

Course Prefix and Number: CAP 5771

Course Title: Data Mining

CRN-52897, Section 700, Credit Hours: 3

COURSE SYLLABUS

Semester: Summer C 2023

Dates: 5/15/2023- 7/21/2023

Delivery method: Online, Off-campus

Location: Online / Off-campus

Instructor: Lawrence O. Hall

Instructor email: lohall@mail.usf.edu

TA/GA: name

TA/GA email:

Virtual Office Hours: Technology used (Microsoft Teams): Tuesday 2-4, Thursday 12-1 or by appointment.

**Minimum Technical Skills & Requirements:**

*In order to take courses online at USF, you will need to be able to demonstrate proficiency at basic computer skills, know some programming language and be able to use statistical software packages, maintain reliable internet access, and meet the* [*computer system requirements*](https://www.usf.edu/it/remote/requirements-for-students) *listed at:*

<https://www.usf.edu/it/remote/requirements-for-students>

# University Course Description

An introductory course to mining information from data. Scalable supervised and unsupervised machine learning methods are discussed. Methods to visualize and extract heuristic rules from large databases with minimal supervision is discussed.

# Course Prerequisites

|  |  |
| --- | --- |
|  | Undergraduate Statistics. |

# Course Purpose

This course will enable you to understand how to build models of data sets. Learning models from data is an important part of AI. In this practical course, you will learn to intelligently analyze data and interpret models of data. The use of supervised machine learning and machine learning without labels will be covered. You will learn classic approaches in using association rules and clustering for unlabeled data. We will cover neural networks and deep neural networks for use with big data. At the end of this course you will have the tools to do data scientist, data mining tasks.

# Instructor Contact Information and Communication

Instructors and TA’s can be contacted by email or phone. Email should be sent from your USF email (not through Canvas to avoid delays). We will also use Piazza for peer to peer communication monitored and augmented by instructors and the TA. Use the Piazza link in the left navigation menu to access our shared instance of Piazza. Check your email at least a couple of times a day.

# First Day Attendance Policy

After watching introductory video(s), there will be a short quiz you must take to avoid being dropped from the class.

# How to Succeed in this Course

Brush up on basic statistics. Be sure you remember how to take a derivative (for neural network learning). Read about the material to be covered in the book before watching a video on the topic. Pay attention to the details as they are critical to scoring well on tests.

# Course Structure

This course is fully online. There will be some small homework projects, quizzes and two exams to evaluate student understanding.

# Course Objectives

* Build models of data sets.
* Intelligently analyze data and interpret models of data.
* Demonstrate supervised machine learning.
* Demonstrate association rules and clustering for use when data lacks labels.

# Student Learning Outcomes

By the end of this course, you will be able to:

* Determine what is in a data set that could be used in a machine learning scenario.
* Choose a learning algorithm and apply it to a data set.
* Determine statistical significance of results and generally evaluate the quality of the result of applying a data mining algorithm.
* Apply deep learning algorithms in appropriate ways to suitable data.

# Required Texts and/or Readings and Course Materials

The textbook is: *Data Mining* Fourth Edition by Ian H. Witten and Eibe Frank, Morgan Kaufmann Publishers. We will use the Weka data mining tool for some experiments. It is downloadable from: <https://waikato.github.io/weka-wiki/downloading_weka/>

For deep learning we will use Python tools in Google Co-lab. Directions on using them will be provided in Canvas.

# Grading Scale

Grading Scale (%)

90-100 A

80 - 89 B

70 - 79 C

60 - 69 D

0 - 59 F

# Grade Categories and Weights

**Graded Items Percent of Final Grade**

Quizzes & Projects 30%

Final Exam 35%

Midterm 35%

# Project Assignments\*

Project 1: Compare several learning algorithms on the same data set in Weka.

Project 2: Use Weka to compare several learning algorithms with 10, 10-fold cross validation followed by the corrected resampled T-Test.

Project 3: Create Association rules from data in Weka.

Project 4: Use Weka to cluster unseen data.

Project 5: Use Google Colab and python code to try solving the exclusive or problem.

Project 6: Deep learning application where you choose parameters to maximize performance on a data set.

