

Text Classification

Aman Berhe

May 9, 2016

1 Tf-Idf

To compute the tf-idf i have used the two documents and the frequency of the terms. Since we only have two documents and the terms can be found either in one or in both of the documents, Simplifying the computation of the log i have done the following

if the term exist in both of the documents: put 0

since $\log(2/2)$ is 0

else multiply the frequency by 0.3

since $\log(2/1)$ is 0.301029996

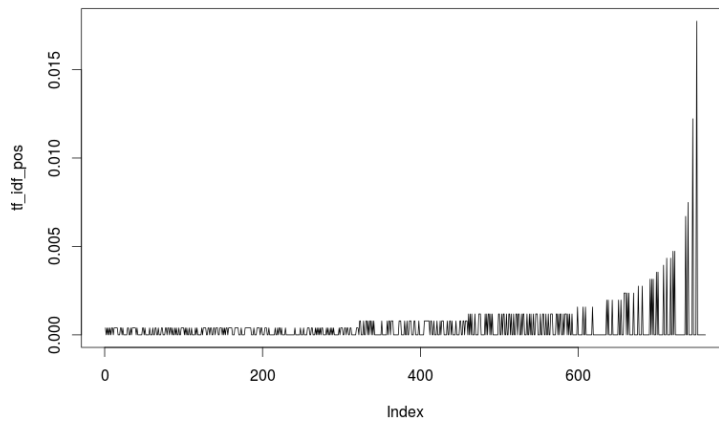
here are the results and plot

1.1 Tf-Idf for positive

[illegible]

```
0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0007884363 0.0000000000 0.0000000000 0.0000000000
0.0000000000 0.0007884363 0.0007884363 0.0000000000 0.0000000000 0.0007884363 0.0007884363 0.0000000000
0.0007884363 0.0007884363 0.0007884363 0.0000000000 0.0011826544 0.0000000000 0.0011826544 0.0000000000
0.0011826544 0.0000000000 0.0000000000 0.0000000000 0.0011826544 0.0000000000 0.0000000000 0.0000000000
0.0000000000 0.0011826544 0.0011826544 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
0.0011826544 0.0000000000 0.0011826544 0.0000000000 0.0000000000 0.0011826544 0.0011826544 0.0000000000
0.0000000000 0.0011826544 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
0.0000000000 0.0000000000 0.0011826544 0.0011826544 0.0000000000 0.0000000000 0.0000000000 0.0011826544
0.0011826544 0.0000000000 0.0000000000 0.0000000000 0.0011826544 0.0000000000 0.0000000000 0.0011826544
0.0000000000 0.0000000000 0.0011826544 0.0000000000 0.0000000000 0.0011826544 0.0000000000 0.0000000000
0.0011826544 0.0000000000 0.0011826544 0.0000000000 0.0000000000 0.0011826544 0.0000000000 0.0000000000
0.0000000000 0.0011826544 0.0011826544 0.0000000000 0.0000000000 0.0000000000 0.0011826544 0.0000000000
0.0000000000 0.0011826544 0.0011826544 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0011826544
0.0000000000 0.0011826544 0.0011826544 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
0.0000000000 0.0011826544 0.0000000000 0.0000000000 0.0011826544 0.0011826544 0.0011826544 0.0000000000
0.0011826544 0.0011826544 0.0000000000 0.0000000000 0.0011826544 0.0000000000 0.0000000000 0.0011826544
0.0000000000 0.0000000000 0.0011826544 0.0000000000 0.0000000000 0.0011826544 0.0000000000 0.0000000000
0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0015768725 0.0000000000
0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
0.0015768725 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0019710907 0.0000000000 0.0019710907 0.0000000000
0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
0.0000000000 0.0000000000 0.0019710907 0.0000000000 0.0000000000 0.0000000000 0.0019710907 0.0000000000
0.0000000000 0.0000000000 0.0019710907 0.0000000000 0.0000000000 0.0000000000 0.0019710907 0.0000000000
0.0000000000 0.0023653088 0.0023653088 0.0023653088 0.0000000000 0.0023653088 0.0000000000 0.0023653088
0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0023653088 0.0000000000 0.0000000000
0.0000000000 0.0000000000 0.0000000000 0.0027595269 0.0000000000 0.0000000000 0.0000000000 0.0000000000
0.0027595269 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
0.0000000000 0.0000000000 0.0031537451 0.0000000000 0.0031537451 0.0000000000 0.0031537451 0.0000000000
0.0000000000 0.0000000000 0.0035479632 0.0000000000 0.0035479632 0.0000000000 0.0000000000 0.0000000000
0.0000000000 0.0000000000 0.0000000000 0.0039421813 0.0000000000 0.0000000000 0.0000000000 0.0043363995
0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0043363995 0.0000000000 0.0000000000 0.0047306176
0.0000000000 0.0047306176 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0067017083
0.0000000000 0.0000000000 0.0074901445 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
0.0122207622 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0177398160 0.0000000000
0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
0.0000000000
```

