Creating and Exploring a Dataset of Pro-Eating Disorder Material on the Internet

Abstract

On the internet, there exists material which encourages people to develop an eating disorder, or to maintain, heighten, or relapse into an existing eating disorder. This material can be very harmful to those who are vulnerable to it. As a project for my Machine Learning class, I created a dataset of pro-eating disorder and pro-recovery material from the internet using text scraping. Then, I experimented with Machine Learning methods on the dataset to attempt to differentiate pro-eating disorder and pro-recovery material.

Introduction / Background and Context

Anorexia is an eating disorder characterized by purposefully not eating enough, low body weight and associated health problems, fear of gaining weight, and distorted body perception. Because of the health problems associated with not eating enough, anorexia is one of the most deadly mental disorders. Bulimia is another eating disorder, characterized by episodes of binging and purging. Although people with bulimia are not underweight, there can be numerous health problems associated with throwing up repeatedly. The main sufferers of eating disorders are teenage girls, but they can happen to anyone.

One of the differences between eating disorders and other psychological disorders is that people with eating disorders often have a sense of pride in having the disorder, which is one of the factors that makes eating disorders very difficult to treat.

One of the wonders of the internet is that is can connect people with similar interests who would never meet in the real world. However, this also presents one of the biggest problems with the internet. When people who are proud of their anorexia and see their disorder as a lifestyle choice post their ideas on the internet, those ideas may be seen by many others. These “pro-ana” and “pro-mia” websites can lead people reading them to develop an eating disorder, or increase their disordered eating behavior.

These websites provide diets encouraging people to eat under 100 calories in a day, and tell people that their self-worth and beauty depends solely on being skinny. They have pictures of people ranging from on the skinnier side of healthy, to severely underweight. These pictures are called “thinspiration”, “thinspo” for short.

Some platforms try to block pro-ana content on their websites, but people can evade those filters by spelling keywords and hashtags differently. For example, some people spell “thinspo” as “thynspo.”

When I looked into making a machine learning algorithm to detect pro-eating disorder content online, I did not find anything that exactly fit the description. However, browser extensions to block certain upsetting content do exist. These browser extensions often use Machine Learning, rather than techniques like looking at hashtags. Using Machine Learning has the advantage of being harder to avoid by just spelling certain words differently, so I think it is the best approach to this particular problem.

Creating the Dataset

The first step in creating my dataset was to find sources. I started by simply Google searching “pro-ana sites”, and found a couple. Many of the top articles were about the harm of pro-ana sites, probably Google’s attempt to deter people from accessing pro-ana sites. I found most of my pro-ana sources by simply clicking past the first page of Google. Some pro-ana sites had links or mentions of others, and I found some that way too. The pro-recovery websites were a lot easier to find, I just searched Google for “recovery from anorexia” and similar queries, and got a lot of articles much more easily.

Once I had all the links I needed, I wrote a program called “downloadData.py” which would scrape the text from those website pages in a given url list, and save the text from each website as a text file, within a directory based on which list the url came from. I used the library “Beautiful Soup” to extract the text from the website’s code, using Beautiful Soup’s function, get_text(). This function returned the plain text and excluded the javascript and html.

The next program I wrote was called “formatData.py”, which has three main functions, and uses the libraries “xlrd” and “xlwt” to read and write to xls files. The first function is “make_sentences”, which separates files into a series of strings, dividing them along, “.” (the periods). The next function is called “makeSpreadsheet”. This function takes in a directory name, iterates through all the files in the directory, splits them up into sentences, and saves them in an .xls file with the same name as the directory. At this step, the data looks like this:

id	label	text
1_0	0	find out more here
2_0	0	i had a long battle with anorexia

The last step in the program, “formatData.py” is to combine the spreadsheets. It does so by reading from each spreadsheet, putting them in a pandas dataframe, putting the dataframes together, shuffling the dataframes, and then writing them to a spreadsheet, which I called “data.xls”.

