0.035484
BounceRates	0.033370
Administrative	0.028572
Administrative_Duration	0.028051
Visitor_Type_Returning_Visitor	0.018375
Informational_Duration	0.007560
Informational	0.004090
SpedalDay	0.002665
Visitor_Type_Other	0.000538

The significant importance of Page Value suggests that customers look at considerably different products and its recommendations. Hence a significant improvement on recommendation engines and bundle packages would bring in more conversions. Including more products exploiting the long tail effect in

e-commerce will also bring in more Targets .Here are the revised pointers than can help improve the conversion rate

1. Being transparent to the visitors about the prices and information of product
2. Improving the stay duration by providing them targeted ads like discounts and offers
3. Reducing the bounce rates through faster refreshing rate and attractive landing page which has highly targeted products exclusive for the visitors

4. Personalized emails for each visitors and engaging the loyal visitors(returning visitor) through coupons and exclusive access of products.

process behind selecting algorithms

We selected classification models to predict customer purchasing intent, and Random Forest proved to be the most effective, offering high accuracy and F1 score. Given the dataset's imbalance, additional tuning was unnecessary to achieve strong results.

We selected Random Forest because it:

- Handles class imbalance well through class weighting, improving focus on the minority class.
 - Captures complex patterns in customer behavior more effectively than Logistic Regression or Naive Bayes.
 - Reduces overfitting better than KNN, providing stable predictive accuracy.
 - Balances key metrics (F1 score), essential for imbalanced datasets where accuracy alone doesn't tell the full story.
-