

UNDERGRADUATE PROGRAM

Training level: undergraduate
Program Code: 7480201
Specialization:

Major: Information Technology

COURSE SYLLABUS

1. General information

1.1	Course title (in Vietnamese)	Khai phá dữ liệu
1.2	Course title (in English):	Data mining
1.2	Course Code:	31231330
1.3	Credits:	3
1.4	Time allocation:	
-	Theory:	30
-	Type 1 Practice:	15
-	Type 2 Practice:	
-	Self-learning:	120
1.5	Lecturers in charge of the course:	
-	Main Lecturer:	Dr. Nguyen Thi Ngoc Anh Email: ntnanh@ued.udn.vn
-	Co-lecturers:	Dr. Nguyen Tran Quoc Vinh Email: ntqvinh@ued.udn.vn
-	Faculty in charge:	Information System
1.6	Conditions for taking part in the course:	
-	Prerequisite course(s):	
-	Previous course(s):	
-	Parallel course(s):	
1.7	Course Type:	<input checked="" type="checkbox"/> Compulsory <input type="checkbox"/> Compulsory elective <input type="checkbox"/> Free elective
1.8	Knowledge Cluster	<input type="checkbox"/> General Education <input type="checkbox"/> Fundamental <input checked="" type="checkbox"/> Specialized <input type="checkbox"/> Pedagogical skills <input type="checkbox"/> Internship and Thesis

2. Course description

The course provides the principles, concepts, fundamental methodologies in data mining. It consists of three main parts: (1) basic concepts, (2) fundamental tasks and techniques in data mining, (3) real applications and modern-techniques in data mining. Specifically, the first part will provide students all about the overall course, concepts, and data mining process. This part also covers data collection, organization and preprocessing. The second part will focus on the principles, methodologies and basic algorithms in classification, clustering, and association rules. The last part will give students the opportunity to research and apply common data mining techniques to solve real-world problems.

3. Course Objectives

3.1. General Course Objective

The course introduces the principles, concepts, fundamental methodologies in data warehouse and data mining. The course also provides solid background and advance in data mining. Student will be able to master the techniques in data mining and apply them to solve real-world topics.

3.2. Specific Course Objectives (COs)

- **CO1:** Providing to students the principles, concepts, fundamental methodologies in data warehouse and data mining.

- **CO2:** Using and applying common techniques in data mining to solve real problems in practice.

- **CO3:** Be able to analyze, discuss and compare among the methodologies on real datasets by choosing the proper algorithms or combine them

4. Course Learning Outcomes (CLOs)

After completing the course, students will be able to:

Course Learning Outcomes (CLOs)	Contents	PIs	Mức độ của PIs
CLO1	Define clearly the definitions, tasks and process in data mining. Also, present clearly related researches and real applications in data mining.	PI6.2, PI6.4	R, M
CL02	Present clearly techniques related to data formation and processing, basic methodologies including classification, clustering, and association rules.	PI6.3, PI6.4	M
CLO3	Analysis basic techniques in data mining to solve some problems.	PI6.4	M
CLO4	Apply data mining knowledge, techniques and tools for real-datasets in practice.	PI6.3, PI6.4	MA

5. Mapping matrix between course learning outcomes (CLOs) and program learning outcomes (PLOs):

The level of contribution and support of CLOs to the PLO is specifically determined as follows:

Fill in one of the levels I, R, M or leave blank (if there is no connection) and enter A in the corresponding box

Course Learning Outcomes (CLOs)			Program Learning Outcomes (PLOs)									
	PLO 1	PLO 2	PLO 3	PLO 4	PLO 5	PLO6			PLO 7	PLO 8	PLO 9	PLO10
						PI6.2	PI6.3	PI6.4				
CLO1						R		M				
CLO2							M	M				
CLO3								M				
CLO4							M,A	M,A				

Note:

- *I (Introduced) – CLO supports the PLO and at the level of introduction/beginners*
- *R (Reinforced) – CLO supports the PLO and at the level higher than the level of introduction/beginners, giving chances for practice, experiment and fieldtrips*
- *M (Mastery) – CLO strongly supports learners in mastering/ grasping or achieving PLO/ PI. If the learner successfully completes this CLO, it is considered that the learner has mastered a program indicator (PI) of the PLO or even mastered the entire PLO.*
- *A (Assessed) – Critical CLO (maximum support for PLO achievement) needs to be collected to measure how well learners achieve the PLO*

6. Course assessment

6.1. Methods and forms of testing - assessment of the course

Student learning outcomes are assessed by the following components: formative assessment, mid-term assessment, end-of-term assessment, and other assessment activities.

Assessment Component	Assessment forms	Assessment methods	Rubric's Criterion	Assessment forms Weight (%)	Assessment components Weight (%)	CLO
A1. Formative Assessment	A1.1 Attendance and short exercise in class	P1.1. Attendance (random 3 times), hand-on/interview	RH1.1	50%	20%	CLO 1, 2, 3, 4
	A1.2 Personal assignments	P1.2 Personal report	RH1.2	50%		CLO 1, 2, 3, 4
A2. Mid-term assessment	A2.1 Mid-term exam	P2.1 Writing assessments	RH2	100%	30%	CLO 1, 2
A3. End of term assessment	A3.1 Final exam	P3.1 Report and Representaion	RH3	100%	50%	CLO 4

Note (11):

- All Assessment Components, Assessment forms, Assessment Methods, Rubric, Assessment Weights, Component Weights should be coded and defined so that the assessment is accurate, reliable and fair;

- Note: W3 = 50% (Weight 0.5);

- Upon the performance of a specific assessment for one or several specific CLOs, attention should be paid to:

a) Ensure that teaching and learning activities for the respective knowledge/skills have been organized

b) The assessment must have a designed question/component, and clearly state what CLO it is being used for.

c) The assessment results are not only used to calculate the GPA but are also used to analyze the CLO achievement of each student and of the entire class or the course (if there are many classes taught in parallel).

No	Contents	Level 1	Level 2	Level 3
R1. On-going Assessment				
1	Data warehouse and Data mining	Concepts Process Identify the goals	- Analyze OLTP and OLAP - Explain the	Establish the relationship between data

	(2 points)	and primary tasks of data mining process	influence of data quality on a data-mining process. -	warehousing and data mining - Explain of big data and data science
2	Preparing the data	- Analyze basic representations and characteristics of raw and large data sets	- Apply different normalization techniques on numerical attributes. - Implement some data preprocessing techniques.	Recognize different techniques for data preparation, including attribute transformation
3	Data reduction (2 points)	- Identify the differences in dimensionality reduction based on features, cases, and reduction of value techniques. - Understand the basic principles of feature selection and feature composition tasks using corresponding statistical methods	- Explain the advantages of data reduction in the preprocessing phase of a datamining process	Apply and compare Fisher ratio and principal component analysis for feature extraction
3	Supervised learning for classification (3 points)	Identify the basic concepts supervised learning for classification.	Apply techniques of supervised learning for classification	Explain the advantages and disadvantages for each techniques.
4	Unsupervised learning (3 points)	Identify the basic concepts unsupervised learning for clustering and association rules.	Distinguish between different representations of clusters and different measures of similarities - Derive the K-means method for partitional clustering and analysis of its	Compare the frequent pattern growth method with the Apriori algorithm

			complexity. Describe the Apriori algorithm and explain all its phases through illustrative examples.	
Tổng	10	6	4	0.5
R2. MID-EXAM				
4	Data warehouse and Data mining (4 points)	Concepts Process Identify the goals and primary tasks of data mining process	- Analyze OLTP and OLAP - Explain the influence of data quality on a data-mining process.	Establish the relationship between data warehousing and data mining - Explain of big data and data science
5	Supervised learning - Classification (6.0 điểm)	Concept and Process of classification problem	Analyze the characteristics of a logic-based approach to classification problems.	
		Identify in depth the C4.5 algorithm for generating decision trees and decision rules	Compare ID3 and C4.5 Apply on real dataset.	Know when and how to use pruning techniques to reduce the complexity of decision trees and decision rules.
		Summarize the limitations and advantages of representing a classification model by kNN, Naïve bayes	Apply on real dataset	
Total		4.75	3.5	1.75
R3. FINAL EXAM				
No	Contents	Level 1	Level 2	Level 3
3.	Data warehouse and Data mining	Concepts Process Identify the goals	- Analyze OLTP and OLAP - Explain the	Establish the relationship between data

