

Solução DojoML

`curiosos_do_ml`

Mateus C. Pedrino

Bruno H. Rasteiro

Kauê L. Moraes





Etapas



1. Verificação de dados ausente;
2. Transformação data e hora (`to_datetime`);
3. Variáveis categóricas;
4. Agrupamento de planilhas (`goupby`, `merge`);
5. Estudo de correlação;
6. Classificação (`GridSearchCV`).





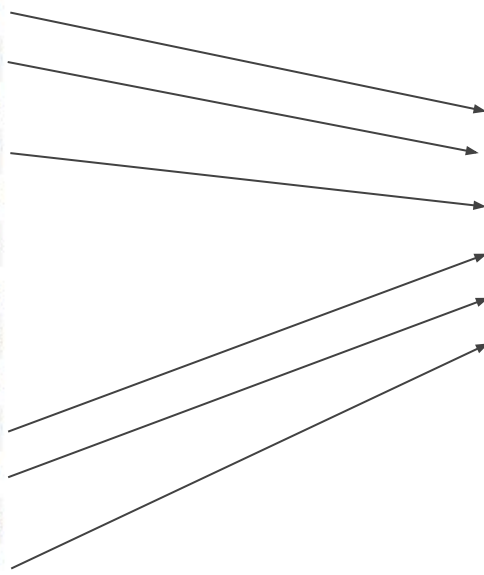
Visão geral

Base

| | maquina | data | falha |
|----|---------|------------|-------|
| 0 | 1 | 2015-01-01 | ok |
| 1 | 1 | 2015-01-02 | ? |
| 2 | 1 | 2015-01-03 | ? |
| 3 | 1 | 2015-01-04 | ok |
| 4 | 1 | 2015-01-05 | ? |
| 5 | 1 | 2015-01-06 | ok |
| 6 | 1 | 2015-01-07 | ok |
| 7 | 1 | 2015-01-08 | ok |
| 8 | 1 | 2015-01-09 | ok |
| 9 | 1 | 2015-01-10 | ok |
| 10 | 1 | 2015-01-11 | ? |
| 11 | 1 | 2015-01-12 | ? |
| 12 | 1 | 2015-01-13 | ok |
| 13 | 1 | 2015-01-14 | ? |
| 14 | 1 | 2015-01-15 | ok |
| 15 | 1 | 2015-01-16 | ok |

Teste

| | maquina | data |
|------|---------|------------|
| 9559 | 1 | 2015-01-02 |
| 1590 | 1 | 2015-01-03 |
| 7875 | 1 | 2015-01-05 |
| 889 | 1 | 2015-01-11 |
| 2634 | 1 | 2015-01-12 |
| 1088 | 1 | 2015-01-14 |



data/hora e dummies

| data | maquina | erro |
|---------------------|---------|--------|
| 2015-01-03 07:00:00 | 1 | error1 |
| 2015-01-03 20:00:00 | 1 | error3 |
| 2015-01-04 06:00:00 | 1 | error5 |
| 2015-01-10 15:00:00 | 1 | error4 |
| 2015-01-22 10:00:00 | 1 | error4 |

data/hora
→
dummy

| dia | horario | maquina | error1 | error2 | error3 | error4 | error5 |
|------------|----------|---------|--------|--------|--------|--------|--------|
| 2015-01-01 | 06:00:00 | 24 | 1 | 0 | 0 | 0 | 0 |
| 2015-01-01 | 06:00:00 | 73 | 0 | 0 | 0 | 1 | 0 |
| 2015-01-01 | 06:00:00 | 81 | 1 | 0 | 0 | 0 | 0 |
| 2015-01-01 | 07:00:00 | 43 | 0 | 0 | 1 | 0 | 0 |
| 2015-01-01 | 08:00:00 | 14 | 0 | 0 | 0 | 1 | 0 |

↓
por dia



| maquina | dia | error1 | error2 | error3 | error4 | error5 |
|---------|------------|--------|--------|--------|--------|--------|
| 1 | 2015-01-03 | 1 | 0 | 1 | 0 | 0 |
| 1 | 2015-01-04 | 0 | 0 | 0 | 0 | 1 |
| 1 | 2015-01-10 | 0 | 0 | 0 | 1 | 0 |
| 1 | 2015-01-22 | 0 | 0 | 0 | 1 | 0 |
| 1 | 2015-01-25 | 0 | 0 | 0 | 1 | 0 |



mais *dummies*...

