

Activity Recognition

Soluções de mineração de dados



[Link projeto github](#)

Prof.: Leandro Almeida

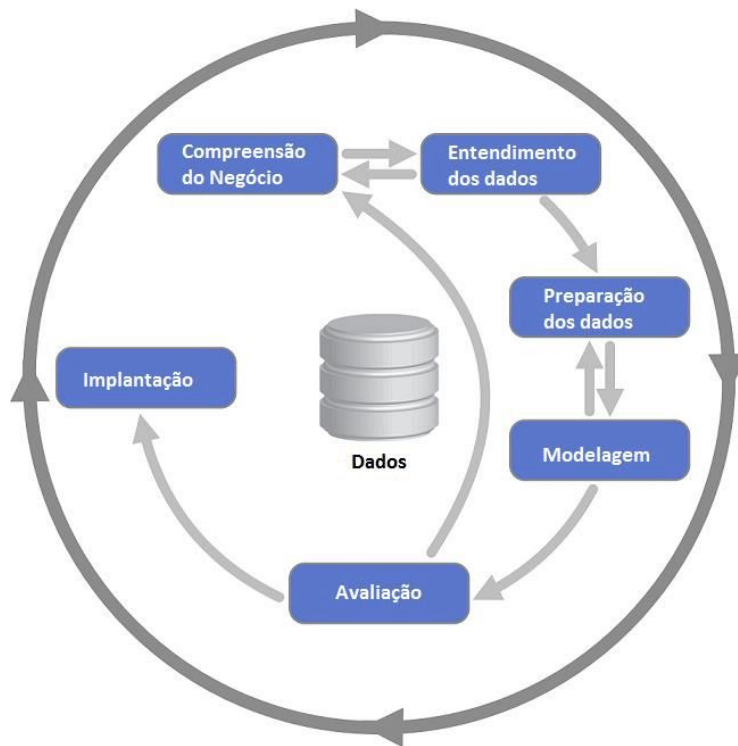
Grupo:

Karl Sousa
kvms@cin.ufpe.br

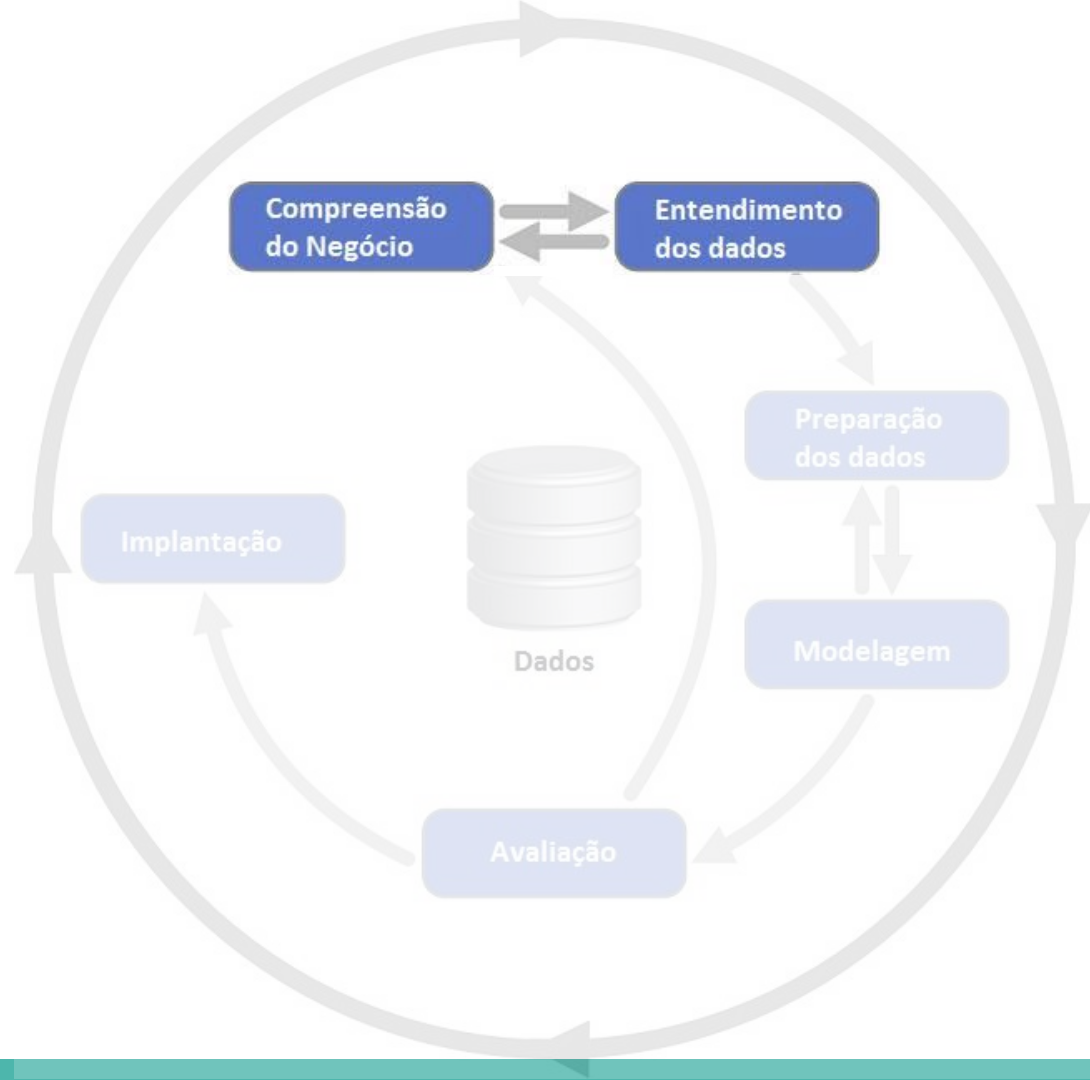
Maria Eugênia
meps@cin.ufpe.br

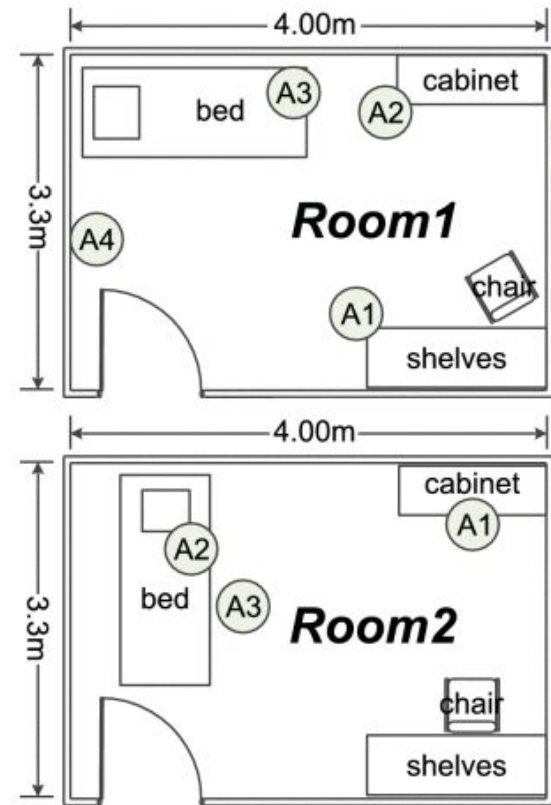
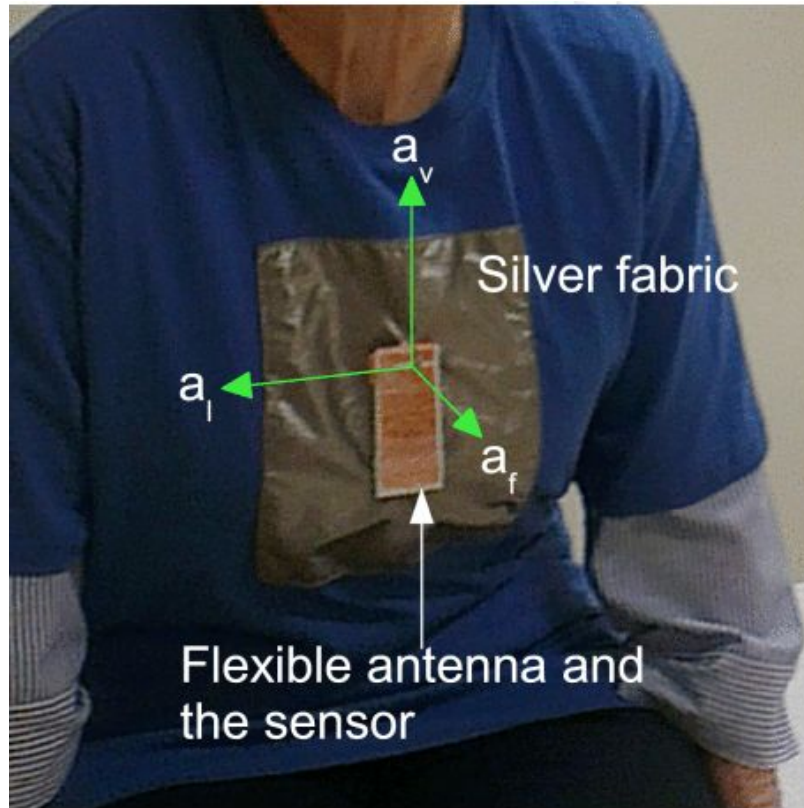
Mateus Silva
mmmps@cin.ufpe.br

