

Preprocess text

Noise filtering (stop words)

Normalization (stemming)



Tokenisation



Vectorization



Count Vectoriser