McCoy College of Business Administration
TEXAS STATE UNIVERSITY
Department of Computer Information Systems and Quantitative Methods
QMST 4373C—TOPICS IN DATA ANALYTICS
SPRING 2017

INSTRUCTOR: Dr. Tahir Ekin   **TEL:** 512-245-3297
OFFICE: McCoy 451
OFF. HOURS: TR: 12:35-1:50 pm & by appointment
M: 3:50-6:20 pm (Round Rock)
E-MAIL: t_e18@txstate.edu

SCHEDULE

Section	Time	Room
251	TR 2:00 -3:20 pm	McCoy 334

COURSE DESCRIPTION:

This course covers the process of transforming big data into information for making decisions. The topics include introduction to data science, analytics and advanced data mining algorithms, and challenges related to analyzing business data. Students will learn how to use software and conduct data analysis.

COURSE MATERIALS:

Required:
- **Access to TRACS:** Lecture notes and required readings will be posted in TRACS
- **RStudio Desktop** (Free open-source software)
- **Tableau** (Instructor will send an email that provides free access)
- **Microsoft Office (Word, Excel)**

    Additional reading will be assigned as necessary.

Recommended:
- **Access to a laptop** (If you do not have a personal laptop, you can check out a laptop from the McCoy 338 open lab by presenting your student ID card.)

Complementary textbook references:
- Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani (2013) *An Introduction to Statistical Learning with Applications in R*. Available at: http://www-bcf.usc.edu/~gareth/ISL/getbook.html

- Zhao, Yanchang (2013) *R and Data Mining: Examples and Case Studies*. Available at: http://www.rdatamining.com/

- Rajaraman, Anand (2014) *Mining of Massive Datasets*, Jure Leskovec and Jeffrey D. Ullman: http://i.stanford.edu/~ullman/mmds/book.pdf

- Trevor Hastie, Robert Tibshirani, Jerome Friedman (2009) The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Available at: http://statweb.stanford.edu/~tibs/ElemStatLearn/

- EMC Education Services (2015). Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data. Wiley. ISBN: 978-1-118-87613-8

- Open Intro Statistics 3rd edition (free) https://www.openintro.org/stat/textbook.php

LEARNING OBJECTIVES

- Understand the concepts of data science
- Apply supervised and unsupervised methods
- Evaluate the performance of algorithms
- Employ appropriate software for data science/data mining.
- Interpret results from algorithms
- Create a quantitative analysis report with the necessary information to make decisions

COURSE POLICIES

A. INSTRUCTIONAL METHODOLOGY:

Lecture, interactive discussion, instructor / group / individual problem solving

B. ATTENDANCE:

Regular class attendance is expected and necessary, and in-class discussion is part of the course assessment. Attendance is required for the guest talk which will be given by an industry analytics practitioner. In case of absence, it is the responsibility of the student to find out what work was missed, and to obtain class notes from another student, should a class be missed. Poor course attendance will result in the instructor not considering borderline grade situations when final grades are assigned.

C. COURSE REQUIREMENTS:

1. **Required reading**: Prior to every session the student needs to review the material discussed during the previous class. For every lecture, corresponding lecture slides are listed in the course outline. Relevant readings will be posted in TRACS Resources and Forums.

   Group Formation: Every student needs to be a part of 2-3 people group. Working in groups will help you to come up with different perspectives. There will be a peer-evaluation at the end of the semester, so be aware that your grade out of group based requirements can be affected by your contribution. Each group is required to email the group members to the instructor by **January, 31st, 2017 2:00 PM.** One email per group is sufficient.

2. **Company/Industry Presentation:** This is group based, and will be 20% of the overall grade. You are expected to select (~2) topics of company/industry, investigate their data mining/analytics practice and present it to your classmates. You will be assessed with respect to organization, content, delivery and professionalism. The presentation topics and respective dates are listed in syllabus. The presentations are expected to be 20-30 minutes. Each group is expected to submit their choice to TRACS/ Forum/ Company-Industry Presentation by **February 2nd, 2017 2:00 pm**.

3. **Class/Forum Discussion:** This is individual. You are expected to be an active contributor to the in-class and forum discussions. Each student is expected to be present and active part of these discussions, and this will be 10% of the overall grade. There will be a discussion during/after each company/industry presentation. The students are also expected to actively engage in discussion using TRACS/Forums. The instructor may facilitate the discussion by posting in the Forum, and the related discussion can take place either in Forums or in-class. Students are advised to check Forums at least 3 times a week, before and after each class. In addition, attendance is required for the guest talk which will be given by an industry analytics practitioner, and will affect the grade for this component.

