MINISTRY OF EDUCATION AND TRAINING
HO CHI MINH CITY OPEN UNIVERSITY

COURSE SYLLABUS

I. GENERAL INFORMATION

1. Course title in Vietnamese: KHAI PHÁ DỮ LIỆU

2. Course title in English: DATA MINING

3. Knowledge / skill categorization:

☐General knowledge ☒ Specialized knowledge

☐Basic knowledge ☐Supplementary knowledge

☐Professional knowledge ☐Graduate project / thesis

4. Number of credits

Total	Theory	Practice	Self-study
3	2	1	3 (2,1,5)

5. In charge of course

a) Faculty / Department / Sub-Department: Information Technology

b) Faculty:

c) Contact email:

d) Address: Room 604. Faculty of Information Technology. 35-37 Hồ Hảo Hớn Street, District 1. HCM City

II. COURSE INFORMATION

1. Course description

Data mining is the process of discovering new knowledge and useful knowledge in the form of potential in existing data sources. The purpose of data mining is to understand raw data by the data processing algorithms. These algorithms are collected from many areas of computer science such as artificial intelligence, machine learning, statistics, and database systems.

2. Conditional courses

#	Conditional courses	Course code
1.	Prerequisites	
	None	

2.	Previous courses	
	Database	ITEC2502
3.	Parallel courses	
	None	

3. Course goals

Course goals	Description	Related Program Learning Outcomes
CO1 (Knowledge)	- Understand and apply the concept of data mining and its advantages and drawbacks. - Understand and apply several common data mining algorithms. - Understand how to exploit knowledge and information from data. - Understand and apply data mining in computer science as well as other related application. - Understand the current research of data mining.	PLO6.11 PLO8.1
CO2 (Skills)	- Applying and installing popular data mining algorithms. - Applying data mining knowledge and developing practical applications. - Ability to further research the theory of data mining algorithms.	PLO6.11 PLO7.2 PLO7.3 PLO8.1
CO3 (Attitude)	- Self-study spirit, self-research. - Constantly updating new research results in the field of data mining.	PLO12.2 PLO12.3

4. Course output standards:

Course goals	Course output standard	Description of course output standards
CO1	CLO1.1	- Understand data mining overview, popular data mining algorithms.
	CLO1.2	- Understand the importance of data mining.
	CLO1.3	- Understand current research and application directions in data mining.
CO2	CLO2.1	- Installed data mining algorithms.
	CLO2.2	- Applying data mining knowledge to develop practical applications.

	CLO2.3	- Ability to further research the theory of data mining algorithms.
CO3	CLO3.1	- Improve self-study, self-research.

Integrated matrix between the course's output standards and the standards of the curriculum.

CLOs	PLO6.11	PLO7.2	PLO7.3	PLO8.1	PLO12.2
1.1	4	3		3	
1.2	4	3		3	
1.3	4	3		3	
2.1	5	4		3	
2.2	5	4		3	
2.3	5	5	4	4	
3.1					
3.2					4

1: Do not meet   *4: Meet more*
2: Merely meet   *5: Meet a lot*
3: Average meet

5.   Reference

a)   *Textbooks*

[1] Charu C. Aggarwal. Data Mining: The Textbook. Springer. 2015. [49465].

[2] Mohammed J. Zaki, Wagner Meira Jr. Data Mining and Analysis Fundamental Concepts and Algorithms. Cambridge University Press. 2014. [49473].

b)   *References (list up to 3 references)*

[3] Jure Leskovec, Anand Rajaraman, Jeffrey David Ullman. Mining of Massive Datasets (2nd edition). Cambridge University Press. 2014. [49466].

c)   *Software*

1) Python 3.7.4

2) Pycharm Community

6. Course evaluation / assessment

Components	Assessment	Point of Time	Course output standard	Rate %
(1)	(2)	(3)	(4)	
A1. Process evaluation	A.1.1. Assessment in class / attendance		CLO1.2, CLO1.2, CLO1.3	20%
	Total: 01			20%
A2. Mid-term review	A2.1. Great exercise		CLO1.2, CLO1.2, CLO1.3	30%
	Total: 01			30%
A3. End-of-term assessment A3.1. Exam written on paper. Total: 01	A3. End-of-term assessment A3.1. Exam written on paper		CLO1.2, CLO1.2, CLO1.3, CLO2.2, CLO2.3	50%
	Total: 01			50%
Total: 02				100%

