

### Trabalho 3

#### Membros:

Isabela de Freitas Lima A Mariz 1720511

Rodrigo Peroba de Souza 1511044

Vinicius Bernardes Bonemer 1620805

#### Análise exploratória:

A zona 1 e 2 tem consumo constante durante todo o ano

A zona 3 parece ter consumo maior em torno de setembro/outubro, quando aumenta a temperatura.

Através do Feature Statistics vemos que zona 3 consome menos que as outras

Sem outlier aparente

Pela correlação, aparentemente a temperatura tem maior influência no consumo elétrico seguida pela humidade.

Há uma forte correlação entre os consumos, mas parece absurdo eles estarem relacionados quando tratamos de zonas diferentes.

#### Seleção de atributos:

Selecionamos apenas um consumo e removemos os outros dois para fazer uma análise separada deles.

#### Geração do modelo:

Rodamos com os modelos Random Forest, Tree, Neural Network, Linear Regression e SVM. Podemos descartar logo no início da análise o modelo Tree por usarmos também o Random Forest e obter um resultado de cara melhor.

Test and Score - Orange

☐ Cross validation  
 Number of folds: 5  
☒ Stratified  
☐ Cross validation by feature  
☐ Random sampling  
 Repeat train/test: 10  
 Training set size: 66 %  
☐ Stratified  
☐ Leave one out  
☐ Test on train data  
☒ Test on test data

Model	MSE	RMSE	MAE	R2
Tree	14204681.183	3768.910	1918.729	0.718
SVM	50430330.761	7101.432	5910.894	-0.003
Random Forest	6905091.991	2627.754	1644.544	0.863
Neural Network	864923316.573	29409.579	28541.869	-16.200
Linear Regression	39344694.808	6272.535	5175.991	0.218

Os outros modelos estavam com um R2 bem baixo, tentamos fazer algumas alterações superficiais para melhorar o resultado. Mudando a função de ativação da rede neural e o kernel da SVM.