

Previsão da Temperatura do Ar no Brasil Utilizando Estações Meteorológicas e Modelos de Aprendizado de Máquina

Marciano Machado Saraiva

Pontifícia Universidade Católica de Minas Gerais
Especialização em Ciência de Dados e Big Data

21 de novembro de 2020

Roteiro

- 1 Introdução
- 2 Coleta de Dados
- 3 Processamento e Tratamento dos Dados
- 4 Análise e Exploração dos Dados
- 5 Criação dos Modelos de Aprendizado de Máquina
- 6 Apresentação dos Resultados
- 7 Conclusão e Trabalhos Futuros

Contextualização

Segundo recentes estudos, os setores mais afetados pela mudança na temperatura são:

- agricultura;
- vegetação;
- recursos hídricos;
- turismo.

O Problema Proposto

O objetivo deste trabalho é prever o comportamento da temperatura média do ar para um intervalo de um ano utilizando séries temporais de temperatura obtidas de estações meteorológicas distribuídas em todo o território brasileiro.

O Problema Proposto

Why?: Variações de temperatura podem ter impacto direto na produção agrícola, geração de energia, turismo e até na saúde da população.

O Problema Proposto

Why?: Variações de temperatura podem ter impacto direto na produção agrícola, geração de energia, turismo e até na saúde da população.

Who?: Os dados foram coletados das estações convencionais e automáticas do Instituto Nacional de Meteorologia do Brasil (INMET) e do Laboratório de Meteorologia da Universidade Federal do Vale do São Francisco (LabMet).

O Problema Proposto

Why?: Variações de temperatura podem ter impacto direto na produção agrícola, geração de energia, turismo e até na saúde da população.

Who?: Os dados foram coletados das estações convencionais e automáticas do Instituto Nacional de Meteorologia do Brasil (INMET) e do Laboratório de Meteorologia da Universidade Federal do Vale do São Francisco (LabMet).

What?: Prever o comportamento da temperatura média do ar para um intervalo de um ano utilizando séries temporais de temperatura obtidas de estações meteorológicas espalhadas por todo o território brasileiro.

O Problema Proposto

Why?: Variações de temperatura podem ter impacto direto na produção agrícola, geração de energia, turismo e até na saúde da população.

Who?: Os dados foram coletados das estações convencionais e automáticas do Instituto Nacional de Meteorologia do Brasil (INMET) e do Laboratório de Meteorologia da Universidade Federal do Vale do São Francisco (LabMet).

What?: Prever o comportamento da temperatura média do ar para um intervalo de um ano utilizando séries temporais de temperatura obtidas de estações meteorológicas espalhadas por todo o território brasileiro.

Where?: Estações meteorológicas espalhadas por todo o território brasileiro.

O Problema Proposto

Why?: Variações de temperatura podem ter impacto direto na produção agrícola, geração de energia, turismo e até na saúde da população.

Who?: Os dados foram coletados das estações convencionais e automáticas do Instituto Nacional de Meteorologia do Brasil (INMET) e do Laboratório de Meteorologia da Universidade Federal do Vale do São Francisco (LabMet).

What?: Prever o comportamento da temperatura média do ar para um intervalo de um ano utilizando séries temporais de temperatura obtidas de estações meteorológicas espalhadas por todo o território brasileiro.

