

Dados utilizados

NASA Turbofan Predictive Maintenance, Composto por **múltiplas séries multivariadas**.

**Dados:** Log de 27 sensores para 290 motores

**Target:** RUL - Remaining Usefull Lifecycle. Quanto ciclos de vida aquele motor tem antes de falhar

Modelagem do problema

Como predição

Prever o RUL de cada motor dadas as leituras dos sensores ao longo do tempo

Como classificação

Para este trabalho, o problema foi transformado em um problema de classificação multilabel:

label 0: RUL> 60

label 1: 20 a 60:

label 2: RUL<= 20

Preparação dos dados

- Transformação de predição para classificação
- Separação das séries por motor - manipulação intensa dos dados em um np.array de 3 dimensões
- Transposição dos dados para o formato de entrada nos algoritmos selecionados

Feature Engineering

De acordo com a representação escolhida

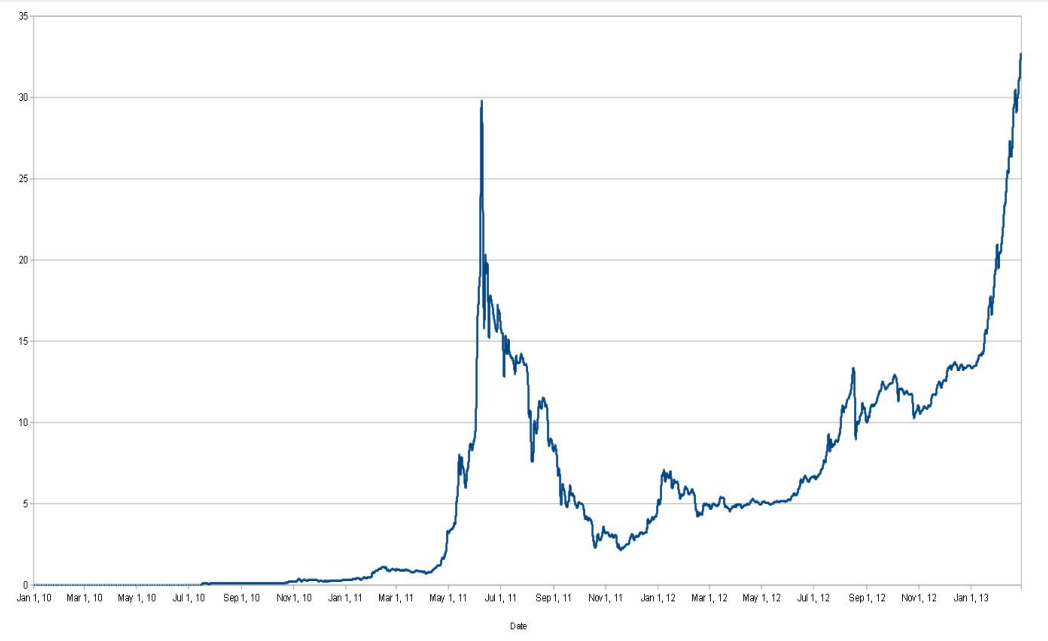


O que é uma série temporal?

É um dado tabular que é coletado várias vezes ao longo do tempo. São basicamente logs

