COMP 5353:

Data Mining

SPRING 2019

Course Syllabus

Table of Contents

Introduction to the Course

About the Course

Meeting Days & Time: Online Course
(Check the webinar timing table)

About the Instructor

Instructor Name: Ihsan Said
Office: 809
Phone: (832) 230-5085
Virtual Office Hours: Monday   11:30AM – 01:00PM (CST)
                Tuesday   01:00PM – 02:30PM (CST)
                Wednesday   11:30AM – 01:00PM (CST)
                Thursday   01:00PM – 02:30PM (CST)
Email: isaid@na.edu

Prerequisites/ Corequisites

None

Statement on Course Materials

Some of the writings, lectures, films, or presentations in this course may include material that conflicts with the core beliefs of some students. Please review the syllabus carefully to see if the course is one that you are committed to taking. If you have a concern, please discuss it with the instructor at your earliest convenience.

Course Description

This course introduces fundamental topics of data mining and knowledge discovery, including statistical foundations, database support, data preprocessing, data warehousing, association discovery, classification, clustering, and mining complex data types.

Instructional Contact Hours/Credits

Lecture - 45 Clock Hours / 3 Semester Credits

Virtual Office Hours

This is an online course. Instructor can be contacted via email and phone calls. Phone calls will be accepted during office hours, and other times by appointment.

Please post your questions of general interest in the **Course Q and A forum**. Even if you don't have any questions of your own right now, check to see what may have been posted by others. If you know the answer to a question, feel free to post a reply. The instructor will be moderating and responding to the questions regularly. Please use email for private correspondence.
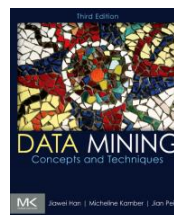
Course Materials

Textbook
Data Mining: Concepts and Techniques
3rd Edition, 2012
by Han, Kamber & Pei
The Morgan Kaufmann, Elsevier Inc

ISBN-13: 978-9380931913
ISBN-10: 9380931913

Library Resources
Since this course includes course activities that require students to gather information using library resources, students are expected to know how to use the library's resources including the available databases, periodicals, and journals. If you are not familiar with using the library, please visit the library's website for more detailed information: http://www.na.edu/library/

Other Resources
- Additional reading material will be provided in class or via course web page. http://www.na.edu/online/ (Moodle)
- Anaconda Distribution: Anaconda 4.2.0 https://www.continuum.io/downloads
- WEKA: http://www.cs.waikato.ac.nz/ml/weka/downloading.html
- R Data Mining: http://www.rdatamining.com/
- Data Camp: https://www.datacamp.com/

Technical Requirements

To access this course, students will need access to the Internet and a supported Web browser (Internet Explorer, Firefox, Safari or Chrome).

Course Technologies
This course requires assignments to be submitted in GitHub, Microsoft Office or Adobe Acrobat (pdf) file formats. Assignments created using other applications, such as Google Documents, Microsoft Works or WordPerfect, are not acceptable.

Additional Software
- Any Python/ R IDE
- Open Source Data Mining Tools (WEKA, R, Orange...Etc.)

Course Format
To complete this course successfully, students need an access to a computer and the Internet daily. All instructions are contained in this syllabus and in the NAU Moodle course management system.

This course will be delivered in class and through NAU Moodle course management system at Office 365 Portal named NAU Moodle. Students can get access through Office 365 portal with their NAU accounts provided by IT department. For login issues and account problems with NAU Moodle, please contact with Distance Education Support via one of the following methods:

- Visit Room 732
- Call 832-464-8691
- Email moodle@na.edu
- Visit http://www.na.edu/nau-distance-education/

For hardware and software problems other than NAU Moodle system, please contact with IT Department via one of the following methods:

- Visit Room 820
- Call 832-230-5541
- Email support@na.edu

Learning Outcomes

Program-Level Outcomes (PLO)

