Who is on your team? Please give their names. \*

Annette Han, Yeseul Kim, Beilei He

Please give a comma separated list of the U-Mich emails of your teammates. For example: winstofb@umich.edu, kirtana@umich.edu, elleobri@umich.edu \*

[hananh@umich.edu](mailto:hananh@umich.edu), [yeseulk@umich.edu](mailto:yeseulk@umich.edu), [beileihe@umich.edu](mailto:beileihe@umich.edu)

What is the subject of your proposal and what questions do you hope to answer or explore? Please introduce your topic in a brief paragraph. \*

The purpose of this proposal is to recommend electives and University of Michigan LSA College Requirements courses for undergraduate students belonging to (a number of majors, TBD) majors within LSA College (we think we should narrow the number of groups of students to limit the scope of the project). The large number and ever-changing course offerings can cause stress to many students to navigate degree requirements, not only within their major but also the board offerings of electives and other college-specific requirements. Mistakes in understanding course requirements of course sequencing can cost students additional time needed to complete the degree, hence can cause unexpected and unnecessary monetary cost, and delay degree completion. The overwhelming selection of courses can also impede students from exploring other suitable courses that they would otherwise enjoy or excel in. In this project, we hope to create a pilot recommender system that reminds students of the courses they should take to meet the college-specific requirements and connect students with courses that they would not have otherwise selected. The overarching goal is to ultimately support students’ on-time graduation while helping them explore other interests and excel to the best of their abilities.

If applicable, please describe the dataset(s) you plan to use. You'll want to share the source, access method, and what features of the data you plan to explore. \*

(We made a diagram but unfortunately it cannot be displayed in the google form)

The datasets we plan to use are Fall 2021 course schedule (from the registrar office), list of degree requirements (from degree audit office, or scrape website), list of LSA College Requirements and courses associated with each requirement (from LSA advising office or registrar office, or scrape website), and history of courses students have taken in the past (from Enrollment Management, Learning Analytics Data Architecture, IRB approval needed). We plan to explore features such as course title, course code, course description, the LSA requirement(s) the course meets, the degree requirement(s) the course meets, the courses other students of the same major have taken in the past, and their grades.

What analyses do you plan to implement as part of your project? For each analysis, please briefly describe how it will support the main goals of your project. \*

Content-based filtering: recommend based on requirements and courses that match the requirements, other feature of courses could be used as well

Collaborative filtering: recommend bases on courses other students have taken in the past, infer a user’s interest and preferences from those of other similar users

Topic modeling: recommend bases on students’ interests and topics of the course (taken from the course description)

What technical challenges do you anticipate in managing your data and analyses, if any? For example, needing to use large datasets (>1 GB), difficult data cleaning tasks, a complicated machine learning pipeline, need for specialized tools/libraries/hardware. \*

I think if we end up not able to obtain the course requirements and list of degree requirements, scraping the website effort will add to our overall time needed to pursue the project. Also, to recommend a course based on students’ interests, we want to know what their interests are. It may be difficult to infer this information from students’ existing data with high accuracy. In this case, we either would use missing value imputation techniques to fill missing information, or we can create synthetic missing values for demonstration purposes.

What ethical challenges or concerns do you expect to encounter, if any, in this project? If there are potential concerns, how do you plan to mitigate them? \*

The data set which includes history of courses students have taken in the past might contain students’ personal information such as names, dates of birth, student id, and age. The personal information will not be used in the analysis, or only used under informed consent. We are also concerned about echo chambers which can happen when popular courses that are taken by other students (such as introductory courses) are reinforced by the system, creating filter bubbles that impede students from exploring other courses that might be just as interesting and enjoyable. A way to solve this problem is to calculate the popularity score of courses (based on the number of students who have taken the course in the past, or the ratio of available seats/number in the waiting list), filter this list by the requirement and interest of student, and present this list in a random order. This will make sure that we present information to students without enforcing or endorsing a particular course (if the student tends to go with the first course on the list).

What are the anticipated outcomes of your analysis & the project? How do you plan to evaluate the outcomes / claim success for the project? \*

Memory based approaches, matrix based approaches are what we practised in the course, some more sophisticated methods like network based approaches, machine learning approaches could also be used to build the course recommendation system. When we use collective filtering method, how to define user similarity. There are many different user representations to choose from.

RMSE will be used to evaluate the recommender system.

Ideally, if the recommender system is deployed for piloting, we want to conduct an A/B test to evaluate the effectiveness of the system.

Briefly explain the planned contributions of each team member. \*

Annette Han: acquiring data (connect with offices and get IRB approval), topic modeling, writing report

Yeseul Kim, data preparation, building the recommender system with data filtering methods, writing report.

Beilei He: data cleaning, feature selections, visualization programming applying methods mentioned above from course Recommendation system. Writing report

Does your team plan to use AWS? If so, please describe your team's past experience with AWS or similar cloud computing platforms.

This project will be largely self-managed, and the instructional team isn't enforcing any deadlines until the due date. That means it's up to you to create internal deadlines for staying on track. Please create a tentative schedule for your team with milestones you'd like to reach by certain dates during the course period. A completely made-up example: "By November 4- have dataset scraped and formatted for analysis. By November 21- Have proof-of-concept pipeline working. By December 12- Have analyses complete and begin writing report." \*

Oct 31 - Have dataset ready for analysis

Nov 7 - Clean and prepare the data for filtering, build a pipeline.

Nov 14 - Start analyses

Nov 21 -

Nov 28 - Complete analyses

Dec 5 - Finish first draft, work on the poster

Dec 12 -

Dec 17 - Submit the report and poster

Is there anything the teaching team should know to support you in this project? Feel free to ask for resources or extra advising on a certain topic- it never hurts to ask!