Série UNIVARIADA

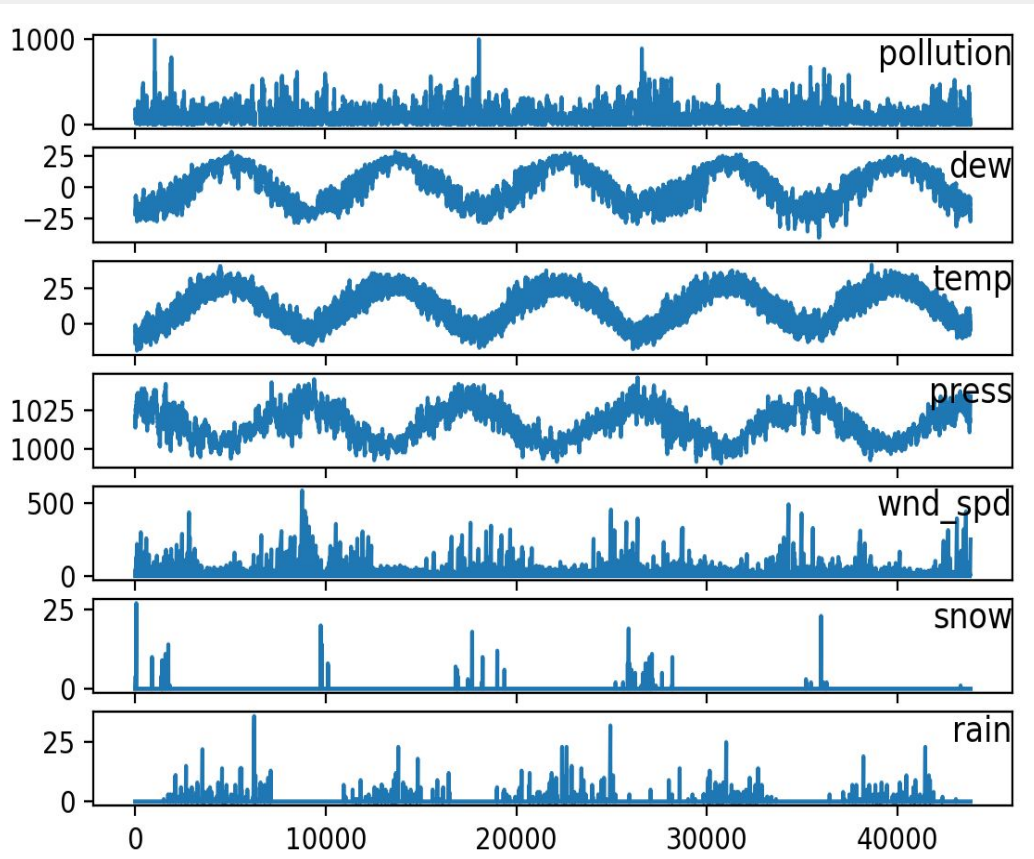
Um valor de interesse que varia ao longo do tempo.



Variação do preço do bitcoin entre 2010 e 2013

Série MULTIVARIADA

Temos **diversos** valores de interesse. Cada indicador é uma série temporal univariada



Variação de indicadores climáticos na embaixada dos EUA na China

# ConvNets para extração automática de features para séries temporais multivariadas - 2 estratégias (1 delas funciona)

A extração de features representa etapa muito importante no desenvolvimento de modelos de aprendizado de máquina. em grande empirismo, Entretanto, é também uma etapa que confia em empirismo, podendo

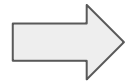
algumas features serem desconsideradas, além de ocupar tempo considerável do pipeline. A **geração automática de features** é um problema muito relevante no cenário atual do mercado de aprendizado de máquina.

----- BASELINE: Representação tabular -----

Modelo  
XGBoost

Acurácia  
86%

id_motor	timestamp	sensor_1	...	sensor_n
1	00:10:00	-0.0007	...	110.00
2	00:10:00	0.0004	...	120.78
1	00:20:00	0.0104	...	100.15
3	00:10:00	0.0004	...	78.47
1	00:30:00	0.0015	...	50.78



id_motor	feature_1	feature_2	...	feature_n
1	0.15	0.78	...	0.35
2	0.25	0.90	...	0.24
3	0.17	0.85	...	0.22
...	...	...	...	...
n	0	0.7	...	0.53

Feature engineering

Linear Growing Lookback Window com extração de estatísticas básicas. **Vantagem:** Pode ser alimentada nos algoritmos estatísticos

