

SmartIFSC:

Transformando dados em modelos

Leonardo Benitez

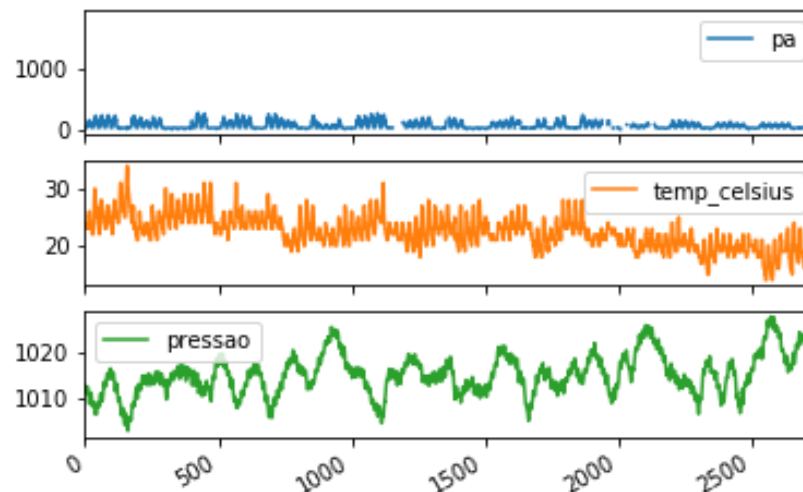
Objetivos

- **Modelar numericamente o consumo elétrico do IFSC**
- **Prever o consumo futuro**
- **Permitir identificar momentos “anormais”: quando o consumo está muito acima do esperado**
- **Abordagem 1: poucos dados, modelo simples**
- **Abordagem 2: muitos dados, modelo complexo**

Abordagem 1

Sobre os dados

- Obteve-se os dados do SmartIFSC
- Potência, temperatura e pressão
- Granularidade de 1h
- Campus Florianópolis, período das 13:30 às 17:30, de 2019-02-18 (segunda semana de aula) até 2019-06-12 (data atual)



Algoritmo

- **Regressão:** prever um valor numérico y (potência), a partir do vetor de entrada X (temperatura e pressão)
- **Regressão linear:** encontra uma função linear que minimize o erro

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

- **Limpeza dos dados:** remover linhas incompletas, outliers, dias não-letivos, etc
- **Validação com a técnica de *cross validation***

Resultados

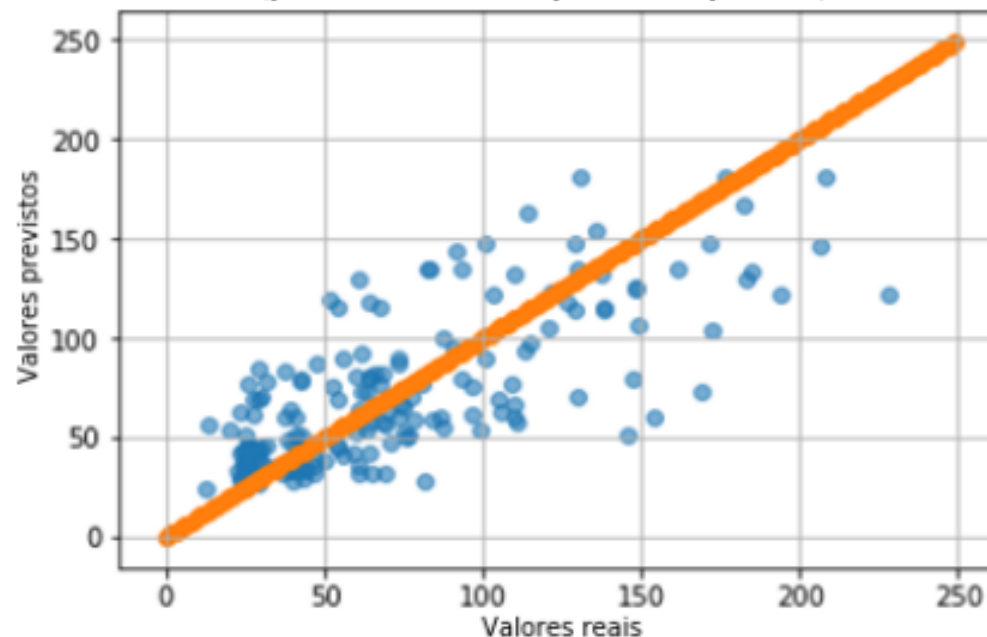
- **MAE: 26,93 KW**
- **RMSE: 34,67 KW**
- **Percentual: 37,51%**

