

Syllabus – Spring 2021 (1/19/2021 – 4/28/2021)

Excluding materials for purchase, syllabus information may be subject to change. These will be announced immediately on HuskyCT and in class.

Course and Instructor Information

Course Title: Big data analytics

Credits: 3 units

Instruction mode: Distance learning (synchronous online lectures on HuskyCT)

Class times: Tuesdays and Thursdays at 09:30AM – 10:45AM

Prerequisites: CSE 3500; MATH 2210; open only to students in the School of Engineering and declared Computer Science and Analytics minors.

Professor: Kriti Bhargava

Email: kriti.bhargava@uconn.edu

Availability: I will do my best to respond to emails within 48 hours.

Office hours: TBD

Office: [virtual WebEx office](#)

Course Materials

Required course materials should be obtained before the first day of class.

Required textbook: Data Mining: Concepts and Techniques, Jiawei Han, Micheline Kamber and Jian Pei, 3rd ed ([e-copy](#))

Course Description

This is the introductory course to the area of data mining and big data analytics. Topics include the following: data visualization, data pre-processing, data warehousing, pattern mining, classification, clustering, and outlier detection. Possible additional topics: time series analysis.

This course is designed to be mostly self-contained. The topics involve hands-on experience with R/Python. We will try to provide sufficient background but learning the basics in advance is recommended.

Course Objectives

By the end of this course, you should understand basic principles, concepts, and techniques of data mining, sufficiently to:

- examine various data mining tools/methods through published books/articles, S/W packages.
- employ basic mining techniques such as pattern mining, clustering, classification to publicly available data.
- evaluate the correctness of data mining models.
- understand current trends in data mining and big data analytics.

Course Outline

A tentative ordering of topics covered is included below.

- Introduction to data mining and its applications
- Data visualization
- Data preprocessing
- Data warehousing
- Frequent pattern mining
- Classification
- Clustering
- Outlier detection
- Time permitting: Time series analysis

Course Requirements and Grading

Summary of Course Grading:

Course Components	Weight
Participation in class and discussion boards	10%
Assignments	30%
Final project + presentation	25%
Midterm exam	15%
Final exam	20%

Participation in class and discussion boards

Students are expected to **keep their webcams on** during a class. Students are expected to actively participate in discussions during the class as well as on discussion boards on various topics throughout the course. Students are expected to respond to prompts as well as other students. Students are graded on the quality of their comments and their ability to spur a thoughtful and respectful discussion.

Following etiquettes must be observed in all discussions during class or otherwise on discussion boards.

- **Be respectful.** While it is easier to say hurtful or disrespectful things without standing face-to-face with someone, it is important to remember that your classmates and teachers are real people who are affected by the words you say and write. It is essential to keep in mind the feelings and opinions of others, even if they differ from your own. If you would not say it to someone's face, don't say it online either.
- **Be aware of strong language, all caps, and exclamation points.** It is easy for written text to be misread and misunderstood. Have you ever sent a text message with good intent, but your recipient thought you were being rude? If so, then you have experienced this firsthand. By being cognizant of strong language, you can identify potential confusions before sending messages. Tip: Read everything out loud before you send it.
- **Be careful with humor and sarcasm.** Certainly, you should not avoid being funny. We love to see your personality shine through in online classes. Many of our teachers are exceptionally funny too. But like mentioned in Rule #2, make sure that it is clear you are being funny and not being rude. Emoticons and smileys can be helpful when conveying humor or sarcasm so that it is read correctly. Just remember to keep the smiley faces away from academic papers. 😊
- **Yes, grammar and spelling matter.** While texting, textspeak can be great for your friends. In an educational setting (even online) however, keep it formal. Your written communication should be professional and reflect proper writing style. Save written shortcuts and less than stellar grammar for Snapchat if you must, but follow grammar rules for school.
- **Cite your sources.** Whenever you are sharing an idea that originated from someone else (even if it is not word for word), it is good practice to cite that source. This applies to discussion forums too. If you read a great thought in your text, share it, but be sure you let your audience know where you saw it first.
- **Do not post or share (even privately) inappropriate material.** Enough said there. Nothing is private online.
- **Do not abuse the chat in HuskyCT or discussion boards.** Be mindful to not distract the others in class by posting material that is of a personal or non-academic nature in HuskyCT chat or discussion boards.
- **Be forgiving.** Remember that not everyone will know these rules before posting. Try to be understanding of others when they struggle with written communication. It is very different than simply talking to a person face-to-face. Try support your fellow students and the instructor as we are all new to this. 😊

Assignments

Individual homework assignments will be given on a bi-weekly basis to cover all topics – data visualization, preprocessing, pattern mining, classification, clustering, and outlier detection. All the homework assignments must be submitted electronically using HuskyCT. For problem sets, the only file format accepted is pdf. For programming assignments, all scripts and results must be submitted as a R/Python notebook.

We encourage high-level discussion of ideas and problems with peers, but assignments must be completed and submitted individually. If there is ambiguity about what constitutes “high-level discussion” or anything else regarding assignments, please ask the instructor. You may not google around hoping to find a similar problem worked out, but you may use other resources to read (like Wikipedia, other lecture notes, textbooks) that you find online. If you are using other resources, you must properly cite them, and you must prove any statement from them that you use. You are not allowed to look for the answers for any of the homework assignments (online or otherwise). If you accidentally find a solution for a question, do not read it (if it is not too late), and indicate that clearly on your assignment (including where you found the solution).

