MIS 545 Data Mining for Business Intelligence

Fall 2019

Instructor Information:
Name: Karthik Srinivasan
Email: karthiks@domain.edu

Teaching Assistant:
Name: TBD
Email: tbd@domain.edu

When and where do we meet?

Class discussions:
Time: Monday & Wednesday - 9:30AM – 10:45AM
Location: TBD
Coaching sessions:
Time: Friday - 9:30AM – 10:45AM after our 12th class
Location: TBD

Instructors office Hours:
Time: Monday: 2:00PM – 3:00PM, Wednesday: 2:00PM – 3:00PM
Office: TBD
TA office Hours:
Time: Wednesday 3:00PM –4:00PM
Office: TBD (discussion area)

Are there any prerequisite(s)?

None

Why should you care about data mining for Business Intelligence?

We are currently in the Digital age where every aspect of our life is related to technology and information systems. As of 2018, an estimated 3 million gigabytes of data are generated every minute from the internet alone and this rate is expected to increase each year. Business growth strategy of every company in the planet entails analysis of data from customers, competitors and internal business processes. Being able to transform data into knowledge and actionable insights constitutes a key lever for competitive advantage in business environments. Irrespective of your role is your future employment, data mining for Business Intelligence not only is a coveted skillset, but also makes you an invaluable asset to your organization.

How will this course help you succeed?

This course is designed to introduce you to the systematic science of collecting, processing, modeling and analyzing different types of data. You will be exposed to state-of-the-art procedures, tools and problem areas in data mining and business intelligence. By the end of the course, you will be able to answer the following questions:

1. What are the different problems in Business Intelligence that require applications of Data Mining? What are the problems you cannot solve using Data Mining?
2. How do I systematically carry out a data mining project?
    a. How do I collect, process and maintain data efficiently?
    b. How do I find efficiently information on the internet as well as scholarly resources?
    c. How do I determine inputs and outputs in a data mining model?
    d. How do I conduct basic data mining tasks using tools such as *Microsoft Excel*, *OpenRefine*, *Data Quality Analyzer*, *Tableau*, *IBM SPSS*, *Weka*, *RapidMiner* and *R*.
3. What are the different categories of Data Mining procedures?
    a. How do I determine if your problem requires clustering, classification, rule mining or regression models?
    b. Which are some of the widely used algorithms in clustering, classification, rule mining and regression models?
4. How do I responsibly report your findings from a data mining task?
    a. How do I check the validity of your data mining model?
    b. How do I compare between different models for a given data mining problem?

What are the learning objectives for this course?

Backward design is a successful learner centric method where one sets the goals for learning and then carries out steps to ensure learning takes place systematically. I have enumerated learning goals for this course. We will also discuss these goals in the first class and revise them based on your collective feedback.

1. Learn to examine business problems and determine if a problem can or cannot be solved using Data Mining.
2. Learn to systematically carry out a data mining project. In particular,
    a. Learn to collect, process and maintain data efficiently.
    b. Learn to efficiently find information on the internet including scholarly resources.
    c. Learn to identify inputs and outputs of a data mining model.
    d. Learn to conduct basic data mining tasks using at least three standard tools.
3. Learn about the different types of Data Mining procedures. In particular,

a. Learn to distinguish data mining problems as clustering, classification, rule mining or regression problems.

b. Learn two algorithms used in each type of data mining procedure.

4. Learn to responsibly report findings from a data mining task. In particular,

    a. Learn to check the validity of a data mining model.

    b. Learn to compare between different models for a given data mining problem.

How will this course help me prepare myself for future courses in Management Information Systems?

Successful completion of the course will prepare you for a career in Management Information Systems in the following ways:

1. You will learn about the different concepts in data analysis and you will be able to choose and apply methods to different real-world problems.
2. Select reading materials will inform you on theoretical aspects of data mining that consist of topics from other domains such as mathematics, computer science and business. This experience is useful to appreciate ideas developed in other fields.
3. You will learn to utilize data mining tools as well as get a basic understanding about the programming languages.
4. Through the course project and final presentation, you will get experience in doing collaborative learning, team work and presenting results in front of an audience.

What is my teaching philosophy?

"Give a man a fish and you feed him for a day; teach a man to fish and you feed him for a lifetime." – Chinese proverb.

My version of this proverb for college-level teaching is as follows. Lecture for hours to transfer information into a student's poor overworked brain. Make a student fall in love with the topic and you may have inspired future leaders in innovation.

I believe that the topic of business intelligence is expansive and cannot be covered in one single course. Data Mining in Business Intelligence (DMBI) is a subset of both Data Mining as well as Business Intelligence, where focus is primarily on application of Data Mining methods and tools into solving day-to-day business problems. Data Mining in its own regards is a mature discipline but has a lot of scope for innovations in methods as well as applications.

