

CSCE 411/811 – Data Modeling for Systems Development

Spring 2020

Course Info

Instructor	Dr. Mohammad Rashedul Hasan hasan@unl.edu Avery Hall 123E https://cse.unl.edu/~hasan/
Office Hours	See Canvas
Textbook	See Canvas
Teaching Assistants	See Canvas

Prerequisite

CSCE 310 or 311. Exposure to multiple programming languages as well as ability to pick up Python and JavaScript. Familiarity with computer-based problem-solving skills, data structures and algorithms, complexity analysis. Probability & Statistics.

Description

This course provides students with practical experience of data analysis and data modeling techniques. More specifically, data analysis topics such as clustering, anomaly detection, classification, regression, dimensionality reduction will be covered in greater detail. Students will learn how to model data through existing techniques including relational and non-relational models. These models can then be used at the center of systems to promote efficient and effective data processing and analysis.

Course Topics

This is a rough sequence of topics that we will cover in this course.

- Data Analysis
 - Exploratory Data Analysis
 - Clustering (K-Means, Gaussian Mixture Model, DBSCAN, etc.)
 - Anomaly/Outlier Detection (GMM, DBSCAN, Local Outlier Factor, Isolation Forest, Minimum Covariance Determinant, etc.)
 - Association-rule Mining
 - Predictive Analytics (Classification & Regression)
 - Complexity of Data Analysis; Possible solutions: Algorithmic & Hardware
- Data Modeling
 - Data modeling and data visualization
 - Low-Dimensional Data Visualization: using D3.js

- High-Dimensional Data Visualization: using Dimensionality Reduction techniques
- Text Analytics, NLP
- Modeling Structured Data
 - o Normalization
 - o End-to-End Data Analysis Workflow
 - o Conceptual modeling
 - o Physical modeling & Memory/storage hierarchy (ISAM, B+ Trees)
 - o E-R Modeling (logical & physical)
- Modeling Semi-Structured Data
 - o NoSQL
 - o Graph Stores, XML, JSON
- Data Analysis & Modeling
 - Focus on visualization
 - Focus on efficiency issues

Text Resources

- Lecture Slides (thorough and extensive) should provide a detailed account of the topics.

Though there are **no required textbooks** for this course. However, following books are recommended.

Data Analysis:

- Machine Learning: A Probabilistic Perspective by Kevin P. Murphy
- Pattern Recognition and Machine Learning by Christopher M. Bishop
- Introduction to Machine Learning (3rd Edition) by Ethem Alpaydin
- Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems (2nd Edition) by Aurélien Géron (**O'Reilley**)

Data Modeling:

- Data Modeling Essentials (3rd Edition) by Graeme Simsion & Graham Witt
- Database Design for Mere Mortals: A Hands-On Guide to Relational Database Design (3rd Edition) by Michael J. Hernandez
- Data Modeling Made Simple: A Practical Guide for Business and IT Professionals (2nd Edition) by Steve Hoberman, Carol Lehn, Michael Blaha, Bill Inmon, Graeme Simsion
- Database Systems: The Complete Book (2nd Edition) by Hector Garcia-Molina, Jeffrey D. Ullman, Jennifer Widom
- Interactive Data Visualization for the Web: An Introduction to Designing with D3 (2nd Edition) by Scott Murray (**O'Reilley**)
- Graph Databases: New Opportunities for Connected Data (2nd Edition) by Ian Robinson, Jim Webber, Emil Eifrem (**O'Reilley**)

The eBook version of any textbook published by O'Reilley can be found from the UNL library.

811 Level

There will be higher expectations with regard to the quality and quantity of work for the 811 level students compared to the 411 level students. As a student in this course, you will be expected to meet or exceed those expectations, which may require more of your time and a more proactive approach to studying and greater diligence in completing assignments. In particular, each assignment contains additional requirements or activities. In the 411 level, these will be opportunities for extra credit/bonus points. However, in the 811 level, these are *required* and will be treated as part of your regular grade.

Grading

Grading will be based on homework assignments, project, and quality of presentation. Note that your course total score will be computed based on the stipulated weights (see below) associated with the following graded components. **The Canvas default total score does not reflect your weighted total score.**

411 Level:

Assignments	85%
Final Project	15%

811 Level:

Assignments	80%
Final Research Project	20%

Scale

Letter grades will be awarded based on the following scale. I will use an **absolute grading policy**. For example, scores within the range between 89.00 and 89.49 will be rounded to 89.00 while scores within the range between 89.50 and 89.99 will be rounded to 90.00. Therefore, a score of 89.49 will earn you a B+ grade but 89.50 will earn an A-(minus) grade.

