DS 3001: Foundations of Data Science

Syllabus, Worcester Polytechnic Institute, D-term 2020

http://web.cs.wpi.edu/~kmlee/ds3001

1. **Credits**: 3
    a. Class Meets: Tuesday and Friday **2:00pm - 3:50pm, Zoom (online learning)**

2. **Instructor**: Kyumin Lee, (435) 797-8420, kmlee@wpi.edu
    Office Hours: **Tuesday 9:30-10:30am** and **Wednesday 4-5pm, Zoom**

    TA: Jianjun Luo, jluo@wpi.edu
    Office Hours**: Monday, Thursday and Friday 10-11am, Zoom**

3. **Textbooks:** Course readings will be drawn from the following textbooks:
    a. Data Mining: Concepts and Techniques. Jiawei Han and Micheline Kamber. Morgan Kaufmann.
    b. Introduction to Data Mining. Pang-Ning Tan, Michael Steinbach and Vipin Kumar. Addison-Wesley.
    c. Mining of Massive Datasets. Jure Leskovec, Anand Rajarman and Jeffrey D. Ullman. Cambridge University Press.
    d. Doing Data Science. Rachel Schutt, Cathy O'Neil. O'Reilly Media.
    e. Mining the Social Web. Mikhail Klassen, Matthew A. Russell. O'Reilly Media.
    f. Data Science from Scratch. Joel Grus. O'Reilly Media.
    g. Introduction to Information Retrieval, Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schutze, Cambridge University Press.

4. **Specific Course Information**:
    a. Course Description: This course provides an introduction to the core ideas in Data Science. It covers a broad range of methodologies for working with and making informed decisions based on real-world data. Students will learn how to manage and analyze data at scale (e.g., big data). Specifically, the students will study big data management and processing techniques, data analytics, statistical methods and models, data visualization, and etc. Students will conduct a team-based project.
    b. Recommended background for this course includes statistics knowledge equivalent to MA 2611 and MA 2612, linear algebra equivalent to MA 2071, and the ability to program equivalent to (CS 1004 or CS 1101 or CS 1102) and (CS 2102 or CS 2119). This course does not fulfill Mathematics, Basic Science or Engineering Science/Design credits.

5. **Specific goals for the course**
    a. Course Objectives
    By the end of the semester you will be able to:

i. Define and explain the key concepts and models relevant to data science.
ii. Design, implement, and evaluate the core algorithms underlying an end-to-end data science workflow, including the experimental design, data collection, mining, analysis, and presentation of information derived from large datasets.
iii. Apply "best practices" in data science, including facility with modern tools (e.g., Hadoop).

6. **Brief list of topics to be covered**
    a. Data Exploration
    b. Data Preprocessing
    c. Mining and Analytics
    d. Visualization
    e. MapReduce
    f. Recommendation

Communication:

All course announcements will be posted via the Canvas course mailing list.

Grading Policy:

The course grading policy is as follows:

40% Assignments
30% Exams
30% Project

The grading scale is A:100-90, B:89.9-80, C: 79.9-70, NR:69.9-0

Assignments:

There are four assignments. Each assignment is proportion to 10% of your grade. You will have total **4 late days** during the semester. You can use **up to 2 late days for each assignment** without penalty. After you consume the total 4 late days or two late days for an assignment (whichever comes first), then you will get penalty proportion to extra late days (e.g., 10% off for the next late day, 20% for the next two late days and so on).

For example, you submitted your first assignment 2 days late. You will not get any penalty, but use 2 out of 4 late days. Or if you submit your first assignment 3 days later than due date, you will use 2 late days (again up to 2 late days for an assignment), and get 10% off penalty because of the third late day.
For each assignment, we will **NOT accept** your solution **more than 3 days late**.

You may discuss an assignment with your colleague, but you should write a program and a report by yourself and should **NOT** copy and paste your colleague's solution. If you discussed an assignment with your colleague, explicitly report the colleague's name and what you discussed in your submission.

Exams:
There are two 'closed book' exams, each of which will be held in class. You may bring one standard 8.5" by 11" piece of paper with any notes you think appropriate or significant (only use front side).

Project:
In this term project, you will apply data science techniques that you learned from this course to your project. The detailed information regarding the term project will be announced in class. You will present and may demonstrate your project in **May 12**.

Add and Drop policy:
A student can add/drop through the 5 day of the term without a fee. On days 6-10 of the term, not including weekends, add/drops are permitted with instructor approval. A $100 late fee will be charged per class (for adds) on days 6-10. No add/drops are allowed after day 10.
Refer to https://www.wpi.edu/offices/registrar/course-registration/add-drop

Learning Aids:
Lecture notes and schedule will be available in the course web page. Your homework solution and project report is supposed to be submitted via Canvas.

Plagiarism and Cheating:
Unless explicitly noted, all work you submit must be your own work. You are encouraged to discuss with others about ideas and material in the course, in preparing for exams, in understanding homework problems, project statements, etc. However, all homework solutions, exams are to be written individually, and the solutions should be your own, unless otherwise specified. Projects encourage teamwork, that is, in that case you are expected to work closely with your partner/(s) to solve problems and prepare a common agreed-upon solution.

Note in particular that copying of any material, may it be a single sentence or a figure, from any location (including the internet) without proper acknowledgement of the source constitutes plagerism. If in doubt, please ask for clarification. Any violation of the WPI's guidelines for academic integrity will result in no credit for the course and referral to the Student Affairs Office for disciplinary action. You should be familiar with the WPI Policy on Academic Honesty.

Students with Disabilities:
If you need course adaptations or accommodations because of a disability, or if you have medical information to share with me, please make an appointment with me as soon as possible. If you have not already done so, students with disabilities who believe that they may need accommodations in this class are encouraged to contact the Office of Disability Services (ODS) as soon as possible to ensure that such accommodations are implemented in a timely fashion. This office in Daniels Hall has phone (508-831-4908) and can be reached by email at disabilityservices@wpi.edu.

Writing Center:

The Writing Center offers one-on-one consultations to help you improve as a writer. Writing Center tutors will read your written work, give you feedback about your document's strengths and weaknesses, and help you chart a path forward as you revise. For D Term, all consultations will take place via online connection and document sharing instead of a face to face, but you'll still get real-time feedback through conversation with a peer tutor. Consultations are free and open to all WPI students for all classes and projects, and tutors will happily work with you at any stage of the writing process (early brainstorming, revising a draft, polishing sentences in a final draft). To learn more about our online tutoring and how to schedule a one-hour appointment, go to the Writing Center homepage: wpi.edu/+writing