Abstract

Unveiling Patterns in E-commerce: A Predictive Analytics Exploration of Online Shoppers Purchasing Intention Data

Md Khaled Saifullah

Student Number: 501230688

Supervisor:

Date of Submission: 23rd January 2024

Institution: Toronto Metropolitan University

In the rapidly evolving digital landscape, understanding user behavior within e-commerce platforms is crucial for businesses aiming to optimize strategies and enhance revenue. This study embarks on an exploration of purchasing intention data within the e-commerce domain, employing Predictive Analytics as a guiding methodology. The primary aim is to decode intricate patterns embedded within user engagement data, anticipate user actions, and thereby bolster online business revenue.

The project confronts the formidable task of deciphering complex patterns enshrouded within copious volumes of clickstream data. Five cardinal research questions have been articulated to address this challenge, focusing on aspects such as the influence of different page categories on user purchases and variations in user behavior during weekends and special days. The endeavor is to glean insights into engagement metrics' predictability for transactions and identify distinct user segments predicated on categorical attributes. The project aims to answer the following research questions:

How do different informative page categories (like Informational and Product Related pages) contribute to the likelihood of a user making a purchase?

Can we predict the likelihood of a user making a purchase based on metrics such as Bounce Rates, Exit Rates, and Page Values?

How does user behavior vary on weekends compared to weekdays in terms of engagement and conversion rates?

What is the impact of special days on user engagement and transaction rates?

This research project will leverage the Online Shoppers Purchasing Intention Dataset sourced from the UC Irvine Machine Learning Repository as the primary dataset for investigative analysis and to address the formulated research questions. This robust dataset, comprising 12,330 instances enriched with 10 numerical and 8 categorical attributes, forms the foundation for this analytical expedition. Attributes encapsulate diverse elements including pages visited, visit durations, bounce rates, exit rates, page values, special day indicators alongside user characteristics like operating systems browsers and regions.

A hybrid analytical approach is adopted to navigate through these research questions, amalgamating exploratory data analysis with statistical tests while leveraging classification models the models will include Decision trees, Random Forests, K-nearest neighbours, and clustering algorithms for deeper insights. Python emerges as the instrumental programming language supported by libraries like Scikit-learn for machine learning applications; Pandas for adept data manipulation; Matplotlib facilitating intuitive visualizations—all encapsulated within Jupyter Notebooks environment ensuring an interactive exploration.

The culmination of this study promises actionable intelligence enabling businesses to refine strategies meticulously; elevate user experiences aligning with individual preferences; optimize revenue streams adapting to dynamic market trends—all underpinned by a comprehensive understanding gleaned from e-commerce data analytics. This integration of Predictive Analytics underscores a paradigm shift towards informed decision-making processes illuminating pathways for enhanced business performance in an increasingly competitive digital marketplace.

References

Online Shoppers Purchasing Intention Dataset

https://archive.ics.uci.edu/dataset/468/online+shoppers+purchasing+intention+dataset