7. Rubrics mid-term review (50%)

Components	Format	Marks	Course Learning Outcomes	Excellent	Good	Average	Fail
Question 1- Question 5	Multi Choice	5	CLO1.1, CLO1.2	5	4	[2-3]	<2
Question 6	Writing	1	CLO1.1	1	0.75	0.5	0
Question 7	Writing	1	CLO1.2	1	0.75	0.5	0
Question 8	Solve Problem	3	CLO1.1, CLO1.2, CLO1.3	3	[2.5-3)	(2.5-2]	<2
			Rank	(9-10]	(6,9)	[4,6]	<4

8. Teaching plans

Week/session	Content	CO	Teaching and learning activities				Reviews / Assessment	Main documents and references
			Offline					
			Theory		Lab			
			Activities	Số tiết	Activities	Số tiết		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1.Week 1/ Theory Session 1	Chapter 1: Overview of data mining 1.1 Overview of data mining	PO1.1 PO1.2 PO1.3	Lecturers: + Introduction to detailed outline.	4.5	Data collection techniques Lecturers:	4.5	A1.1 A2.1 A3.1	[1]

			Lecturers: + Lecture + Ask questions, exercises. + Emphasize the main points. + Outline the requirements for the next class. Student: + Learning in class: listening to lectures, answering questions, solving the posed exercises, taking notes. + Study at home: watch lectures, summarize key knowledge, learn related knowledge. + On LMS system: answer theoretical multiple-choice questions, participate in discussion on forums.		Demo program or give a brief tutorial on how to do the lessons. Student: + Reworked the trainer sample demo / tutorial. +Do DIY exercises based on sample demos and learned theory.			
1.2 Data form and pattern need mining focus. 1.3 Methods and applications for the data mining problem 1.4 Influencing factors.								

2.Week 2/ Theory Session 2	Chapter 2: Pretreatment data
	2.1 Survey data
	2.2 Statistical
	2.3 Visualization
	2.4 Measure the similarity.
	2.5 Data cleaning
	  2.5.1 Handling of missing data
	  2.5.2 Incorrect and inconsistent data handling
	  2.5.3 Standardize and scale up.
		PO2.1
		PO2.2
		PO2.3
			Lecturers:
			+ Introduction to detailed outline.
			+ Lecture
			+ Ask questions, exercises.
			+ Emphasize the main points.
			+ Outline the requirements for the next class.
			
			Student:
				4.5	Lecturers:
					Demo program or give a brief tutorial on how to do the lessons.
					
					Student:
					+ Reworked the trainer sample demo / tutorial.
					+ Do DIY exercises based on sample demos and
						4.5	A1.1
							A2.1
							A3.1
								[1]

	2.6 Data integration and reduction 2.6.1 Sampling 2.6.2 Select a characteristic subset. 2.6.3 Reduced direction with spindle 2.6.4 Reduce dimension with style conversions. 2.7 Data transformation and discrete		+ Learning in class: listening to lectures, answering questions, solving the posed exercises, taking notes. + Study at home: watch lectures, summarize key knowledge, learn related knowledge. + On LMS system: answer theoretical multiple-choice questions, participate in discussion on forums.		learned theory.			
3. Week 3/ Theory Session 3	Chapter 3: Exploiting common patterns. 3.1 Harvest the set. 3.1.1 Common practice and association rules 3.1.2 Mining algorithms set. 3.1.3 General association rules 3.2 Exploit the sequence. 3.2.1 Popular order 3.2.2 Exploit common sequences. 3.2.3 Extraction of child chains through suffix trees 3.3 Mining Graph patterns 3.3.1 Support and isomorphic	PO2.1 PO2.2 PO2.3	Lecturers: + Introduction to detailed outline. + Lecture + Ask questions, exercises. + Emphasize the main points. + Outline the requirements for the next class. Student: + Learning in class: listening to lectures, answering questions, solving the posed exercises, taking notes. + Study at home: watch lectures, summarize key knowledge,	4.5	Data classification algorithms Lecturers: Demo program or give a brief tutorial on how to do the lessons. Student: + Reworked the trainer sample demo / tutorial. + Do DIY exercises based on sample demos and learned theory.	4.5	A1.1 A2.1 A3.1	[1][3]

			learn related knowledge. + On LMS system: answer theoretical multiple-choice questions, participate in discussion on forums.					