Program Name-Program Learning Outcomes (PLO)	Assessment
1. Demonstrate skills of computing theory and algorithms.	U
2. Develop knowledge of operating systems and hardware.	
3. Apply practices of software design and development cycle.	I
4. Analyze data using advanced computing tools.	U
5. Develop state-of-the-art skills of computer networks.	
6. Engage and apply secure practices in various digital environments.	I
Key: I=Introduced; E=Emphasized; U=Utilized; A=Comprehensive Assessment	

Course-Level Outcomes

Course-Level Outcomes (CLO)	Meets PLO
1. Describe a data warehouse.	4
2. Discuss data clearing and reduction issues.	4
3. Utilize an advanced data mining query language.	1,4
4. Apply statistical measures in data analysis.	1,4

5. Explain data mining association rules.	1,4
6. Utilize advanced decision tree techniques for data analysis.	1,4
7. Employ improved data visualization.	4
8. Perform cluster analysis.	4
9. Discuss ethical issues in data mining.	6
10. Apply advanced data mining techniques to solve real life problems	4

Grading

Grading Table

The following table summarizes the requirements and grading of the assignments in this course. The specific instructions for each activity are included in the appropriate forum, assignment, or quiz on the course NAU Moodle website.

Assignment	Quantity	Unit	Maximum Total
Discussion Forum	6	1	6
Webinar Participation	6	1	6
Homework Assignments	6	3	18
Quizzes	6	3	18
Course Project	1	30	30
Research Papers	2	11	22
Course Total:		**100**	

Grade Distribution

Grade Distribution Point (or Percentage Achieved)	96~100	91~95	86~90	81~85	76~80	71~75	66~70	0~65
Course Grade	A	A-	B+	B	B-	C+	C	F

Time Requirements:

The activities in this course are based on a 15-week instruction schedule. Course topics will be demonstrated and discussed in class; however, additional time outside of class is required to achieve learning objectives.

Students are expected to spend approximately 15 hours a week, on average, completing class activities such as assignments and projects. This meets the Federal Government's expectation of 2 hours of homework for each hour of lecture. The average time commitment range calculation for this course (3 Semester Credit Hour) is shown in the

following table:

Class Activities	Weekly Minimum Expected Average Time
Readings	5 hours
Assignments/Quizzes/Tests	3 hours
Class Participation	2 hours
Project	5 hours
Weekly Total:	**15 hours**
Term Total:	**225 clock-hours**

Late Submissions

Technology issues cannot be used as a reason for late assignments. You must have back-up plans for technology issues, such as technical problems with your computer, Internet server provider problems, etc.

Late work will be accepted if and only if arrangements are made with the instructor prior to time due and based on the following deduction system

Due Date Sunday 11:00PM	Submission				
	Before Sunday 11:00PM	Before Monday 11:00PM	Before Tuesday 11:00PM	Before Wednesday 11:00PM	Before Thursday 11:00PM
	Maximum Grade 100/100	Maximum Grade 90/100	Maximum Grade 80/100	Maximum Grade 70/100	Maximum Grade 60/100
Submissions after Thursdays 11:00PM of the same week will be not graded.					

There will be **no make-up quizzes or tests/exams**. If you miss an exam and have a valid, documented, university-approved excuse, that exam score will be replaced by the average of the other exams.

Assignment Guidelines

Discussion Forums

Students should participate in discussion forums and complete discussion activities with correct English usage, accurate spelling, and standard grammar. Students are expected to post a response to the discussion questions early in the week and maintain a thread of discussions throughout the week as they respond to their classmates.