   All course related questions should be posted in TRACS-Forums-Course Material Related Questions instead of communicating via email so that everyone can benefit from the exchange. There is another discussion forum created in which you can collaborate with each other regarding your profiles and class matters. For personal issues and confidential matters, you are more than welcome to email the instructor.

   Your grade will be assessed based on the level of your effort, engagement and contribution to discussion. A student who is present in class (who may have absences with verified excuses), engaged in class discussion and reasonably active in the TRACS Forum discussion is expected to get full credit.

4. **Project**: The project is group based. You are expected to select a topic of your preference, explore data, utilize a data mining tool with the statistical software and communicate your findings via well-written report and a presentation. Periodically, you will be asked to complete tasks. There will be certain check points in project evaluation, which are listed in the outline page of the syllabus. For the final report, a hard copy (print-out) of the written report that summarizes the results and includes the supporting graphs is due to the beginning of the lecture on the due date. You will also communicate your findings to your classmates in a presentation. The due date of the final reports and the date for the presentations are listed in the outline page of this syllabus. The project will account for 30 % of your overall grade.

   You need to be present during each checkpoint and work coherently within your group. Please see Project Guidelines page of this syllabus for more details.

5. **Quizzes**: These quizzes are individual. In the outline page of the syllabus, the dates are listed. There are 2 quizzes. Each quiz score will count for 10 % of your overall grade, for a total of 20 % of the overall grade. More details about the format of the quizzes will be provided.

6. **Final Exam:** The final exam is individual. This will be administered with respect to the final exam schedule of the university. Overall, this will account for 20 % of your overall grade. More details about the format of the final exam will be provided.

Test conduct: Breaking any of the following exam rules could result in a grade of zero (0) for the test or the entire course.

- The proctor of the test may ask you to show your student ID card.
- You may not consult other people, classmates during the exam.
- Any sort of communication is not allowed.

Absence: If an absence is due to medical reasons, emergencies or school related activities, the student is required to provide appropriate written documentation within three business days of the absence. An absence that cannot be documented is unlikely to be approved. Unless the absence is due to approved reasons, missed test will result with a grade of zero.

Make-up Examinations: Read the schedule and make your arrangements. As a rule, make-up exams will not be given unless the student has a **valid and verifiable** excuse that result in an approved absence. If absence is approved, the instructor will decide the date, place and content of the make-up examination. Exams missed due to school sponsored activities, such as athletics, etc., will be excused, per university policy.

Grade Evaluation: The student's overall performance will be determined using the weights below from each respective area:

Company/Industry Presentation	**20 % (Group)**
Class/Forum Discussion	**10 % (Individual)**
Quiz 1	**10% (Individual)**
Quiz 2	**10% (Individual)**
Project	**30% (Group)**
Final Exam	**20% (Individual)**

The student's total score is the final weighted average score and the letter grade will be assigned according to the following table:

Range	Grade
[90%, 100%]	A
[80%, 90%)	B
[70%, 80%)	C
[60%, 70%)	D
[0%, 60%)	F

The instructor holds the right to consider overall performance and make adjustments to the letter grades that would be in the favor of students. The grades will be posted in TRACS. Borderline grade situations will only be considered if the student has excellent attendance, and has observed the civility rules presented on section D.

D. CLASSROOM CIVILITY:

Disruptive behavior in the classroom is prohibited in Section 2.02 of Texas State's Code of Student Conduct and includes behavior that substantially or repeatedly interferes with the conduct, instruction, and education of a class. The complete Conduct of Classes policy is available at http://www.provost.txstate.edu/pps/policy-and-procedure-statements/4-teaching/pps4-02.html.

Students are expected to come to class each day prepared to participate in class discussions and activities. It is expected that all students will exhibit **professional behavior** during the class. That includes being in class on-time, being prepared and attentive, respecting others and their ideas, turning off cell phones, and other electronic devices, **not surfing the internet** or **checking email**, and staying in class (unless you have an emergency or have cleared it in advance with the professor). The instructor will take appropriate disciplinary action against violators of classroom civility, including the removal of offending parties from the class and reporting disruptive students to the Dean of the college and Student Justice.

Students **must** attend the class section for which they are officially registered and complete all exams, assignments, and other course-work in that section.

More information regarding classroom responsibilities can be found at **Students' Responsibilities on Advising and Learning found** at http://advising.mccoy.txstate.edu/about/learningpolicy.html.

- Please note that NO FOOD OR DRINKS are allowed in any McCoy Hall classrooms.
- If you need to use the cell phone, leave the classroom until you are done.
- If you need to text message, leave the classroom until you are done.
- No sleeping during class.
- No reading or working on unrelated material (e.g., newspaper, material from other classes).

