

**Universiti Teknologi MARA**

**Comparative Analysis of k-Means and  
k-Medoids for Clustering Exam  
Questions**

**Nurul Zafirah binti Mokhtar**

**Thesis submitted in fulfilment of the requirements  
for  
Bachelor of Information Technology (Hons.)  
Information Systems Engineering  
Faculty of Computer and Mathematical Science**

**January 2016**

## **SUPERVISOR'S APPROVAL**

### **COMPARATIVE ANALYSIS OF K-MEANS AND K-MEDOIDS FOR CLUSTERING EXAM QUESTIONS**

By

**NURUL ZAFIRAH BINTI MOKHTAR  
2013591329**

This report was prepared under the supervision of the project supervisor, Nur Suhailayani Suhaimi. It was submitted to the Faculty of Computer and Mathematical Science and was accepted in partial fulfilment of the requirement for the degree of Bachelor of Information Technology (Hons) Information Systems Engineering.

Approved by

.....  
Nur Suhailayani Suhaimi  
Project Supervisor

FEBRUARY 4, 2016

## **STUDENT DECLARATION**

I certify that this report and the project to which it refers is the product of my own work and any idea or quotation from the work of other people, published or otherwise are fully acknowledged in accordance with the standard referring practice of the discipline.

.....  
NURUL ZAFIRAH BINTI MOKHTAR  
2013591329

FEBRUARY 4, 2016

## ABSTRACT

Clustering has become more needed as a technique to cluster with intent to provide better grouping due to several problems. Clustering dynamic data is a challenge in identifying and forming groups. This unsupervised learning usually leads to undirected knowledge discovery. The cluster detection algorithm searches for clusters of data which are similar to one another by using similarity measures. Determining the suitable algorithm which can bring the optimized group clusters could be an issue. K-Means and k-Medoids are popular technique used in the world of clustering. Grouping an exam questions is a confusing tasks as it have to deal with the attributes and parameters of the data. Both techniques also may resulted in different outcomes. Depending on the parameters and attributes of the data, the results obtained from using both k-Means and k-Medoids could be varied. Each and every attribute and parameters selected undergo several process of data mining starting from pre-processing until the analysis of the data. The attributes and parameters that takes part in grouping the questions are marks, cognitive level and also the topics of the questions. Then the results is compared to determine which technique will produce higher accuracy results. This paper presents a comparative analysis of both algorithm in different data clusters to lay out strength and weaknesses of both. The grouping of an exam questions encompass low, medium and high level questions. Throughout the studies that conducted in determining the cluster, ITS570 course was used as a data and a set of cluster rules that hold the centroid and medoids value for both algorithm were produced at the end of this project for both techniques. The studies had found that k-Medoids produced higher accuracy result with 0.11% higher than k-Means. As a conclusion, with this type of data, k-Medoids algorithm had shown higher accuracy result rather than k-Means.

# TABLE OF CONTENT

<b>CONTENTS</b>	<b>PAGE</b>
<b>SUPERVISOR'S APPROVAL</b>	ii
<b>STUDENT DECLARATION</b>	iii
<b>ACKNOWLEDGMENTS</b>	iv
<b>ABSTRACT</b>	v
<b>TABLE OF CONTENTS</b>	vi
<b>LIST OF FIGURES</b>	xi
<b>LIST OF TABLES</b>	xiii
<b>LIST OF EQUATION</b>	xiv
<b>LIST OF ABBREVIATION</b>	xv
<b>CHAPTER ONE: INTRODUCTION</b>	1
1.1 Project Background	1
1.2 Problem Statement	3
1.3 Project Aim and Objectives	4
1.4 Project Scope	4
1.5 Project significance	5
1.6 Organization Chapter	5
1.7 Summary	6
<b>CHAPTER TWO: LITERATURE REVIEW</b>	7
2.1 Exam Questions Type	7