\* Note projects will be submitted as part of quizzes. The last project will count for part of the Final

# Instructor Feedback Policy & Grade Dissemination

The instructor will respond to email communication relevant to the subject matter within 48 hours of the date received. Instructor will provide feedback on assignments within one week of the posted deadline, and feedback on final papers within two weeks of the posted deadline. [You can access your scores at any time using "Grades" in Canvas](https://community.canvaslms.com/t5/Student-Guide/How-do-I-view-my-grades-in-a-current-course/ta-p/493).

# Course Schedule

|  |  |  |
| --- | --- | --- |
| **Date** | **Study Assignments** | **Practice and Graded Assignments** |
| Week 1 | * Read syllabus, read up on standard deviation * Read Chapter 1,2 - What is Data Mining, Weka example * Watch videos in Module 1 * Read Chapters 2, 3 Data descriptions, Knowledge Representation * Watch Videos in Module 2 | * First Day Attendance due [May 16, 2023] * Take Quiz before 7pm. * Watch video on Piazza and do introduction on Piazza * Take Quiz 1 * Take Quiz 2 |
| **Week 2** | Chapter 3,4 Knowledge Representation, Trees, Weka basics  Watch Videos in Module 3.  Chapter 4, notes in Weka, Algorithm basics, supervised learning, Association rules. Watch Videos in Module 4. Pay special attention to Weka Video | Take Quiz 3  Take Quiz 4 |
| **Week 3** | Chapter 4, 5.1-5.5, simple neural networks, instance-based learning and unsupervised clustering and and Evaluation.  Watch Videos in Module 5 | Take Quiz 5 |
| **Week 4** | Chapter 5.6-12, 6.1 Credibility and Decision Trees, Watch Videos in Module 6 Chapter 6.2, 6.3, Ripper Rules, Advanced Association Rules, Review for exam  Watch Videos in Module 7 | Take Quiz 6  Take Quiz 7 |
| Week 5 | Chapter 7.1-2 Instance based learning, Support Vector Machines, Test 1  Watch Videos in Module 8 | Take Exam (Test 1) |
| Week 6 | Chapter 7.2 - 7.3 Neural Networks, Numeric prediction  Watch Videos in Module 9  Chapter 8, Attribute Selection and Modification  Watch Videos in Module 10 | Take Quiz 8  Take Quiz 9 |
| Week 7 | Chapter 10, Deep Neural Networks overview  Watch Videos in Module 11 | Take Quiz 10 |
| Week 8 | Chapter 10, notes, Big data and deep networks  Watch Videos in Module 12 | Do Autoencoders assignment |
| Week 9 | Notes, Modern convolutional neural nets/applications, Final Project assigned  Watch Videos in Module 13  Chapter 12, Utilizing ensembles of classifiers, Bagging and Boosting, Snapshot learning and augmentation  Watch Videos in Module 14 | Take Quiz 11  Take Quiz 12 |
| Week 10 | Active and Semi-supervised learning and Multi-Instance Learning, Review for Final  Watch Videos in Module 15 |  |
| **7/21 Fri.** |  | **Final Exam, 2:30pm-4:30pm and Project due** |

\* Note: The Schedule is subject to revision

# USF Core Syllabus Policies

USF has a set of central policies related to student recording class sessions, academic integrity and grievances, student accessibility services, academic disruption, religious observances, academic continuity, food insecurity, and sexual harassment that **apply to all courses at USF**. These may be accessed on the [USF Core Syllabus Policy Statements page](https://www.usf.edu/provost/faculty/core-syllabus-policy-statements.aspx) at https://www.usf.edu/provost/faculty/core-syllabus-policy-statements.aspx.

# Course Policies: Grades (as applicable)

## Late Work Policy

No late work is accepted!

## 

## Extra Credit Policy

There is no extra credit!

## Grades of “Incomplete”

The current university policy concerning incomplete grades will be followed in this course.

*For graduate courses:* An Incomplete grade ("I") is exceptional and granted at the instructor’s discretion only when students are unable to complete course requirements due to illness or other circumstances beyond their control. The course instructor and student must complete and sign the "I" Grade Contract Form that describes the work to be completed, the date it is due, and the grade the student would earn factoring in a zero for all incomplete assignments. The due date can be negotiated and extended by student/instructor as long as it does not exceed two semesters for undergraduate courses and one semester for graduate courses from the original date grades were due for that course. An "I" grade not cleared within the two semesters for undergraduate courses and one semester for graduate courses (including summer semester) will revert to the grade noted on the contract.

