

Data Mining

CAP-5771-001

Syllabus, Fall 2020

When: M,W 11:00 AM-12:15 PM

Where: CPR 115

Instructor: Larry Hall, (lohall@mail.usf.edu)

Office: ENB 330, Hrs. in Canvas M 3:00-4:00, W: 9-10am, R 1:00-2:00pm or by appointment.

Phone: 974-4195 or 974-3652 (Dept.)

TA: Ahmad Babaeian Jelodar (ajelodar@usf.edu) Virtual Office Hours: Th 3:30-5:00 pm, F 3:30-5:00 pm

All students must comply with university policies and posted signs regarding COVID-19 mitigation measures, including wearing face coverings and maintaining social distancing during in-person classes. Failure to do so may result in dismissal from class, referral to the Student Conduct Office, and possible removal from campus.

Grading: There will be two midterms which will each count 30% of the final grade. A final will be 20% of the grade and will include a project portion. Homeworks, projects and quizzes will be 20% of the grade. Grading scale: $A \geq 90$, $B \geq 80$, $C \geq 70$, $D \geq 60$, $F < 60$.

General: The textbook is: *Data Mining* Fourth Edition by Ian H. Witten and Eibe Frank, Morgan Kaufmann Publishers. Each topic should be read about, before the lecture which pertains to it. Available for free electronically at: <http://ezproxy.lib.usf.edu/login?url=https://ebookcentral.proquest.com/lib/usf/detail.action?docID=4708912>

USF Core Syllabus Policy Statements:

<https://www.usf.edu/provost/faculty/core-syllabus-policy-statements.aspx>

No late work is accepted!! Any academic dishonesty will result in an **F** for the course.

Course Objectives:

- o Understand how to build models of data sets.
- o Understand how to intelligently analyze data and interpret models of data.
- o Understand supervised machine learning.
- o Understand association rules and clustering for use when data lacks labels.

Week 1: Chapter 1,2

Week 2: Chapter 2, 3

Week 3: Chapter 3,4

Week 4: Chapter 4, notes

Week 5: Chapter 4, 5.1-5.5

Week 6: Chapter 5.6-12, 6.1

Week 7: Chapter 6.2,

Week 8: Chapter 6.3, 7.1-2

Week 9: Chapter 7.3

What is Data Mining, Weka example

Data descriptions, Knowledge Representation

Knowledge Representation, Trees, Weka basics

Algorithm basics, supervised learning & Association rules

Clustering and Evaluation

Probabilities, Cost, ROC analysis, Decision Trees

Test, Ripper Rules

Association Rules, Instance based learning, Support

Vector Machines, neural networks

Numeric prediction

Week 10: Chapter 8	Attribute Selection and Modification
Week 11: Chapter 10	Deep Neural Networks overview
Week 12: Chapter 10, notes	Big data and deep networks, Test 2
Week 13: notes	Modern convolutional neural nets/applications
Week 14: Chapter 12	Utilizing ensembles of classifiers, Bagging and Boosting
	Snapshot learning and augmentation
Weeks 15: Chapter 11	Active and Semi-supervised learning
	and Multi-Instance Learning
Final: Monday December 7: 10:00am-12pm	