CSCE 474/874: Introduction to Data Mining

Spring 2014

General Information

Instructor

Ashok Samal
Avery Hall, Room 361
Phone: 472-2217
Email: samal@cse.unl.edu
Office Hours: Monday and Wednesday 11:00 – 12:00 (Tentative)

Junjie Qian: Teaching Assistant, jqian@cse.unl.edu
Office Hours: To be announced, Avery 113

Lectures

Location: Avery 119
Time: MW 3:30 - 4:45 pm

Text Book:

Data Mining Concepts and Techniques by Han, Kamber & Pei (Morgan Kaufmann), 3rd Edition, 2012.

Data Mining: Practical Machine Learning Tools and Techniques, by Witten, Frank and Hall, (Morgan Kaufmann), 3rd Edition, 2011. (Optional)

Prerequisites: CSCE310, STAT 380 or permission or equivalent.

1. **Mastery** of a high-level programming language.
2. **Familiarity** with calculus and statistics.
3. **Exposure** to linear algebra.

Course Objectives:

Data mining is the process of automatically finding useful *hidden* patterns in data. The objective of data mining is to use discovered patterns to help *explain* current behaviors or to *predict* the future outcomes. Upon completing this course, students will understand:

- **Mastery** of basic data mining algorithms, including association rule mining, clustering, prediction and classification.
- **Familiarity** with data warehousing and its applications.
- **Familiarity** with role of data mining in knowledge discovery.
- **Familiarity** with data preprocessing and reduction techniques.
- **Exposure** to applications of data mining to practical problems.

Assignments

There will be four assignments for this course. Together they will count for 30% of the overall grade. The assignments are due at the class time on the indicated dates. We will use a flexible, slip date system for the assignments. Each student is given an automatic extension of four (4) calendar days. You can use these on any assignment(s) during the course. However, the total number of late days cannot exceed 4. Slack days cannot be used for the project. After you have used your ``late'' days, if you submit an assignment late, you will not get any credit for it. Then, it is better to submit a partially completed homework than a late one. If you have a special reason for being late, get permission well ahead of the due date.

You will be required to `handin' an online copy. Do not `handin' any hardcopies. You are required to 'handin' all source code, executable, etc. Make sure that you do `handin' all these files.

While it can be informative to discuss the assignments with others, you should write your final programs independently. Make sure that all your files are protected. You are also responsible if somebody copies your files and hands them in. You should also read the policy document on *"Academic Integrity in the Department of Computer Science and Engineering."*

Tests and Quizzes

There will be no exams in this class. However, there will be periodic quizzes (both announced and unannounced). 15% of the overall grade is devoted to quizzes.

Software

We will use Weka (http://www.cs.waikato.ac.nz/ml/weka/index.html) system for assignments (and projects if you choose). You may download a copy for personal use.

Project

Students are required to work on a project on some aspect of data mining. You must select a project from a list assigned in the class. Each team should consist of three students. All project documents will be kept online in Blackboard.

There will be a group project that will be worth 50% of the overall grade.

Attendance

Attendance in class is required. 5% of the overall grade is reserved for class attendance and participation.

Academic Misconduct

Violations of academic integrity will result in automatic failure of the class and referral to the proper university officials. The work a student submits in a class is expected to be the *student's own work and must be work completed for that particular class and assignment*. Students wishing to build on an old project or work on a similar topic in two classes must discuss this with both professors. Academic dishonesty includes: handling in another's work or part of another's work as your own, turning in one of your old papers for a current class, or turning in the same or similar paper for two different classes. Using notes or other study aids or otherwise obtaining another's answers for an examination also represents a breach of academic integrity. Sanctions are applied whether the violation was intentional or not.

Those who share their code and those who copy other's code will be penalized in the same way; both parties will be considered to have plagiarized.

Final Grade

The final grade will be based on the following scale.

≥ 90%: A	**≥ 84%: B+**	**≥ 74%: C+**	**≥ 64%: D+**	
≥ 88%: A-	**≥ 80%: B**	**≥ 70%: C**	**≥ 60%: D**	**< 50%: F**
	≥ 78%: B-	**≥ 68%: C-**	**≥ 50%: D-**	