## Make-up Exams Policy

If a student cannot take an examination for a valid reason (validity to be determined by the instructor), a make-up exam will be given only if the student has notified the instructor in advance that s/he cannot be present for the exam. Make-up exams are given at the convenience of the instructor.

## Exam Retention Policy

After exams are graded, the instructor will review the examination during a live synchronous session. The exams will be digitally retained for one semester following the current one, and then they will be destroyed.

## Final Examinations Policy

All final exams will be scheduled in accordance with the University’s final examination policy*.*

# Course Policies: Technology and Media (as applicable)

## Canvas

## This course is offered via USF's learning management system (LMS), Canvas. If you need help learning how to perform various tasks related to this course or other courses being offered in Canvas, please consult the [Canvas Student Guides](https://community.canvaslms.com/t5/Student-Guide/tkb-p/student). You may also contact USF's IT department at (813) 974-1222 or [help@usf.edu](mailto:help@usf.edu)*.*

You should regularly check Canvas for any updates on assignments which will also be communicated in Piazza (so do use it).

## Online Proctoring

All students must review the syllabus and the requirements, including the online terms and video testing requirements, to determine if they wish to remain in the course. Enrollment in the course is an agreement to abide by and accept all terms. Any student may elect to drop or withdraw from this course before the end of the drop/add period. Online exams and quizzes within this course may require online proctoring. Therefore, students will be required to have a webcam (USB or internal) with a microphone when taking an exam or quiz. Students understand that this remote recording device is purchased and controlled by the student and that recordings from any private residence must be done with the permission of any person residing in the residence. To avoid any concerns in this regard, students should select private spaces for the testing. The University library and other academic sites at the University offer secure private settings for recordings and students with concerns may discuss location of an appropriate space for the recordings with their instructor or advisor. Students must ensure that any recordings do not invade any third-party privacy rights and accept all responsibility and liability for violations of any third-party privacy concerns. Students are strictly responsible for ensuring that they take all exams using a reliable computer and high-speed internet connection. Setup information will be provided prior to taking the proctored exam. To use Honorlock students are required to download and install the [*Honorlock Google Chrome extension*](https://static.honorlock.com/install/extension). For additional information please visit the [*USF online proctoring student FAQ*](https://www.usf.edu/innovative-education/digital-learning/digital-learning-resources/proctoring-student-faq.aspx) and [*Honorlock student resources*](https://honorlock.com/students/).”

# Course Policies: Student Expectations

## Title IX Policy

Title IX provides federal protections for discrimination based on sex, which includes discrimination based on pregnancy, sexual harassment, and interpersonal violence. In an effort to provide support and equal access, **USF has designated all faculty (TA, Adjunct, etc.) as Responsible Employees, who are required to report any disclosures of sexual harassment, sexual violence, relationship violence or stalking.** The Title IX Office makes every effort, when safe to do so, to reach out and provide resources and accommodations, and to discuss possible options for resolution.  Anyone wishing to make a Title IX report or seeking accommodations may do so online, in person, via phone, or email to the Title IX Office. For information about Title IX or for a full list of resources please visit: <https://www.usf.edu/title-ix/gethelp/resources.aspx>. *If you are unsure what to do, please contact Victim Advocacy – a confidential resource that can review all your options – at 813-974-5756 or*[*va@admin.usf.edu*](mailto:va@admin.usf.edu)*.*

## Course Hero / Chegg Policy

The [USF Policy on Academic Integrity](https://usf.app.box.com/v/usfregulation3027) specifies that students may not use websites that enable cheating, such as by uploading or downloading material for this purpose. This does apply specifically to Chegg.com and CourseHero.com – any use of these websites (including uploading proprietary materials) constitutes a violation of the academic integrity policy.

## End of Semester Student Evaluations

All classes at USF make use of an online system for students to provide feedback to the University regarding the course. These surveys will be made available at the end of the semester, and the University will notify you by email when the response window opens. Your participation is highly encouraged and valued.

