Introduction to Data Mining
Spring 2020

Instructor: *Dr. Ge, Ling*
Office: *BLB 312J*
Office phone: *(940) 565-3104*
E-mail: *Ling.Ge@unt.edu*

Office hours: *Tuesday 10-12 and Thursday 10-11 or by Appointment*
Tutor Hours: *Friday 9:00am-2:00pm*
Tutor Email: *Thatavarthi, yashwanth (venkatasaiyashwanththatavarthi@my.unt.edu)*

Class Schedule:

Sec	Time	Days	Location
001	2:00pm-4:50pm	T	BLB005

Fall 2019 Academic Calendar

What happens	Dates
First day of classes	Monday, January 13
MLK Day	Monday, January 20
Last day to drop a course without a 'W'	March 30
May request an incomplete	April 6
Last day to withdraw (Grade W)	April 17
Spring Break	March 9-13
Last day of classes	April 30

Course materials:

1. Please bring your own computers to class for SAS OnDemand in-class practices.
2. Recommended Books:
   - Kattamuri Sarma, *Predictive Modeling with SAS Enterprise Miner*, Third Edition, SAS Press 2017, ISBN: 978-1-62960-264-6 (also available as an e-Book and as a downloadable PDF file). If you buy it as an e-book from their website, you can ask for 20% student discount. https://www.sas.com/en_us/learn/academic-programs/resources/academic-books.html
   - *Applied Analytics Using SAS Enterprise Miner*, by Peter Christie, Jim George, Jeff Thompson and Chip Wells. SAS e-book (Book AAEM, free PDF, available on Canvas)
   - *Data Mining Using SAS Enterprise Miner, A Case Study Approach, 4th Edition*, SAS Publishing 2018 (free PDF, available on Canvas)
   - *Getting Started with SAS Enterprise Miner 14.3*, SAS Publishing 2017 (free PDF, available on Canvas)
   - *Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner*, 2e, by Galit Shmueli, Nitin Patel and Peter Bruce. Wiley, ISBN-10: 0470526823, ISBN-13: 978-0470526828

Course description:

Large volume of data could reveal useful information about customers, products or other strategic aspects. This course aims to equip students with the knowledge and skills to extract patterns in vast amounts of data and discover actionable insights in business contexts. Students will learn to identify and solve data related business problems in today's information-rich environments. Upon completing this course, students will be able to apply their analytical capabilities to build innovative business solutions using large datasets.
The course will cover, but are not limited to, the following:

1) Data mining techniques (e.g., regression, classification, clustering, association rule analysis)

2) Business applications (e.g., direct marketing, market-basket analysis, fraud detection, credit scoring, and loan application)

Learning Objectives

- Understand the problems and opportunities when dealing with large data sets.
- Review statistical and data mining software used for identifying and interpreting complex patterns in multidimensional data. Learn to identify what information is useful and what is not.
- Provide an understanding of predictive models and algorithms, as well as exploratory algorithms.
- Examine all phases of decision making, including discovery and data query, data cleaning, data analysis and confirmation, presentation, and implementation of results.

Grading

Evaluation Items	Percentage
Exams	50%
Project (3 individual parts, 1 group part)	20%
Assignments (4 individual assignments)	30%
Total	**100%**

Exams:

There will be two exams: exam1 and exam2. Both exams will be held in class and equally weighted. You will be tested over all material covered in the lectures, course readings, labs and exercises. Make-up exams are **NOT** given in general. If you have a conflicting event or appointment, please see the instructor well in advance. Make-ups requested AFTER the exam date require verifiably extraordinary circumstances. You will be asked to provide evidence or documentation such as hospital papers, an obituary, or a doctor's notice.

Project:

The details of project requirements will be available in a separate document. The goal of this project is to successfully implement one or more of the business analytic techniques to solve a real problem in a business context. You will work individually for the first three parts.

Groups: During the last part of the project you will work in groups. **The maximum group size will be 5**. Groups will be self-managed. If the group is not satisfied with some member's contribution they may choose to dismiss that person from the group. In such a case, alternative individual assignment will be given to the dismissed group member.

Final part: The group will work together to finish the final part of the project, which includes a final report and a class presentation. The group will present the project in the last session of the class. If the project is not completed by the final deadline, students will be evaluated on their progress and on the deliverables that have been submitted prior to this deadline.

