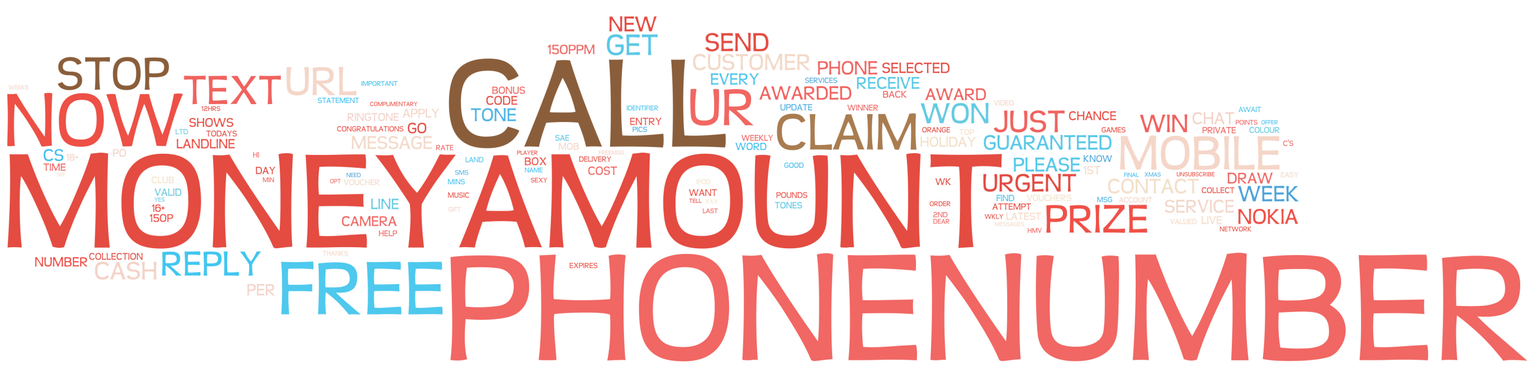
Wide use of SMS results increasing amount of spam messages. Classification of SMS message for spam detection helps manage our messages efficiently and reduce company's cost on message service.

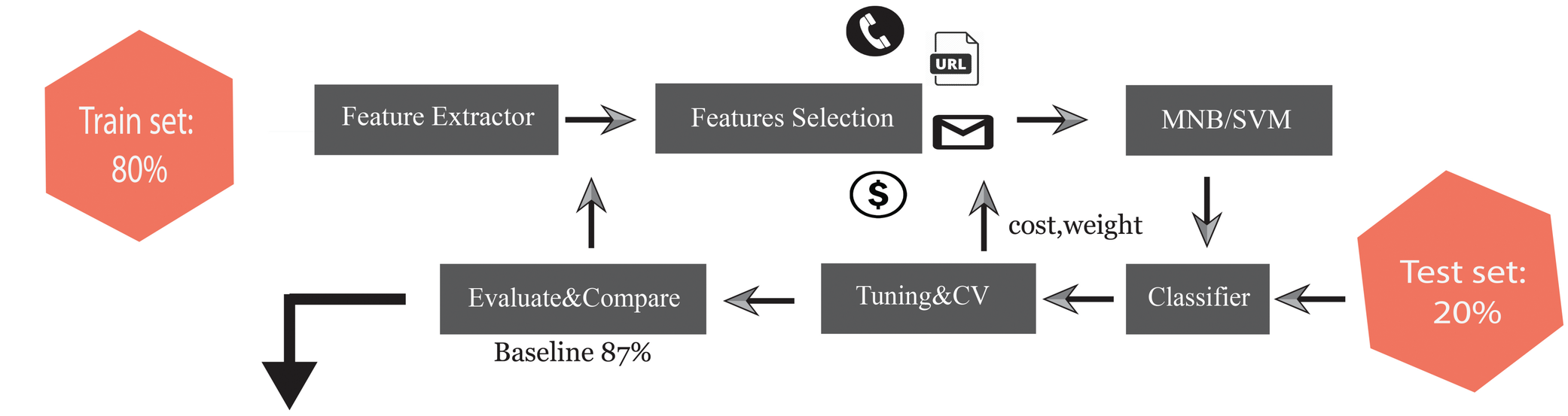
Dataset 3000 orignial messages with labels from Kaggle

