Course Syllabus

1.	Course Number:	2110446
2.	Course Credit:	3 (3-0-6)
3.	Course Title:	Data Science and Data Engineering
4.	Department	Computer Engineering
5.	Semester	Second Semester
6.	Academic Year	2016
7.	Instructor	Natawut Nupairoj, Ph.D. Veera Muangsin, Ph.D. Peerapon Vateekul, Ph.D.
8.	Condition	-
9.	Status	Elective
10.	Curriculum	Computer Engineering
11.	Degree	B.Eng.
12.	Hours/Week	3-Hour Lecture & Lab

13. Course Description

Data Science is the study of the discovery of knowledge from data. Being a data scientist requires an integrated skill set spanning mathematics, statistics, machine learning, databases and other branches of computer science along with a good understanding of the craft of problem. Data Engineering is the study of how to engineer or process data, i.e., data cleansing, data storing, etc.

This course will focus on analyzing very large amount of data (Big Data). There are four main parts in this course:

- Infrastructure Module
- Data Exploration & Statistical Analysis
- Data Analytics (Data Mining)
- Data visualization

14. Course Outline

14.1. Learning Objectives

- Describe what Data Science and Data Engineering are and the skill sets needed
- Understand the concept of Big Data Analytics
- To be able to install and setup infrastructure to process large scale of data
- To be able to explore and understand collected data
- To be able to conduct proper statistical hypothesis tests
- To be able to analyze Big Data by applying data mining techniques
- To be able to visualize data in relation to spatial and temporal points of views

14.2. Learning Contents (Room: ENG1-305/1)

Week	Date	Contents	Instructor
1	10-Jan-17	Introduction to Data Science and Data Enginering	Aj.Peerapon
2	17-Jan-17	Sport Week	no class
3	24-Jan-17	Hadoop (HDFS & MapReduce)	Aj.Natawut
4	31-Jan-17	Spark	Aj.Natawut
5	7-Feb-17	Data Ingestion	Aj. Natawut
6	14-Feb-17	Data Preprocessing	Aj. Peerapon
7	21-Feb-17	Statistical Analysis	Aj. Peerapon
8	28-Feb-17	Midterm (27/02 – 03/03)	TBD
9	7-Mar-17	Traditional Data Mining	Aj. Peerapon
10	14-Mar-17	CU Academic Expo 2017	no class
11	21-Mar-17	Data Mining on Spark	Aj. Peerapon
12	28-Mar-17	Case Studies in Data Science	Dr. Virote (Guest Speaker)
13	4-Apr-17	Data Visualization	Aj. Veera
14	11-Apr-17	Graph Analysis	Aj. Veera
15	18-Apr-17	Spatial Analysis	Aj. Veera
16	25-Apr-17	Project Presentation	
17	2-May-17	Final Exam	

14.3. Method: Lecture and Lab

14.4. Learning Media: PowerPoint presentation

Software:

- Hadoop, Spark, R, Rapid Miner, VirtualBox

14.5. Evaluation

Module1 Assignment	10%
Module2 Assignment	10%
Module3 Assignment	10%
Module4 Assignment	10%
Project1	10%
Project2	20%
Final	30% (Lab Test)

15. Reading List

15.1. Required Text: N/A

15.2. Electronic Media or Websites:

- CourseVille: "wongnai2016" (self-registration)
- Facebook: "2110446 Data Science and Data Engineering (2016/2)"