

UNIVERSITI TEKNOLOGI MARA

SUSPICIOUS ONLINE CAR ADVERTISEMENT  
DETECTOR USING PRINCIPLE COMPONENT  
ANALYSIS (PCA)

FATHIAH HUSNA BINTI FIRDAUS

BACHELOR OF COMPUTER SCIENCE (HONS.)

JULY 2019



## **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. Ali Seman. He is a supportive supervisor with all the wise words that always motivate me. He said, final year project is a journey that everyone must go through with all the ups and downs. A true hard work will be paid off with a set of knowledge gained through the learning process along with the skills sharpened in order to accomplished the project.

Special appreciation also goes to my beloved parents who always understand and support me in everything I do. Also, thank you and appreciation to all my colleagues for being together and kept supporting me along this journey.

Last but not least, I would like to give my gratitude to my dearest friend, Muhammad Syazwan bin Sabran that had always motivate me in developing this project and finish up this research.

## **ABSTRACT**

E-shopping or also known as Electronic Commerce (E-Commerce) is an online shopping website that have been evolved in these past years. The evolution of online shopping website has been beneficial to both of the sellers and customers mainly in terms of time and cost. Despite of the goods of online shopping, there is possibility that the advertisement in the website is considered suspicious and a potential scam. Hence, this study is to propose suspicious advertisement detector by using data mining technique which is Principal Component Analysis (PCA). PCA is a dimensionality reduction technique that is robust and ease the visualization task. The result of PCA will be used to identify the outliers from the PCA scatter plot. The further analysis is done by apply statistical box plot method, K-Means Clustering and compare the distance of outliers and its centroid. The outliers from the result are the suspicious advertisements which could be potential scammer or genuine seller that differs in value of price and mileage of a car in the advertisement.

## TABLE OF CONTENTS

CONTENT	PAGE
SUPERVISOR APPROVAL.....	i
STUDENT DECLARATION.....	ii
ACKNOWLEDGEMENT.....	iii
ABSTRACT.....	iv
TABLE OF CONTENTS.....	v
LIST OF FIGURES.....	viii
LIST OF TABLES.....	x
INTRODUCTION.....	1
1.1 Project Background.....	1
1.2 Problem Statement.....	2
1.3 Research Questions.....	3
1.4 Objective.....	3
1.5 Project Scope.....	3
1.6 Project Significance.....	3
LITERATURE REVIEW.....	5
2.0 Introduction.....	5
2.1 Online Advertisement.....	5
2.2 Machine Learning.....	6
2.3 Application of Machine Learning.....	7
2.3.1 Machine Learning for Online Shopping.....	7
2.3.2 Machine Learning for Cybercrimes.....	9
2.4 Principal Component Analysis (PCA).....	11

<b>2.4.1 Health</b> .....	11
<b>2.4.2 Biological Sciences</b> .....	12
<b>2.4.3 Geological and Earth Science</b> .....	12
<b>2.4.4 Engineering</b> .....	13
<b>2.4.5 Industries</b> .....	13
<b>METHODOLOGY</b> .....	15
<b>3.0 Introduction</b> .....	15
<b>3.1 Requirement</b> .....	16
<b>3.1.1 Data Retrieval</b> .....	16
<b>3.1.2 Data Cleaning</b> .....	18
<b>3.1.3 Data Transformation</b> .....	19
<b>3.2 Design</b> .....	21
<b>3.2.1 Data analysis design by using PCA</b> .....	21
<b>3.2.2 User Interface design</b> .....	24
<b>3.3 Development</b> .....	27
<b>3.4 Experiment</b> .....	27
<b>RESULT AND DISCUSSION</b> .....	30
<b>4.0 Introduction</b> .....	30
<b>4.1 Case Study 1: Perodua Myvi</b> .....	30
<b>4.1.1 Detecting Suspicious Advertisements using PCA</b> .....	30
<b>4.1.2 Detecting Suspicious Advertisements using Cluster Analysis</b> .....	34
<b>4.2 Case Study 2: Perodua Kancil</b> .....	36
<b>4.2.1 Detecting Suspicious Advertisements using PCA</b> .....	36
<b>4.2.2 Detecting Suspicious Advertisements using Cluster Analysis</b> .....	40
<b>4.3 Case Study 1: Honda Civic</b> .....	42
<b>4.3.1 Detecting Suspicious Advertisements using PCA</b> .....	42
<b>4.3.2 Detecting Suspicious Advertisements using Cluster Analysis</b> .....	46