A project using NLP on the dataset “Comments on articles published in the New York Times” (<https://www.kaggle.com/aashita/nyt-comments>).

The ultimate goal of this project is to build a web interface in which data for a new article (topic, summary, title, tags) could be entered, and the model will:

1. generate statistics such as the number of comments that this article would receive and the number of abusive comments (reported by other readers).
2. This model will also generate a sample of possible comments which would be made on this article, as well as an estimate of how many recommendations each comment would receive.

For step 1), the model will use metadata about articles published in the New York Times including headlines, topic keywords, word count, and snippets of text from the article to predict the total number of comments that a test article will receive.

I can train my model by comparing the known comment count to the predicted count.

For step 2), the article will be trained on a dataset of comments for these articles to generate comments for the above articles and evaluating how many recommendations each comment will receive.

This project should ultimately be hosted on a simple web server which can take an input through text boxes, or through uploading a file. The output can be displayed on the page or downloaded.