

Universiti Teknologi MARA

**Customer Profiling Using K-Means
Clustering Method**

Nik Asyrani Asna Binti Nik Mohd Asri

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

JANUARY 2024

ACKNOWLEDGEMENT

Alhamdulillah, praises and thanks to Allah because of His Almighty and His utmost blessings, I was able to finish this research within the time duration given. Firstly, my special thanks goes to my supervisor, Dr. Norzehan Sakamat, who not only helped me finish this research but also devoted her time and effort into doing so. In addition to that, I would like to express my gratitude to Madam Ummu Fatimah Binti Mohd Bahrin, who served as my lecturer for CSP600/CSP650 and shared her knowledge with me, provided support, and inspired me for the entirety of the semester.

Special appreciation also goes to my beloved parents for the unending support they have given me, both physically and mentally, as well as encouragement and even financial assistance when I needed it the most. Nobody has been more important to me in the pursuit of this project than my family members. I am grateful to all of those with whom I have had the pleasure to work during this project.

Last but not least, I would like to give my gratitude to my dearest classmates for lending me a helping hand and providing emotional support in order to finish the final year project successfully.

ABSTRACT

Businesses must understand customer behavior in today's ever-changing business environment in order to properly customize their marketing strategy. With the use of the K-means clustering approach, this research seeks to improve customer profiling by allowing businesses to divide their customers into discrete groups according to shared behaviors and preferences. Through the analysis of various customer data sets, such as people, products, promotion, place, the K-means algorithm can detect clusters that correspond to consistent client groups. The next phase of the project will concentrate on creating a thorough customer profiling system that makes use of these clusters to produce insightful data on customer preferences. This will allow companies to create customized marketing campaigns that appeal to certain target customers. By providing more relevant content, this strategy not only increases customer satisfaction but also improves marketing effectiveness and boosts conversion rates. In order to facilitate smooth business interactions with the profiling system, the project will incorporate the K-means clustering technique for consumer segmentation and optimize it. The end goal is to provide organizations with an effective tool that helps them better understand their customer segments, enabling them to develop more individualized and successful communication strategies in the context of a competitive market.

TABLE OF CONTENTS

CONTENT	PAGE
SUPERVISOR APPROVAL	ii
STUDENT DECLARATION	iii
ACKNOWLEDGEMENT	iv
ABSTRACT	v
TABLE OF CONTENTS	vi
LIST OF FIGURES	ix
LIST OF TABLES	xi
CHAPTER ONE: INTRODUCTION	1
1.1 Background of Study	1
1.2 Problem Statement	2
1.2.1 Lack of Customer Understanding.....	2
1.2.2 Inefficient Marketing Strategies.....	2
1.2.3 Limited Customer Segmentation.....	3
1.3 Objectives	4
1.4 Project Scope	4
1.4.1 User	4
1.4.2 Data	4
1.4.3 Algorithm	5
1.5 Project Significance	5
1.5.1 Improved customer engagement	5
1.5.2 Optimize marketing resources.....	5
1.5.3 Better customer understanding.....	5
1.6 Overview of Research Framework	6
1.7 Conclusion	7
CHAPTER TWO: LITERATURE REVIEW	8
2.1 Customer Profiling.....	8
2.2 Clustering.....	9
2.2.1 Overview of K-means Clustering Algorithm	11

2.2.2	Implementation of K-means in Various Problem	12
2.3	Similar Work Section	22
2.4	Implication of Literature Review	33
2.5	Conclusion	33
CHAPTER THREE: METHODOLOGY		34
3.1	Overview of Research Framework Methodology	34
3.2	Preliminary Study	38
3.2.1	Knowledge Acquisition.....	39
3.3	Data Finding	41
3.3.1	Data Analysis	42
3.3.2	Data Collection.....	42
3.4	Design and Implementation Phase.....	44
3.4.1	Software Development Life Cycle (SDLC).....	46
3.4.2	System Architecture.....	47
3.4.3	Flowchart	48
3.4.4	Pseudocode.....	50
3.4.5	User Interface Design.....	51
3.5	System Requirement	52
3.5.1	Hardware Recommendation	52
3.5.2	Software Recommendation	52
3.6	Evaluation Phase.....	54
3.6.1	Accuracy of Algorithm	55
3.7	Documentation.....	58
3.8	Gantt Chart	59
3.9	Chapter Summary	60
CHAPTER FOUR: RESULT AND FINDING		61
4.1	Conceptual Framework.....	61
4.2	Results for Objective 1	63
4.2.1	Analysis of Literature Review on K-means Clustering	63
4.2.2	Dataset Preprocessing	64
4.2.2.1	Importing Dataset Python Libraries	64
4.2.2.2	Transmitting Dataset into Python Coding	65
4.2.2.3	Data Preparation.....	66