VIETNAM NATIONAL UNIVERSITY – HO CHI MINH CITY

UNIVERSITY OF INFORMATION TECHNOLOGY

SYLLABUS
IS252 – DATA MINING

1. GENERAL INFORMATION

Course name (Vietnamese):	Khai thác dữ liệu
Course name (English):	Data Mining
Code:	IS252
Type of course:	Speciality
Department:	Information Systems
Instructor:	PhD. Nguyen Dinh Thuan, MSc. Mai Xuan Hung,
	Email: thuannd@uit.edu.vn, hungmx@uit.edu.vn,
Number of credits:	4
Theory:	3
Lab:	1
Prerequisite:	Fundamentals of Database Systems, Probability and Statistics

2. COURSE DESCRIPTION

The course provides basic concepts of data mining and data mining process as well as main phases of data mining process.

Students are equipped many commonly used methods such as: classification, clustering, regression, association rules… Moreover, students are practised using real knowledge discovery tools.

3. COURSE GOALS

Table 1.

Gx	Goal description	Program outcomes
G1	Discuss about topics, team creation and action plan of team project.	3.1.1, 3.1.2, 3.2.2, 3.2.3
G2	Understand basic concepts, importance of data mining	1.2.6

G3	Know applying of data mining in real life fields	1.3.6
G4	Understand data mining process	1.3.6
G5	Understand and have ability to apply techniques for determining frequent itemsets, association rules. Have ability to analyse and evaluate results of these techniques comparing to the results used SQL server tool.	1.3.6, 2.1.1, 2.1.3, 4.1.1, 4.3.2
G6	Understand and have ability to apply techniques for mining frequent episodes. Have ability to analyse and evaluate results of these techniques comparing to the results used SQL server tool.	1.3.6, 2.1.1, 2.1.3, 4.1.1, 4.3.2
G7	Understand and have ability to apply rough set techniques for data reduction.	1.3.6, 2.1.1, 2.1.3, 4.1.1, 4.3.2
G8	Understand and have ability to apply techniques for classification. Have ability to analyse and evaluate results of these techniques.	1.3.6, 2.1.1, 2.1.3, 4.1.1, 4.3.2
G9	Understand and have ability to apply techniques for clustering. Have ability to analyse and evaluate results of these techniques.	1.3.6, 2.1.1, 2.1.3, 4.1.1, 4.3.2
G10	Understand and have ability to data mining in database technology	1.2.10, 1.3.6, 2.1.1, 2.1.3

4. COURSE LEARNING OUTCOMES

Table 2.

Course outcomes	**Descriptions**	**Level of teaching**
G1.1	Create and manage team	U
G1.2	Participate on team work discussion based on course topics	U,A
G2.1	Understand basic concepts: data mining, and its importance.	T
G2.2	Understand the importance of data mining in real file.	T,U
G3.1	Know applying data mining techniques in various real life fields.	T,U

G4.1	Understand phases of data mining process.	T, U
G4.2	Understand and have ability to apply data preprocessing phase on data mining	T,U
G5.1	Understand and have ability to use techniques for determining frequent itemsets, maximal frequent itemsets.	T,U
G5.2	Understand and have ability to use techniques for determining association rules based on maximal frequent itemsets, confidence of associate rule. Analyse and evaluate results of these techniques comparing to the results used SQL server tool.	T,U, A
G6.1	Understand and have ability to use techniques for determining frequent parallel episodes, frequent serial episodes.	T,U
G6.2	Understand and have ability to use techniques for determining association rules based on frequent episodes.	T,U
G7.1	Understand concepts: rough set, indiscernibility relation, lower approximation, upper approximation.	T,U
G7.2	Understand and have ability to determine discernibility matrix, discernibility function as well as reduct discernibility function.	T,U
G8.1	Understand concept: classification	T,U
G8.2	Understand and have ability to use classification technique based on Bayesian network. Analyse and evaluate results of this technique comparing to the results used recent tools.	T,U,A
G8.3	Understand and have ability to use classification technique based on Decision tree. Analyse advantages and disadvantages of the method comparing to Bayesian network.	T,U,A
G8.4	Understand and have ability to use classification technique based on Neural network. Analyse advantages and disadvantages of the method comparing to Bayesian network and Decision tree.	T,U,A
G9.1	Understand concept: clustering	T,U
G9.2	Understand and have ability to use clustering technique based on Kmean network. Analyse advantages and disadvantages	T,U,A

	of the method.	
G9.3	Understand and have ability to use clustering technique based on Kohonen network. Analyse advantages and disadvantages of the method. Analyse and evaluate results of the method comparing to the results used Kmean network.	T,U,A
G10.1	Understand great support of data mining on database technology	T,U

5. COURSE CONTENT, LESSON PLAN

a. Theory

Table 3.

