اعتماد
NCAAA

T4

2020

Course Specifications

Course Title:	Data Mining
Course Code:	CSI 512
Program:	Computer Sciences and Information Technology
Department:	Computer Science and Information
College:	Science at Al-Zulfi
Institution:	Majmaah

VISION
2030
المملكة العربية السعودية
KINGDOM OF SAUDI ARABIA

تميز
NCSEE

مسار
MASAR

اعتماد

قياس
qiyas

www.etec.gov.sa

Table of Contents

A. Course Identification

1. Credit hours: 3
2. Course type
a. University ☐ College Department ■ Others
b. Required Elective
3. Level/year at which this course is offered: Level 9
4. Pre-requisites for this course (if any)**:** CSI 314 Database
5. Co-requisites for this course (if any)**:** Nil

6. Mode of Instruction (mark all that apply)

No	Mode of Instruction	Contact Hours	Percentage
1	Traditional classroom	44	80 %
2	Blended	3	5 %
3	E-learning	3	5 %
4	Correspondence	3	5 %
5	Other	3	5 %

7. Contact Hours (based on academic semester)

No	Activity	Contact Hours
1	Lecture	30
2	Laboratory/Studio	30
3	Tutorial	-
4	Others (specify)	-
	Total	60

B. Course Objectives and Learning Outcomes

1. Course Description

This course introduces an introduction to the fundamentals of data mining principlesand a basic understanding of data analysis techniques required to tackle real-world, data-rich problems in business and academia, including:

- Data set acquisition, cleaning, ,aggregation , dimension reduction , FeatureSelection and extraction based on PCA.
- Similarity and dissimilarity of objects
- Exploratory data analysis and visualization
- Data classification methods: Decision tree, Rule-based, Bayes' theorem , K-Nearest Neighbor and Support Vector Machine (AVM).

Clustering methods: K-means

2. Course Main Objective

1. Students will develop relevant programming abilities.
2. Students will demonstrate proficiency with statistical analysis of data.
3. Students will develop the ability to build and assess data-based model classifications.
4. Students will execute data analysis with professional software.
5. Students will demonstrate skill in data analysis and management

3. Course Learning Outcomes

	CLOs	Aligned PLOs
1	**Knowledge and Understanding**	
1.1	Acquire familiarity with the basic concepts of data mining .	
1.2	Identify the problems and tasks including dataset collection, data preprocessing and data analysis.	
1.3	An understanding of problems solvable and an ability to attack them from a statistical perspective.	K2
1.4	An understanding of when to use supervised and unsupervised learning methods on labeled and unlabeled data-rich classification/clustering problems.	
2	**Skills :**	
2.1	Distinguish between different kinds of data and identify challenges related tobig data.	
2.2	The ability to create/extract features from big data and applications in Matlab.	S1
2.3	An ability to extract rules involving relations, trees for numeric prediction,instance based classification.	
2.4	Familiarity with the Matlab various tools needed to classify and clusteringbig data based on using supervised and unsupervised learning methods.	S2
3	**Values:**	
3.1	An ability to implement and use rules for numeric prediction, instance basedrepresentation and cluster data.	
3.2	Applying proficiency with statistical analysis of data using Matlab.	
3.3	Applying supervised and unsupervised statistical learning methods on datasetusing Matlab tools.	C1
3.1	An ability to implement and use rules for numeric prediction, instance basedrepresentation and cluster data.	

C. Course Content

No	List of Topics	Contact Hours
1	**Data Mining Introduction** o What is data mining o Data mining functionalities o Data mining applications Data mining systems	8

2	**Data Preparation**
	    o Descriptive data summarization
	    o Data cleaning
	    o Data integration and transformation
	    o Data preprocessing : Dimension reduction & Feature extraction
	    o Data discretization and concept hierarchy generation
	     Lab: Data analysis applications with Weka software or Matlab.
		8	

3	**Data Warehouse and OLAP**
	    o Data warehouse introduction
	    o Multidimensional data model
	                       Data cube and OLAP
		8	

4	**Frequent Pattern Mining**
	    o Basic concepts
	    o Efficient and scalable frequent item set mining methods
	    o Correlation analysis
	    o Sequential pattern mining
	    o Graph and tree mining
	     Lab: Data exploration applications with Weka software or Matlab.
		12	

5	**Classification and Prediction**
	    o Decision tree induction
	    o Bayesian classification
	    o Support vector machines
	    o K-Nearest neighbor methods
	    o Other classification methods
	    o Performance evaluation
	    Lab: Advanced topics and applications with Weka software or Matlab.
		12	

6	**Clustering Analysis**
	    o Partition methods
	    o Hierarchical methods
	    o Density-based methods
	    o Outlier analysis
	
