University of New Haven

TAGLIATELA COLLEGE OF ENGINEERING

Electrical and Engineering and Computer Science

CSCI 4401/ CSCI 6401
Data Mining
Fall 2024

Meeting Times and Location(s): MW 3:30 pm - 4:45 pm Kaplan Hall 107
Credit Hours: 3

Dr. Shivanjali Khare, Assistant Professor
Faculty Contact Information:

Office: Maxcy Room 120B

Phone: 203-479-4872 Email: skhare@newhaven.edu

Office Hours: MW 2:00 pm - 3:15 pm, (other hours: Zoom on appointment)

Department Chair: Dr. Ali Golbazi AGolbazi@newhaven.edu

COURSE SYLLABUS

This syllabus is informational in nature and is not an express or implied contract. It is subject to change due to unforeseen circumstances, as a result of any circumstance outside the University's control, or as other needs arise. If, in the University's sole discretion, public health conditions or any other matter affecting the health, safety, upkeep or wellbeing of our campus community or operations requires the University to make any syllabus or course changes or move to remote teaching, alternative assignments may be provided so that the learning objectives for the course, as determined by the University, can still be met. The University does not guarantee that this syllabus will not change, nor does it guarantee specific in-person, on-campus classes, activities, opportunities, or services or any other particular format, timing, or location of education, classes, activities, or services.

Course Description:

The course will include curating and cleaning the data as well as choosing the proper data mining techniques in response to a research question. Emphasis is on data warehousing, time series data analysis, web mining, text mining techniques, and detecting abnormal data, reporting research findings in a scientific way. Students enhance skills by continuous search, practice, and collaboration with their corresponding team.
Undergraduate: https://catalog.newhaven.edu/index.php?catoid=30
Graduate: https://catalog.newhaven.edu/index.php?catoid=31

Prerequisites: CSCI 6651 - Introduction to Script Programming

Optional Text(s):

Although the lecture slides are designed to be self-contained, it is recommended (but not required) to reference the textbook: Han, J., Kamber, M., & Pei, J. (2011). Data mining: Concepts and techniques (3rd ed.). Waltham: Morgan Kaufmann. You can download a PDF version from Data mining: Concepts and techniques (3rd ed.) for free. Note that these are all the chapters related to the topics covered in this course, so the free PDF version of the chapters is sufficient for this course.

Other:

- T. Hastie, R. Tibshirani, and J. Friedman (2001) The Elements of Statistical Learning:data mining, inference and prediction. Springer Verlag. (https://web.stanford.edu/~hastie/Papers/ESLII.pdf)
- Stock, J. H., & Watson, M. W. (2015). Introduction to Econometrics, Third Update, Global Edition. Pearson Education Limited. (https://www.econometrics-with-r.org/#)
- Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data (2nd ed.), Bing Liu,Springer,2011,ISBN:978-3642194597 (https://link.springer.com/book/10.1007/978-3- 642-19460-3utlier Analysis, Aggarwal C.C,Springer, 2017. https://doi.org/10.1007/978-3-319-47578-3_1 (https://link.springer.com/chapter/10.1007/978-3-319-47578-3_1)

Mandatory Tools: You should install a Oracle Virtual Box on your computer and start to learn the Oracle Virtual Box on your own time. Hacking Club is one recourse for help learning Oracle Virtual Box. In advanced programming and networking courses you will use Linux environments to manage your code, some of you from the command line.

Course Structure/Course Format:

Course structure is lecture-based and includes lab and discussion; group learning projects; and presentations.

Extended Course Description:

- **On-Ground:** Fully on-ground course with every student meeting in-person.

Course Objectives:

- principles of data warehousing
- Text mining techniques
- Data analysis (Classification and clustering)
- Predictive Time series analysis and outlier detection

Student Learning Outcomes:

By the end of this course, students will:

- Select the proper data mining technique to address distinct research questions
- Perform data cleaning, exploration, modeling, and optimization techniques
- Solve real-world data mining problems and report the results

Course Requirements & Assessment:

Please see official University of New Haven Academic Policies located in the links below:

Choose appropriate link(s).