Assignments:

We have four homework assignments throughout the semester. The fourth one is optional if you are satisfied with the grades of your first three. I will take the three with highest grades. It is expected that you finish the assignments on individual basis. **Copying off from others' work is strictly prohibited.**

Late submission:

Policy on late submission is as shown in the table below. Hardware failure or inaccessibility is not a valid excuse for late work. This means that if the computer eats your assignment, or if the labs are full, you will not be excused from handing in an assignment on time.

Submission Time	Percentage of full grade
Within 24 hours after deadline	70%
Within 48 hours after deadline	50%
Beyond 48 hours after deadline	0%

Incomplete grades:
The grade of "I" is not given except for rare and very unusual emergencies, as per University guidelines. An "I" grade cannot be used to substitute your poor performance in class. If you think you will not be able to complete the class satisfactorily, please drop the course.

Letter Grades:

90% or more	A
80% or more	B
70% or more	C
60% or more	D
Below 60%	F

Class policies:

1. **Class attendance**: Regular class attendance and informed participation are expected.
2. **Academic integrity**: Business professionals must be trusted, because they may have access to a wide variety of confidential and private information. Everyone is expected to maintain the highest degree of ethical standards when taking exams or doing assignments.
   This course adheres to the UNT policy on academic integrity. The policy can be found at http://deanofstudents.unt.edu/conduct. If you engage in academic dishonesty related to this class, you will receive a failing grade on the exam or assignment, or a failing grade in the course. In addition, the case may be referred to the Dean of Students for appropriate disciplinary action.
3. **Accessibility**: The University complies with the Americans with Disabilities Act in making reasonable accommodations for qualified students with special needs. If you have any special learning or testing requirements please let us know as soon as possible so special arrangements can be made. It is better that we register with the UNT Office for Disability Accommodation (ODA) as soon as possible.
4. **Statement Regarding Email as Official Means of Communication**: Each student must have a UNT email address to be able to communicate with the professor, the TA, and students in the group or the class. Email communication will be sent only to official UNT email addresses. If you use a different email address, please ensure that your UNT email is forwarded correctly.
5. **Canvas**: The class will also rely heavily on Canvas for class communication, material distribution and discussion. It is important for you to visit Canvas at least **twice a week** for class information.
6. **Responsibilities of students**: It is YOUR responsibility to make sure you are current with course happenings. If you must miss a class, it is your responsibility to recover any missed materials with your classmates. It is your responsibility that all assignments and projects must be completed and submitted according to the instructions provided. Failure to follow instructions may result in a failing grade.
7. **Meeting Behavior**: A repeatedly disruptive student will have his or her letter grade reduced. Side conversations during lectures or when another student is speaking are to be kept to a minimum. The exception to this expectation is when students are working in teams. In this situation, I would prefer that you work with your teammates before asking the instructor. This practice will keep the class moving along at a faster pace.
8. **Leaving Class Early**: If you want to leave class earlier due to a good reason such as 1) seeing a doctor, 2) job interview, 3) important student organization activities, etc., please inform the professor before class and select a seat so that you will minimally disrupt the class.
9. **Student Perceptions of Teaching (SPOT)**: Student Perceptions of Teaching (SPOT) is a requirement for all organized classes at UNT. This short Web-based survey will be available to you at the end of the semester, providing you a chance to comment on how this class is taught. I am very interested in this feedback from my students, as I work to continually improve my teaching.
10. **Changes in the Syllabus**: The professor reserves the right to improve the materials and requirements as the semester unfolds, with sufficient warning concerning exams, and assignments.

Tentative Course Schedule – Spring 2020

This is a tentative schedule. Dates and topics of lecture are subject to change. Any changes will be announced.

Week	Date	Lecture	Due	Readings
1	01/14	Course Introduction	Form groups	
2	01/21	Basics / Data visualization/Explorative Analysis		AAEM 1
3	01/28	Explorative Analysis	Assignment 1	AAEM 2
4	02/04	Predictive Modeling-Fundamentals	Project Part 1	AAEM 3
5	02/11	Predictive Modeling - Decision Tree		AAEM 3
6	02/18	Predictive Modeling - Regression	Assignment 2	AAEM 4
7	02/25	**Exam 1**		
8	03/03	Predictive Modeling - Logistic Regression & Neural Network		AAEM 4 & 5
9	03/10	*Spring Break*		
10	03/17	Break		AAEM 6
11	03/24	Model Assessment	Assignment 3	
12	03/31	Predictive Modeling - Neural Network		
13	04/07	Unsupervised Learning – Clustering	Project Part 2	AAEM 8
14	04/14	Unsupervised Learning - Clustering		AAEM 8
15	04/21	Association Rules/Exam Review	Assignment 4	
16	04/28	**Exam 2**	Project final submission	