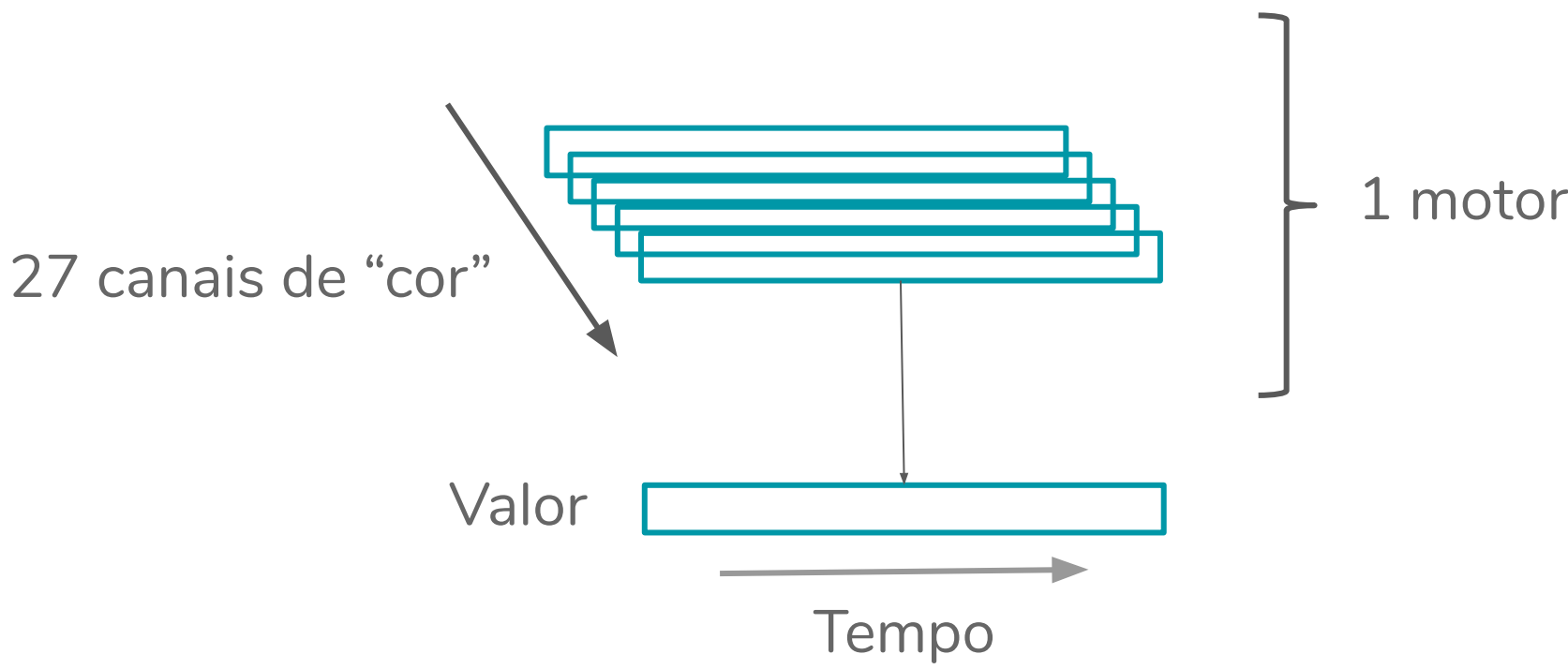
**Desvantagem:** extração manual de features