Additionally,

- everyone can have bad days, so, the lowest homework assignment will be dropped. You will still be responsible for the material on the exams.
- everyone will be allotted 3 no-questions-asked late days to be used at any point unless explicitly prohibited. These are calendar days, not business days (Saturdays and Sundays count towards the 3 days). You do not need to inform us when you are taking the late days – we will automatically apply them. Inform the instructor if observance of religious holidays, illness, disability, or other valid reasons interferes with your work so we can try our best to accommodate you.

Final project and presentation

Students will be assigned to a team of 2-3 members, depending on the class enrollments, to apply the data mining tools learnt in class to publicly available datasets (to be approved by the instructor). Each team will be required to give three in-class presentations: proposal (What topics are you going to work on? Why do you want to work on that topic? What results are you expected to get? What reference papers are you going to read to do a comprehensive survey on that topic? What data sets are you going to use and how to evaluate?), progress (What is the progress on the project? What are the results so far? Is it aligned with what was expected? Challenges so far? Road ahead?), and final presentation (Complete report on what was done. What data set was analyzed? What was the hypothesis? What were the techniques used? What is the accuracy of analysis and correctness of methodology? What are the findings?) that includes their understanding and discovery. The presentations will be followed by a Q&A session in class by the instructor and other students (will be counted towards class participation). Each team must submit a final report that includes the script and outcomes as a R/Python notebook after the final presentation.

Midterm and Final exam

[Respondus Lockdown Browser with Monitor](#) would be used for both exams, which would primarily consist of multiple-choice questions. Please ensure your environment is set up correctly using the practice test at the start of the semester. *The use of Respondus Lockdown Browser with Monitor is at the discretion of the instructor. All students must have the required working equipment. No exceptions to this will be allowed.*

The midterm exam would be scheduled during class hours around week 6. This will be announced a few days in advance and would last for ~1 hour.

The final exam will be scheduled in the week of May 3 to May 8 (based on the University exam time-table).

Grading Scale:

Grade	Letter Grade	GPA
93-100	A	4.0
90-92	A-	3.7
87-89	B+	3.3
83-86	B	3.0
80-82	B-	2.7
77-79	C+	2.3
73-76	C	2.0
70-72	C-	1.7
67-69	D+	1.3
63-66	D	1.0
60-62	D-	0.7
<60	F	0.0

Due Dates and Late Policy

All course due dates will be identified on HuskyCT. All assignments are due by 11:59 pm Eastern Time (ET). If you are in a different time zone, please adjust your submission times accordingly. *The instructor reserves the right to change dates accordingly as the semester progresses. All changes will be communicated in an appropriate manner.* Late Assignments will be penalized.

Feedback and Grades

I will make every effort to provide feedback and grades within a week. For each assignment, project, and exam, regrade requests will *only be open 2 days after the grades have been published*. Please provide a detailed explanation for your regrade request. To keep track of your performance in the course, refer to My Grades in HuskyCT.

Weekly Time Commitment

You should expect to dedicate approximately 9-10 hours a week to this course. This expectation is based on the various course activities, assignments, and assessments and the University of Connecticut's policy regarding credit hours. More information related to hours per week per credit can be accessed at the [Online Student website](#).

Student Authentication and Verification

The University of Connecticut is required to verify the identity of students who participate in online courses and to establish that students who register in an online course are the same students who participate in and complete the course activities and assessments and receive academic credit. Verification and authentication of student identity in this course will include:

1. Secure access to the learning management system using your unique UConn NetID and password.
2. Online student authentication and five proctored tests.

Student Responsibilities and Resources

Formally, we follow the university policy on academic integrity to discourage and penalize academic misconduct.

Academic misconduct is dishonest or unethical academic behavior that includes, but is not limited to, misrepresenting mastery in an academic area (e.g., cheating), failing to properly credit information, research, or ideas to their rightful originators or representing such information, research, or ideas as your own (e.g., plagiarism). – UConn Community Standards, Academic Misconduct

Students found in violation of academic integrity may be subject to failing the assignment, the course, and/or review by the academic integrity hearing board. See the [student code](#) and references therein for more information.

Informally, all academic work you submit must be your own. Discuss classwork, exercises, and problems with peers in a manner that helps all parties understand a problem or possible path towards its solution. Do not copy solutions from other students or from any other resource. You will put your future self in the best position to succeed in academia or industry if you dedicate yourself to learning and retaining the material.

Students with Disabilities

The University offers many services to its students with disabilities through the [Center for Students with Disabilities](#). Eligibility for these services is determined individually based on documented need. If you have a diagnosed disability (physical, learning, or psychological) that will make it difficult for you to carry out the course work as outlined, or that requires accommodations such as recruiting note-takers, readers, or extended time on exams or assignments, please advise the instructor during the first two weeks of the course so that we may review possible arrangements for reasonable accommodations.

Device Requirements (with Accessibility and Privacy Information)

Please see the attached Device Requirements document linked to HuskyCT. You must abide by these.

Help

[Technical and Academic Help](#) provides a guide to technical and academic assistance.

This course is completely facilitated online using the learning management platform, [HuskyCT](#). If you have difficulty accessing HuskyCT, you have access to the in person/live person support options available during regular business hours through the [Help Center](#). You also have [24x7 Course Support](#) including access to live chat, phone, and support documents.

Evaluation of the Course

Students will be provided an opportunity to evaluate instruction in this course using the University's standard procedures, which are administered by the [Office of Institutional Research and Effectiveness](#) (OIRE).

Additional informal formative surveys may also be administered within the course as an optional evaluation tool.