CRISP-DM

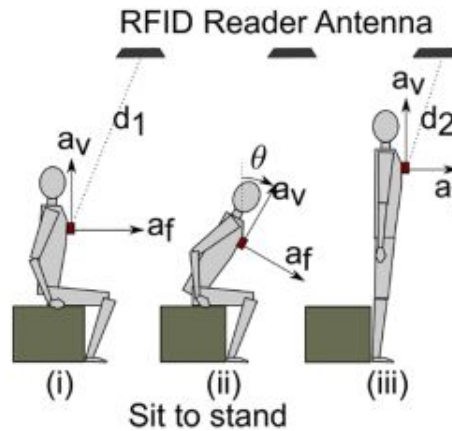
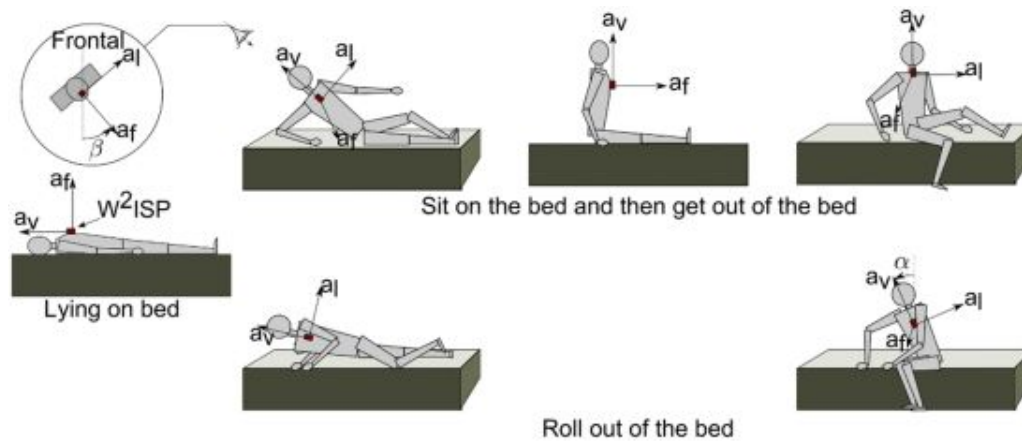


[Link projeto github](#)





A. Wickramasinghe, D. C. Ranasinghe, C. Fumeaux, K. D. Hill, R. Visvanathan, "**Sequence learning with passive RFID sensors for real-time bed-egress recognition in older people**", *IEEE J. Biomed. Health Inform.*, vol. 21, no. 4, pp. 917-929, Jul. 2017.

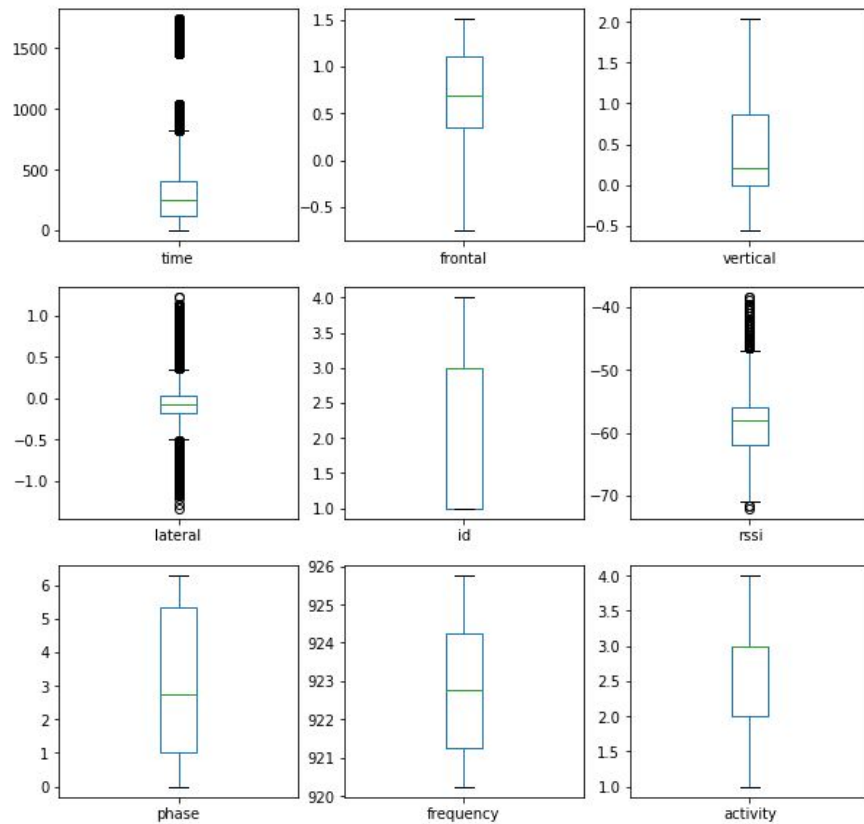


A. Wickramasinghe, D. C. Ranasinghe, C. Fumeaux, K. D. Hill, R. Visvanathan, "**Sequence learning with passive RFID sensors for real-time bed-egress recognition in older people**", *IEEE J. Biomed. Health Inform.*, vol. 21, no. 4, pp. 917-929, Jul. 2017.

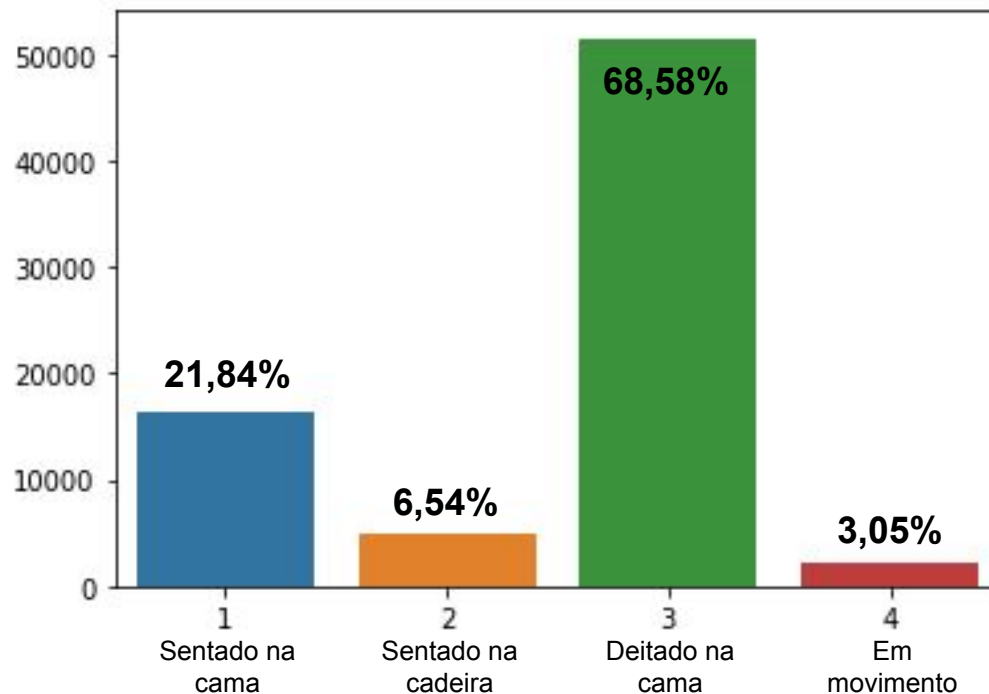
Dados da base

- **Número de exemplos:** 75.128
- **Número de atributos:** 8
- **Número de arquivos:** 60 (S1), 27 (S2)
- **Média de exemplos (linhas) por arquivo:** 863

