Abstract

Our project aims to identify a potential ethically sound framework for universities to more effectively gauge student well-being using publicly available resources such as social media in addition to the typical survey methods used to identify student pain points during the semester.

Acquiring Public Data

The team is in the process of collecting relevant Tweets and Reddit posts using search criteria involving the use of hashtags and keywords related to the University in a specific time frame. We plan on aggregating the sentiment of these posts to gauge overall student well-being.

Survey Timeline

The McWell Center conducted a wellness survey in Fall 2020 (the height of covid related stress for students). Our timeline for collecting data reflects this same time frame. We plan on aggregating the sentiment of these posts to gauge overall student well-being.

Ethical Concerns

- Students may not be aware or comfortable with the University actively monitoring social media
- Ensuring data collection does not single out individuals
- Certain demographics may be more inclined to post, potentially biasing the data towards certain groups

Using iTREDS training to keep our project ethically sound

Good intentions are not enough when it comes to data-oriented projects. We must ensure that our data ethics practices are incorporated into this project to maintain the integrity of our solution. Data privacy, bias, and student comfort will all be considered in-depth.

Twitter API

Using the Twitter API, we are able to collect tweets geotagged from Notre Dame, Saint Marys, and Holy Cross. As Notre Dame Football is a national brand, we will filter based on content and timing in order to only analyze tweets by students, rather than tweets about our football team. These tweets will provide us with a view into the tri-campus community in the past two years, as well as a live “pulse” for future use.

Reddit Data

Using a reddit scraper, we have collected 2509 submissions, and 25,030 comments made on the /r/NotreDame subreddit between March 1, 2020 and October 15, 2021. Between January 1st, 2019 and March 1, 2020, we have collected 762 submissions and 5187 comments as a baseline. We are also able to scrape the subreddits of other universities across the country, as well as general college communities.

Privacy

All social media data scraped will be from publicly available sources, and be anonymized to protect privacy. Individual social media posts will not be shared with any outsiders.