



# Modelo de *Machine Learning* para Predição de Correntes de Fuga em Isoladores de Distribuição Classe 25 kV

Apresentado ao Curso de Especialização em Ciência de Dados e Big Data na Pontifícia Universidade Católica de Minas Gerais (PUC MG), como requisito parcial à obtenção do título de especialista.

Rafael Henrique Pinotti



## Ferramentas



*NumPy*

*matplotlib*



PUC Minas



# Trabalho base

- Modelo Matemático de Predição de Correntes de Fuga em Isoladores de Distribuição Classe 25 kV.
- Dissertação de Mestrado de Milton Augusto Pinotti.
- Apresentado ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Regional de Blumenau - FURB.
- Utilizou a ferramenta Matlab para criar os modelos matemáticos de regressão linear e não linear.



## Escopo do trabalho

1. Definição do problema

2. Coleta de dados

3. Processamento/tratamento dos dados

4. Análise/exploração dos dados

5. Criação de modelo de ML

6. Interpretação dos resultados

7. Comunicação dos resultados

Fonte: PUC MG

# Definição do problema



- Grande preocupação das concessionárias de energia elétrica em fornecer este insumo com alta qualidade.
- Dentre os principais equipamentos utilizados no sistema elétrico estão os isoladores elétricos.
- As falhas em isoladores são responsáveis por muitos dos desligamentos não programados na rede elétrica.
- A saúde destes isoladores é monitorada através da medição de correntes de fuga.
- O objeto deste estudo é o isolador tipo Pilar Polimérico - HTV.
- Os dados foram gerados em uma estação de monitoramento de sistemas isolantes (EMSI) localizada na Praia Brava em Itajaí/SC.
- O período de coleta disponibilizado foi de 01 a 28/02 de 2010.



# Coleta e processamento dos dados

## Aquisição dos dados

```
# Importar dados de arquivo csv, onde o separador de campos é '.' e o separador de casas decimais é ','  
df = pd.read_csv('dados/dados_isolador40.csv', sep=';', decimal=',')
```

executed in 66ms, finished 16:27:22 2020-05-15

dados\_isolador40.csv

```
1  umidade;temperatura;pressao;vento_ang;vent_veloc;chuva;corrente_fuga  
2  70;29,6;1016;278;1,4;0;0,024992  
3  69;29,7;1016;297;1,4;0;0,022767  
4  70;29,8;1016;273;1,6;0;0,022607  
5  68;29,8;1016;243;2,6;0;0,022426  
6  68;29,7;1016;257;2,8;0;0,031797  
7  69;29,9;1016;261;2,4;0;0,032152  
8  67;29,9;1016;254;4,5;0;0,032233  
9  68;29,9;1016;287;3,4;0;0,032337  
10 68;29,9;1016;291;2,8;0;0,032357  
11 68;30;1016;254;2,6;0;0,03189  
12 67;30,1;1016;258;2,6;0;0,032369  
13 68;30,2;1016;263;2,6;0;0,032091  
14 66;30,3;1016;305;1,6;0;0,031828  
15 66;30,5;1016;233;2,6;0;0,032115  
16 66;30,5;1016;334;2,2;0;0,03256  
17 65;30,6;1016;285;1,8;0;0,032550
```

	umidade	temperatura	pressao	vento_ang	vent_veloc	chuva	corrente_fuga
0	70	29.6	1016	278	1.4	0	0.024992
1	69	29.7	1016	297	1.4	0	0.022767
2	70	29.8	1016	273	1.6	0	0.022607
3	68	29.8	1016	243	2.6	0	0.022426
4	68	29.7	1016	257	2.8	0	0.031797

# Análise/exploração dos dados

```

# formato do dataframe (registros/colunas)
df.shape

executed in 9ms, finished 16:27:22 2020-05-15

```

(40316, 7)

```

# Quantidade de registros com valor zero em cada atributo.
(df == 0).sum()

executed in 44ms, finished 16:27:22 2020-05-15

```

```

umidade          0
temperatura      0
pressao          0
vento_ang        92
vent_veloc      14511
chuva           38064
corrente_fuga    0
dtype: int64

