

Universiti Teknologi MARA

**Customer Segmentation Using
Clustering Techniques**

MOHAMAD AMIR SALIHIN BIN MUSTAFA

**Thesis submitted in fulfilment of the requirements
for Bachelor of Computer Science (Hons.)
College of Computing, Informatics and
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 go to my supervisor, Madam Norulhidayah Binti Isa. This thesis will not be possible to be completed without the guidance, support and advice from her. She did not only help me finish this research but also devoted her time in doing so. I would also want to express my sincere gratitude to Madam Ummu Fatimah Binti Mohd Bahrin, who served as my CSP600 and CSP650 lecturer. She shared her valuable knowledge and guided me through the entire semester.

Special thanks to my family members who always support me from behind. My dear parents deserved my unending thanks as they not only supported me physically and emotionally but also through financial assistance. I am also grateful to my fellow friends for lending me their helping hand when I was in need so that I will be able to finish my final year project successfully.

ABSTRACT

This study explores clustering techniques for customer segmentation, focusing on the K-Means algorithm in particular, and uses a dataset that was obtained from the customer data of an international supermarket company. Studying clustering methodologies, creating a K-Means clustering model, and assessing its quality for efficient client segmentation are the main goals. The Elbow method is employed to determine the ideal number of clusters (K value), resulting in the segmentation of customers according to their buying patterns. Customer profiling is the result of the segmentation process. The Silhouette Score is used to assess the quality of the clustering model, and it achieves a good value of 0.54. By demonstrating a good balance between cluster cohesiveness and cluster separation, this score shows that the K-Means method is successful in identifying unique customer categories. The supermarket company can improve overall business performance and customer satisfaction by using tailored customer engagement and targeted marketing strategies made possible by the insightful customer profiles that are produced.

TABLE OF CONTENTS

CONTENT	PAGE
SUPERVISOR APPROVAL	ii
STUDENT DELCARATION	iii
ACKNOWLEDGEMENT	iv
ABSTRACT	v
TABLE OF CONTENTS	vi
LIST OF FIGURES	ix
LIST OF TABLES	xi
LIST OF ABBREVIATIONS	xii
CHAPTER ONE: INTRODUCTION	
1.1 Background of Study	1
1.2 Problem Statement	3
1.3 Objective	4
1.4 Project Scope	4
1.5 Project Significance	5
1.6 Overview of Research Framework	6
1.7 Conclusion	9
CHAPTER TWO: LITERATURE REVIEW	
2.1 Introduction	10
2.2 Customer Segmentation	11
2.3 Clustering Techniques	13
2.3.1 K-Means and How Does It Works	13
2.3.2 Flowchart of K-Means Algorithm	14
2.3.3 DBSCAN and How Does It Works	15
2.3.4 Flowchart of DBSCAN Algorithm	15

2.3.5 Hierarchical Clustering and How Does It Works	17
2.3.6 Flowchart of Hierarchical Clustering	18
2.4 Implementation of Clustering Techniques in Various Problem	19
2.5 Similar Works	28
2.6 The Implication of Literature Review	36
2.7 Conclusion	37

CHAPTER THREE: RESEARCH METHODOLOGY

3.1 Overview Of Research Framework Methodology	38
3.1.1 Detailed Research Framework	38
3.2 Business Understanding	42
3.2.1 Literature Study	42
3.3 Data Understanding	43
3.3.1 Data Collection	43
3.4 Data Preparation	44
3.4.1 Data Selection and Cleaning	44
3.4.2 Data Reduction	44
3.4.3 Data Transformation	45
3.5 Modelling	45
3.5.1 Algorithm Design	46
3.5.2 Customer Segmentation Model Architecture Design	47
3.5.3 Model Development	48
3.6 Evaluation	49
3.6.1 Clustering Model Evaluation	49
3.7 Deployment	50
3.8 Gantt Chart	50
3.9 Conclusion	51

CHAPTER FOUR: RESULT AND FINDING

4.1 Conceptual Framework	52
4.2 Analysis of Literature Review on K-Means	54