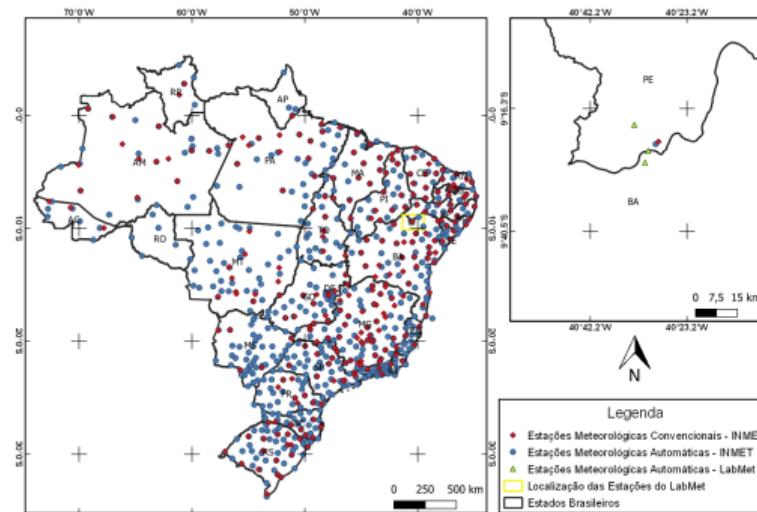
Where?: Estações meteorológicas espalhadas por todo o território brasileiro.

When?: De 1961 à 2019 para o conjunto de dados das estações convencionais do INMET, de 2000 à 2019 para os dados obtidos das estações automáticas do INMET e de 2007 à 2019 para dados obtidos das estações automáticas do LabMet.

Coleta de Dados

Figura 1: Distribuição espacial das estações utilizadas.

Conjunto de Dados	Qtd. Estações	Período	Qtd. Registros
Estações Convencionais do INMET	265	1961 - 2019	~12M
Estações Automáticas do INMET	610	2000 - 2019	~54M
Estações Automáticas do LabMet	3	2007 - 2019	~9K

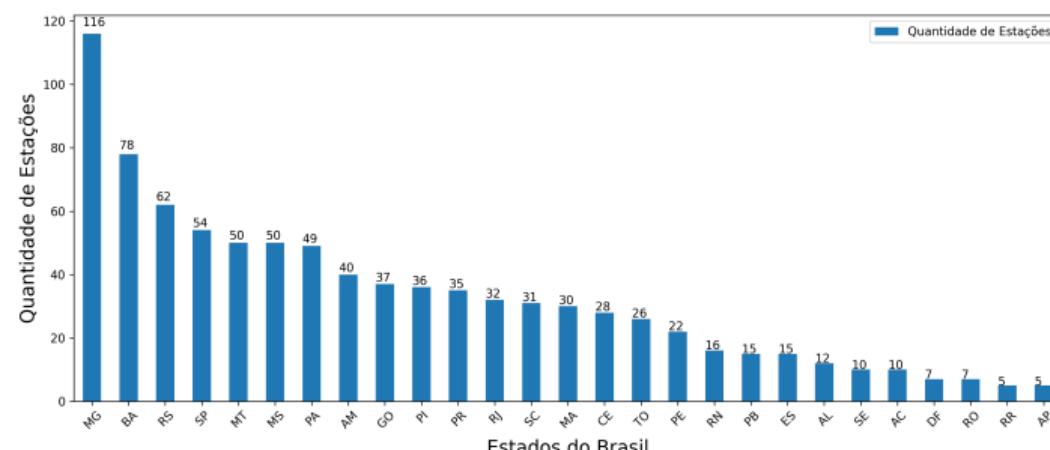


Processamento e Tratamento dos Dados

- Conversão dos conjuntos de dados para formato estruturado;
- Conversão dos conjuntos de dados para frequência diária;
- Eliminando valores inconsistentes: inconsistências de limites, inconsistências lógicas e inconsistências temporais.

