University of North Texas
Department of Computer Science and Engineering
CSCE 4380/5380 – Data Mining
Session: Spring/2023

Classes:

4380-001 (**8444**) - MoWe 2:30PM - 3:50PM - NTDP B192
5380-001 (**7004**) - MoWe 2:30PM - 3:50PM - NTDP B192
5380-002 (**9429**) - TuTh 2:30PM - 3:50PM - NTDP B190

Instructor:
Dr. Moawia Eldow
Office: **E250E**
Email: **moawia.eldow@unt.edu**
Office hours: **TuTh 12:30-2:30 PM**

TAs/Graders:

4380-001/5380-001:
  Name: Himanvitha Katragadda
  Email: HimanvithaKatragadda@my.unt.edu
  Office & Office hours: TBA

5380-002:
  Name: Varun Reddy Kaitha
  Email: VarunReddyKaitha@my.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, 4th Edition, by Jiawei Han, Micheline Kamber and Jian Pe. (DMCT – Chapter 12)

Grades and grading policy:

Participation	**5%**
Homework Assignments	**30%**
Project	**25%**
Midterm Exam	**20%**
Final Exam	**20%**

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

Grade	4380/5380
A	90 and Above
B	[80-90)
C	[70-80)
D	[60-70)
F	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 **optiona**l 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 polices from university, college and department.

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

Week	Reading chapters and Topics	Homework (HW) & Project (Due dates)
Jan 16-20	Overview of Class Ch1 – Introduction	
Jan 23-27	Ch2 – Data & Data Preprocessing	
Jan 30-Feb 3	Ch3 – Basic Concepts & Decision Trees	**HW1 (02/10/2023)** **Project (Instructions posted on Canvas)**
Feb 6-10	Ch3 – Model Overfitting & Evaluation	
Feb 13-17	Ch4 – Rule-based Classifiers	**HW2 (02/24/2023)** **Project (Proposal & groups are due on 02/17/2023)**
Feb 20-24	Ch4 – Naïve Bayes Classifier & Belief Networks	**HW3 (03/03/2023)**
Feb 27-Mar 3	Ch4 – K-Nearest Neighbor & Logistic Regression	
Mar 6-10	**Mid-Term Exam (Online):** *03/06/2023 (MoWe class - 4380-1/5380-1), Available from 2:00 pm to 4:30 pm* *03/07/2023 (TuTh class – 5380-2), Available from 2:00 pm to 4:30 pm*	
Mar 13-18	**Spring Break (*No classes*)**	
Mar 20-24	Ch4 - Artificial Neural Networks	
Mar 27-31	Ch4 – Support Vector Machine & Class Imbalance Problem	**HW4 (04/07/2023)**
Apr 3-7	Ch5 – Basic Association Analysis	**HW5 (04/14/2023)**
Apr 10-14	Ch7 – Basic Cluster Analysis	**HW6 (04/21/2023)**
Apr 17-21	Ch9 – Anomaly Detection	**Project (Final reports are due on 04/21/2023)**
Apr 24-28	Ch12 (DMCT) - Trends on Data Mining Project Presentations	
May 1-5	Review Week	
May 8-12	**Final Exam (Online):** *05/08/2023 (MoWe class - 4380-1/5380-1), Available from 1:00 pm to 3:30 pm* *05/09/2023 (TuTh class – 5380-2), Available from 1:00 pm to 3:30 pm*	