**Assignment 2**

*Sonny Desai*

February 2020

**Identify the problem, describe data, clarify**

Clustering is an essential part of machine learning and natural language processing. The main point of clustering is to group similar items together so that each group or cluster, contains items that are similar to one another. In this corpus, movie reviews have been compiled from different films into a large text document. The issue is to find of all of the nearly 100 films, which can be grouped together based upon similarities in their movie reviews. The data for this experiment is interesting, because on its face, the data has very little similarities between each other. The corpus is compiled of movie reviews of each film which have been written by different writers, with different editors, along a large period of time (nearly forty years!)

**Summarize algorithm/key features of code**

**Review results**

**Interpret results**

**Summarize insights and findings**