WQD7005 (Data Mining) – (30 %)

2019/2010 Semester 2

This assignment should not only have learned in theory, but also have experience predictive analyzing and prescriptive analyzing data with the main objective of solving real-world problems. There are four key skills needed to succeed in data science, which we refer to as **c**reating, **c**onnecting, **c**omputing and deploying.

It consists of **5 parts with 5 milestones:**

1. Acquisition of data in your familiar domain (Group : Max 2 students)

* Web crawling the real time data by using Python (WQD7004 Programming for Data Science)

Samples:

<https://drive.google.com/file/d/1CeguhXl6ZNIhC4YCQze90QwhAHEW4BGV/view>

https://github.com/GHEEJIAN/data-mining-project

Due Date : Week 3 (5 marks)

1. Management of data (Group : Max 2 students)

* Store data into hive data warehouse (WQD7007 Big Data Management)

or store data into data lake

Due Date : Week 5 (5 marks)

1. Processing of data (Group : Max 2 students)

* Accessing hive data warehouse or data lake using Python

Due Date : Week 8 (5 marks)

1. Interpretation of data & Communication of Insights of data (Individual)

(WQD 7003 DATA ANALYTICS, WQD7006 MACHINE LEARNING FOR DATA SCIENCE AND WQD7009 BIG DATA APPLICATIONS & ANALYTICS)

1. Deployment in web (Flask) and mobile app (Kivy) , Presentation and documentation (Max 2 students. One for flask deployment and another for Kivy moble app deployment.

Due Date: Week 13 (5 marks)