----- 1 Séries Temporais como imagens -----

Modelo

Conv2D + MaxPooling  
Conv2D + MaxPooling  
Dense + Dropout

Acurácia  
53%



Algo de errado não está certo

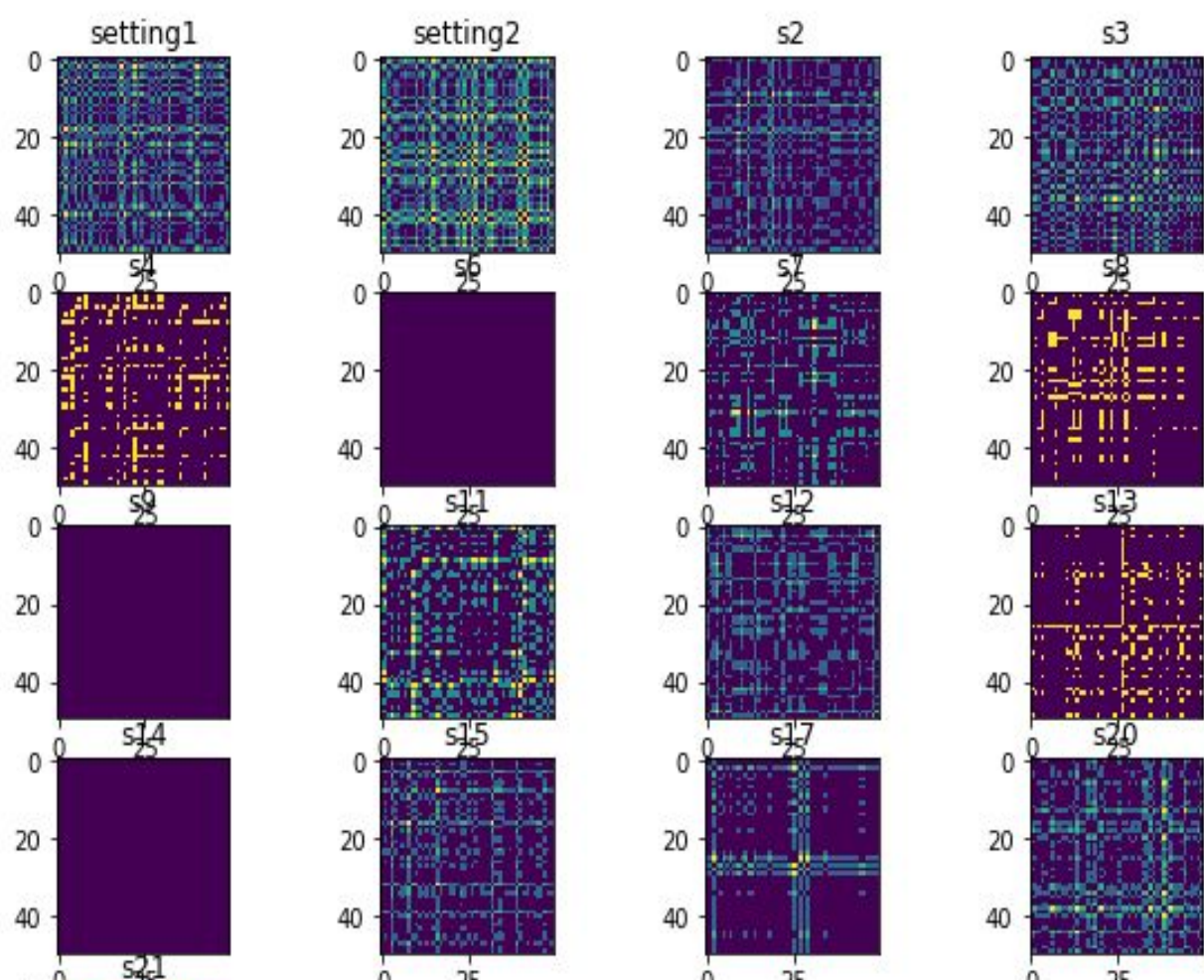
A grande diferença entre imagens e séries temporais é se a ordem importa. A CNN 2D assume que a ordem não importa (spacial invariance). O que passa longe de ser verdade no caso de séries temporais...

----- 2 Recurrence Plots -----

Modelo

Conv2D + MaxPooling  
Conv2D + MaxPooling  
Dense + Dropout

Acurácia  
92%



Recurrence Plots

Forma de visualização de séries temporais multivariadas que permite entender os padrões de repetição ao longo do tempo. “Recurrence plots, which provide a way to visualize the periodic nature of a trajectory through a phase space.” (Hatami et al, 2017)

Nima Hatami, Yann Gavet, Johan Debayle. **Classification of Time-Series Images Using Deep Convolutional Neural Networks**. 2017. Arxiv. <Disponível em: <https://arxiv.org/pdf/1710.00886.pdf>>