```

```

# Valores missing
df.isnull().any()

executed in 20ms, finished 16:27:22 2020-05-15

```

```

umidade          False
temperatura      False
pressao          False
vento_ang        False
vent_veloc       False
chuva            False
corrente_fuga    False
dtype: bool

```

```

# Registros nulos
df.isnull().sum()

executed in 19ms, finished 16:27:22 2020-05-15

```

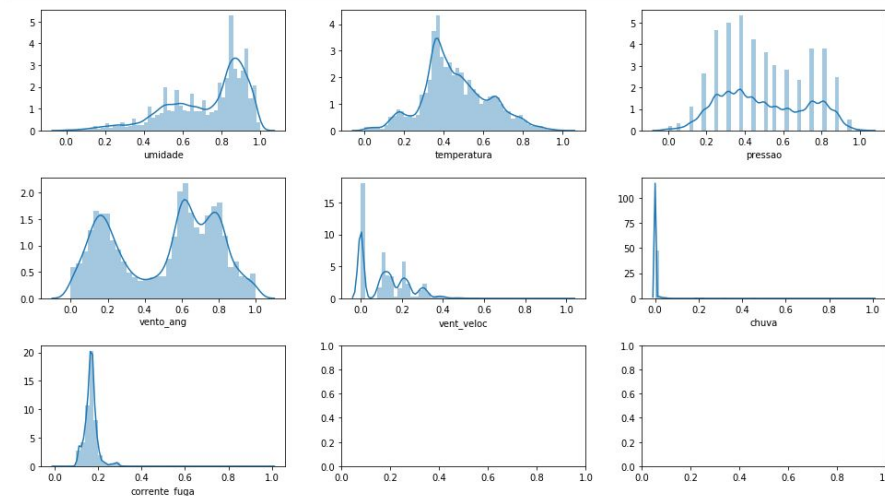
```

umidade          0
temperatura      0
pressao          0
vento_ang        0
vent_veloc       0
chuva            0
corrente_fuga    0
dtype: int64

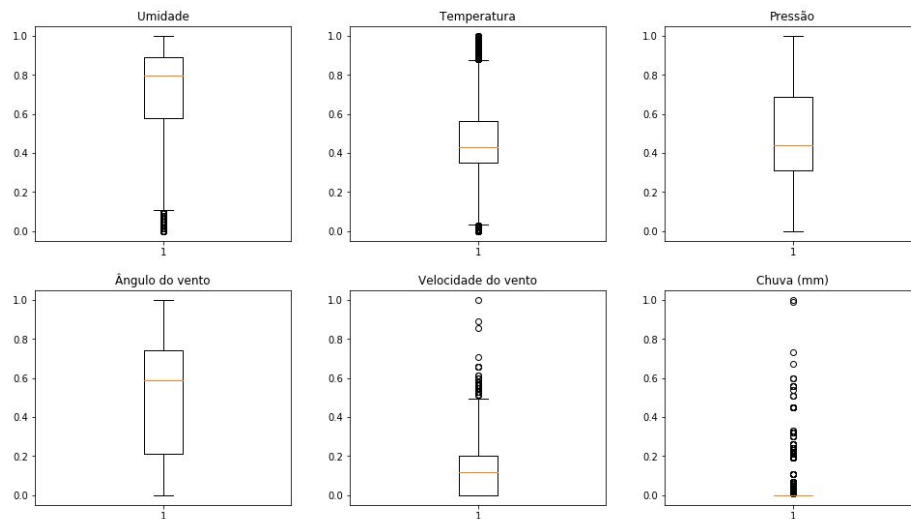
```

```
min_scaler = MinMaxScaler(feature_range=(0,1))
scaled_data = pd.DataFrame(min_scaler.fit_transform(df), columns=df.columns)
executed in 22ms, finished 16:27:28 2020-05-15
```

```
distplots_dataframe(scaled_data)
executed in 5.57s, finished 16:27:34 2020-05-15
```



