UNIVERSITI TEKNOLOGI MARA

CLASSIFICATION AND VISUALIZATION OF E-COMMERCE PRODUCT REVIEWS COMPARISON USING SUPPORT VECTOR MACHINE

NUWAIRAH AIMI BINTI AHMAD KUSHAIRI

BACHELOR OF COMPUTER SCIENCE (HONS.)

AUGUST 2023

ACKNOWLEDGMENT

Alhamdulillah, all praise, and thanks be to The Almighty, Allah S.W.T, for His blessings and for granting me the strength and determination to successfully complete this project. I extend my deepest gratitude to my supervisor, Ts. Dr. Khyrina Airin Fariza Binti Hj Abu Samah, for her unwavering guidance and helpful advice throughout the entire journey of this project. I am truly grateful for all the insightful feedbacks provided, endless long discussions and for her willingness to guide me until the successful completion of the project. I would also like to show my appreciation towards my CSP lecturer, Dr. Raihah Binti Aminuddin for all the knowledge and guidance in completing the project throughout the whole two semesters. To my examiner, Madam Nurazian Mior Dahalan, thank you for providing insightful comments and valuable suggestions, which have significantly contributed to the quality of my project.

I would like to express my heartfelt appreciation towards my dear friends, whose unwavering support have been instrumental in completion of this project and throughout our degree journey. To Nur Amielia Natasha, Aisyah Husna, Nur Eisyah Farihah, and Muhammad Taufeeq, I feel incredibly fortunate to have shared this experience with such amazing group of friends. To Nur Maherah Az Zaharah, thank you for being an ever-supportive friend and roommate. Your presence during my lowest moments provided immense comfort, and your positivity served as a constant source of strength. Special appreciation goes to Muhammad Azri for his comforting presence and endless words of encouragement, which motivated me to push through and strive to complete this project successfully. Special credit to SEVENTEEN for being a constant source of joy and motivation. Your songs filled with positivity have brightened my days, giving me the strength to overcome obstacles and persevere through this journey. To all of you, I am immensely grateful for the lasting impact you've had on my academic journey and for being a significant part of this project completion.

Lastly, my deepest appreciation goes to my dearest loved ones, my family. Their unwavering encouragement and boundless support have been the foundation of my strength throughout this journey. To my beloved parents, thank you for your endless comfort and unconditional love. I am truly grateful for all the efforts you put into ensuring my happiness and well-being throughout this journey. To my sisters, thank you for the wise words and constant companionship, which supported me through difficult times. Thank you for reassuring me in times of doubt and lifting my spirits when everything seemed challenging. I dedicate this to all of you, with all my love and heartfelt appreciation.

ABSTRACT

E-commerce has experienced significant growth as a platform for online shopping, offering convenience and cost-saving benefits. Especially in Malaysia, Shopee is known to be one of the leading e-commerce platforms, attracting millions of monthly visitors. These days, it has come to our attention that online product reviews play a crucial role in influencing consumer behaviour by building trust, identifying customer needs, and improving satisfaction. It was agreed by 97.3% of 186 respondents from a questionnaire survey to rely on product reviews before purchasing any product. Nevertheless, 96.2% of the respondents agreed that not all product reviews are helpful when shopping online. Online customers or shoppers rely on these reviews as decision-making to purchase products. It becomes time-consuming to read through the reviews, especially when it is not product related. Moreover, the large number of reviews could lead to information overload, which exhausts customers to decide. Therefore, this project aims 1) to design a web-based system that can classify the comparison of useful and not useful product reviews from Shopee using the Support Vector Machine (SVM) algorithm and visualize the comparison, 2) to develop the designed system, and 3) to test the functionality and usability of the system. Users can enter a maximum of six product links into the system. The system classifies the reviews into "Useful" or "Not Useful" based on the review text, star rating, duplicated spam, and sentiment score. Afterward, the system recommends the best shop to purchase from and visualizes the reviews to compare products from different shops. The SVM classifier model successfully classified the reviews with an accuracy of 96.8% during the testing stage of the classification. Aside from that, the system was thoroughly evaluated for its functionality, which passed all test cases with expected performance. From the Mann-Whitney U Test in reliability testing, the obtained *p-value*, 0.008, is below the significance level, $p\text{-value}<0.05$, hence, rejecting the null hypothesis. It means a significant difference exists between the time taken for manual evaluation and the evaluation using the web-based system. As a result, the system has the potential to help users make informed decisions while making purchases.

