Comparing Computational and Predictive performance of Learning Algorithms

Sruthi Annapureddy
Department of Computer Science
Blekinge Institute of Technology Karlskrona, Sweden.
sran18@student.bth.se

I. Introduction

The main aim is to classify whether the message is spam or ham. We use three supervised classification algorithms to perform the classification. Three algorithms are Logistic Regression, Naïve Bayes, K- nearest neighbors are the best performing algorithms when it comes to supervised classification.

II. Method

The data is divided into 10 equal folds, so each fold has uniform class distribution. And later the considered three models will be trained and tested on each fold. Next step is to calculate and record the training time, accuracy, and f1-score for each fold. Based on the scores in each fold the considered algorithms will be ranked. Friedman's test with a significance level $\alpha = 0.05$ will be performed after the completion of ranking and calculating the average ranks for the training time, accuracy, and f1-score. If the statistics of Friedman's test is less, then the critical value (7.8) then the null hypothesis will be accepted. Here all the selected algorithms perform in a similar way, or else we can reject it. We can use Nemenyi test if the null hypothesis is reflected. This test is to find whether the considered algorithm is performing in a different way compared from others. Here, we calculate the average rank difference in a pair wise between the algorithms. We can say algorithms are performing differently when the average rank difference is greater than the critical value.

.

III. Results

The obtained results are presented in the tables as shown below.

Results for accuracy:

Fold	LR	NB	KNN
1.	0.9305 (1)	0.7915 (2)	0.7592(3)
2.	0.9175 (1)	0.7939(2)	0.7809(3)
3.	0.9175 (1)	0.8091 (2)	0.7852(3)
4.	0.9434 (1)	0.8347 (2)	0.8(3)
5.	0.9326 (1)	0.8282 (2)	0.8195(3)
6.	0.9347 (1)	0.7760(3)	0.8239(2)
7.	0.9521 (1)	0.7782 (3)	0.8086(2)
8.	0.9391 (1)	0.8152 (2)	0.8152(2)
9.	0.8496 (1)	0.6928(3)	0.7145(2)
10.	0.8562 (1)	0.7429(2)	0.7342(3)
Avg	0.9173 (1)	0.7863(2)	0.7841(3)
Avg rank	1	2.5	2.9

Results for f1-score:

Fold	LR	NB	KNN
1.	0.9080 (1)	0.7241 (2)	0.6873 (3)
2.	0.8914 (1)	0.7214(2)	0.7186 (3)
3.	0.8938 (1)	0.7485 (2)	0.7179(3)
4.	0.9277 (1)	0.7877 (2)	0.7325(3)
5.	0.9412 (1)	0.7749 (2)	0.7700(3)
6.	0.9193 (1)	0.7146 (3)	0.784(2)
7.	0.9367 (1)	0.7213 (3)	0.7333 (2)
8.	0.9217 (1)	0.7521 (3)	0.7632 (2)
9.	0.8179 (1)	0.6618 (2)	0.6618 (2)
10.	0.8135 (1)	0.6775 (2)	0.6775 (2)
Avg	0.8944 (1)	0.7284(2)	0.7249(3)
Avg rank	1	2.5	2.8

Results for Training:

Fold	LR	NB	KNN
1.	0.0588 (3)	0.0049 (1)	0.0210 (2)
2.	0.05889 (3)	0.0019(1)	0.0329 (2)
3.	0.1376 (3)	0.0020 (1)	0.0149 (2)
4.	0.0587 (3)	0.0019(1)	0.0120 (2)
5.	0.0987 (3)	0.0020 (1)	0.0139 (2)
6.	0.0658 (3)	0.0021 (1)	0.0199 (2)
7.	0.1156 (3)	0.0030(1)	0.0100 (2)
8.	0.0648 (3)	0.0029 (1)	0.0159 (2)
9.	0.0438 (3)	0.0020 (1)	0.0270 (2)
10.	0.0558 (3)	0.0019 (1)	0.0149 (2)
Avg	0.0758 (3)	0.0025 (1)	0.0182(2)
Avg rank	3	1	2

Friedman's statistic for accuracy is 15.84 (> critical value 7.8). So, we can see the difference way of algorithms. The average ranks between Logistic regression and NB and average rank between LR and KNN is greater than the critical difference 1.047(Namanya test). So, we can say that their performance is not similar.

Friedman's statistic for F1-score is 15.20 (> critical value 7.8). So, we can see the difference way of algorithms. The average ranks difference between Logistic regression and K- nearest neighbors, LR and NB is greater than the critical difference 1.047(Namanya test). So, we can say that their performance is not similar.

Friedman's statistic for training time is 15.84 (> critical value 7.8). So, we can see the difference way of algorithms. The average ranks difference between Logistic regression and NB, LR and KNN is greater than the critical difference 1.047(Namanya test). So, we can say that their performance is not similar.