Advanced Data Mining (CS 573)

Welcome to CS573! This course is intended for graduate students interested in research and applications in data mining. Both first year graduate students and more senior graduate students are welcome. I also motivate undergraduate seniors who are interested in. I expect it to be fun working with you in the coming spring semester! 😊

Instructor

Dr. Yanfang (Fanny) Ye
Email: yanfang.ye@mail.wvu.edu
Office Hours: T/R 05:00pm-06:00pm, or by appointment, at ESB - 935
Lectures: T/R 3:30pm-4:45pm at ESB - E401

Course Description and Objectives

Recent advances in database technology along with the phenomenal growth of the Internet have resulted in an explosion of data collected, stored, and disseminated by various organizations. Because of its massive size, it is difficult for analysts to sift through the data even though it may contain useful information. Data mining holds great promise to address this problem by providing efficient techniques to uncover useful information hidden in the large data repositories, which has recently gained a substantial interest among practitioners in a variety of fields and industries.

The purpose of this graduate course is two-fold: (1) broadening graduate students' knowledge and understanding of the current frontiers of data mining research, and (2) teaching data mining research methodology and skills. The core topics to be covered in this course include data preprocessing, data visualization, association analysis, classification, clustering, anomaly detection and deep learning. This course will also provide extensive hands-on experience in applying the concepts to real-world applications.

Upon successful completion of this course, the students should be able to:

- Understand basic theory of common data mining algorithms.
- Have knowledge of advanced principles required for solving data mining problem.
- Have knowledge of principles required for the design, implementation and analysis of complex data mining experiments.
Be familiar with advanced concepts of data mining.
Be able to communicate ideas effectively: in writing.
Be able to communicate ideas effectively: verbally.
Be able to work and learn effectively as members of a team.
Be exposed to programming data miners.

At the end of this course, students are expected to possess the fundamental theory and skills needed to conduct their own research in data mining or to apply data mining techniques to their own research fields.

**Course Prerequisites**

CS230 and CS35. I do often grant access to the course to people who are interested in! Please write me email or come to see me for the prerequisite waivers.

**Text Book**

Course lectures, notes and other materials will be found under the course materials section on e-campus or WVU MIX System.


**Topics (tentative)**

- Introduction to Data Mining
- Data preparation
- Data visualization
- Data warehouse and data cube
- Association analysis
- Classification
- Clustering
- Anomaly detection
- Deep learning
- Applications in Data Mining
Grading

A 90-100  B 80-89  C 70-79  D 60-69  F ≤60

**Homework (50%)** You will be handed several homework assignments. You may discuss homework with other students, but each student must write up solutions in their own words without assistance from anyone. Any submitted work that it copied from any source or too similar to be an independent write-up will not be given credit.

**Group Project (50%)** You will be assigned one group research project. Project topics could be related to diverse data mining application domains, e.g., bioinformatics, cyber security, social network analysis, or IoT, etc. You will be required to use cutting edge data mining techniques to solve the proposed research problems.

- 2-3 students per group (ideally, 2 students per group)
- Select a seed idea for your group project
- Fully motivate the problem (5%)
- Survey related work (10%)
- Develop your own solutions – substantial novel algorithm development, theoretical analysis, and implementation are expected (25%)
- A thorough empirical evaluation, preferably using large real data sets, and comparing with baseline methods (20%)
- A fully developed project report (25%): 12 pages in ACM SIG Tighter Alternate style: [http://www.acm.org/sigs/publications/proceedings-templates#aL2](http://www.acm.org/sigs/publications/proceedings-templates#aL2)
- Project presentation (15%): 20min presentation + 10min Q/A

*’+’ and ‘-’ grade may be reported if the score is near boundary.*

* There is no mid-term or final exam.

Attendance Policy

Students are expected to attend and participate in class lectures and discussions, and should note that students will be responsible for course material and information that may be conveyed through lectures and class discussion whether or not that material or information is contained in handouts, instructor provided notes, or assigned or optional readings. Students should also note that a significant portion of the course content will be conveyed through class lectures, in-class activities and class discussions. Students should
also note that a portion of the student’s grade will be based on course participation: after three noted absences, students will lose two points per absence; after three noted late arrivals, students will lose one point per late arrival.

**Academic Integrity Statement**

The integrity of the classes offered by any academic institution solidifies the foundation of its mission and cannot be sacrificed to expediency, ignorance, or blatant fraud. Therefore, I will enforce rigorous standards of academic integrity in all aspects and assignments of this course. For the detailed policy of West Virginia University regarding the definitions of acts considered to fall under academic dishonesty and possible ensuing sanctions, please see the Student Conduct Code at: [http://www.arc.wvu.edu/admissions/integrity.html](http://www.arc.wvu.edu/admissions/integrity.html).

Should you have any questions about possibly improper research citations or references, or any other activity that may be interpreted as an attempt at academic dishonesty, please see me before the assignment is due to discuss the matter.

**Social Justice Statement**

West Virginia University is committed to social justice. The instructor concurs and expects to foster a nurturing learning environment, based upon open communication, mutual respect and non-discrimination. Our University does not discriminate on the basis of race, sex, age, disability, veteran status, religion, sexual orientation, color or national origin. Any suggestions to further such a positive and open environment in this class will be appreciated and given serious consideration. If you are a person with a disability and anticipate needing any type accommodation, in order to participate in this class, please advise me of the same and make appropriate arrangements with Disability Services (293-6700). If you feel that you are being treated inappropriately or unfairly in any way, please feel free to bring your concerns to my attention; rest assured that doing so will not prejudice the grading process. In return, I expect you to behave professionally and ethically.