TABLE OF CONTENTS

CONTENT	PAGE
SUPERVISOR APPROVAL	**i**
STUDENT DECLARATION	**ii**
ACKNOWLEDGMENT	**iii**
ABSTRACT	**iv**
TABLE OF CONTENTS	**v**
LIST OF FIGURES	**viii**
LIST OF TABLES	**xii**
LIST OF ABBREVIATIONS	**xiv**
CHAPTER ONE: INTRODUCTION	**1**
1.1 Background of Study	1
1.2 Problem Statement	2
1.3 Objectives	4
1.4 Scope	4
1.5 Significance of the Study	5
1.6 Summary	5
CHAPTER TWO: LITERATURE REVIEW	**6**
2.1 E-Commerce Platforms	7
2.2 Shopee E-Commerce	7
2.3 Electronic Products	8
2.4 Attributes of Useful Online Product Reviews	9
2.5 Web Scraping	10
2.6 Machine Learning Approaches	12
2.6.1 Supervised Learning	13
2.6.2 Unsupervised Learning	13
2.6.3 Comparison Between Machine Learning Approaches	14
2.7 Classification Algorithms	15
2.7.1 Logistic Regression	15
2.7.2 Naïve Bayes	16
2.7.3 Decision Tree	16
2.7.4 Random Forest	17
2.7.5 K-Nearest Neighbors	17
2.7.6 Support Vector Machine	17
2.7.7 Comparison Between Classification Algorithms	18
2.8 Keyword Extraction	20
2.8.1 Term Frequency and Inverse Document Frequency	20
2.8.2 Rapid Automatic Keyword Extraction	21

2.8.3 Comparison of Keyword Extraction 21
2.9 Visualization Techniques 22
2.9.1 Line Chart 23
2.9.2 Bar Chart 23
2.9.3 Pie Chart 24
2.9.4 Scatter Plot 25
2.9.5 Box and Whisker Plot 26
2.9.6 Word Cloud 27
2.9.7 Comparison of Visualization Techniques 28
2.10 Development Approaches 29
2.10.1 Web Development Approach 29
2.10.2 Native Development Approach 30
2.10.3 Comparison of Development Approaches 30
2.11 Related Work 31
2.11.1 Detecting Truthful and Useful Reviews for Products Using Opinion Mining 31
2.11.2 Assessing The Helpfulness of Online Hotel Reviews: A Classification-based Approach 32
2.11.3 Features Comparison of Related Works 34
2.12 Summary 35

CHAPTER THREE: METHODOLOGY 36

3.1 Introduction 36
3.2 Modified Waterfall Methodology 36
3.3 Requirement Phase 40
3.4 Design Phase 46
3.5 Implementation 46
3.6 Software and Hardware Requirements 48
3.7 Testing Phase 48
3.7.1 Functionality Testing 49
3.7.2 Reliability Testing 50
3.8 Summary 51

CHAPTER FOUR: DESIGN AND DEVELOPMENT 52

4.1 System Design 52
4.1.1 Flowchart 52
4.1.2 Use Case Diagram 55
4.1.3 User Interface Design 57
4.2 Back End Development 60
4.2.1 Data Collection 61
4.2.2 Data Pre-processing 63
4.2.3 Support Vector Machine Classifier Model 69
4.2.4 Model Deployment 73
4.3 Front End Development 74
4.3.1 Home Page 74
4.3.2 Track Product Page 74