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: **Enterprise Data Mining**

Course Code: **MS4224**

Course Duration: **One Semester**

Credit Units: **3**

Level: **B4**

Proposed Area:
(for GE courses only)

☐ Arts and Humanities
☐ Study of Societies, Social and Business Organisations
☐ Science and Technology

Medium of Instruction: **English**

Medium of Assessment: **English**

Prerequisites:
(Course Code and Title)   **MS3252 Regression Analysis**

Precursors:
(Course Code and Title)   **CB2200 Business Statistics or equivalent**
MS3251 Analytics Using SAS or equivalent

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

Exclusive Courses:
(Course Code and Title)   **MS4424 Data Mining and Modelling**

Part II   Course Details

1.   Abstract
(A 150-word description about the course)

This course aims to:
- Provide fundamental concepts and techniques of using data mining in the context of business applications.
- Develop students' analytical ability to identify, formalize and solve the real world problem with business intelligence
- Prepare students for a position in managing business activities for data modelling in the commercial and government sectors in both local and global environment

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.	Discuss the history, nature role and relevant concepts of data mining.	10%		✓	
2.	Evaluate a wide range of emerging and newly-adopted methodologies and technologies to facilitate the knowledge discovery.	40%		✓	✓
3.	Critically discuss the pros and cons of various processes, methodologies in knowledge discovery.	20%		✓	
4.	Master the SAS Enterprise Miner software to perform data mining tasks	30%	✓	✓	✓
* If weighting is assigned to CILOs, they should add up to 100%.		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 knowledge of Data mining are explained	✓	✓	✓				
Tutorial	How to master SAS Enterprise Miner software are demonstrated. Students apply the learned concepts, techniques and SAS Enterpriser Miner skills on exercise questions	✓	✓	✓	✓			

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: _60___%								
Mid-term Test The test will assess the students' basic understanding of the material learnt in the first half of the course.	✓	✓	✓				20	
Group Project Students will work in groups; prepare on selected topics; students are required to formulate a problem and apply relevant data mining tools.	✓	✓	✓	✓			20	
Individual presentation and Q&A Students will deliver presentations on selected topics; Students are required to answer the data mining questions during the Q&A session	✓	✓	✓	✓			20	
Examination: __40__% (duration: 3 hours , if applicable)								
Written Examination The exam will assess the students' understanding of the material learnt in the course and their ability to apply subject related knowledge.	✓	✓	✓				40	
The weightings should add up to 100%.							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 data mining	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 for implementing the solutions	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 data mining	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.)

   Data Preparation for Analysis
   Summary Statistics; Data Visualization; OLAP and Multidimensional Data Analysis; Data Warehouse; Star Schema; Analysis Subject; Single view of the customer; Subject-oriented data; Data preparation for data mining; Dimension reduction;

   Concepts of Data mining - SEMMA processes of Data mining; Supervised and unsupervised Statistical Learning;

   Predictive Modelling - Logistic Regression; Decision Tree; Artificial Intelligence; Neural Network; Ordinal and Multinomial Logit Model; Tree-based Regression model; Learning Algorithm;

   Pattern Recognition/Customer Behaviour Recognition – Similarity Measures; Clustering Analysis, RFM Analysis; Association Analysis; Market Basket Analysis; Apriori Algorithm; Multi-level association rules; Sequential Pattern Mining;

   Model validation – Goodness of fit; Model tuning; Model assessment and implementation
   Business Analytics/Intelligence - Knowledge Discovery, Credit Scoring, Credit model development; Reject Inference;

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.	Tan, P.N., Steinbach, M. and Kumar, V., *Introduction to Data Mining*. Pearson, 2014.
2.	Thomas, L., Crook, J. and Edelman, D., Credit Scoring and its Applications, Second Edition, 2017. SIAM
3.	Forster Provost and Tom Fawcett, 2013. Data Science for Business. O'Reilly Media, Inc.

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

1.	Paolo Giudici, Applied Data Mining: *Statistical Methods for Business and Industry*, John Wiley & Sons, 2003
2.	Matignon, Randall. *Data Mining Using SAS Enterprise Miner*. Second Edition. Wiley, 2007
3.	Cerrito, Patricia, *Introduction to Data Mining Using SAS Enterprise Miner*. SAS Institute, 2007
4.	Michael Berry, & Gordon Linoff, *Data mining techniques: For marketing, sales, and customer support*, John Wiley & Sons, 2004
5.	Patricia B. Cerrito, *Introduction to Data Mining Using SAS Enterprise Miner*, SAS Institute, 2006.
6.	Michael Berry, & Gordon Linoff, *Mastering Data Mining, John Wiley & Sons, 2000.* *Jiawei Han, & Micheline Kamber, Data mining: Concepts and techniques*, Morgan Kaufmann Pub., 2000
7.	Bart Baesens, *Analytics in a BIG DATA WORLD – The essential guide to data science and its applications.* WILEY, 2014
8.	Bart Baesens, *Credit Risk Modeling Using SAS,* SAS Institute, 2011