# Análise/exploração dos dados



	◇ umidade ◇	temperatura ◇	pressão ◇	vento_ang ◇	vent_veloc ◇	chuva ◇	corrente_fuga ◇
umidade	1.0	-0.44157	-0.20368	0.32259	-0.50396	0.11214	0.69658
temperatura	-0.44157	1.0	-0.39874	-0.43925	0.45762	-0.11037	-0.34919
pressão	-0.20368	-0.39874	1.0	-0.011539	-0.045943	0.044076	-0.064254
vento_ang	0.32259	-0.43925	-0.011539	1.0	-0.34307	0.0424	0.22667
vent_veloc	-0.50396	0.45762	-0.045943	-0.34307	1.0	-0.0043842	-0.3078
chuva	0.11214	-0.11037	0.044076	0.0424	-0.0043842	1.0	0.29679
corrente_fuga	0.69658	-0.34919	-0.064254	0.22667	-0.3078	0.29679	1.0



▶ 01. Regressão Linear

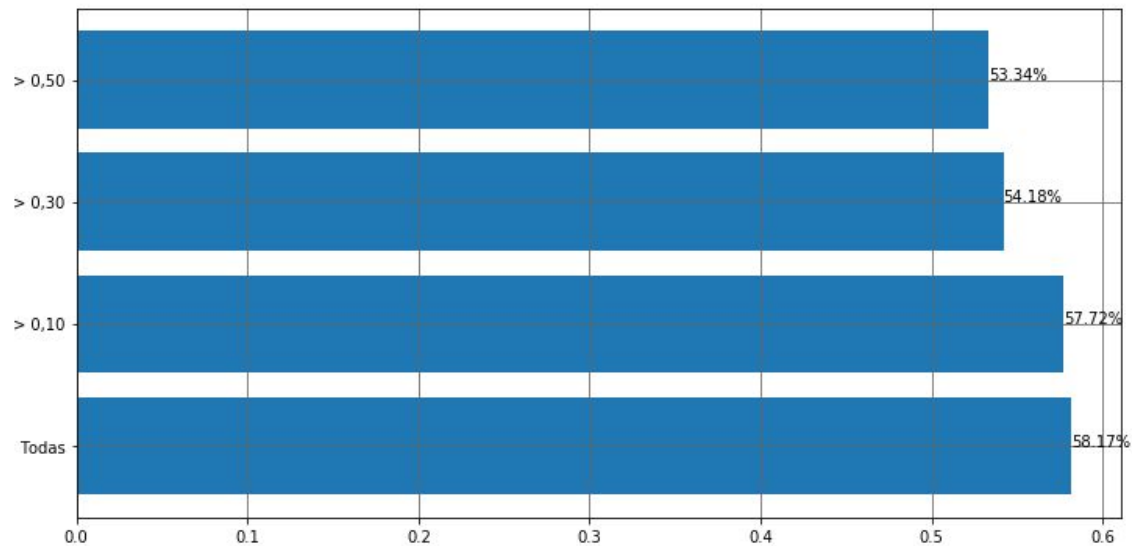
▶ 02. Árvore de Decisão

▶ 03. Random Forest

▶ 04. Random Forest com ajustes nos hiperparâmetros

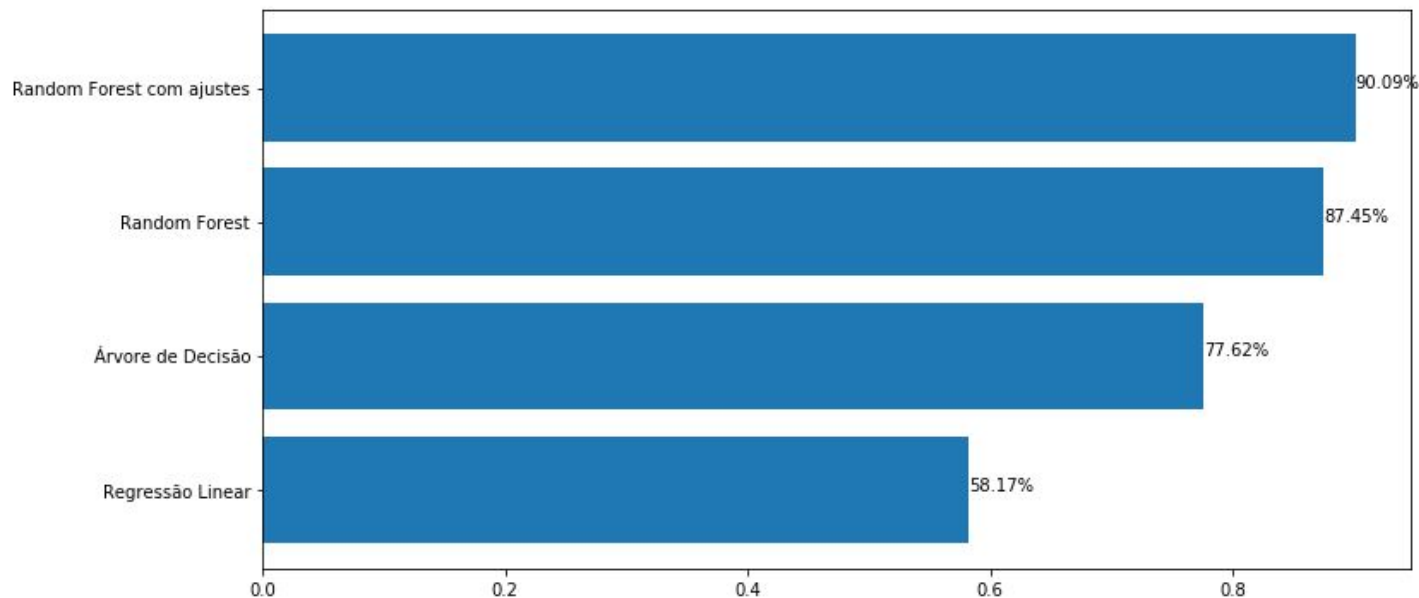
# Criação dos modelos de ML

Comparação entre os scores - Regressão Linear

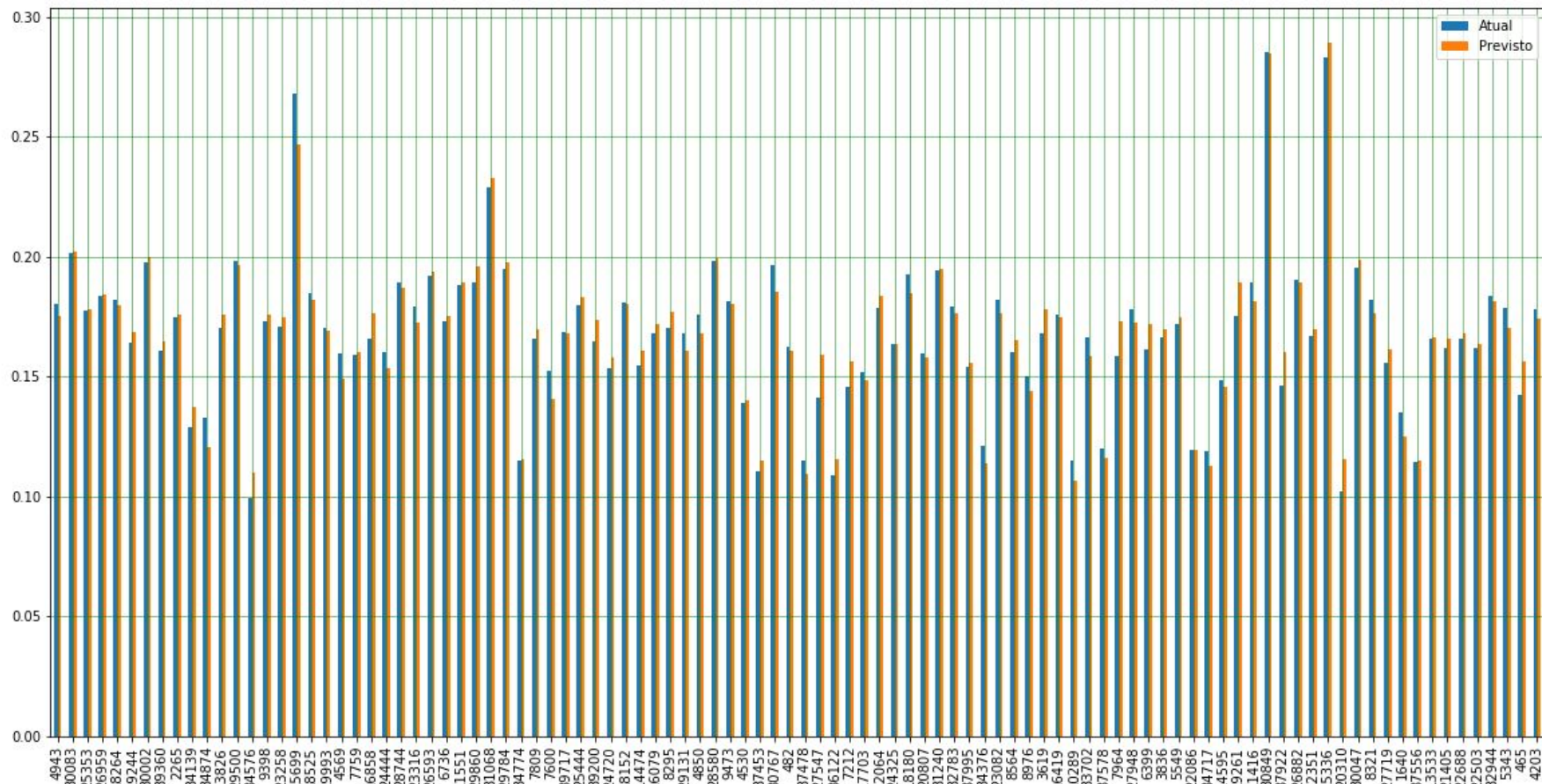


# Interpretação dos resultados

Comparação entre as diversas abordagens



◆	Atual ◆	Previsto ◆
4943	0.180253	0.175453
30083	0.201466	0.202456
25353	0.177424	0.178007
26959	0.183708	0.184582
18264	0.181854	0.179802
19244	0.164301	0.168584
