IS252 - Data mining

Module designation	**IS252 - Data mining** The course provides basic concepts of data mining, data mining process, data mining techniques, trends as well as challenges in data mining. The course also equips knowledge and practical skills in many commonly used methods such as: frequent itemset and association rules, rough set, outlier detection, classification, clustering, and text mining. Practical exercises help students understand the theoretical contents and know how to use data mining tools. In addition, students carry out a group project to solve data mining problems in real life.
Semester(s) in which the module is taught	6
Person responsible for the module	Prof. PhD. Do Phuc, Assoc. Prof. PhD. Nguyen Dinh Thuan, PhD. Cao Thi Nhan, MSc. Trinh Minh Tuan, MSc. Mai Xuan Hung.
Language	English, Vietnamese
Relation to curriculum	Specialisation
Teaching methods	Lecture, discussion, seminar, project.
Workload (incl. contact hours, self-study hours)	(Estimated) Total workload: 195 hours. - Contact hours: Lecture: 45 hours, Discussion: 30 hours. - Private hours: 120 hours
Credit points	Number of credits: 4 - Lecture: 3 - Laboratory: 1
Required and recommended prerequisites for joining the module	Pre-study: Databases, Probability and Statistics

Module objectives	CLO	CLOs description	ILOs
	G1	- Understand and present basic concepts of data mining, the importance of data mining application in real life. - Understand and present data mining process, the data preprocessing steps. - Understand and present data mining methods such as: frequent itemset and association rules, rough set, outlier detection, classification, clustering, and text mining.	ILO2 (2.2)
	G2	- Have ability to identify and state problems (conceive ideas) - Have ability to apply data mining techniques to solve the given problems (from collecting and pre-processing data, choosing algorithms, evaluating achieved results, and integrating into application).	ILO3 (3.1, 3.2, 3.3)
	G3	- Have English reading, writing and presentation skills.	ILO6 (6.1, 6.2)

	CLO	ILO	CLOs description	Competency level
Module intended learning outcomes	G1.1	2.2	- Understand and present basic concepts of data mining, the importance of data mining application in real life. - Understand and present data mining process, the data preprocessing steps. - Understand and present data mining methods such as: frequent itemset and association rules, rough set, outlier detection, classification, clustering, and text mining.	K4
	G2.1	3.1	- Have ability to identify and state data mining topic: context, boundary, propose collection data method, pre-processing data techniques, algorithms, evaluation achieved results metrics, and integration into application. - Describe the problem with medium complexity	S4
	G2.2	3.2 3.3	- Have ability to apply data mining techniques to solve the given problems: from collecting and pre-processing data, choosing algorithms, evaluating achieved results, and integrating into application. - Have ability to apply analytical techniques and evaluate results using tools	S4
	G3.1	6.1 6.2	- Apply reading skills to read documents and articles in English. - Apply writing skills to complete report and presentation skills in English.	S2
	(Competency level: K: Knowledge, S: Skill, A: Attitude)			

Content	**Theory**				

Theory

Week/ Duration 4 hours)	Content		CLOs	Assessment elements
1	Chapter 1: Overview		G1.1	
2	Chapter 2: Data preprocessing		G1.1, G2.1, G2.2	A2
3	Chapter 3: Frequent itemsets and associate rules		G1.1, G2.1, G2.2	A2
4	Chapter 4: Frequent sequence		G1.1, G2.1, G2.2	A2
4, 5	Chapter 5: Rough set		G1.1, G2.1, G2.2, G3.1	A2
6, 7	Chapter 6: Classification		G1.1, G2.1, G2.2, G3.1	A2
8, 9	Chapter 7: Clustering		G1.1, G2.1, G2.2, G3.1	A2
10	Chapter 8: Text mining		G1.1, G2.1, G2.2, G3.1	A2
11	Review			

Lab

Week/ Duration (5 hours)	Content	CLOs	Assessment elements
1	- Data Mining Tools - Data types - Data preprocessing	G1.1, G2.1, G2.2	A1
2	- Frequent itemsets and associate rules	G1.1, G2.1, G2.2	A1
3	- Classification - Guide to Project Implementation	G1.1, G2.1, G2.2, G3.1	A1
4	- Clustering - Text mining	G1.1, G2.1, G2.2, G3.1	A1
5	- Guide to Project Implementation	G2.1, G2.2, G3.1	
6	- Present final project	G1.1, G2.1, G2.2, G3.1	A1

Examination forms		Assessment element	CLOs	Percentage	
		A1. Labs and Project	G1.1, G2.1, G2.2, G3.1	50%	
		A2. Final exam	G1.1, G2.1, G2.2, G3.1	50%	

Study and examination requirements	- Register and do Projects in groups of up to 4 students. - Students participate activities in class: lecture and discussions. - Students must read slides of the lesson before each class, and seriously do the registered group project. - Students must complete exercises in practical sections. - Students must attend 80% of the class sessions, must participate in the group's project presentation.
Reading list	[1] Do Phuc, Data mining, Vietnam National University – Ho Chi Minh city Publishing House, 2020. [2] Vu Huu Tiep, Fundamental of Machine Learning, Science and Technology Publishing House, 2019. [3] Jiawei Han, Micheline Kamber, and Jian Pei, Data Mining Concepts and Techniques, 3rd edition, Morgan Kaufmann Publishing house, Elsevier, 2012.