

THỰC HÀNH LAB06

Họ tên: Nguyễn Mạnh Đức

MSSV: 20521196

Bài 1: Sử dụng lại bộ dữ liệu Human Activity Recognition Using Smartphones Data set

Trung bình 5 lần thực hiện:

```
> human_activity
      KNN      SVM      RF
1 0.972222 0.944444 1.000000
2 0.972222 0.944444 1.000000
3 0.972222 0.944444 1.000000
4 0.972222 0.944444 1.000000
5 0.972222 0.944444 0.972222
> mean_human_activity <- summarise_if(human_activity, is.numeric, mean)
> mean_human_activity
      KNN      SVM      RF
1 0.972222 0.944444 0.994444
```

Đánh giá mô hình:

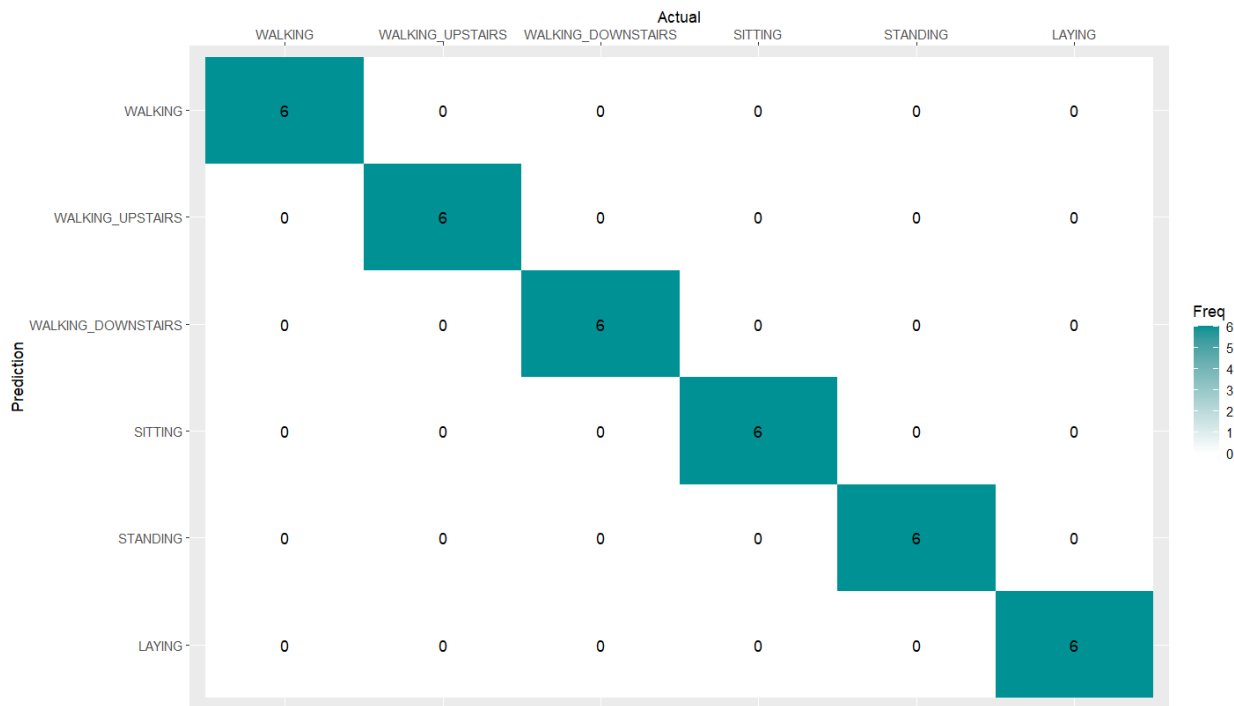
```
Call:
summary.resamples(object = results)

Models: knn, svm, rf
Number of resamples: 10

Accuracy
      Min.    1st Qu.    Median      Mean 3rd Qu.    Max.   NA's
knn 0.9333333 1.0000000 1.0000000 0.9933333      1      1      0
svm 0.7142857 0.8821429 0.9642857 0.9295238      1      1      0
rf  0.9285714 1.0000000 1.0000000 0.9928571      1      1      0

Kappa
      Min.    1st Qu.    Median      Mean 3rd Qu.    Max.   NA's
knn 0.9193548 1.0000000 1.0000000 0.9919355      1      1      0
svm 0.6564417 0.8580742 0.9567901 0.9152443      1      1      0
rf  0.9130435 1.0000000 1.0000000 0.9913043      1      1      0
```