## Netiquette Guidelines

* Act professionally in the way you communicate. Please treat your instructors and peers with respect, the same way you would do in a face-to-face environment. Respect other people’s ideas and be constructive when explaining your views about points you may not agree with.
* Be sensitive. Be respectful and sensitive when sharing your ideas and opinions. There will be people in your class with different linguistic backgrounds, political and religious beliefs or other general differences.
* Proofread and check spelling. Doing this before sending an email or posting a thread on a discussion board will allow you to make sure your message is clear and thoughtful. Avoid the use of all capital letters, it can be perceived as if you are shouting, and it is more difficult to read.
* Keep your communications focused and stay on topic. Complete your ideas before changing the subject. By keeping the message on focus you allow the readers to easily get your idea or answers they are looking for.
* Be clear with your message. Avoid using humor or sarcasm. Since people can’t see your expressions or hear your tone of voice, meaning can be misinterpreted.

## Email and Discussion Board Guidelines

1. Use the subject line effectively by using a meaningful line of what your email or discussion is about.
2. Keep your emails and postings related to the course content. You should not post anything personal on a discussion board, unless is requested by the instructor.
3. Any personal, course or confidential issues should be directly communicated to the instructor via email. The discussion boards are public spaces; therefore, any issues should not be posted there.

# Learning Support and Campus Offices

## Academic Accommodations

Students with disabilities are responsible for registering with Student Accessibility Services (SAS) in order to receive academic accommodations. For additional information about academic accommodations and resources, you can visit the SAS website.

[SAS website for the Tampa and Sarasota-Manatee campuses.](https://www.usf.edu/student-affairs/student-accessibility/)

[SAS website for the St. Pete campus.](https://www.stpetersburg.usf.edu/student-life/resources/student-accessibility-services/)

## Academic Support Services

The USF Office of Student Success coordinates and promotes university-wide efforts to enhance undergraduate and graduate student success. For a comprehensive list of academic support services available to all USF students, please visit the [Office of Student Success website](https://www.usf.edu/student-success/undergrads/).

## Canvas Technical Support

If you have technical difficulties in Canvas, you can find access to the Canvas guides and video resources in the “Canvas Help” page on the homepage of your Canvas course. You can also contact the help desk by calling 813-974-1222 in Tampa or emailing[*help@usf.edu.*](mailto:help@usf.edu)

[IT website for the Tampa campus.](https://www.usf.edu/it/)

[IT website for the St. Pete campus.](https://lib.stpetersburg.usf.edu/c.php?g=728128&p=5199235)

[IT website for the Sarasota-Manatee campus.](https://www.sarasotamanatee.usf.edu/academics/academic-resources/technology-services/index.aspx)

## Center for Victim Advocacy

The [Center for Victim Advocacy](https://www.usf.edu/student-affairs/victim-advocacy/contact-us/index.aspx) empowers survivors of crime, violence, or abuse

by promoting the restoration of decision making, by advocating for their rights, and by

offering support and resources. Contact information is available online.

## Counseling Center

The Counseling Center promotes the wellbeing of the campus community by providing culturally sensitive counseling, consultation, prevention, and training that enhances student academic and personal success. Contact information is available online.

[Counseling Center website for the Tampa campus.](https://www.usf.edu/student-affairs/counseling-center/about-us/contact-us.aspx)

[Counseling Center website for the St. Pete campus.](https://www.stpetersburg.usf.edu/student-life/wellness/about/schedule-appointment.aspx)  
[Counseling Center website for the Sarasota-Manatee campus.](https://www.sarasotamanatee.usf.edu/campus-life/health-and-safety/counseling-and-wellness-center/counseling-services.aspx)

# Important Dates to Remember

All dates and assignments are tentative and can be changed to maximally facilitate learning at my discretion. For important USF dates, see the [Academic Calendar](http://www.usf.edu/registrar/calendars/) at <http://www.usf.edu/registrar/calendars/>

*Drop/Add Deadline:                Fri, May 19, 2023*

*Memorial Day Holiday:           Mon, May 29, 2023*

*Mid-term Grading Opens:      Mon, Jun 19, 2023*

*Juneteenth Holiday: Mon, Jun 19, 2023*

*Mid-term Grading Closes       Tue, Jun 27, 2023*

*Withdrawal Deadline:            Sat, Jul, 1, 2023*

*Independence Day Holiday:  Tue, Jul 4, 2023*

*Summer C Classes end:      Fri, Jul 21, 2023*