

Syllabus

Logistics

- Time: Mon/Wed 4pm-5:15pm
- Location: E206 Westgate
- Instructor: [Dongwon Lee](#) (*dul13*). Office Hour: Mon 2-4pm @ E353 Westgate
- TA: [Yiming Liao](#) (*yfl5119*). Office Hour: Wed 2-4pm @ E343 Westgate

Overview

The course will introduce students to the principles of machine learning and data mining, representative machine learning algorithms, and their applications in data sciences. Topics to be covered include:

- Principled approaches to supervised learning
- Principled approaches to unsupervised learning
- Feature engineering
- Dimensionality reduction
- Performance assessment of models
- Relative strengths and weaknesses of alternative algorithms

Representative algorithms to be covered include:

- Naïve Bayes
- Decision trees
- Regression
- SVM
- Ensemble methods (boosting, boosting, random forests)

The course will include several projects to provide students hands-on experience to apply learned algorithms to problems from several domains.

Learning Objectives

At the completion of this course, students are expected to obtain the following:

- Broad understanding of the principles of data mining/machine

learning, representative data mining/machine learning algorithms and their applications in data analytics and data sciences

- Capability to identify, formulate and solve exploratory data analysis and predictive modeling problems that arise in practical applications
- Understanding of the strengths and weaknesses of alternative algorithms
- Capability to adapt or combine key elements of existing algorithms to design new algorithms as needed
- Hands-on experience with the applications of several representative algorithms in a high-level programming language (e.g., Java, Python)
- Hands-on experience in participating in online data mining/machine learning competitions (e.g., Kaggle)

Textbook

The following textbook is used as the main reference of the course. Note that we use the latest 4th edition, but if you happen to have an earlier edition of the book, that is OK.

- [Data Mining \(4th Edition\): Practical Machine Learning Tools and Techniques](#), by Ian Witten, Eibe Frank, Mark Hall, and Christopher Pal, 2016

In addition, the following textbooks are also useful to complement the main textbook above:

- [Machine Learning: The Art and Science of Algorithms that Make Sense of Data](#), Peter Flach, 2012
- [Data Mining: Concepts and Techniques](#), Jiawei Han, Micheline Kamber, and Jian Pei, 2011
- [Introduction to Data Mining](#), Pang-Ning Tan, Michael Steinbach, Vipin Kumar, 2016

Projects

Two hands-on projects are planned:

- Individual Project: Clickbait tweet detection
- Team Project: Online competition

- [Kaggle](#)
- [CrowdAI](#)
- [DrivenData](#)

Grading Weights

- Attendance and Class participation: 5%
- Homework assignments: 15%
- Project: 55%
 - Individual project: 20%
 - Team project: 35%
- Final Exam: 25%

Assignment Submission Policy

- Homework and Projects are usually assigned during WED class
- Dues are by default SUN 11:59pm (EST)
- Students can submit late with the penalty of 25% deduction for every 12 hours late (up to 2 days)
- After 2 days, no more late submission is allowed

Academic Integrity

According to the Penn State Principles and University Code of Conduct: Academic integrity is a basic guiding principle for all academic activity at Penn State University, allowing the pursuit of scholarly activity in an open, honest, and responsible manner. In according with the University's Code of Conduct, you must not engage in or tolerate academic dishonesty. This includes, but is not limited to cheating, plagiarism, fabrication of information or citations, facilitating acts of academic dishonesty by others, unauthorized possession of examinations, submitting work of another person, or work previously used without informing the instructor, or tampering with the academic work of other students. Any violation of academic integrity will be investigated, and where warranted, punitive action will be taken. For every incident when a penalty of any kind is assessed, a report must be filed.

Plagiarism (Cheating): Talking over your ideas and getting comments on your writing from friends are NOT examples of plagiarism. Taking

someone else's words (published or not) and calling them your own IS plagiarism. Plagiarism has dire consequences, including flunking the paper in question, flunking the course, and university disciplinary action, depending on the circumstances of the office. The simplest way to avoid plagiarism is to document the sources of your information carefully.

Disability Access Statement

Americans with Disabilities Act: The School of Information Sciences and Technology welcomes persons with disabilities to all of its classes, programs, and events. If you need accommodations, or have questions about access to buildings where IST activities are held, please contact us in advance of your participation or visit. If you need assistance during a class, program, or event, please contact the member of our staff or faculty in charge. Access to IST courses should be arranged by contacting the Office of Human Resources, 332 IST Building: (814) 865-8949.

Students with Disabilities: It is Penn State's policy to not discriminate against qualified students with documented disabilities in its educational programs. (You may refer to the Nondiscrimination Policy in the *Student Guide to University Policies and Rules*.) If you have a disability-related need for reasonable academic adjustments in this course, contact the Office for Disability Services (ODS) at 814-863-1807 (V/TTY). For further information regarding ODS, please visit the Office for Disability Services Web site at <http://equity.psu.edu/ods/>.

In order to receive consideration for course accommodations, you must contact ODS and provide documentation (see documentation guidelines at <http://equity.psu.edu/ods/guidelines/documentation-guidelines>). If the documentation supports the need for academic adjustments, ODS will provide a letter identifying appropriate academic adjustments. Please share this letter and discuss the adjustments with your instructor as early in the course as possible. You must contact ODS and request academic adjustment letters at the beginning of each semester.

Statement on Nondiscrimination & Harassment (Policy AD42)

The Pennsylvania State University is committed to the policy that all persons shall have equal access to programs, facilities, admission and employment without regard to personal characteristics not related to ability, performance, or qualifications as determined by University policy or by state or federal authorities. It is the policy of the University to maintain an academic and work environment free of discrimination, including harassment. The Pennsylvania State University prohibits discrimination and harassment against any person because of age, ancestry, color, disability or handicap, national origin, race, religious creed, sex, sexual orientation, gender identity or veteran status. Discrimination or harassment against faculty, staff or students will not be tolerated at The Pennsylvania State University. You may direct inquiries to the Office of Multicultural Affairs, 332 Information Sciences and Technology Building, University Park, PA 16802; Tel 814-865-0077 or to the Office of Affirmative Action, 328 Boucke Building, University Park, PA 16802-5901; Tel 814-865-4700/V, 814-863-1150/TTY.

For reference to the full policy (Policy AD42: Statement on Nondiscrimination and Harassment): <http://guru.psu.edu/policies/AD42.html>