

Report

Sai krishna prasad Kurapati

Sxk190015@utdallas.edu

Total no of training samples:

No of Documents	Ham Samples	Spam Samples
Train Documents =463	340	123
Test Samples = 478	348	130

Total Stop Words Taken = 429

Naïve Bayes Results

List of Accuracy's	Without Stop words	Using Stop Words
Ham Accuracy	97.701	97.126
Spam Accuracy	85.384	89.230
Total Accuracy	94.351	94.979

Output in Console

```
In [55]: runfile('C:/Users/SAI KRISHNA/Desktop/NB.py', wdir='C:/Users/SAI KRISHNA/Desktop')
TotalAccuracy: 94.35146443514645
HamAccuracy: 97.70114942528735
SpamAccuracy: 85.38461538461539

In [56]: runfile('C:/Users/SAI KRISHNA/Desktop/NBsw.py', wdir='C:/Users/SAI KRISHNA/Desktop')
Using Stop Words
TotalAccuracy: 94.97907949790795
HamAccuracy: 97.12643678160919
SpamAccuracy: 89.23076923076924

In [57]:
```

Here we can see that the **overall accuracy** has been increased when executed using stop words slightly compared to without using stop words

And also **spam accuracy** has been increased significantly by using the stop words whereas the **ham accuracy** has decreased slightly with stop words which is due to elimination of some useful stop words

Logisitic Regression

Total iterations : 100 , learning rate 0.001

for lamda = 0.001

List of Accuracy's	Without Stop words	Using Stop Words
Ham Accuracy	95.142	94.182
Spam Accuracy	88.281	93.162
Total Accuracy	93.305	93.933

for lamda = 0.01

List of Accuracy's	Without Stop words	Using Stop Words
Ham Accuracy	95.142	94.182
Spam Accuracy	88.281	93.162
Total Accuracy	93.305	93.933

for lamda = 2

List of Accuracy's	Without Stop words	Using Stop Words
Ham Accuracy	94.871	93.922
Spam Accuracy	88.188	93.103
Total Accuracy	93.096	93.723

Here we have calculated Logistic Regression for different **lamda values** and fixed learning **rate 0.001** for taking the smaller steps possible to converge and have performed **100 iterations**

As lamda value is increasing we can see that the Ham and Spam accuracies are changing slightly and decreasing due to elimination some useful stop words. And even without using stop words we can see that values are decreasing by the change of the lamda value

Shown in console

```
In [70]: runfile('C:/Users/SAI KRISHNA/Desktop/LR.py', wdir='C:/Users/SAI KRISHNA/Desktop')
```

```
TotalAccuracy : 93.30543933054393
```

```
HamAccuracy : 95.14285714285714
```

```
SpamAccuracy : 88.28125
```

```
TotalAccuracy : 93.30543933054393
```

```
HamAccuracy : 95.14285714285714
```

```
SpamAccuracy : 88.28125
```

```
TotalAccuracy : 93.09623430962343
```

```
HamAccuracy : 94.87179487179486
```

```
SpamAccuracy : 88.18897637795276
```

Using Stop words

```
In [71]: runfile('C:/Users/SAI KRISHNA/Desktop/LRsw.py', wdir='C:/Users/SAI KRISHNA/Desktop')
```

Using Stop Words

TotalAccuracy : 93.93305439330544

HamAccuracy : 94.18282548476455

SpamAccuracy : 93.16239316239316

TotalAccuracy : 93.93305439330544

HamAccuracy : 94.18282548476455

SpamAccuracy : 93.16239316239316

TotalAccuracy : 93.72384937238493

HamAccuracy : 93.92265193370166

SpamAccuracy : 93.10344827586206

Thank you