I am hopeful that the benefits of this course experience continue a life time for each of you. The coursework for DMBI is a combination of theory and practical applications. Though most of the topics will be at an introductory level, at the end of the course, you will master skills of data

inspection, problem assessment and plug-and-use approach to data analysis. Post course completion, I urge students to continue reading relevant DMBI articles and keep updated with innovative applications. This course covers a broad range of topics, and hence the course is structured to facilitate interleaved learning and support retrieval of topic understanding. Collaborative activity in each following class ensures that students interact with their peers and discuss the basic concepts of each topic. Some of you may not be exposed to tools, programming whereas some of you may already have some prior knowledge or experience in data mining. Therefore, this course is designed to provide a multi-faceted learning experience for students in business process understanding, programming, algorithms, problem solving, team building, application and presentation.

What do you have to read before, during and after class?

We do not have one single textbook that we will refer for the entire course. All the materials provided including slides, articles, website links should be sufficient as primary reference materials. Reading assignments will be uploaded in advance in the learning management system and are to be completed before beginning of each class. The readings are mandatory and will enable you to participate in intelligible conversation with your peers during collaborative tasks.

Some of us feel comfortable doing a bit of parallel reading during the course and own a book that summarizes our learning from a course. I would recommend the following book for parallel reading and future reference:

> *Data Science for Business – What you need to know about Data Mining and Data-Analytic Thinking*, by Foster Provost & Tom Fawcett. O'Reilly Media, 2013. ISBN: 978-1449361327.
> - Available in major online e-commerce websites for purchase in the price range (15-25 dollars).

Data Mining and Business Intelligence, like most other disciplines constantly get updated with new theories and applications. The internet is the best repository to keep yourself updated on data mining problems, relevant tools and solution guidelines. We will discuss about strategies to mine the internet on how to mine our data for any given problem. Searching the internet and finding what you are looking in a limited time is a fun skill that can be mastered with practice.

There are some other resources that may serve as optional reading for those who are more curious and hungry for knowledge:

> *Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner*, by Galit Shmueli, Nitin R. Patel, Peter C. Bruce. Wiley, Second Edition; 2010. ISBN: 978-0470526828.

This is a quick read and assists in learning the XLMiner tool and its application to Data Mining for Business Intelligence.

> *Data Mining: Concepts and Techniques*, by Jiawei Han, Micheline Kamber, Jian Pei. Morgan Kaufmann Series, Third Edition, 2011. ISBN: 978-9380931913.

This is a comprehensive textbook on data mining that includes several advanced concepts. The first author is the most influential researcher in Data Mining for all time as per the Microsoft Academic Search.

How will you and I evaluate your progress?

The list of deliverables and the scale for overall grade are given below:

Deliverable		Percentage of grade
Class Participation		5
Reflections	Reflection on Collaborative learning activity - 1,2,3,4,5	10
	Reflection on Collaborative case-study analysis	2
	Course reflection blog	3
Assignments	Assignment 1	10
	Assignment 2	10
Course Project	Course Project – Preliminary analysis	5
	Course project – Final presentation slides	5
	Course project report	10
	Course project – Peer evaluation summary	10
Exams	Mid-term Exam	10
	Final Exam	20
Extra credit	Tweeting feedback to other groups	2

Class Participation

In order for the design of this class to work, we all must be actively involved in the class. Repeated absence will reduce your class participation grade. Always ask questions if you are not clear. It is my duty to ensure that each and every student follows concepts taught in class, and nobody is left behind. I wouldn't know if you do not understand if you do not ask. But if you have a lot of questions, I can always stay after class and teach them to you again. Use the office hours wisely, it is your privilege. Late, incomplete or improper peer assessment by a student will result in reduction in class participation score.

Collaborative learning activities and reflections

Collaborative learning is a great way to learn and involves active engagement in discussions with your peers. We will form groups of 4-5 members to solve case-studies related to the reading topics in class 3 to class 8. Each collaborative learning activity will commence at the beginning of a class and last for around 20 minutes. The activities may involve role-play, creating artifacts such as concept-maps or have a short presentation summarizing initial group discussions.

Each of you is required to summarize your understanding of the topic from the readings and the group activity in a one-page reflection. Reflections are due on the day of the following class after each group discussion (refer schedule). In addition, you need to summarize your learning from the collaborative case-study activity as a one-page reflection, as well as your overall learning statement at the end of the course in the form of a personal blog (min. 1000 words).

