澳門理工大學
Universidade Politécnica de Macau
Macao Polytechnic University

FACULTY OF APPLIED SCIENCES

BACHELOR OF SCIENCE IN COMPUTING

LEARNING MODULE OUTLINE

Academic Year	2024/2025	Semester	2
Module Code	COMP323		
Learning Module	Data Warehousing and Data Mining		
Pre-requisite(s)	COMP211 Database Design		
Medium of Instruction	English		
Credits	3	Contact Hours	45 hrs
Instructor	Dr. Jie Zhang	Email	jpeter.zhang@mpu.edu.mo
Office	Room N56, 5/F, WUI CHI Building, Main Campus	Office Phone	8599-6837

MODULE DESCRIPTION

This comprehensive learning module delves into the fundamental principles and practical applications of data warehousing and data mining. It equips students with a deep understanding of designing, implementing, and leveraging data mining techniques within an enterprise framework. Through a structured approach, students will explore various facets including data visualization, preprocessing, and a wide array of data mining algorithms. By the end of this module, students will have acquired the essential skills to proficiently navigate the realms of data warehousing and data mining, enabling them to extract valuable insights and drive informed decision-making processes within organizational settings.

MODULE INTENDED LEARNING OUTCOMES (ILOS)

On completion of this learning module, students will be able to:

M1.	Understand the fundamental concepts of data warehousing and data mining, including the architectures involved, and gain proficiency in using Python for data warehousing and analysis. (EA1p, EP2p)
M2.	Master fundamental data warehousing and preprocessing techniques using a variety of Python packages. (EA1p, D1p, D2p)
M3.	Utilize advanced data analysis and visualization techniques to analyze multidimensional data effectively (EA1p, D1p, D2p, EP2p)
M4.	Apply a range of data mining algorithms proficiently to uncover intricate patterns within data warehouses. (EA1p, D2p)

These ILOs aims to enable students to attain the following Programme Intended Learning Outcomes (PILOs):

PILOs		**M1**	**M2**	**M3**	**M4**
P1.	Select and apply proven methods, tools and techniques to the effective and efficient implementation of information systems;			✓	✓
P2.	Evaluate computer systems in a local area network, and understand the				

	additional requirements for connection to other networks through wide area networks;				
P3.	Be competent in system development in the Internet and the web platform;				
P4.	Work independently to design and implement a relational database, with an emphasis on how to organise, maintain and retrieve information from a DBMS;	✓	✓	✓	
P5.	Acquire essential knowledge in specific fields of computing disciplines including multimedia, security and artificial intelligence;				
P6.	Acquire the perceptive skills needed to understand information presented in the form of UML diagram, flow chart or other industry standard formats;	✓	✓	✓	
P7.	Understand the need for and use of the necessary mathematical techniques;				
P8.	Work independently to develop an understanding of, and the knowledge and skills associated with the general support of computer systems and networks;				
P9.	Work as an effective member of a team in the analysis, design and development of software systems;			✓	✓
P10.	Use project planning and management techniques in systems development;				
P11.	Understand the fundamental and operational issues of computer systems in business environments;	✓	✓	✓	✓
P12.	Equip with adequate written, oral communication and interpersonal skills;				
P13.	Build the capacity and desire for lifelong learning and to learn advanced and emerging technologies on one's own;				
P14.	(For Enterprise Information Systems specialization) Gain an in-depth understanding of the information technology related to enterprise information systems, with an emphasis on development of such systems to support business processes;	✓	✓	✓	✓
P15.	(For Gaming Technology specialization) Acquire the general and advanced knowledge of current technologies and operating environment in the gaming industry;				
P16.	(For Computer Education specialization) Acquire the general and practical knowledge of computer education and its practicing environment in secondary education.				

