

Universiti Teknologi MARA

**KUALA TERENGGANU HOMESTAY
RECOMMENDER SYSTEM USING CONTENT
BASED FILTERING ALGORITHM**

Nur Atierah Binti Nor Azman

**Thesis submitted in fulfilment of the requirements for
Bachelor of Computer Science (Hons.)
College of Computing, Informatics and Mathematics**

January 2024

ACKNOWLEDGEMENT

The study was successfully conducted within the designated timeframe, and I express gratitude to Allah for His magnificence and the abundant blessings bestowed upon me. Firstly, I would like to extend my appreciation to my supervisor for her invaluable guidance and support throughout this process. This research would not have been achievable without the invaluable support and guidance provided by Miss Nik Marsyahariani Binti Nik Daud. Her assistance not only facilitated the completion of this study but also demonstrated her unwavering commitment, time and dedication. Furthermore, I would like to extend my sincere appreciation to Madam Ummu Fatimah, who fulfilled the role of my instructor for CSP600 and CSP650. Throughout the entire semester, she generously imparted her expertise, offered unwavering support, and served as a source of inspiration. Other than that, I would like to thank all of my lecturers that has been consistently reminding, encouraging and ideas sharing throughout the journey.

My family members also have played a crucial role in supporting me throughout every step of this project. I would like to express my profound gratitude towards my parents for their unwavering provision of support either by physical and mental aspects, as well as their consistent encouragement and financial aid during times of utmost necessity. I also would like to extend my gratitude to my esteemed friends for their invaluable assistance and unwavering emotional support, which played a pivotal role in the successful completion of my final year project. Last but not least, may this project become one of my practices where I can share my knowledge to the readers, to the best of my knowledge.

ABSTRACT

Kuala Terengganu's tourism industry is expanding quickly, which has increased demand for acceptable hospitality and lodging alternatives. However, finding the ideal homestay that aligns with user preference can be time taking task. Hence, Content-Based filtering is capable to find the matching homestay based on the user preferrable criteria of the homestay such as price, locations and availability. Then in order to identify the distinguishing features or term of the homestay, TF-IDF will be calculating the importance term in the homestay review or description. By analyzing this textual content, TF-IDF will creates a representation of each homestay in the form of a corpus vector. It is required to find and compares the similarity between the user's preferences and the characteristics of the homestay. The similarity measure needs to be done in order to generate the recommendations. Cosine Similarity and TF-IDF will allow the prototype to identify the homestays that align closely with the user's desired features. The system's performance is meticulously evaluated through precision at 79.3%, recall at 93%, F1-score at 86%, RMSE at 9.95 and MAE at 9.42. These results demonstrate the system capability to offer accurate and relevant homestay suggestions. The project has successfully achieved all of the objectives and able to demonstrated the practical application of content-based filtering algorithm.

TABLE OF CONTENTS

CONTENTS	PAGE
TITLE PAGE	ii
SUPERVISOR APPROVAL	iii
STUDENT DECLARATION	iv
ACKNOWLEDGEMENT	v
ABSTRACT	vi
LIST OF FIGURES	x
LIST OF TABLES	xiii
LIST OF ABBREVIATIONS	xiv

CHAPTER ONE: INTRODUCTION

1.1	Background of Study	1
1.2	Problem Statement	2
1.3	Research Objectives	4
1.4	Research Scope	4
1.6	Project Framework	5
1.7	Conclusion	6

CHAPTER TWO: LITERATURE REVIEW

2.1	Machine Learning	7
2.2	Recommender System	8
	2.2.1 Phase of Recommender System	8
	2.2.1.1 Information Collection Phase	9
	2.2.1.2 Learning Phase	9
	2.2.1.3 Recommendation Phase	10
2.3	The Use of Recommender System	11

2.4	Filtering Techniques in Recommender System	12
2.4.1	Content-Based Filtering	13
2.4.2	Collaborative Filtering	13
2.4.3	Hybrid Filtering	14
2.4.4	Comparison Between CBF, CF and Hybrid Filtering	16
2.5	Content-Based Filtering Algorithm in Recommender System	17
2.5.1	Content-Based Filtering Architecture	18
2.5.2	Cosine Similarity Measure	19
2.5.3	Term Frequency-Inverse Document Frequency (TF-IDF)	20
2.6	Similar Works in Content Based Filtering	21
2.7	Implication of Literature Reviews	25
2.9	Conclusion	31

CHAPTER THREE: PROJECT METHODOLOGY

3.1	Overview of Research Framework Methodology	32
3.1.1	Detailed Research Framework	32
3.2	Project Planning	37
3.2.1	Preliminary Study	37
3.2.2	Literature Review	38
3.3	Requirement Analysis	38
3.3.1	User Requirement Analysis	39
3.4	Data Collection	42
3.4.1	Data Filtering	43
3.4.2	Text Tokenization	44
3.4.3	Removing Stop Words	45
3.5	Prototype Design and Development	46
3.5.1	Prototype Flowchart	47
3.5.2	Prototype architecture	48
3.5.3	User Interface	49
3.6	Evaluation Phase	51
3.6.1	Accuracy Metrics	51
3.7	Documentation	51
3.8	Gantt Chart	52