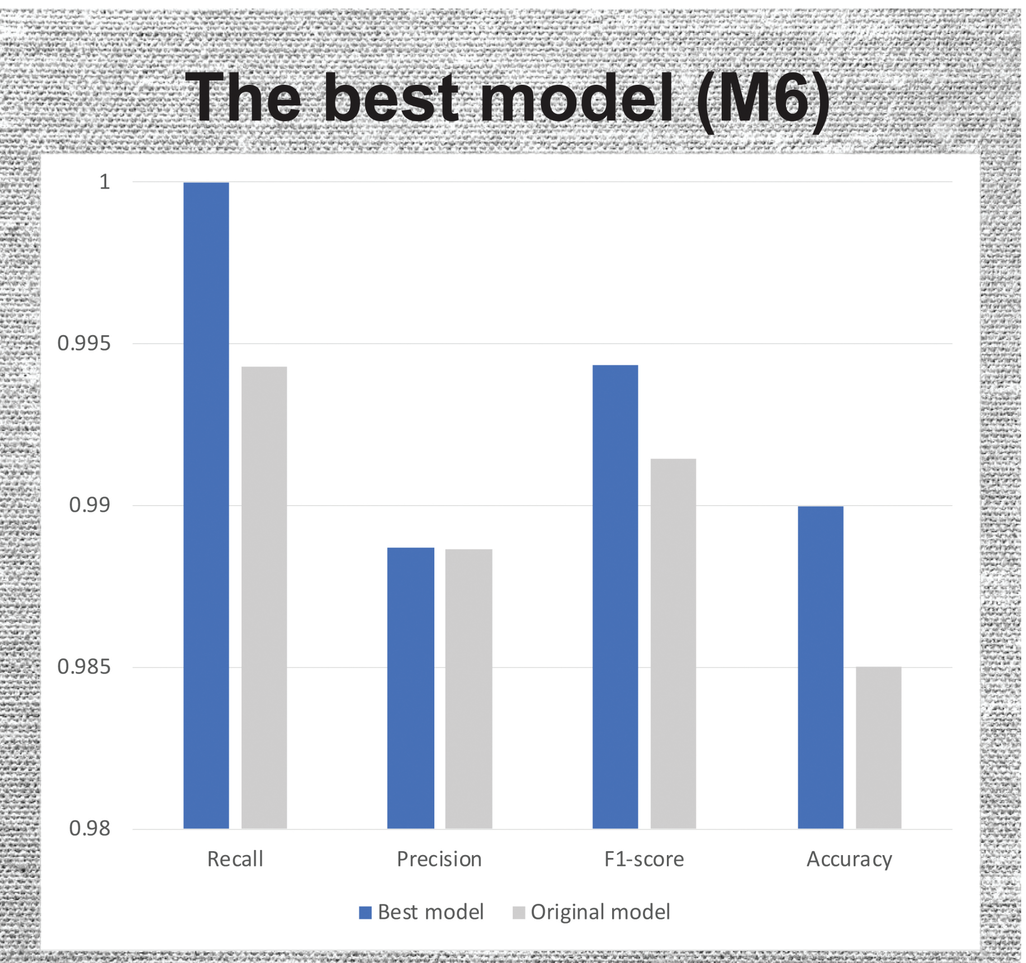
Processing Removing stop word lists and setting mean of document term frequency=5



SMS CLASSIFICATION FOR SPAM DETECTION

Yunteng Gu, Nan Sun, Jiarong Yang





Result1
Result2
Result3
Result4
Result5

M0: Unigram with TF

Recall: 0.984
Precision: 0.994
F1-score: 0.989

517	8
3	72

Naive Bayes

522	3
6	69

SVM

Recall: 0.994
Precision: 0.989
F1-score: 0.991

M1: M0+Phone

518	7
2	73

NB

522	3
5	70

SVM

↗ improved

M2: M0+URL

517	8
3	72

NB

521	4
6	69

SVM

M3: M0+Email

517	8
3	72

NB

522	3
6	69

SVM

M4: M0+Money

517	8
3	72

NB

523	2
6	69

SVM

↗ improved

!!Phone number and Money Amount can improve perfor-mance!!

M6: M1+M4+FullRecall

525	0
6	69

Recall: 1.000
Precision: 0.989
F1-score: 0.994

?

M5: M1+M4

523	2
2	73

Recall: 0.996
Precision: 0.996
F1-score: 0.996

But recall is more important!