Gradvisor

Team Members

Mohit Sarin - 934008352
Pankaj Kumar Tiwari - 734004077
Shaunak Hemant Joshi - 934008405
Shwetima Sakshi - 934008489

1. **What exactly is the function of your tool? That is, what will it do?**

   Gradvisor serves as an education platform tailored for students aiming to pursue studies abroad. It provides comprehensive tools for university search, aiding students in selecting institutions that align with their academic goals and preferences. Gradvisor streamlines the complex process of international education by offering resources and information tailored to each student's profile and aspirations.

2. **Why would we need such a tool and who would you expect to use it and benefit from it?**

   Our product, Gradvisor, is designed to address the different demands of students desiring to pursue higher education abroad, capitalizing on education's expanding globalization and growing demand for overseas study possibilities. At its core, Gradvisor is a sophisticated university research tool that allows students to thoroughly investigate a wide range of educational institutions throughout the world. Gradvisor offers invaluable insights into various universities by meticulously analyzing student profiles, which include criteria such as educational background and domain-specific interests. Using powerful algorithms and data analytics, our platform recommends six colleges divided into three categories: safe, moderate, and ambitious, which fit with each student's specific objectives and interests. Furthermore, Gradvisor allows students to filter university suggestions based on their favorite category, resulting in an individualized approach to university choices. In future we want Gradvisor to offer enhanced functionality allowing students to refine their university search based on specific geographical preferences, including narrowing down options by state.

3. **Does this kind of tool already exist? If similar tools exist, how is your tool different from them? Would people care about the difference? How hard is it to build such a tool? What is the challenge?**

   Our proposed tool aims to enhance existing platforms like Yocket and YMGrad by employing advanced machine-learning techniques such as content-based and collaborative filtering. This approach ensures superior quality recommendations tailored to users' specific needs and preferences. The key differentiator lies in providing more accurate and personalized suggestions, enhancing users' decision-making regarding educational and career pursuits. Building such a tool necessitates extensive research and development efforts, including exploring various ML algorithms, data gathering, preprocessing, and rigorous testing. Challenges include implementing complex ML algorithms, integrating them seamlessly into the user interface, and ensuring scalability and efficiency. Despite these challenges, the potential benefits, including improved recommendation quality and user satisfaction, make the development of such a tool a compelling endeavour.

To enhance a recommender system beyond Yocket, we consider combining a hybrid approach combining collaborative and content-based filtering. Additionally, integrating techniques like matrix factorization, deep learning models such as neural collaborative filtering (NCF), or advanced algorithms like factorization machines (FM) could improve recommendation accuracy and personalization.

4. **How do you plan to build it? You should mention the data you will use and the core algorithm that you will implement.**

Initially, we will perform data cleaning during the preprocessing phase. This cleaning process will encompass both the user input data and the training data. This stage will encompass tasks such as removing redundant columns, handling missing values, and ensuring uniformity in data formats. Subsequently, we will delve into determining the similarity between users and the content, for which we will implement content-boosted collaborative filtering and user-user filtering algorithms. These algorithms are adept at alleviating the impact of data sparsity by assigning predictive scores to all items for each user within our dataset.

Moving forward, the important features identified during the preprocessing phase, along with the scores derived from filtering, will be integrated into the deep learning model. This model will undergo training to develop recommendations for universities. Our current plan involves recommending a selection of six universities, categorized into groups of safe, moderate, and ambitious options. For evaluating our model, we will opt for k-fold cross-validation over a randomized train-test split. This decision mitigates the potential issue of irregular data distribution within the test dataset. Our dataset, sourced from Kaggle, will undergo certain modifications to better suit our project's requirements.

Moreover, we aim to develop a user-friendly web application to facilitate easy access to our university recommendation system. This application will feature a form allowing users to input their academic details. Subsequently, the trained model will leverage this information to generate personalized university recommendations tailored to each user's profile and preferences. Through this holistic approach encompassing data processing, model development, and application deployment, we endeavour to create an efficient and effective platform for assisting aspiring students in their pursuit of higher education opportunities.

5. What existing resources can you use?

We will be utilizing a variety of existing resources including research papers from, Kaggle datasets, as well as materials from class lectures and readings and hands-on experience from basic implementation of recommender models in assignments. These resources will provide a comprehensive understanding and facilitate the development of effective solutions.

Related research papers:

- Content-Based Filtering:
  - "Item-based collaborative filtering recommendation algorithms" by Sarwar, Badrul M., et al. (Link: https://dl.acm.org/doi/pdf/10.1145/371920.372071)
  - "Learning to Rank Recommendations with the k-Order Statistic Loss" by Wang-Cheng Kang, et al. (Link: https://arxiv.org/abs/2005.10941)
- Collaborative Filtering:
  - "Collaborative filtering for implicit feedback datasets" by Hu, Yifan, et al. (Link: https://dl.acm.org/doi/10.1145/1864708.1864721)
  - "Factorization Meets the Item Embedding: Regularizing Matrix Factorization with Item Co-occurrence" by He, Xiangnan, et al. (Link: https://dl.acm.org/doi/10.1145/3292500.3330664)
  - "Neural Collaborative Filtering" by Xiangnan He, et al. (Link: https://arxiv.org/abs/1708.05031)
- Hybrid Recommendation System:
  - "Hybrid Recommender System: Integrating Content-Based Filtering with Collaborative Filtering for Enhanced Recommendations" by Yassine Afoudi, Mohamed Lazaar, Mohammed Al Achhab (Link: https://www.sciencedirect.com/science/article/abs/pii/S1569190X21000836)

Class readings:

- Netflix Prize:

  https://www.cs.uic.edu/~liub/KDD-cup-2007/proceedings/The-Netflix-Prize-Bennett.pdf
- Matrix Factorization Techniques for Recommender Systems: https://datajobs.com/data-science-repo/Recommender-Systems-%5BNetflix%5D.pdf

6. How will you demonstrate the usefulness of your tool?

To gauge our algorithm's effectiveness, we will employ RMSE as an evaluation metric, highlighting its accuracy in generating recommendations. Additionally, we will develop a user-friendly web application to showcase the algorithm's functionality. Users can input preferences and receive personalized suggestions, demonstrating the tool's practical utility in guiding educational and career decisions. This combined approach emphasizes both technical proficiency, as assessed by RMSE, and practical value, as demonstrated through user interaction with the web application.

7. **A rough timeline to show when you expect to finish what. List a couple of milestones.**

Task	Timeline
Data Augmentation and Cleaning	03/05/2024 - 03/12/2024
Data Preprocessing	03/13/2024 - 03/23/2024
Exploring different algorithms	03/24/2024 - 03/31/2024
Finalizing the algorithm	04/01/2024 - 04/09/2024
Testing / Validation	04/10/2024 - 04/13/2024
Model Evaluation	04/14/2024 - 04/19/2024
UI Development & Backend integration	04/20/2024 - 04/27/2024