I also wrote a file called “sortData.py”. It creates a terminal interface that will present the sorter with a sentence, ask them for a decision about which category it belongs in, and record that row of the spreadsheet in a different file with a different id based on the sorter’s decision. The code seems to work, but after sorting through a couple dozen sentences, I realized that it is difficult to tell whether a sentence is actually pro-ana, pro-recovery, or neutral is very difficult without the context. If I were to do this project again, I would divide by paragraph, or by a certain large number of characters, and sort the data in larger chunks. However, that seemed to be too large of a modification for me to make during the last week of my project.

Once I had the file, “data.xls”, I opened the file and saved it as a .csv. Then I uploaded the .csv file to Google Drive and got a sharable link for the file. Using the sharable link, I replaced the file id in the following line of code with the id from the sharable link.

gdown.download('https://drive.google.com/uc?authuser=0&id=10Te32ZdtwxPbKdqd5YZw6Ztoa4g7FYDB&export=download', 'finalprojectdata.csv', quiet=False)

Then, I was able to convert the csv into a pandas dataframe object using this line:

df = pd.read_csv('finalprojectdata.csv', header=0, delimiter=',')

From there, I was able to work with my dataset.

One of my project goals was to create a usable dataset and put it in the world for people to use, so I have put my data on the website Kaggle, so people can use my data for their own ML models. I also used a creative commons license, so that people can use my data without worrying about copyright.

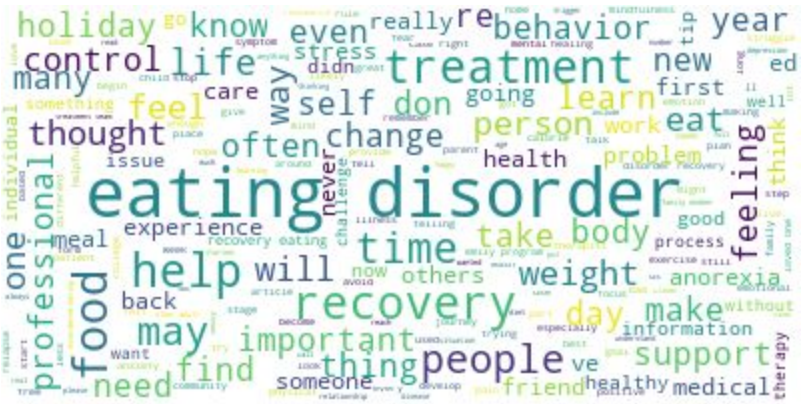
Here is the link to my dataset on Kaggle:
https://www.kaggle.com/kfoster608/proanavsprorecovery

You can also find the dataset by using the search terms “pro-ana” or “eating disorders” on Kaggle.

Visualizing the Dataset

Word Clouds

Pro-Ana Word Cloud



Pro-Recovery Word Cloud

During my project check-in, the professors suggested that it would be very helpful to make word clouds of the different sections of my dataset in order to understand and visualize it. Word Clouds display the most common words in datasets with a size that is proportionate to how often the word occurs. I found a tutorial that used a dataset from Kaggle, and pandas and matplotlib in order to create a word cloud using the library “wordcloud”. The tutorial was very helpful, as it used tools that I had already worked with, so it was easy to implement the tutorial on my own dataset.

I decided to make a word cloud for each of the pro-ana and pro-recovery datasets. I split the data from my dataset up into the two categories by using the code,

" ".join(sentence for sentence in df[df["label"]==1].text)

Where the label of 1 means that is is pro-ana content. A label of 0 would be pro-recovery content. Then I used the function WordCloud.generate() on the data that I had organized.

