

MIS 6356: BUSINESS Analytics with R

Fall, 2016

Section	MIS6356.501, Monday: 7:00pm - 9:45pm, JSOM 11.210 MIS6356.502, Wednesday: 7:00pm - 9:45pm, JSOM 12.222
Instructor	Gregory G. MacDonald, PhD – Office: JSOM 3.604 Office Hours: Monday 6:00 pm - 7:00pm, Wednesday 6:00 pm – 7:00pm. <i>Please use eLearning for emailing the instructor and the TA</i>
TA	TBD, will notify all via email

Strongly Recommended

- 1) *Introduction to Data Mining*, by Pang-Ning Tan, Michael Steinbach and Vipin Kumar. Pearson, ISBN 0-321-32136-7.
- 2) *Data Mining with Rattle and R*, by Graham Williams. Springer, ISBN 978-1-4419-9889-7 (also free on the web)

Optional

- 3) *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 (**available as an eBook from the UTD library**)
- 4) *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management, 3e*, by Gordon Linoff and Michael Berry. Wiley, ISBN-10: 0470650931, ISBN-13: 978-0470650936 (**available as an eBook from the UTD library**)

Additional Reading	Lecture notes and additional learning material will be posted in eLearning as the semester proceeds.
Required Software	R, RStudio, and “Rattle” (all free!). Instructions on how to install these are provided as separate handouts. Use of other tools also encouraged (Weka, RapidMiner, SAS).
Laptop Computer	This course is <i>heavily</i> hands-on with in-class exercises. <i>Students are <u>required</u> to bring laptops to each class thereafter</i> (except for exam days).
Course Description	Most organizations are data rich and information poor. For instance, by 2014 Facebook is adding 600 TB of new data daily. The rate at which data has been accumulating has only drastically increased with newer sources like social networks and RFID. These large volumes of data potentially could reveal useful information about the target of interest, such as customers and products. The primary objective of this course is to introduce you to various techniques available to extract useful information (business intelligence) from the large volumes of data an organization might possess. At the end of the semester, you will not only appreciate the substantial opportunities that exist in the BI realm,

but also learn techniques that will allow you to exploit these opportunities. The course will cover general concepts in the BI field, along with popular BI techniques including association rules, clustering and classification. The focus will be on how the techniques are to be used, and the details of the methodologies will be covered only to the extent necessary to understand when and how each technique can be used.

- Learning Objectives**
1. To gain a general understanding of business intelligence / data mining, and to appreciate the data rich environment of today's global economy.
 2. To gain a practical understanding of key methods integral to data mining.
 3. To gain an understanding of when to use which technique.
 4. To become aware of some current trends in the use of BI.
 5. To gain the intellectual capital required for business analytics services.
 6. To experience one of the more popular open source software choices.

Groups This class involves several group assignments and one group project. Groups should comprise 4-5 students, and should be formed ASAP. Once formed, altering the groups will not be possible except in very special circumstances.

Group Project TBD

Assessment	Group Homework Assignments	30%
	Exam 1	25%
	Exam 2	25%
	Group Project	20%

Final Grade ($\geq 90\%$) \Rightarrow A, ($\geq 87.5\%$ & $< 90\%$) \Rightarrow A-,
 ($\geq 85\%$ & $< 87.5\%$) \Rightarrow B+, ($\geq 80\%$ & $< 85\%$) \Rightarrow B, ($\geq 77.5\%$ & $< 80\%$) \Rightarrow B-,
 ($\geq 75\%$ & $< 77.5\%$) \Rightarrow C+, ($\geq 65\%$ & $< 75\%$) \Rightarrow C,
 ($< 65\%$) \Rightarrow F

Comments *You are responsible for any announcements made in class or through eLearning, including changes to the schedule. You are also responsible for material distributed in class or through eLearning (so check eLearning regularly).*

All assignments and project reports must be submitted through eLearning. Alternative submission methods, e.g., by paper, by e-mail or on disk/USB drive, is not acceptable unless prior permission of instructor is obtained. Submissions after the deadline will not be accepted, and accordingly a grade of zero will be automatically applied for a missing submission after the deadline.