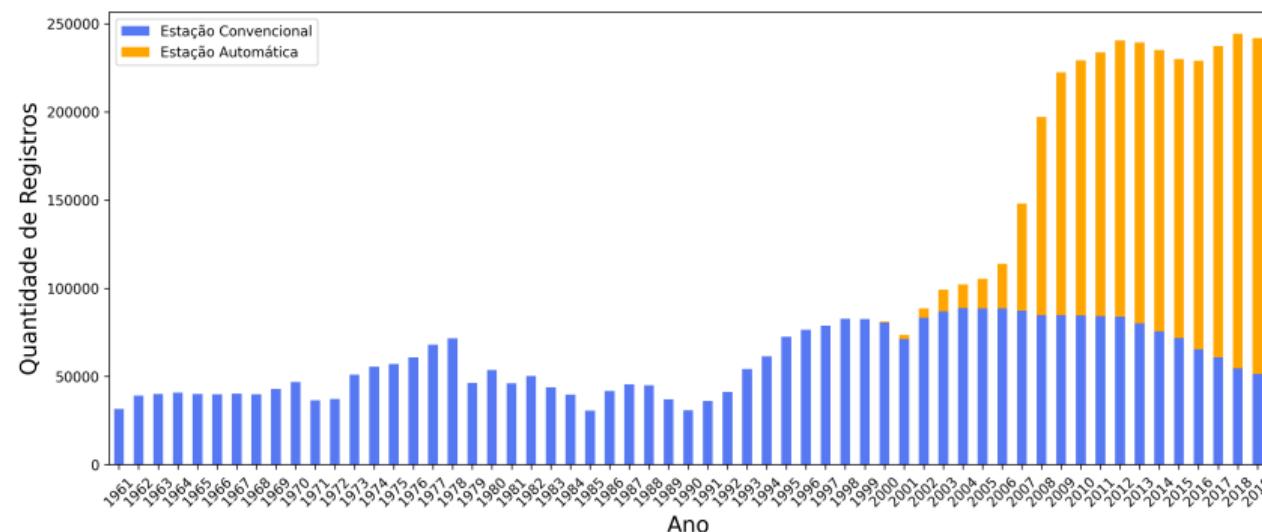
Análise e Exploração dos Dados

Figura 2: Distribuição, por estado, das estações meteorológicas analisadas.



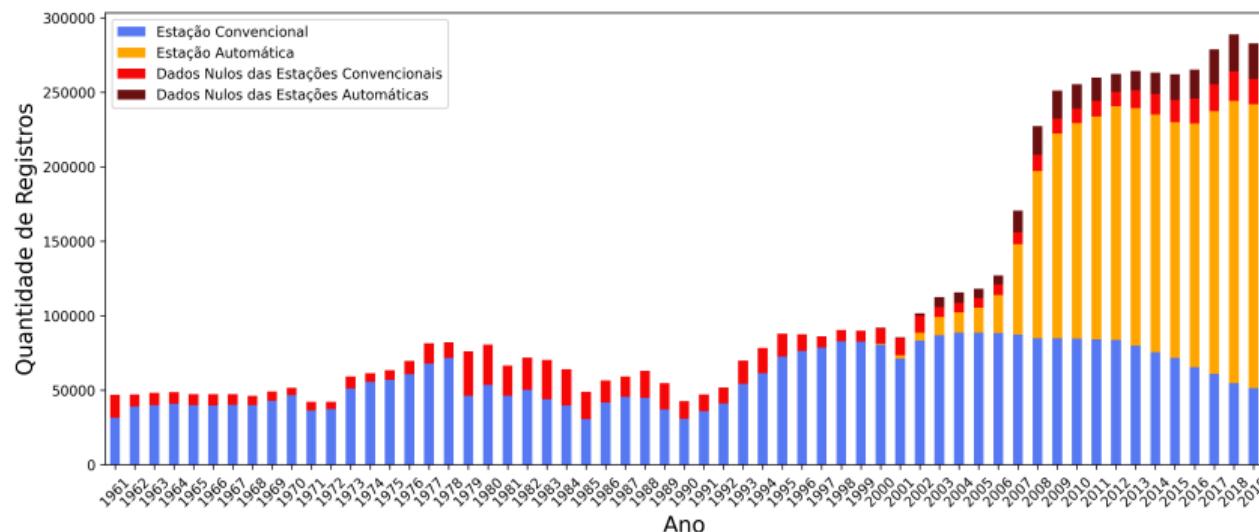
Análise e Exploração dos Dados

Figura 3: Quantidade de registros disponíveis ao longo de todo o período analisado.



