

Course Name and Number: DATA 612 - Special Topics: Recommender Systems

Credits: 3 cr.

Prerequisite(s): DATA 607, DATA 606, and DATA 602. Students with strong R or Python skills may have the DATA 602 pre-requisite waived. More advanced program course work will be helpful, but not assumed.

How is this course relevant for data analytics professionals?

Recommender systems and related ranking applications—are widely regarded as one of the most widely adopted commercial implementations of data science.

Course Description:

In this course, students will learn to build and evaluate different kinds of recommender systems, using both R or Python. Some of the course work will be done using Apache Spark.

Course Learning Outcomes:

By the end of the course, students should be able to:

- Build content management and collaborative filtering recommender systems in R or Python.
- Implement recommender systems using matrix factorization techniques (SVD and ALS).
- Implement hybrid recommendation systems.
- Understand business and deployment issues related to “explainability” and managing “attacks” on recommender systems.
- Implement recommender systems at scale under Apache Spark.



THE PROBLEM WITH
AVERAGING STAR RATINGS

Source: <https://xkcd.com/937/>

Assignments and Grading:

Projects (4)	40%
Research Discussion Assignments (4)	16%
Recommender Systems in Context Presentation (1)	10%
Final Project Proposal (1)	4%
Final Project (1)	20%
Final Project Presentation (1)	10%
TOTAL	100%

Quality of Performance	Letter Grade	Range %
Excellent - work is of exceptional quality	A	93 - 100
	A-	90 - 92.9
Good - work is above average	B+	87 - 89.9
Satisfactory	B	83 - 86.9
Below Average	B-	80 - 82.9
Poor	C+	77 - 79.9
	C	70 - 76.9
Failure	F	< 70

Notes

- All projects and assignments, unless otherwise noted, are due end of day New York time on Sundays.

Late work is not accepted. However, there are **five** projects and **five** research discussion assignments. Your final grade is based on your **four** highest-scoring projects and your **four** highest scoring research discussion assignments.

- **Course Completion Requirements.** To pass this course, you must minimally turn in at least four projects, and the final.
- You're required to make both a "Recommender Systems in Context" presentation and a final presentation. If you cannot deliver your presentation in the final Meetup on July 17th, you'll need to make available a video recording of your final presentation before that meetup. For students who opt to record their "Recommender Systems in Context" presentation, this presentation must be submitted before the start of our 6th meetup on July 10th.
- **Reproducibility Requirement, Testing Requirement, But Not Perfection!** Students are responsible for providing all code and data so that I can test your work. All assignments—projects and research discussion assignments—need to be delivered in either a Jupyter Notebook or an R Markdown format. This code should be posted into your GitHub repository, and you should include a link to the GitHub repository in your assignment submission.
- If you turn in code that does not run, you will not receive full credit, unless you also include an explanatory note at the time of submission. At the same time, you don't need to turn in perfect code. Generous partial credit will be given for deliverables that timely, tested, and reproducible. Cutting corners—as long as they are documented at the time of submission—is also acceptable.
- Students that complete all work in a satisfactory and timely manner will earn a maximum grade of A-. To earn a grade of A in this course, you'll need to regularly demonstrate work above and beyond what is expected.

Policy on Sharing and "Borrowed" Code. In this course, you may collaborate and you may take base code from whatever sources you wish. But you *must* document what you started with, and what you added, so you are graded on your own contributed work! Other courses in the program may have stricter policies—always check with your instructor in advance.

Course Learning Materials



Required Text:

- *Building a Recommendation System with R*, Suresh K Gorakala and Michele Usuelli, Packt Publishing, 2015.

Relevant Software, Hardware, or Other Tools:

We'll look at materials from a wide variety of sources. In particular, the videos from the University of Minnesota / Coursera course on recommender systems will be optional viewing throughout the course. These is an on-demand course with a “free” option, and provides a conceptual (non-programming) overview of recommender systems.

