Part A: Course Information

1.	Code & Name of Course:	UECS3213 / UECS3453 Data Mining
2.	Year of Study (Programme):	Y2 (Bachelor of Science (Hons) Software Engineering), Y2 (Bachelor of Science (Hons) Applied Mathematics with Computing)
3.	Credit Value:	3
4.	Lecturer:	Dr. Simon Lau Boung Yew
5.	Tutor:	Dr. Ashanira binti Mat Deris
6.	Year and Trimester:	2019 January
7.	Synopsis:	Advances in data generation and collection are producing data sets of massive size in commerce and a variety of scientific disciplines. Data analysis techniques are becoming a necessity. This course presents fundamental concepts and algorithms of data mining techniques in detail thus providing the students with the background for the application of data mining to real problems. It introduces programming tools for students to apply their concepts to implement and evaluate data mining techniques to solve problems.
8.	Course Outcomes (CO):	Course Outcomes: upon completion of this course, a student shall be able to: 1. Identify the key technological foundations of data mining 2. Create programming solutions using data mining techniques for given problem 3. Evaluate performance of data mining solutions for a given problem 4. Construct a data mining project as a team 5. Recognize the importance of data mining techniques and its applications in the industry
9.	References:	Main References 1. J. Grus (2015). *Data Science from Scratch: First Principles with Python.* O'Reilly Media. 2. C. C. Aggarwal. (2015). *Data Mining: The Textbook*. Springer Additional References: 1. Witten, I.H, Franck, E, and Hall, M. A. (2011). Data Mining: Practical Machine Learning Tools and Techniques. (3rd ed.). Morgan Kaufmann. 2. Richert, W. and Coelho, L.P. (2013). Building Machine Learning Systems with Python. Packt Publishing. 3. Russel M.A. (2013). Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More. (2nd Ed). O'Reilly Media.

Part B: Methods of Assessment

No.	CO	Programme Outcomes (PO)	Domain & Taxonomy Level# *(e.g. A2 / C3 / P5)*	Delivery Methods *(e.g. Lecture, Tutorial, Practical, Case Study, Site Visit, Role Play, Class Discussion etc.)*	Assessment Methods & Mark Breakdown*					
					Assign 1	Assign 2	Assign 3 (Part 1)	Assign 3 (Part 2)	Assign 3 (Part 3)	Final Exam
					10%	10%	10%	10%	10%	50%
1	CO1	PO1	C2	L, P	x					x
2	CO2	PO2	P3	L, P		x	x			x
3	CO3	PO6	C5	L, P		x		x		x
4	CO4	PO5	A5	L, P					x	
5	CO5	PO6	C4	L, P	x					x

 # Domain – Affective (A), Cognitive (C), Psychomotor(P); Taxonomy Level – A (Level 1 – 5), C (Level 1 – 6), P (Level 1 – 5)

**Assessment methods and the mark breakdown must be tally with the teaching workload approved by Senate.*
**ONE type of Assessment Question (AQ) can be mapped to ONLY one CO.*

Part C: Continual Quality Improvement (CQI)

No.	Proposed Improvement Action (from Previous Course Report)	Implementation Action for Current Trimester **Must show in Part D with evidence (if necessary)*
1	Give more emphasis on evaluation aspect of classifiers in the lectures and practicals	Content on performance evaluation of classifiers to be added in the lectures and practicals

Part D: Weekly Plan (Lecture, Tutorial/Practical and Assessment)

