**Report**

**Sai krishna prasad Kurapati**

[**Sxk190015@utdallas.edu**](mailto:Sxk190015@utdallas.edu)

**Total no of training samples:**

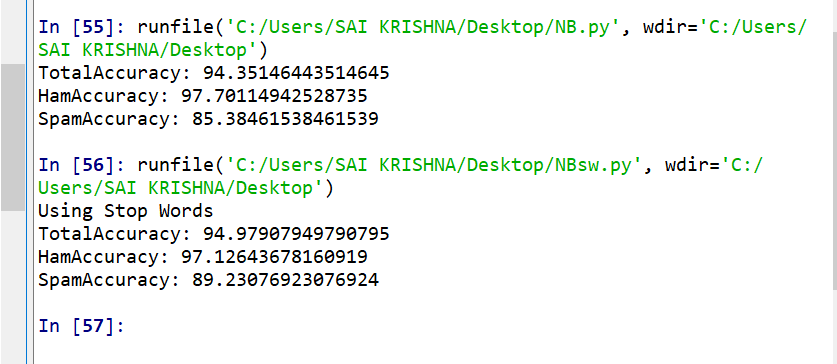
|  |  |  |
| --- | --- | --- |
| **No of Documents** | **Ham Samples** | **Spam Samples** |
| **Train Documents =463** | 340 | 123 |
| **Test Samples = 478** | 348 | 130 |

**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**

****

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** |

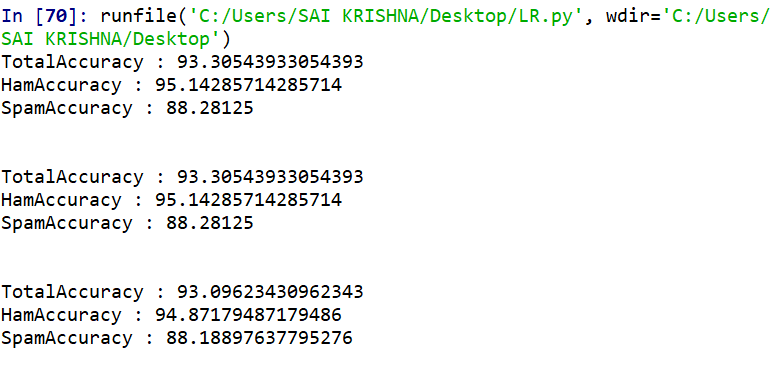
for lamda = 2

|  |  |  |
| --- | --- | --- |
| **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** |

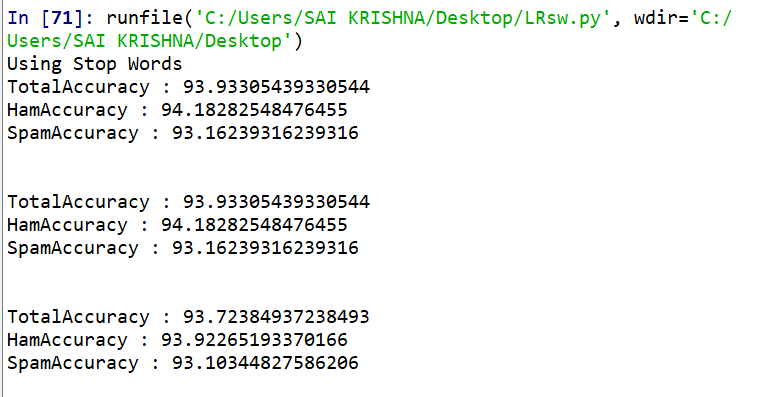
Here we have calculated Logistic Regression for different **lamda values** and fixed learning **rate 0.001** for taking the smaller steps possible to converge and have performed **100 ite**rations

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



Using Stop words



Thank you