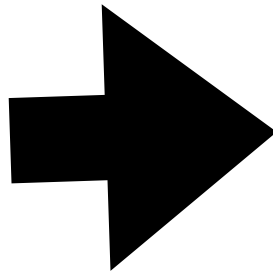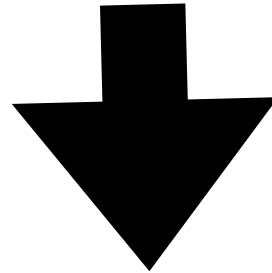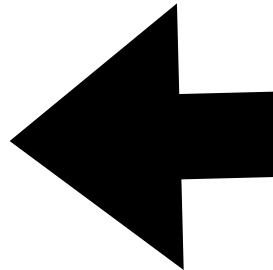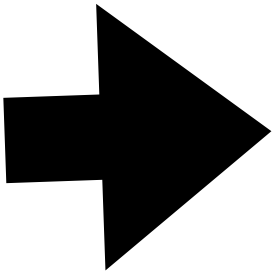Training set of text documents.

Apply preprocessing
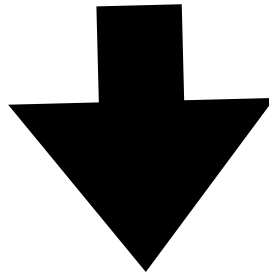(Removing stop-words, stemming,
lemmatized)

Extract-feature
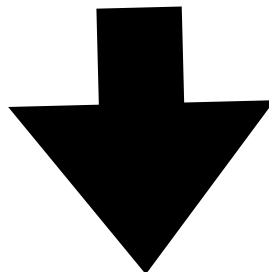(Using TFIDF, Bag of words)
get dominant topic as feature for

Implement topic modelling to
classification
(Using LDA with TFIDF, and LDA
with bag of word)

Encode keyword get from dominant topic for classification
(CountVectorizer)

Select appropriate machine learning model
(Naive Bayes, Logistic Regression, Support Vector Machine)

train classifier and test classifier