July 2019

The Hong Kong Polytechnic University
Hong Kong Community College

Subject Description Form

Subject Code	SEHH3164
Subject Title	Introduction to Data Mining
Level	3
Credit Value	3
Medium of Instruction	English
Pre-requisite / Co-requisite/ Exclusion	Nil
Objectives	This subject aims to equip students with fundamental data mining concepts and methodologies. It provides students with the knowledge of the current methodologies and technologies in data mining. It also equips students with techniques of using data mining in the context of business applications. It helps students develop an understanding of the strengths and limitations of popular data mining techniques and to be able to identify promising applications of data mining.
Intended Learning Outcomes	Upon completion of the subject, students will be able to: (a) describe data mining process and apply the principles and practices of different stages of data mining process; (b) apply different data mining techniques to achieve data mining goals; (c) formulate procedures with the associated software to perform data mining tasks; (d) discuss future development of data mining technologies.
Subject Synopsis/ Indicative Syllabus	**Overview of data mining** Data Mining Process; Fundamental Tasks, Issues and Paradigms of Learning Models from Data. **Data mining techniques** Data Preparation; Clustering; Logistic Regression; Decision Tree; Neural Network; Association Analysis. **Data mining software** Use of the Associated Tools to Perform Data Mining Tasks: such as SAS programming, SAS Enterprise Miner Software. **Applications** Data Mining and Society; Data Warehousing; Future Directions.

Teaching/Learning Methodology

Lectures will emphasise data mining concepts and techniques that are commonly used in data mining applications, and will demonstrate how to master data mining tools. Lectures are designed to equip students with knowledge of the real-life/industrial data mining applications and skills in performing data mining tasks. Group discussions and activities may be arranged to stimulate students' interest or their awareness of the practical implications of some concepts.

Tutorials will provide students with the opportunity to deepen and consolidate their understanding of the concepts taught in lectures. They will also help students to apply the theories to different stages of the data mining process and enhance their knowledge and skills in operating data mining tools. The activities in tutorials may include exercises, student presentations, and discussions of problems.

Assessment Methods in Alignment with Intended Learning Outcomes

A variety of assessment tools will be used to develop and assess students' achievement of the subject intended learning outcomes.

Specific assessment methods/tasks	% weighting	Intended subject learning outcomes to be assessed			
		a	b	c	d
Continuous Assessment*	**50**				
▪ Test	20	✓	✓		✓
▪ Assignment 1	15	✓	✓	✓	
▪ Assignment 2	15	✓	✓	✓	✓
Final Examination	**50**	✓	✓	✓	✓
Total	**100**				

**Continuous assessment items and/or weighting may be adjusted by the subject team subject to the approval of the College Programme Committee.*

To pass this subject, students are required to obtain Grade D or above in **both** the Continuous Assessment and Final Examination.

Student Study Effort Expected

Class contact	**Hours**
▪ Lecture	26
▪ Tutorial	13
Other student study effort	
▪ Self-study	52
▪ Continuous Assessment	39
Total student study effort	**130**

Reading List and References	**Recommended Textbook** Tan, P.N., Steinbach, M. and Kumar, V. (2019), *Introduction to Data Mining*. (2nd ed.), Pearson. **References** Bramer, M. (2016), *Principles of Data Mining*. (3rd ed.), Springer. Han, J., Kamber, M. and Pei, P. (2011), *Data Mining: Concepts and Techniques*. (3rd ed.) (The Morgan Kaufmann Series in Data Management Systems), Morgan Kaufmann. Kovalerchuk, B. and Vityaev, E. (2013), *Data Mining in Finance: Advances in Relational and Hybrid Methods*. Springer. Linoff, G.S. and Berry M.J.A. (2011), *Data mining techniques: For marketing, sales, and customer support*. (3rd ed.), John Wiley & Sons. Liu, B. (2011), *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data*. (2nd ed.), Springer. Witten, I.H., Frank, E. and Hall, M.A. (2016), *Data Mining: Practical Machine Learning Tools and Techniques*. (4th ed.) (The Morgan Kaufmann Series in Data Management Systems), Morgan Kaufmann.