Homework Assignments

The effective assimilation of the technical course material requires repeated exposure and practice. The homework assignments are designed to encourage students to adopt the habit of working actively with the course material. The objectives of the homework assignments are to help students understand the course material and to help both the students and the instructor recognize any points that are not yet completely understood. You are encouraged to diligently attempt each homework problem. However, if you get seriously stuck on a problem, make a note of the specific difficulty you are having and move on. You may discuss the homework problems with your classmates before they are due. You can also contact me, or the teaching assistant and we can point you to the directions in case you uncertain on how to proceed. However, ***you should work independently when you actually write up your work***. It is essential that you individually understand the homework assignments in order to do well on the exams. **Assignments are to be uploaded by 11:59am (noon) on d2l for respective due dates**. Late submissions will attract 20% penalty for 24 hours post deadline and will not be graded thereafter.

Course Project

The course project gives you the opportunity to work in groups of five in a data mining project in which you choose your own problem that resembles or is derived from an actual business problem. You apply most of the concepts discussed in the class including data collection, data processing, analysis, model comparison and validation, reporting, team work and project management principles. You will present your preliminary analysis that describes your problem formulation, describes your data and summarizes your solutions strategy for subsequent analysis. The final presentation can include the preliminary analysis, the data mining models, the results and your inference from the analysis. Assessment will have a peer assessment component that sums up individual scores from students who are not in your group (between-group) as well as students who are in your group (within-group). We will discuss this in detail in the first class as well as before the project presentations so that each of you completely understand the peer assessment component. The project report (one per group) should have at least 3500 words and provide a comprehensive analysis of your project. Submit your presentation slides along with the report to receive full grade for the presentation and report.

Exams

The first Exam will count as 10% of your final grade and covers the first part of the course. It will consist of multiple choice questions ad is closed book. The final exam will account for 20% of your final grade. It will include an in-class data mining task and answering a descriptive question. You will need to use your computer for the Final Exam.

We will follow an absolute scoring system for this course as follows:

Overall score	Grade	Student performance
>= 90	A	Excellent
80 – 89.9	B	Very good
70 – 79.9	C	Satisfactory
60 – 69.9	D	Needs significant improvement
40 – 59.9	E	Performance significantly below expectations
< 40	F	Fail

Do we have a schedule for this course?

Yes. Let us plan ahead for an exciting learning experience by listing out set of topics we can discuss in each assigned class for the course. However, we can also revise these when required. The most updated schedule should be available in d2l under the 'Course basics' section. The learning objectives for each class will be shared in the last slide of the previous class as well as summarized in the beginning of each class discussion.

Date	Topics	Assigned	Due
Aug. 21	Introduction to the course		
Aug. 23	- *Data Mining as a mini-world in the universe of Business Intelligence* - *What is and What is not Data Mining* – Brainstorming business problems with the data mining lens		
Aug. 28	- Collaborative learning activity 1 - *A fruitful journey from data to knowledge* – Nuances over the topics of data collection, processing, storage and management	Reflection on Collaborative activity 1	
Aug. 30	- Collaborative learning activity 2 - *Me and my friends formed a rock band* – How to work out a data mining project, efficiently manage time, resource and manpower	Reflection on Collaborative activity 2	Reflection on Collaborative activity 1
Sep. 4	- Collaborative learning activity 3 - *Seeking out internet seers* – Why, how and when to seek information on internet in relation to your data mining projects	Reflection on Collaborative activity 3	Reflection on Collaborative activity 2
Sep. 6	- Collaborative learning activity 4 - *Knowing your variables right* – What are inputs and outputs in a data mining model - How to present a summary of your data	Reflection on Collaborative activity 4	Reflection on Collaborative activity 3
Sep. 11	- Collaborative learning activity 5	Reflection on Collaborative activity 5	Reflection on Collaborative activity 4

	- *Asking the right questions* – Categorizing data mining problems into clustering, classification, rule mining and regression		
Sep. 13	- *Knowing your options* – Summarizing tools and programming languages in data mining. - *The axiom of choice* – Choosing between tools and programs for data mining problems		Reflection on Collaborative activity 5
Sep. 18	***Hands on session*** – Using Microsoft Excel data mining add-ons, OpenRefine, Tableau, IBM SPSS, Weka, RapidMiner.	Assignment I	
Sep. 20	***Hands on session*** – Using *Microsoft Excel data mining add-ons, OpenRefine, Data Quality Analyzer, Tableau, IBM SPSS, Weka, RapidMiner*.		
Sep. 25	***Hands on session*** – Introduction to *R, Java, Python, Spark* and other programming languages for Data Mining.		
Sep. 27	Midterm exam - MCQs		
Oct. 2	- Midterm exam solutions discussion - Data visualization and exploratory analysis		Assignment I
Oct. 4	*Getting to know clustering better* – Algorithms and developments		
Oct. 9	*Getting to know classification better* – Algorithms and developments		
Oct. 11	*Getting to know regression better* – Algorithms and developments		
Oct. 16	*Getting to know rule mining better* – Algorithms and developments		
Oct. 18	- *And they say it works* – Assessing model fit, model validation and model comparison. - *Etiquette of telling your mining stories* – Responsible reporting of findings from data mining analysis		
Oct. 23	*Now your turn* – Presentation of preliminary analysis results of group projects	- Assignment II - Callouts for tweet-based feedback from audience	
Oct. 25	*Now your turn* – Presentation of preliminary analysis results of group projects	Callouts for tweet-based feedback from audience	
Oct. 30	*Now your turn* – Presentation of preliminary analysis results of group projects	Callouts for tweet-based feedback from audience	

