

References:

- <https://mtsamples.com/>
- https://en.wikipedia.org/wiki/Medical_transcription
- <https://www.simplilearn.com/regression-vs-classification-in-machine-learning-article>
- <https://smltar.com/tokenization.html>
- <https://itnext.io/introduction-to-natural-language-processing-nlp-tools-for-python-cf39af3cfc64>
- <https://www.geeksforgeeks.org/using-countvectorizer-to-extracting-features-from-text/>
- <https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761>
- <https://medium.com/analytics-vidhya/understanding-logistic-regression-b3c672deac04>
- <https://machinelearningmastery.com/logistic-regression-for-machine-learning>
- <https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/>