| | maquina | data | falha |
|-------|---------|------------|-------|
| 22293 | 1 | 2015-01-01 | ok |
| 602 | 1 | 2015-01-04 | ok |
| 2544 | 1 | 2015-01-06 | ok |
| 11899 | 1 | 2015-01-07 | ok |
| 13940 | 1 | 2015-01-08 | ok |



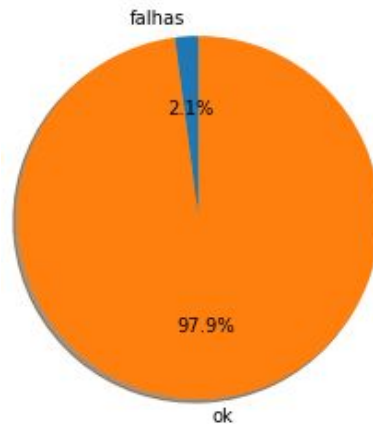
| | maquina | comp1 | comp2 | comp3 | comp4 | ok |
|-------|---------|-------|-------|-------|-------|----|
| 22293 | 1 | 0 | 0 | 0 | 0 | 1 |
| 602 | 1 | 0 | 0 | 0 | 0 | 1 |
| 2544 | 1 | 0 | 0 | 0 | 0 | 1 |
| 11899 | 1 | 0 | 0 | 0 | 0 | 1 |
| 13940 | 1 | 0 | 0 | 0 | 0 | 1 |

treino

| maquina | modelo | idade |
|---------|--------|-------|
| 1 | model3 | 18 |
| 2 | model4 | 7 |
| 3 | model3 | 8 |
| 4 | model3 | 7 |
| 5 | model3 | 2 |



| maquina | idade | model1 | model2 | model3 | model4 |
|---------|-------|--------|--------|--------|--------|
| 1 | 18 | 0 | 0 | 1 | 0 |
| 2 | 7 | 0 | 0 | 0 | 1 |
| 3 | 8 | 0 | 0 | 1 | 0 |
| 4 | 7 | 0 | 0 | 1 | 0 |
| 5 | 2 | 0 | 0 | 1 | 0 |





Agrupando *dummies* em todo período

| maquina | idade | model1 | model2 | model3 | model4 | error1 | error2 | error3 | error4 | error5 | comp1 | comp2 | comp3 | comp4 | ok |
|---------|-------|--------|--------|--------|--------|--------|--------|--------|--------|--------|-------|-------|-------|-------|-------|
| 1 | 18 | 0 | 0 | 1 | 0 | 11 | 6 | 9 | 5 | 4 | 1.0 | 1.0 | 0.0 | 3.0 | 247.0 |
| 2 | 7 | 0 | 0 | 0 | 1 | 5 | 10 | 5 | 5 | 3 | 0.0 | 2.0 | 0.0 | 0.0 | 261.0 |
| 3 | 8 | 0 | 0 | 1 | 0 | 11 | 10 | 8 | 8 | 2 | 1.0 | 3.0 | 0.0 | 0.0 | 248.0 |
| 4 | 7 | 0 | 0 | 1 | 0 | 10 | 11 | 8 | 2 | 0 | 2.0 | 3.0 | 0.0 | 0.0 | 248.0 |
| 5 | 2 | 0 | 0 | 1 | 0 | 14 | 8 | 7 | 6 | 3 | 2.0 | 3.0 | 0.0 | 0.0 | 245.0 |





Variáveis numéricas (sensores)

| | data | maquina | voltagem | rotacao | pressao | vibracao |
|--|---------------------|---------|------------|------------|------------|-----------|
| | 2015-01-01 06:00:00 | 1 | 176.217853 | 418.504078 | 113.077935 | 45.087686 |
| | 2015-01-01 07:00:00 | 1 | 162.879223 | 402.747490 | 95.460525 | 43.413973 |
| | 2015-01-01 08:00:00 | 1 | 170.989902 | 527.349825 | 75.237905 | 34.178847 |
| | 2015-01-01 09:00:00 | 1 | 162.462833 | 346.149335 | 109.248561 | 41.122144 |
| | 2015-01-01 10:00:00 | 1 | 157.610021 | 435.376873 | 111.886648 | 25.990511 |