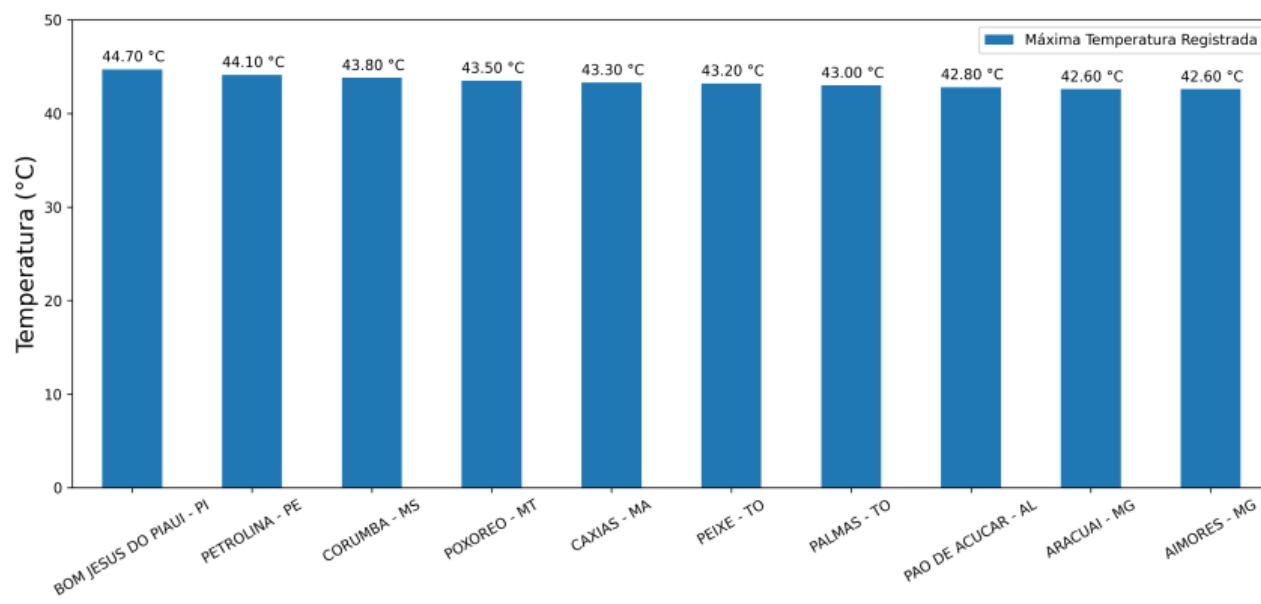
Análise e Exploração dos Dados

Figura 4: Quantidade de registos disponíveis e nulos ao longo de todo o período analisado.



