THE UNIVERSITY OF TEXAS AT EL PASO
COLLEGE OF SCIENCE
DEPARTMENT OF MATHEMATICAL SCIENCES

Course #:	STAT 5494 (CRN 17195)	
Course Title:	Introduction to Data Mining	
Credit Hrs:	4.0	
Term:	Fall 2019	
Course Meetings & Location:	MW 3:00–4:20 pm (lecture) in HUDS 213	
	MW 4:30–5:20 pm (lab) in BELL 130	
Prerequisite Courses:	STAT 4380 and STAT 5380 or Department/Instructor Approval. Linear regression and generalized linear models (GLM) knowledge and some programming experiences would be desirable.	
Course Fee:	N/A	
Instructor:	Dr. Michael Pokojovy	
Office Location:	Bell Hall 224	
Contact Info:	Phone: (915) 747-6761	
	E-mail address: mpokojovy@utep.edu	
	Fax # 915-747-6502 (Math Department)	
	Emergency Contact: 915-747-5761 (Math Department)	
Office Hrs:	W 2:00–3:00 pm or by appointment	
Textbook(s), Materials:	Required:	Hastie, T., Tibshirani, R., and Friedman, J. H. (2008), Elements of Statistical Learning, 2nd Edition Chapman and Hall. ISBN-13: 978-0387848570 http://www-stat.stanford.edu/~tibs/ElemStatLearn/
	Recommended:	1) Vapnik, V. N. (1998). Statistical Learning Theory. Wiley 2) Clarke, B., Fokoue, E., and Zhang, H. (2009). Principles and Theory for Data Mining and Machine Learning. Springer. ISBN-13: 978-0387981345 3) Bishop, C. M. (1996). Neural Networks for Pattern Recognition. Oxford University Press: USA. ISBN-13: 978-0198538646.
Course Description and Learning Outcomes:	*Contents:*	
	Data Mining emerges as an interdisciplinary field with joint inputs from statistics, computer science, machine learning, and artificial intelligence. This course is intended to cover some commonly-used data mining techniques, with more focus on the most technical part - statistical learning algorithms. The materials are arranged in two main categories: unsupervised learning and supervised learning. A tentative outline of the specific topics is provided below.	
Course Activities/Assignments:	In-class lectures accompanied by program demonstrations and class discussions will be given. Lecture notes will be provided. Computer labs will be mostly student driven. Sporadically, in-class quizzes/exams will be administered. Homework projects will assigned bi-weekly.	

For the final project, students will be expected to independently locate/ select a dataset (from any field) and propose an interesting research question before asking for the instructor's approval of the project. Each student will have the opportunity to present his or her work in class.

Assessment of Course Objectives: Academic performance in this class will be the only factor used in determining the course grade. No extra credit work will be available to improve on any grade. Student performance will be evaluated based on the grades (see Grading Policy below) for in-class quizzes/exams, homework projects and the final project/presentation. No late coursework will be accepted (with exception of some extreme situations as deemed by the instructor).

Tentative Course Schedule:
- Duration: 8/26/2019 – 12/5/2019
- Final project presentation: Mon, 12/9/2019 at 1:00 pm –3:45pm
- Course drop/withdrawal deadline: Fri, 11/1/2019 (No "W" will be assigned for dropping the course after the deadline!)
- Grades officially available online: Thu, 12/19/2019

Tentative List of Topics:
1. An Overview of Data Mining;
2. Intro to Optimization;
3. Intro to Reproducing Kernel Hilbert Space (RKHS);

Part I: Unsupervised Learning
4. Association Rules: Market Basket Analysis; the Apriori Algorithm; Generalized Association Rules.
5. Outlier Detection: Statistical Discordance Tests, Nearest Neighbor and Distance-Based; Partitioning-Based; Influential Points.
6. Non-Negative Matrix Factorization: Archetypal Analysis.
7. Web Mining: Google PageRank

Part II: Supervised Learning
8. Parametric Nonlinear Regression: Numerical Optimization methods;
9. Nonparametric Smoothing I - the Kernel Method: K-nearest neighbor (KNN); kernel regression;
10. Nonparametric Smoothing II - the Spline Method: Polynomial Regression and Piecewise Polynomials; (Natural) Cubic Splines; B-Splines; Smoothing Splines - the penalized likelihood.
11. Generalized Additive Models (GAM): Additive Models and GAM.
12. Multivariate Adaptive Regression Splines (MARS); Generalization to different types of responses.
13. Projection Pursuit Regression (PPR): the backfitting algorithm
14. Artificial Neural Networks - I: Multilayer Perceptron (MLP): ANN history and architectures; Single-Layer Perceptron and MLP; Back Propagation (Backprop) and other optimization methods; deep learning