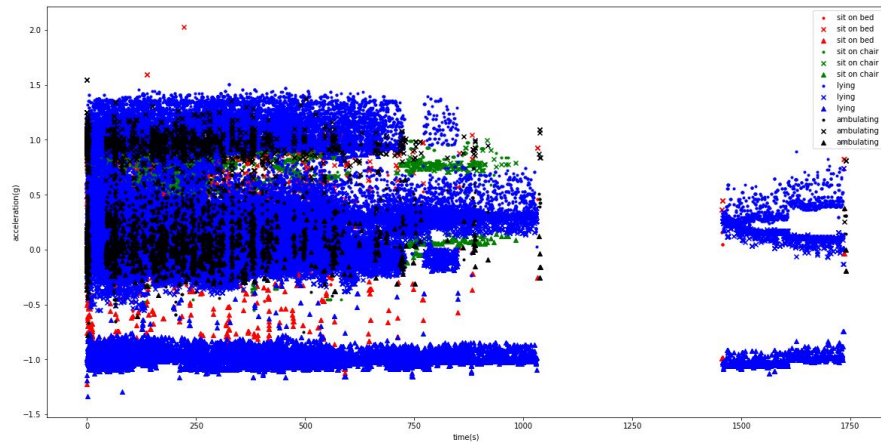
Boxplot de atributos



Distribuição de classes



Acelerações x tempo



Dados de um único ensaio

Dados de todos os exemplos

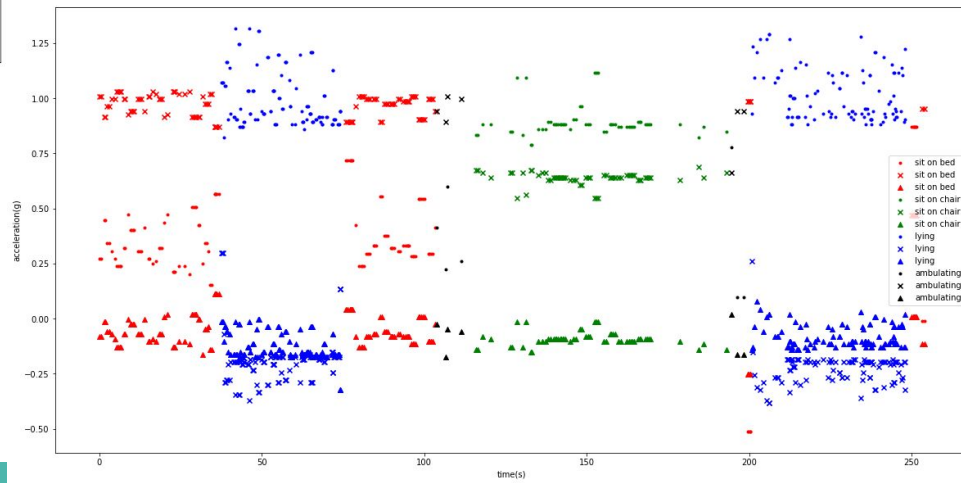
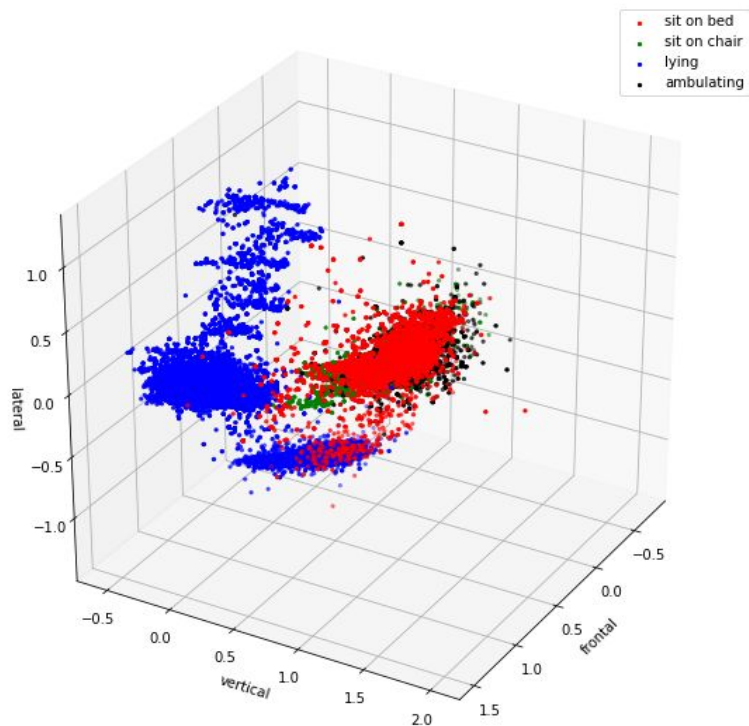
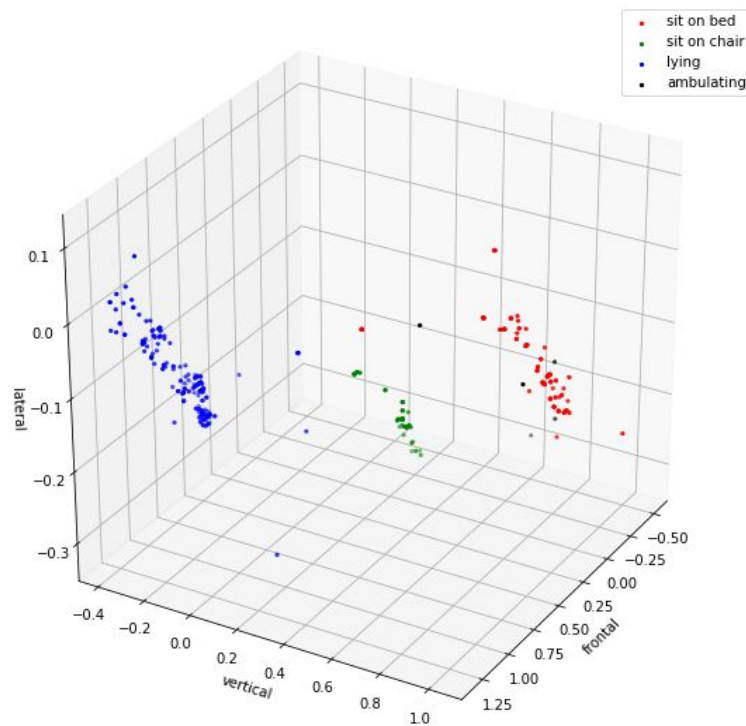


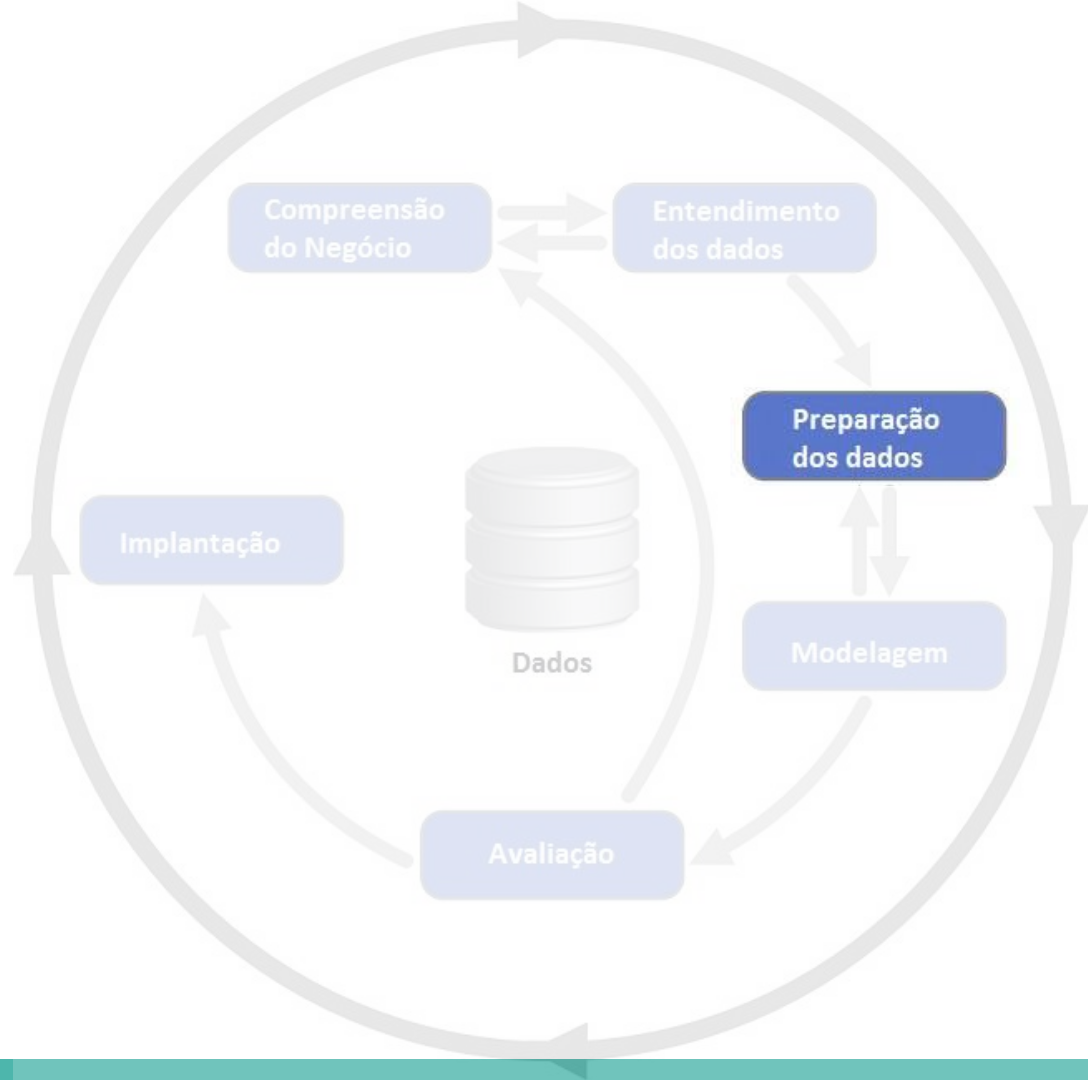
Gráfico das acelerações



Dados de todos os exemplos



Dados de um único ensaio



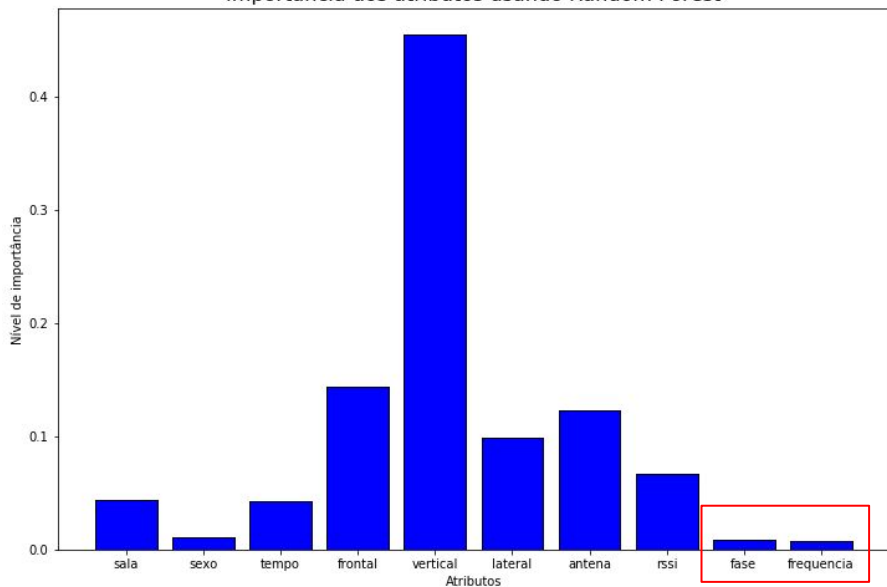
Normalização

MinMaxScaler (Intervalo de 0 a 1)

