A Yelp for Universities

Text Mining Online Student Reviews of Postsecondary Institutions Using Large Language Models (LLMs)

Kathy Yin
kathy.yin@stanford.edu
M.S. in Education Data Science
Stanford Graduate School of Education

Introduction

- College searching and enrollment decisions are critical for students and families. Increasingly, they are turning to **online school reviews as a source of information**, in addition to schools and social networks.
- Yet, the usage of online school ratings **diverges** significantly among SES backgrounds, influencing how parents gather and **leverage information for educational decisions**.[1]
- College information not only facilitates better decision-making but also enhances student satisfaction and success in college.[2]
- Topic modeling and sentiment analysis are popular ways to extract opinions from texts. LLMs and other natural language processing techniques can be employed in this context.[3]

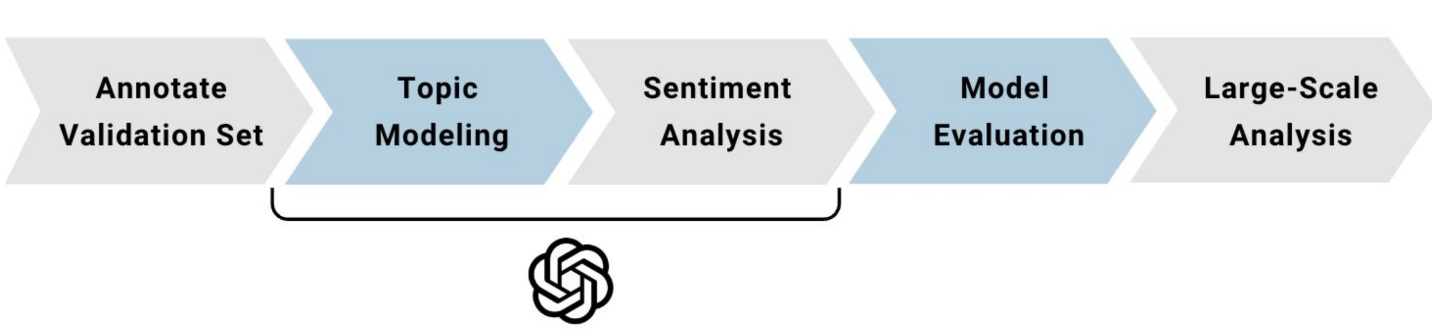
Research Questions

- How accurately can an AI model (like GPT) identify the main topics and opinions in university reviews?
- Is there a relationship between the mention of specific topics in reviews and the schools' overall ratings?

