TCS: Tailored Course Suggester

Aditya Tanwar
200057

September 6, 2022

Abstract

We present a simple and lightweight GNN-based model that aims to improve the quality of life of IIT-Kanpur students by assisting them in selecting the best course bucket for their upcoming semester, taking into account personal preferences, sociological factors, compulsory courses, and individual strengths.

Introduction

There is often widespread confusion among the students during the Pre-registration and Add-drop periods prior to a semester. During the few days they have to request elective courses, the bulk of their time is spent settling schedule conflicts, leaving them with little opportunity to choose relevant or preferred courses. This results in students compromising their preferences in order to meet the per-semester credit minimum dictated by Institute standards[†], which in turn forces teachers to instruct less motivated students.
Improving the experience of students in courses is a further persistent issue. Our system aims to provide students with improved course suggestions prior to enrollment, effectively assisting them in selecting "better" courses for them.
This paper describes an application that utilises Deep Learning techniques (particularly Natural Language Processing and Graph Neural Networks) to provide the students of IIT-K with courses that they are most likely to thoroughly enjoy. It also includes *a modest proposal* to the Students' Senate, IIT Kanpur to fund the initiative for the benefit of the college's student body.

[†]a student should do at least 35 credits worth of coursework each semester as per the *UG Manual*

Motivation

The solution works with two empirical observations at its core. These are:

- Students choose classes that match with their interests and do better in those subjects
  The majority of students pursue a personal interest throughout the semester, while others attempt to explore different fields within their grasp. If they were to study this "passion" in a class, they would be naturally motivated to improve their performance in the quest of satisfaction.
  It is assumed that students do not deviate excessively from their previous interests/ hobbies while exploring new domains. The application aims to discover and understand this past history of each student via a series of interactions after their admission to college, such as participation in clubs/ club activities, projects done (both under a student body and a professor), courses requested, performance in various courses, etc. How this data shall be acquired and processed by the application is explained later.

- The presence of friends in a classroom enriches the experience and helps learning
  Throughout a student's time in college, they are encouraged to explore as much as possible and to make as many "connections" as possible. Almost all students adhere to this recommendation in some capacity. Oftentimes, the friends people create are a mirror of their own interests and personalities, and these companions influence their decisions and preferences. These friends may have been created through groupings in classes, wings, clubs, sports teams, events, etc.
  The interests of one's friends are thus deemed

pivotal to a student's own interests. Not only can friends influence one's interests, but they can improve experience in courses as well. For these two reasons, a student's social network assumes great significance in determining their interests and probability of enjoying a course.

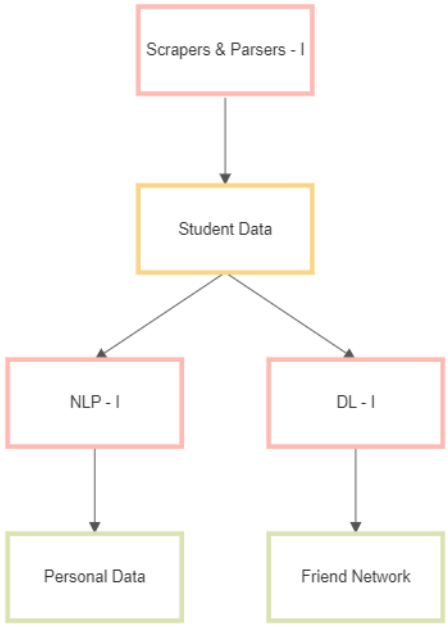
Novelty

The students already have access to some of the data like previous course offerings, grade statistics, courses taken by others, etc. But, to be able to make sense of this quantity of data is not humanly possible. It is also not possible to individually cater to each student by counseling, given the intake of the students each year, which is only going to increase with time.
It is thus, necessary that students have access to an application which can provide tailored course recommendations, taking their interests, and friends' interests into account, which is precisely the goal of our solution.

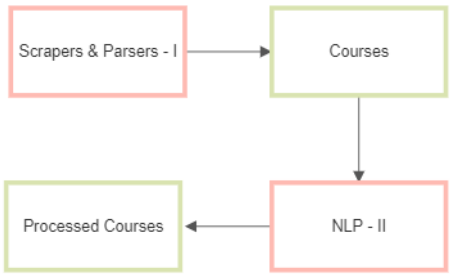
Architecture & Implementation

The tentative architecture and implementation plans have been provided below, subject to minor adjustments as the development of the project progresses.
The whole architecture has been broken down to smaller, modularized architectures for both ease in understanding and implementational purposes:

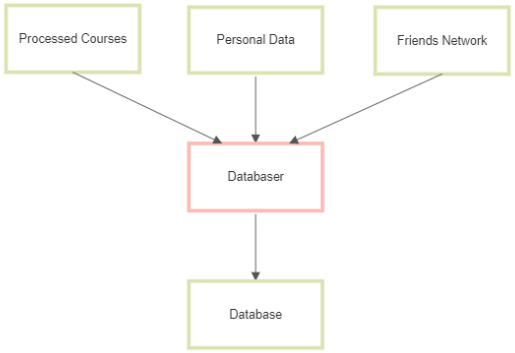
- Stage I: In this stage the student data is parsed by a subroutine implemented by software developers and then processed by NLP and DL models made by the ML researchers. It takes data from different resources such as groups in courses, teams in events, club participation, wingmates, etc.