E. OTHER:

- **Course material, test announcements, report announcements, learning videos and presentation slides will be posted on TRACS**. Access to TRACS, at least twice a week is recommended.

- You will need to come to class in order to receive the lecture and to participate. "Personalized" lectures are not delivered during office hours. This is **not** a distance-learning course.

UNIVERSITY/COLLEGE POLICIES:

A. **DROP:** Dropping means that the student will remain enrolled in at least one hour in the current semester. A "W" will be automatically assigned if the drop procedure is completed on or before **11:59 p.m. on March 28, 2017**. After this deadline the student will be unable to drop individual classes and will receive the grade earned in the course (see AAPPS 4.07 for a list of grades). It is suggested that students consult the instructor prior to dropping from the class.

WITHDRAWAL: Withdrawal means that the student is going to <u>zero hours</u> for the current semester. A "W" will be automatically assigned if the withdrawal procedure is completed on or before **11:59 p.m. on <mark>March 28, 2017</mark>**. After this deadline, the student may withdraw on or before **5:00 p.m. on <mark>April 20, 2017</mark>**. If the student is passing the class on the official date of withdrawal, a "W" grade will be assigned. If the student is failing the class on the date of withdrawal, a "U" grade will be assigned.

B. **ACADEMIC HONESTY**: Submission of any work for a grade for which unauthorized help has been received is termed academic dishonesty and will be grounds for a failing grade in the course. "Unauthorized" is a term used here to designate stealing, copying (with or without permission), collaboration with other individuals, or sharing programming code outside of sanctioned group activities. Students are strongly encouraged to refer to the Texas State student handbook, available at http://www.dos.txstate.edu/handbook.html for policies related to academic dishonesty. **This instructor views any such act as a clear violation of ethical standards and will take appropriate disciplinary and punitive action.**

C. **HONOR CODE:** All students are required to abide by the Texas State University Honor Code found in UPPS 07.10.01 under attachment I. The pledge for students states:

> **Students at our University recognize that, to insure honest conduct, more is needed than an expectation of academic honesty, and we therefore adopt the practice of affixing the following pledge of honesty to the work we submit for evaluation:**
>
> **I pledge to uphold the principles of honesty and responsibility at our university.**

D. **FINANCIAL AID:** Federal regulations require students to meet certain minimum academic and attendance standards in order to remain eligible for financial aid assistance. Other program-specific requirements may also exist. Additional information is available at www.finaid.txstate.edu.

E. **STUDENTS WITH DISABILITIES:** A student with a disability may require an accommodation(s) to participate in the course. They must contact the instructor as soon as possible, typically within the first two weeks of the semester. They will be asked to provide documentation from the Office of Disability Services (ODS) at that time. Failure to contact the instructor in a timely manner will delay any accommodations they may be seeking. Ongoing care by a physician does not automatically qualify you as an ODS special needs student. Students are required to file paperwork for accommodations with ODS each semester. Accommodations granted one semester do not automatically carry forward to the next. See UPPS No. 07.11.01 for additional information.

PROJECT GUIDELINES

(1) GROUP FORMATION

One member of the group is required to email the instructor the the group members (2-3 people) by ==**January, 31st, 2017 2:00 PM.**== One email per group is sufficient. Grade 1 or 0

(2) PROJECT OUTLINE:

In this part of the project, you will prepare a one-page outline describing the data set and the problem to be analyzed. You are expected to use an analytical method of choice. You will list the variables of interest and their relevance in the analysis, specify the method of choice. Be ready to show your data. In case you need some ideas, a number of data sources will be posted in TRACS. The due date for the project outline is, ==**March, 30th, 2017 2:00 PM**==. There will be a discussion with each group. Grade 1 or 0

(3) PROJECT MID-REVIEW

At this checkpoint, you are required to have done a preliminary analysis. The instructor will answer questions and will help with the potential challenges. There will be a discussion with each group in class time, ==**April, 20th, 2017 2:00 PM**==. Grade 1 or 0

(4) FINAL REPORT

A sample report may include the sections such as

    a) Executive Summary of your findings which will be discussed in detail in your report
    b) Introduction that describes your problem and your approach
    c) Data description including a discussion of the data pre-processing and descriptive statistics
    d) Steps of your analysis
    e) Model selection and evaluation steps: validation analysis
    f) A conclusion that summarizes your findings

The maximum number of pages for the report is 15 pages. The print out of the final report is due to the beginning of the final day of classes, ==**April 27th, 2017 2:00 PM**==. Grade out of 20.

