**College of Computing and Informatics, Drexel University**

**INFO 371 Data Mining Applications**

**(3-0-3)**

**Term:** 2021-2022 Fall

**Instructor:** Dr. Yuan An

**Office:** Room 1112, 3675 Market Street

**Office Hours:** By appointment

**Contact:**[ya45@drexel.edu](mailto:ya45@drexel.edu)

**Course Location**:

**Course Time**: TR 11:00am-12:20pm

**Contacting the Instructor**

Email is the best way to contact me. **Be sure to include a subject line, and start the subject with a reference to the course.**  For example: “(INFO 371) Question” would work. Email without a clear subject may be deleted with spam

**Catalog Course Description**

Introduces students to basic data mining approaches using machine learning tools. Focuses on machine learning algorithms for information inference and knowledge discovery from data. Covers major applications in data/text/web processing, analysis and mining.

Curriculum Role

This is a required or elective course for Information Science BS majors. It can be taken as an elective by other CCI majors and may be of interest to students in other majors in the University. It offers students concepts and tools to deal with technical challenges related to (big) data processing and mining.

Course Rationale

It has become crucial to automatically store, process, mine, and analyze growing volumes of data for information discovery and decision-making. This course offers an algorithmic perspective on data processing, analysis, and mining. It will not only expose students to state-of-the-art data mining techniques but also prepare them with ways of thinking in the design and implementation of machine-learning-based solutions. Both theory and practice will be emphasized with domain applications.

**Course Outcomes**

Upon successful completion of this course, a student will be able to:

* Describe major machine learning approaches to data processing and mining
* Apply machine learning theories and concepts in finding data mining solutions.
* Design automated systems to infer information and discover knowledge from data.
* Evaluate data mining processes and technologies in domain applications.
* Use or integrate machine learning tools to support data (pre-)processing, mining, and analysis.

**Prerequisites**

STAT 201 Introduction to Business Statistics

**Recommended Textbooks**

* Robert Layton*. Learning Data Mining with Python. Packt, 2017.* (Available through Drexel Library)

**Additional References**

* Wes McKinney*. Python for Data Analysis. O’Reilly, 2017.* (Available through Drexel Library)
* Ian H. Witten, Eibe Frank, and Mark A. Hall*. Data Mining: Practical Machine Learning Tools and Techniques (Third or Fourth Edition). Morgan Kaufmann.*
* Kevin P. Murphy*. Machine Learning A Probabilistic Perspective.* MIT Press. 2012

Tentative Schedule

The schedule is tentative and is likely to vary somewhat based on knowledge of students in the class section. The table below shows the initial schedule for the term.

|  |  |  |  |
| --- | --- | --- | --- |
| **Week** | **Topics** | **Reading (Chapters in Third Edition)** | **Tasks** |
| 1 | * Introduction to machine learning and data mining * Concepts and examples | Chapter 1 |  |
| 2 | * Types of machine learning * Classifying with Scikit-Learn Estimator * K-nearest neighbors | Chapter 2 | Assignment 1 |
| 3 | * Performance evaluation * Data preprocessing, cleansing, and transformation * Decision tree | Chapter 3 |  |
| 4 | * Feature engineering * Decision tree * Ensemble and Random Forest | Chapter 3 | Assignment 2 |
| 5 | * Association rule mining * A-priori algorithm | Chapter 4 |  |
| 6 | * A-priori algorithm and association rule mining in Python | Chapter 4 | Assignment 3 |
| 7 | * Text representation * Introduction to text mining | Chapters 6 |  |
| 8 | * Probabilistic model * Naïve Bayes classification | Chapter 6 | Project Proposal |
| 9 | * Data mining applications: author analysis * Text mining | Chapter 9 | Assignment 4 |
| 10 | * Unsupervised learning * Data clustering * K-Means | Chapter 10 |  |
| 11 | * Evaluation | Chapter 10 | Final Project report |

**Grading**

Your course grade will be based on: **homework assignments, team project, and an exam**. The grade is computed as follows:

|  |  |  |
| --- | --- | --- |
| **Task** | **Weight** | **Due** |
| Assignment 1 | 15% | Week 2 |
| Assignment 2 | 15% | Week 4 |
| Midterm Exam | 20% | Week 5 |
| Assignment 3 | 15% | Week 6 |
| Assignment 4 | 15% | Week 8 |
| Team project | 20% | End of the term |

All work is graded on a numeric scale of 0 to 100. Assignments are graded based on their content (whether you answered the questions correctly and well) and their form (did you produce a submission that has a professional appearance that follows course requirements). A single grade is computed for the assignment based on those two factors.

Conversion from points to letters is given in the following table:

|  |  |  |  |
| --- | --- | --- | --- |
| >=99 | A+ | 77-79 | C+ |
| 93-98 | A | 73-76 | C |
| 90-92 | A- | 70-72 | C- |
| 87-89 | B+ | 67-69 | D+ |
| 83-86 | B | 63-66 | D |
| 80-82 | B- | <=62 | F |

## Exam

## There will be one midterm exam in this course. Tentatively, the midterm exam will be held in week 5. The specific date and the format of the exam will be announced in class.

**Assignments**

The assignments will be designed as individual work*.*

**Course Project**

## The course project can be done by teams each of which consists of up to 3 students.

**Submitting Assignments**

1. You must submit **assignments** electronically via Blackboard Learn **no later than 11:59pm** on the due day. The **name(s)** and **student ID(s)** of all submitters of the assignment along with the **course number** and **assignment number** must be clearly printed on the first page.
2. University rules and policies regarding academic honesty are followed to the letter.

