

DS-GA 3001: Search and Discovery

Brian McFee

Fall, 2020

E-mail:	brian.mcfee@nyu.edu
Class hours:	F 14:00–15:40
Lab hours:	F 15:45–16:35
Room:	SILV 206
Office hours signup:	http://bit.ly/brian-office-hours or by appointment
Section leaders:	Guido (Sid) Petri gp1655@nyu.edu Angela (Amber) Teng at2507@nyu.edu

Pre-requisites

Pre-requisites: DS-GA 1003 (or equivalent by instructor approval) **OR** MPATE-GE 2623.

What is search and discovery?

This class will introduce you to modern tools for indexing, exploring, searching, and recommending items in large collections. Familiar examples might be movie recommendation, news feeds, web indexing, music streaming, or social network link prediction. Specifically, we will cover:

Fundamentals of information retrieval and recommender systems

Common machine learning methods for search and recommendation

Quantitative and qualitative evaluation

Critical analysis of theoretical and practical limitations of recommendation

Broader societal impacts of search and recommendation

Course Structure

Due to the ongoing COVID-19 situation, this course will run in *blended mode*. This means that there will be a mixture of students attending in-person and remotely over Zoom. Moreover, some remote students will not be able to attend synchronously, due to time zone differences, internet connectivity, *etc.* This is clearly not an ideal situation, but we're going to try to make the best of it.

To ensure fair treatment of all students, this course will use a *flipped classroom* setup. All lectures will be pre-recorded and uploaded in advance, usually in short segments. The class meetings will then be used to facilitate group discussion about the lecture material and assigned reading.

Your grade for the class will be determined by the following break-down:

25% reading summaries

25% lab assignments

50% project

Additionally, the following policies will be in place:

- Most weeks in the calendar (listed below) have assigned reading. For each reading, you will write a short summary, which will be due the day before the class meeting. This will help us structure and focus the discussions during the class meeting. Reading summaries are to be completed **individually**. The lowest two (2) scores will be dropped.
- There will be a small number of lab assignments in the first half of the semester. These also must be completed **individually**: the goal here is to ensure that you all have basic fluency with the tools necessary to implement your group project. You will have two (2) slip days which can be used to turn in assignments late with no penalty. Otherwise, a 20% penalty will be incurred for each day a lab assignment is late. Late assignments will not be accepted after 5 days.
- The project will consist of designing, implementing, and evaluating a recommender system. The specifics (choice of data/domain, methods, software frameworks, etc.) will be up to you, but your proposal must be carefully written and approved by the instructor. You may collaborate in a **group of up to 4 students**.

Exceptions to the above policies may be granted in case of emergencies, but any requests must be cleared by the instructor prior to the original due date.

Course Policies

Who should I contact?

Miscellaneous policy questions: (e.g., how do assignment due dates work?) please re-read this course information document and the course website. If you still have questions, please post them to the discussion forum on the course website.

Help with assignments or course topics: post on the discussion forum, or ask the instructor or section leaders during office hours.

Anything sensitive or confidential: (e.g., health issues, emergencies): email brian.mcfee@nyu.edu.

Anything else: I'm happy to talk to any student during office hours about various topics (course advising, questions about research, and so on).

Class environment

It is my job to make this course inclusive and equitable for all students. Please do your part by seeking to promote the success of others, and by treating each other in ways that respect the diversity of talent that is drawn to this field. Here are a few specific things that you should know about my policies on creating an inclusive and equitable class environment (both in the classroom and on the course website/forum):

- **Preparation:** Students come to this class from a wide range of backgrounds, and greatly varying previous exposure to mathematics, programming, and data science more generally. I want to assure students who may feel out of place here that you are indeed prepared to succeed in this class! If you feel that there are gaps in your knowledge, please speak with the course staff and we will help you find additional materials as needed.

- **Classroom environment:** For some reason, it is common in technical or programming-oriented classes that some students ask “questions” that are not really questions so much as opportunities to demonstrate knowledge of jargon, or facts that are beyond the scope of the topic at hand. This can have discouraging effects on other students who may not be familiar with those terms, and worry that this indicates that they are less prepared to do well in the class. (Note: this is rarely the case: knowing terms outside the scope of the course is not a good predictor of success.) If you find yourself wanting to make such a question or comment in lecture, I encourage you to consider whether office hours would be a better venue for exploring that topic. The course staff are more than happy to discuss tangentially related topics in office hours, when they would not distract from lecture or alienate other students.
- **Accessibility:** If you have any accessibility requirements, please present a letter from the [Moses Center](#) to me at your earliest convenience, so that I can ensure that materials and staff comply with your needs. I am always willing to do what it takes to support you, but I ask that you have your exam scheduling requests submitted no later than 1 week prior to the exam so that we have sufficient time to make any necessary arrangements.
- **Names and pronouns:** If your name and/or pronoun doesn’t match the class roster delivered from the registrar, please let me know and I will ensure that you are addressed correctly in our class. You are always welcome to use your preferred form of address on all class assignments and exams; just be sure to include your NYU netID number so that we can properly link records.
- **Class expenses:** If obtaining any material for use in our class presents a financial hardship for you, please let me know and I will do my best to arrange for loaner materials.
- **Feedback** If you have pressing or specific issues, please do not hesitate to let me know if any aspect of our course or class community can be improved.

