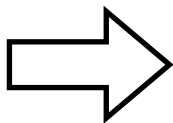
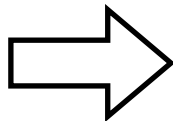


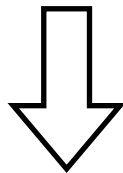
Combined words to
form dictionary of
words



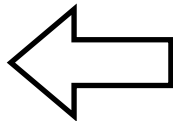
Reduced words in
dictionary by
combining similar
words



Each document was
represented as a
feature vector



Dimension was
reduced further using
PCA



This data was finally
fed to a classifier