	   Lab: Advanced topics and applications with Weka software or Matlab
		12	

Total		60	

D. Teaching and Assessment

1. Alignment of Course Learning Outcomes with Teaching Strategies and Assessment Methods

Code	Course Learning Outcomes	Teaching Strategies	Assessment Methods
1.0	**Knowledge and Understanding**		
1.1	Acquire familiarity with the basic concepts of data mining .	▪ **Direct Teaching:** Lectures, PowerPoint slides and discussion.	- Homework tasks - Quiz - Midterms - Final Exam
1.2	Identify the problems and tasks including dataset collection, data preprocessing and data analysis.		
1.3	An understanding of problems solvable andan ability to attack		- E-learning - Internet search

Code	Course Learning Outcomes	Teaching Strategies	Assessment Methods
	them from a statistical perspective.	**Aimed Teaching** Discovery and Oral Questions	- Oral Exam
1.4	An understanding of when to use supervisedand unsupervised learning methods on labeled and unlabeled data-richclassification/clustering problems.		
2.0	**Skills**		
2.1	Distinguish between different kinds of data and identify challenges related to big data.	**Indirect Teaching:** Brainstorming - Free Discovery –Inquiry	- HW Exercises - Lab Exam - Oral Exam - Presentations
2.2	The ability to create/extract features from big data and applications in Matlab.		
2.3	An ability to extract rules involvingrelations, trees for numeric prediction, instance based classification.		
2.4	Familiarity with the Matlab various toolsneeded to classify and clustering big data based on using supervised and unsupervisedlearning methods.		
3.0	**Values**		
3.1	An ability to implement and use rules fornumeric prediction, instance based representation and cluster data.	**Course Project: (Work group)** critical thinking and ability to seek solutions.	Introduce group project and case study approachesto enable students to have an experience in problem solving situations.
3.2	Applying proficiency with statistical analysis of data using Matlab.		
3.3	Applying supervised and unsupervised statistical learning methods on dataset usingMatlab tools.		

2. Assessment Tasks for Students

#	Assessment task*	Week Due	Percentage of Total Assessment Score
1	Homework 1	2	2%
2	QUIZ 1	3	5%
3	Homework 2	4	2%
4	QUIZ 2	5	5%
5	Midterm 1	6	10%
6	Homework 3	7	2%
7	QUIZ 3	8	5%
8	Homework 4	9	2%

#	Assessment task*	Week Due	Percentage of Total Assessment Score
9	QUIZ 4	10	5%
10	Midterm 2	11	10%
11	Lab Exam/ Project Evaluation	14	12%
12	Final Exam	16	40%

*Assessment task (i.e., written test, oral test, oral presentation, group project, essay, etc.)

E. Student Academic Counseling and Support

Arrangements for availability of faculty and teaching staff for individual student consultations and academic advice:
- Determine meeting appointments for the weak' students to solve their problems and givethem academic advices.
- One office hour daily
- Dealing a workshops.
- Motivate students

F. Learning Resources and Facilities

1.Learning Resources

Required Textbooks	Tan, P., Steinbach, M., and Kumar, V., Introduction to Data Mining, Pearson Education, Inc., 2006.
Essential References Materials	Han, J., Kamber, M., Pei, J., Data Mining: Concepts and Techniques, Third Edition, 2011
Electronic Materials	https://hanj.cs.illinois.edu/bk3/
Other Learning Materials	Matlab toolboxes: Data mining/ Data Science / Learning machine

2. Facilities Required

Item	Resources
Accommodation (Classrooms, laboratories, demonstration rooms/labs, etc.)	Classroom - Laboratory
Technology Resources (AV, data show, Smart Board, software, etc.)	Data show – Smart Board
Other Resources (Specify, e.g. if specific laboratory equipment is required, list requirements or attach a list)	Matlab software – Weka – Python Programming

G. Course Quality Evaluation

Evaluation Areas/Issues	Evaluators	Evaluation Methods
1. Questionnaires (course evaluation) filled by the students and acquired electronically by the University	Students	Indirect Assessment
2. Students-faculty management meetings		
3. Departmental internal review of the course.	Department Council	Questionnaires
4. Discussion with the industrial partnersto enhance the courses in order to meet their needs.	Stockholders	Meetings
5. Midterms and Final Exam	Course CoordinatorStaff	Direct Assessment
6. Project Evaluation		

Evaluation areas (e.g., Effectiveness of teaching and assessment, Extent of achievement of course learning outcomes, Quality of learning resources, etc.)
Evaluators (Students, Faculty, Program Leaders, Peer Reviewer, Others (specify)
Assessment Methods (Direct, Indirect)

H. Specification Approval Data

Council / Committee	
Reference No.	
Date	