(5) PRESENTATION

You are expected to present your findings and a summary of your report at the end of the semester, ==**April 27th, 2017 2:00 PM.**== You will be assessed with respect to organization, content, delivery and professionalism. Grade out of 10

(6) PEER EVALUATION

The feedback of your group members about your performance in your group will affect your final grade. Your final grade. Grade between (0.5 and 1) out of 1

Your project grade will be the product of the three grades from (1), (2), (3), (4+5) and (6).

COMPANY/INDUSTRY PRESENTATION TOPICS

1) Feb 14: Social Media/Marketing Analytics in general with potential emphasis on a company of choice such as Google, Yahoo, Linkedin, Tinder, Snapchat, Baidu (any such company other than Facebook/Twitter)
2) Feb 14: Social Media/Marketing Analytics at Facebook (or similar) with an emphasis on newsfeed, modeling activity such as number of likes, marketing mix
3) Feb 21: Social Media/Marketing Analytics at Twitter (or similar) with an emphasis on text mining
4) Feb 23: Retail Analytics: Walmart or a similar company
5) Feb 23: Retail Analytics: Target or a similar company
6) Feb 28: Entertainment Analytics: Netflix or a similar company
7) Feb 28: Entertainment Analytics: Pandora or a similar company
8) Mar 7: Healthcare Analytics in general with an extra emphasis on a company of choice such as Truven Health or a specific group of applications such as clinical analytics
9) Mar 7: Audit Analytics: Government/company which may be IRS, Medicare or similar
10) Mar 30: Information Security Analysis in general with a potential focus on recent issues such as Sony, Target, US Elections or with an emphasis on an information security company
11) Apr 4: Big Data Analytics: IBM Watson with potential discussion of smart cities
12) Apr 4: Big Data Analytics: Internet of Things
13) Apr 11: Transportation Analytics: Uber or a similar company with potential discussion of self-driving cars
14) Apr 11: Transportation Analytics: Fedex or a similar company
15) Apr 11: Transportation Analytics: Amazon
16) Apr 20: Future of Analytics

COURSE OUTLINE

(This is a tentative schedule. Additional reading will be assigned as necessary. LS= Lecture Sets)

WEEK	DATE	DAY	TOPIC	MATERIAL	DUE
1	17-Jan	T	Introduction	LS1	
1	19-Jan	R	Analytics Overview	LS1	
2	24-Jan	T	Algorithms and Applications	LS2	
2	26-Jan	R	Introduction to R: open lab session	LS2	Install Rstudio/R
3	31-Jan	T	Data/Variables	LS3	Group Formation
3	2-Feb	R	Data preprocessing/transformation	LS3	Presentation Choice
4	7-Feb	T	Explanatory data analysis	LS4	
4	9-Feb	R	Explanatory data analysis	LS4	

WEEK	DATE	DAY	TOPIC	MATERIAL	DUE
5	14-Feb	T	Social Media/Marketing Analytics	LS5	
5	16-Feb	R	Social Media/Marketing Analytics: Regression	LS5	
6	21-Feb	T	Social Media/Marketing Analytics: Text Mining	LS5	
6	23-Feb	R	Retail Analytics	LS6	
7	28-Feb	T	Entertainment Analytics	LS6	
7	2-Mar	R	Retail and Entertainment Analytics: Association/Classification	LS6	
8	7-Mar	T	Healthcare/Audit Analytics	LS7	
8	9-Mar	R	Healthcare/Audit Analytics: Clustering/Bayesian methods/Outlier detection	LS7	
9	14-Mar	T	No Class-Spring Break		
9	16-Mar	R	No Class-Spring Break		
10	21-Mar	T	R review: open lab session		
10	23-Mar	R	Quiz 1		
11	28-Mar	T	Sales Analytics: Advanced Visualization	LS8	Install Tableau
11	30-Mar	R	Information Security Analytics: Outlier detection	LS9	Project Outline
12	4-Apr	T	Big Data Analytics	LS10	
12	6-Apr	R	Big Data Analytics: Databases and Parallel Computing	LS10	
13	11-Apr	T	Transportation Analytics	LS11	
13	13-Apr	R	Transportation Analytics: Neural Networks/Deep Learning/Optimization	LS11	
14	18-Apr	T	Quiz 2		
14	20-Apr	R	Future of Analytics and Project Discussion		Project Mid-Review
15	25-Apr	T	Guest lecture: Igor Frolow, USAA		Attendance
15	27-Apr	R	Project Presentations		Project Report/ Presentation
16	4-May	R	FINAL EXAM 2 -4:30 PM		