

CSC 5380/4380 – Data Mining

Session: Spring/2022

Classes:

4380-001 (11197) - MoWe 2:30PM - 3:50PM - NTDP B192
5380-001 (8750) - MoWe 2:30PM - 3:50PM - NTDP B192
5380-002 (13217) - TuTh 2:30PM - 3:50PM - NTDP B190

Instructor:

Dr. Moawia Eldow

Office: E250E

Tel: TBA

Email: moawia.eldow@unt.edu

Office hours: MoWe_12pm – 2pm

TAs/Graders:

4380-001/5380-001:

Name: Sumukha Parameshwara Bhat (Grader)

Email: SumukhaParameshwaraBhat@my.unt.edu

Office & Office hours: F232, Mo 1:00PM - 3:00PM

5380-002:

Name: Sita Swapna Lakshmi Challapalli (Grader)

Email: SitaSwapnaLakshmi.Challapalli@unt.edu

Office & Office hours: TBA

Description:

This course focuses on fundamental concepts, principles and techniques related to data mining. We will study important topics of data mining, including data preprocessing, frequent pattern and association rule mining, classification, clustering, anomaly detection, and some recent developments and trends.

Prerequisites:

Programming with one of the high-level languages such as C, C++, or Java; Introductory courses on data structures and algorithm, linear algebra and probability theory.

Textbooks:

- 1- Introduction to Data Mining, 2nd Edition, by Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. (IDM – Chapters 1, 2, 3, 4, 5, 7 and 9)
- 2- Data Mining: Concepts and Techniques, 3rd Edition, by Jiawei Han, Micheline Kamber and Jian Pe. (DMCT – Chapter 13)

Grades and grading policy:

Homework Assignments	30%
Project	30%
Midterm Exam	20%
Final Exam	20%

The letter grade will be assigned based on the following scale:

<u>Grade</u>	<u>4380</u>	<u>5380</u>
A	85 and Above	90 and Above
B	[75-85)	[80-90)
C	[65-75)	[70-80)
D	[55-65)	[60-70)
F	Below 55	Below 60

Homework Assignments:

Written individual homework assignments/exercises **will be due at 11:59 p.m. on Fridays.** Assignments must be turned in using the dropbox on canvas.

Project:

Group project must be submitted and may be presented in the class (2 to 4 students per project). Presentation of the projects is mandatory for graduates (CSCE5380) and optional for undergraduates (CSCE4380). Components of the project may include a proposal and final project report. Projects must be turned in using the dropbox on canvas. Instructions will be posted in Canvas.

Exams:

There will be one midterm exam during the semester at the normal lecture time, which will cover the first half of the class topics. There will also be a final exam during finals week, which will cover the second half of the class topics.

Late Submission Policy:

Assignments may be turned in late, but not more than two weeks. All the late submissions **may lose a percentage of their graded point values** according to the following schedule:

On time :	0%
1-3 days :	10%
4-7 days :	20%
8-14 days :	40%
> 14 days :	100%

Announcements

Stay tuned and make sure to check Canvas frequently. Important announcements will be posted there.

Academic Policies

No cheating or plagiarism is allowed in assignments and exams. Academic dishonesty will result in a final course *grade of "F"*. "Sharing/reuse" of solutions to assignment problems is strictly prohibited. All work turned in with your name on it must be your own work.

Other Policies:

Students should refer to any other policies from university, college and department.

CSCE 5380/4380 - Course Outline (Tentative Schedule):

<u>Week</u>	<u>Reading chapters and Topics</u>	<u>Homework (HW) & Project (Due dates)</u>
Jan 17-21	Overview of Class Ch1 – Introduction	
Jan 24-28	Ch2 – Data & Data Preprocessing	
Jan 31-Feb 4	Ch3 – Basic Concepts & Decision Trees	HW1 (02/11/2022) Project (<u>Instructions posted on Canvas</u>)
Feb 7-11	Ch3 – Model Overfitting & Evaluation	
Feb 14-18	Ch4 – Rule-based Classifiers	HW2 (02/25/2022) Project (<u>Proposal & groups are due on 02/18/2022</u>)
Feb 21-25	Ch4 – Naïve Bayes Classifier & Belief Networks	HW3 (03/04/2022)
Feb 28 -Mar 4	Ch4 – K-Nearest Neighbor & Logistic Regression	
<u>Mar 7-11</u>	Mid-Term Exam	03/07/2022 (MoWe class) 03/08/2022 (TuTh class)
Mar 14-18	Spring Break (No classes)	
Mar 21-25	Ch4 - Artificial Neural Networks	
Mar 28-April 1	Ch4 – Support Vector Machine & Class Imbalance Problem	HW4 (04/08/2022)
April 4-8	Ch5 – Basic Association Analysis	HW5 (04/15/2022)
April 11-15	Ch7 – Basic Cluster Analysis	HW6 (04/22/2022)
April 18-22	Ch9 – Anomaly Detection	Project (<u>Final reports are due on 04/22/2022</u>)
April 25-39	Ch13 - New Trends on Data Mining (DMCT) & Project Presentations	
May 2-6	Review and Project Presentations Cont.	
<u>May 9-13</u>	Final Exam (TBA)	