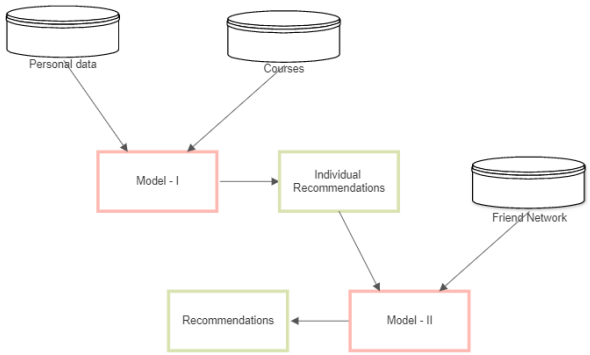
- Stage II: This stage shall run in parallel with Stage I and parse course contents and process them to create labels according to their course contents. It also tracks common course policies, grading records, semesters of offering, credits, pre-requisites, etc.



- Stage III: The databaser takes outputs from the previous two stages and compiles them into a database based on personal data (of students), friend networks, and courses.



- Stage IV: This stage involves two deep learning models; the first of them uses the personal data of students and the course database to make individual course recommendations. The second model operates on the output of the first model and uses the friend network to make recommendations, taking into account course policies involving group formation, team work, intersecting interests, etc.



Resources

Given the nature of the application, the two main components required are the student data and the computational power to process it completely. Some of these are expected to be received from the concerned bodies at IIT-K, and some of them are requested from the Student's Senate, IIT-K along with monetary funding. Both of them have been elaborated and indicated below:

- *Data:* Naturally, the application requires data of students in different forms such as:
  - Courses: The courses completed by students and the grades they were awarded in them, as well as the courses they had requested. It is expected that DoAA, IIT-K will agree to feeding this data to the application.
  - Projects: These include both projects mentored by professors and students alike. The former can be extrapolated from the courses they have done, and the latter is *requested* from the Students' Gymkhana, IIT-K.
  - Clubs and Teams: The participation of a student will help figure out possible friend(s). This is *requested* from the Students' Gymkhana, IIT-K.
  - Wingmates: Helpful in establishing friend connections since most students have friends in their wings. It is expected that Counseling Service, IIT-K will agree to feed this data to the application.
  - Groups: Some courses often have group-based assignments. The choice of teammates of a student in a course will thus be relevant in making the friends graph. It is expected that DoAA, IIT-K will agree to feed this data to the application.
- *Technical Requirements:* The computation required to make sense of the data above will be immense. To help make it feasible, some devices will be required:
  - Servers: The application will be hosted as a web application and thus required the servers to keep it up as well as support the huge database.
  - GPU: Helpful in facilitating the Deep Learning and NLP algorithms used in the application.

The other resource required would be manpower (students) to make the application, the roles are:

- Product Manager: Will look over the whole project and coordinate the different departments. *(1)*
- Software Developers: Responsible for the full-stack development of the application. *(3)*
- ML researchers: Responsible for coming up with algorithms in NLP and GNN for the ML side of the application. *(6)*
- Volunteers: Lend personal data for the training phase of the model. *(50)*

