CIS 600: Fundamentals of Data and Knowledge Mining

Description

This course will introduce popular data mining and statistical methods for extracting knowledge from data. The principles and theories of data mining methods will be discussed and will be related to the issues in applying data mining to real-world problems. Students will also acquire hands-on experience using Python programming language to develop data mining solutions for scientific, social, or business problems. The focus of this course is on understanding data and how to identify the right data mining techniques and formulate data mining tasks in order to solve problems using the data aided by data mining techniques.

Additional Course Description

The topics of the course will include the key tasks of data mining, including data preparation, concept description, exploration and discovery, statistical methods, visualization, association rule mining, classification, clustering, performance evaluation and analysis, and anomaly detection. Through the exploration of the concepts and techniques of data mining and practical exercises, students will develop skills that can be applied to data challenges.

Credits

3

Learning Objectives

Upon completion of this course, students are expected to be able to:

- Understand the fundamental processes, concepts, and techniques of data mining
- Develop familiarity with data mining techniques and be able to apply them to real-world problems
- Advance your understanding of contemporary data mining systems

Bibliography/Texts/Supplies: Required

Pang-Ning T., Steinbach, M., Anuj Karpatne & Kumar, V. (2018). *Introduction to data mining*. (Free sample chapters available at authors' website https://www-users.cs.umn.edu/~kumar001/dmbook/index.php)

TEXTBOOKS (OPTIONAL)

Jake VanderPlas. *Python Data Science Handbook.* (https://jakevdp.github.io/PythonDataScienceHandbook/)

Francois Chollet. *Deep learning with Python.* (2017) ISBN 9781617294433

Grading

Your final grade is determined by your performance on the items in the table below.

Assessment Item	Weight %
Homework assignments (3)	40
Final exam	20
Final project and presentation	30
Live Session Participation	10
Total	100

Homework assignment

Three homework assignments will be given throughout the semester. Assignments must be professionally prepared and submitted electronically to the LMS. All assignments should be submitted on time. Late submissions will be penalized by 10% of the grade for every additional 24 hours' delay until all the points are deducted. Your homework schedule could be found in the course schedule on the last page of this syllabus.

Exams

The final exam will be taken asynchronously during the weeks as specified in the course schedule. You will have two full days to complete and submit the exam. Note: there is no mid-term exam for the course.

Final Project

The objective of the project is to apply the data mining techniques covered in the course to solve real-world data mining problems. Students can choose the subject matters and datasets for the project. Students will work in groups with up to FOUR members. The final project deliverables consist of a paper and a presentation. The final project report should describe the data mining problem, its significance and broader impact, the data mining approaches and methods, results, and interpretation of the discovered patterns. At the end of the semester, groups will present their findings (duration of the presentation to be determined) during week 10 live session.

While all components must be submitted, they will not be graded individually; they serve the purpose of allowing your instructor to track your progress. You will receive one overall grade for the final project at the end of the semester. For more information on the final project, please consult your final project assignment guide.

Grade	Points	Grade	Points	Grade	Points	Grade	Points
		B+	87–89	C+	77–79	D	60–69
A	93–100	B	83–86	C	73–76	F	0–59
A–	90–92	B–	80–82	C–	70–72		

Participation

You are expected to constructively participate in class. Actively contribute to the class discussions, and feel free to ask questions.

Communications

This course will use the 2U course management site as the main communication platform. Students are required to check their SU accounts on a regular basis. Important announcements will be posted to the Announcements board, which automatically sends the announcement to students' Syracuse e-mail accounts. Failure to read the class announcements will not be considered a suitable excuse for not being informed.
All e-mails to the instructor should be sent with **subject line starting with "CIS600."**

The course wall will be primarily used for making course-related announcements. Please reach out to the faculty in case of any questions rather than posting those questions on the course wall to avoid important course announcements buried in other discussion threads.

Course Specific Policies on Attendance, Late Work, Make-Up Work, Examinations If Outside Normal Class Time

Academic Integrity Policy

Syracuse University's academic integrity policy reflects the high value that we, as a university community, place on honesty in academic work. The policy defines our expectations for academic honesty and holds students accountable for the integrity of all work they submit. Students should understand that it is their responsibility to learn about course-specific expectations, as well as about university-wide academic integrity expectations. The university policy governs appropriate citation and use of sources, the integrity of work submitted in exams and assignments, and the veracity of signatures on attendance sheets and other verification of participation in class activities. The policy also prohibits students from submitting the same written work in more than one class without receiving written authorization in advance from both instructors. The presumptive penalty for a first instance of academic dishonesty by an undergraduate student is course failure, accompanied by a transcript notation indicating that the failure resulted from a violation of academic integrity policy. The presumptive penalty for a first instance of academic dishonesty by a graduate student is suspension or expulsion. SU students are required to read an online summary of the university's academic integrity expectations and provide an electronic signature agreeing to abide by them twice a year during pre-

term check-in on MySlice. For more information and the complete policy, see http://academicintegrity.syr.edu/.