$$MAE(\hat{y}) = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

$$RMSE(\hat{y}) = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

$$Perc(\hat{y}) = \frac{MAE(\hat{y})}{média(y)},$$

- **Valores previsto pelo algoritmo vs valores reais:**

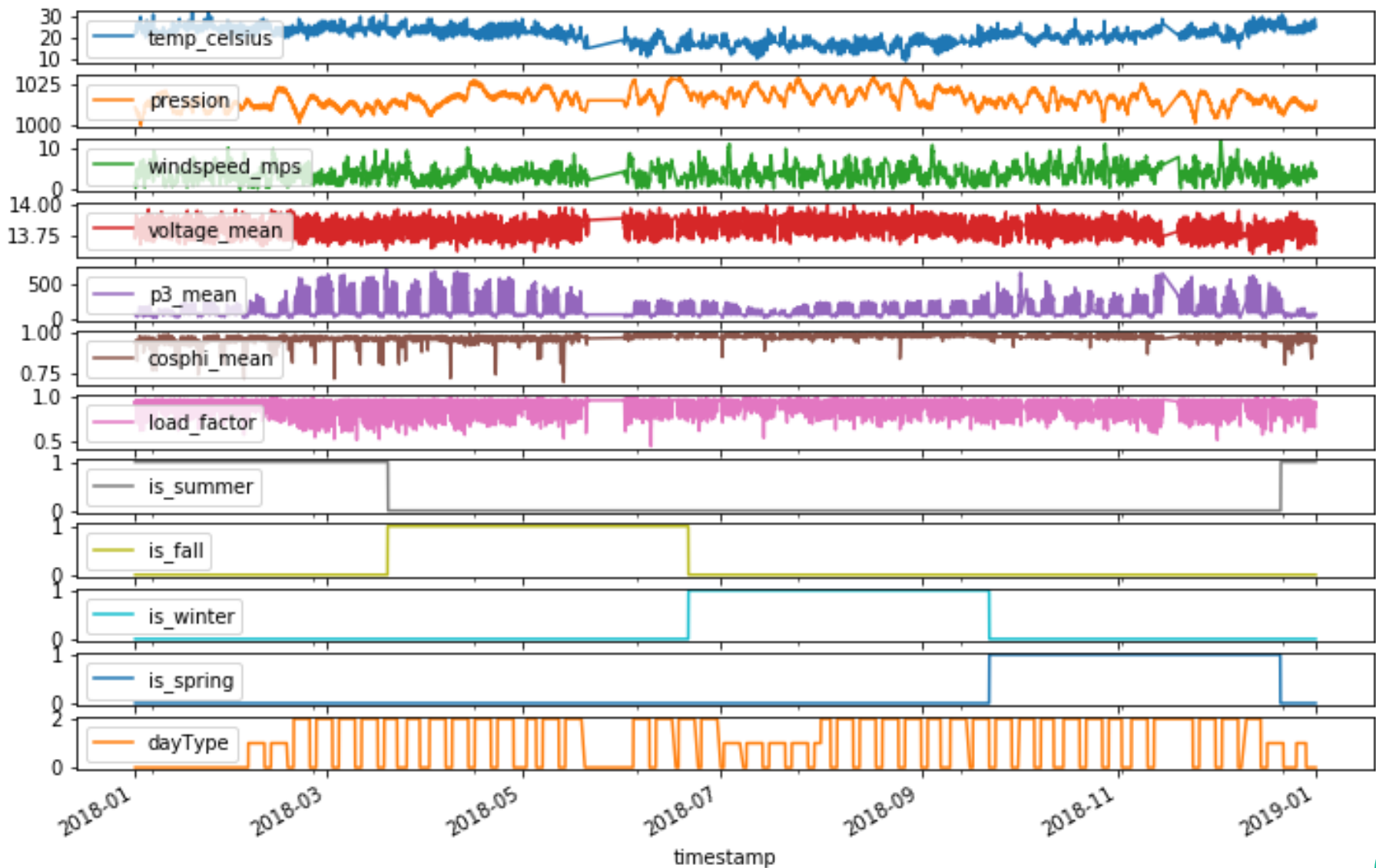


Abordagem 2

Sobre os dados

- Dados de 2018, do campus Florianópolis
- Granularidade de 1h
- Descrevem o perfil instantâneo do campus por meio de variáveis elétricas e climáticas
- Temperatura (°C), pressão (hPa), velocidade do vento (m/s), tensão (V), fator de potência, fator de carga (máximo/médio), tipo de dia (não letivo, letivo e somente administrativo) e estação do ano