30002	0.197662	0.199776
39360	0.160629	0.165030
2265	0.174859	0.175861
34139	0.129026	0.137641



# Title: Modelo de Machine Learning para Predição de Correntes de Fuga em Isoladores de Distribuição Classe 25kV

## 1 Problem Statement

What problem are you trying to solve?  
What larger issues do the problem address?

- Eliminar as quedas na distribuição de energia elétrica causadas por falhas nos isoladores elétricos.
- Problemas nestes equipamentos causam rompimento na distribuição de energia elétrica.

## 2 Outcomes/Predictions

What prediction(s) are you trying to make?  
Identify applicable predictor (X) and/or target (y) variables.

- Antecipar o momento em que os isoladores podem falhar, através da previsão das correntes de fuga.
- Analisando os fatores ambientais podemos prever a corrente de fuga.
- Variável alvo: corrente de fuga.
- Variáveis preditoras: umidade, temperatura, pressão, ângulo do vento, velocidade do vento, quantidade de chuva.

## 3 Data Acquisition

Where are you sourcing your data from?  
Is there enough data? Can you work with it?

- Dados coletados em estação de monitoramento de sistemas isolantes (EMSI), na Praia Brava em Itajaí/SC.
- Os dados disponibilizados são do período de 01 a 28/02 de 2010.
- O dataset possui 7 atributos com 40.136 registros.
- O dataset foi disponibilizado através de arquivo CSV.

## 4 Modeling

What models are appropriate to use given your outcomes?

- Os modelos de regressão são os ideais para este tipo de problema.
- Três algoritmos serão testados:
  1. Regressão Linear.
  2. Árvore de Decisão.
  3. Random Forest.
- Na Regressão Linear serão criados diversos modelos de acordo com critérios de correlação.
- Para o algoritmo de Random Forest serão aplicados ajustes nos hiperparâmetros.

## 5 Model Evaluation

How can you evaluate your model's performance?

- Para cada modelo é calculado um score, permitindo uma comparação através da acurácia dos modelos.
- O score calculado é o coeficiente  $R^2$  da predição.

## 6 Data Preparation

What do you need to do to your data in order to run your model and achieve your outcomes?

- A verificação e tratamento de campos nulos e vazios é necessária para que não haja inconsistências.
- O tratamento dos outliers pode ser necessário para garantir a qualidade do modelo.