Undergraduate Grading System

Graduate Grading System

Assignments/Projects/Examinations:

- It is important to read the project instructions carefully and follow them precisely. They will guide you through learning the things you need to know. Please let me know if youthink there is an error in the instructions.
- You will select a project for this course and report your progress twice a week via assignments. It is expected to submit your assignments by the deadline. There will be penalty points for late submissions.
- The project must be the result of a team working. It is expected that each member of a group can present his/her role. Also, you will get the project's related scores if your team participates in presentations.

Participation

Active learning will constitute as much as 50% of the class. Participation will be recorded based on engagement in in-person discussions and attendance.

Grading:

Edit the statement and charts below for your course.

Grades earned are based on your performance on homework, quizzes, exams and the final exam.

Homeworks assignments and quizes	05
Programming activity interactions & submissions	05
Phase 1: Selecting your group	02
Phase 2: Selecting a research question & a dataset	08
Phase 3: Report a review on the related work	03
Phase 4: Data Exploration	05
Phase 5: Modeling Data	05
Phase 6: Optimization	05
Phase 7: Submit the final report in the format of academic paper	10
Phase 8: Presentation	12
Midterm Exam	20
Final Exam	20
Total**	100

**Final Grades are assigned with the following scale:

Typical Undergraduate Scale

93 to 100	A
90 to Less than 93	A-
87 to Less than 90	B+
83 to Less than 87	B
80 to Less than 83	B-
77 to Less than 80	C+
73 to Less than 77	C
70 to Less than 73	C-
67 to Less than 70	D+
63 to Less than 67	D
60 to Less than 63	D-
Less than 60	F

Typical Graduate Scale

93 to 100	A
90 to Less than 93	A-
87 to Less than 90	B+

83 to Less than 87	B
80 to Less than 83	B-
77 to Less than 80	C+
73 to Less than 77	C
70 to Less than 73	C-
Less than 70	F

The calculation of final grades is determined by the faculty member. The calculated grade in the total column in Canvas may or may not be reflective of your final grade.

Expectations:

Students are expected to spend at least two hours on academic studies outside, and in addition to, each hour of class time.

Attendance Policy Statement: Compulsary in-class attendance. Lecture recordings will be available after the lecture. Zoom attendance is not allowed.

TCoE Academic Lab reservation form

As a TCoE student, you have access to reserve academic lab spaces for academic purposes where you need access to specific equipment. Example approved uses might include time for a team meeting to finish a team project or a study-session with a TA. For more information or to submit your reservation, please visit: https://forms.office.com/r/EUeJT36ZFr

Course Outline/Schedule:

https://www.newhaven.edu/academics/calendar/

08/26	Introduction
08/28	Data warehouse and OLAP
09/02	Labor day Break
09/04	BI Association and Rule Mining + weka intro
09/09	BI Association and Rule Mining + weka activity
09/11	Term Project activity
09/16	Classification + Performance measures
09/18	Classification + Overfitting
09/23	Classification + python activity
09/25	Clustering + python activity
09/30	Clustering + python activity

10/02	Regression + python activity
10/07	Mid-term
10/09	Regression
10/14	Regression + python activity
10/16	Term project activity
10/21	Fall Break
10/23	Time-series Analysis
10/28	Time-series Analysis + python activity
10/30	Meta Algorithms
11/04	Meta Algorithms
11/06	Model Evaluation + Class Activity
11/11	Model Evaluation + Optimization Class Activity
11/13	Deep Learning + Class Activity
11/18	Deep Learning + Class Activity
11/20	Scaling up Deep Learning
11/25	Recommendation systems
11/27	Thanksgiving Break
12/02	Presentations - I
12/04	Presentations - II
12/09	Presentations – III + Doubt clearing session

Diversity Statement

The University of New Haven embraces diversity and recognizes our responsibility to foster a diverse, inclusive, and welcoming environment in which all members of the Charger community of all backgrounds and identities can learn, work, and live together. We benefit from the academic, social, and cultural developments that arise from a diverse campus that is committed to equity, inclusion, belonging, and accountability.

