University of North Carolina at Charlotte
College of Computing and Informatics
Department of Software and Information Systems

Course Number and Title: DSBA/HCIP 6162: Knowledge Discovery in Databases (KDD), Spring 2024

Credits: 3 Graduate Credits

Time, Days, and Location:
12:30 pm-3:15 pm on Tuesdays at the College of Education 010

Faculty Information: Xi (Sunshine) Niu, Ph.D., Associate Professor
Office: Woodward 310G
Email: xniu2@charlotte.edu

Teaching Assistant: 1. Ms. Sravya Reddy Gangi Reddy
Email: sgangir2@charlotte.edu

2. Ms. Kajari Bhaumil
Email: kbhaumik@charlotte.edu

Course Description:

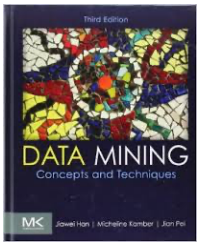
This course is about data mining. It is an essential part of AI, which is one of the hottest topics in computer science today. Data mining is a fast-evolving field, especially for recent five years.

The availability of large amounts of data has created unprecedented opportunities to leverage computational and statistical approaches to turn data into actionable knowledge. This course covers general techniques for analyzing large amounts of **numerical** and **text** data. The entire data mining process is covered in this course: setting up a problem, data preprocessing, model constructions, model evaluations, and interpretations in decision making.

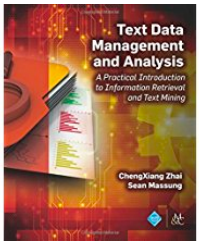
This course covers both classical data mining approach (e.g., Apriori, Random Forest, etc) as well as the recent deep learning models (e.g., RNN, CNN, BERT). In addition, the recent rise of large language models (LLMs), especially ChatGPT, has brought global excitement. We have LLMs as one of our topics.

Required Textbooks and Papers:

We will use **academic papers, online learning materials,** and the **textbooks** as our learning materials. For all the papers, Dr. Niu will provide the full-text versions. For textbooks, Dr. Niu will provide the electronic copies. The two textbooks are listed as below:



Title: Data Mining: Concepts and Techniques
Author(s): Jiawei Han, Micheline Kamber, and Jian Pei
Edition: 3rd Edition
Publisher: Morgan Kaufmann
Year: 2011



Title: Text Data Management and Analytics: A Practical Introduction to Information Retrieval and Text Mining
Author(s): ChengXiang Zhai and Sean Massung
Publisher: ACM and Morgan & Claypool Publishers
Year: 2016

Evaluation Methods:

Course grading will be based on these activities.

Activities	**Point**
In-Class Quizzes	3 points x 13 = 39 points
After-Class Homework	4 points x 14 = 56 points
Class attendance	5 points
Total	**100 points**

Grade Scale:
A = 90 points – 100points
B = 80 points – 89 points
C = 70 points – 79 points
U = Below 70 points

Weekly Lesson Schedule:

Date	Contents
Jan 16	Syllabus Lesson 1: Getting to Know Your Data
Jan 23	Lesson 2: Principal Component Analysis
Jan 30	Lesson 3: Pattern Mining
Feb 6	Lesson 4: Machine Learning 1
Feb 13	Lesson 5: Machine Learning 2
Feb 20	Lesson 6: Cluster Analysis
Feb 27	Lesson 7: Word Association Mining
Mar 5	Student Spring Recess – No Classes
Mar 12	Lesson 8: Topic Modeling
Mar 19	Lesson 9: Introduction to Deep Learning
Mar 26	Lesson 10: RNN and CNN
Apr 2	Lesson 11: Word Embedding
Apr 9	Lesson 12: Transformers and BERT
Apr 16	Lesson 13: Large Language Models 1
Apr 23	Lesson 14: Large Language Models 2

Course Policies:

Course Credit Workload:
This 3-credit course requires 9-12 hours effort (including the class time) for this course each week for 14 weeks. Efforts may include but is not limited to: required reading, homework assignments, and studying for quizzes.

Class Attendance Policy:
Attending every class is mandatory. Class attendance entails being prepared, present, and attentive for the entire class period. Two or fewer absence (excused or unexcused) is OK without losing the class attendance points (5 points), but you will lose the in-class quiz points for those absence(s). Three absences will lose your class attendance points (5 points). Four or more absences in total will result in U in the course. For each absence, the student is responsible for catching up with all covered materials and assignments.

Late Submissions:
For assignments, unexcused late submission (according to the Canvas timestamp and the "late" flag) will receive a grade of 0. You should plan sufficiently for completing and submitting assignments. Should an emergency arise that greatly disrupts one's ability to complete an assignment, please send an email to Dr. Niu **before** the due date with a plan for submission after the due date. You need to receive Dr. Niu's permission for late submission.

Special Needs and Religious Accommodation:
If you have a documented disability and require accommodation in this course, contact the Office of Disability Services (https://ds.uncc.edu/students/academic) the first week of the semester. Accommodations for learning will be arranged by that office and communicated to Dr. Niu.
It is the obligation of students to provide faculty with reasonable notice of the dates of religious observances on which they will be absent by submitting a Request for Religious Accommodation Form to their instructor prior to the census date for enrollment for a given semester. The census date for each semester (typically the tenth day of instruction) can be found in UNC Charlotte's Academic Calendar (https://registrar.uncc.edu/printable-calendar).

Copyright and Permissions:
My lectures and course materials, including presentations, quizzes, code demonstrations, homework problems and answers are protected by copyright. I am the exclusive owner of copyright in those materials I create. You may make copies of course materials for your own educational use. However, you may not, nor may you knowingly allow others to reproduce or distribute these materials publicly without my written consent. This includes providing materials to commercial websites such as CourseHero, Chegg, and other similar services. Students who publicly distribute or display or help others publicly distribute or display copies or modified copies of my course materials are in violation of University Policy 406, The Code of Student Responsibility.

University Policies:

Code of Student Responsibility: https://legal.uncc.edu/policies/up-406
Code of Student Academic Integrity: https://legal.uncc.edu/policies/up-407
Diversity and Inclusion: https://diversity.uncc.edu/
Sexual Misconduct and Interpersonal Violence:
https://legal.uncc.edu/policies/up-502
Standard for Responsible Use: https://oneit.uncc.edu/iso/standard-responsible-use