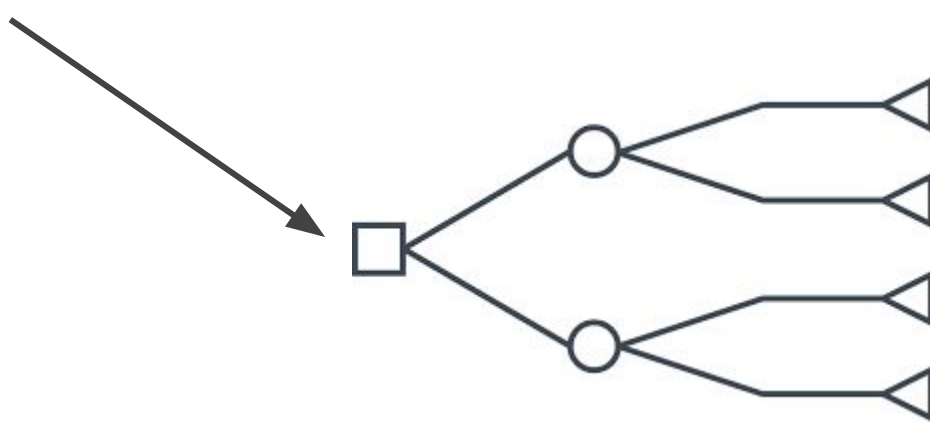
média por dia

| | maquina | data | voltagem | rotacao | pressao | vibracao |
|--|---------|------------|------------|------------|------------|-----------|
| | 1 | 2015-01-01 | 167.576533 | 440.515328 | 98.522345 | 40.049623 |
| | 1 | 2015-01-02 | 169.795758 | 446.832666 | 98.454608 | 39.271645 |
| | 1 | 2015-01-03 | 171.862244 | 459.204742 | 97.998233 | 48.074091 |
| | 1 | 2015-01-04 | 174.792428 | 448.743201 | 101.452266 | 52.190268 |
| | 1 | 2015-01-05 | 171.018408 | 454.822750 | 102.363114 | 43.330311 |



Incrementando o treino

| falha | idade | error1 | error2 | error3 | error4 | error5 | mean_voltagem | mean_rotacao | mean_pressao | mean_vibracao | model1 | model2 | model3 | model4 |
|-------|-------|--------|--------|--------|--------|--------|---------------|--------------|--------------|---------------|--------|--------|--------|--------|
| ok | 17 | 0 | 0 | 0 | 0 | 0 | 165.678044 | 436.241599 | 100.054536 | 38.053647 | 0 | 0 | 1 | 0 |
| ok | 17 | 0 | 0 | 0 | 0 | 0 | 167.984429 | 449.209447 | 100.395188 | 40.178632 | 0 | 0 | 1 | 0 |
| ok | 17 | 0 | 0 | 0 | 0 | 0 | 173.296031 | 461.113762 | 96.854299 | 39.133315 | 0 | 0 | 1 | 0 |
| ok | 17 | 0 | 0 | 0 | 0 | 0 | 167.224425 | 450.492274 | 99.447270 | 40.829714 | 0 | 0 | 1 | 0 |
| ok | 17 | 1 | 0 | 0 | 0 | 0 | 173.641802 | 441.383124 | 98.744748 | 39.439531 | 0 | 0 | 1 | 0 |



- Interpretável
- Sem GridSearch
- Baixo Custo



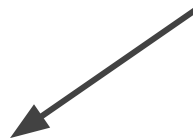
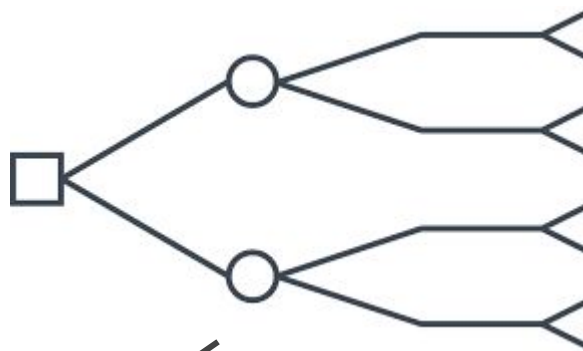
“ok”





Novas *features*

- Mínimo;
- Máximo;
- Desvio Padrão;
- RMS.