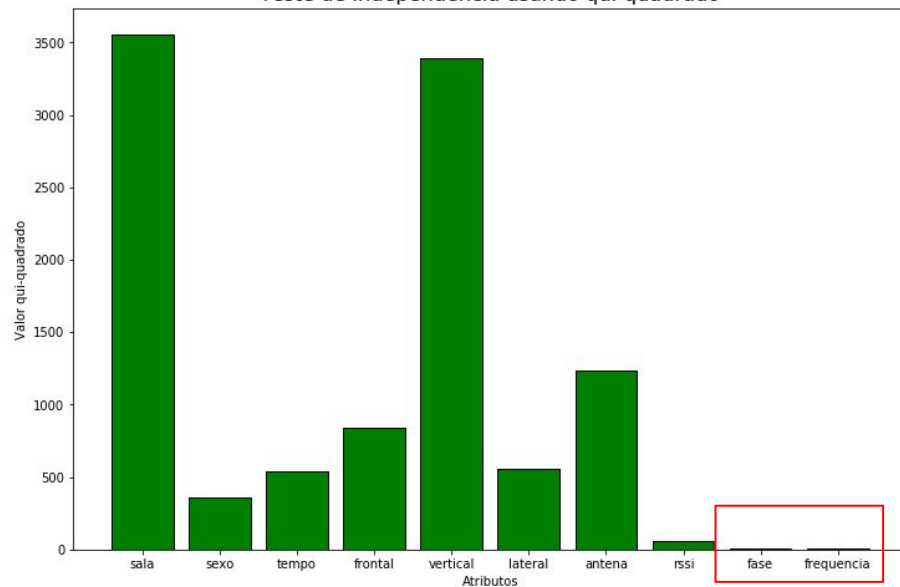
$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Seleção de atributos

Importância dos atributos usando Random Forest



Teste de independência usando qui-quadrado





Visualização dos dados

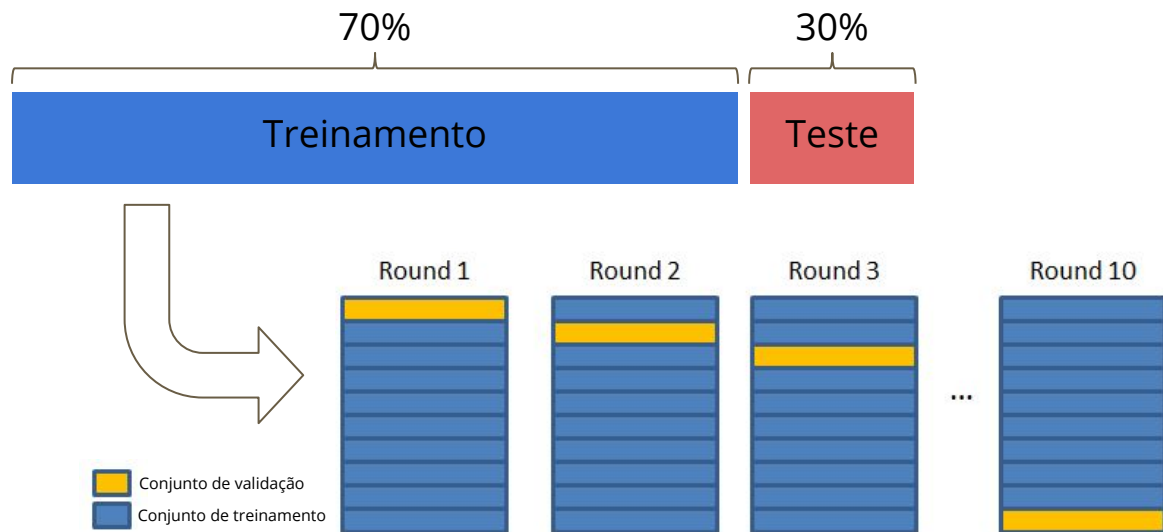
Antes:

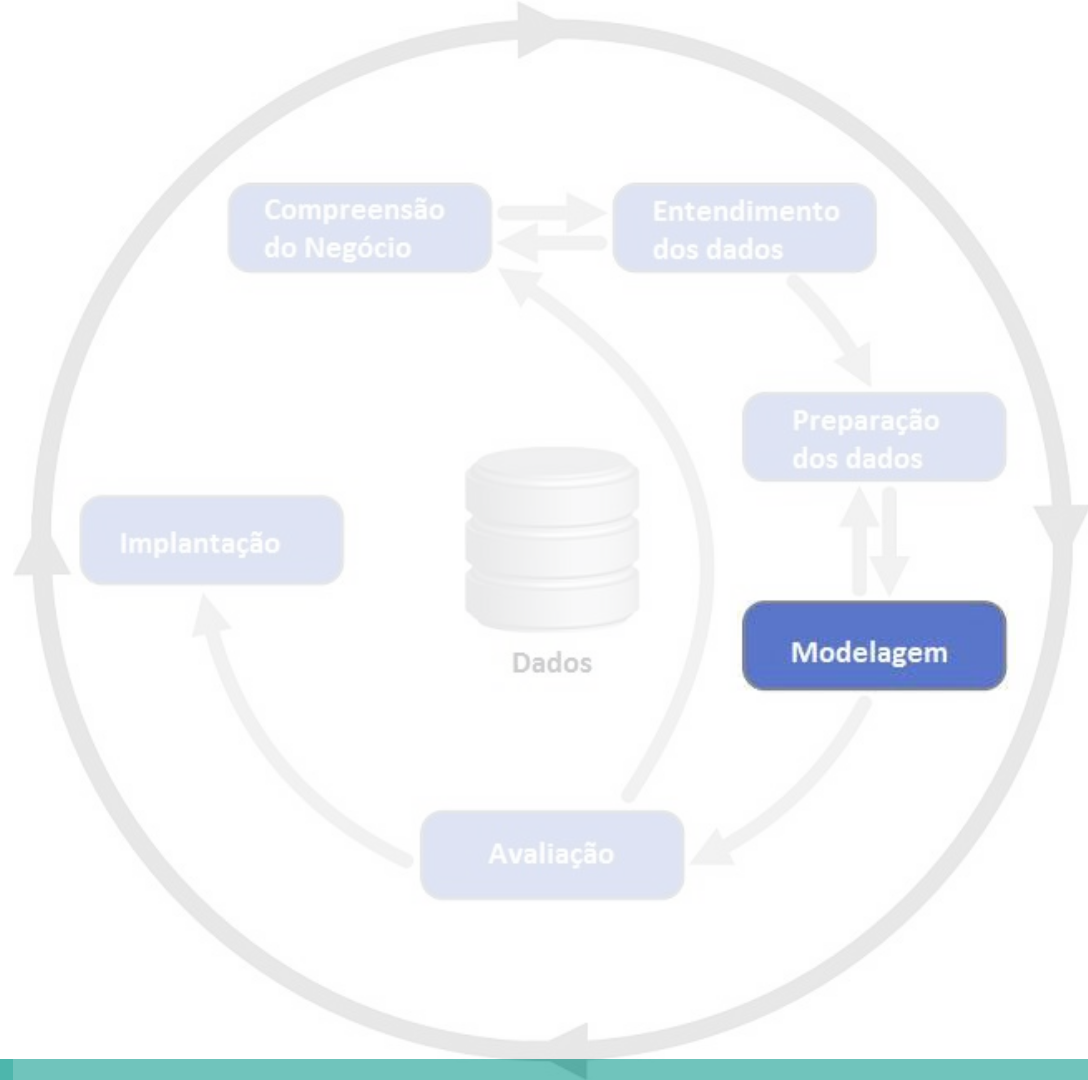
	tempo	frontal	vertical	lateral	antena	rss	 fase	 frequencia	atividade
0	0.00	0.27203	1.00820	-0.082102	1	-63.5	2.4252	924.25	1
1	0.50	0.27203	1.00820	-0.082102	1	-63.0	4.7369	921.75	1
2	1.50	0.44791	0.91636	-0.013684	1	-63.5	3.0311	923.75	1
3	1.75	0.44791	0.91636	-0.013684	1	-63.0	2.0371	921.25	1
4	2.50	0.34238	0.96229	-0.059296	1	-63.5	5.8920	920.25	1

Depois:

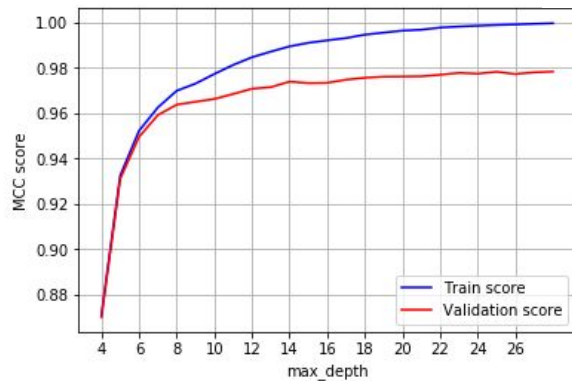
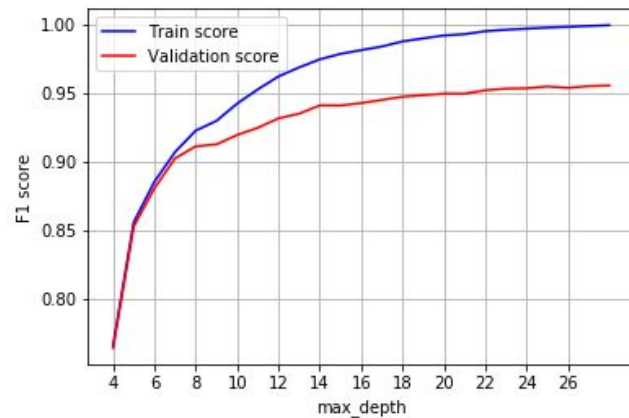
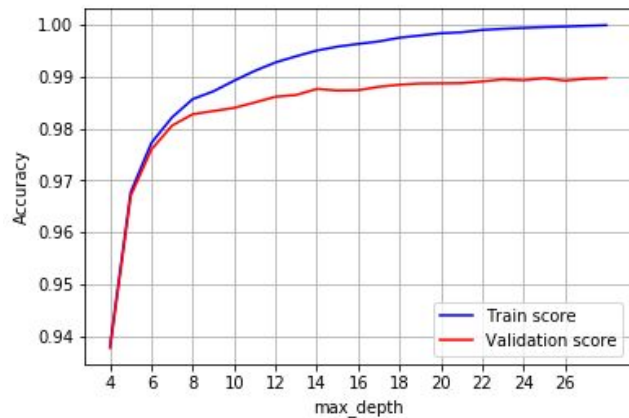
	 sala	 sexo	tempo	frontal	vertical	lateral	antena	rss	atividade
0	0	0	0.160849	0.881062	0.168890	0.481822	0.000000	0.432836	3
1	1	1	0.292244	0.383782	0.324447	0.090887	0.333333	0.701493	3
2	1	1	0.435208	0.427024	0.306670	0.104520	0.333333	0.641791	3
3	0	1	0.023140	0.859441	0.235558	0.704556	0.000000	0.388060	3
4	0	0	0.139703	0.751351	0.235558	0.477276	0.666667	0.656716	3

Separação dos dados

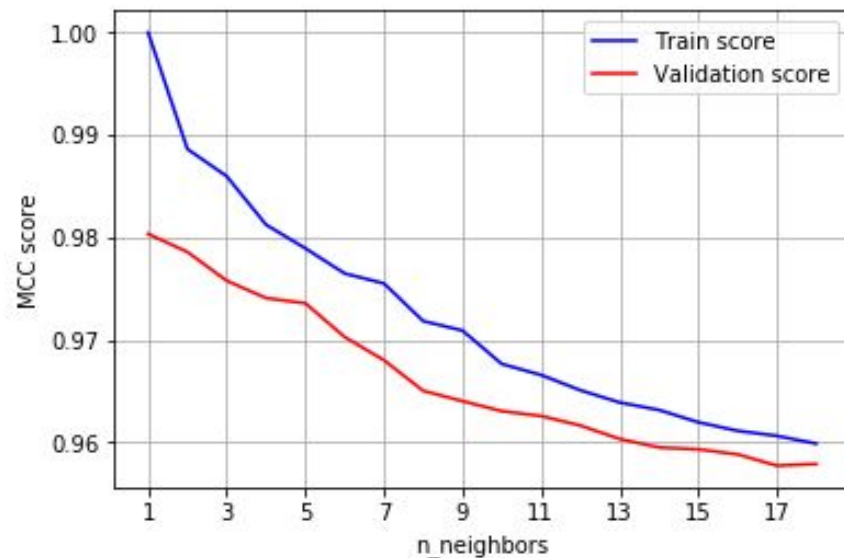
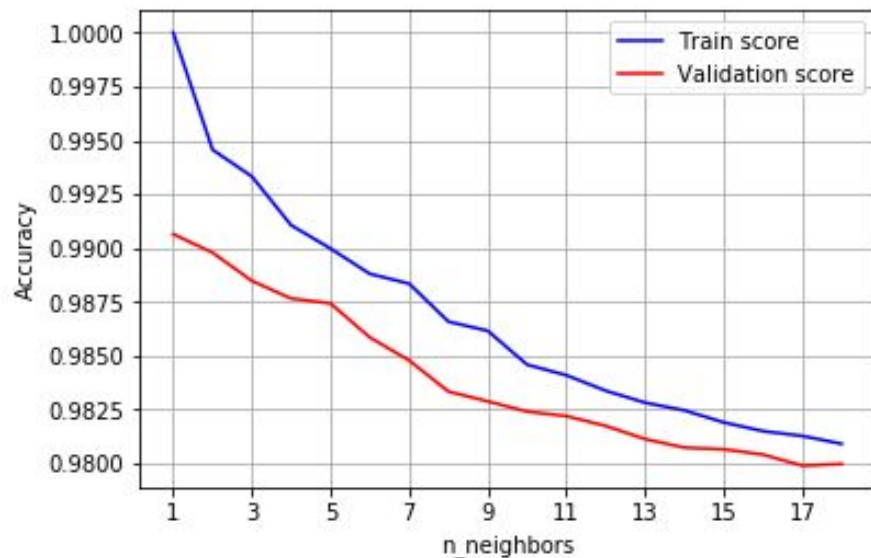