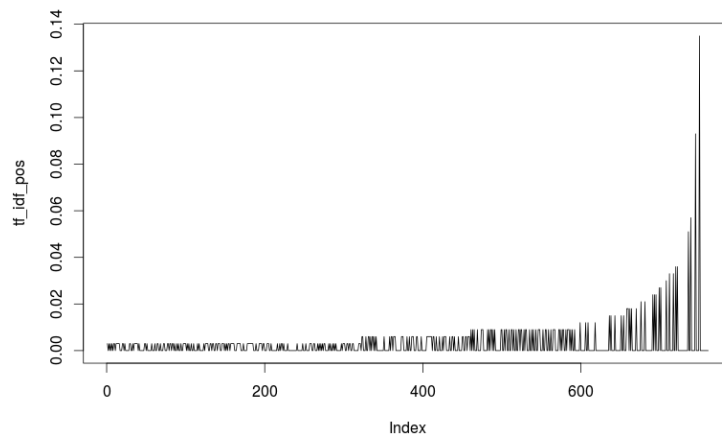
the Plot looks like this



1.2 Tf-Idf for positive with the number of sentences

```
0.003 0.000 0.003 0.000 0.003 0.000 0.003 0.000 0.003 0.003 0.000 0.003 0.003 0.003 0.003 0.000 0.000
0.000 0.003 0.003 0.000 0.003 0.000 0.000 0.000 0.000 0.003 0.003 0.000 0.000 0.003 0.000 0.003 0.003
0.003 0.003 0.003 0.000 0.003 0.000 0.000 0.000 0.000 0.000 0.003 0.003 0.000 0.003 0.000 0.000 0.000
0.000 0.000 0.003 0.000 0.000 0.000 0.003 0.000 0.000 0.003 0.003 0.000 0.000 0.000 0.000 0.000 0.003
0.003 0.000 0.000 0.000 0.003 0.003 0.000 0.003 0.003 0.000 0.003 0.000 0.003 0.000 0.003 0.000 0.000
0.003 0.000 0.000 0.003 0.000 0.000 0.003 0.003 0.003 0.003 0.000 0.003 0.000 0.003 0.000 0.000 0.003
0.000 0.003 0.000 0.000 0.000 0.000 0.003 0.000 0.003 0.000 0.000 0.000 0.000 0.000 0.000 0.003 0.003
0.003 0.003 0.000 0.000 0.003 0.003 0.000 0.003 0.003 0.000 0.000 0.003 0.003 0.000 0.003 0.003 0.000
0.000 0.000 0.003 0.003 0.003 0.003 0.000 0.000 0.000 0.003 0.000 0.000 0.000 0.000 0.000 0.003 0.003
0.003 0.003 0.003 0.003 0.003 0.000 0.000 0.000 0.000 0.003 0.000 0.000 0.000 0.000 0.000 0.000 0.003
0.003 0.000 0.000 0.000 0.003 0.003 0.003 0.000 0.000 0.003 0.000 0.000 0.000 0.000 0.000 0.000 0.003
0.000 0.000 0.003 0.000 0.003 0.003 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
0.000 0.000 0.000 0.000 0.000 0.003 0.003 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.003 0.000 0.003
0.003 0.000 0.003 0.000 0.003 0.003 0.000 0.000 0.000 0.000 0.003 0.000 0.003 0.000 0.000 0.003 0.000
0.000 0.003 0.000 0.000 0.000 0.000 0.000 0.000 0.003 0.000 0.003 0.003 0.003 0.000 0.003 0.003 0.000
0.003 0.003 0.000 0.000 0.003 0.003 0.000 0.000 0.000 0.000 0.000 0.000 0.003 0.003 0.003 0.000 0.006
0.000 0.000 0.000 0.000 0.006 0.000 0.006 0.006 0.000 0.006 0.000 0.006 0.006 0.000 0.006 0.006 0.000
0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.006 0.000 0.000 0.000 0.000 0.000 0.000 0.006 0.006
0.006 0.000 0.006 0.006 0.006 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.006 0.006 0.000 0.000
0.000 0.006 0.000 0.000 0.006 0.000 0.000 0.006 0.006 0.006 0.000 0.000 0.000 0.006 0.006 0.000 0.000
