The Library

Train on CPP-2012.txt has 7 chapters.

Test on CPP-2016.txt has 56 chapters.

Train on NDC-2012.txt has 8 chapters.

Test on NDC-2016.txt has 38 chapters.

Train on NPP-2012.txt has 8 chapters.

Test on NPP-2016.txt has 18 chapters.

**Confusion Matrix**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Confusion Matrx  Actual/Predicted | CPP | NDC | NPP | Error Rate |
| CPP | 4 | 1 | 2 | 0.4285714 |
| NDC | 1 | 7 | 0 | 0.125 |
| NPP | 1 | 0 | 7 | 0.125 |

Interpretation

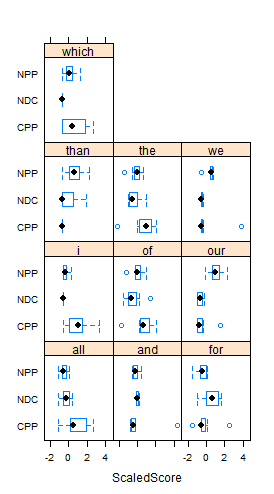
After training the randomForest model with the 2012 party manifestos, it was noticed that the model could clearly distinguish the writing styles of NDC and NPP since 7 out 8 chapters of both parties’ manifestos were clearly distinguished. For CPP, it was slightly successful at distinguishing the party’s manifesto as belonging to CPP’s writing style.

**The OOB Error Rate**

OOB estimate of error rate: 21.74%. This is a relatively high estimate and shows that the model found difficulty in distinguishing between the writing styles of the various parties.

Graph 1

Form the stopwords graph, CPP used the words are “which”, “the”, “we”, “i”, “of”, “all”, “and”, and “for” than any other party. NPP used the words “than” and “our” more than any other party in their manifestos. NDC did not use any stopword more than the other parties. Hence it can be concluded that CPP used stopwords more in their manifestos than any other party.



Interpretation of Graph 1

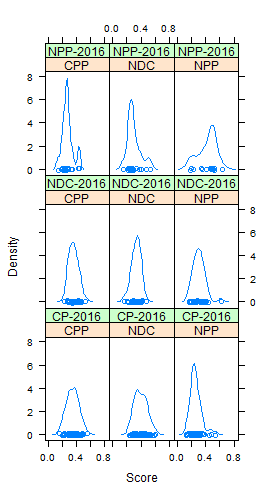
**Guess**

|  |  |  |  |
| --- | --- | --- | --- |
|  | CPP | NDC | NPP |
| CPP | 31 | 20 | 5 |
| NDC | 12 | 17 | 9 |
| NPP | 4 | 2 | 12 |

Interpretation

For the testing files, the randomForest model was not able to clearly distinguish the writing styles of any of the parties although it did a good work to some extent. For instance, from the 56 chapters tested for CPP, 31 out of them matched CPP’s writing style and 20 matched NDC’s writing style. Similar trends were observed for the other parties and this shows that the three political parties do not have an entirely unique writing style.

Graph 2



**Interpretation**