We will make use of the R programming environment and the RStudio IDE, the Python 3.x data science stack, and Jupyter Notebooks. You'll have the opportunity to work in Apache Spark in the cloud, using Databricks' Community Edition. Motivated students who want to instead run their code on their own virtual machines or on another hosting service (such as Amazon, Web Services, Microsoft Azure, or Google Cloud) may do so, but they will be responsible for their own setups, configurations, and possible service charges.

Faculty Contact Information

David Stern david.stern@sps.cuny.edu david.c.stern@gmail.com (please copy both for quickest response!)	
--	--

We won't have standing weekly office hours for this course, but students are encouraged to send me an e-mail to set up a mutually agreeable time to talk. I'm happy to talk with everyone who wants one-on-one each week.

How This Course Works:

- CUNY requires that all summer courses take place in an accelerated seven week format, so this course will cover the material of a 15 week course, at about twice the normal pace. Please allocate your time accordingly.
- Meetups take place every week on Tuesdays from 7:00 p.m. to 8:00 p.m. EDT. Please see course site for specific dates and connection information. You are strongly encouraged to attend; all meet-ups will be recorded.
- Office Hours can be scheduled by appointment. If you need extra help and are willing to invest the time and effort to be successful, we'll make the time to help you. But... you should not be asking for extra help on a project the day before it's due, since this indicates that you're not investing the time and effort to be successful.
- You are encouraged to ask questions on the "Ask Your Instructor" forum on the course discussion board where other students will be able to benefit from your inquiries. For the most part, you can expect me to respond to questions by email within one business day.

Unit	Topic	Core Readings	Deliverables
Week 1 Jun 4 - Jun 11	Overview of Recommender Systems	<i>Building a Recommendation System with R</i> , chapter 1	Meetup on 6/5 at 7:00 p.m. EDT Environment Setup
Week 2 Jun 12 - Jun 18	Content Management and Collaborative Filtering	<i>Building a Recommendation System with R</i> , chapter 2 and 3	Meetup on 6/12 at 7:00 p.m. EDT Project 1
Week 3 Jun 19 - Jun 25	Matrix Factorization Techniques		Meetup on 6/19 at 7:00 p.m. EDT Project 2
Week 4 Jun 26 - Jul 2	Evaluating Recommenders	<i>Building a Recommendation System with R</i> , chapters 4 and 5	Meetup on 6/26 at 7:00 p.m. EDT Project 3
Week 5 Jul 3 - Jul 9	Spark and Distributed Computing		Meetup on 7/3 at 7:00 p.m. EDT Project 4
Week 6 Jul 10 - Jul 16	New Techniques and New Domains		Meetup on 7/10 at 7:00 p.m. EDT Project 5
Week 7 Jul 17 - Jul 20	Final Projects	None	Meetup on 7/17 at 7:00 p.m. EDT Final Projects and Presentations

Accessibility and Accommodations

The CUNY School of Professional Studies is firmly committed to making higher education accessible to students with disabilities by removing architectural barriers and providing programs and support services necessary for them to benefit from the instruction and resources of the University. Early planning is essential for many of the resources and accommodations provided. Please see: http://sps.cuny.edu/student_services/disabilityservices.html

Online Etiquette and Anti-Harassment Policy

The University strictly prohibits the use of University online resources or facilities, including Blackboard, for the purpose of harassment of any individual or for the posting of any material that is scandalous, libelous, offensive or otherwise against the University's policies. Please see: http://media.sps.cuny.edu/filestore/8/4/9_d018dae29d76f89/849_3c7d075b32c268e.pdf

ACADEMIC INTEGRITY

Academic dishonesty is unacceptable and will not be tolerated. Cheating, forgery, plagiarism and collusion in dishonest acts undermine the educational mission of the City University of New York and the students' personal and intellectual growth. Please see: http://media.sps.cuny.edu/filestore/8/3/9_dea303d5822ab91/839_1753cee9c9d90e9.pdf

STUDENT SUPPORT SERVICES

If you need any additional help, please visit Student Support Services: http://sps.cuny.edu/student_resources/