**Late or Missing Work**

Each item of course work you turn in has a due date and time:

* Deliverables are due at the time and date indicated by instructions. Written assignments are due no later than **11:59pm** on the due day.
* There will be a 10% (absolute value) deduction for each day of lateness, to a maximum of 3 days; assignments will not be accepted beyond that point. Missing work will earn a zero grade.

**Re-Marking**

If you are dissatisfied with a grade or point deduction, you can request re-marking. All re-marking requests must be done through written (paper or email) descriptions of why you think the grade is in error. Please note that it is very rare that changing disputed grades on assignments or exams actually affects a student's calculated letter course grade. The corrections are usually insignificant. If you wish to appeal a course grade, school policies apply.

**Academic Honesty**

The Drexel University Academic Honesty Rules and Procedures (as stated in the student handbook) will be adhered to strictly. Students who commit plagiarism or cheat on assignments may receive an F grade for both the assignment and the course.

Academic Dishonesty Policy: <http://www.drexel.edu/provost/policies/academic_dishonesty.asp>

In order to avoid plagiarizing material, observe the following:

If you work on an assignment with another student or a group of students, be certain that your final, individual paper is your own work or the work of your project group (for group projects) unless otherwise specified by the professor. While you might want to discuss the assignment with other students, you must, in your paper, express your own ideas in your own way.

If you use printed or electronic resources in your papers, be sure to attribute the sources you have used. This can be done by quoting the material or by paraphrasing the material and, in either case, listing the source in an annotated bibliography. Use standard notation when citing references.

**Class Attendance and Participation**

You will get the most out of this course if you come to class on time, participate and conduct yourself in a professional manner. Skipping classes will hinder your ability to do well in the course and to learn this important subject. You will also miss handouts, assignments, and announcements. Participation in class is important; there are no stupid questions or comments.

Class attendance is expected. Attendance includes arriving when the class is scheduled to start and staying for the duration of the class period. Roll may be taken in this course and you are expected to acknowledge your attendance at least 90% of the time that attendance is taken in the course**. I reserve the right to lower any earned course grade if a student fails to meet attendance requirements.**

**Disabilities**

**Accommodation of Special Needs** - Students with disabilities requesting accommodations and services at Drexel University need to present a current accommodation verification letter (AVL) to faculty before accommodations can be made. AVL's are issued by the Office of Disability Services (ODS). For additional information, contact ODS at 3201 Arch St., Street, Suite 210, Philadelphia, PA 19104, 215.895.1401 (V), or 215.895.2299 (TTY). <http://www.drexel.edu/oed/disabilityResources/students/>

**Withdrawal of the Course**

For dropping or withdrawing from the course, please refer to the university policies at:

<http://www.drexel.edu/provost/policies/course_drop.asp> <http://drexel.edu/provost/policies/course-withdrawal/>

**Class Cancellation**

On rare occasions, instructors may be delayed or unable to attend a scheduled class due to unforeseen circumstances. In the event that an instructor does not appear in class and has not notified the class of his/her expected arrival time, class is cancelled 15 minutes after the scheduled start of class. More information about class cancellations can be found at:

<http://drexel.edu/provost/policies/cancellation_instructor_absence/>

**Class Lecture Recording**

Lectures and class discussions may be audio-recorded and streamed or rebroadcast for educational purposes only.

**Incomplete Policy**

Incomplete grades are contingent upon instructor approval and will only be considered in extenuating circumstances beyond the student’s control. The instructor is under no obligation to offer an incomplete grade. At least 80% of the graded coursework must have already been completed in order for an incomplete grade to be considered (per the recommendation of the Provost’s Office). An incomplete contract with an instructor-determined due date for delivery of the completed work must be completed by the student and the instructor. It can be found here:

<http://www.drexel.edu/provost/policies/pdf/forms/incomplete.pdf>.

**Support for Equality and Diversity**

Drexel University strives to promote an environment of equality of opportunity and compliance with University policies and federal, state and local laws prohibiting discrimination based upon race, color, religion, gender (sex), marital status, pregnancy, national origin, age, disability and veteran status. Students, faculty, and staff with questions about or complaints concerning discrimination, harassment, and/or retaliation should contact the Office of Equality and Diversity at (215) 895-1403 or <http://www.drexel.edu/oed/>

**Student Conduct and Community Standards**

Drexel University expects that all students as well as student organizations will conduct themselves responsibly and in a manner that reflects favorably upon themselves and the University. For more information see: <http://www.drexel.edu/studentlife/community_standards/overview/>

**Strategies for Success**

* Attend class regularly and complete all assignments on time.
* Complete the assigned reading for each week before the class of that week.
* Take good notes in class.

**Course Evaluation**

Your feedback about the course and instructor is the only way instructors and academic units can improve the quality of a course and its content. Courses administered by the College of Computing and Informatics are evaluated electronically via AEFIS. Students will receive all necessary information via email by the 8th week of classes (or by the 4th week of classes in case of accelerated courses). The evaluations are entirely confidential and will preserve your anonymity.

**Syllabus Changes**

The instructor reserves the right to make changes to this syllabus if circumstances warrant such change. All changes will be provided to students in writing.

**Guidance to COVID-19**

Guidance regarding COVID-19 can be found at Drexel’s Response to Coronavirus website: <https://drexel.edu/coronavirus/>