# Lab 4

In this project, 7 classifiers were used as an Ensemble model. These are Random Forest, AdaBoost, Gaussian Naïve Bayes, Feed forward Neural Net, K-Nearest neighbors, Logistic Regression, and Decision Tree. Individual accuracies of all the models are between 74% and 82% when using my own data split. Accuracies mostly increase when using the recommended train/test csv files. The Ensemble over all models tops that by having an accuracy of 82.67%.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Random  Forest | AdaBoost | Gaussian  NB | Neural  Network | K-Nearest  Neighbours | Logistic  Regression | Decision Tree |
| Using my own stratified data split | | | | | | | |
| % | 82.00% | 79.33% | 74.67% | 80.00% | 78.67% | 78.67% | 77.33% |
| M | 106 81  9 17 | 114 0 31 5 | 105 9 29 7 | 107 7 23 13 | 111 3 29 7 | 112 2  30 6 | 102 1222 14 |
| Using recommended train.csv and test.csv | | | | | | | |
| % | 83.06% | 80.40% | 80.07% | 81.40% | 78.41% | 80.73% | 81.40% |
| M | 228 10 41 22 | 236 2 57 6 | 226 12 48 15 | 231 7 49 14 | 229 9 56 7 | 232 6 52 11 | 221 1739 24 |

## Hyper parameters:

RandomForestClassifier(n\_jobs=2,random\_state=68,n\_estimators=30,max\_depth=5)

AdaBoostClassifier(n\_estimators=80,learning\_rate=0.1,algorithm='SAMME.R')

GaussianNB()

MLPClassifier(solver='lbfgs', alpha=1e-4,hidden\_layer\_sizes=(6, 9), random\_state=1)

KNeighborsClassifier(n\_neighbors=9)

LogisticRegression(random\_state=0, solver='lbfgs',multi\_class='multinomial')

DecisionTreeClassifier(max\_depth=3)

## Ensemble experiments:

### On the data I split myself:

Using the ensemble classifier on the 5 models, I get a consistent accuracy of just 80.67%. For some reason, this value doesn’t change whether I apply proportional weights or use the accuracies themselves as weights, or use equal weights (unweighted).

For the 7 models however, weighted ensemble gives me an accuracy of 82.67%, larger than all the models I have. I believe this means the more models I have the better my ensemble performs.

### Using recommended data split

Using the ensemble classifier on the 5 models, I get a an accrucay of 81.40% when using equal weights (unweighted). Weighted ensemble gives me an accuracy of 83.06%. This is better than all the five models used in the ensemble.

For the 7 classifiers, unweighted ensemble has an accuracy of 80.73%, and the weighted one out weighs all individual classifiers used with 83.39%.

## Conclusion

From the above experiments I notice that weighted ensembles are capable of increasing the accuracy of a bunch of weaker models. Furthermore, the recommended (provided) data split performed better with most models.