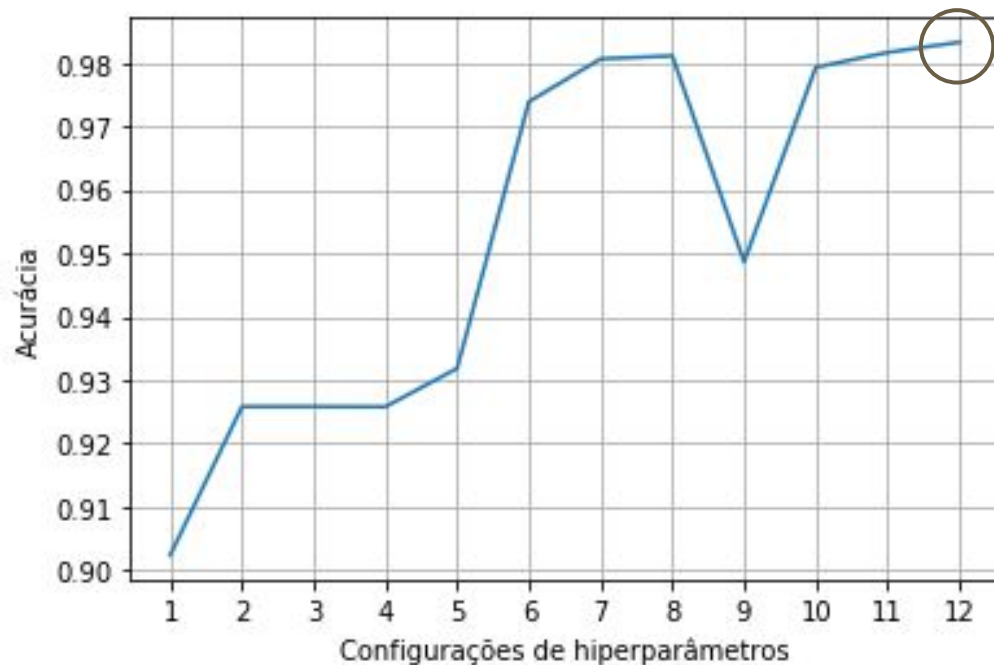
Árvore de decisão



K-Nearest Neighbors

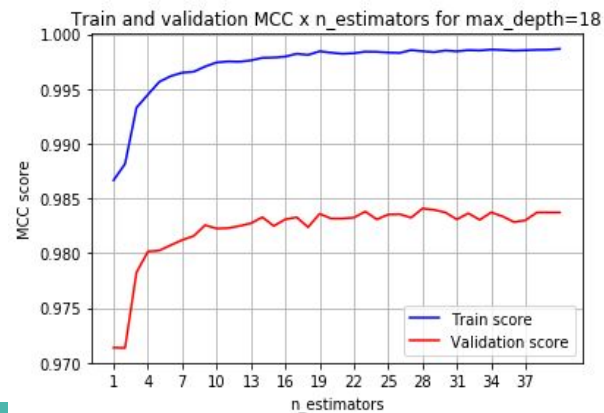
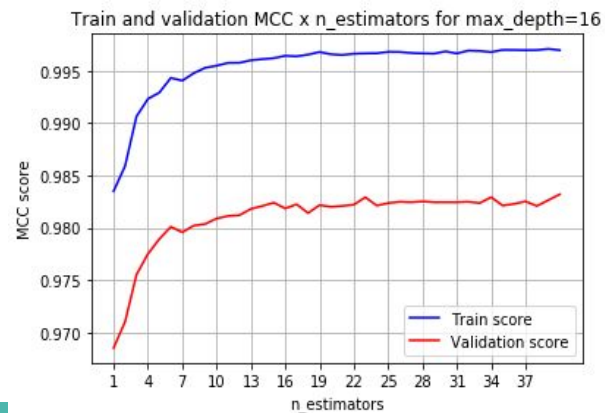
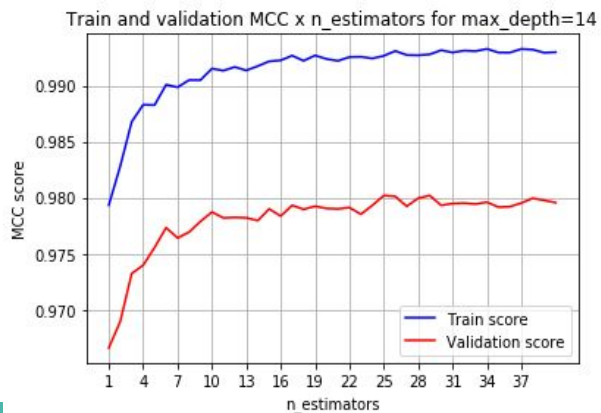
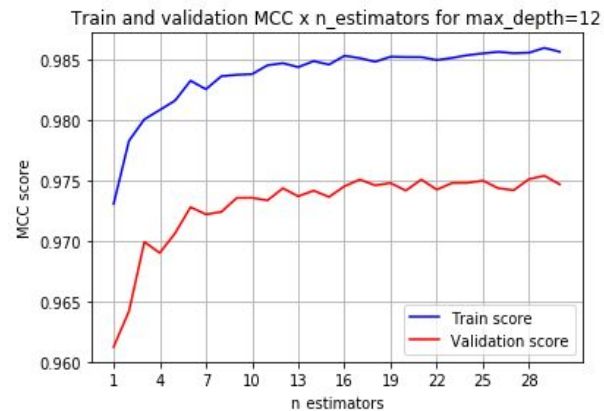
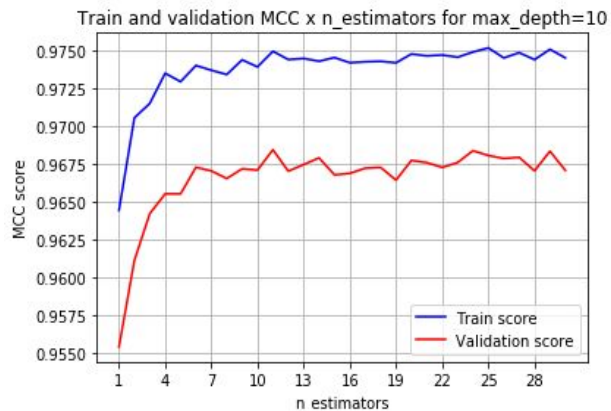
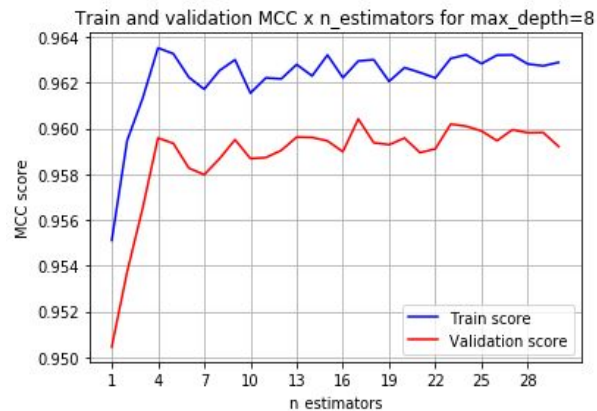


Rede Neural MLP

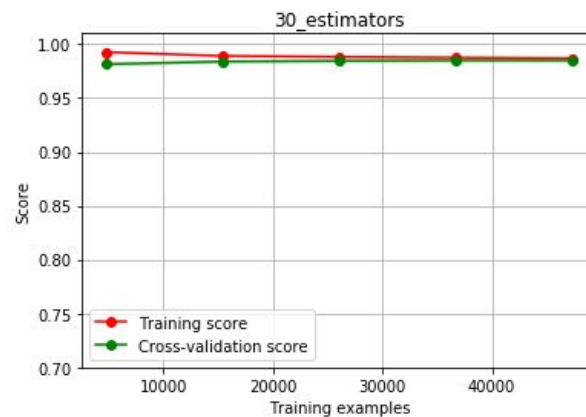
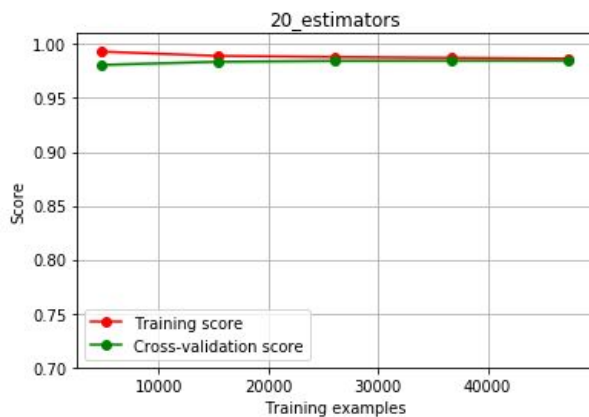
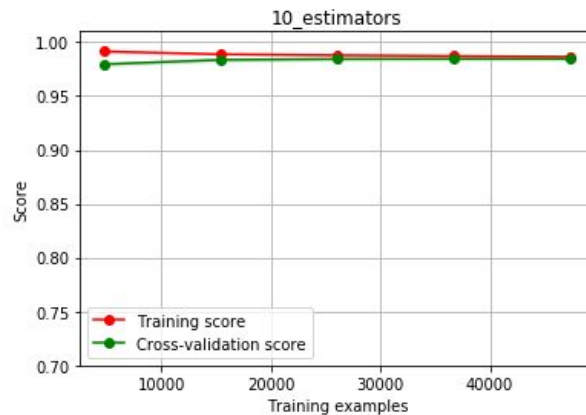
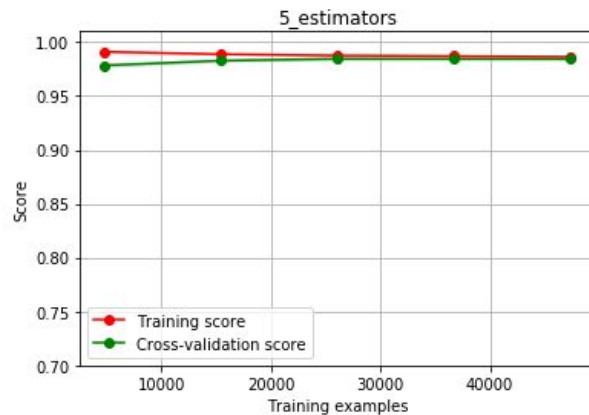


hidden_layer_sizes: (20,20),
activation: "tanh",
max_iter: 400,
solver: "lbfgs",
média da acurácia: 0.9832475

Random Forest



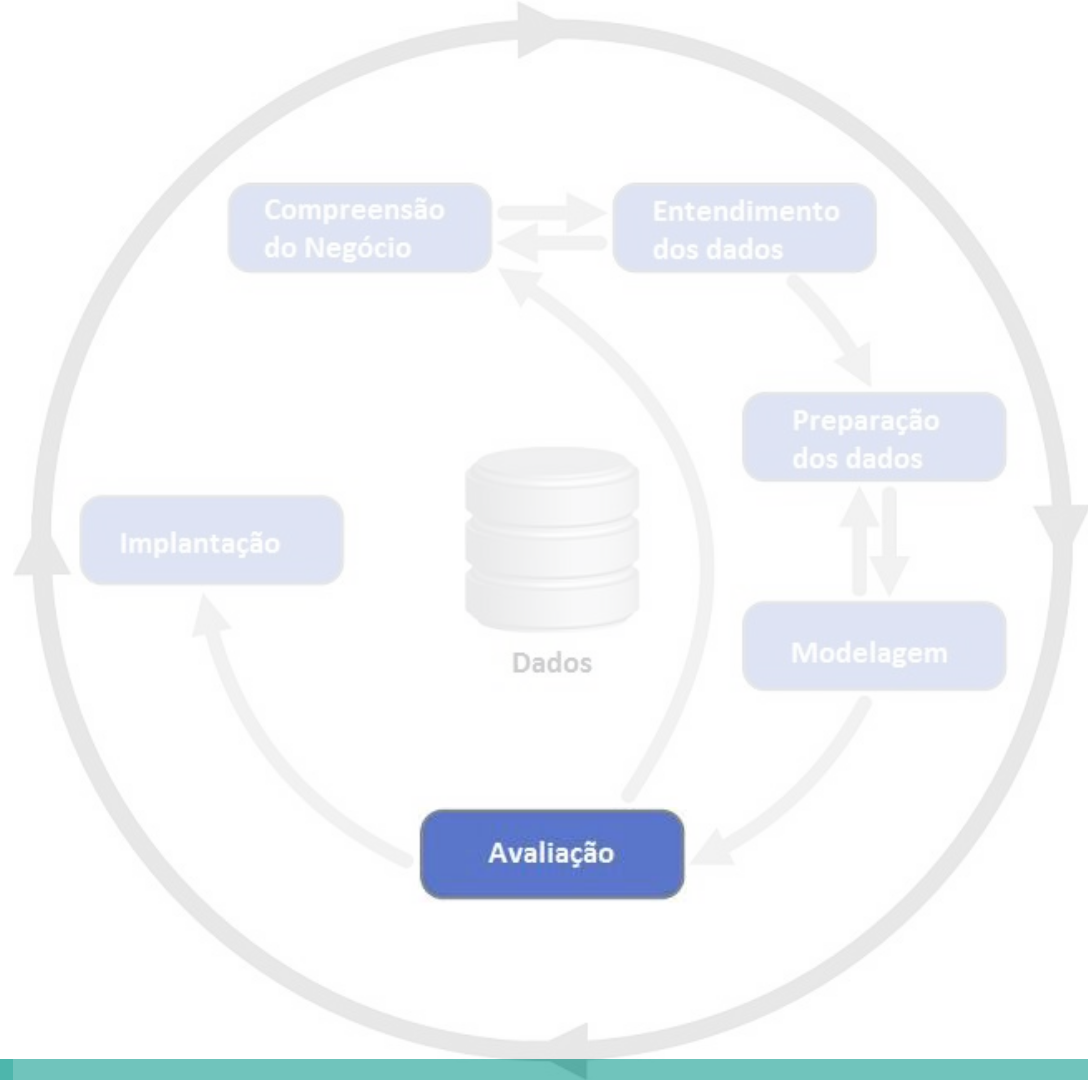
Comité MLP



Comitê Heterogêneo

Voting Classifier (voting: 'soft'):

