<u>CSE4022</u>

<u>Natural Language Processing</u>

<u>J-Component Review 1</u>

Shivansh Mishra 17BCB0112

Het Patel 17BCE0567

Project Title:

Commentator Trait identification using NLP

<u>Overview:</u>

We plan to use the power of Natural Language processing to our advantage by using it on a dataset of cricket commentary (Cricbuzz/ESPN CricInfo). The first step would be that of finding the keywords from the commentary set. This will be done by classification of most common words and removal of stop words. We will find the top monograms, bigrams and tri grams to do this. After this, we would convert this into a vector table of word counts which would help us find the frequencies of the word. This would finally give us the keywords.

The keywords can be classified into various emotions like angry, sad, happy. A good application of this would be in selecting the best commentators together, or when selecting new personnel to do a sports related segment. Since polar opposites would likely do better, a good mix of emotions will be favourable. We would try finding those people who use the most different class of words and pair them together.