Course Number and Title: ITIS 6162: Knowledge Discovery in Databases, Fall 2019

Credits, Days/Time, Location: 3 Graduate Credits; Thursdays 2:30–5:15 p.m. in CHHS 281 or online

Faculty Information: Xi(Sunshine) Niu, Ph.D., Assistant Professor
Office: Woodward 310G
Office Hours: By appointment
Email: xniu2@uncc.edu (preferred way of contact)

TA: Ms. Chandni Jivani
Email: cjivani@uncc.edu
Office Hours: 2pm – 4pm Wednesdays at Woodward 331

Catalog Description – include Pre and/or Co-requisites:
The entire knowledge discovery process is covered in this course. Topics include: setting up a problem, data preprocessing, data mining in search for knowledge, knowledge evaluation, visualization and application in decision making.

Prerequisite(s): ITCS 6160 or permission of instructor.
Cross-listed as: DSBA 6162, HCIP 6162

Course Objectives:
About two thirds of the course content focuses on numeric data analytics and the other third text data analytics. Both R and Python will be used as the toolkits. Topics include: setting up a problem, data preprocessing, statistical analyses, pattern mining, classification, clustering, and several common text mining and analytics techniques. The main focuses of this course are both concepts and implementations/applications.

Specifically, there are 8 objectives:
- **State** the key stages in data mining process
- **Recall** concepts in each stage
- **Recognize** each data mining technique
- **Explain** the differences between supervised learning and unsupervised learning
- **Use** Python or R to construct data mining models
- **Interpret** the reasons behind the Python or R coding
- **Compare** different data mining techniques
- **Criticize** each machine learning technique

Teaching Strategies: This course is in a hybrid format with 13 lessons in total. 9 of them are face-to-face and the other 4 lessons are online.
For the face-to-face lessons:
Teaching methods include a combination of lectures for concepts, demos for R or Python implementations, and in-class quizzes for learning assessment, and activities for sharing ideas. The class time is used in this way:
- 2:30pm ~ 3:20pm Lecture
- 3:20pm ~ 3:30pm Break
- 3:30pm ~ 4:20pm Lecture
- 4:30pm ~ 5:15pm Quiz & Reflections

Students should complete the weekly assigned readings before each class. During the class, usually toward the end of each class, the students will be given a quiz. After the class, students need to finish the homework and work on group assignments. During the semester, students will form groups of three and work on a mid-term toolkit demo and a final project.

For the online lessons:
Each online class will have a series of short videos and reading materials, after that, the students need to complete a quiz and a homework. In addition, the students are encouraged to participate in online discussion forum through Piazza.

Required Textbooks:
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

Title: Discovering Statistics Using R
Author(s): Andy Field, Jeremy Miles, and Zoe Field
Publisher: Sage Publication Ltd
Year: 2012
(The instructor will provide the electronic copies of all the textbooks)

**Evaluation Methods:**

**Course grading will be based on these activities.**

<table>
<thead>
<tr>
<th>Activities</th>
<th>Point</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quizzes</td>
<td>2 points x 13 = 26 points</td>
</tr>
<tr>
<td>Exercises</td>
<td>3 points x 13 = 39 points</td>
</tr>
<tr>
<td>Midterm demo (group)</td>
<td>10 points</td>
</tr>
<tr>
<td>Final presentation (group)</td>
<td>10 points</td>
</tr>
<tr>
<td>Final report (group)</td>
<td>10 points</td>
</tr>
<tr>
<td>Peer reviews*</td>
<td>2.5 points x 2 = 5 points</td>
</tr>
</tbody>
</table>

* Peer reviews: each student needs to provide their reviews for other group’s midterm demos and final presentations. Based on the quality of the provided reviews, the student will earn up to 2.5 points for the midterm demos and final presentations respectively.

**Grade Scale:**

A = 90 - 100
B = 80 - 90
C = 70 - 80
U = Below 70

**Topical Outline (Weekly Schedule):**

<table>
<thead>
<tr>
<th>Date</th>
<th>Lecture</th>
<th>Assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lesson 1: Aug 23,</td>
<td>• Discussing Syllabus</td>
<td>• Homework 1</td>
</tr>
<tr>
<td>Lesson 2: Aug 30,</td>
<td>• Data Preprocessing</td>
<td>• Reading: Field Chapter 6.1-6.5 and Field Chapter 7.1-7.9</td>
</tr>
<tr>
<td>Lesson 3: Sept 6,</td>
<td>• Linear Regression &amp; Logistic Regression</td>
<td>• Homework 3</td>
</tr>
<tr>
<td></td>
<td>• Getting to Know Your Data</td>
<td>• Reading: Han Chapter 1-3</td>
</tr>
<tr>
<td></td>
<td>• Getting Familiar with R</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Getting Familiar with Python</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Quiz 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Quiz 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Quiz 3</td>
<td></td>
</tr>
</tbody>
</table>
Course Policies