Sobre os dados



Sobre os dados

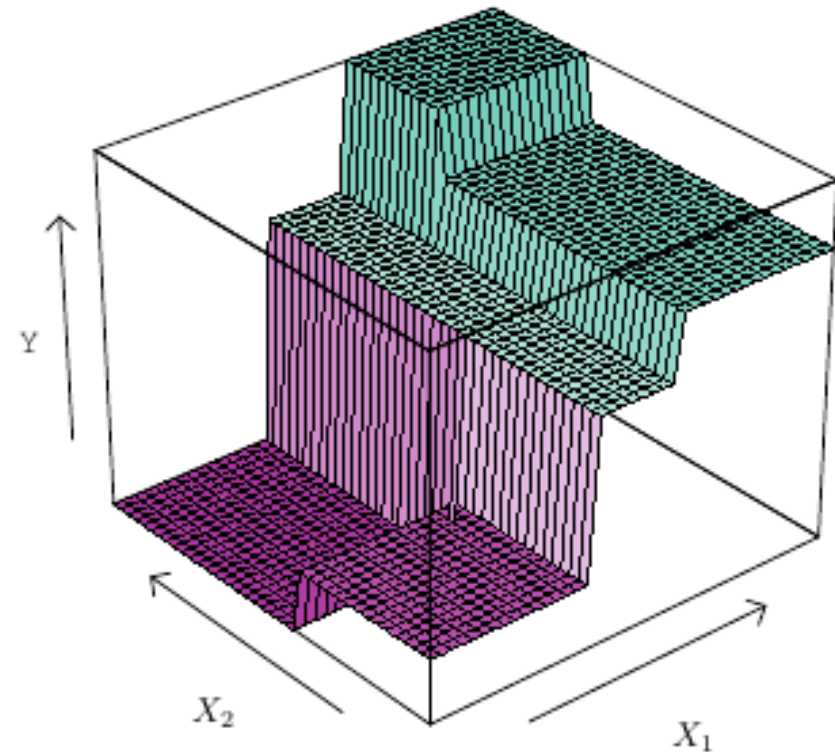
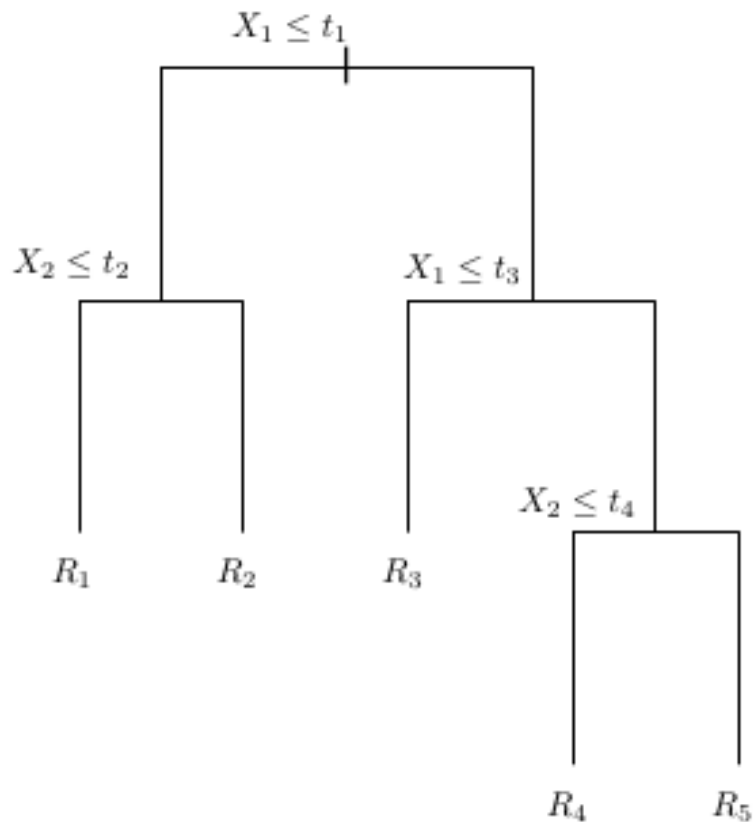
	temp_celsius	pression	windspeed_mps	voltage_mean	p3_mean	cosphi_mean	load_factor	is_summer	is_fall	is_winter	is_spring	dayType
count	7629.000000	7629.000000	7629.000000	7629.000000	7629.000000	7629.000000	7629.000000	7629.000000	7629.000000	7629.000000	7629.000000	7629.000000
mean	21.034605	1015.678215	3.780325	13.811183	155.874310	0.974422	0.901446	0.256128	0.242103	0.262289	0.239481	1.162669
std	3.833874	5.149767	1.858141	0.061987	127.979651	0.022231	0.073233	0.436522	0.428384	0.439908	0.426795	0.925313
min	9.000000	999.300000	0.000000	13.603576	13.904076	0.702013	0.440432	0.000000	0.000000	0.000000	0.000000	0.000000
25%	18.000000	1012.100000	2.300000	13.765798	67.877782	0.967909	0.870611	0.000000	0.000000	0.000000	0.000000	0.000000
50%	22.000000	1015.400000	3.600000	13.811567	87.371403	0.977661	0.925378	0.000000	0.000000	0.000000	0.000000	2.000000
75%	24.000000	1019.100000	4.900000	13.857204	218.355411	0.987969	0.954601	1.000000	0.000000	1.000000	0.000000	2.000000
max	32.000000	1029.900000	11.800000	13.995569	732.414639	0.999565	0.992561	1.000000	1.000000	1.000000	1.000000	2.000000

Algoritmo

- **Random Forest Regression**
- **Regressão:** prever um valor numérico y (potência), a partir do vetor de entrada X
- **Random forest:** fazer essa previsão a partir da agregação de diversas árvores de regressão, cada uma treinada em um subconjunto aleatório do vetor de entrada X
- **Árvore:** estrutura baseada em busca que divide os dados em regiões sucessivamente mais simples, até que esta região possa aproximar um único valor de saída

