**Implementing Machine Learning Algorithms to Predict Customer Purchase Behaviour in the Online Retail Domain (A case Study of Amazon)**

**Developing a Predictive Analytics Model for Order Rejection in E-commerce: A Machine Learning Approach Using Amazon Data**

# Chapter 1: Introduction

## 1.1 Background of the Study

The advent of e-commerce platforms such as Amazon has significantly transformed the retail industry, leading to a period marked by the digitization of the consumer's shopping journey. Consequently, the retail industry has become a highly productive environment for generating data, capturing a wide range of consumer interactions that include browsing patterns and purchasing decisions (Michael, 2022). The abundance of data has facilitated the utilisation of advanced analytical methods that offer the potential to unravel the intricacies of consumer behaviour. Due to the increase in online transactions, e-commerce platforms gather a vast amount of data that, when analysed efficiently, can reveal valuable information about consumer preferences, behavioural trends, and purchasing habits (Santos et al., 2022). In fact, the potential of big data goes beyond transactional analysis and allows for a deeper understanding of post-purchase behaviour, including order rejections.

Order rejections, which occur when customers retract their purchase decisions after the transaction has been completed, are a crucial turning point for businesses. Indeed, it is a dimension frequently overlooked in comparison to the more visible aspects of e-commerce operations, yet it presents a significant obstacle. For instance, despite the rapid expansion of online shopping, online retailers are increasingly worried about the significant number of product returns. They experience a return rate of over 20%, which is much higher than the rate of 9% for traditional brick-and-mortar retailers (De, Hu and Rahman, 2013). In fact, the estimated cost of returns in the United States in 2020 was 101 billion dollars (Wang, Yu, & Chen, 2023). Worst still, the retailer may incur a cost of 15 to 30% of the original purchase price for each return.

This is evidence that while these platforms have made shopping more efficient, they also face challenges due to cancelled orders and returns. Each occurrence adds to the logistical difficulties and financial losses. These consequences include both immediate financial costs related to operations and logistics, as well as subtler and widespread impacts on customer satisfaction, loyalty, and trust in the platform.

Studies have demonstrated that post-purchase dissonance often occurs due to a mix of psychological, procedural, and product-related elements (Olejniczak, 2017; Hasan and Nasreen, 2014; Wang, Yu and Chen, 2023). Buyer's remorse, a psychological trait characterised by feelings of regret regarding a purchase, is a notable factor. However, procedural complications such as payment failures, shipping delays, and inadequate return policies can also lead to the rejection of an order (Hasan and Nasreen, 2014). Furthermore, if there is a disparity between consumer expectations and the actual product caused by inaccurate product descriptions, inferior digital representations, or variations in perceived value, it may lead to dissatisfied customers choosing to return their purchases after making a transaction.

Hence, this study aims to thoroughly investigate the complex aspects of consumer behaviour after making a purchase, with a specific emphasis on the factors that trigger the outcomes of rejecting orders in the e-commerce industry. This study undertakes a thorough investigation of the various factors that affect customer behaviour, with a specific emphasis on the data-intensive setting of Amazon. The study aims to utilise machine learning algorithms to move beyond descriptive analytics and venture into the domain of predictive modelling. The aim is to analyse and interpret the detailed data patterns found in Amazon's large transactional datasets, in order to gain a detailed understanding of the factors that lead to order rejections. The research aims to convert complex data insights into a predictive framework that can anticipate and subsequently decrease the occurrence of order rejections. This strategic application aims to enhance consumer analytics in the e-commerce industry, specifically within the dynamic and ever-changing online retail market. It seeks to promote a more robust and customer-focused business model for major online retailers like Amazon.

## 1.2 Statement of the Problem

Order rejection in e-commerce is an under-addressed phenomenon that carries significant implications for business efficiency and consumer trust. Amazon, with its expansive market reach, encounters a substantial number of these incidents, each contributing to resource strain and potential revenue loss. Despite the vast data available, there remains a deficiency in predictive models that can effectively anticipate and mitigate the occurrence of order rejections. The challenge lies in crafting a model that not only predicts the likelihood of rejections but also identifies the triggering factors that lead to such outcomes.

## 1.3 Research Aims and Objectives

The primary aim of this study is to build and validate a predictive model that forecasts the likelihood of order rejections on Amazon, delineating the triggers and consumer behaviours associated with such events. The objectives crafted to achieve this aim are as follows:

1. To explore and analyse the existing body of literature on consumer post-purchase behaviour, with an emphasis on order rejections in e-commerce.
2. To harness and pre-process Amazon's order transaction data to ascertain patterns and commonalities leading to order rejections.
3. To develop a machine learning algorithm that accurately predicts the probability of order rejection, identifying key triggers.
4. To critically assess the model's predictive accuracy using a suite of statistical measures.
5. To craft a strategic framework for Amazon to proactively manage and reduce order rejections, enhancing customer satisfaction and operational efficiency.

## 1.4 Research Questions

This study is driven by the following pivotal research questions:

1. Which factors are most indicative of a pending order rejection on Amazon?
2. How can machine learning models utilize Amazon data to predict the likelihood of order rejections?
3. What strategies can be derived from the predictive model to minimize the incidence of order rejections?

## 1.5 Justification of the Research

The justification for this research is twofold: academic and practical. From an academic perspective, it expands the existing discussion on consumer behaviour after making a purchase, specifically focusing on instances where orders are rejected. This specific area has not been extensively studied in scholarly research especially in relation to predictive modelling. Essentially, it fulfils an important requirement for businesses like Amazon by providing a predictive tool that helps make proactive decisions. This tool has the potential to reduce costs and enhance customer retention by improving the overall customer experience.

## 1.6 Methodology Overview

This study is focused on a quantitative analysis that revolves around a predictive model created using machine learning techniques. The research methodology entails a meticulous procedure of data selection, purification, and analysis, utilising Amazon's vast transaction datasets. The machine learning algorithms will be customised to detect patterns and predictive factors that lead to order rejections. The validation process will involve using appropriate statistical methods. Furthermore, strategic recommendations for Amazon will be developed to demonstrate how the findings can be applied in a practical business context.

## 1.7 Structure of the Dissertation

This dissertation is meticulously organised to facilitate a logical examination of predictive modelling for order rejections on Amazon. Chapter 2 gathers the current body of literature, establishing a strong theoretical foundation for the study. Chapter 3 provides a comprehensive explanation of the research methodology, describing the data collection, pre-processing, model development, and validation processes. Chapter 4 presents the empirical findings, showcasing the predictive performance of the models and the insights obtained from the analysis. Chapter 5 provides an analysis of the consequences of these discoveries, presenting conclusions and suggesting strategic measures for both academic advancement and practical implementation in the e-commerce industry.