

Course Syllabus
IE/SE/MFG 5390: Special Topics (Industrial data analytics)
Fall 2022

(Class will be delivered online, Synchronously, class materials will be posted on Blackboard every Tuesday before 6:00pm)

Instructor	Dr. Md Fashiar Rahman
Location and Time	Online via Blackboard, Tuesday 6.00 PM to 8.50 PM
Instructor's Office	A 243
Instructor's Office Hours	Wednesday and Thursday: 12.00 PM – 1.00 PM
Email	mrahman13@utep.edu
Teaching Assistant	TBD

Course web page:

<https://blackboard.utep.edu/>

Course Motivation:

21st century has been called by many “The Century of Data.” We see more and more data collected with the expectation (justified by empirical evidence!) that analyzing these data will give organizations a competitive edge and will help them to excel. The amount of data collected is enormous and growing. In many cases, analyzing these data lags behind. Mark Twain wrote once “A man who does not read has no advantage over a man who cannot read.” A similar sentence is most certainly true: “A company who does not analyze his data has no advantage over a company who has no data.” Methods for analyzing data, called collectively “data analytics” are, therefore, crucial in for businesses to gain competitive advantage.

Data mining (Data analytics), also called Knowledge Discovery in Databases (KDD), is the field of discovering novel and potentially useful information from large amounts of data. Data mining has been applied in a great number of fields, including retail sales, bioinformatics, healthcare, manufacturing and counter-terrorism, among others. Data analytics is an important, fast-growing field that has quickly become a key basis of productivity growth, innovation, and consumer surplus. There is an increasing need for analytics-savvy employees who can think uniquely across disciplines to transform data into relevant insights for making better business decisions. It is a highly inter-disciplinary area representing the confluence of machine learning, statistics, operations research, database systems and high-performance computing, among other fields.

Course Description (3 Credits)

This course is an **introductory course** open to both graduate and senior undergraduate students. As an introductory course on data analytics, this course introduces the concepts, algorithms, techniques, and systems of data mining, including (1) an introduction to data analytics, (2) data preprocessing, (3) mining frequent patterns and associations, (4) classification, (5) cluster analysis, and (6) learning about software used in data mining and (7) demonstration of how to apply data analytics techniques using Python. The course will provide students with basic understanding of common data mining and analytics techniques.

Prerequisites:

Basic engineering statistics and basic calculus and linear algebra. Having basic programming knowledge would be an advantage. However, this is not a course of mathematical statistics or advanced calculus.

Course Objectives:

- Understand the importance of Data Analytics
- Students will learn about fundamental data analytic techniques and tools used to analyze data
- Be capable of confidently applying common data analytic algorithms in practice
- Be capable of performing data mining experiments using programming in Python
- Be capable of performing experiments in using real-world data

Learning Outcomes:

At the end of this course, students should be able to

- Understand modern views in data analysis;
- Explain the data mining methodology.
- Use visual techniques to describe data.
- Explain and apply the different regression, clustering, classification, and association rule mining techniques.
- Able to explain the assumption and interpret the output of the above data analytic techniques

Textbook

There is no required textbook. However, necessary reading materials will be provided during the semester.

Useful References:

- Data Mining: Concepts and Techniques, 3rd edition, written by Jiawei Han, Micheline Kamber, Jian Pei.
- An Introduction to Statistical Learning, written by Daniela Witten, Trevor Hastie, Robert Tibshirani
- An Introduction to R by W. N. Venables, D. M. Smith and the R Core Team
- Materials not from the textbook will be provided in lecture notes.

Technology Requirements:

This course has been designated as a laptop course. Most class activities will be conducted using laptops. You will need to bring your laptop every class unless otherwise noted explicitly. In general, I assume that each student has a laptop with the appropriate software. You will need to have access to a computer/laptop, scanner, webcam, and microphone. You will need to download or update the following software: Microsoft Office, Adobe Acrobat Reader, Windows Media Player, QuickTime, and Java. Check that your computer hardware and software are up-to-date and able to access all parts of the course. Suppose you do not have word-processing software. In that case, you can download Word and other Microsoft Office programs (including Excel, PowerPoint, Outlook and more) for free via UTEP's Microsoft Office Portal. Click the following link for more information about [Microsoft Office 365](#) and follow the instructions.

Student also need to install **Python** on their laptop. Python is an open source and free programming language compatible with any kind of operating systems. Click the following to download and install the [Python](#). **I will provide the further installation guideline during the class.**

IMPORTANT: If you encounter technical difficulties beyond your scope of troubleshooting, please contact the [UTEP Help Desk](#) as they are explicitly trained in assisting with the technological needs of students. Please do not contact me for this type of assistance. The Help Desk is much better equipped than I am to assist you!

