UNIVERSITI TEKNOLOGI MARA

ANALYTICAL FRAMEWORK FOR PREDICTING ONLINE PURCHASING BEHAVIOR IN MALAYSIA USING A MACHINE LEARNING APPROACH

NURUL AIN BINTI MUSTAKIM

Thesis submitted in fullfillment
of the requirements for the degree of
Doctor of Philosophy
(Computer Science)

College of Computing Informatics and Mathematics

March 2025

ABSTRACT

In the fast-changing world of digital commerce, predicting online purchasing behavior is essential for improving e-commerce strategies. This study presents a post-pandemic machine learning-based analytical framework specifically designed for Malaysia's e-commerce market. The framework addresses a gap in predictive analytics by combining computational techniques, consumer behavior theories, and demographic data to better understand and forecast purchasing trends. The framework uses machine learning methods, including classification, clustering, feature selection, and parameter tuning, to improve accuracy and reliability. It is organized into three phases: preliminary investigation, implementation and analysis, and validation. The descriptive analysis examines purchasing behavior through correlation and regression analyses, while the predictive model uses decision trees (J48, Random Tree, REPTree), rule-based algorithms (JRip, OneR, PART), and clustering (K-Means) to identify patterns and predict trends. Feature selection techniques, such as WrapperSubsetEval, were used to improve focus on key attributes, and parameter tuning further optimized performance. Among the three datasets analyzed (D1, D2, and D3), Dataset 3, which emphasizes psychological and emotional factors, achieved the highest accuracy and predictive performance. Validation included both machine learning evaluation and expert feedback. Machine learning results showed strong performance, with Random Tree and PART delivering the most reliable predictions. Adjustments, such as tuning K-Values and binary splits, further enhanced results. Expert feedback validated the framework's alignment with the Howard-Sheth consumer behavior model and suggested improvements, such as including additional behavior related factors and refining segment categories. Key findings highlight the growing participation of female consumers in online shopping, signaling changing trends in Malaysia's e-commerce landscape. This study also introduces a unique dataset on Malaysian online consumer behavior, filling a critical gap and providing valuable insights for businesses. The framework aligns with Malaysia's Science, Technology, Innovation, and Economic (MySTIE) framework and supports SDG 8 and KEGA 2 (Digital Economy) goals. By integrating classification, clustering, and consumer behavior models, this framework offers practical tools for understanding and predicting consumer behavior. Its validation, including evaluation and expert feedback, ensures reliability and adaptability, making it a valuable resource for improving e-commerce strategies in Malaysia.

ACKNOWLEDGEMENT

TABLE OF CONTENTS

CHAPTER 1
INTRODUCTION

1.1   Preamble

This chapter introduces the study by presenting the background of the research. It then highlighting the critical issues through the problem statement, followed by the research question, objectives, the scope of study and significance. Lastly, a summary of the structure of the whole thesis is presented.

1.2   Research Background

In the late 1990s and early 2000s, the advent of e-commerce marked a pivotal period in the business arena. During this time, businesses were predominantly focused on establishing a digital presence and guaranteeing the security of online transactions, signifying a noteworthy departure from traditional in-store shopping models. Due to the novelty of e-commerce, the reliability of operations played a crucial role in fostering consumer trust (Yuliarto & Gusti, 2022). This phase witnessed the implementation of secure payment methods and data protection protocols, laying the foundation for the security standards prevalent in contemporary online transactions. Additionally, there was a significant emphasis on enhancing website usability. Early e-commerce platforms also prioritized user-friendly interfaces, highlighting clear navigation, easily accessible product information, and streamlined checkout processes to ensure that online shopping was as intuitive and convenient as possible. This assertion is corroborated by Patrada and Andajani (2021), who determined that e-customer satisfaction is influenced by factors such as security, convenience, and website design. Furthermore, during this era, analytics served as an initial and investigative tool in the rapidly evolving digital landscape. The primary focus of analytics was on guaranteeing website functionality and transaction security. This emphasis stemmed from the fact that both individuals and companies were still adjusting to the concept of e-commerce, emphasizing the crucial importance of establishing a secure and reliable online environment.