Trabalho 5: Árvores de Decisão
Reconhecimento de Padrões – 2021/1

Nesse trabalho sobre árvores de decisão pede-se que, utilizando *random forest* e procura em grid, que se preveja a máxima temperatura na cidade de Seatle, usando informações passadas do clima. Estão disponíveis 6 anos de dados. Cinco variáveis distintas devem ser usadas para a predição: year, ws_1, temp_2, temp_1, average. Essas variáveis são encontradas em cada registro da tabela temps_extended.xlsx. Essa tabela, de onde extraímos os cinco registros mostrado abaixo, contém outras variáveis além das cinco acima mencionadas. A variável *actual* corresponde a temperatura máxima verificada no dia.

Tabela 1: Cinco registros extraídos da tabela extended.xlsx

	year	month	day	weekday	ws_1	prcp_1	snwd_1	temp_2	temp_1	average	actual	friend
0	2011	1	1	Sat	4.92	0.00	0	36	37	45.6	40	40
1	2011	1	2	Sun	5.37	0.00	0	37	40	45.7	39	50
2	2011	1	3	Mon	6.26	0.00	0	40	39	45.8	42	42
3	2011	1	4	Tues	5.59	0.00	0	39	42	45.9	38	59
4	2011	1	5	Wed	3.80	0.03	0	42	38	46.0	45	39

O significado das 5 variáveis selecionadas é descrito a seguir:

year – ano em que o registro ocorreu
ws_1 – velocidade média do vento um dia antes
tem_2 – temperatura máxima dois dias antes
temp_1 – temperatura máxima um dia antes
average – histórico de média de temperaturas máximas

O que se pede:

1. Que você avalie o desempenho de uma árvore utilizando *random forest*. Você tem os seguintes hiperparâmetros de procura utilizando a biblioteca do *scikit learn*.

```
{'bootstrap': [True, False],
'max_depth': [10, 20, 30, 40, 50, 60, 70, 80, 90, 100,
None],
'max_features': ['auto', 'sqrt'],
'min_samples_leaf': [1, 2, 4],
'min_samples_split': [2, 5, 10],
'n_estimators': [200, 400, 600, 800, 1000, 1200, 1400,
1600, 1800, 2000]}
```

   Para esse experimento, divida o conjunto de dados inicial em 75% para treinamento e 25% para teste. Faça alguns experimentos em busca de um modelo que resulte numa maior precisão. Mostre o melhor modelo obtido em termos do erro absoluto médio e do erro quadrático médico no conjunto de teste.

Fazer o upload no sistema de EAD até 14/06/2021