**Tokenization**

Eliminate words over a certain length

Limit to plain English

Removing non English words

Non English words that are jargon

**Vectorization**

Hashing?

How to get plain English back out

**Features**

How to get context from other columns (labelled data)

Number of words

**Containing**

Slack

Facebook

Twitter

How to store parameters (These knobs that can be tweaked at a later time.

Reasonable scope of work