All exams are *closed-book, closed-notes* and require *individual-effort*. Unless under extreme circumstances, make-up exams will *not* be arranged. A make-up exam will include significant discussion, essay, and short answer questions.

The participation portion of your grade is affected by your involvement in the classroom discussions and exercises, as well as by your contribution to your group (we will conduct in-group peer evaluation toward the end of the semester). *In extreme cases*, such as when a student contributes significantly less to group

assignments or projects throughout the semester, a penalty will be imposed in the form of directly deducting assignment and/or project points.

Tentative Schedule

* Topics and timelines are subject to change at the discretion of the professor

* Bring your laptop to every class except for exam days

* Check eLearning regularly for announcements, lecture notes, and due dates

* Due time is right before the corresponding class for your section

<u>Week</u>	<u>Topic(s) for Sections 501/502</u>	<u>Assignments</u>
Aug 22	Course Introduction, Problems w/ Data, Model Assessment	Form groups, get RStudio, R, and "Rattle"
24	Course Introduction, Problems w/ Data, Model Assessment	
Aug 29	Classification: Decision Trees, Introduction to "Rattle"	Start on Homework 1
31	Classification: Decision Trees, Introduction to "Rattle"	Start on Homework 1
Sep 05	No class (Labor Day)	
07	Classification: Nearest Neighbors, Artificial Neural Networks	
Sep 12	Ensemble Methods and the Class Imbalance Problem (RF)	Homework 1 due
14	Ensemble Methods and the Class Imbalance Problem (RF)	Homework 1 due
Sep 19	Association Analysis: <i>Apriori</i> Principle, Itemsets, Support/Confidence	
21	Association Analysis: <i>Apriori</i> Principle, Itemsets, Support/Confidence	
Sep 26	Similarity/Dissimilarity Measures, Missing Data, MDS, Review	Start on Homework 2
28	Similarity/Dissimilarity Measures, Missing Data, MDS, Review	Start on Homework 2
Oct 03	Exam 1 (Mid-term)	
05		
Oct 10	Cluster Analysis: Types of Clustering, Project Disc.	Homework 2 due
12	Cluster Analysis: Types of Clustering, Project Disc.	Homework 2 due
Oct 17	Neural Networks and SOMs	
19	Neural Networks and SOMs	
Oct 24	Support Vector Machines, Project Checkpoint	Start on Homework 3
26	Support Vector Machines, Project Checkpoint	Start on Homework 3

Oct 31		
Nov 02	Text Analytics/Mining Fundamentals, Novelty Detection Text Analytics/Mining Fundamentals, Novelty Detection	Homework 3 due Homework 3 due

Nov 07 09	Misc. Topics and Review	
-----------------	-------------------------	--

Nov 14	Exam 2 (Final Exam)	
16	Exam 2 (Final Exam)	

Nov 21	No Class (Fall Break)	
23	No Class (Fall Break)	

Nov 28	Project Presentations	
30	Project Presentations	

Dec 05	Project Presentations	
07	Project Presentations	

Dec 12 / Dec 14 Make-Up

MIS 6356 BI Software and Techniques, Fall 2016

Gregory G. MacDonald

Comet Creed

This creed was voted on by the UT Dallas student body in 2014. It is a standard that Comets choose to live by and encourage others to do the same:

“As a Comet, I pledge honesty, integrity, and service in all that I do.”

UT Dallas Syllabus Policies and Procedures

The information contained in the following link constitutes the University’s policies and procedures segment of the course syllabus.

Please go to <http://go.utdallas.edu/syllabus-policies> for these policies.

Descriptions and timelines in this syllabus are subject to change at the discretion of the professor.