15. Artificial Neural Networks - II: Radial Basis Function Networks (RBF): Ordinary RBF and Normalized RBF.
16. Support Vector Machines (SVM)

Part III: Additional Topics (if time permits)
17. Hidden Markov Models (HMM)
18. Reinforcement Learning

Grading Policy: 60% Homework/project assignments
10% In-class quizzes and/or exams and attendance
30% Final project (both implementation and presentation)

The usual grading scale will be used for this course:
90–100% = A
80–89% = B
70–79% = C
60–69% = D
0–59% = F

Make-up Policy: If class is missed for a valid (as deemed by the instructor) & documented reason and the instructor is informed beforehand, the in-class assignments may be made-up for full credit. All other assignments must be turned in on time.

Attendance Policy: You are expected to attend class so that you may turn in the in-class assignments and homework projects. Lecture and lab attendance is further required and helpful to decide the grade in "borderline" situations. Five or more unexcused absences (from the lab and/or the lecture) may result in an instructor-initiated drop or final grade reduction. *Your academic advisor may be consulted before final action is decided and taken.* If a student has to be absent from a class, he/she will be responsible for catching up with course material. Late arrivals are not permitted. Being late by 10 minutes or more or leaving the classroom before the class is dismissed will be considered an absence.

Academic Integrity Policy: The University policy is that all suspected cases or acts of alleged scholastic dishonesty must be referred to the OSCCR for investigation and appropriate disposition. Any student who commits an act of scholastic dishonesty is subject to discipline. Scholastic dishonesty includes, but is not limited to cheating, plagiarism, collusion, the submission for credit of any work or materials that are attributable in whole or in part to another person, taking an examination for another person, any act designed to give unfair advantage to a student or the attempt to commit such acts. Each student is responsible for notice of and compliance with the provisions of the Regents' Rules and Regulations, which are available for inspection electronically at
http://www.utsystem.edu/bor/rules/homepage.htm

All students are expected and required to obey the law, to comply with the Regents' Rules and Regulations, with System and University rules, with directives issued by an administrative official in the course of his or her authorized duties, and to observe standards of conduct appropriate for the University. A student who enrolls at the University is charged with the obligation to conduct himself/herself in a manner compatible with the University's function as an educational institution.

Any student who engages in conduct that is prohibited by Regents' Rules and Regulations, U. T. System or University rules, specific instructions issued by an administrative official or by federal, state, or local laws is subject to discipline, whether such conduct takes place on or off campus or whether civil or criminal penalties are also imposed for such conduct.

Civility Statement: This is a class where participation is required. You will be seated at your desk or in front of a computer for the duration of your class and you are expected to follow the lecture/discussion and at various times complete in-class assignments. You are not allowed to browse the Internet during class time or work on any other material. If you regularly do not complete in-class assignments in a satisfactory manner, participate in class, or if you work on other material in class you will have points deducted from your in-class assignments portion of your grade. All class participants must and will be treated with respect.

Please do not use cell phones, iPods, MP3 players, blue tooth devices, etc. during class. Cell phones and pagers should be set to silent or vibrate, and any calls should be taken outside of class. Please do not wear headsets or blue tooth devices during class. Please don't talk in class. Cell phone calculators may not be used on quizzes or exams. Active participation in class is expected, teamwork in class will be implemented.

Disability Statement: If a student has or suspects she/he has a disability and needs an accommodation, he/she should contact The Center for Accommodations and Support services (CASS) at 747-5148 or at <cass@utep.edu> or go to Room 106 Union East Building. The student is responsible for presenting to the instructor any CASS accommodation letters and instructions.

Military Statement: If you are a military student with the potential of being called to military service and/or training during the semester, please contact me by the end of the first week of class

College of Science Policies: All grades of Incomplete must be accompanied by an Incomplete Contract that has been signed by the instructor of record, student, Department Chair, and the Dean. Although UTEP will allow a maximum of one year to complete this contract, the College of Science requests it be limited to month based upon completion data. A grade of Incomplete is only used in extraordinary circumstances confined to a limited event such as a missed exam, project, or lab. If the student has missed a significant amount of work (e.g., multiple assignments or tasks), a grade of Incomplete is not appropriate or warranted.

Last updated on 8/28/2019

Disclaimer:
This syllabus may be subject to changes if these are deemed necessary by the instructor.
Despite all efforts, this syllabus may contain typos and errors.