We have a responsibility as a community and as individuals to address and remove barriers, achieve success, and sustain a culture of inclusivity, empathy, kindness, and compassion. We encourage, welcome, and embrace participation in ongoing dialogue, engagement, and education to critically examine and thoughtfully respond to the changing realities of our community. Diversity, equity, inclusion, acceptance, and belonging enrich the Charger community and are instrumental to institutional success and fulfillment of the University mission.

Reporting Bias Incidents

At the University of New Haven, there is an expectation that all community members are committed to creating and supporting a climate which promotes civility, mutual respect, and open-mindedness. There also exists an understanding that with the freedom of expression comes the responsibility to support community members' right to live and work in an environment free from harassment and fear. It is expected that all members of the University community will engage in anti-bias behavior and refrain from actions that intimidate, humiliate, or demean persons or groups or that undermine their security or self-esteem.

If you have an immediate safety concern for yourself or others, and/or believe someone poses an immediate threat to themselves or others, please contact University Police at 203-932-7070 or call 911. Community members can report bias-motivated incidents by completing the form at **www.newhaven.edu/biasreporting**. Community members are encouraged to complete this form if they are the target of bias or harassing behaviors, witness such behaviors, or gain knowledge of these behaviors occurring within the University community. All matters concerning bias and harassment will be handled by the Dean of Students Office and Human Resources Office.

University-wide Academic Policies

A continually-updated list of University-wide academic policies and descriptions of key university student resources, can be found on Canvas. You can access them by simply clicking on the (?) help button.

The University-wide academic policies include (but are not limited to) the University's attendance policy, procedures for both adding / dropping a course and course withdrawals, an explanation for the sorts of circumstances where incomplete (INC) grades could be considered by the faculty, and the academic integrity policy (among others).

It is strongly advised to include a statement regarding the use of generative artificial intelligence tools in your course, along with any other actions that could lead to academic misconduct violations and subsequent reporting. This statement serves to

inform students about the expectations and it will be beneficial in the reporting process should any violations occur throughout the course. A violation should be reported by the instructor for any instance in which a sanction was imposed. Please refer to the *Faculty Guide for Academic Misconduct* *https://facultysupport.newhaven.edu/academic-integrity-information-for-faculty/* to directly access the Academic Integrity Policy.

University Student Support Services

The University recognizes that students can often use some help outside of class and offers academic assistance through several offices.

Accessibility Resources Center

The University of New Haven seeks to maintain a supportive academic environment for all students inclusive of those with disabilities including chronic health-related conditions and military service-connected disorders. If you feel that you may need reasonable accommodations to enable your full participation in this course, please provide me with your Verification of Reasonable Accommodations letter through AIM found in MyCharger or contact the Accessibility Resources Center to begin the process to ensure that accommodations can be made available to you. Reasonable accommodations are not required to be provided retroactively and may not be made without written verification from the Accessibility Resources Center. The Accessibility Resources Center is located in Sheffield Hall on the ground floor in the rear of the building, and can be reached by email at ARC@newhaven.edu or by phone at (203) 932-7332.

Center for Learning Resources (CLR)

The Center for Learning Resources (CLR), located on the lower level of the Peterson Library, comprises four units: the **Learning Lab**, which provides content tutoring for all undergraduate students; the **Writing Center**, which offers one-on-one writing support for undergraduate and graduate students; the **Grad Lab**, which provides tutoring and other academic support for graduate students; and the **Learning Assistant Program**, which places high-achieving undergraduate learning assistants in selected sections of challenging courses to help students reach their academic potential.

In addition to course-based tutoring, we offer skills-based tutoring (e.g., for computer programs and programming languages), and we also offer workshops on a wide variety of academic subjects.

The CLR sees between one-third and one-half of the student body in any given year, and we pride ourselves on being an encouraging communal space where all students feel welcome.

