

City University of Hong Kong
Course Syllabus

offered by Department of Management Sciences
with effect from Semester A 2020/21

Part I Course Overview

Course Title: Big Data Analytics

Course Code: MS4252

Course Duration: One Semester

Credit Units: 3

Level: B4

Arts and Humanities

Proposed Area:
(for GE courses only)

Study of Societies, Social and Business Organisations

Science and Technology

Medium of Instruction: English

Medium of Assessment: English

Prerequisites:
(Course Code and Title)

IS2240 Python Programming for Business,

Precursors:
(Course Code and Title) **MS3251 Analytics Using SAS,**
MS3252 Regression Analysis

Equivalent Courses:
(Course Code and Title) **Nil**

Exclusive Courses:
(Course Code and Title) **Nil**

Part II Course Details

1. Abstract

(A 150-word description about the course)

This course aims to:

- Provide fundamental concepts and techniques of using big data, social network analysis and advanced data mining techniques in the context of enterprise strategic decision making;
- Develop students' analytical ability to identify, formalize and solve the real world problem with big data, social network analytics and advanced data mining techniques;
- Develop students' hands-on experience of construction of big data, social network analysis, and advanced data mining techniques using professional software packages;
- Prepare students to demonstrate generic skills in interpersonal interaction, communication, working both individually and in teams.

2. Course Intended Learning Outcomes (CILOs)

(CILOs state what the student is expected to be able to do at the end of the course according to a given standard of performance.)

No.	CILOs	Weighting* (if applicable)	Discovery-enriched curriculum related learning outcomes (please tick where appropriate)		
			A1	A2	A3
1.	Understand and apply a wide range of big data, social network analytics and advanced data mining techniques, and recognize their characteristics, strengths and weaknesses	10%		✓	
2.	Evaluate a wide range of emerging and newly-adopted methodologies and technologies to facilitate the knowledge discovery.	40%		✓	✓
3.	Formulate and prepare statistical-oriented data for business solutions	20%		✓	✓
4.	Perform big data, social network analysis and advanced data mining techniques using professional software (e.g. SAS/DIS, SAS/EM, Python, R)	30%	✓	✓	✓
		100%			

* If weighting is assigned to CILOs, they should add up to 100%.

A1: Attitude

Develop an attitude of discovery/innovation/creativity, as demonstrated by students possessing a strong sense of curiosity, asking questions actively, challenging assumptions or engaging in inquiry together with teachers.

A2: Ability

Develop the ability/skill needed to discover/innovate/create, as demonstrated by students possessing critical thinking skills to assess ideas, acquiring research skills, synthesizing knowledge across disciplines or applying academic knowledge to self-life problems.

A3: Accomplishments

Demonstrate accomplishment of discovery/innovation/creativity through producing /constructing creative works/new artefacts, effective solutions to real-life problems or new processes.

3. Teaching and Learning Activities (TLAs)

(TLAs designed to facilitate students' achievement of the CILOs.)

TLA	Brief Description	CILO No.						Hours/week (if applicable)
		1	2	3	4			
Lecture	Concepts and techniques of big data and social network analytics	✓	✓	✓				

	using SAS or Python or R are explained;							
Tutorial	Hands-on practice to enhance their skills in big data and social network analytics using SAS or Python or R so that learning difficulties can be identified and tackled. Identify the business case issues regarding how to enhance the data retrieval and preparation for big data and social network analytics, performing different analytics technique to analyze the big data and generate different levels of statistical reporting.	✓	✓	✓	✓			

4. Assessment Tasks/Activities (ATs)

(ATs are designed to assess how well the students achieve the CILOs.)

Assessment Tasks/Activities	CILO No.						Weighting*	Remarks
	1	2	3	4				
Continuous Assessment: <u>60</u> %								
Mid-term Test The test is designed to assess students' big data knowledge and ability in applying big data techniques to solve the problems	✓	✓	✓				20%	
Group project Students work together in groups to complete a big data project. Assessments/Evaluations are carried out at various stages of the project, including the design of the project objectives, data modelling, project presentation and project report.	✓	✓	✓	✓			20%	
Individual presentation and Q&A Students have to provide a project presentation, answer the questions, and give recommendations to address the problems, justified by the data modelling, in relation to the project at hand.	✓	✓	✓	✓			20%	
Examination: <u>40</u> % (duration: 3 hours , if applicable)								
Examination The exam is designed to assess students' professional knowledge and ability in applying big data techniques to solve business problems	✓	✓	✓				40%	
							100%	

* The weightings should add up to 100%.

5. Assessment Rubrics

(Grading of student achievements is based on student performance in assessment tasks/activities with the following rubrics.)