	(2 points)	and primary tasks of data mining process	influence of data quality on a data-mining process.	warehousing and data mining - Explain of big data and data science
4.	Classification (3.0 điểm)	Concept and Process of classification problem	Apply and implement supervised approaches to classification problems.	Summarize the limitations of representing a classification models
5.	Clustering (3.0)	Concepts Kmean	Distinguish between different representations of clusters and different measures of similarities	Discuss why validation of clustering results is difficult problem
6.	Association rules (1.0 điểm)	Concepts Apriori FP-growth	Explain the local modeling character of association rule techniques	Compare the frequent pattern growth method with the Apriori algorithm
7.	Data mining Applications (1.0 điểm)	Identify the tools of data mining technology	Setup and implement techniques in data minings	
		Use tool for real dataset for tasks in data mininf		

6.2. Student's tasks

Students must perform the following tasks:

- Attending at least 80% of the lessons of the course;
- Participating in group work activities as prescribed by the course;
- Self-studying the problems assigned by the lecturer outside the class time;
- Completing all course assessment forms.
- Attending mid-term and final examination.

7. Lesson plan and content

Week/ Session (3)	Detailed content of the Lesson / chapter	No. of periods (Theory/ Type 1	Teaching and learning activities		Asses- ment form	CLOs
			Teaching	Learning method		

periods/ session)		Practice / Type 2 Practice)	method			
1.	Chapter 1. Data warehouse and Data mining: basic concepts, tasks, modelling.	3	Oral presentation	Read material 1, chapter I, 1.1, 1.2	A1.1, A1.2	CLO 1, 2
2.	Chapter 1. Data mining: related domain, applications, process of data mining, (cont'd)	3	Oral presentation	Read material 1, chapter I, 1.1, 1.2	A1.1, A1.2	CLO 1, 2
3.	In-class practice	3	Oral presentation and practice guides	- Practice - Reports and presentation	A1.1 A1.2 A2.1	CLO 1,2
4.	Chapter 2: Learning from the data, Preparing the data, Data reduction	3	Oral presentation	Read material 1, chapter II, 2.1, 2.2		
5.	Chapter 3: Supervised learning (Classification): Decision Tree (ID3, C4.5), K-Nearest Neighbors, Naïve Bayes	3	Oral presentation	Read material 1, chapter III, 3.1- 3.4		
6.	Chapter 4: Supervised learning (Classification): Decision Tree (ID3, C4.5), K-Nearest Neighbors, Naïve Bayes	3	Oral presentation	Read material 1, chapter III, 3.1- 3.4		
7.	In-class practice	3	Oral presentation and practice guides	- Practice - Reports and presentation	A1.1 A1.2 A2.1	CLO 1, 2, 3

