Steps to clean text related to twitter:

1. Clean text html tags
   1. Full html tags
   2. Partial html tags
2. Handle non-ascii characters
   1. accented characters
   2. incorrectly encoded characters
   3. determine when to add spaces or auto correct words
3. Remove any URLs
4. Remove new lines and tabs
5. Remove twitter handles/unique references
   1. Retweets
   2. Mentions
   3. Hastags
   4. Email addresses
6. fix words with unnecessary repeating characters
7. Remove emoticons/emojis
8. Handle slang terms
   1. Convert abbreviations
   2. Handle slang terms
   3. Auto correct words
9. Convert dates to format that doesn’t include strings
   1. Convert to mm/dd/year for length 3
   2. Convert to mm/dd for length 2
10. Expand contractions
11. Remove numbers and special characters
12. Remove extra spaces
13. Remove stop words and lemmatize data

Simple machine learning approach

Best model was using Bernoulli Navie Bayes classifier

Deep learning approach