Machine Learning DV2578 Assignment 2

Sai Teja Palla
20000325-T097
sapb20@student.bth.se

***Index Terms*—Machine Learning,Spam filter,Statistics**

I. Introduction

Our task is to compare models that can distinguish between spam emails and not-spam emails.We have used UCI Machine Learning Repository [2] for the data that has been used for training our model.Our data has 4000 examples of mails that are classified as spam or not-spam mails.We have used Decision Tree, Naive Bayes classifier, Logistic Regression based classifier,algorithms for training our models that classify mails as spam or ham.There are 58 columns total, with columns 1 through 57 representing class properties and column 58 representing the target variable.

II. Concepts And Hypothesis

We obtain classifiers from standard libraries in python like sklearn,numpy and pandas to train models on the dataset. We use 10-fold cross validation sets method for training the classifier.However, in order to implement these models for various use cases, it may be necessary to know their relative performance. Statistical tests are used to discriminate between different models.If the test statistic falls below a certain threshold, one of the hypotheses is rejected; otherwise, the other is accepted.This method is known as hypothesis testing.The test used for distinguishing these models is Friedman test. Measure used for comparing the models are accuracy, training time and f1-score. The following hypothesis for friedman test:

- **Null Hypothesis :** There is no difference in behaviour between three models.
- **Alternative Hypothesis :** There is statistically significant difference in behaviour between the three models.

III. Results

We use measures that evaluates the performance of the models and it is used for comparision.Accuracy,Training Time and F1-Score.

- **Accuracy :**

We have obtained the Friedman Statistic as 15.2.The Threshold value for at alpha = 0.05 level is 7.8 [1]. Since our statistic has crossed the threshold value,we should reject the null hypothesis.Since we have rejected the null hypothesis therefore we have to conduct Nemeyi test to analyze these differences on a pairwise basis.The critical difference value from the Nemeyi test is 1.047.Decision Tree and Naive bayes do not perform equally,Logistic regression and Naive Bayes do not perform equally.

- **Training Time :**

We have obtained the Friedman Statistic as 20.The Threshold value for at alpha = 0.05 level is 7.8 [1]. Since our statistic has crossed the threshold value,we should reject the null hypothesis.Since we have rejected the null hypothesis therefore we have to conduct Nemeyi test to analyze these differences on a pairwise basis.The critical difference value from the Nemeyi test is 1.047.Logistic regression and Naive bayes do not perform equally.

- **F-1 Score :**

We have obtained the Friedman Statistic as 15.2.The Threshold value for at alpha = 0.05 level is 7.8 [1]. Since our statistic has crossed the threshold value,we should reject the null hypothesis.Since we have rejected the null hypothesis therefore we have to conduct Nemeyi test to analyze these differences on a pairwise basis.The critical difference value from the Nemeyi test is 1.047.Logistic regression and Naive bayes do not perform equally,Decision Tree and Naive Bayes do not perform equally.

IV. Conclusion

We have obtained values of measures for the selected models.Based on those values we could determine that Logistic regression performed well in terms of F1-score and accuracy,but it took longest to train.Decision tree performed average on all measures compared to other algorithms.Naive Bayes took lowest training time but performed poorly when compared to other algorithms in accuracy and F1-score.

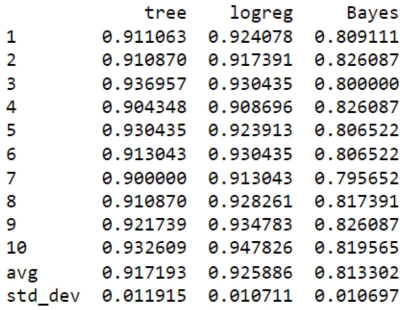
References

[1] Flach, Peter. Machine learning: the art and science of algorithms that make sense of data. Cambridge University Press, 2012

[2] Dua, D. and Graff, C. (2019). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: Univer- sity of California, School of Information and Computer Science.

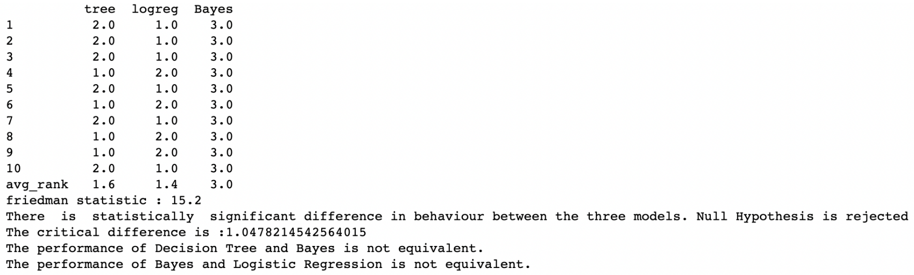
V. Appendix

Accuracy



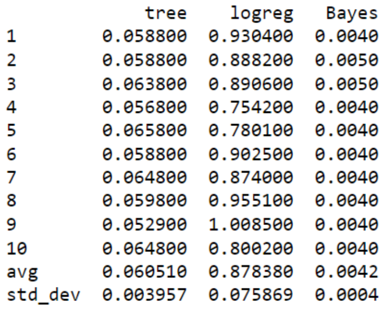
```
              tree    logreg     Bayes
1         0.911063  0.924078  0.809111
2         0.910870  0.917391  0.826087
3         0.936957  0.930435  0.800000
4         0.904348  0.908696  0.826087
5         0.930435  0.923913  0.806522
6         0.913043  0.930435  0.806522
7         0.900000  0.913043  0.795652
8         0.910870  0.928261  0.817391
9         0.921739  0.934783  0.826087
10        0.932609  0.947826  0.819565
avg       0.917193  0.925886  0.813302
std_dev   0.011915  0.010711  0.010697
```

