University of Computer Studies, Yangon

Faculty of Information Science

2019-2020 Academic Year

Diploma in Computer Science (DCSc.)

Subject Code	IS-204	**Subject Name**	Data Mining (Elective)
Credit point	3	**Course Coordinator**	Dr. Nang Saing Moon Kham Professor
Prerequisite	IS-205	**Semester**	Second
Online Lecture hours	21 hrs	**Online Self Learning hours**	60 hours
Practical Hour	24 hours (1.5hr*16 Week))	**Tutorial Test hours**	3 hours (1.5 hrs*2 Week)

Course Description

This course is an introductory course on data mining. It introduces the basic concepts, principles, methods, implementation techniques, and applications of data mining. Data Mining studies algorithms and computational paradigms that allow computers to find patterns and regularities in databases, perform prediction and forecasting, and generally improve their performance through interaction with data. This course also provides hands-on experience with data mining methods using WEKA software tool.

Course Objectives:

This course is intended to

- Introduce the basic applications, concepts and techniques of data mining.
- Develop skills of using recent data mining software for solving practical problems.
- Develop ability to analyze and construct knowledge from data.
- Develop and apply critical thinking, problem solving and decision making skills.

Learning Outcomes:

Students who complete the course should be able to

- Know the latest development of knowledge discovery and data mining concepts and techniques.
- Utilize the theories and algorithms for data mining and knowledge discovery.
- Manipulate the possibilities and fundamental limitations that are included in data preprocessing steps of data mining.

- Demonstrate an understanding of the basic machine learning algorithmic methods that support knowledge discovery.
- Apply the data mining method using WEKA tools for analysis.

Course Contents

1. Introduction to Data Mining
2. Getting to know your data
3. Data Preprocessing
4. Mining frequent patterns, Associations and Correlations, Basic Concept and Method
5. Classification Basic Concept
6. Cluster Analysis: Basis Concept and Method

Tools: WEKA software

Text Book

- Jiawei Han, MichelineKamber and Jian Pei, Data Mining – Concepts and Techniques. Morgan Kaufmann, Third Edition, 2011.

Reference Books

- Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Introduction to Data Mining, Addison Wesley, 2006.
- Ian H. Witten and Eibe Frank, Data Mining – Practical Machine Learning Tools and Techniques (2nd Ed.), Morgan Kaufmann, 2005.
- S.M. Weiss and N. Indurkhya, Predictive Data Mining, Morgan Kaufmann, 1998.
- Margaret H. Dunham, Data Mining – Introductory and Advanced Topics, Prentice Hall, 2003.

Learning Assessment

Paper Exam	:	50%
Tutorials/Test	:	10%
Quiz/Discussion	:	10%
Project /Presentation	:	15%
Assignment/Lab	:	10%
Class Participation	:	5%

Course Policy

Assigned Readings: The student is expected to read Lectures and assignments to prepare for scheduled discussions of the material.

Attendance: The student is expected to attend orientation classes, online Lecture and Practical Class, the exam meetings, and scheduled project presentations. Regular class and/or online participation should ensure that expectations are understood, and provide feedback to monitor and

assess progress. The student is responsible for accessing the course website to obtain assignments and related materials.

Participation: The student is expected to take part in regular class and online discussions too, implement and test lab assignments and assist class members with technical issues.

Lab projects: It is expected that the student will begin each project when assigned or as topics are approved, then present system components by the scheduled progress reporting dates. The Lab project schedule and book format are going to be confirmed during the lecture course.

Exams and Tutorial test: The student is expected to complete each exam and tutorial test at the scheduled time. All exams and tutorial are based upon all learning objectives to be reached before the scheduled date. Final Exam date is declared at the time table of course schedule and the tutorial test schedule are going to be confirmed during the lecture course.

Intellectual Honesty: By departmental policy, the discovery of plagiarism (i.e. copying from another's assignment paper or lab project) will result in a reduction of grade result.

University of Computer Studies
2019-2020 Academic Year
Diploma in Computer Science (DCSc.)

IS-204 **: Data Mining** **(Second Semester)**

Text Book **: Data Mining Concepts and Techniques (Jiawei Han, Micheline Kamber, Jian Pei)**

Period - 48 periods for 18 Weeks including Practice (50 minutes for 1 period)

No.	Chapter	Page	Period	Remark
1	**1 Introduction**	1	1	
	1.1 Why Data Mining?	1		
	1.2 What is Data Mining?	5		
	1.3 What Kinds of Data Can Be Mined?	8	1	
	1.4 What Kinds of Patterns Can Be Mined?	15	1	
	1.5 Which Technologies Are Used?	23	0.5	
	1.6 Which Kinds of Applications Are Targeted?	27	0.5	
	1.8 Summary	33	2	
	1.9 Exercises	34		
2	**2 Getting to Know Your Data**	39		
	2.1 Data Objects and Attributes Types	40	1	
	2.2 Basic Statistical Description of Data	44	1	
	2.3 Data Visualization	56	1	
	2.4 Measuring Data Similarity and Dissimilarity	65	1	
	2.5 Summary	79	2	
	2.6 Exercises	79		
3	**3 Data Preprocessing**	83		
	3.1 Data Preprocessing: An Overview	84	1	
	3.2 Data Cleaning	88	1	
	3.3 Data Integration	93	1	
	3.4 Data Reduction	99	1	
	3.5 Data Transformation and Data Discretization	111	1	
	3.6 Summary	120	2	
	3.7 Exercises	121		
4	**6 Mining Frequent Patterns, Associations, and Correlation: Basic Concepts and Methods**	243		
	6.1 Basic Concepts	243	2	
	6.2 Frequent Itemset Mining Methods	248	2	
	6.3 Which Patterns Are Interesting? Pattern Evaluation Methods	264	2	