1. Create Unigrams + Bigrams + Trigrams Feature vector

**Responsible: Kasun**

1. Stemming

Responsible: Gerard

1. Combine Synonyms

**Responsible: Syed**

1. Optimal feature selection

Possibly Genetic Algorithm

**Responsible: Kasun**

1. Retaining additional Features to augment information of the feature vector
   1. Emoticons

Possibly replace the emoticons with a string which has little chance of appearing anywhere else

*E.g. Replace “☺” with ”HappyFaceEmoticon” or ”Emoticon1” etc.*

**Responsible: Gerard**

1. JSON (Hope it’s correct) format for Tweets

**Responsible: Gerard**