

**McCoy College of Business Administration**  
**TEXAS STATE UNIVERSITY**  
**Department of Computer Information Systems and Quantitative Methods**  
**QMST 3339—DATA MINING AND VISUALIZATION**  
**FALL 2016**

**INSTRUCTOR:** Dr. Tahir Ekin **TEL:** 512-245-3297  
**OFFICE:** McCoy 451  
**OFF. HOURS:** TR 10:55 am-12:20 pm & TR 1:55 pm- 3:00 pm & by appointment  
**E-MAIL:** t\_e18@txstate.edu

| <b>SCHEDULE</b> |                    |             |
|-----------------|--------------------|-------------|
| <b>Section</b>  | <b>Time</b>        | <b>Room</b> |
| 001             | TR 09:30 -10:50 am | McCoy 224   |

**COURSE DESCRIPTION:**

This course introduces data mining concepts and practical skills for applying data mining techniques to solve business problems. It emphasizes data visualization and data analysis algorithms (e.g., prediction, classification, clustering), systematic evaluation and model assessment for big data sets.

**COURSE MATERIALS:**

**Required:**

- **Textbook (e-book and learning software):** EMC Education Services (2015). Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data. Wiley. ISBN: 978-1-118-87613-8
- **Access to a laptop with RStudio every class**
- **Access to TRACS: Lecture notes will be posted in TRACS**
- **R programming language, RStudio Desktop**
- **Microsoft Office (Word, Excel)**

Additional reading will be assigned as necessary.

**Recommended:**

**Complementary 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/>.

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

## **LEARNING OBJECTIVES:**

What you will acquire in this course is the ability to:

- Learn basic and advanced visualization techniques in data mining
- Understand how to choose algorithms for different analysis tasks. •
- Apply basic algorithms for supervised and unsupervised learning. •
- Evaluate the performance of the data mining algorithms. •
- Employ appropriate software for data mining/machine learning. •
- Interpret results from data mining algorithms. •
- Create a quantitative analysis report with the necessary information to make decisions •

## **SKILLS AND KNOWLEDGE YOU ARE EXPECTED TO BRING INTO THIS COURSE:**

- The ability to solve word problems •
- Basic understanding of probability and statistics concepts •
- The ability to read algebra for meaning •
- The ability to solve algebra equations •
- Knowledge of computer operating system (Windows or UNIX or LINUX), wordprocessing •

## **PREREQUISITES:**

QMST 2333 or equivalent.

## **COURSE POLICIES:**

### **A. INSTRUCTIONAL METHODOLOGY:**

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

### **B. ATTENDANCE:**

Regular class attendance is expected and necessary. 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 and the textbook are listed in the course outline.
2. **Team formation:** Every student needs to be a part of 3-5 people team (group). The homeworks, project and industry presentation are group-based requirements. 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. One member of the group is required to email the instructor the team members by **September, 6<sup>th</sup>, 2016 5:00 PM**. One email per group is sufficient.
3. **Homeworks:** These homeworks are group-based. However, all team members are expected to understand them thoroughly as some portions of the homeworks may be used in the quizzes. 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. These due dates are listed in the outline page of this syllabus.

There are a total of 5 assignments. Your submission that results with your lowest grade will not be considered for the final grade computation. And the rest of each four will be counted for 5 % of your overall grade. The homeworks in total will account for 20 % of your overall grade.

4. **Quizzes:** These quizzes are individual. In the outline page of the syllabus, the dates are listed. There are a total of 3 quizzes. Each quiz will be counted for 10 % of your overall grade. The quizzes in total will account for 30 % of your overall grades.

If an absence is due to **medical reasons or school related activities**, the student is required to provide appropriate written documentation within three business days of the absence. And the instructor holds the right to determine an action. Otherwise, a missed quiz will result with a grade of zero.

5. **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. 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. The due date is listed in the outline page of this syllabus. This will account for 20 % of your overall grade. You need to be present during each checkpoint and work coherently within your team to be eligible for 20 % of your overall grade. Please see Project Guidelines page of this syllabus for more details.
6. **Industry Presentation:** The industry presentation is group based. You are expected to select a topic and industry/company of your preference, investigate their data mining/analytics practice and present it to your classmates. The presentation date is listed in the outline page of this syllabus. In addition, you are required to be in attendance for the guest lecture that will be given by an industry representative.