Course Materials and Office Hours

I will post lectures, links to other relevant reading materials, homework questions, and project details on Blackboard. All submissions **MUST** be submitted through Blackboard. Paper submissions will not be accepted. The office hours are on Wednesday and Thursday between 12.00 PM and 1.00 PM. However, I can also meet with you using video conferencing services such as Microsoft Teams based/Zoom on a pre-determined meeting time.

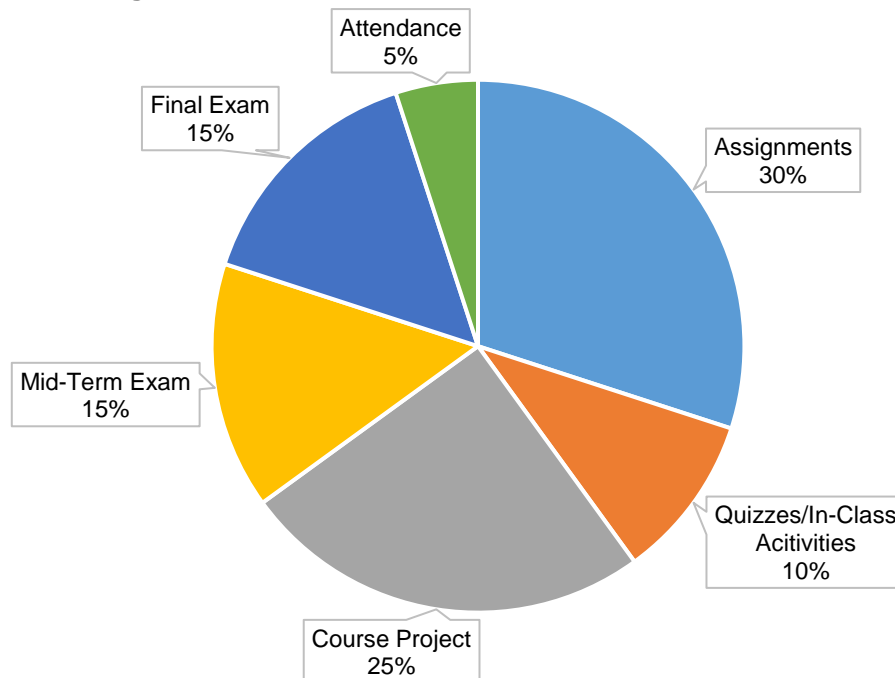
Email:

UTEP e-mail is the best way to contact me. I will make every attempt to respond to your e-mail within 24-48 hours of receipt. When e-mailing me, be sure to email from your UTEP student account and please put the course number in the subject line. In the body of your e-mail, clearly state your question. At the end of your e-mail, be sure to put your first and last name, and your university identification number.

Grades:

- **A:** 90 and above
- **B:** ≥ 80 and < 90
- **C:** ≥ 70 and < 80
- **D:** ≥ 60 and < 70
- **F:** < 60

Course Grading Scheme:



Homeworks/assignments and Make-Up Work Policy:

Homeworks/Assignments will be announced every week. Make-up work will be given only in the case of a documented emergency. Note that make-up work may be in a different format than the original work, may require more intensive preparation, and **may be graded**

with penalty points. If you miss an assignment and the reason is not considered excusable, you will receive a zero. It is therefore important to reach out to me—in advance if at all possible—and explain with proper documentation why you missed a given course requirement. Once a deadline has been established for make-up work, no further extensions or exceptions will be granted.

Quizzes/In-Class Activities

The quizzes/in-class activities will be in class without previous notice. The contents of the quiz will be covered before on the same class. So, all the students are suggested to be regular in class. Note that make-up quiz may be in a different format than the original work, may require more intensive preparation, and may be graded with penalty points. If you miss a quiz and the reason is not considered excusable, you will receive a zero.

Course Project

All students are required to do a course project in this course. The purpose of the term project is to provide an opportunity to apply and/or further explore a topic area related to the course. Some examples of types of projects that are acceptable are (but not limited to):

- Research proposal incorporating engineering analysis that demonstrates creativity and originality
- Apply different models to a specific problem and compare results (interest is in models not covered in class, maybe use some from the class and present a new one)
- Literature review of data mining methods used in a problem area that demonstrates breadth and organization of the literature

Each team is required to identify a topic of interest and submit a one-page abstract by the first quarter of the semester. It is expected that the group will review recently published literature on the topic of interest.