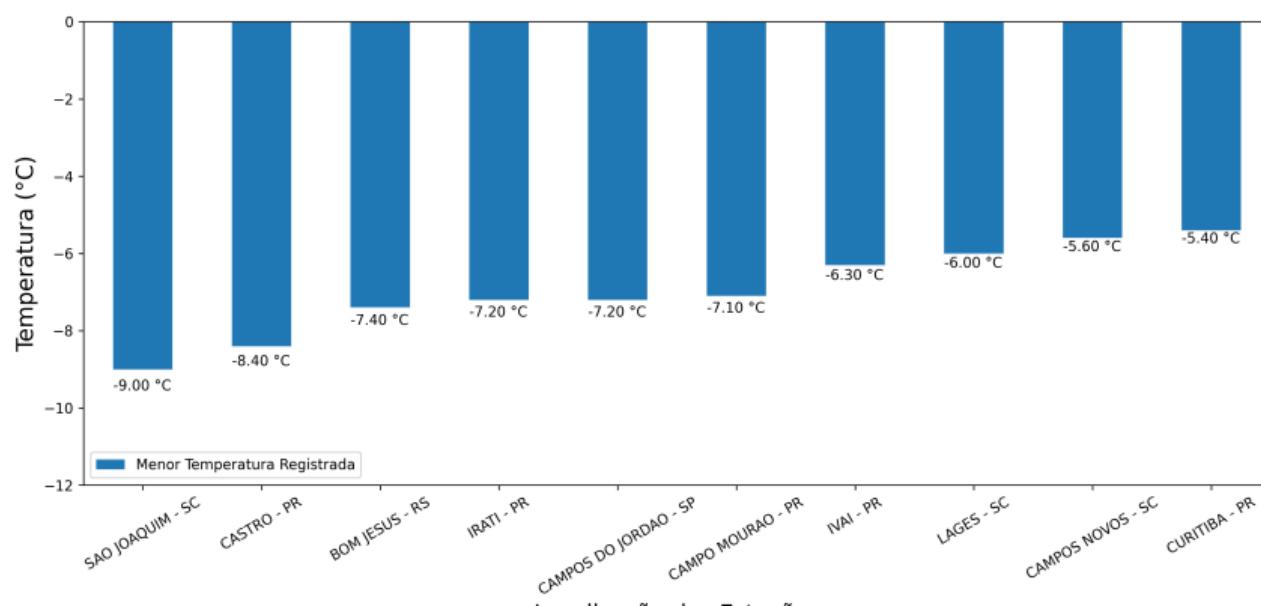
Análise e Exploração dos Dados

Figura 5: Recorde das maiores temperaturas já registradas no período de 1961 a 2019 nas estações convencionais do INMET.



Análise e Exploração dos Dados

Figura 6: Recorde das menores temperaturas já registradas no período de 1961 a 2019 nas estações convencionais do INMET.

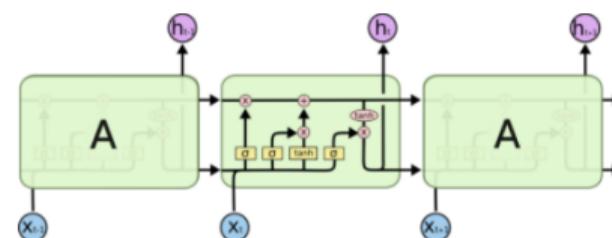


Criação dos Modelos de Aprendizado de Máquina

Figura 7: Autoregressive integrated moving average (ARIMA).

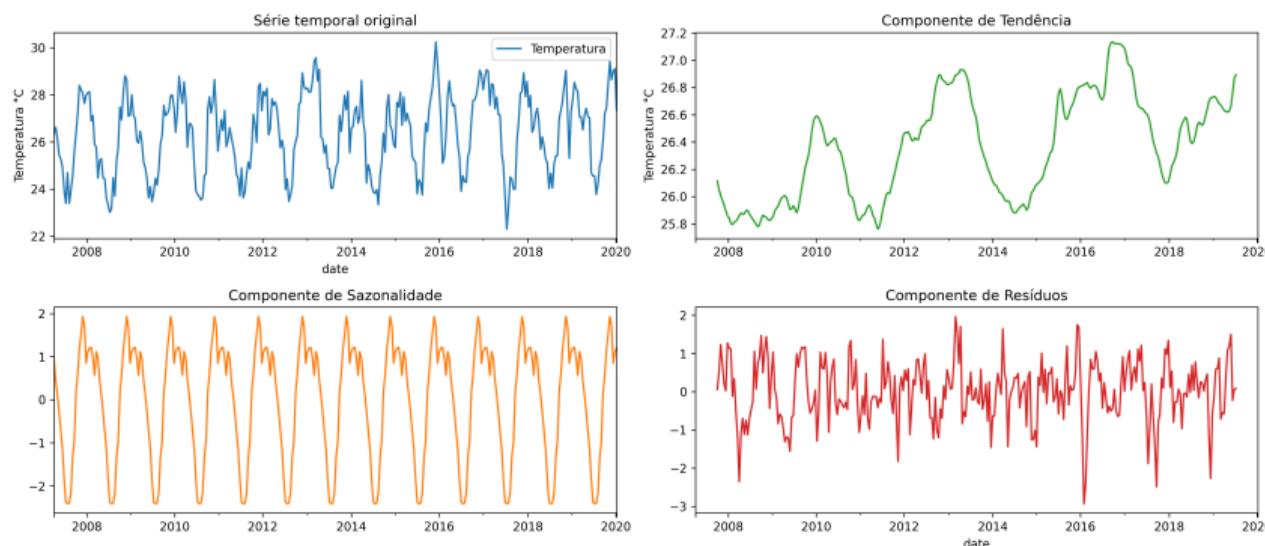


Figura 8: Long Short-Term Memory (LSTM).