Week	Lecture Topic	Tutorial / Practical Topic	Assessments / Specific Task
1	Topic 1: Introduction to Data Mining • Overview of Data Science • What is data mining? What is not data mining? • DIKW Knowledge Pyramid • Terminologies • Origin of data mining • Why data mining is important? • Data Mining Methods • Challenges of Data Mining • Data Mining Applications	Lab 1: Familiarization with Data Mining Tools: Python, R ● Anaconda Python, Jupyter Notebook ● R, RStudio Tutorial 1 - Introduction to Data Mining	
2	Topic 2a: Data • What is data? • Attribute and Attribute Value • Types of Data Sets • Structured vs Unstructured Data • Data Processing • Curse of Dimensionality • Similarity and Dissimilarity Measures	Lab 2: Introduction to Python Programming & Python Data Science Libraries: Pandas, NumPy, SciPy, matplotlib, scikit-learn Tutorial 2 - Data	Assignment 3 question release
3	Topic 2b: Data Exploration • What is data exploration? • Exploratory Data Analysis (EDA) • Summary Statistics • Visualization • Visualization Techniques • OLAP	Lab 3: Introduction to NumPy Library Tutorial 2 - Data	
4	Revision	Revision	-
5	Topic 3a: Classification (Part 1) • Supervised vs Unsupervised Learning • What is classification? • Examples of Classification Tasks • Classification Techniques • k-NN	Lab 4: Introduction to Pandas DataFrame Tutorial 3 - Classification	
6	Topic 3a: Classification (Part 1) • Decision Tree Classifier ■ Definition ■ Examples ■ Decision Tree Induction Algorithms ■ Measure of Node Impurity ■ Strength ■ Weakness	Lab 5: Introduction to Data Visualization Methods in Python using matplotlib Tutorial 3 - Classification	Assignment 1: Industry Talk - Summary of Talk - Opinions Tentative date: 18/2/2019

Week	Lecture Topic	Tutorial / Practical Topic	Assessments / Specific Task
7	Topic 3b: Classification (Part 2) • Naive Bayes Classifier ■ Bayes Classifiers ■ Bayes Theorem ■ The Naïve Bayes Model ■ Examples ■ Strength ■ Weakness • Support Vector Machine (SVM) ■ Definition ■ Discriminative vs Generative Classifiers ■ Examples	Lab 6: Implementing K-Nearest Neighbors in scikit-learn Tutorial 3 - Classification	Assignment 2 release - Programming in Python, numpy, pandas, matplotlib, scikit-learn
8	Topic 3b: Classification (Part 2) • Support Vector Machine (SVM) ■ Linear SVM ■ Non-linear SVM ■ Strength ■ Weakness • Ensemble Methods ■ Definition ■ Examples ■ Types of Ensemble Methods ■ Bagging ■ Boosting ■ Random Forest	Lab 7: Implementing Decision Tree using scikit-learn Tutorial 3 - Classification	
9	Topic 4: Regression Analysis • What is a model? • Deterministic Models • Probabilistic Models • Regression model • Correlation • What is regression? • Simple Linear Regression • Model Assessment • Multiple Linear Regression • Logistic Regression	Lab 8: Naïve Bayes implementation using scikit-learn Tutorial 4 - Regression Analysis	
10	Topic 5: Cluster Analysis and Anomaly Detection • What is Cluster Analysis? • Applications of Cluster Analysis • Types of Clustering • Hierarchical Clustering • Partitional Clustering • Types of Clusters	Lab 9: SVM implementation using scikit-learn Tutorial 5 - Cluster Analysis	Assignment 1 submission Assignment 2 submission Tentative date: 22/3/2019

Week	Lecture Topic	Tutorial / Practical Topic	Assessments / Specific Task
	• Clustering Algorithms		
11	Topic 5: Cluster Analysis and Anomaly Detection • Anomaly / Outlier Detection • Causes of Anomalies • Distinction Between Noise and Anomalies • Model-Based Anomaly Detection • Anomaly Detection Techniques ■ Proximity-based ■ Density-based ■ Pattern matching	Lab 10: k-Means clustering implementation using scikit-learn Tutorial 5 - Cluster Analysis	
12	Topic 6: Association Analysis • Association Rule Mining • Association Rule Mining Task • Mining Association Rules • Frequent Itemset Generation Strategies • Apriori principle	Tutorial 6 - Association Analysis	
13	Project Presentation		Project submission - Group Presentation - Programming - Report Tentative date: 12/4/2019
14	Project Presentation / Revision		

(Please ensure your course teaching plan covers all of the topics as per syllabus)

This Teaching Plan is:

Prepared by:	Moderated by:	Approved by:
_____ (Name: SIMON LAU BOUNG YEW) Course Coordinator	_____ (Name: TOO CHIAN WEN) Moderator	_____ (Name: MADHAVAN NAIR) Head of Department
Date: _____	Date: _____	Date: _____