Students are strongly encouraged to form teams (maximum two members per team) to work on the project. You can use course's discussion boards to find your team members if you don't know anyone else. It is highly advised to actively participate in project collaboration. **Two complaints of non-performance/non-participation about the other teammates will result in reduction of grade by one grade. Four cumulative complaints will result in an F-grade.**

Blackboard Submission

I strongly suggest that you submit your work with plenty of time to spare in the event that you have a technical issue with the course website, network, and/or your computer. If you are experiencing difficulties submitting your work through the course website, please contact the UTEP Help Desk. It is your responsibility to submit the assignments in Blackboard before the due date

Netiquette

As we know, sometimes communication online can be challenging. It's possible to miscommunicate what we mean or to misunderstand what our classmates mean given the lack of body language and immediate feedback. Therefore, please keep these netiquette (network etiquette) guidelines in mind. Failure to observe them may result in disciplinary action.

- Always consider audience. This is a college-level course; therefore, all communication should reflect polite consideration of other's ideas.
- Respect and courtesy must be provided to classmates and to the instructor at all times. No harassment or inappropriate postings will be tolerated.

- When reacting to someone else's message, address the ideas, not the person. Post only what anyone would comfortably state in a face-to-face situation.
- Blackboard is not a public internet venue; all postings to it should be considered private and confidential. Whatever is posted on in these online spaces is intended for classmates and professor only.

Accommodations Policy:

The University is committed to providing reasonable accommodations and auxiliary services to students, staff, faculty, job applicants, applicants for admissions, and other beneficiaries of University programs, services and activities with documented disabilities in order to provide them with equal opportunities to participate in programs, services, and activities in compliance with sections 503 and 504 of the Rehabilitation Act of 1973, as amended, and the Americans with Disabilities Act (ADA) of 1990 and the Americans with Disabilities Act Amendments Act (ADAAA) of 2008. Reasonable accommodations will be made unless it is determined that doing so would cause undue hardship on the University. Students requesting an accommodation based on a disability must register with the UTEP Center for Accommodations and Support Services (CASS). Contact the Center for Accommodations and Support Services at 915-747-5148, or email them at cass@utep.edu, or apply for accommodations online via the CASS portal.

Scholastic Integrity:

Academic dishonesty is prohibited and is considered a violation of the UTEP Handbook of Operating Procedures. It includes, but is not limited to, cheating, plagiarism, and collusion. Cheating may involve copying from or providing information to another student, possessing unauthorized materials during a test, or falsifying research data on laboratory reports. Plagiarism occurs when someone intentionally or knowingly represents the words or ideas of another as ones' own. Collusion involves collaborating with another person to commit any academically dishonest act. Any act of academic dishonesty attempted by a UTEP student is unacceptable and will not be tolerated. All suspected violations of academic integrity at The University of Texas at El Paso must be reported to the Office of Student Conduct and Conflict Resolution (OSCCR) for possible disciplinary action. To learn more, please visit HOOP: Student Conduct and Discipline.

Plagiarism Detecting Software

Some of your course work and assessments may be submitted to SafeAssign, a plagiarism detecting software. SafeAssign is used to review assignment submissions for originality and will help you learn how to properly attribute sources rather than paraphrase.

Copyright Statement for Course Materials

All materials used in this course are protected by copyright law. The course materials are only for the use of students currently enrolled in this course and only for the purpose of this course. They may not be further disseminated.

Class Recordings:

The use of recordings will enable you to have access to class lectures, group discussions, and so on in the event you miss a synchronous or in-person class meeting due to illness or other extenuating circumstance. Our use of such technology is governed by the Federal Educational Rights and Privacy Act (FERPA) and UTEP's acceptable-use policy. A recording of class sessions will be kept and stored by UTEP, in accordance with FERPA and UTEP policies. I will not share the recordings of our class activities outside of course participants. You may not share recordings outside of this course. Doing so may result in disciplinary action.

Programming Language – Python:

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built-in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python is simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms, and can be freely distributed. <https://www.python.org/>

Civility Statement:

Please be respectful of all students' right to learn without disruptions. In line with this statement please make an active effort to keep the talking to a minimum during lectures and presentations. Also make an active effort to either turn cell phones off or turn them to vibrate mode prior to the start of class. Appointments with instructor should be made in advance.

Tentative Course Schedule (It may change, based on feedback or progress)

Day	Date	Topics
1	08/23	Course syllabus and introduction to data analytics
2	08/30	Getting to know your data
3	09/06	Data Preprocessing
4	09/13	Lab # 1: Python installation and basics
5	09/20	Lab # 2: Working with Panda library and data preprocessing
6	09/27	Linear regression
7	10/04	Logistic regression
8	10/11	Mid-term Exam
9	10/18	Lab # 3: Linear and Logistics regression
10	10/25	Clustering – Partitioning method
11	11/01	Density-based Clustering
12	11/08	Classification Techniques
13	11/15	Lab #4: Clustering and Classification
14	11/22	Association Rule Mining
15	11/29	Final Project Presentation
16	12/06	Final Exam