I found the word clouds very interesting to look at. In the Pro-recovery dataset, some of the most common words were, “eating”, “disorder”, “recovery”, and “treatment”. The Pro-Anorexia dataset focused on words like, “food”, “diet”, “eat”, “calorie”, and “don’t”. I predicted some of these patterns, like the word “calorie” appearing far more in the pro-ana dataset. The word “don’t” appearing more in the pro-ana dataset was surprising, but makes a lot of sense in hindsight, since a lot of the websites said “don’t eat x”. Due to my filtering punctuation out of the data when creating the dataset, “don’t” appears as “don” in the word cloud. There was also a lot more variety of words used in the pro-recovery dataset, with fewer big words, and more small words appearing in the word cloud.

Here is the link to the notebook in which I made word clouds:
https://colab.research.google.com/drive/14SNWbQjCuB1m1TehLpPKpjH9vZ_lgjbo

Analysis and Models

N-Grams

The first approach I attempted to differentiate pro-ana and pro-recovery material was n-grams. N-grams are used by splitting a sample of text into phrases with n words in them. For example, bigrams would be phrases with two words. The n-grams are contiguous, which means they overlap with each other. For example, the phase, “hot cross buns” would have the bigrams “hot cross” and “cross buns”. Machine Learning models can compare the n-grams of different text to determine things like sentiment, like we did earlier this semester.

The first step in my approach of n-grams was to split the sentences into a “bag of words”, splitting the sentences around the character of “ “ (space). Then, I used the nltk python library to transform the “bag of words” into n-grams. Then, CountVectorizer gets the top bigrams,and uses those bigrams to classify the sentences. I got an accuracy of 99% on the training set, and 84% on the testing set, which is not bad, but clearly the model is overfitting.

Here is the link to the notebook where I used N-Grams.
https://colab.research.google.com/drive/1MIqo7BcVb-oCm4EaNbFbRS6YYv7s21ZP

Discussion of Ethics

What would be the use of a machine learning model that can detect pro-eating disorder material online? Well, there are several cases in which it could be used, but it is important to consider the side effects and ethics of using such a model.

One possible implementation would be to make a browser extension that could blur our pro-ana content. This approach has the fewest ethical problems, since people would be participating in it voluntarily and for their own benefit. The only consideration I can think of is that if a person using this extension might become more sensitive to pro-eating disorder material through lack of exposure. However, I think the good of this approach would outweigh the harm, and people should decide for themselves what they want to see online, as long as what they see does not hurt anyone else.

Another possible implementation is again as a browser extension or app, but to alert the person's parents of healthcare professionals when the person views too much pro-eating disorder material. Viewing such material is often a sign of a relapse, and it is helpful to catch and treat a relapse as early as possible. This raises the ethical dilemma of invasion of privacy. It is doubtful that someone who intends to view pro-ana material would voluntarily download such a program. Many people with eating disorders are minors, so the parents could force their child to download the program, but is it fair to give tools to allow parents to violate their children's privacy. However, if the other option is that the parent takes away all connection to the internet, using a program like this might be a good compromise. Using this ML program would allow people to remain connected to their friends and positive online communities, which would encourage recovery, without the chance of going back to pro-ana websites.

Social Media platforms could also use a similar algorithm to block pro-ana content on their platforms. This application raises ethical questions of free speech, like every attempt to block harmful content on a social media platform does. Although blocking certain content on a social media site, like asking someone to leave your house for saying certain things, does not violate the constitutional right to free speech, it can prevent having a completely honest discussion with all points of view. The social media platform has to ask themselves what is more important, full honesty, or protecting people from harm.

Sources

Background Information:
https://en.wikipedia.org/wiki/Pro-ana
https://www.newsweek.com/2016/07/01/pro-ana-websites-anorexia-nervosa-473433.html
https://www.wired.com/story/how-pro-eating-disorder-posts-evade-social-media-filters/
https://www.cbsnews.com/news/despite-social-media-bans-of-pro-ana-websites-pages-persist/
https://www.nytimes.com/2002/09/08/magazine/the-way-we-live-now-9-8-02-phenomenon-a-secret-society-of-the-starving.html

Word Cloud Tutorial:
https://www.datacamp.com/community/tutorials/wordcloud-python