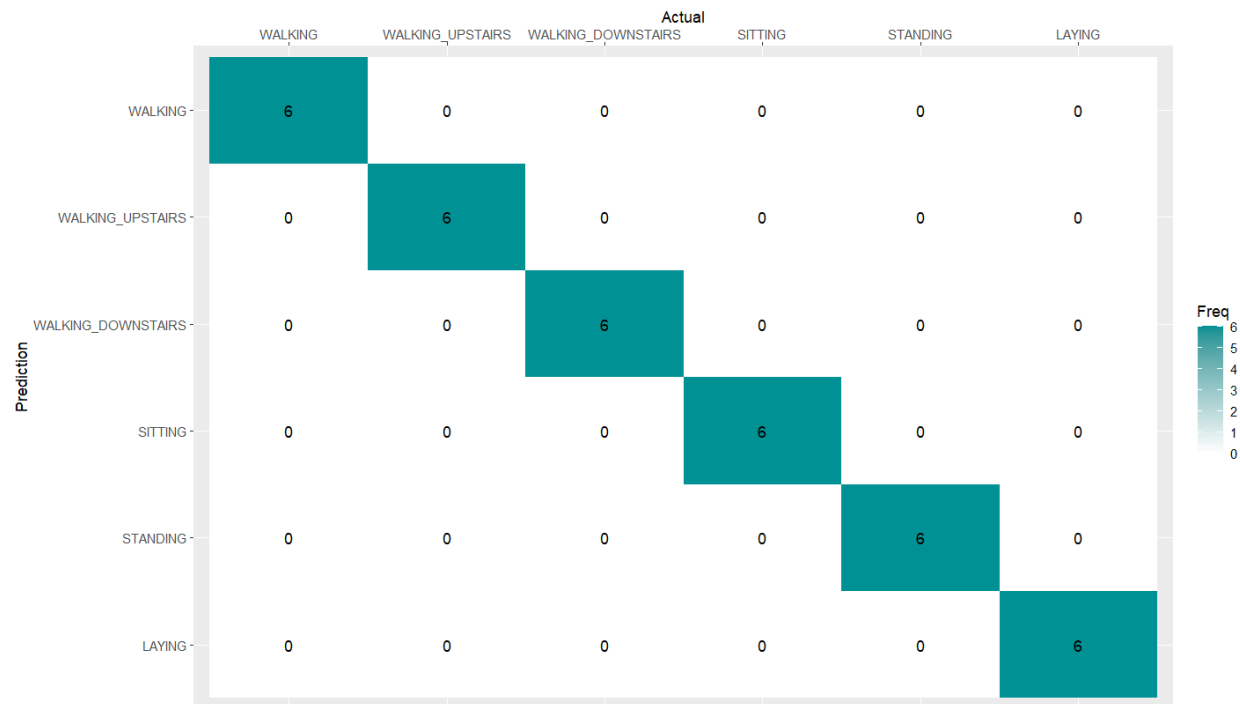
Ma trận nhầm lẫn KNN



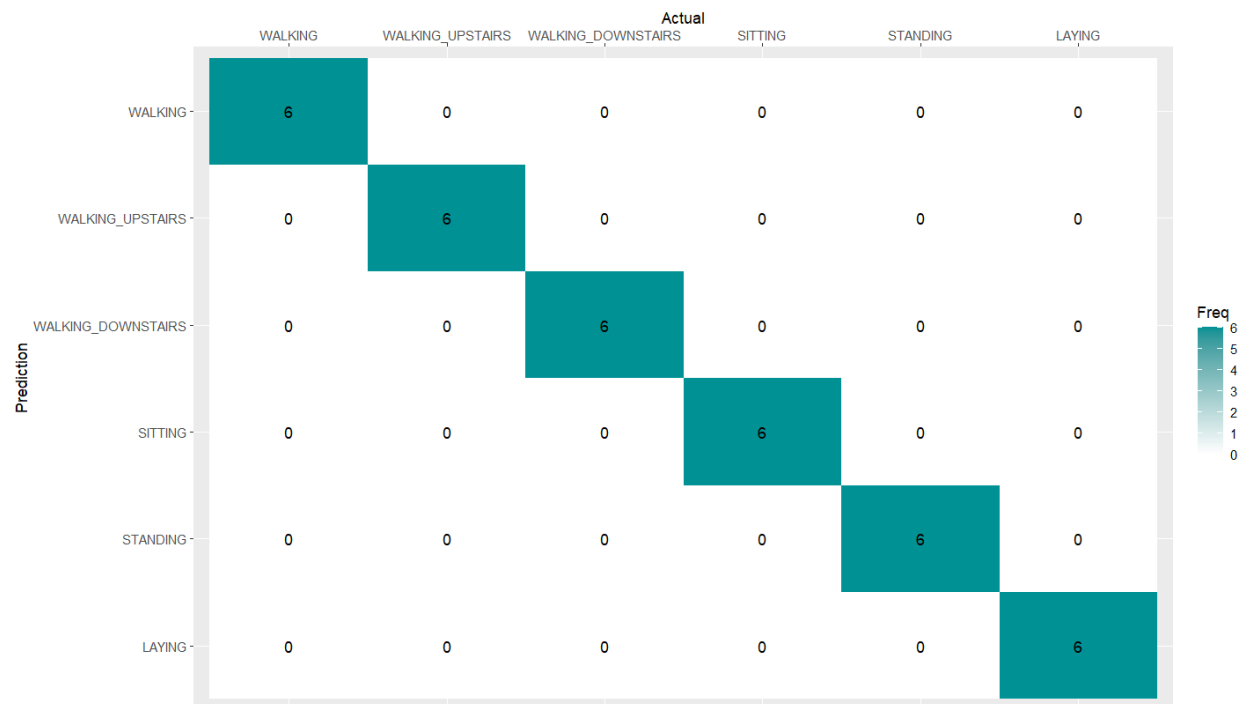
Ma trận nhầm lẫn SVM:



Ma trận nhầm lẫn Random forest:



Mã trận nhầm lẫn LDA:



So sánh:

1) Độ chính xác: Random forest tốt nhất.

- 2) Hiệu quả dự đoán trên từng nhãn: SVM có nhầm lẫn, các mô hình còn lại dự đoán đúng tuyệt đối.
- 3) Tất cả nhãn dự đoán đều tốt.
- 4) Tỷ lệ nhầm lẫn nhãn cao nhất: WALKING_DOWNSTAIRS – Mô hình SVM dự đoán sai 2

Bài 2: Thực hiện các yêu cầu trên với bộ dữ liệu Iris.

Trung bình 5 lần chạy:

```
> Iris
      KNN      SVM      LDA      RF
1 0.9310345 0.9310345 0.9655172 0.9310345
2 0.9310345 0.9310345 0.9655172 0.9310345
3 0.9310345 0.9310345 0.9655172 0.9310345
4 0.9310345 0.9310345 0.9655172 0.9310345
5 0.9310345 0.9310345 0.9655172 0.9310345
> mean_Iris <- summarise_if(Iris, is.numeric, mean)
> mean_Iris
      KNN      SVM      LDA      RF
1 0.9310345 0.9310345 0.9655172 0.9310345
```

So sánh giữa các mô hình:

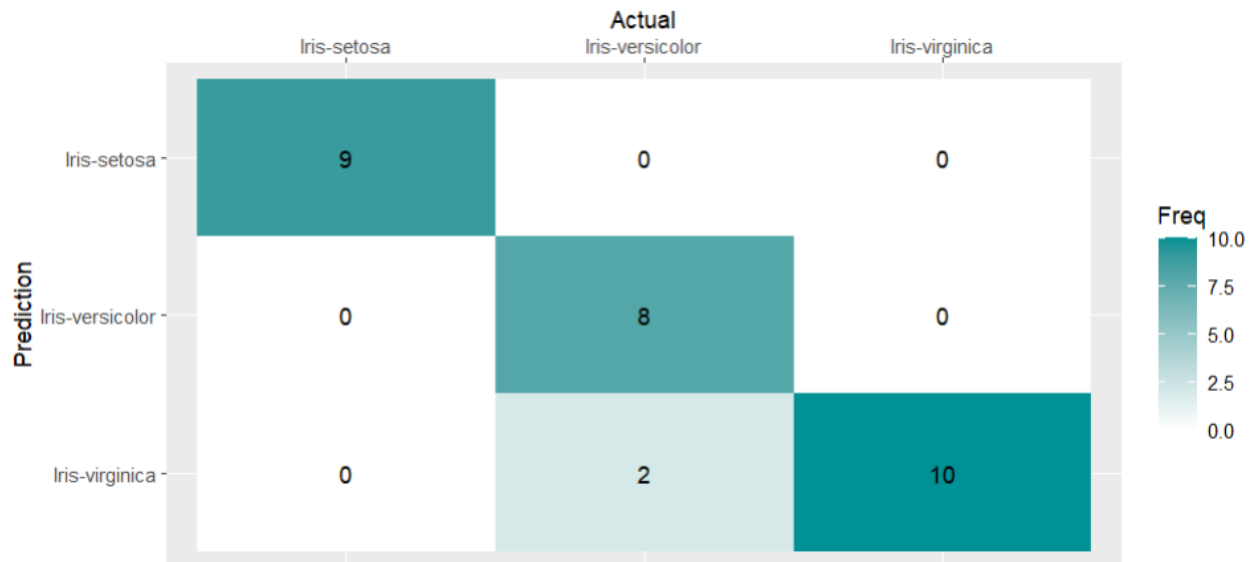
```
Call:
summary.resamples(object = results)

Models: knn, svm, lda, rf
Number of resamples: 10

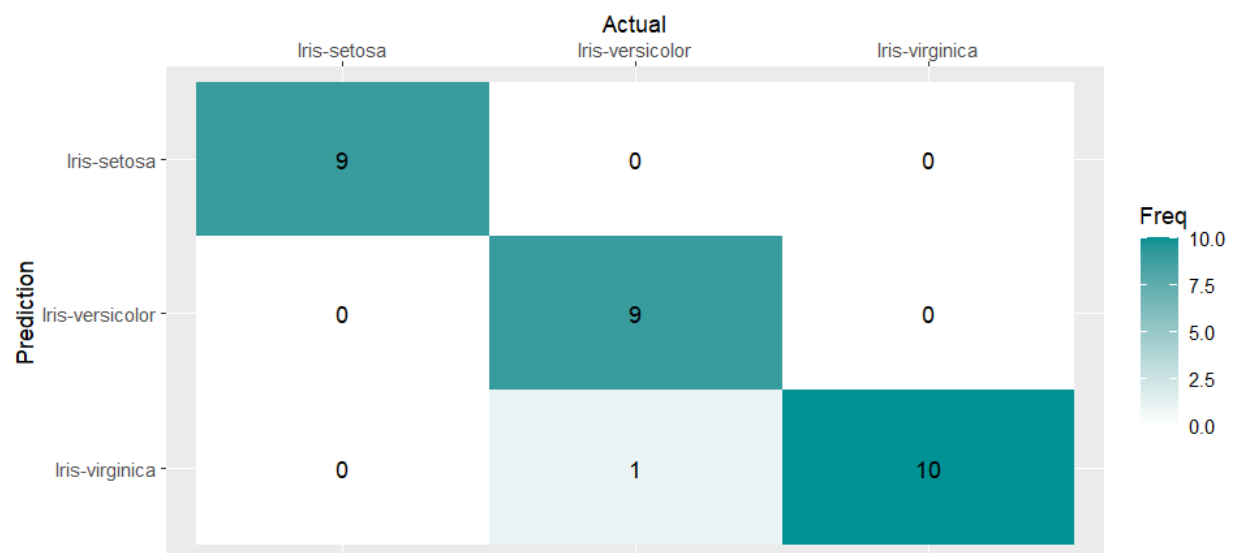
Accuracy
      Min. 1st Qu. Median      Mean 3rd Qu.  Max. NA's
knn 0.9166667 0.9375      1 0.9750000      1      1      0
svm 0.8333333 0.9375      1 0.9583333      1      1      0
lda 0.9166667 1.0000      1 0.9833333      1      1      0
rf  0.8333333 1.0000      1 0.9750000      1      1      0

Kappa
      Min. 1st Qu. Median      Mean 3rd Qu.  Max. NA's
knn 0.875 0.90625      1 0.9625      1      1      0
svm 0.750 0.90625      1 0.9375      1      1      0
lda 0.875 1.00000      1 0.9750      1      1      0
rf  0.750 1.00000      1 0.9625      1      1      0
```

Ma trận nhầm lẫn của KNN và SVM và Random forest (giống nhau)



Ma trận nhầm lẫn LDA:



Nhận xét: mô hình LDA hoạt động tốt nhất trên bộ dữ liệu Iris

Hiệu quả dự đoán trên từng nhãn của 3 mô hình KNN, SVM và Random forest ngang nhau, kém hơn 1 chút so với LDA

Tỉ lệ nhầm lẫn nhãn cao nhất là Iris-versicolor

Bài 3: Thực hiện các yêu cầu trên với bộ dữ liệu WineQuality

Trung bình 5 lần chạy:

Ma trận nhầm lẫn KNN

Ma trận nhầm lẫn SVM

Ma trận nhầm lẫn LDA

Ma trận nhầm lẫn Random forest

Nhận xét: mô hình hoạt động tốt nhất trên bộ dữ liệu winequality

Hiệu quả dự đoán trên từng nhãn