Nov. 1	- *Responsible data mining* – Privacy and mining ethics		Assignment II
Nov. 6	***Hands on session*** – Case-studies in clustering		
Nov. 8	***Hands on session*** – Case-studies in classification		
Nov. 13	***Hands on session*** – Case-studies in regression		
Nov. 15	***Hands on session*** – Case-studies in rule mining		
Nov. 20	Collaborative case-study analysis. Each student will be assigned to two groups. In the first group, students will discuss about a case-study and come up with a solution. Then each student summarizes their first group solution to their respective second groups	Reflection on Collaborative case-study analysis	
Nov. 22	*Now your turn* – Final group project presentations (Day 1)	Peer assessment form for Day 1 groups	Reflection on Collaborative case-study analysis
Nov. 27	*Now your turn* – Final group project presentations (Day 2)	Peer assessment form for Day 2 groups	Peer assessment form for Day 1 groups
Nov. 29	*Now your turn* – Final group project presentations (Day 3)	Peer assessment form for Day 3 groups	Peer assessment form for Day 2 groups
Dec. 4	*Now your turn* – Final group project presentations (Day 4)	Peer assessment form for Day 4 groups	Peer assessment form for Day 3 groups
Dec. 6	Summarizing our learning		Peer assessment form for Day 4 groups
Dec. 10	Final Exam – Open book, Open laptop		- Course reflection blog - Peer evaluation form Group project report (one per group) Group project presentation slides

How can we ensure a healthy learning environment?

Be punctual and do not be a chronic procrastinator. If you have obligations that conflict with exam or assignment due dates, you should make arrangements with me as soon as possible. Feel free to email homework assignments if necessary.

Missed Exams/Assignment. Make-up exams will be given only in extraordinary circumstances. If you expect to miss an exam or to be unable to meet another requirement, please discuss this with the instructor **before** the scheduled date.

Score/Grade appeals. It is important to recognize that a grade reflects another's judgment of your work. In this sense, all grading is subjective. Score changes are at the discretion of the instructor. It is important to understand that your score may go up **or down** based upon a complete review of the work in question. It is usually the case that changing a few points on an assignment rarely makes a difference in the final grade. Time is much better spent discussing and clarifying the information content presented in the course. However, if you would like to appeal your grade, you must submit the appeal in writing to me within 48 hours, but no earlier than 24 hours, of receiving the graded work.

Academic Conduct. You are expected to abide by the **Eller Student Integrity Oath:**

As a student of the Eller College of Management I promise:

To foster an honorable and academic environment for current and future students

To uphold the principles outlined in The University of Arizona Code of Conduct and Code of Academic Integrity

To neither engage in nor tolerate: cheating, plagiarism, lying, or stealing and

To accept the lifelong responsibility of upholding the highest levels of honesty, integrity, and respect for others

I accept this oath and make these promises freely and upon my honor.

Incomplete Grades Policy. A grade of Incomplete may be awarded to students who have a legitimate reason for needing additional time to complete a course. Legitimate reasons include emergencies or extenuating circumstances that prevent a student from completing the course requirements within the normal time frame. Students must initiate the request for an incomplete prior to the end of the semester and must be in good standing in the course. In no case will a grade of incomplete be awarded to someone seeking more time to master the course material in order to improve their grade.

***Accommodations for Students with Special Needs*.** If you anticipate the need for reasonable accommodations to meet the requirements of this course, you must register with the Disability Resource Center and request that the DRC send me official notification of your accommodation needs as soon as possible. Please notify me by the end of the first week of class and plan to meet with me by appointment or during office hours to discuss accommodations and how my course requirements and activities may impact your ability to fully participate.

Last day to drop online with a "W" is Oct 29, 2019. After Oct 29th, course withdrawals may be processed only with the approval of the student's college Dean on a Late Change Petition now through November 17th. Students must have an extraordinary reason for approval. (Please note: effective Fall 2014, there is an 18 unit cap on the number of W units).

Inclusive Excellence. Inclusive Excellence is a fundamental part of the University of Arizona's strategic plan and culture. As part of this initiative, the institution embraces and practices diversity and inclusiveness. These values are expected, respected and welcomed in this course.

This course supports elective gender pronoun use and self-identification; rosters indicating such choices will be updated throughout the semester, upon student request. As the course includes group work and in-class discussion, it is vitally important for us to create an educational environment of inclusion and mutual respect.