



Pré-processamento e Aprendizado de Máquina para Séries Temporais

Jorge

Pré-processamento

- Dados Ausentes
- Padronização
- Extração de características

Objetivo das técnicas de pré-processamento

- Tornar os dados mais precisos e significativos, levando a melhores soluções, decisões e resultados ao serem entregues para os modelos de AM
- Métodos que não distorcem sistematicamente os dados

Atenção: LOOKAHEAD

- Indica qualquer conhecimento do futuro da Série temporal.
- Ter as informações do “futuro” X_n , observando os dados em X_{n+1} pode ser tendencioso
- Falaremos disso na explicação das técnicas de Pré-processamento

1) Dados Ausentes - Imputação

- Forward Fill (mais simples computacionalmente)
- Interpolação
- K-NN (custo computacional maior)
- Média Móvel (reduz a variância e melhora o cálculo de R^2)

Comparação Geral: EQM

Forward Fill: 2.3454407622435633

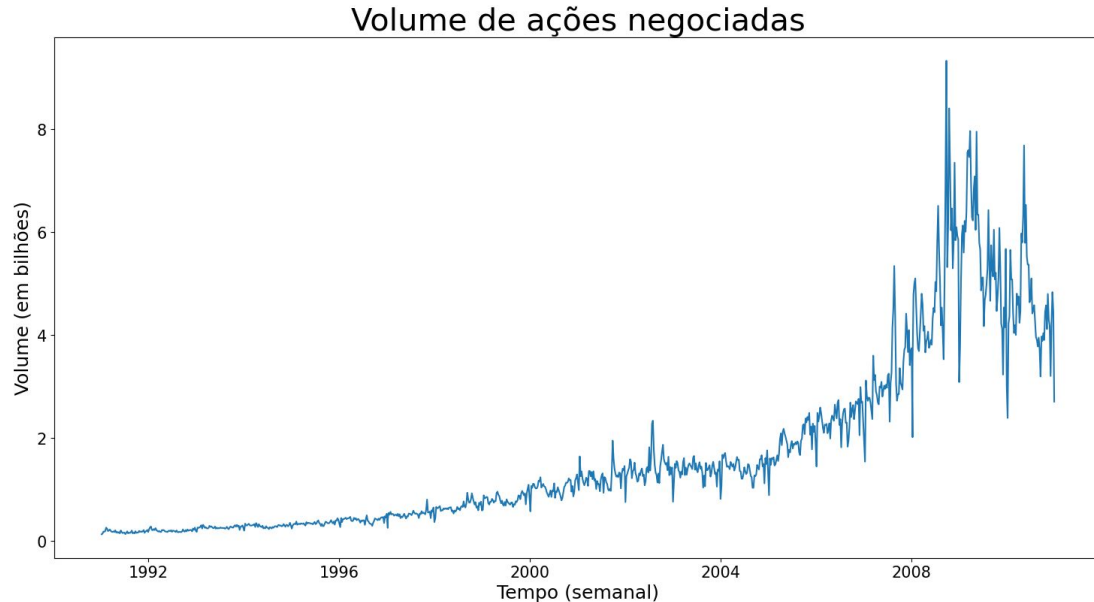
Interpolação Linear: 2.260973153369053

KNN: 1.6724754991671567

Média Móvel: 2.01100130628661

2) Padronização dos dados

- Modelos como Redes neurais possuem sensibilidade a magnitude dos dados fornecidos
- Por exemplo: Valores de fechamento semanal da bolsa de valores S&P 500 (EUA)