Disability-Related Accommodations

If you believe that you need accommodations for a disability, please contact the Office of Disability Services (ODS), http://disabilityservices.syr.edu/, located in Room 309 of 804 University Avenue, or call (315) 443-4498, TDD: (315) 443-1371 for an appointment to discuss your needs and the process for requesting accommodations. ODS is responsible for coordinating disability-related accommodations and will issue students with documented Disabilities Accommodation Authorization Letters, as appropriate. Since accommodations may require early planning and generally are not provided retroactively, please contact ODS as soon as possible.

Syracuse University values diversity and inclusion; we are committed to a climate of mutual respect and full participation. My goal is to create learning environments that are useable, equitable, inclusive, and welcoming. If there are aspects of the instruction or design of this course that result in barriers to your inclusion or accurate assessment or achievement, I invite any student to meet with me to discuss additional strategies beyond accommodations that may be helpful to your success.

Religious Observances Notification and Policy

SU religious observances notification and policy, found at http://hendricks.syr.edu/spiritual-life/index.html, recognizes the diversity of faiths represented among the campus community and protects the rights of students, faculty, and staff to observe religious holidays according to their tradition. Under the policy, students are provided an opportunity to make up any examination, study, or work requirements that may be missed due to a religious observance provided they notify their instructors before the end of the second week of classes for regular session classes and by the submission deadline for flexibly formatted classes.

For fall and spring semesters, an online notification process is available for students in **My Slice/StudentServices/Enrollment/MyReligiousObservances/Add a Notification**. Instructors may access a list of their students who have submitted a notification in My Slice Faculty Center.

Student Academic Work Policy

SU policy on student academic work may be found at:
http://coursecatalog.syr.edu/content.php?catoid=3&navoid=270#Student_Academic_Work

Student work prepared for university courses in any media may be used for educational purposes, if the course syllabus makes clear that such use may occur. You grant permission to have your work used in this manner by registering for, and by continuing to be enrolled in, courses where such use of student work is announced in the course syllabus.

If you intend to use student work for educational purposes during the current semester:

Educational use of student work: I intend to use academic work that you complete this semester for educational purposes in this course during this semester. Your registration and continued enrollment constitute your permission.

If you intend to use student work for educational purposes in subsequent semesters:

Educational use of student work: I intend to use academic work that you complete this semester in subsequent semesters for educational purposes. Before using your work for that purpose, I will either get your written permission or render the work anonymous by removing all your personal identification.

Course Schedule

Week	Session/Unit	Readings
1	Data Mining Overview	Chapter 1
2	Data Types, Preparation and Exploration	Chapter 2
3	Decision Tree Induction	Chapter 3.1 - 3.3
4	Model Evaluation	Chapter 3.4 - 3.8
5	Bayes' Theorem	Chapter 4.4 - 4.5
6	KNN and Ensemble Learning	Chapter 4.3, 4.10
7	Artificial Neural Network and Deep Learning	Chapter 4.7-4.8
8	Association Rule Learning	Chapter 5.1 - 5.3, 5.7
9	Cluster Analysis: *K*-Means and Hierarchical	Chapter 7.1 - 7.3
10	Alternative Cluster Analysis and Performance Evaluation	Chapter 7.4 - 7.5
11	Final Exam Week	No Live Session

Course Assignments' Key Dates

Week	Assignment	Live Session Time	Assignment Due Date
Week 1		9-Oct	
Week 2	HW1 released	16-Oct	
Week 3	HW1 due	23-Oct	26-Oct
Week 4	HW2 released	30-Oct	
Week 5	HW2 due	6-Nov	9-Nov
Week 6	HW3 released	13-Nov	
Week 7	HW3 due	20-Nov	23-Nov
Week 8	Final Project Proposal due	27-Nov	27-Nov
Week 9		4-Dec	
Week 10	Final Project Slides due	11-Dec	11-Dec
Week 11	Final Project Report due	NA	16-Dec
	Final Exam	NA	18-Dec