This program requires python 3.4 with jupyter :

Using python program with [scikit-learn](https://mail.python.org/mailman/listinfo/scikit-learn) library to do this problem and write details comments for each steps:

I would like the code be more understandable. I do not want complex code

**Part 1**

I have file call “dibates3.csv” I want to use 10 fold cross validation to find result for model by applying each one of this algorithm:

1. I want to apply libsvm and kernel is RBF and gamma (g) is 0.005
2. I want to apply random forest
3. I want to apply nearest neighbor
4. I want to apply naïve Bayes

**Show me result for each one of these algorithms with confusion matrix**

**I want you to call built function for each one of these measurements from library so there will be no incorrect during multiplication.**

# Calculate accuracy

from sklearn import metrics

print((TP + TN) / float(TP + TN + FP + FN))

print(metrics.accuracy\_score(y\_test, y\_pred\_class))

#calculate **Sensitivity**

print(TP / float(TP + FN))

print(metrics.recall\_score(y\_test, y\_pred\_class))

#calculate specificity

print(TN / float(TN + FP))

#calculate precision

print(TP / float(TP + FP))

print(metrics.precision\_score(y\_test, y\_pred\_class))

# calculate f1 score

F1 = 2 \* (precision \* recall) / (precision + recall)

metrics.f1\_score(y\_test,y\_pred\_class, labels=None, pos\_label=1, average='binary', sample\_weight=None)

#calculate matthews correlation

from sklearn.metrics import matthews\_corrcoef

matthews\_corrcoef(y\_test,y\_pred\_class)

result should display as the following table:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Accuracy | Sensitivity | Specificity | Precision | Matthew’s correlation | F1 score |
| Mean ± standard variance | Mean± standard variance | Mean ± standard variance | Mean± standard variance | Mean ± standard variance | Mean ± standard variance |

I used Weka Explorer with 10 fold cross validation and I got result as show for each algorithms. The python result should be similar to Weka results for each algorithm:

Result for LibsvM in Weka

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Accuracy | Sensitivity | Specificity | Precision | Matthew’s correlation | F1 score |
| 69.7917 % | 0.276 | 0.924 | 0.661 | \_ | 0.389 |

Result for Random Forest in Weka

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Accuracy | Sensitivity | Specificity | Precision | Matthew’s correlation | F1 score |
| 74.349 | 0.571 | 0.836 | 0.651 | 0.447 | 0.608 |

Result for K-nearest neighbor in Weka

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Accuracy | Sensitivity | Specificity | Precision | Matthew’s correlation | F1 score |
| 70.1823 % | 0.53 | 0.794 | 0.58 | 0.331 | 0.554 |

Result for Naïve Bayes in Weka

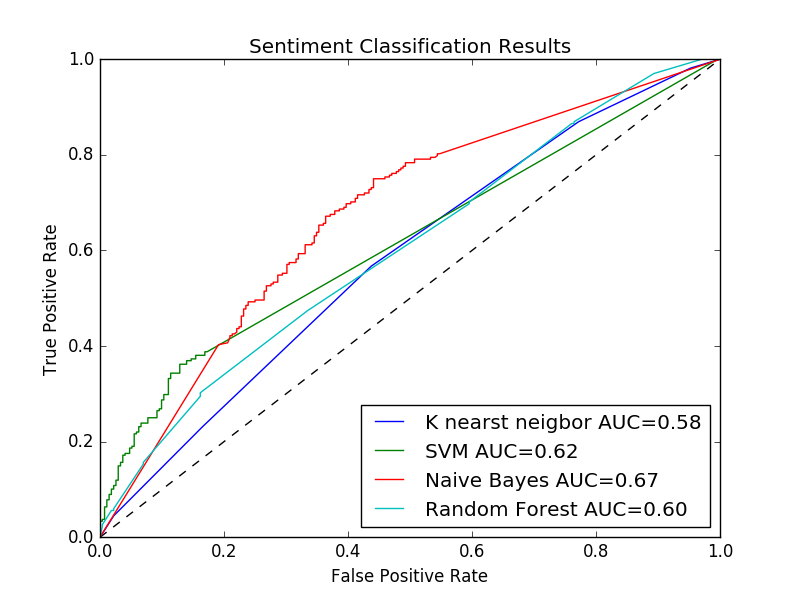
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Accuracy | Sensitivity | Specificity | Precision | Matthew’s correlation | F1 score |
| 76.3021 | 0.612 | 0.844 | 0.678 | 0.468 | 0.643 |

If results are too different from Weka result so there are a problems in the code.

**Part 2**

Now I want to write code to combine four algorithms in ROC curve and show AUC for each algorithm

The curve should be Similar to this Picture:



After you finish send me python code as file py and I also want some details comments on code explain how does it works. I also want you to copy the code in word file with output.

If you have any question please let me know