Classificador de smartphone

Parte 2 - Lucas Silva de Mendonça (lsm5)

ORGANIZAÇÃO DO CÓDIGO

- BeatifulSoup (remoção das tags HTML)
- Normalização de caracteres (NFKD)
- Remoção de caracteres
  - Símbolos
  - Números
  - > 3
  - Stopwords Br
  - Stopwords En
- Bag of Words
  - Random Forest (n_estimators=100)
  - Feature Selection (>= 0.005)
- Modelos de classificação

CLASSIFICADORES

Naive Bayes

```
              precision    recall  f1-score   support

         0       0.93      0.77      0.84        35
         1       0.84      0.96      0.90        45

  accuracy                          0.88        80
 macro avg       0.89      0.86      0.87        80
weighted avg     0.88      0.88      0.87        80

acurácia de: 0.875
```

```
CPU times: user 4.72 ms, sys: 1.02 ms, total: 5.73 ms
Wall time: 5.37 ms
GaussianNB()
```

Naive Bayes - Hiperparâmetros
var_smoothing=0.01

```
              precision    recall  f1-score   support

         0       0.83      0.83      0.83        35
         1       0.87      0.87      0.87        45

  accuracy                          0.85        80
 macro avg       0.85      0.85      0.85        80
weighted avg     0.85      0.85      0.85        80

acurácia de: 0.85
```

```
CPU times: user 6.5 ms, sys: 973 µs, total: 7.47 ms
Wall time: 13.6 ms
GaussianNB(var_smoothing=0.01)
```

CLASSIFICADORES

Decision Tree

	precision	recall	f1-score	support
0	0.91	0.89	0.90	35
1	0.91	0.93	0.92	45
accuracy			0.91	80
macro avg	0.91	0.91	0.91	80
weighted avg	0.91	0.91	0.91	80

acurácia de: 0.9125

```
CPU times: user 7.61 ms, sys: 38 µs, total: 7.65 ms
Wall time: 10.1 ms
```

Decision Tree - Hiperparâmetros
criterion = "entropy", splitter = "random", max_depth = 10

	precision	recall	f1-score	support
0	0.91	0.89	0.90	35
1	0.91	0.93	0.92	45
accuracy			0.91	80
macro avg	0.91	0.91	0.91	80
weighted avg	0.91	0.91	0.91	80

acurácia de: 0.9125

```
CPU times: user 3.77 ms, sys: 0 ns, total: 3.77 ms
Wall time: 3.81 ms
```

CLASSIFICADORES

Decision Tree - Hiperparâmetros
criterion = "gini", splitter = "random", max_depth = 100

	precision	recall	f1-score	support
0	0.94	0.89	0.91	35
1	0.91	0.96	0.93	45
accuracy			0.93	80
macro avg	0.93	0.92	0.92	80
weighted avg	0.93	0.93	0.92	80

acurácia de: 0.925

```
CPU times: user 4.38 ms, sys: 922 µs, total: 5.3 ms
Wall time: 5.65 ms
```

Decision Tree - Hiperparâmetros
criterion = "gini", splitter = "random", max_depth = 50

	precision	recall	f1-score	support
0	0.97	0.89	0.93	35
1	0.92	0.98	0.95	45
accuracy			0.94	80
macro avg	0.94	0.93	0.94	80
weighted avg	0.94	0.94	0.94	80

acurácia de: 0.9375

```
CPU times: user 3.2 ms, sys: 0 ns, total: 3.2 ms
Wall time: 3.21 ms
```

CLASSIFICADORES

SVM

```
              precision    recall  f1-score   support

           0       0.97      0.89      0.93        35
           1       0.92      0.98      0.95        45

    accuracy                           0.94        80
   macro avg       0.94      0.93      0.94        80
weighted avg       0.94      0.94      0.94        80

acurácia de: 0.9375
```

```
CPU times: user 5.44 ms, sys: 0 ns, total: 5.44 ms
Wall time: 9.02 ms
SVC()
```

SVM - Hiperparâmetros
gamma='auto', kernel="linear", decision_function_shape='ovo'

```
              precision    recall  f1-score   support

           0       1.00      0.83      0.91        35
           1       0.88      1.00      0.94        45

    accuracy                           0.93        80
   macro avg       0.94      0.91      0.92        80
weighted avg       0.93      0.93      0.92        80

acurácia de: 0.925
```

```
CPU times: user 7.01 ms, sys: 0 ns, total: 7.01 ms
Wall time: 7.63 ms
SVC(decision_function_shape='ovo', gamma='auto', kernel='linear')
```

CLASSIFICADORES

Logistic Regression

```
              precision    recall  f1-score   support

           0       0.97      0.83      0.89        35
           1       0.88      0.98      0.93        45

    accuracy                           0.91        80
   macro avg       0.92      0.90      0.91        80
weighted avg       0.92      0.91      0.91        80

acurácia de: 0.9125
```

```
CPU times: user 50.7 ms, sys: 0 ns, total: 50.7 ms
Wall time: 127 ms
```

Logistic Regression - Hiperparâmetros
random_state=42, solver='newton-cg'

```
              precision    recall  f1-score   support

           0       1.00      0.83      0.91        35
           1       0.88      1.00      0.94        45

    accuracy                           0.93        80
   macro avg       0.94      0.91      0.92        80
weighted avg       0.93      0.93      0.92        80

acurácia de: 0.925
```

```
CPU times: user 50 ms, sys: 1.03 ms, total: 51 ms
Wall time: 62.1 ms
```

CLASSIFICADORES

Multilayer Perceptron

```
              precision    recall   f1-score    support

         0       0.97       0.86       0.91         35
         1       0.90       0.98       0.94         45

  accuracy                             0.93         80
 macro avg       0.93       0.92       0.92         80
weighted avg     0.93       0.93       0.92         80

acurácia de: 0.925
```

```
CPU times: user 497 ms, sys: 483 ms, total: 981 ms
Wall time: 1.35 s
```

Multilayer Perceptron - Hiperparâmetros
random_state=1, hidden_layer_sizes=(200,200)

```
              precision    recall   f1-score    support

         0       1.00       0.91       0.96         35
         1       0.94       1.00       0.97         45

  accuracy                             0.96         80
 macro avg       0.97       0.96       0.96         80
weighted avg     0.96       0.96       0.96         80

acurácia de: 0.9625
```

```
CPU times: user 924 ms, sys: 546 ms, total: 1.47 s
Wall time: 764 ms
```