4. Week 4/ Theory Session 4	Chapter 4: Data Classification
	4.1 Mathematical background
	4.2 Decision tree
	  4.2.1 Decision trees
	  4.2.2 Criteria for separation
	  4.2.3 Pruning trees and stopping criteria.
	  4.3 The Bayes Algorithm
		PO2.1
		PO2.2
		PO2.3
		PO3.1
		PO3.2
			Lecturers:
			+ Introduction to detailed outline.
			+ Lecture
			+ Ask questions, exercises.
			+ Emphasize the main points.
			+ Outline the requirements for the next class.
			
			Student:
			+ Learning in class: listening to lectures, answering questions, solving the posed exercises, taking notes.
			+ Study at home: watch lectures, summarize key knowledge, learn related knowledge.
			+ On LMS system: answer theoretical multiple-choice questions, participate in discussion on forums.
				4.5	Data classification algorithms (cont.)
					
					Lecturers:
					Demo program or give a brief tutorial on how to do the lessons.
					
					Student:
					+ Reworked the trainer sample demo / tutorial.
					+ Do DIY exercises based on sample demos and learned theory.
						4.5	A3.1	[1][2]

| 5.Week 5/ Theory Session 5 | Chapter 4: Data Classification (cont.) 4.4 Categorize data based on rules. 4.4.1 Initialize rules from decision trees. 4.4.2 Sequential coverage algorithms 4.4.3 Prune the laws. 4.4.4 Combined categories 4.5 Precision techniques. | PO2.1 PO2.2 PO2.3 PO3.1 PO3.2 | Lecturers: + Introduction to detailed outline. + Lecture + Ask questions, exercises. + Emphasize the main points. + Outline the requirements for the next class. Student: + Learning in class: listening to lectures, answering questions, solving the posed exercises, taking notes. + Study at home: watch lectures, summarize key knowledge, learn related knowledge. + On LMS system: answer theoretical multiple-choice questions, participate in discussion on forums. | 4.5 | Data clustering algorithms Lecturers: Demo program or give a brief tutorial on how to do the lessons. Student: + Reworked the trainer sample demo / tutorial. + Do DIY exercises based on sample demos and learned theory. | 4.5 | A3.1 | [1][2] |
| 6.Week 6/ Theory Session 6 | Chapter 5 Grouping of data clusters. 5.1 Mathematical background 5.2 K-mean partition 5.3 Group clusters based on a nested structure | PO2.1 PO2.2 PO2.3 PO3.1 PO3.2 | Lecturers: + Introduction to detailed outline. + Lecture + Ask questions, exercises. + Emphasize the main points. + Outline the requirements | 4.5 | Data clustering algorithms (cont.) Lecturers: Demo program or give a brief tutorial on how to do the lessons. | 4.5 | A3.1 | [1][2] |

			for the next class. Student: + Learning in class: listening to lectures, answering questions, solving the posed exercises, taking notes. + Study at home: watch lectures, summarize key knowledge, learn related knowledge. + On LMS system: answer theoretical multiple-choice questions, participate in discussion on forums.		Student: + Reworked the trainer sample demo / tutorial. + Do DIY exercises based on sample demos and learned theory.			
7.Week 7/ Theory Session 7	Chapter 5 Grouping data clusters (cont.) 5.4 Gather clusters based on mesh. 5.5 Evaluate the efficiency of clustering	PO2.1 PO2.2 PO2.3 PO3.1 PO3.2	Lecturers: + Introduction to detailed outline. + Lecture + Ask questions, exercises. + Emphasize the main points. + Outline the requirements for the next class. Student: + Learning in class: listening to lectures, answering questions, solving the posed exercises, taking notes.	3.0		3.0	A3.1	[1][2]

| | | | + Study at home: watch lectures, summarize key knowledge, learn related knowledge. | | | | | |
| | | | + On LMS system: answer theoretical multiple-choice questions, participate in discussion on forums. | | | | | |

9. Regulations of the course

- Students participate fully in theory and practice.

- Students must submit large assignments through LMS and participate in question-and-answer sessions.

DEAN OF FACULTY
(Sign and specify full name)

EDITOR
(Sign and specify full name)