0.000 0.006 0.000 0.000 0.000 0.000 0.000 0.000 0.006 0.006 0.006 0.006 0.006 0.006 0.000 0.000 0.000
0.000 0.000 0.006 0.000 0.000 0.000 0.000 0.006 0.000 0.000 0.000 0.006 0.006 0.006 0.006 0.000 0.000
0.000 0.006 0.000 0.000 0.006 0.006 0.000 0.006 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.006
0.006 0.000 0.000 0.006 0.006 0.000 0.006 0.006 0.006 0.000 0.000 0.000 0.009 0.000 0.009 0.000 0.000
0.009 0.000 0.000 0.000 0.000 0.009 0.009 0.009 0.000 0.000 0.000 0.000 0.000 0.000 0.009 0.000 0.000
0.009 0.009 0.000 0.000 0.009 0.000 0.000 0.009 0.009 0.000 0.000 0.000 0.000 0.000 0.009 0.000 0.009
0.000 0.009 0.009 0.000 0.000 0.009 0.000 0.009 0.009 0.000 0.000 0.000 0.009 0.000 0.009 0.000 0.009
0.000 0.000 0.009 0.000 0.000 0.009 0.009 0.009 0.000 0.000 0.000 0.009 0.000 0.009 0.009 0.000 0.000
0.009 0.000 0.000 0.009 0.000 0.000 0.009 0.009 0.009 0.000 0.000 0.000 0.009 0.009 0.000 0.009 0.000
0.009 0.009 0.000 0.000 0.009 0.000 0.009 0.009 0.009 0.000 0.009 0.000 0.009 0.000 0.009 0.000 0.000
0.000 0.000 0.000 0.000 0.012 0.000 0.000 0.000 0.000 0.000 0.000 0.012 0.000 0.000 0.012 0.000 0.000
0.000 0.000 0.000 0.000 0.000 0.012 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
0.000 0.000 0.000 0.000 0.000 0.015 0.000 0.015 0.000 0.000 0.000 0.000 0.015 0.000 0.000 0.000 0.000
0.000 0.000 0.015 0.000 0.000 0.015 0.000 0.000 0.000 0.018 0.018 0.018 0.000 0.018 0.000 0.018 0.000
0.000 0.000 0.000 0.018 0.000 0.000 0.000 0.000 0.000 0.021 0.000 0.000 0.000 0.000 0.021 0.000 0.000
0.000 0.000 0.000 0.000 0.000 0.000 0.024 0.000 0.024 0.000 0.024 0.000 0.000 0.000 0.027 0.000 0.027
0.000 0.000 0.000 0.000 0.000 0.030 0.000 0.000 0.000 0.033 0.000 0.000 0.000 0.000 0.033 0.000 0.036
0.000 0.036 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.051 0.000 0.000
0.057 0.000 0.000 0.000 0.000 0.000 0.093 0.000 0.000 0.000 0.000 0.135 0.000 0.000 0.000 0.000 0.000
0.000 0.000 0.000 0.000 0.000
```

the Plot looks like this



1.3 Tf-Idf for negative

[illegible]

the Plot looks like this



1.4 Tf-Idf for negative with the number of sentences

[illegible]

the Plot looks like this

text classification