# Assignment 4 - Text Classification

## Dharmam Buch DSB150030

* Data Set Chosen :-

1. Atheism(Class 1)
2. Motorcycles (Class 2)
3. For sale (Class 3)
4. Sports hockey (Class 4)
5. Electronics (Class 5)

* After removing duplicate file names :-

1. Total number of training data – 2388.
2. Total number of test data – 1381.

* Attached in the Zip is Data separated in Test and Training folders, however, the code uses consolidated data which is in **CommonData Folder.**
* **CommonLabels** is the label/index file for the CommonData Folder.
* Classifiers choosed :-

1. SVM
2. Neural Network
3. Maximum Entropy
4. Decision Tree
5. Boosting

* **Result Summary :-**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Accuracy** | **Precision** | **Recall** | **Fscore** |
| **SVM** | 0.87693 | 0.9 | 0.882 | 0.89 |
| **Boosting** | 0.833577 | 0.846 | 0.804 | 0.812 |
| **Neural Network** | 0.7389 | 0.872 | 0.82 | 0.836 |
| **Decision Tree** | 0.7314 | 0.624 | 0.634 | 0.606 |
| **Max Entropy** | 0.9833 | 0.916 | 0.904 | 0.908 |

* **R Code Result :-**

> source('C:/Users/Dharmam/Desktop/Machine Learning/Assignment 4/Data/Assignment4\_RCode.R')

Loading required package: RTextTools

Loading required package: SparseM

Attaching package: ‘SparseM’

The following object is masked from ‘package:base’:

backsolve

[1] "Data Read."

[1] "Container Ready."

[1] "Model Creation started."

[1] "SVM Model Ready."

[1] "NN Model Ready."

[1] "MaxEnt Model Ready."

[1] "Boosting Model Ready."

[1] "DT Model Ready."

[1] "Model Creation Completed, Predictiction Started."

[1] "SVM Predcition Done."

[1] "NN Predcition Done."

[1] "MaxEnt Predcition Done."

[1] "Boosting Predcition Done."

[1] "DT Predcition Done."

[1] "Predictiction Completed, Analysis of results Started."

ENSEMBLE SUMMARY

n-ENSEMBLE COVERAGE n-ENSEMBLE RECALL

n >= 1 1.00 0.92

n >= 2 1.00 0.92

n >= 3 0.96 0.95

n >= 4 0.86 0.97

n >= 5 0.68 0.99

ALGORITHM PERFORMANCE

SVM\_PRECISION SVM\_RECALL SVM\_FSCORE LOGITBOOST\_PRECISION LOGITBOOST\_RECALL LOGITBOOST\_FSCORE

0.900 0.882 0.890 0.846 0.804 0.812

GLMNET\_PRECISION GLMNET\_RECALL GLMNET\_FSCORE TREE\_PRECISION TREE\_RECALL TREE\_FSCORE

0.872 0.820 0.836 0.624 0.634 0.606

MAXENTROPY\_PRECISION MAXENTROPY\_RECALL MAXENTROPY\_FSCORE

0.916 0.904 0.908