Accuracy Rank Based



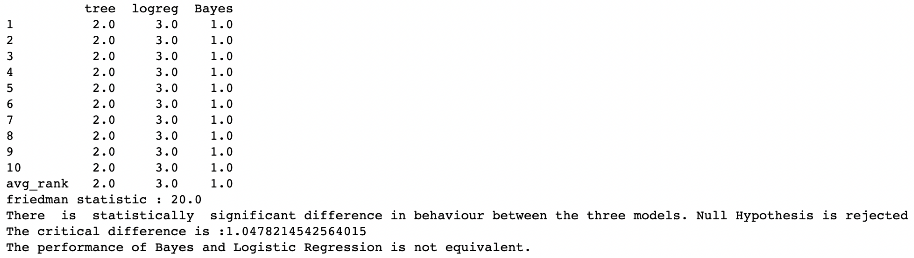
```
           tree  logreg  Bayes
1           2.0     1.0    3.0
2           2.0     1.0    3.0
3           2.0     1.0    3.0
4           1.0     2.0    3.0
5           2.0     1.0    3.0
6           1.0     2.0    3.0
7           2.0     1.0    3.0
8           1.0     2.0    3.0
9           1.0     2.0    3.0
10          2.0     1.0    3.0
avg_rank    1.6     1.4    3.0
friedman statistic : 15.2
There  is  statistically  significant difference in behaviour between the three models. Null Hypothesis is rejected
The critical difference is :1.0478214542564015
The performance of Decision Tree and Bayes is not equivalent.
The performance of Bayes and Logistic Regression is not equivalent.
```

Training Time



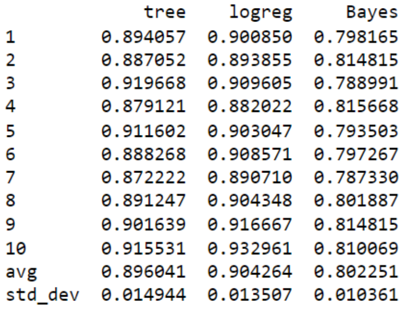
```
              tree    logreg   Bayes
1         0.058800  0.930400  0.0040
2         0.058800  0.888200  0.0050
3         0.063800  0.890600  0.0050
4         0.056800  0.754200  0.0040
5         0.065800  0.780100  0.0040
6         0.058800  0.902500  0.0040
7         0.064800  0.874000  0.0040
8         0.059800  0.955100  0.0040
9         0.052900  1.008500  0.0040
10        0.064800  0.800200  0.0040
avg       0.060510  0.878380  0.0042
std_dev   0.003957  0.075869  0.0004
```

Training Time Rank Based



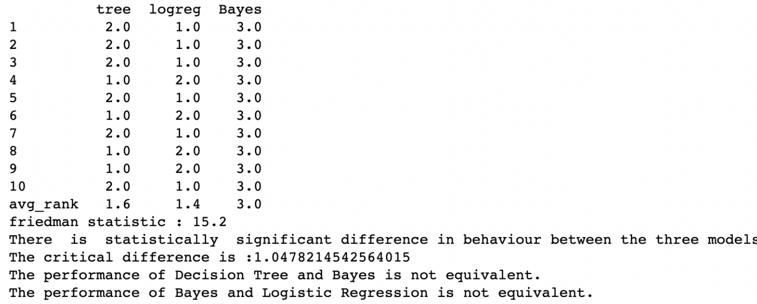
```
           tree  logreg  Bayes
1           2.0     3.0    1.0
2           2.0     3.0    1.0
3           2.0     3.0    1.0
4           2.0     3.0    1.0
5           2.0     3.0    1.0
6           2.0     3.0    1.0
7           2.0     3.0    1.0
8           2.0     3.0    1.0
9           2.0     3.0    1.0
10          2.0     3.0    1.0
avg_rank    2.0     3.0    1.0
friedman statistic : 20.0
There  is  statistically  significant difference in behaviour between the three models. Null Hypothesis is rejected
The critical difference is :1.0478214542564015
The performance of Bayes and Logistic Regression is not equivalent.
```

F1 Score



```
              tree    logreg     Bayes
1         0.894057  0.900850  0.798165
2         0.887052  0.893855  0.814815
3         0.919668  0.909605  0.788991
4         0.879121  0.882022  0.815668
5         0.911602  0.903047  0.793503
6         0.888268  0.908571  0.797267
7         0.872222  0.890710  0.787330
8         0.891247  0.904348  0.801887
9         0.901639  0.916667  0.814815
10        0.915531  0.932961  0.810069
avg       0.896041  0.904264  0.802251
std_dev   0.014944  0.013507  0.010361
```

F1 Score Rank Based



```
           tree  logreg  Bayes
1           2.0     1.0    3.0
2           2.0     1.0    3.0
3           2.0     1.0    3.0
4           1.0     2.0    3.0
5           2.0     1.0    3.0
6           1.0     2.0    3.0
7           2.0     1.0    3.0
8           1.0     2.0    3.0
9           1.0     2.0    3.0
10          2.0     1.0    3.0
avg_rank    1.6     1.4    3.0
friedman statistic : 15.2
There  is  statistically  significant difference in behaviour between the three models
The critical difference is :1.0478214542564015
The performance of Decision Tree and Bayes is not equivalent.
The performance of Bayes and Logistic Regression is not equivalent.
```