

MSIA 414 Independent Project Part 1 – Anthony Colucci

I am planning to use my independent project to do research on a set of historical Jeopardy questions. The dataset I am considering is available online at https://drive.google.com/file/d/0BwT5wj_P7BKXUI9tOUJWYzVvUjA/view and contains 216,930 questions from the trivia TV show airing between 1985 and 2015. Each observation includes a textual question and answer.

The idea that interests me most for this data set is to use Named Entity Recognition over the questions on this dataset. I imagine that this was a significant piece of the success of the Watson model, and as an avid fan of Jeopardy and trivia, I have a lot of interest in the background of that project. I am also potentially interested in extending the project to create word embeddings of the question and to use the “math” of word embeddings that has been shown in class and our homework to measure distance from the correct answers, though I am not yet certain if that is something that is possible.