**Attendance for the face-to-face lessons is mandatory.** A basic requirement of this course is that you will participate in all class meetings, and complete all required course activities and

| Lesson 4: Sept 13, Friday | • Pattern Mining  
|                          | • Quiz 4  
|                          | • Homework 4  
|                          | • Reading: Han Chapter 8  |
| Lesson 5: Sept 20, Friday | • Machine Learning 1  
|                          | • Quiz 5  
|                          | • Homework 5  
|                          | • Reading: Han Chapter 9  |
| Lesson 6: Sept 27, Friday | • Machine Learning 2  
|                          | • Quiz 6  
|                          | • Homework 6  |
| Lesson 7: Oct 4, Friday – Oct 11, Friday | • Word Association Mining  
|                          | • Quiz 7  
|                          | • Homework 7  |
| Lesson 8: Oct 11, Friday – Oct 18, Friday | • Topic Analysis I  
|                          | • Quiz 8  
|                          | • Homework 8  |
| Midterm Day Oct 25, Friday | • Midterm Toolkit Demo Presentation  |
| Lesson 9: Oct 25, Friday – Nov 1 Friday | • Topic Analysis II  
|                          | • Quiz 9  
|                          | • Homework 9  |
| Lesson 10: Nov 1, Friday – Nov 8, Friday | • Text Clustering & Categorization  
|                          | • Quiz 10  
|                          | • Homework 10  |
| Lesson 11: Nov 8, Friday | • Deep Learning with TensorFlow  
|                          | • Quiz 11  
|                          | • Homework 11  
|                          | • Read Field Chapter 17  |
| Lesson 12: Nov 15, Friday | • PCA  
|                          | • Quiz 12  
|                          | • Homework 12  
|                          | • Read Han Chapter 10  |
| Lesson 13: Nov 22, Friday | • Cluster Analysis  
|                          | • Quiz 1  
|                          | • Homework 13  |
| Nov 29, Friday | Thanksgiving No class  |
| Final Presentation Day Dec 6, Friday | Final Project Presentation  |
| Final Day Dec 13, Friday | No class  |
assignments. Class attendance is required. It entails being present and attentive for the entire class period. Attendance shall be taken in every class. Missing class reduces your grade through the following grade reduction policy: You are allowed two absences. Each additional absence, results in 5 points reduction from the class participation points. More than three (four or above) absences result in U in the course. For all absences, the student is responsible for all covered materials and assignments. Absences must be explained with the email sent to the instructor and cc’ed to the TA BEFORE the beginning of a class.

Late submissions. For individual work, late submission (according to the Canvas timestamp and the “late” flag) will receive a grade of 0. Group members should plan sufficiently for completing group assignments. Should an emergency arise that greatly disrupts a group’s ability to complete an assignment, permission and documentation must be received BEFORE the due date with a plan for submission after the due date.

Diversity Statement. No student will be discriminated against in the class based upon age, race, nationality, religion, sexual orientation, gender identity/expression, veteran’s status, country of origin, or group affiliation. Likewise, all participants in this class will be expected to respect other members who fall into these categories. Any student who does not behave in a respectful manner with their classmates will be withdrawn from the class.

UNIVERSITY, COLLEGE AND DEPARTMENTAL POLICIES

Code of Student Responsibility:
“The UNC Charlotte Code of Student Responsibility (the Code) sets forth certain rights and responsibilities in matters of student discipline. The Code defines these responsibilities and guarantees you certain rights that ensure your protection from unjust imposition of disciplinary penalties. You should familiarize yourself with the provisions and procedures of the Code” (Introductory statement from the UNC Charlotte brochure about the Code of Student Responsibility). The entire document may be found at this Internet address: http://legal.uncc.edu/policies/ps-104.html

Academic Integrity:
All students are required to read and abide by the Code of Student Academic Integrity. Violations of the Code of Student Academic Integrity, including plagiarism, will result in disciplinary action as provided in the Code. Students are expected to submit their own work, either as individuals or contributors to a group assignment. Definitions and examples of plagiarism and other violations are set forth in the Code. The Code is available from the Dean of Students Office or online at: http://www.legal.uncc.edu/policies/ps-105.html.

Integrity Statement is required for every weekly Homework. At the beginning of each homework, there should be a statement:
“I completed this Homework in compliance with the Code of Student Academic Integrity.” Please sign your electronic signature after this statement.

Course Credit Workload:
This 3-credit course requires three hours of classroom with direct faculty instruction and six hours of out-of-class student work each week for approximately 15 weeks. Out-of-class work may include but is not limited to: required reading, resource research, coding, written assignments, studying for quizzes, and working on group projects.

Special Needs:
If you have a documented disability and require accommodation in this course, contact Disability Services, Fretwell 230, phone: 687 4355 voice/TDD) the first week of the semester. Information about available services may be found at http://legal.uncc.edu/policies/ps-51.html. Accommodations for learning will be arranged by that office and communicated to the Instructor. If you speak English as a second language, please inform the instructor.

Diversity Statement:
UNC Charlotte strives to create an academic climate in which the dignity of all individuals is respected and maintained. Therefore, we celebrate diversity that includes, but is not limited to ability/disability, age, culture, ethnicity, gender, language, race, religion, sexual orientation, and socio-economic status.

All students are required to abide by the UNC Charlotte Sexual Harassment Policy (http://www.legal.uncc.edu/policies/ps-61.html) and the policy on Responsible Use of University Computing and Electronic Communication Resources (http://www.legal.uncc.edu/policies/ps-66.html). Sexual harassment, as defined in the UNC Charlotte Sexual Harassment Policy, is prohibited, even when carried out through computers or other electronic communications systems, including course-based chat rooms or message boards.

Religious Accommodation:
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 http://legal.uncc.edu/policies/ps-134.html. The census date for each semester (typically the tenth day of instruction) can be found in UNC Charlotte’s Academic Calendar (http://registrar.uncc.edu/calendars/calendar.htm).