A+	>= 97	B+	>= 87	C+	>= 77	D+	>= 67	F	<60
A	>= 93	B	>= 83	C	>= 73	D	>= 63		
A-	>= 90	B-	>= 80	C-	>= 70	D-	>= 60		

Homework Assignments

There will be 4/5 programming assignments. Assignments are due at the beginning of class on the due date. Code and other relevant files must be submitted using CSE's webhandin (<http://cse.unl.edu/handin>). If there are written portions of the programming homework, they should be typed. You should typeset code snippets using a monotype font (Courier for example) for readability. Figures may be hand drawn, but you are encouraged to use some sort of software to render them.

Coding

There will be **no support for the coding, compilation, or debugging.**

Final Project

You will have to do a **programming-based end-to-end project** incorporating data analysis and modeling techniques for systems development on your chosen topic. For 811 students, the project must demonstrate **significant depth**. The project proposal must be approved in the middle of the semester and the results must be presented during the last two weeks of semester.

Exams

There are no exams in this course.

Grading Policy

We make every effort to grade and return materials within one academic week of the due date. You are expected to pick up your graded material in a timely manner. If you have questions about grading or believe that points were deducted unfairly, you must first address the issue with the grader it to see if it can be resolved. Such questions should be made within a reasonable amount of time after the graded assignment has been returned. **No further consideration will be given to any assignment a week after it has been graded and returned (regardless of whether you fail to pick it up when handed back).**

It is your responsibility to **check Canvas regularly** and report missing/incorrect entry within one week of submission of your assignment or exam.

You need to report any missing/incorrect Canvas grade entry or questions related to grading **before I post your final grade** in Canvas. **I will not take any question/complaint after the final grade is posted.**

Late Work

In general, late submission will not be allowed. Exceptions may be made in certain circumstances such as health or emergency, but you must make every effort to get prior permission. Documentation may also be required.

Homework assignments have a strict *in-class* (at the beginning) due date. Any written portions should be handed in hardcopy in class while softcopies should be handed in using webhandin as specified in individual homework assignments. The webhandin program that you will use enforces a *strict* handin time based on the CSE server's clock. Programs that are even a few seconds past the due date/time will be considered late.

It is understandable that unforeseen events may interfere with your ability to submit all homework assignments on time. As such, this course allows the following late work policy: you may hand in **any one assignment up to 7 days late**. Any submissions after 7 days will not be considered and will be given an automatic zero. These 7 “free passes” can be used on multiple assignments.

Exceptions: You *CANNOT* use a late pass for the first written and programming assignment as well as on the last assignment and the final project.

Academic Integrity

All homework assignments, programs, and exams must be your own work unless otherwise stated. No collaboration with fellow students, past or current, is allowed unless otherwise permitted on specific assignments or problems. The Computer Science & Engineering department has an Academic Integrity Policy. All students enrolled in any computer science course are bound by this policy. You are expected to read, understand, and follow this policy. Violations will be dealt with on a case-by-case basis and may result in a failing assignment or a failing grade for the course itself. **If any student performs academic dishonesty (complete or partial), as defined by the UNL academic integrity policy, in any graded component of the course, her/his grade will be set to zero for that assignment/exam and she/he will be reported to the academic integrity committee of the CSE department.** The most recent version of the Academic Integrity Policy can be found at

<http://cse.unl.edu/academic-integrity-policy>

Communication

The best way to communicate with your Professor is through email. The Professor and teaching assistants will communicate with you either directly or through the Canvas email system. You are responsible for ensuring that the email associated with your Canvas account is up-to-date and that you are regularly checking it. Generally speaking, you should talk to Dr. Hasan for questions about course material and lecture. You should talk to the TA about grading and homework assignments.

In addition, the Department of Computer Science & Engineering also maintains an anonymous suggestion box that you may use to voice your concerns about any problems in the course or department if you do not wish to be identified. It is available at the following URL:

<http://cse.unl.edu/department/suggestion.php>

Finally, I will hold regular office hours and will make myself available by appointment; please email me to set one up.

Disabilities

Students with disabilities are encouraged to contact Christy Horn for a confidential discussion of their individual needs for academic accommodation. It is the policy of the University of Nebraska-Lincoln to provide flexible and individualized accommodation to students with documented disabilities that may affect their ability to fully participate in course activities or to meet course requirements. To receive accommodation services, students must be registered with the Services for Students with Disabilities (SSD) office, 132 Canfield Administration, 472-3787 voice or TTY.