

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

Introductions

- **Jana Doppa, Huie-Rogers Endowed Chair Associate Professor of Computer Science**
- At WSU since Fall-2014
- PhD from Oregon State University (2014)
- Masters from IIT Kanpur (2006)
- Passionate about artificial intelligence , machine learning, data mining, computing and data-driven solutions for real-world applications. Doing this for ~17 years.
- Like teaching courses on these topics

Introductions

- **Jana Doppa, George and Joan Berry Distinguished Associate 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 ~17 years.
- **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
- Supervised learning algorithms
 - ▲ Classification and regression tasks
- Clustering algorithms
- Outlier and anomaly detection algorithms
- Computational Advertising
- Mining Data Streams
- Responsible data mining

CptS 315 vs. CptS 437 vs. CptS 475

- CptS 437: Introduction to Machine Learning
 - CptS 475/575: Data Science
- Data Mining, Machine Learning, and Data Science are very closely related
 - There will be **some overlapping parts**
 - ▲ Supervised learning (classification and regression)
 - ▲ Clustering algorithms
- CptS 315 is a mandatory class for BS in Data Analytics
 - ▲ These students may not take CptS 437 and/or CptS 475

Student Learning Outcomes and Assessment

- **Learning Outcomes**

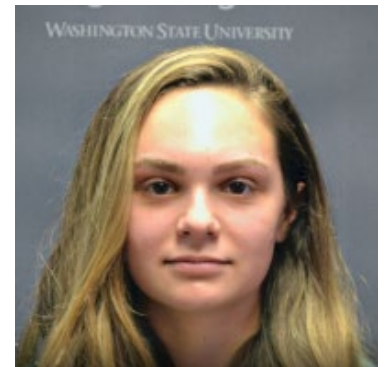
- ▶ Understand the foundations, algorithms, applications, and challenges of data mining
- ▶ Ability to apply basic data mining algorithms for solving real-world problems

- **Assessment Methods**

- ▶ Homework assignments
- ▶ Exams
- ▶ Course project

Course Logistics

- **Class Timings:** Tue and Thu 9:10 to 10:25am
- **Class Location:** Brian 305
 - ▲ Video recordings available through Canvas
- **Instructor:** Prof. Jana Doppa
 - ▲ Office hours: EME 133, Mon 4-5pm
- **Teaching Assistants:**
 - ▲ Chibuike Ugwu (Graduate TA)
 - Dana 122, Wed 2-3pm
 - ▲ Emma Mickas (UG TA)
 - TBD



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
- **Canvas will be used for handling submissions and posting grades**

Grading Policy

- **3 Homeworks (30%)**
- **2 Exams (40%)**
 - ▲ Mid-term #1 and Mid-term #2
- **1 Course Project (25%)**
 - ▲ Individual or group of two students
- **Class Participation (5%). (Easy part. Don't miss it)**
 - ▲ Piazza and in-class
 - ▲ Attendance (Instructor will take attendance on 4 random days. To get attendance related grade, should be present in at least 2 occasions)

Late Policy

- All assignments, project proposal/report are due at midnight of the deadline day.
- **Late Policy**
 - ▲ 0-24 hours late -- 80% of the final score
 - ▲ 24-48 hours late -- 50% of the final score
 - ▲ Beyond 48 hours -- 0%

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

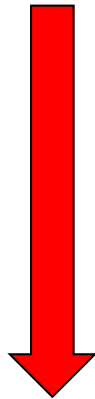
- **Assume strong programming experience**
 - ▲ You can choose any high-level programming language of your choice (e.g., Python, Java, C++)
 - ▲ Strongly encourage you to consider Python (data analytics community is built around Py tools)
- **Assume knowledge of the following**
 - ▲ Basic data structures (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**
 - ▶ **[MMD]** Mining of Massive Datasets: Jure Leskovec, Anand Rajaraman, and Jeff Ullman. Available online at <http://infolab.stanford.edu/~ullman/mmds/book0n.pdf>
 - ▶ **[CIML]** A Course in Machine Learning: Hal Daume
<http://ciml.info/>
 - ▶ I will assign reading material from MMD and CIML

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

Health etc.

- If you are sick or tested positive for COVID, please stay at home. I will record and make the lecture videos available on Canvas. You can watch them at your own convenience.
- If you need accommodation with class deadlines due to health-related challenges, please let me know in advance and I will accommodate.