If an absence is due to **medical reasons or school related activities**, the student is required to provide appropriate written documentation within three business days of the absence. Otherwise, this will result with a grade of zero. Overall, the industry presentation and attendance of the guest lecture will account for 10% of your overall grade.

7. **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.

**Test (quiz and final exam) 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 exam may ask you to show your student ID card.
- **You may not consult other people, classmates during the exam.**
- No talking, whispering, or communication of any sort with anyone else.
- Disruptive behavior will NOT be tolerated.

**Make-up Examinations:** Read our schedule and make your arrangements. As a rule, make-up exams will not be given unless the student has a **valid and verifiable** excuse. An approved absence is one due to **medical reasons or emergencies**. The instructor will determine what constitutes an emergency. Appropriate written documentation, pertaining to the absence, is required and will be provided by the student in order for the instructor to document and determine whether or not the absence is approved. An absence that cannot be documented is unlikely to be approved. If 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 final grade will be determined using

|                              |  |
|------------------------------|--|
| <b>Quizzes</b>               | <b>30% (10 % each)</b>                         |
| <b>Homeworks</b>             | <b>20% (5% each, lowest one to be dropped)</b> |
| <b>Project</b>               | <b>20%</b>                                     |
| <b>Industry Presentation</b> | <b>10%</b>                                     |
| <b>Final Exam</b>            | <b>20%</b>                                     |

Your grade will be based on your final weighted average score and the letter grade will be assigned according to the following table:

| <b>Range</b> | <b>Grade</b> |
|--------------|--------------|
| [90%, 100%]  | A            |
| [80%, 90%)   | B            |
| [70%, 80%)   | C            |
| [60%, 70%)   | D            |
| [0%, 60%)    | F            |

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.

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.

#### **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/pp4-02.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 October 30, 2016**. 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 zero hours for the current semester. A "W" will be automatically assigned if the withdrawal procedure is completed on or before **11:59 p.m. on October 30, 2016**. After this deadline, the student may withdraw on or before **5:00 p.m. on November 29, 2016**. 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](http://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.

## **Financial Aid Information**

---

### **Non-Attendance and Financial Aid**

If you are a Pell Grant, Iraq-Afghanistan Service Grant (IASG) or TEACH Grant recipient, federal regulations require you to have begun attending the courses for which you are enrolled and receiving these grants. If on the census date roster (e.g., 12<sup>th</sup> day of each fall and spring semester) you are reflected as not attending a course, you are assumed (for financial aid purposes) not to have begun attendance for that course. Your grant will then be adjusted or cancelled based on the courses you have actually begun attending.

### **Unofficial Withdraws and Financial Aid**

If you fail to earn a passing grade in at least one of your courses (i.e., all U's, all I's or a combination of all U's, W's or I's) during a semester, you are considered to have, for purposes of federal Title IV funds, unofficially withdrawn from the university. As a result, a federal withdrawal calculation must be performed to determine the amount of Title IV funds that you must repay. Once the amount you must repay is determined, Financial Aid and Scholarships will mail you a letter with the repayment details.

Note: A grade of U (Unearned Failing) is awarded to students who do not officially withdraw from but fail to complete a course (i.e., did not take a final exam, stopped attending, etc.) and failed to achieve the course objectives.

### **Satisfactory Academic Progress (SAP) and Financial Aid**

Federal regulations require you to meet certain minimum academic standards in order to remain eligible for financial assistance. The requirements are that you: 1) maintain a minimum cumulative Texas State GPA; 2) complete at least 70% of all your coursework; and 3) not exceed a maximum limit of attempted hours toward your degree or certificate program.

Additional program-specific requirements also exist (e.g., TEXAS Grant). You can view these SAP criteria in more detail at [www.finaid.txstate.edu](http://www.finaid.txstate.edu) by selecting *Undergraduate Aid* or *Graduate Aid* from the dropdown menu and then *Maintain My Eligibility*.

