StudyMatch- a data-driven tool for Chinese students applying to overseas universities

Each year in China, millions of students opt to pursue their university and graduate studies overseas. Most of them have no experience living or traveling abroad, knowing little about evaluating and selecting schools. To ease the concerns, they often engage international education counselors, paying a premium for advice and strategies that can shape their future trajectories. As a former member of this group, I distinctly remember the sense of helplessness I felt when deciding which universities to apply to. In the aftermath, like many of my schoolmates, I began to doubt the effectiveness of international education counselors and their plans, which often seem to rely heavily on their personal intuition and preferences. Based on my understanding of this and current statistical knowledge, I plan to develop a data-driven tool for students. It will provide reliable and cost-effective guidance by predicting their chances of admission to specific institutions. This product seeks to offer precise information and specific strategies, moving away from vague response from educational counselors like 'you can try it' or 'you might succeed'.

Firstly, I decide to scrape data from a website: https://www.liuxue315.cn/. It collected historical data on the educational background of Chinese students and the overseas institutions they applied to.

Here is an example of the data table in this page.



Second, I will gather data from other public sources, including U.S. News and QS Universities Rankings, to establish a comprehensive database that consolidates all the information.

Next, using that dataset, I intend to develop a logistic regression model aiming to provide students with clarity and precision about their chances of admission to specific institutions, while offering practical strategies to balance their budget and associated risks.

Finally, I will provide the information regarding the best-fitting colleges to enrich students' insights into the institutions, their offered majors, and campus life.

Addendum:

1. Currently, the data sources encompass three websites, but I may incorporate additional information as necessary.

https://www.liuxue315.cn/

https://www.usnews.com/best-colleges/

http://www.cuaa.net/

2. The information in these websites is not personal identifiable, which reduces potential ethic concerns.

3. While unstructured data can be crucial in the college application process, like a student's statement of purpose and CV, analyzing it often requires advanced machine learning techniques, such as natural language processing. However, this product will mainly concentrate on leveraging statistical expertise, so we won't be incorporating those unstructured data elements.

4. While this model is initially designed for Chinese prospective international students, it can be adapted to assist students from other countries, provided we have access to study abroad data from those nations to refine the model.