# Good Word Attacks on Spam Filters

### Implementation

The author implements the paper using 2 different classifiers.

### Phase one: data collection and analyzing the data

### In the Phase 1, the author utilizes the dataset from spamarchive.org from the month of April 2004. But neither the dataset nor the website mentioned in citation of the project seems to be valid anymore. Hence, to make the project doable we have chosen an alternate dataset. We have chosen the dataset which contains spam messages/ SMS’s instead of Spam emails. As part of the preparation of the dataset for the classification, we have arranged the dataset in such a way that the first column of the data contains whether its spam / ham (legitimate). This dataset is in tab separated values, this can be read by read\_csv using the separator “\t”. The dataset consists of 5572 records

### Phase two: training the model on individual data sets

### Phase three: testing the models