## **PROJECT GUIDELINES**

### (1) PROJECT OUTLINE:

In this part of the project, you will prepare a one-page outline describing the data set and problem you will investigate (analyze) by using at least one of the methods you learn in this class. In case you need some ideas, a number of data sources will be posted in TRACS. You will list the variables of interest and their relevance in the analysis. Specify the method you will use. You will of course list the members of your group and learn how you will evaluate each other's efforts. Be ready to show your data. The due date is posted in syllabus. There will be a discussion with each group in class time (the date is provided in the syllabus.). Grade 1 or 0

### (2) PROJECT MID EVALUATION

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 (the date is provided in the syllabus.). Grade 1 or 0

### (3) FINAL REPORT

The maximum number of pages for the report CANNOT EXCEED 15 PAGES. A sample report may include the sections such as

- a) An Executive Summary of your findings which will be discussed in detail in your report
- b) An Introduction that describes your problem and your approach
- c) Description of data including a discussion of the data pre-processing, descriptive statistics, outliers
- d) Steps of your analysis
- e) Validation analysis, model choice and evaluation
- f) A conclusion that summarizes your findings

The print out of the final report is due to the beginning of the final day of classes. Grade out of 20.

### (4) PEER EVALUATION

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

Your final grade will be the product of the three grades from (1), (2), (3) and (4).



## COURSE OUTLINE

\*This is a tentative schedule. Everything is subject to change if circumstances warrant. Additional reading will be assigned as necessary. **T=Textbook chapter, S= lecture Sets (Slides) of the instructor, HW: homework**

| WEEK | DATE  | DAY | TOPIC  | MATERIAL       | DUE             |
|------|-------|-----|--|----------------|-----------------|
| 1    | 8/30  | T   | Introduction to Data Mining                                      | S1, T1, 2      |                 |
| 1    | 9/1   | R   | Introduction to Data Mining / Introduction to R                  | S2, T 2, 3.1   | Install R       |
| 2    | 9/6   | T   | Introduction to R / Exploratory Data Analysis                    | S2, T 3.2, 3.3 | Teams Set       |
| 2    | 9/8   | R   | Exploratory Data Analysis  | S3, T 3.2, 3.3 |                 |
| 3    | 9/13  | T   | Basic Visualization  | S3, T 3.2      |                 |
| 3    | 9/15  | R   | Basic Visualization  | S3, T 3.2      |                 |
| 4    | 9/20  | T   | Advanced Visualization   | S3             |                 |
| 4    | 9/22  | R   | Review of HW 1, Conduct of Quiz 1                                |                | HW1             |
| 5    | 9/27  | T   | Clustering   | S4, T4         |                 |
| 5    | 9/29  | R   | Clustering   | S4, T4         |                 |
| 6    | 10/4  | T   | Clustering, Review of HW 2                                       | S4, T4         | HW2             |
| 6    | 10/6  | R   | Association  | S5, T5         |                 |
| 7    | 10/11 | T   | Association  | S5, T5         |                 |
| 7    | 10/13 | R   | Review of HW 3, Conduct of Quiz 2                                | S5, T5         | HW3             |
| 8    | 10/18 | T   | Regression   | S6, T6         |                 |
| 8    | 10/20 | R   | Regression   | S6, T6         |                 |
| 9    | 10/25 | T   | Regression   | S6, T6         |                 |
| 9    | 10/27 | R   | Regression, Review of HW 4                                       | S6, T6         | HW4             |
| 10   | 11/1  | T   | Classification   | S7, T7         |                 |
| 10   | 11/3  | R   | Classification   | S7, T7         |                 |
| 11   | 11/8  | T   | Review of HW 5, Conduct of Quiz 3                                | S7, T7         | HW5             |
| 11   | 11/10 | R   | Project In Class Discussion1                                     |                | Project Outline |
| 12   | 11/15 | T   | CLASS DOES NOT MEET/ Assigned Readings                           |                |                 |
| 12   | 11/17 | R   | Industry Presentations Group 1                                   |                | Presentations   |
| 13   | 11/22 | T   | Industry Presentations Group 2                                   |                | Presentations   |
| 13   | 11/24 | R   | CLASS DOES NOT MEET-THANKSGIVING                                 |                |                 |
| 14   | 11/29 | T   | Model Evaluation and Assessment, Project In Class Discussion2    | S8, T12        |                 |
| 14   | 12/1  | R   | Guest Lecture: Analytics at Workplace: by Michael Willette, USAA |                | Attendance      |
| 15   | 12/6  | T   | Model Evaluation and Assessment, Project In Class Discussion2    | S8, T12        |                 |
| 15   | 12/8  | R   | Review, Discussion of Advanced Topics                            | T9             | Project Report  |
| 16   | 12/13 | T   | CLASS DOES NOT MEET  |                |                 |
| 16   | 12/15 | R   | FINAL SCHEDULED FOR 08:00-10:30 AM                               |                |                 |