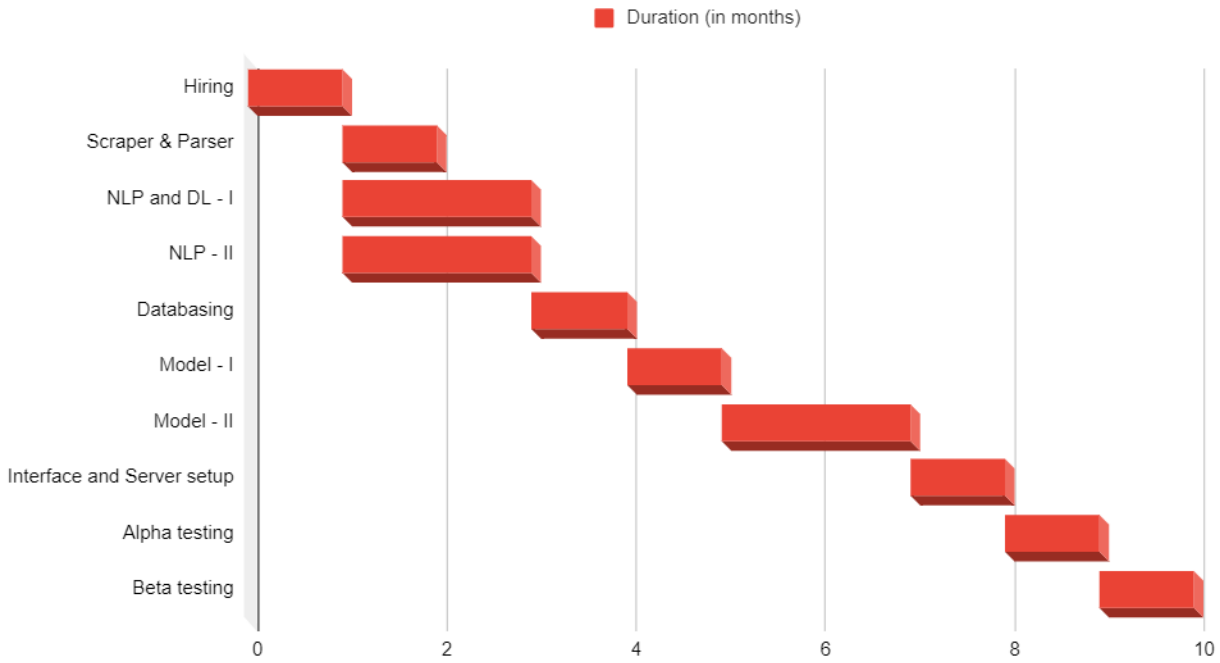
Budget

The costs of the resources have been compiled to calculate the total budget requirements of the model:

Asset	Quantity	Price	Total Cost
Servers	1	5,00,000 INR	5,00,000 INR
GPU	4	3,00,000 INR	12,00,000 INR
Product Manager	1	15,000 INR pm.	1,35,000 INR
Software Developer	3	10,000 INR pm. *(5 months)*	1,50,000 INR
ML researcher	6	10,000 INR pm. *(5 months)*	3,00,000 INR
Volunteers	50	300 INR	15,000 INR
Buffer	-	2,00,000 INR	2,00,000 INR
Total			**25,00,000 INR**

Deliverables & Timeline

The timeline of the project is as follows:



The deliverables shall be:

1. Data Scraper and Parser: Responsible for scraping and parsing all the course data (grades, contents, professors, semesters offered, etc.) and personal data of students (interests/ hobbies, friends, grades, etc.)

2. NLP and DL Model - I: The NLP model shall be responsible for finding the relevant data on students. The DL model shall be responsible for calculating the probability of two students being friends.

3. NLP Model - II: This NLP model reads through the course contents and buckets them into domains of study.

4. Database: The software team shall take the output of the above models and use them to create a master database (courses and students combined).

5. Model - I: Deep learning model that operates on the database as input and outputs course recommendations for a student on an individual level. Trained on volunteer data.

6. Model - II: Graph neural network that operates on the individual student recommendations, incorporates the friend network of a student along with course policies to enhance the recommendations. Trained on volunteer data.

7. Application: All the deep learning models accessible via an interface, ready for usage by any student registered in IIT-K. Eligible for operation on complete data. Expected to be delivered by the end of **11 months** (taking a buffer of 1 month).

Concerns

There are a lot of concerns due to the amount of sensitive data used by the application, ethical and logistical alike. This section tries to explain how the model will take care of them and respect privacy of everyone involved.

- Why should the creators of the application be trusted with such sensitive data?
  The users shall never be able to access the data of the application, except the data that the volunteers provide. The volunteers are given complete choice on the extent of data they are willing to provide for the training of the model. Moreover, neither the developers nor the application shall ever use the actual name of any volunteer, rather, they will be labelled with numbers to maintain anonymity.
  As per the data of the general body of students, the application will be handed over to the Senate after deployment, so the developers shall never get direct access to that data. In order to prevent abuse of data by any admin, the application will be converted into a completely blackboxed interface (which can only be used for feeding data and querying recommendations) so that it is impossible to retrieve its database directly.

- Why would the bodies like DoAA, Counseling Service, Students' Gymkhana IIT-K agree to let the application access student data?
  The privacy concerns have been answered above. As for productivity, the application aims to improve the quality of life of students, which is inline with the principles of the institution. Improving the QoL of students has direct consequences on the quality of education absorbed by the students, the grades they get, and the QoL of the instructors as well, since they will get more motivated students in their courses.

Future Development

The application certainly has a scope for introducing more useful and proper data, but at the cost of added complexities. Some of these advancements include giving the student probabilities of getting a course request accepted, probability of finding a friend in a particular course, and advising a whole template instead of just the immediately next semester.

Ofcourse, these plans are speculatory, and will be refined during and after the course of development.