Points will be given for:
- Following the timeline and quantity of posts to discussion questions
- Offering ideas or resources and inviting a review of them based on readings
- Articulating, explaining and supporting positions on ideas
- Exploring and supporting issues by adding explanations and examples

- Reflecting on and re-evaluating personal opinions based on readings
- Offering a review, challenging, discussing and expanding ideas of others
- Negotiating interpretations, definitions, and meanings
- Responding to at least two other students' responses

Grading:

Reply to Instructor's post (3/5 of the pts)	Reply to 2 other students' posts (1/5 of the pts for each)
Post a response to the instructor's discussion question. To secure credit, the response should be thoughtful; that is, it must refer to the weekly readings and information from other pertaining resources. **Post your response by 11:00 p.m. Thursday.**	respond to two other students' responses. Responses should be thoughtful and should go beyond simple "I agree" posts. They need to expand the idea presented and contribute to the discussion. Post your responses in the appropriate forum **by 11:00 p.m. Sunday.**

Bi-Weekly Assignments

The purpose of this assignment is to encourage students to read, do a research, practice, and analyze the course materials. Students are expected to respond to the bi-weekly prompt(s) provided by the instructor with both source code and writing content.

Rubric for Bi-Weekly Assignments:

Criterion	Unacceptable	Acceptable	Outstanding
	Score 0	Score 1	Score 2
Organization	No organization is evident.	Writing is generally clear and consistent.	The writing is clear, logical, and internally consistent.
Use of Source code	Information presented is sloppy or unclear of source code.	Not complete but fit the work	Complete work with efficient source code
Using of Data Tools	Diagrams are neither completed nor neat.	Diagrams are somehow completed and neat.	Diagrams are perfectly completed and neat.

Webinars

The instructor will be conducting 30 to 70 minutes' webinar sessions every other week. The webinars will include lecturing, reviewing course requirements, and answering any questions students may have. **Participation is optional but will be graded.** To get the full credit from the webinars:

1. Join the webinar, and type in your full name into chat box, **OR**
2. Watch the recorded session on the course page and submit a brief summary of the webinar, a synopsis, in 100-150 words.

The first webinar will be held on Tuesday January 15, 2019, at 5:00 PM CST.

Webinar Schedule:
Webinar links will be posted on course website. Make sure that you installed Gotomeeting software/app to participate in the webinar sessions. Please refer to our technical support website to get more technical information: http://www.na.edu/nau-distance-education/video-tutorials/

Webinar	Date	Time
Session 1	Tuesday January 15th	5:00PM CST
Session 2	Tuesday January 29th	5:00PM CST
Session 3	Tuesday February 12th	5:00PM CST
Session 4	Tuesday February 26th	5:00PM CST
Session 5	Tuesday March 19th	5:00PM CST
Session 6	Tuesday April 9th	5:00PM CST

Students Rights and Responsibilities

Students are responsible for:
- reading any assigned reading as stated in the weekly blocks of the course page
- reviewing the web-page resources posted in the NAU Moodle course site for each week
- assuring that their computer/device is compatible and working to engage effectively in this online course
- uploading assignments before or on the assigned due date/time

Students can expect:
- the instructor will return email and phone communications within 24 hours unless otherwise announced in the course page.
- Assignments/Quizzes/Project will be graded after 5 days of the due date.

Academic Honesty
Each student assumes the responsibilities of being a member of the NAU academic community. All acts of plagiarism are not tolerated including: cheating, claiming one's work as their own, fabrication and helping one to commit any of these acts. Any violations of academic honesty will receive strict disciplinary action, which can include suspension and even expulsion from NAU.

ADA Statement
When possible, and in accordance with 504/ADA guidelines, we will attempt to provide reasonable academic accommodations to students who request and require them. Please call North American University at (832) 230-5555 for more assistance.

