

**UNIVERSITI TEKNOLOGI MARA**

**DATA SCRAPING TO ANALYZE TOURIST  
ATTRACTIONS IN MALAYSIA**

**MUHAMMAD ASLAM BIN MD ADAM**

**Thesis submitted in fulfillment of the requirements for  
Bachelor of Computer Science (Hons.) with the supervision of  
Sir Ahmad Farid Bin Najmuddin and co-ordinate by  
Faculty of Computer and Mathematical Sciences**

**JANUARY 2019**

## **SUPERVISOR'S APPROVAL**

### **DATA SCRAPING TO ANALYZE TOURIST ATTRACTION IN MALAYSIA**

By

**MUHAMMAD ASLAM BIN MD ADAM  
2016565959**

This thesis was prepared under the direction of thesis supervisor, Sir Ahmad Farid Bin Najmuddin. It was submitted to the Faculty of Computer and Mathematical Sciences and was accepted in partial fulfilment of the requirements for the degree of Bachelor of Computer Science (Hons).

Approved by:

.....  
Ahmad Farid Bin Najmuddin  
Thesis Supervisor

JANUARY 8, 2019

## **DECLARATION**

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

.....

MUHAMMAD ASLAM BIN MD ADAM

2016565959

JANUARY 8, 2019

## **ACKNOWLEDGEMENT**

Alhamdulillah, praises and thanks to Allah SWT because of His Almighty and His utmost blessings, I was able to finish this final year project within the time duration given. Firstly, I would like to give my special thanks of gratitude to my supervisor, Sir Ahmad Farid Bin Najmuddin and my co-supervisor, Sir Muhamad Nasir bin Abdullah who had monitored, advised, giving guidance and supported me from beginning until completion of this project.

Also, special thanks to my lecturer CSP600 and CSP650, Madam Azilawati binti Azizan and Madam Masurah Binti Mohamad who had taught and shared her knowledge to me and my friends. All of them never give up on supporting and spending lots of time with me although during their busy time.

Furthermore, I would like to extend my special gratitude towards my beloved parents who always support me and give me encouragement since the beginning. There are my main source of strength and motivation in completing this final year project.

Last but not least, I also would like to give my gratitude thanks to my fellow friends who assisted and led me in completing of this project. This project would not have been completed without their help and support. All of the comments, criticism and advices from them are really useful for me. Thank you.

## **ABSTRACT**

In Malaysia, there is a lot of interesting places to visit especially for the foreign traveler. Usually, before start traveling, people always plan and looking for information about the place that they want to visit. With the Internet, every information can be easily obtained. There are many websites that provide information about popular places that can be visited. However, not all websites are updated, with some of them are lastly updated in years ago, and some of them do not provide a review section. This can lead to information inaccuracy and is questionable to the people who are desired to visit the place. It also causes a problem to the initial plan and it takes a lot of time to search through the internet to find good information about the location. By using the information about the attraction place that has been posted on social media such as Twitter, this application will extract the data, analyze the information and display the information to the end user to help the user in making decision, planning and getting a good understanding about the places. Sentiment analysis also has been conducted by this application in order to give a quick review and reliable information to the end user. In this project, R programming language is used since R language is one of the programming languages that are suitable in producing and conducting analysis, where it has multiple functions and packages that can be used to make the analysis easier. Twitter API allows the data to be extract and the R language help in processing those data which give a good analysis result about the place of attraction in the type of web application.