Week (3 hours)	Contents	Course learning outcomes	Activities	Assessment element
1	**Chapter 1: Overview of Data Mining** 1.1. Knowledge discovery process 1.2. Concepts 1.3. Significance of data mining 1.4. Data mining applications on diversified fields. 1.5. Make groups to do project.	G1.1, G2.1, G2.2, G3.1	- Teaching: teach, give situations and discuss. Lecture explains based on examples.	Groups propose applications used data mining in real life.
2	**Chapter 2: Data preporcessing** 2.1. Introduction 2.2. Data cleaning 2.3. Data integration 2.4. Data transformation 2.5. Data reduction 2.6. Data discretization 2.7. Generating concept hierarchies 2.8. Data representation	G4.1, G4.2	- Teaching: teach, guide teamwork skills. Lecture summarizes, evaluates and concludes. - Learning: Group exercise in class.	Based on discussion and presentation of group exercise.
3,4	**Chapter 3: Frequent itemsets and association rule** 3.1. Introduction 3.2. Presentation of association rules 3.3. Frequent itemsets 3.4. Determine association rules based on Apriori algorithm and its variations 3.5. Determine association rules based on maximal frequent itemsets 3.6. Confidence of association rule	G5.1, G5.2	- Teaching: Lecture gives situations and explains. Lecture summarizes, evaluates and concludes. - Learning: Students do exercise 1 and submit in class.	Based on discussion and presentation of the exercise 1.
5	**Chapter 4: Frequent episodes** 4.1. Introduction 4.2. Frequent parallel episodes	G6.1, G6.2	- Teaching: summarize the exercise 1;	Based on discussion and presentation

	4.3. Frequent serial episodes 4.4. Determine association rules based on frequent episodes		Lecture teaches, gives situations, and explains. Lecture summarizes, evaluates and concludes. - Learning: Students do exercise 2 and submit in class.	of the exercise 2.
6,7	**Chapter 5: Rough set** 5.1. Introduction 5.2. Indiscernibility relation 5.3. Upper approximation 5.4. Lower approximation 5.5. Dependency of attributes 5.6. Determine discernibility matrix 5.7. Determine discernibility function 5.8. Determine conductions based on discernibility functions	G7.1, G7.2	- Teaching: summarize the exercise 2; Lecture teaches, gives situations, and explains. Lecture summarizes, evaluates and concludes. - Learning: Students do exercise 3 and submit in class.	Based on discussion and presentation of the exercise 3.
8,9	**Chapter 6: Data classification** 6.1. Introduction 6.2. Decision tree 6.3. Bayesian network 6.4. Neural network	G8.1, G8.2, G8.3, G8.4	- Teaching: summarize the exercise 3; Lecture teaches, gives situations, and explains. Lecture summarizes, evaluates and concludes. - Learning: Students do exercises 4, 5 and submit in class.	Based on discussion and presentation of the exercises 4, 5.
10,11	**Chapter 7: Data clustering** 5.1. Introduction 5.2. Kmean network 5.3. Kohonen network 5.4. Various methods for data clustering	G9.2, G9.3	- Teaching: summarize the exercises 4, 5; Lecture teaches, gives situations, and explains. Lecture summarizes, evaluates and concludes. - Learning: Students do	Based on discussion and presentation of the exercise 6, 7.

			exercises 6, 7 and submit in class.	
12,13	**Chapter 8: Data mining and database technology** 8.1. Introduction to database technology 8.2. Data mining capability of database technology 8.3. Query language for data mining 8.4. Current Database Management Systems with data mining support	G10.1, G10.2	- Teaching: summarize the exercise 6, 7; Lecture teaches, gives situations, and explains. Lecture summarizes, evaluates and concludes. - Learning: Students do exercise 8 and submit in class.	Based on discussion and presentation of the exercise 8.
14	**Project Presentation**		- For each group, students present their own project, other teams together discuss, assess the project.	Based on presentation, response to questions, interface and function of the demo.
15	**Review**		- Reviewing - Fixing common errors. - Students give questions.	

b. Lab

Table 4.