- **Árvore de decisão** - 0.9717752
- **KNN** - 0.97851257
- **MLP** - 0.96212
- **Comitê Heterogêneo**: 0.976



Comparação de modelos

- **Árvore de decisão**

max_depth=12

- **KNN**

n_neighbors=2

- **MLP**

hidden layers: (20, 20)

activation='tanh'

- **Random Forest**

max_depth=12

n_estimators = 20

- **Comitê MLP**

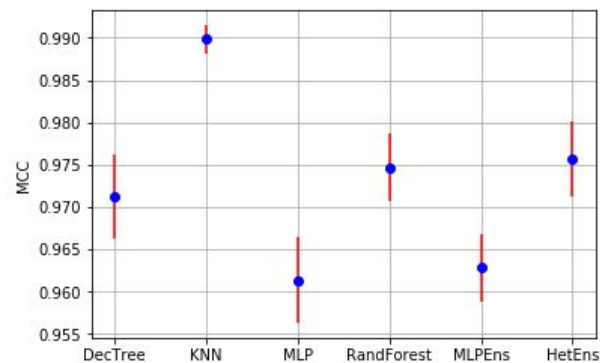
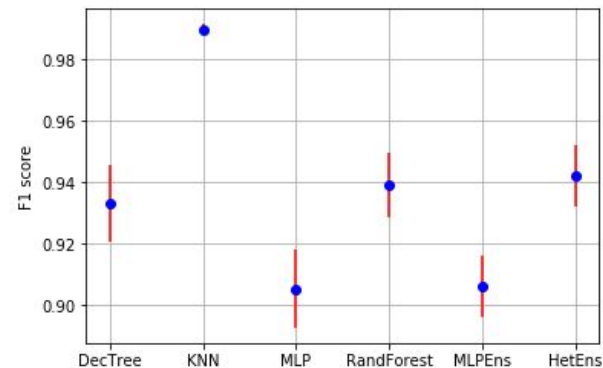
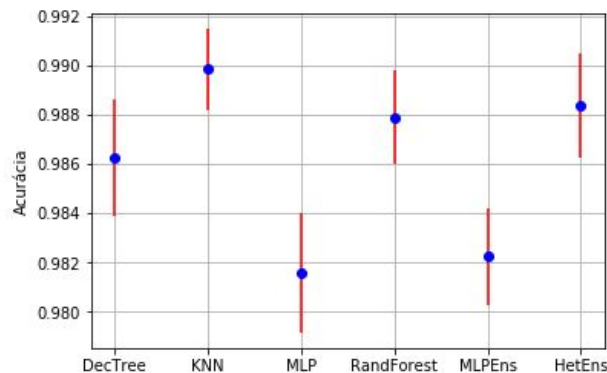
n_estimators = 5

- **Comitê heterogêneo**

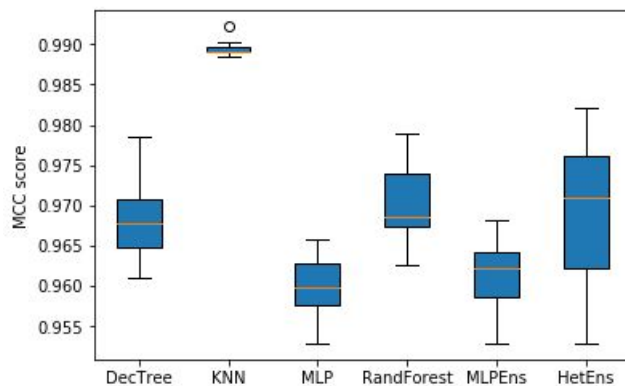
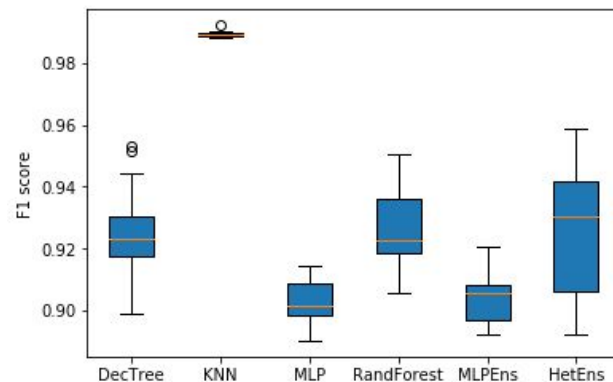
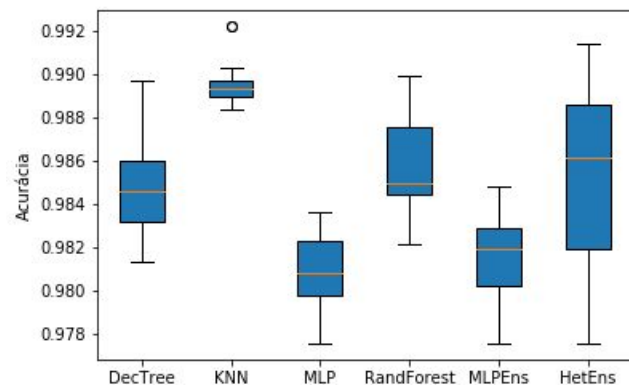
Voting classifier, *soft*

Comparação de modelos (média e desvio padrão)

CV 20-fold



Comparação de modelos



Testes estatísticos

Kruskal Wallis

```
Kruskal-Wallis test for cross validation with k=20  
  
Accuracy: statistics=81.161, p=0.000000000000000047964  
F1 score: statistics=95.234, p=0.000000000000000000053  
MCC: statistics=93.056, p=0.0000000000000000000153  
  
Different distributions (reject H0)
```

Testes estatísticos

Posthoc Nemenyi

1	Árvore de decisão
2	KNN
3	MLP
4	Random Forest
5	Comitê MLP
6	Comitê Heterogêneo

Posthoc Nemenyi for accuracy

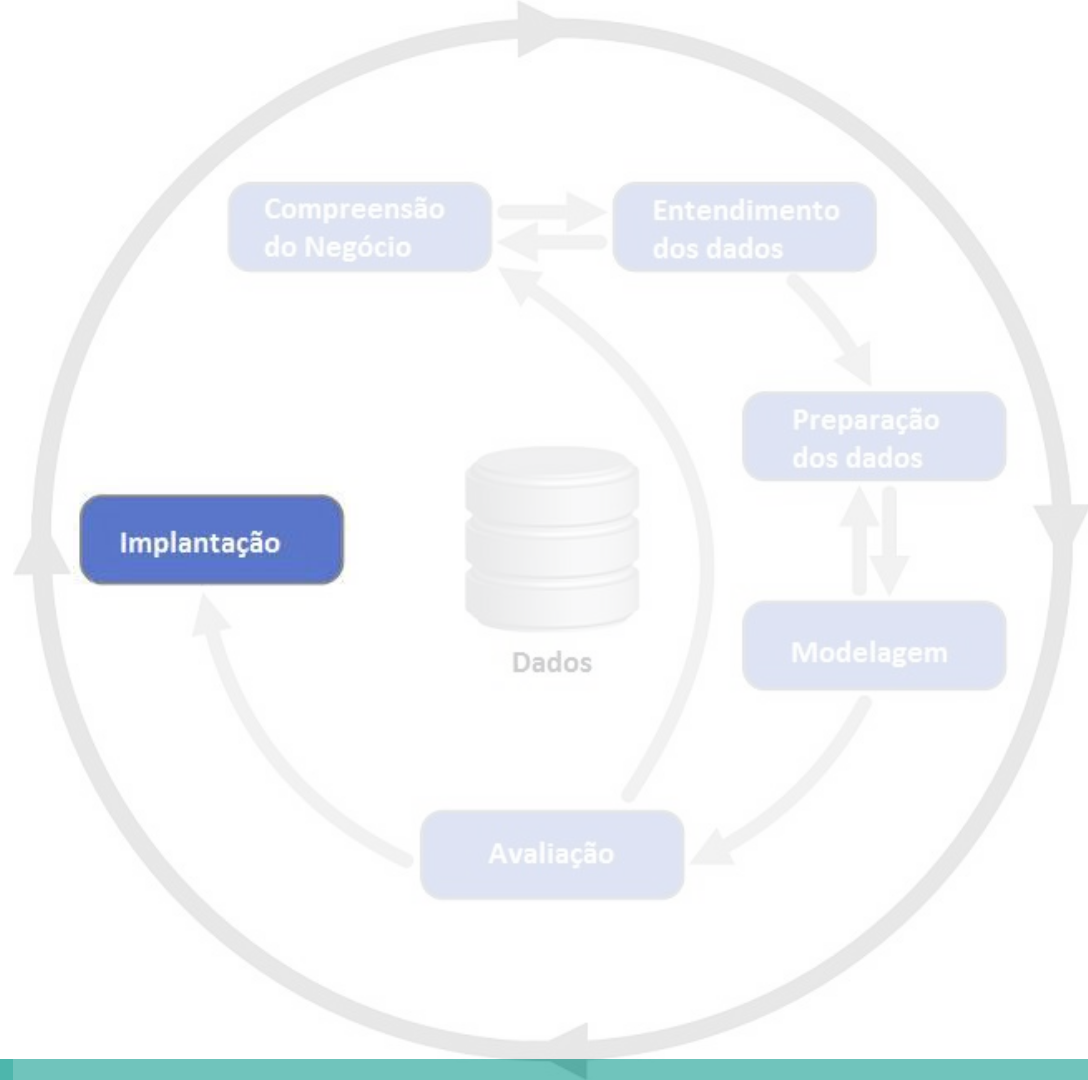
	1	2	3	4	5	6
1	-1.000000	4.278506e-02	3.067420e-02	0.817408	7.287382e-02	0.480139
2	0.042785	-1.000000e+00	4.355513e-09	0.609191	3.578990e-08	0.900868
3	0.030674	4.355513e-09	-1.000000e+00	0.000139	9.997870e-01	0.000007
4	0.817408	6.091906e-01	1.387272e-04	-1.000000	5.766781e-04	0.995426
5	0.072874	3.578990e-08	9.997870e-01	0.000577	-1.000000e+00	0.000036
6	0.480139	9.008676e-01	6.798665e-06	0.995426	3.554310e-05	-1.000000