Course Schedule

Week	Day	Chapter	Topic, Learning Activities	Assignments, and Due Dates
Week 1	Tuesday Jan 15th	Chapter 1 Introduction	About the Course Why Data Mining? What is Data Mining? What Kinds of Data Can Be Mined? What Kinds of Patterns Can Be Mined?	Webinar #1 (on Tuesday @4:00PM) Discussion #1 (due date: Thursday @11:00PM) Discussion #1 Replies (due date: Sunday @11:00PM)
Week 2	Tuesday Jan 22nd		Which Technologies Are Used? Which Kinds of Applications Are Targeted? Major Issues in Data Mining Data Mining Tools	Quiz #1 (due date: Sunday @11:00PM) Assignment #1 (due date: Sunday @11:00PM)
Week 3	Tuesday Jan 29th	Chapter 2 Getting to Know Your Data	Data Objects and Attribute Types Basic Statistical Descriptions of Data Data Visualization Measuring Data Similarity and Dissimilarity	Webinar #2 (on Tuesday @4:00PM) Discussion #2 (due date: Thursday @11:00PM) Discussion #3 Replies (due date: Sunday @11:00PM)
Week 4	Tuesday Feb 5th	Chapter 3 Data Preprocessing	Data Preprocessing: An Overview Data Cleaning Data Integration	Quiz #2 (due date: Sunday @11:00PM) Assignment #2 (due date: Sunday @11:00PM)
Week 5	Tuesday Feb 12th		Data Reduction Data Transformation Data Discretization	Webinar #3 (on Tuesday @4:00PM) Discussion #3 (due date: Thursday @11:00PM) Discussion #3 Replies (due date: Sunday @11:00PM)
Week 6	Tuesday Feb 19th	Chapter 4 Data Warehousing and Online Analytical Processing	Data Warehouse: Basic Concepts Data Warehouse Modeling: Data Cube and OLAP Data Warehouse Design and Usage	Quiz #3 (due date: Sunday @11:00PM) Assignment #3 (due date: Sunday @11:00PM)
Week 7	Tuesday Feb 26th		Data Warehouse Implementation Data Generalization by Attribute-Oriented Induction	Webinar #4 (on Tuesday @5:30PM) Discussion #4 (due date: Thursday @11:00PM) Discussion #4 Replies (due date: Sunday @11:00PM)
Week 8	Tuesday Mar 5th			Quiz #4 (due date: Sunday @11:00PM) Assignment #4

				(due date: Sunday @11:00PM)
		Research Paper I		
Week 9	Tuesday Mar 12th	SPRING BREAK		
Week 10	Tuesday Mar 19th	Chapter 5 Data Cube Technology	Data Cube Computation: Preliminary Concepts Data Cube Computation Methods Processing Advanced Kinds of Queries by Exploring Cube Technology Multidimensional Data Analysis in Cube Space	Webinar #5 (on Tuesday @5:30PM) Discussion #5 (due date: Thursday @11:00PM) Discussion #5 Replies (due date: Sunday @11:00PM)
			Advanced Topics in Data Cube Technology	
Week 11	Tuesday Mar 26th	Chapter 6 Mining Frequent Patterns, Associations, and Correlations	Basic Concepts Frequent Itemset Mining Methods Which Patterns Are Interesting? Pattern Evaluation Methods	Quiz #5 (due date: Sunday @11:00PM) Assignment #5 (due date: Sunday @11:00PM)
			Advanced Topics in Mining Frequent Patterns, Associations, and Correlations	
Week 12	Tuesday Apr 2nd	Chapter 8 Classification	Basic Concepts Decision Tree Induction Bayes Classification Methods	
Week 13	Tuesday Apr 9th		Rule-Based Classification Model Evaluation and Selection Techniques to Improve Classification Accuracy	Webinar #6 (on Tuesday @5:30PM) Discussion #6 (due date: Thursday @11:00PM) Discussion #6 Replies (due date: Sunday @11:00PM)
			Advanced Topics in Classification	
Week 14	Tuesday Apr 16th	Chapter 10 Cluster Analysis	Basic Concepts Partitioning Methods Hierarchical Methods Density-Based Methods Grid-Based Methods Evaluation of Clustering	Quiz #6 (due date: Sunday @11:00PM) Assignment #6 (due date: Sunday @11:00PM)
Week 15	Tuesday Apr 23rd		Advanced Topics in Cluster Analysis	
Week 16	Tuesday May 1st	Course Project Presentation (online)		
		Course Project Presentation (online)		
TBD		Research Paper II		

Syllabus subject to change

This syllabus may change as needed to support the student learning outcomes for this course.

This syllabus was last reviewed on December 19th,2018