Week (4 hours)	**Contents**	**Course learning outcomes**	**Activities**	**Assessment element**
01	Introduction to data mining tools	G3.1, G4.1, G10.1, G10.2	❖ Teaching: – Introduction to data mining tools: Analysis Services Multidimensional and Dataming, Weka Data Mining Software ❖ Classroom activities: – Use functions of these tools ❖ Homework: – Install and find out about Analysis Services Multidimensional and	Based on utilizing of tool functions by students

			Dataming, Weka Data Mining Software.	
02,03	Practise data preprocessing for applying association rules method	G5.1, G5.2	❖ Teaching: – Create data source, data source view, targeted mailing mining model structure. – Determine data types and content types – Mining association rules using SQL Server. ❖ Classroom activities: – Practise following the reference SQL Server 2012 Tutorials - Analysis Services Data Mining (lesson 01) ❖ Homework: – Do homework 1	Based on the results of determining, frequent itemsets and association rules
04,05	Practise data import for some mining models	G8.1, G8.2 G9.1	❖ Teaching: – Guide data import for clustering mining model, Naive Bayes mining model, … ❖ Classroom activities: – Practise following the reference SQL Server 2012 Tutorials - Analysis Services Data Mining (lesson 03) ❖ Homework: – Do homework 2.	Based on the homework 1
06,07	Practise data mining for Exploring the Targeted Mailing Models	G8.2, G8.3 G9.1	❖ Teaching: – Guide some models: decision tree, clustering, naive Bayes. – Discuss on homework ❖ Classroom activities: – Practise following the reference SQL Server 2012 Tutorials - Analysis Services Data Mining (lesson 04)	Based on the homework 2

			❖ Homework: – Do homework 3	
08 **(2 hours)**	Practise predictions	G10.1, G10.2	❖ Teaching: – Guide mining steps for prediction problems. – Discuss on homework ❖ Classroom activities: – Practise following the reference SQL Server 2012 Tutorials - Analysis Services Data Mining (lesson 05, 06) ❖ Homework: – Do report, documentation for homework 2, 3.	Based on the homework 3

6. COURSE ASSESSMENT

Table 5.

Assessment element	Course learning outcomes (Gx)	Percentage (%)
Attendance, homework	G5,G6,G7,G8,G9,G10	**10%**
Midterm exam	G5,G6,G7	**20%**
Project	G1,G5,G6,G7,G8,G9,G10	**30%**
Final exam	G5,G6,G7,G8,G9,G10	**40%**

7. COURSE REQUIREMENTS AND EXPECTATIONS

- Register group for doing project (2-3 students each group)
- Students pay attention to lectures, discuss actively team exercises in class. Students have to find out about lecture in advance of each class as well as study seriously the project.
- Students have to do contents in the practical section.
- Student have to participate at least 80% classes and the presentation of team project.

8. COURSE MATERIALS

1. Do Phuc, *Textbook: Data Mining*, Vietnam National University, 2006.
2. Do Phuc, *Data Mining (Lectures)*, Vietnam National University, 2006.
3. Ho Tu Bao, *Introduction to knowledge discovery and data mining*, IOIT, 2001.
4. Jiawei Han and Micheline Kamber, *Data Mining Concepts and Techniques, University of Illinois, Morgan Kaufmann Publishers, 2006.*
5. X. Wu, V. Kumar, J. Ross Quinlan, *Top 10 Algorithms in Data Mining*, Chapman & Hall/CRC, Taylor & Francis Group, LLC, 2009.
6. ZhaoHui Tang and Jamie MacLennan, *Data Mining with SQL Server 2005*, Wiley Publishing, 2005.

7. Graham J. Williams and Simeon J. Simoff, *Data Mining: Theory, Methodology, Techniques, and Applications*, Springer-Verlag, 2006.

9. SOFTWARE, TOOLS

1. Microsoft SQL Server.
2. Weka.

Date:

Faculty Head **Instructor**