To make an appointment, call us at 203-932-7215, write to us at clr@newhaven.edu, download the Navigate app, or just walk in and tell us how we can help.

Center for Student Success (CSS)

The Center for Student Success can help you refine your study skills and develop new academic strategies. CSS staff assists with enhancing your time management and organizational skills. They provide understanding of your GPA, degree audit, and transcripts, and can answer general questions about academic policies. They also can connect you to campus resources and assist you with resolving issues as they arise. During registration periods, CSS advisors work in conjunction your faculty advisor to provide assistance with the advising and registration process. Finally, at various points throughout the semester, CSS works to provide students with progress reports from their instructors. Students can make an appointment to see a CSS staff member through Navigate; the Center for Student Success can be reached via email at css@newhaven.edu.

Counseling & Psychological Services (CAPS)

CAPS offers confidential, mental health care which is included with tuition in order to support student mental health and wellbeing. The services include individual and group therapy, support groups, consultations, and 24/7 crisis support. We are available in person at Charger Plaza and remotely, and are in the office M-F, 8:30-4:30. Please call us to schedule an appointment or with any questions at 203-932-7333; you can also schedule online. If you experience a mental health crisis after hours, you can call our main number for support.

Myatt Center

The Myatt Center for Diversity and Inclusion is committed to creating a multicultural environment through intentional education, campus community engagement, and valuing the unique identities of each member of the Charger Community. Our commitment to diversity is driven by the core values of connection, belonging, inclusivity, equity, acceptance, and accountability. The Myatt Center's focus is to create a respectful and inclusive environment based our awareness and ability to engage with others who are different on many levels including ethnicity, race, sexual orientation, gender, military, religious belief, and life experiences.

Our space is open Monday through Friday from 8:30 – 9:00 P.M for students and organizations to meet, perform schoolwork or to find a safe place for downtime. If you are looking for further information on mentorship, multicultural programmatic opportunities or educational opportunities, please contact us at cdi@newhaven.edu or stop by!

Marvin K. Peterson Library

The library is here to support all members of the University with their research needs. We are proud to serve as an academic hub on campus; providing high quality resources, space for study and exploration, and staff expertise to support all stages of the research process.

Library staff can assist in locating relevant sources of information for course assignments and projects from intro courses to graduate level assignments. To work with a librarian, you can stop by the Library Service desk on the main level of the Peterson Library, use our "Ask a Librarian" chat service, or email libraryhelp@newhaven.edu. Appointments for in-person or online consultations can also be made by using the Navigate app or completing the "Research Consultation" form on the library website.

For 24/7 online support, library staff have created LibGuides to assist with research and citation. These guides contain overviews of resources available through the library by subject, as well as tutorials on a variety of research related topics.

The Peterson Library has three floors with a mix of individual quiet study space, collaborative group study space, and bookable study rooms with technology. The main level of the library provides access to scanners, printers, and both Dell and iMac computers.

Military & Veteran Services

The Military & Veteran Affairs team is here to answer any questions Student Veterans (both current and prospective), active duty/reserve/national guard members, and military family members have regarding transitioning to higher education, VA educational benefits, formal advising, or to listen to issues pertaining to class. The University of New Haven's Military & Veterans Affairs team consists of full-time staff, part time student employees, and VA Work Study students whose aim is to assist and support the student veteran population both on and off campus. These individuals have a dedication to the development, success, and well-being of the student veteran population on campus which includes veterans, active-duty military, service members in the reserves or national guard, and dependents using VA Education Benefits. The office advises, guides, and supports this student population and is available to assist at a moment's notice to address the needs and concerns of this population.

Certified Green Zones

The following offices and resources are "Green Zone Certified," as staff members equipped to assist and support student veterans and military affiliated students and their unique needs. For more information on Green Zone Training please email veteranservices@newhaven.edu.

- Accessibility Resources Center
- Center for Learning Resources (CLR)
- Center for Student Success (CSS)
- Counseling & Psychological Services (CAPS)
- Marvin K. Peterson Library
- Military & Veteran Affairs