I Compared the size of each string, how many similar words are in there, and checked the Damerau Levenshtein Distance. In addition, I transformed the text to remove all double quotes and make everything lowercase as a way to better standardize the output.

I used the C# NuGet Machine learning library, which contains the currently used models for both Machine Learning and Deep Learning. I used only the Machine learning libraries. The model I used was Logistic Regression. While SVM is historically better with text comparison, for some reason I was unable to get it to work, and as a result used LR instead.

I learned so many useful things, including some critical errors that I found out with only a few minutes to spare after reviewing my output. I know if I was aware of these errors in advance I would have had a much better outcome compared to what I will most likely be getting, as I went until the last day thinking I had a near 80% accuracy, just to find out I was reversing the proportions of data I was actually trying to use. Regardless, it was a great experience to get hands-on practice with machine learning, and I know the final project will be much more successful.