“ok”, “comp1”, “comp2”,
“comp3”, comp4”



Janelas de tempo

base

| | maquina | data | error1 | error2 | error3 | error4 | error5 | mean_voltagem | std_voltagem | min_voltagem | max_voltagem | rms_voltagem |
|---|---------|------------|--------|--------|--------|--------|--------|---------------|--------------|--------------|--------------|--------------|
| 0 | 1 | 2015-01-06 | NaN | NaN | NaN | NaN | NaN | 171.025033 | 17.623759 | 121.502718 | 213.787102 | 171.893040 |
| 1 | 1 | 2015-01-07 | NaN | NaN | NaN | NaN | NaN | 174.139410 | 10.611608 | 151.641444 | 192.154544 | 174.448985 |
| 2 | 1 | 2015-01-08 | NaN | NaN | NaN | NaN | NaN | 172.135931 | 13.829868 | 143.551174 | 200.741045 | 172.667526 |
| 3 | 1 | 2015-01-09 | NaN | NaN | NaN | NaN | NaN | 169.350052 | 14.600352 | 143.506562 | 197.147081 | 169.952136 |
| 4 | 1 | 2015-01-10 | 0.0 | 0.0 | 0.0 | 1.0 | 0.0 | 171.006246 | 12.461183 | 146.279048 | 193.249751 | 171.440798 |

treino (janela de 5 dias)

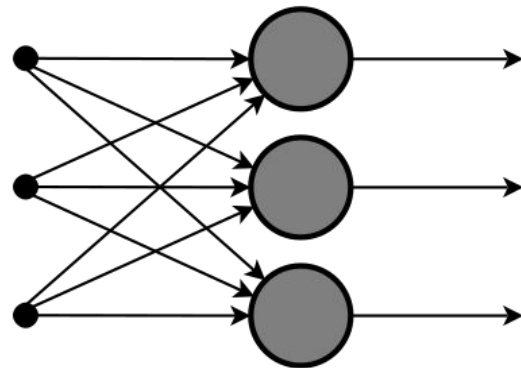
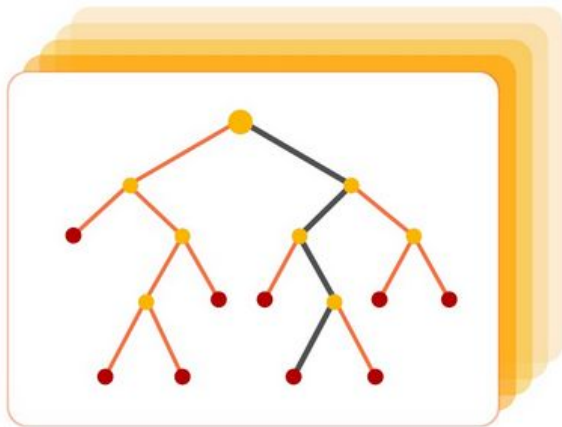
| | maquina | data | error1 | error2 | error3 | error4 | error5 | mean_voltagem | std_voltagem | min_voltagem | max_voltagem | rms_voltagem |
|-------|---------|------------|--------|--------|--------|--------|--------|---------------|--------------|--------------|--------------|--------------|
| 21701 | 1 | 2015-01-10 | 0 | 0 | 0 | 1 | 0 | 171.531334 | 13.825354 | 141.296189 | 199.415904 | 172.080497 |



Classificadores (*GridSearchCV*)

- sem janela;
- janelas de 1 a 5 dias.