Posthoc Nemenyi for F1 score

	1	2	3	4	5	6
1	-1.000000	1.026202e-03	3.073091e-02	0.970662	3.397480e-02	0.863218
2	0.001026	-1.000000e+00	1.415531e-12	0.025357	1.869587e-12	0.078294
3	0.030731	1.415531e-12	-1.000000e+00	0.001337	1.000000e+00	0.000229
4	0.970662	2.535662e-02	1.337256e-03	-1.000000	1.536817e-03	0.999253
5	0.033975	1.869587e-12	1.000000e+00	0.001537	-1.000000e+00	0.000268
6	0.863218	7.829439e-02	2.292648e-04	0.999253	2.680839e-04	-1.000000

Posthoc Nemenyi for MCC

	1	2	3	4	5	6
1	-1.000000	2.953991e-04	5.677129e-02	0.869864	1.389019e-01	0.733418
2	0.000295	-1.000000e+00	8.081112e-13	0.034400	1.541104e-11	0.075804
3	0.056771	8.081112e-13	-1.000000e+00	0.000653	9.995402e-01	0.000177
4	0.869864	3.439999e-02	6.525745e-04	-1.000000	2.916309e-03	0.999855
5	0.138902	1.541104e-11	9.995402e-01	0.002916	-1.000000e+00	0.000905
6	0.733418	7.580410e-02	1.772567e-04	0.999855	9.053353e-04	-1.000000

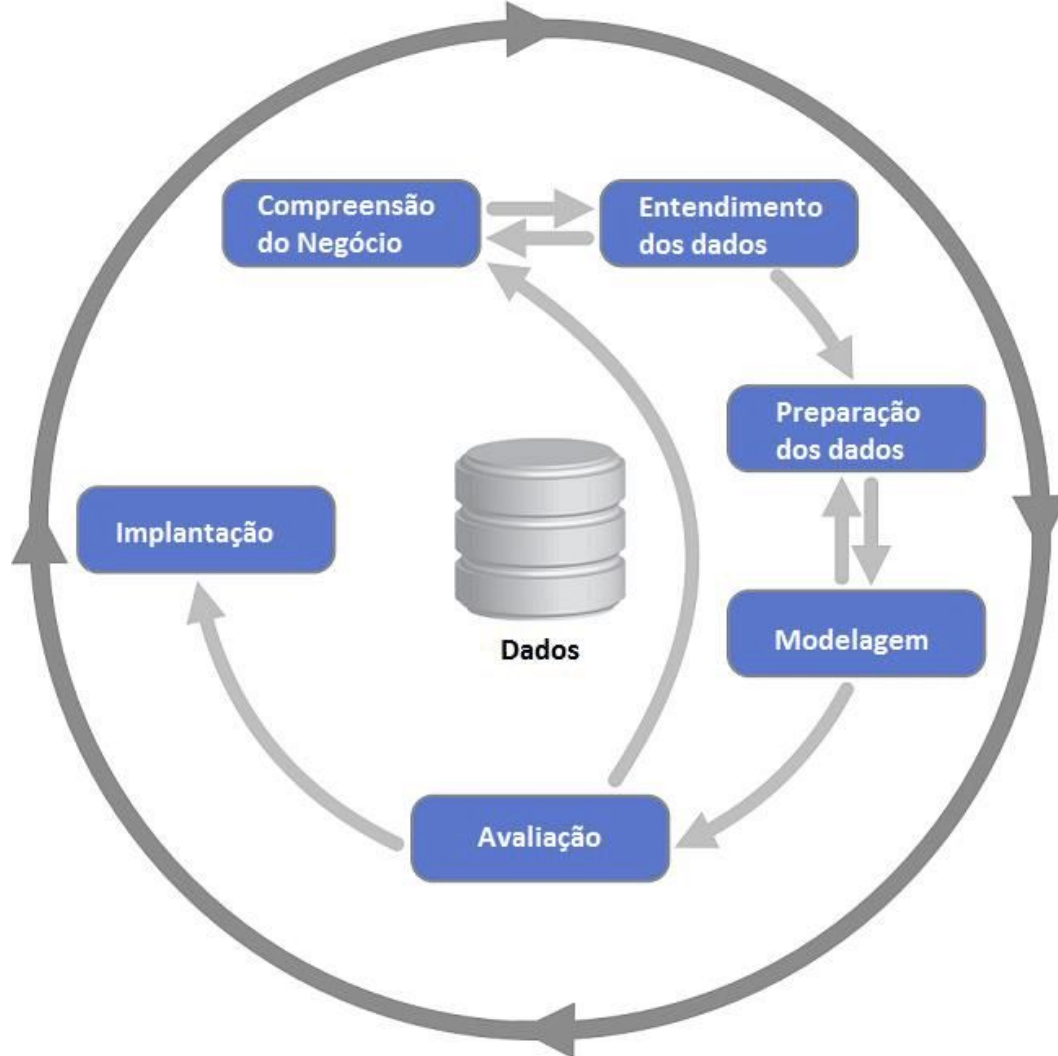


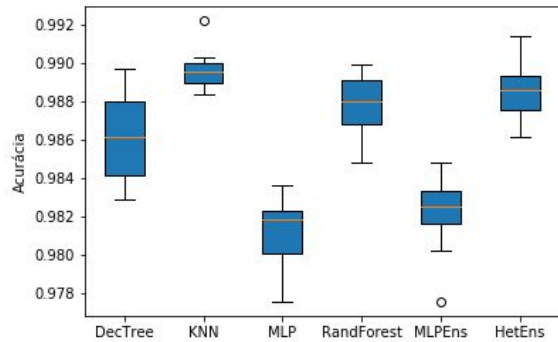
Resultado final

K-Nearest Neighbor com K=2

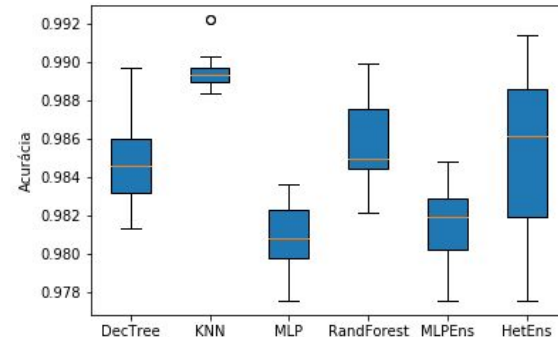
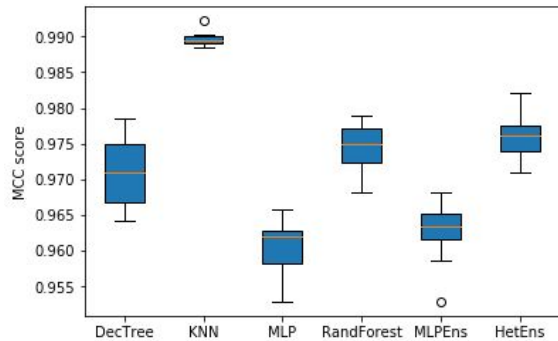
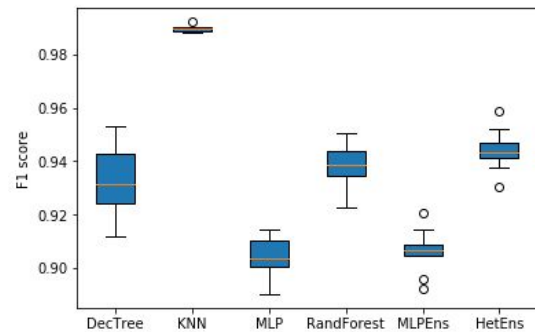
Predito	Real			
	11441	0	0	0
	4	3474	0	0
	34	1	36046	0
	175	57	6	1351

- **Acurácia:** 99,47%
- **F1 score:** 97,51%
- **MCC score:** 0,98898





Validação cruzada
10-fold



Validação cruzada
20-fold

