UNIVERSITI TEKNOLOGI MARA

CLUSTERING ON CS230'S ALUMNI DATA FOR EMPLOYMENT USING K-MEANS

SITI NORAINUL FARAHA BT MUHAMMAD ANUAR

BACHELOR OF COMPUTER SCIENCE (Hons.)

JULY 2015

UNIVERSITI TEKNOLOGI MARA

CLUSTERING ON CS230'S ALUMNI DATA FOR EMPLOYMENT USING K-MEANS

SITI NORAINUL FARAHA BT MUHAMMAD ANUAR

**Thesis submitted in fulfilment of the requirements for Bachelor of Computer Science (Hons)
Faculty of Computer and Mathematical Sciences**

JULY 2015

SUPERVISOR'S APPROVAL

CLUSTERING ON CS230'S ALUMNI DATA FOR EMPLOYMENT USING K-MEANS

By

SITI NORAINUL FARAHA BT MUHAMMAD ANUAR

2012458384

This report was prepared under the supervision of project supervisor, Hajar Izzati bt Mohd Ghazalli. It was submitted to the Faculty of Computer and Mathematical Sciences and was accepted in partial fulfilment of the requirement for the degree of Bachelor of Computer Science (Hons).

Approved by

…………………………..

HAJAR IZZATI BT MOHD GHAZALLI

PROJECT SUPERVISOR

JULY 30, 2015

STUDENT'S DECLARATION

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

…………………….....

SITI NORAINUL FARAHA BT MUHAMMAD ANUAR

2012458384

JULY 30, 2015

ABSTRACT

Computer Science (CS) is a subject or a course for some universities. In Universiti Teknologi Mara (UiTM), CS or Bachelor Degree of Computer Science is known as CS230. There is a problem where it is not confirmed if the CS230's alumni work in that field or not. Hence, a research that includes gathering information of big number of important data of alumni may be done to solve the problem. The proposed system will include data gathering, data extraction, data analysis by using K-Means technique and data visualization. The extraction will be done on the data of alumni that is obtained from Office of Industry Community and Alumni Network (ICAN). The data involved is Cumulative Grade Point Average (CGPA) and employment. Then, the result will be presented as visualization or graphic. This may shows the important results clearly.

Keywords : computer science, big data, analysis, clustering, K-Means, visualization