XGBoost



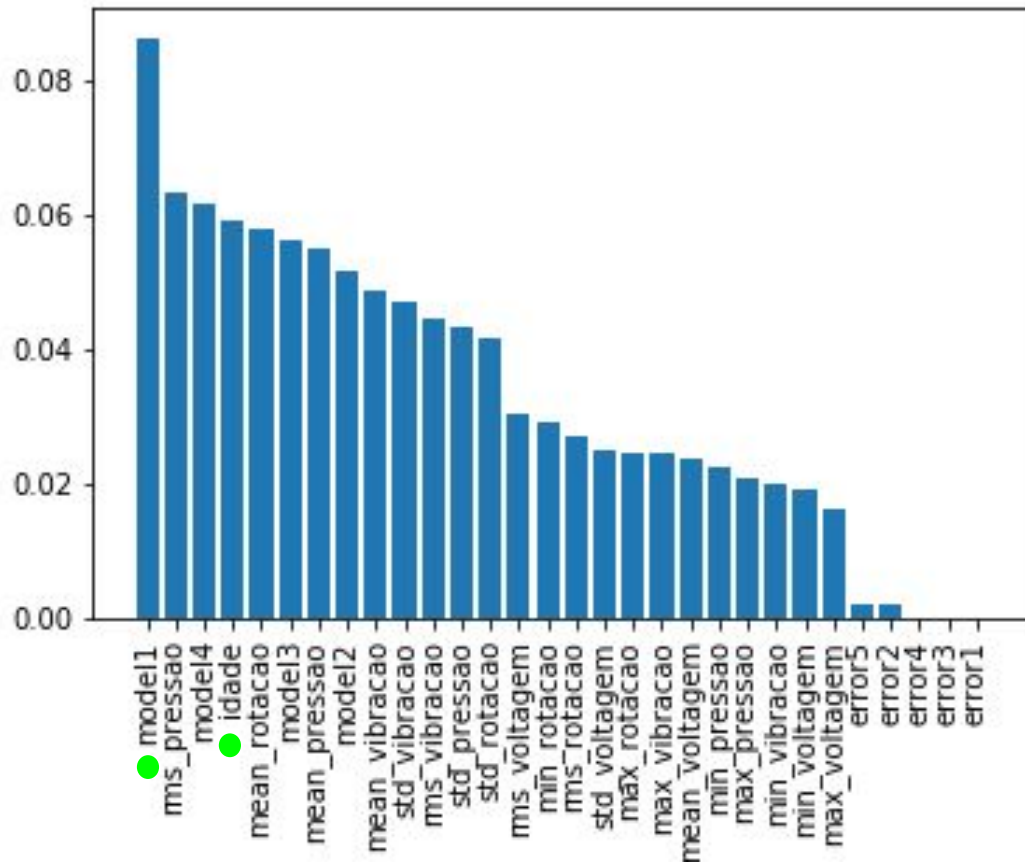


Sem janelamento

XGBoost

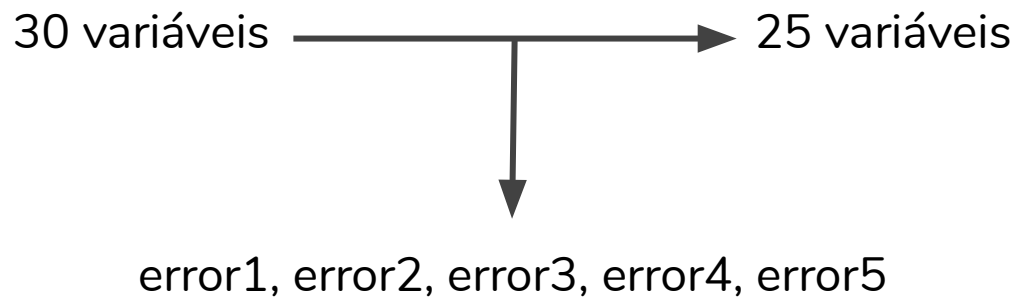
| | |
|----------------------|------|
| <i>learning_rate</i> | 0.05 |
| <i>max_depth</i> | 7 |
| <i>max_features</i> | 5 |
| <i>n_estimators</i> | 15 |

- F1 treino : 0.98015
- F1 teste : 0.97816





RFE com K-fold





Perspectivas de melhoras

- Pensar agrupamentos de outras maneiras (variáveis, dinâmica);
- Reamostrar base de dados (natural, SMOTE, etc...);
- PCA;
- Apurar modelos.