Predicting Customer Purchase Intentions in E-commerce

Student Name: Leticia Andrade Vieira

Student Number: 2025304

Date: 29/10/2025

GitHub Link: https://github.com/leticiavieira03/-CA1-Strategic-Thinking-student2025304

Table of content:

Introduction:

In an increasingly digital world, online stores and e-commerce have become essential for business success. From small companies to large corporations, e-commerce is no longer an option but a necessity to reach customers, increase brand visibility, and expand market presence. Today's consumers highly value convenience and speed, making a strong online presence crucial for any business.

This shift has created a new type of consumer — more demanding, connected, and seeking personalized experiences. With millions of users visiting online stores every day, companies are now focused on understanding customer behavior to improve their marketing strategies and boost sales performance.

Every online interaction from the pages visited to the time spent browsing generates valuable data that can reveal important behavioral patterns. By analyzing this data, companies can predict the likelihood of a customer completing a purchase and design more effective and personalized marketing campaigns.

This project, **"Predicting Customer Purchase Intentions in E-commerce,"** aims to apply data analysis and machine learning techniques to estimate the probability of a purchase based on user browsing behavior. The dataset used in this project comes from the **UCI Machine Learning Repository** and contains user session information such as page views, bounce rates, and visitor types.

Problem Definition:

In the e-commerce context, companies collect large volumes of interaction data such as page views, session durations, and visitor types. When properly analyzed, these data can reveal important patterns in purchasing intentions. Understanding and predicting customer behavior is crucial to improving marketing efficiency and enhancing the user experience.

However, many organizations still don't effectively utilize this information to anticipate consumer behavior. As a result, they miss opportunities to increase sales, reduce costs, and optimize the performance of their marketing strategies.

This project proposes the application of data analysis and machine learning techniques to identify behavioral patterns that influence purchasing decisions. The results will support strategic thinking, providing insights that help companies create more effective, personalized, and evidence-based marketing strategies.

Objectives:

The main objective of the first phase of this Capstone Project is to explore and understand customer behavior data from e-commerce websites to identify the key factors that influence purchase intentions. This stage focuses on **data exploration and preparation**, forming the foundation for the predictive modelling.

this assignment aims to explore the following objectives:

1- Review relevant literature on customer purchase behavior and predictive analytics in e-commerce

2- Collect, clean, and understand the dataset obtained from the UCI Machine Learning Repository.

3- Conduct exploratory data analysis (EDA) to identify behavioral patterns and relationships between variables that may influence purchase decisions.

4- Summarize initial findings and highlight the main factors affecting customer purchase intentions.

Scope:

Over the first semester, the scope of the project is to explore the following topics and try to answer the following questions:

i. **Customer Behavior Analysis:** Which factors influence whether a visitor completes a purchase? How do variables such as page views, session duration, and visitor type affect purchase intentions?

ii. **Patterns in Online Shopping Data:** What trends or patterns can be identified from the dataset? Are there specific behaviors that indicate a higher likelihood of purchase?

iii. **Data Quality and Reliability:** Are there missing, inconsistent, or biased data points that need to be addressed before further analysis?

At the end of this first-semester phase, the project aims to provide a **comprehensive understanding of the dataset**, uncover preliminary insights about customer behavior, and establish a solid foundation for further predictive modeling in subsequent stages.

Methodology

The project will follow a structured data analysis process divided into two main phases.

1. Data Collection and Preparation (Semester 1)

The dataset used for this project is the *Online Shoppers Purchasing Intention Dataset* from the UCI Machine Learning Repository. This dataset contains information about user sessions on an e-commerce website, including metrics such as page views, bounce rates, exit rates, time spent on site, and visitor type.

During this phase, the data will be cleaned and pre-processed to remove missing values, inconsistencies, and irrelevant variables. Exploratory Data Analysis (EDA) will be conducted to understand the relationships between features and the target variable (“Revenue”), which indicates whether a purchase occurred.

Visualization techniques such as histograms, correlation heatmaps, and box plots will be used to identify trends and patterns.

Data Sources

The dataset used in this project is the **Online Shoppers Purchasing Intention Dataset**,

obtained from the **UCI Machine Learning Repository**. It contains 12,330 records of user sessions collected from a real online retail store over a 12-month period.
The dataset includes features such as the number of pages visited, duration of each session, bounce and exit rates, visitor type (new or returning), operating system, browser, and month of visit. The target variable, *Revenue*, indicates whether a purchase was made during the session (1 for purchase, 0 for no purchase).
The data is publicly available and anonymized, meaning no personally identifiable information (PII) is included. Permission for academic use is granted under the dataset's open license on the UCI repository.

In-depth analysis:

The data that has been found as potential for the project are from open resources, so they are allowed to be used according to their terms and conditions.

Ethical considerations:

While this project on predicting customer purchase intentions in e-commerce does not involve the use of sensitive personal data, ethical considerations remain a fundamental part of the research process.

The dataset used — the **Online Shoppers Purchasing Intention Dataset** from the UCI Machine Learning Repository — contains only anonymized and aggregated information, ensuring that no individual user can be identified.

The project will prioritize **transparency, accuracy, and integrity** in the use and interpretation of data. All analyses will be conducted objectively, avoiding manipulation or misrepresentation of results.

All third-party materials, including datasets and academic references, will be appropriately cited following **Harvard Referencing guidelines**, ensuring proper acknowledgment of the contributions of others and the avoidance of plagiarism.

Furthermore, the report will emphasize the importance of **responsible data handling and ethical use of information** for academic and educational purposes, promoting respect for data privacy and responsible analytics practices in the field of data science.

References: