CptS 315 Introduction to Data Mining: Content, Logistics, and Administrivia

Introductions

- **Jana Doppa, Assistant Professor of Computer Science**
- At WSU since Fall-2014
- PhD from Oregon State University (2006-2014, 8 yrs!!)
- Masters from IIT Kanpur (2004-2006)
- Passionate about artificial intelligence , machine learning, data mining, computing and data-driven solutions for real-world applications. Doing this for ~12 years.
- Like teaching courses on these topics

Introductions

- **Jana Doppa, Assistant Professor of Computer Science**
- At WSU since Fall-2014
- PhD from Oregon State University (2006-2014, 8 yrs!!)
- Masters from IIT Kanpur (2004-2006)
- Passionate about artificial intelligence , machine learning, data mining, computing and data-driven solutions for real-world applications. Doing this for ~12 years.
- Like teaching courses on these topics

- Tell the class about you
  - Name, background, and what are your goals for taking this class

Course Contents

- **Introduction to the field of data mining**
  - Automatically analyze data using computers for discovering knowledge and insights

  - Computational problems motivated from real-world applications
  - Computational algorithms to solve data analysis problems
  - Real-world applications for each of the data analysis problems

Tentative Syllabus

- Mining frequent item sets and association rules
- Recommendation algorithms
- Clustering algorithms
- Supervised learning algorithms
  - Classification: predicting discrete outputs
  - Regression: predicting continuous outputs
- Outlier and anomaly detection algorithms
- Graph mining and social network analysis
- Ethics, fairness, and transparency in data mining

Student Learning Outcomes and Assessment

- **Learning Outcomes**
  - Understand the foundations, algorithms, applications, and challenges of data mining
  - Ability to apply basic data mining algorithms for solving real-world problems
- **Assessment Methods**
  - Homework assignments
  - Exams
  - Course project

Course Logistics

- **Class Timings:** Tue and Thu 12 to 1:15pm
- **Class Location:** Sloan 175
  - For make up classes, location may change?
- **Instructor:** Prof. Jana Doppa
  - Office: EME 133
  - Office hours: Tue and Thu 3-4pm (EME 133)
- **Teaching Assistant:** Mohammad Omar Faruk
  - Office hours: Mon and Fri 4-5pm (Sloan 335)

Course Logistics (contd.)

- **Piazza for course announcements and discussions**
  - Great user interface for communication
  - Post all your questions directly on Piazza: you will get faster response from either your classmates or instructor or TA
  - Collaborative learning from the crowd
- **All course material will be posted on Piazza**
  - Lecture notes, homework assignments and solutions, reading materials
- Blackboard will be used ONLY for posting grades

Grading Policy

- **5 Home works (36%)**
  - Best four will be counted towards your grade
- **2 Exams (40%)**
  - Mid-term #1 and Mid-term #2
- **1 Course Project (22%)**
  - Individual or group of two students
- **Class Participation (2%)**
  - Piazza and in-class

Late Policy

- All assignments, project proposal/report are due at the **start** of the class
- **Late Policy**
  - 0-24 hours late -- 80% of the final score
  - 24-48 hours late -- 50% of the final score
  - Beyond 48 hours -- 0%
- If you are late, please slip the assignment through my office door with a time-stamp

Grading Policy

- [100-93] A
- (93-90] A-
- (90-86] B+
- (86-83] B
- (83-80] B-
- (80-76] C+
- (76-73] C
- (73-70] C-
- (70-66] D+
- (66-60] D
- (60-0] F

Note: I may decide to move the thresholds down based on the distribution of final percentages

Course Pre-requisites

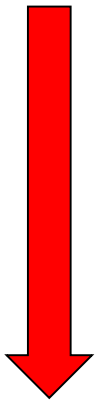
- **Assume strong programming experience**
  - You can choose any high-level programming language of your choice (e.g., Python, Java, C++)
- **Assume knowledge of the following**
  - Basic data structures (CptS 223)
  - Basic algorithms
  - Time/space complexity analysis

Course Materials

- **We will NOT follow a fixed textbook for this course**
  - Instructor will provide slides and lecture notes
  - Slides and reading materials will be posted on Piazza site
- **Optional Textbooks**
  - **[DM]** Data Mining: Concepts and Techniques (Third Edition): Jiawei Han, Micheline Kamber, and Jian Pei
  - **[MMD]** Mining of Massive Datasets: Jure Leskovec, Anand Rajaraman, and Jeff Ullman. Available online at http://infolab.stanford.edu/~ullman/mmds/book.pdf
  - I will assign reading material from MMD

Email Etiquette

- Please use Piazza for all the class related communication as much as possible
- If there is anything you cannot ask on Piazza, please send me an email with subject ``CptS 315: <what are you writing about>''
- You can address me as one of the following:
  - Prof. Jana
  - Prof. Doppa
  - Dr. Jana
  - Dr. Doppa



Academic Integrity

- WSU definitions and procedures for cases of academic dishonesty are given at conduct.wsu.edu
  - These procedures will be followed rigorously
- Do not copy solutions/code from other students and/or internet
  - Will get ``F'' if caught with conclusive evidence
- **Bottom line:** Don't even think of cheating!

Students with Disabilities

- Reasonable accommodations are available for students with documented disability. If you have a disability and may need accommodations to fully participate in this class, please visit the Access Center (Washington Building 217) to schedule an appointment with an Access Advisor. All accommodations MUST be approved through the Access Center. Additional information can be viewed at http://drc.wsu.edu

Campus Safety

- Classroom and campus safety are of paramount importance at Washington State University, and are the shared responsibility of the entire campus population. WSU urges students to follow the "Alert, Assess, Act" protocol for all types of emergencies and the "Run, Hide, Fight" response for an active shooter incident. Remain ALERT (through direct observation or emergency notification), ASSESS your specific situation, and ACT in the most appropriate way to assure your own safety (and the safety of others if you are able). The Campus Safety Plan, which can be found at http://safetyplan.wsu.edu, contains a comprehensive listing of university policies, procedures, statistics, and information relating to campus safety, emergency management, and the health and welfare of the campus community