Week/ Session	Detailed content of the Lesson / chapter	No. of periods	Teaching and learning activities		Asses- sment	CLOs
8.	Mid-term exam		Writing assessment	Writing assessment	<i>AI.1</i> <i>AI.2</i> <i>A3.1</i>	CLO 1,2,3
9.	Chapter 5: Unsupervised learning: clustering	3	Oral presentation	Read material 1, chapter III, 3.1- 3.4	<i>AI.1</i> <i>AI.2</i> <i>A3.1</i>	CLO 2,3, 4
10.	Chapter 5: Unsupervised learning: clustering	3	Oral presentation	Read material 1, chapter III, 3.1- 3.4	<i>AI.1</i> <i>AI.2</i> <i>A3.1</i>	CLO 2,3, 4
11.	In-class practice	3	Oral presentation and practice guides	- Practice - Reports and presentation	<i>AI.1</i> <i>AI.2</i> <i>A2.1</i>	CLO 2,3,4
12.	Chapter 5: Unsupervised learning: Association rules	3	Oral presentation	Read material 1, chapter V, 5.1- 5.4	<i>AI.1</i> <i>AI.2</i> <i>A3.1</i>	CLO 2,3, 4
13.	In-class practice	3	Oral presentation and practice guides	- Practice - Reports and presentation	<i>AI.1</i> <i>AI.2</i> <i>A2.1</i>	CLO 2,3,4
14.	Chapter 5: Advances in Data mining, Tools and Applications	3	Oral presentation	Read material 1, chapter V, 5.1- 5.4	<i>AI.1</i> <i>AI.2</i> <i>A2.1</i>	CLO 2,3,4
15.	Chapter 5: Advances in Data mining, Tools and Applications (cont'd)	3	Oral presentation	Read material 1, chapter VI, 5.1- 5.4	<i>AI.1</i> <i>AI.2</i> <i>A2.1</i>	CLO 2,3,4
16.	In-class practice	3	Oral	- Practice	<i>AI.1</i>	CLO 2,3,4

Week/ Session	Detailed content of the Lesson / chapter	No. of periods	Teaching and learning activities		Asses- sment	CLOs
			presentation and practice guides	- Reports and presentation	A1.2 A2.1	
17.	Final examination		Presentation	Report/ Interview	A1.1 A1.2 A3.1	CLO 4.

Note:

- Determine the number of periods (Theory/Practice): Determine the number of theoretical, practical/experimental periods of each chapter

- Teaching methods to achieve the CLOs: Name the teaching methods used in each chapter to achieve the CLOs.

- Students' learning methods: Identify the contents students need to prepare at home (which document to read, from what page, working in groups to do exercises, working on projects); Activities in class (group discussion, regular exercise No....).

8. Learning materials

8.1. Course books, textbooks, references materials

No.	Author	Publishing year	Title of book, textbook, title of article, document	Publishing house, journal /place of publication
Required books, lectures and textbooks				
1	Lê Văn Phùng, Quách Xuân Trường	2012	Khai phá dữ liệu	NXB Thông tin và Truyền thông
2	Jiawei Han, Micheline Kamber and Jian Pei	2013	Data Mining Concepts And Techniques	Morgan Kaufmann
3	Mehmed Kantaedzic	2020	Data mining: Concepts, Models, Methods, and Algorithms	Wiley, IEEE press
Reference books and textbooks				
1	P. N. Tan, M. Steinbach, V. Kumar	2006	Introduction to Data Mining	Addison-Wesley

No.	Author	Publishing year	Title of book, textbook, title of article, document	Publishing house, journal /place of publication
2	Paulraj Ponnian	2008	Data Warehousing Fundamentals	Addison-Wesley

8.2. List of website addresses for references

No.	Contents	Website	Access date
1	Weka toolkit: Data Mining System with Free Open Source Software in Java.	https://www.cs.waikato.ac.nz/~ml/weka/index.html	20/7/2021
2	Qlik tool : Data mining and visualisation tool	https://www.qlik.com/us/products/qlik-sense	20/7/2021
3	Orange Data mining	https://orangedatamining.com/	20/7/2021
4	Rapid miner	https://rapidminer.com/	20/7/2021
5	UCI dataset	https://archive.ics.uci.edu/ml/datasets.php	20/7/2021
6	Data mining competition	https://www.kaggle.com/competitions	20/7/2021

9. Facilities

No	Classroom, lecture hall, laboratory, practice room...	Equipment, tool, software...		For Content/Chapter
		Item	Quantity	
1	Classroom	Computer, Projector	01	
2	Classroom	Computer, Projector	01	

Da Nang, ...

Dean	Head of Division	Lecturer in charge