COM 527 – Assignment

Using naïve bayes to classify text documents is a simple approach that assumes that all the conditions are independent meaning that the words used within each of the documents are not considered to be in any specific order or follow a sentence structure, the algorithm uses the probability that each word shows up in data sets for each classification.

An advantage of the naïve Bayes algorithm is that if the conditional independence assumption is correct the classifier can converge very fast requiring much less training data than other models.

The algorithm is incredibly scalable across thousands of words which is typical of the average vocabulary of a human, due the simplicity of the algorithm It scales linearly.