Assessment Task	Criterion	Excellent (A+, A, A-)	Good (B+, B, B-)	Fair (C+, C, C-)	Marginal (D)	Failure (F)
1. Mid-term Test	1.1 ABILITY to EXPLAIN the key concepts and fundamental knowledge of big data and social network analysis	High	Significant	Moderate	Basic	Not even reaching marginal levels
2. Group project	2.1 CAPACITY for COLLABORATING with students to carry out problem-based activities based on real world problems.	High	Significant	Moderate	Basic	Not even reaching marginal levels
	2.2 ABILITY to EXPLAIN in DETAIL and with ACCURACY methods in analysing the relationship between business and sustainability solutions.	High	Significant	Moderate	Basic	Not even reaching marginal levels
	2.3 CAPACITY for SELF-DIRECTED LEARNING to find solutions to the problems and make recommendations	High	Significant	Moderate	Basic	Not even reaching marginal levels
3. Individual presentation and Q&A	3.1 ABILITY to UNDERSTAND the knowledge of big data and social network analysis	High	Significant	Moderate	Basic	Not even reaching marginal levels
4. Examination	4.1 ABILITY to EXPLAIN the key concepts and fundamental knowledge of big data and social network analysis	High	Significant	Moderate	Basic	Not even reaching marginal levels

Part III Other Information (more details can be provided separately in the teaching plan)

1. Keyword Syllabus

(An indication of the key topics of the course.)

1. Issue of big data analytics for business

Success factors for big data and social network analytics, The Analysis Process, Business Point of view in big data, Analytic Complexity;

2. Structured and unstructured big data management

Unstructured and structured big data management; Probabilistic matching for unstructured data; Map Reduce and Hadoop; Analytics Process Model;

3. Big data, Social Network Analytics, and Advanced Data Mining technique

3.1 Information Retrieval and Web Search:

Vector Space Model; Statistical Language Model; Singular Value Decomposition;

3.2 Social Network Analysis:

Network Measures and Structure; Graph Theory; Centrality (Degree, Eigenvector, etc); Prestige (Degree, Proximity, etc.); Network propagation and models (Epidemic, Threshold, Stochastic); Link Prediction; Path Analysis; Link Analysis; etc.

3.3 Statistical foundations of recommendation systems:

Content-based filtering; neighbourhood-based collaborative filtering; Model-based Collaborative Filtering;

3.4 Advanced Data Mining Tools and Technologies

Statistical foundations of Naïve Bayes classification; Support Vector Machine; Memory-based Reasoning; Ensemble and Random Forests;

3.5 Text Mining and Modelling:

Natural Language Processing; Wordcloud; Term Frequency - Inverse Document Frequency (TF-IDF); Maximum Entropy; Text-Clustering; Expectation-Maximization; Text Association Rule; Opinion Mining and Sentiment Analysis

3.6 Lifetime Modelling:

Lifetime value model; Survival Analysis; A/B testing;

2. Reading List

2.1 Compulsory Readings

(Compulsory readings can include books, book chapters, or journal/magazine articles. There are also collections of e-books, e-journals available from the CityU Library.)

1.	EMC Education Services, 2015. Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data. WILEY
2.	Jennifer Golbeck 2013. <i>Analyzing the Social Web</i> . Elsevier Inc.
3.	Goutam Chakraborty et al, 2013. <i>Text Mining and Analysis: Practice Methods, Examples, and case studies using SAS</i> . Cary, NC: SAS Institute Inc.
4.	David Ghan, 2016. <i>Introduction to SAS and Hadoop: Essentials Course Notes</i> . Cary, NC: SAS Institute Inc.
5.	Tan, P.N., Steinbach, M. and Kumar, V., 2014 <i>Introduction to Data Mining</i> . Pearson.
6.	Marc Huber & Michael Patetta, 2013 <i>Survival Analysis Using the Proportional Hazards Model Course Notes</i> , SAS Institute.

2.2 Additional Readings

(Additional references for students to learn to expand their knowledge about the subject.)

1.	Lin, Jimmy. 2010, <i>Data-Intensive Text Processing with MapReduce</i> , Morgan & Claypool Publishers.
2.	Svolba, Gerhard. 2006. <i>Data preparation for analytics using SAS</i> . Cary, NC: SAS Institute Inc.
3.	Michael Berry, & Gordon Linoff, 2004. <i>Data mining techniques: For marketing, sales, and customer support</i> , John Wiley & Sons.
4.	Madhavan Samir 2015. <i>Mastering Python for Data Science</i> . Packt Publishing.
5.	Marco Bonzanini 2016. <i>Mastering Social Media Mining with Python</i> . Packt Publishing.
6.	Bing Liu 2013, <i>Web Data Mining</i> . Springer