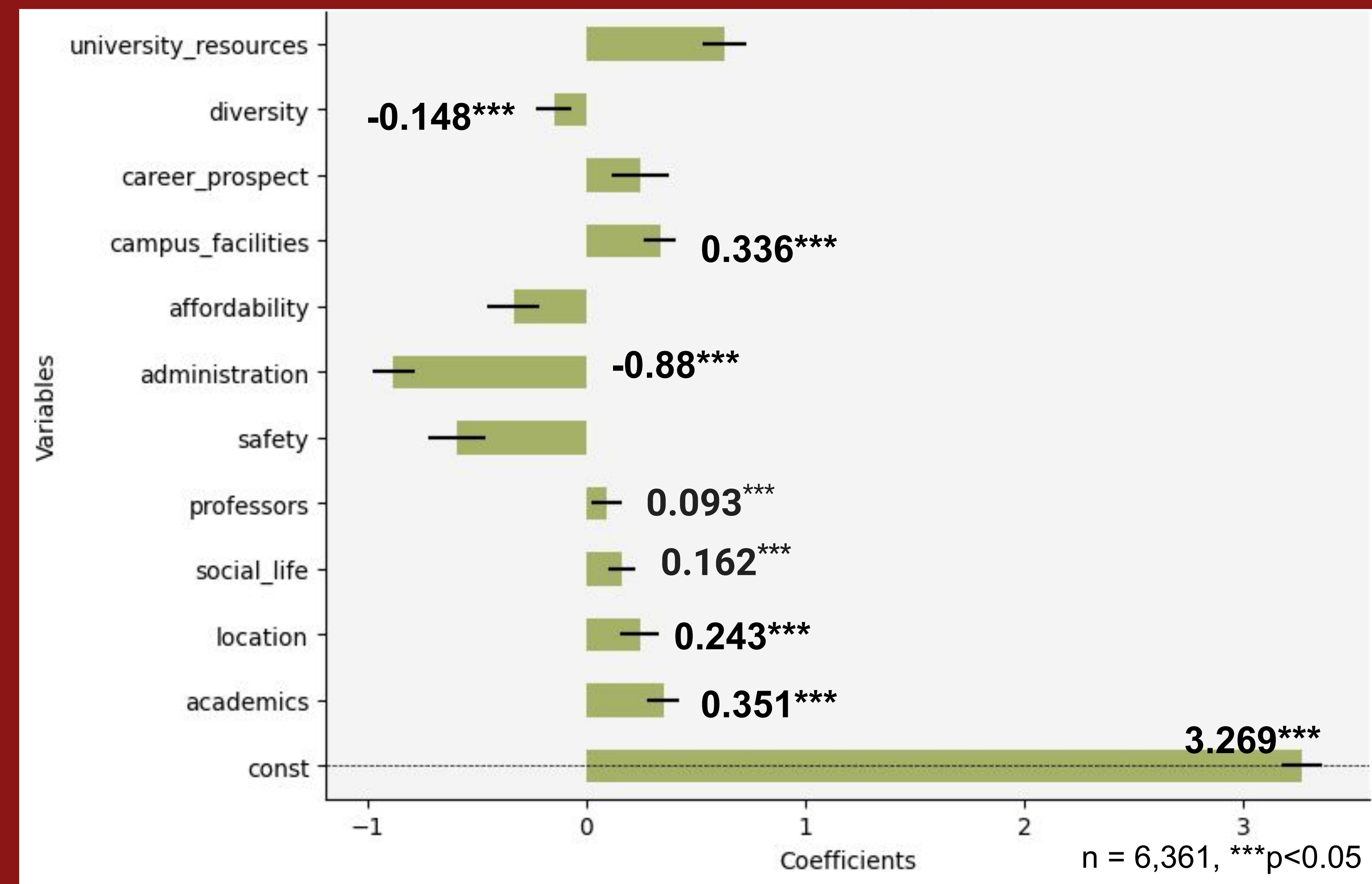
Data

- **94** four-year non-for-profit postsecondary **institutions**
- **6,361** user-generated reviews from a popular college review site
- Each institution has a mean of 64 reviews with a std. of 14.
- A total of 285 reviews are manually labeled for evaluation
- Distribution of reviews by rating category at a scale of 1-5:

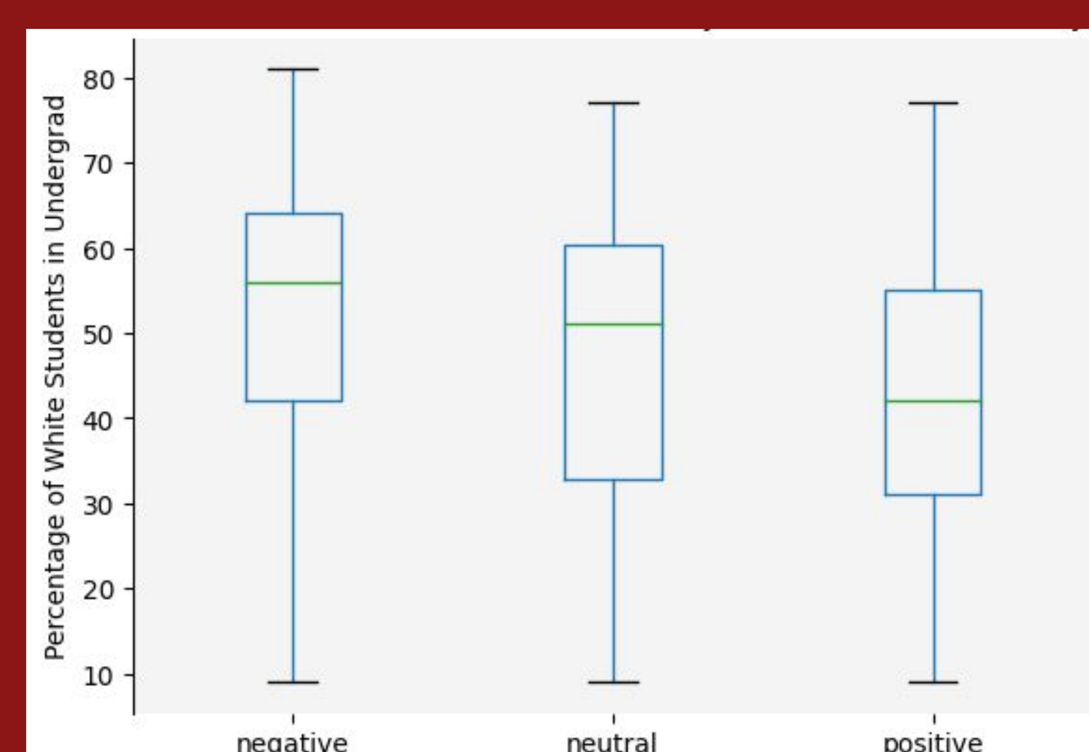
Methods



Students show **positive** feelings about *facilities*, *academics*, and *social life*, while discussions of *administration* and *diversity* often reflect dissatisfaction.