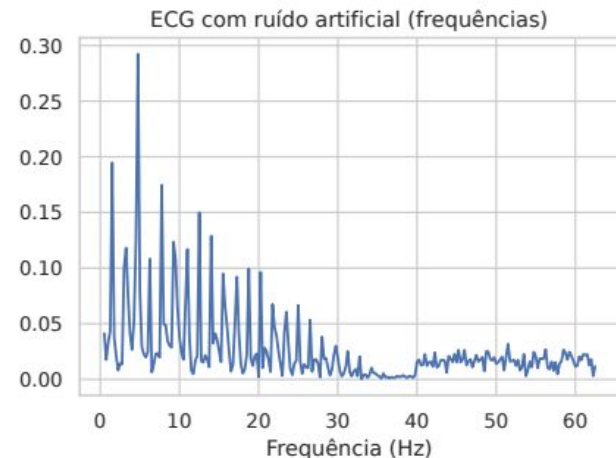
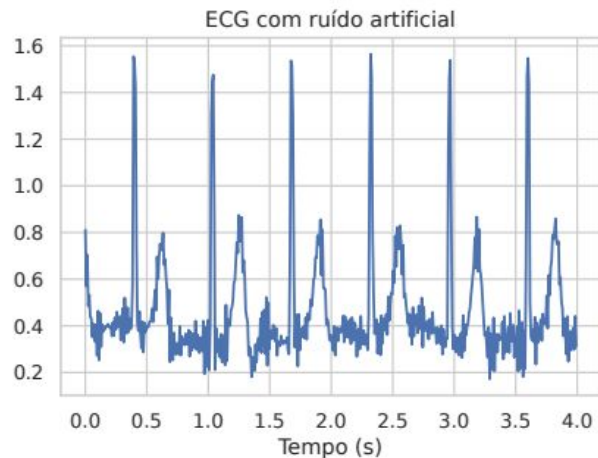
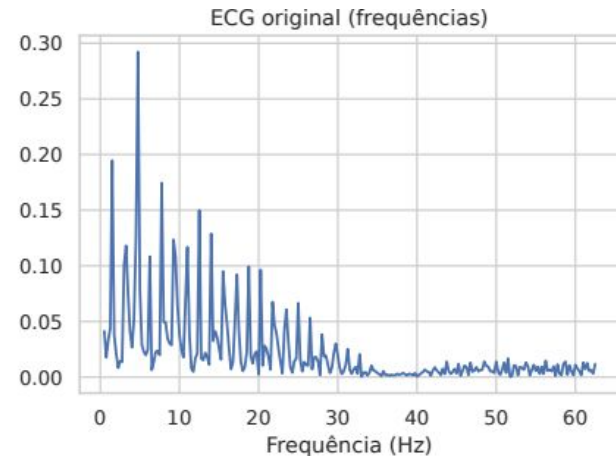
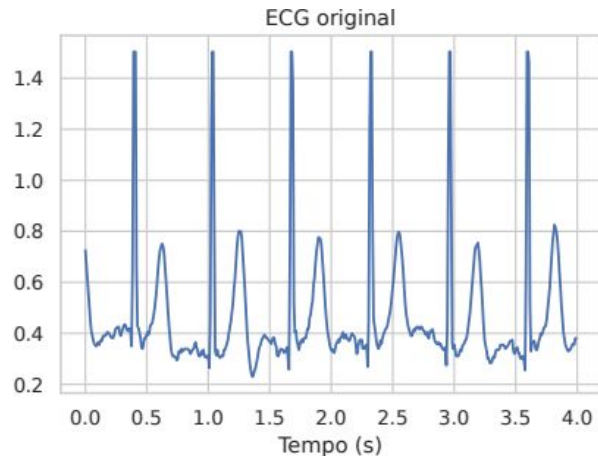


2) Padronização dos dados

```
import numpy as np
from sklearn.preprocessing import MinMaxScaler

scaler = MinMaxScaler(feature_range=(0,1))
train_data_norm = scaler_rnn.fit_transform(np.array(train_data_rnn).reshape(-1,1))
test_data_norm = scaler_rnn.transform(np.array(input_data_rnn).reshape(-1,1))
val_data_norm = scaler_rnn.transform(np.array(test_data_rnn).reshape(-1,1))
```

3) Filtragem



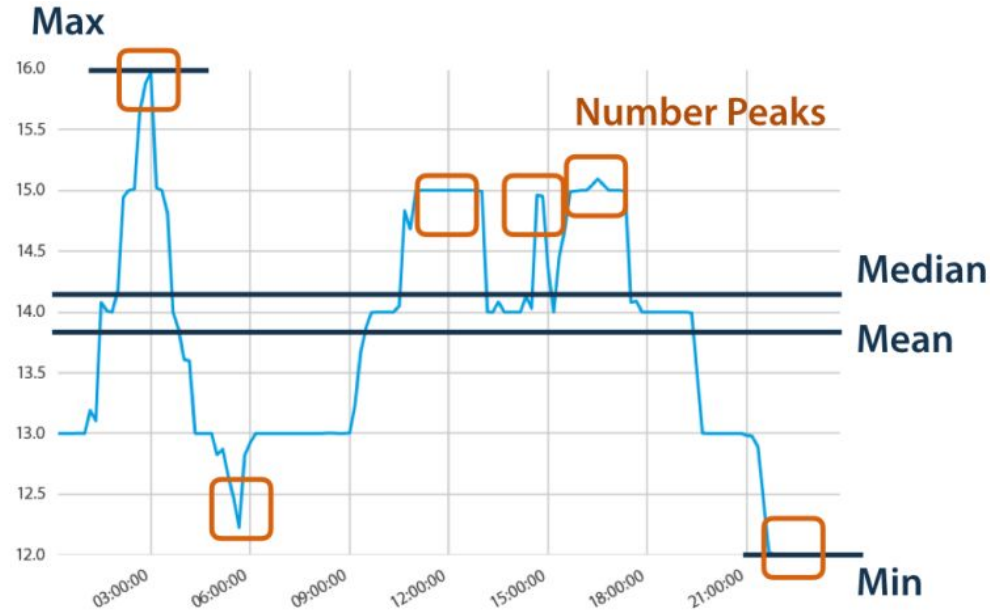
3) Extração e geração de características



Features:

- *tendência*
- *sazonalidade*
- *autocorrelação*
- *Trans. dom. freq.*

Fourier



Exemplo de extração de algumas características de uma série temporal

Aprendizado de máquina

- Cenários possíveis de aplicação:

Classificação

Clusterização

Classificação



Classificação e Clusterização

- Registro de EEG (Eletroencefalograma)
- Extrair características através dos 5 casos possíveis.
- Modelos baseados em árvores de decisão

