Laboratório 2
Estimativa de Volume de Chuva

O arquivo RadarChuvaJan2021.csv contém 107136 linhas de dados de radar de chuva para o mês de Janeiro de 2021. Os dados são capturados a cada 15 minutos e cinco características são armazenadas: **DBZH, DBZV,KDP, ZDR, RHOHV**. O valor estimado de chuva (**Tp_est**) é capturado por um pluviômetro.

A base de dados é fortemente desbalanceada com a maior parte dos valores para **Tp_est** igual a zero. Outro fato a ser observado nessa base é a ausência de valores em algumas características. Quando o valor do radar não pode ser capturado, NaN é colocado no valor da característica, como ilustrado na figura abaixo.

	Data	DBZH	DBZV	KDP	ZDR	RHOHV	Tp_est
10	2021-01-01 02:30	26.410000	24.109999	NaN	2.79	0.8652	0.2
13	2021-01-01 03:15	NaN	NaN	NaN	4.49	0.3401	0.8
14	2021-01-01 03:30	22.359999	19.180000	NaN	3.68	0.8570	0.2
18	2021-01-01 04:30	23.480000	20.059999	NaN	3.91	0.8806	0.2
261	2021-01-03 17:15	5.190000	7.070000	-0.5	-1.37	0.5333	0.8

Para esses casos, existem duas alternativas: eliminar as linhas que tem alguma variável NaN ou preencher esses valores de alguma forma, como por exemplo com a média dos valores daquela característica. Esse processo deve ser realizado com cuidado, pois a base de dados tem um aspecto temporal e média geral pode não ser a melhor alternativa.

Seu trabalho consiste em treinar um regressor para estimar **Tp_est** para um dia qualquer da base. Por exemplo, se o seu dia de teste for 31, você poderá usar de 1 a 30 para treinar o modelo.

Para o treinamento e teste do modelo, você deve desprezar todas as linhas que em que **Tp_est** = 0.

Importante: Lembre-se que o dia escolhido para o teste não deve ser usado para escolha dos parâmetros do classificador nem para o processo de limpeza e correção dos valores NaN. Isso configura vazamento de dados (*data leaking*).

Nos seus experimentos você deve reportar o erro médio quadrado (MSE) e o erro médio absoluto (MAE). Para cada experimento, apresente também um gráfico de dispersão no qual o eixo X tem o valor de **Tp_est** e o eixo Y tem o valor predito pelo seu algoritmo.

Você pode usar todos os algoritmos vistos em sala de aula, ou seja, não utilize técnicas de ensemble como Random Forest e Gradient Boosting.

O que deve ser entregue:
Um arquivo python (jupyter notebook) documentado.