MODULE SCHEDULE, COVERAGE AND STUDY LOAD

Week	Content Coverage	Contact Hours
1-2	**1. Overview of Data Warehousing and Data Mining**	6
	1.1. Overview of Data Warehousing	
	1.2. Overview of Data Mining	
	1.3. Architectures	
	1.4. Introduction to Python for Data Warehousing and Analysis	

3-6	**2. Data Visualization**	12
	2.1. Overview of Data Visualization	
	2.2. Data Visualization with Matplotlib	
	2.3. Data Visualization with Seaborn	
	2.4. Data Visualization with Plotnine	
7-10	**3. Data Warehousing and Preprocessing**	12
	3.1. Data Structures in Numpy: Array Operations and Manipulations	
	3.2. Data Structures in Pandas: Series, DataFrame	
	3.3. Data Wrangling, Cleaning and Transformation	
	3.4. Data Normalization and Standardization	
	3.5. ETL: Extraction, Transform, and Load	
11-15	**4. Data Mining Algorithms**	15
	4.1. Clustering Algorithms	
	4.2. Dimensionality Reduction Algorithms	
	4.3. Association Rule Mining Algorithms	
	4.4 Classification Algorithms	
	4.5 Regression Algorithms	

TEACHING AND LEARNING ACTIVITIES

In this learning module, students will work towards attaining the ILOs through the following teaching and learning activities:

Teaching and Learning Activities	**M1**	**M2**	**M3**	**M4**
T1. Lectures	✓	✓	✓	✓
T2. Tutorials		✓	✓	✓

ATTENDANCE

Attendance requirements are governed by the Academic Regulations Governing Bachelor's Degree Programmes of the Macao Polytechnic University. Students who do not meet the attendance requirements for the learning module shall be awarded an 'F' grade.

ASSESSMENT

In this learning module, students are required to complete the following assessment activities:

Assessment Activities	Weighting (%)	AHEP3 LOs	ILOs to be Assessed
A1. Assignment	20	D1p, D2p, EP2p	M1, M2
A2. Assignment	20	D1p, D2p, EP2p	M3, M4
A3. Test	20	EA1p, D1p, D2p, EP2p	M1, M2, M3
A4. Examination	40	EA1p, D2p	M1, M2, M3, M4

The assessment will be conducted following the University's Assessment Strategy (see www.mpu.edu.mo/teaching_learning/en/assessment_strategy.php). Passing this learning module indicates that students will have attained the ILOs of this learning module and thus acquired its credits.

Students with an overall score of less than 35 in the coursework must take the re-sit examination even if the overall score for the module is 50 or above.

Students with a score of less than 35 in the final examination must take the re-sit examination even if the overall score for the module is 50 or above.

Students with an overall final grade of less than 35 are NOT allowed to take the re-sit examination.

REQUIRED READINGS

There is no official text for this module. Module notes are distributed in the module.

Textbook(s)

1. Jiawei Han, Micheline Kamber, Jian Pei. (2011) Data Mining: Concepts and Techniques (The Morgan Kaufmann Series in Data Management Systems) 3rd Edition. Morgan Kaufmann.
2. Trevor Hastie, Robert Tibshirani, Jerome Friedman. (2011) The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition (Springer Series in Statistics) 2nd Edition. Springer.
3. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani. (2023) An Introduction to Statistical Learning: with Applications in Python (Springer Texts in Statistics) 2023rd Edition. Springer. (https://github.com/qx0731/Sharing_ISL_python)

STUDENT FEEDBACK

At the end of every semester, students are invited to provide feedback on the learning module and the teaching arrangement through questionnaires. Your feedback is valuable for instructors to enhance the module and its delivery for future students. The instructor and programme coordinators will consider all feedback and respond with actions formally in the annual programme review.

ACADEMIC INTEGRITY

The Macao Polytechnic University requires students to have full commitment to academic integrity when engaging in research and academic activities. Violations of academic integrity, which include but are not limited to plagiarism, collusion, fabrication or falsification, repeated use of assignments and cheating in examinations, are considered as serious academic offenses and may lead to disciplinary actions. Students should read the relevant regulations and guidelines in the Student Handbook which is distributed upon the admission into the University, a copy of which can also be found at www.mpu.edu.mo/student_handbook/.