Academic Integrity and Honesty

All students are expected to do their own work for labs and reading summaries, and should cite external sources if used. Questions regarding acceptable collaboration should be directed to the class instructor **prior** to the collaboration.

Please refer to the general [NYU academic integrity statement](#).

Calendar

This calendar is tentative, and content may be rearranged or substituted to suit the needs of the class. All assigned readings will be made available through the class website.

Week 01, 09/04: Introduction and course overview.

- Reading: Chris Anderson. *The long tail: Why the future of business is selling less of more*. Hachette Books, 2006, chapter 3.

Week 02, 09/11: Information retrieval and recommendation

- Reading: Francesco Ricci, Lior Rokach, and Bracha Shapira. Recommender systems: Introduction and challenges. In *Recommender Systems Handbook*, pages 1–34. 2015. doi: 10.1007/978-1-4899-7637-6_1. URL https://doi.org/10.1007/978-1-4899-7637-6_1.

Week 03, 09/18: Latent factor models for recommendation

- Reading: Yehuda Koren and Robert M. Bell. Advances in collaborative filtering. In *Recommender Systems Handbook*, pages 77–118. 2015. doi: 10.1007/978-1-4899-7637-6_3. URL https://doi.org/10.1007/978-1-4899-7637-6_3.

Week 04, 09/25: Evaluation metrics for ranking and retrieval

- Reading: Asela Gunawardana and Guy Shani. Evaluating recommender systems. In *Recommender Systems Handbook*, pages 265–308. 2015. doi: 10.1007/978-1-4899-7637-6_8. URL https://doi.org/10.1007/978-1-4899-7637-6_8.

Week 05, 10/02: Novelty and diversity

- Reading: Pablo Castells, Neil J. Hurley, and Saul Vargas. Novelty and diversity in recommender systems. In *Recommender Systems Handbook*, pages 881–918. 2015. doi: 10.1007/978-1-4899-7637-6_26. URL https://doi.org/10.1007/978-1-4899-7637-6_26.
- We'll also discuss the project requirements.

Week 06, 10/09: Context-aware recommendation

- Reading: Gediminas Adomavicius and Alexander Tuzhilin. Context-aware recommender systems. In *Recommender Systems Handbook*, pages 191–226. 2015. doi: 10.1007/978-1-4899-7637-6_6. URL https://doi.org/10.1007/978-1-4899-7637-6_6

Week 07, 10/16: Neural recommendation

- Reading: Maurizio Ferrari Dacrema, Paolo Cremonesi, and Dietmar Jannach. Are we really making much progress? a worrying analysis of recent neural recommendation approaches. In *Proceedings of the 13th ACM Conference on Recommender Systems*, pages 101–109, 2019

DUE:
Project pro-
posal

Week 08, 10/23: Multi-stakeholder recommendation

- Reading: Robin D Burke, Himan Abdollahpouri, Bamshad Mobasher, and Trinadh Gupta. Towards multi-stakeholder utility evaluation of recommender systems. In *UMAP (Extended Proceedings)*, 2016. URL http://ceur-ws.org/Vol-1618/SOAP_paper2.pdf.
- Reading: Himan Abdollahpouri and Robin Burke. Multi-stakeholder recommendation and its connection to multi-sided fairness. In *Workshop on Recommendation in Multistakeholder Environments (RMSE)*, 2019. URL <http://ceur-ws.org/Vol-2440/paper3.pdf>.

Week 09, 10/30: Counterfactual evaluation

- Reading: Tobias Schnabel, Adith Swaminathan, Ashudeep Singh, Navin Chandak, and Thorsten Joachims. Recommendations as treatments: Debiasing learning and evaluation. In *Proceedings of the 33rd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016*, pages 1670–1679, 2016. URL <http://proceedings.mlr.press/v48/schnabel16.html>.

Week 10, 11/06: Project updates

- No reading this week.
- Groups will present project status updates.

Week 11, 11/13: Feedback loops and filter bubbles

- Reading: Ray Jiang, Silvia Chiappa, Tor Lattimore, András György, and Pushmeet Kohli. Degenerate feedback loops in recommender systems. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, AIES '19*, page 383–390, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450363242. doi: 10.1145/3306618.3314288. URL <https://doi.org/10.1145/3306618.3314288>.

Week 12, 11/20: Privacy

- Reading: Arik Friedman, Bart P. Knijnenburg, Kris Vanhecke, Luc Martens, and Shlomo Berkovsky. Privacy aspects of recommender systems. In *Recommender Systems Handbook*, pages 649–688. 2015. doi: 10.1007/978-1-4899-7637-6_19. URL https://doi.org/10.1007/978-1-4899-7637-6_19.

Week 13, 11/27: *Thanksgiving recess: no meeting***Week 14, 12/04:** User-centric evaluation

- Reading: Bart P. Knijnenburg and Martijn C. Willemsen. Evaluating recommender systems with user experiments. In *Recommender Systems Handbook*, pages 309–352. 2015. doi: 10.1007/978-1-4899-7637-6_9. URL https://doi.org/10.1007/978-1-4899-7637-6_9.

Week 15, 12/11: Final project presentations