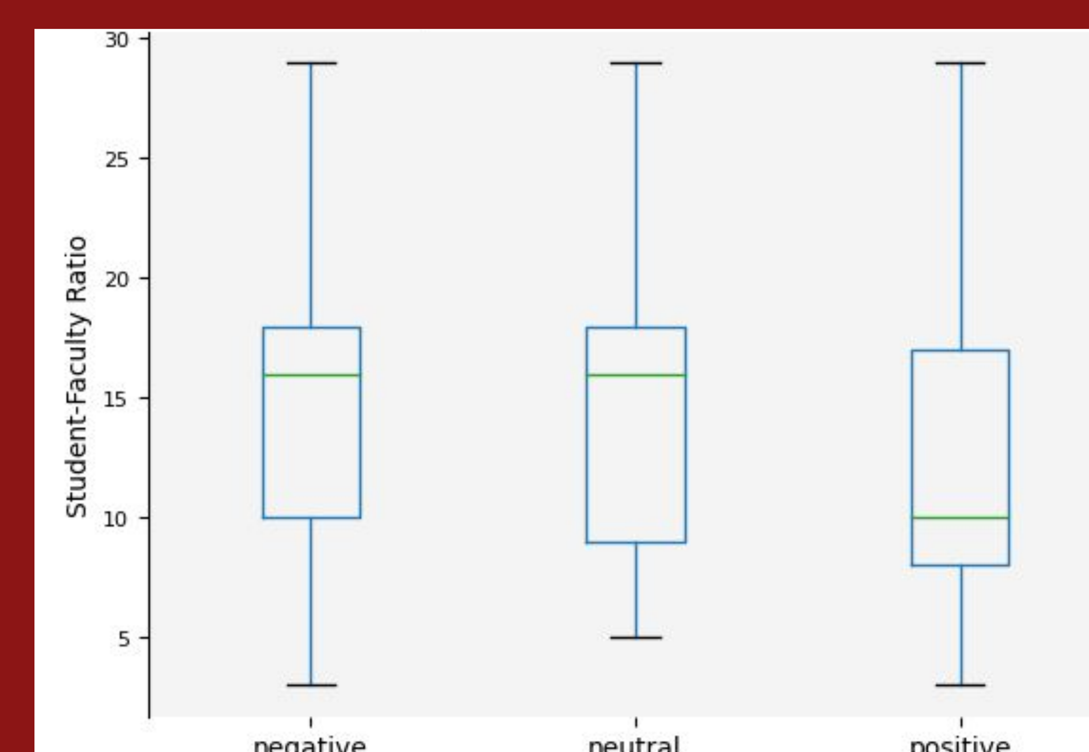


The **sentiments** expressed in online reviews mirror the **characteristics of universities**, such as the *diversity of student population*, *student-faculty ratio*, etc.



Schools that are frequently mentioned in a **positive** context regarding **diversity** tend to have **less homogenous** student body (n = 1,071).

ANOVA test: F-statistic = 58.08, p-value = 1.13e-24***



Schools that are frequently mentioned in a **positive** context regarding **professors** tend to have a **lower student-faculty ratio** (n = 1,669).

ANOVA test: F-statistic = 37.63, p-value = 1.06e-16***

Methods: Prompt Engineering

> You are an assistant who reviews school reviews and identifies the main topics mentioned. Please, define the main topics in the review separated by {delimiter}. The assigned topic label should be strictly within this list: ['academics', 'location',...,'university_resources']. The output should be a list with the following format: [<topic1>, <topic2>]. School review: {delimiter} {review} {delimiter} An example of defining topics in review: {delimiter} School review: {example_1} Output: {output_1} {delimiter}

System Prompt User Prompt Input Example

Results

- **GPT-3.5 Turbo with zero-shot prompting** reports a decent performance in multi-label classification of topics **(F1 score = 0.66)** and is highly accurate in determining positive or negative sentiments **(F1 score = 0.97)**.

Discussions and Conclusion

- GPT-based topic classification achieves up to 66% overlap with topic labels tagged by human annotators, demonstrating the capability of consistently creating outputs in the domain of university reviews.
- The analysis of reviews highlights a strong correlation between the diversity of the population and the perceived school climate, emphasizing the role of inclusivity in shaping student satisfaction.
- Online reviews provide valuable insights into academics and college experiences. This is particularly beneficial for students from underrepresented backgrounds who lack social networks.

References

1. Hasan, S., & Kumar, A. (2018). Digitization and Divergence: Online School 2. Ratings and Segregation in America. SSRN Electronic Journal.
2. Cabrera, A.F. and La Nasa, S.M. (2000), Understanding the College-Choice Process. New Directions for Institutional Research, 2000: 5-22.
3. Chumakov, S., Kovantsev, A., & Surikov, A. (2023). Generative approach to Aspect Based Sentiment Analysis with GPT Language Models. Procedia Computer Science, 229, 284–293.

Acknowledgement

I gratefully acknowledge the support and feedbacks from Dr. Sanne Smith, Dr. Dora Demszky, Hansol Lee and many more mentors and professors. Special thanks to Alice Zhang and Chelsea Dixon for review annotation.