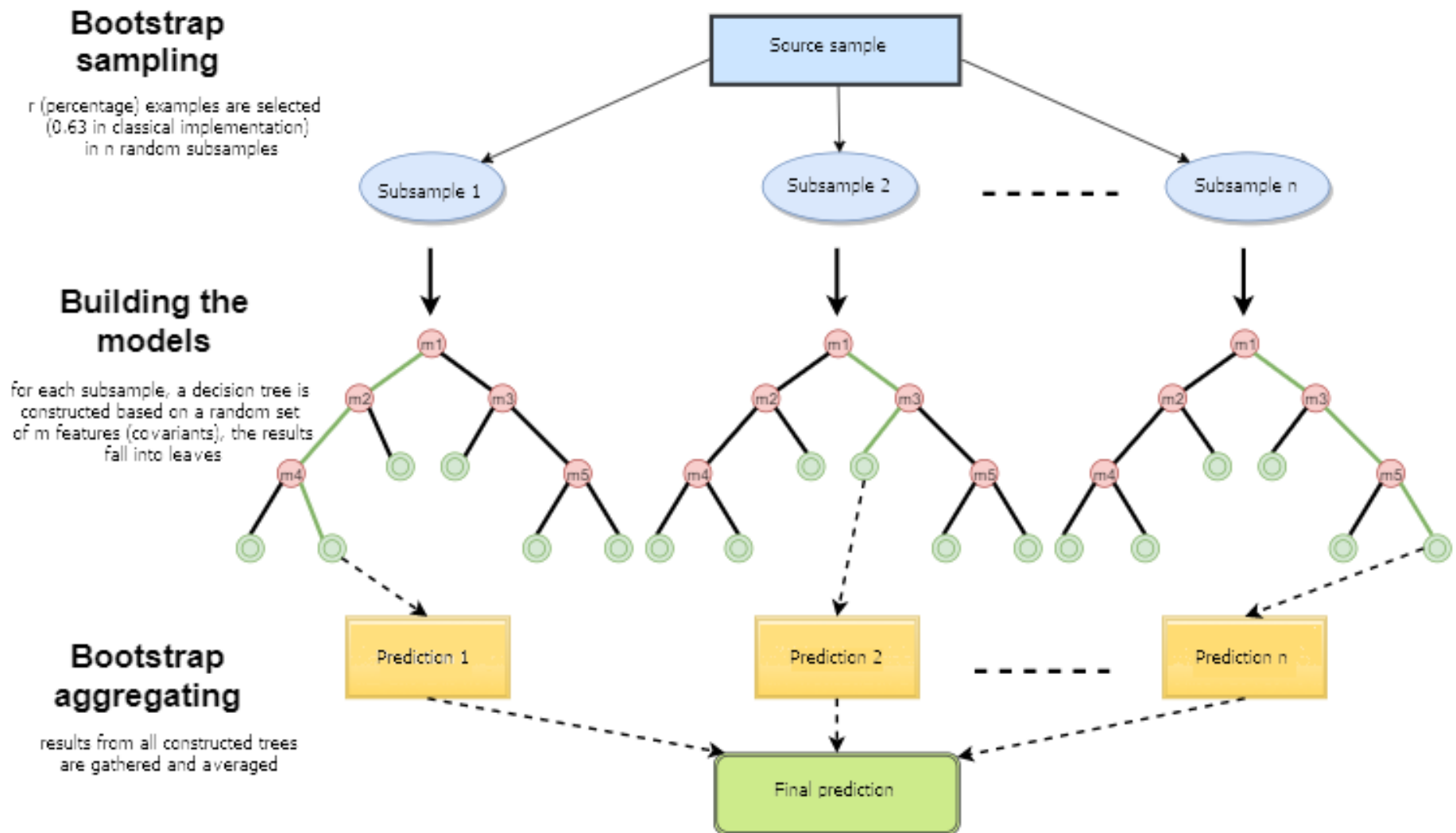
Algoritmo

- **Árvore de regressão (exemplo):**



Algoritmo

- **Random Forest Regression (exemplo):**



Resultados

- **MAE: 29,7 KW**
- **RMSE: 53,0 KW**
- **Percentual: 18,9 %**

$$MAE(\hat{y}) = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

$$RMSE(\hat{y}) = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

$$Perc(\hat{y}) = \frac{MAE(\hat{y})}{m\acute{e}dia(y)},$$

- **Valores previsto pelo algoritmo vs valores reais:**