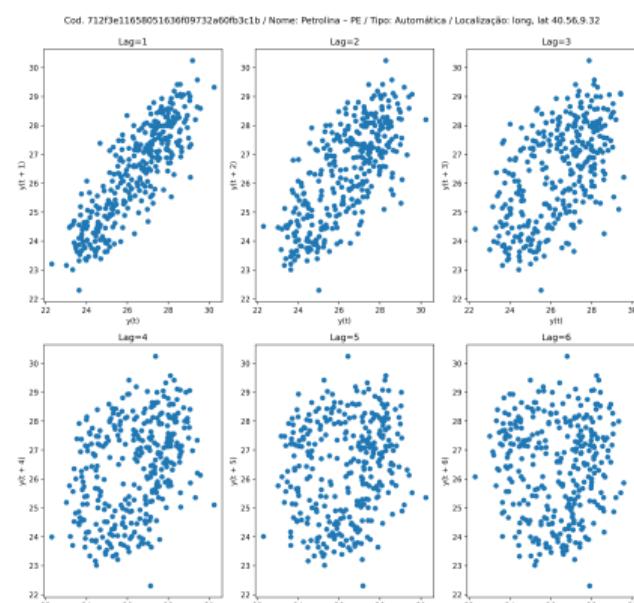
Modelo ARIMA

Figura 9: Exemplo da decomposição de uma série temporal.



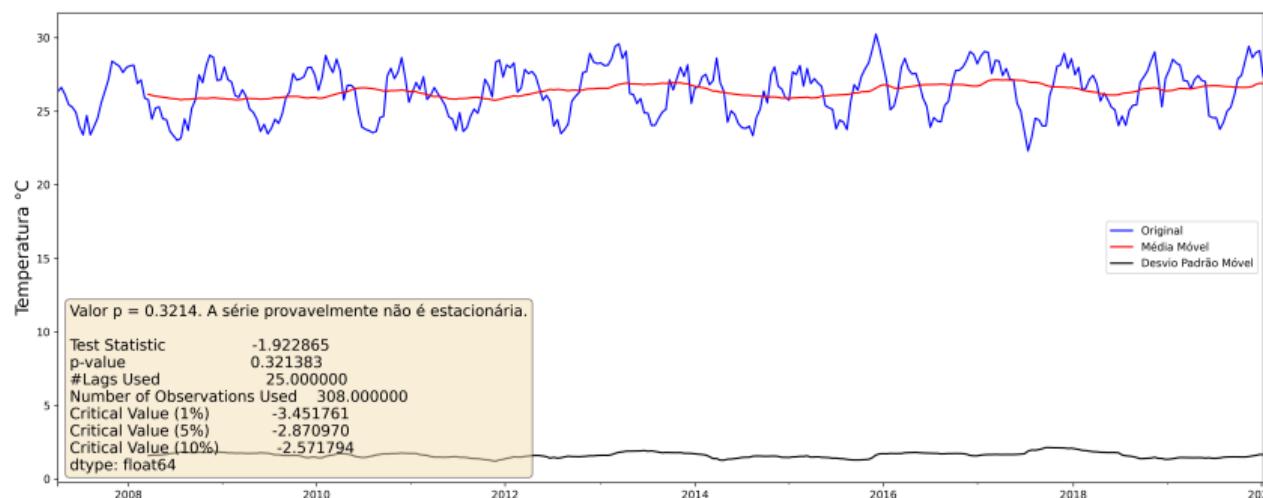
Modelo ARIMA

Figura 10: Exemplo de autocorrelação para uma série temporal.



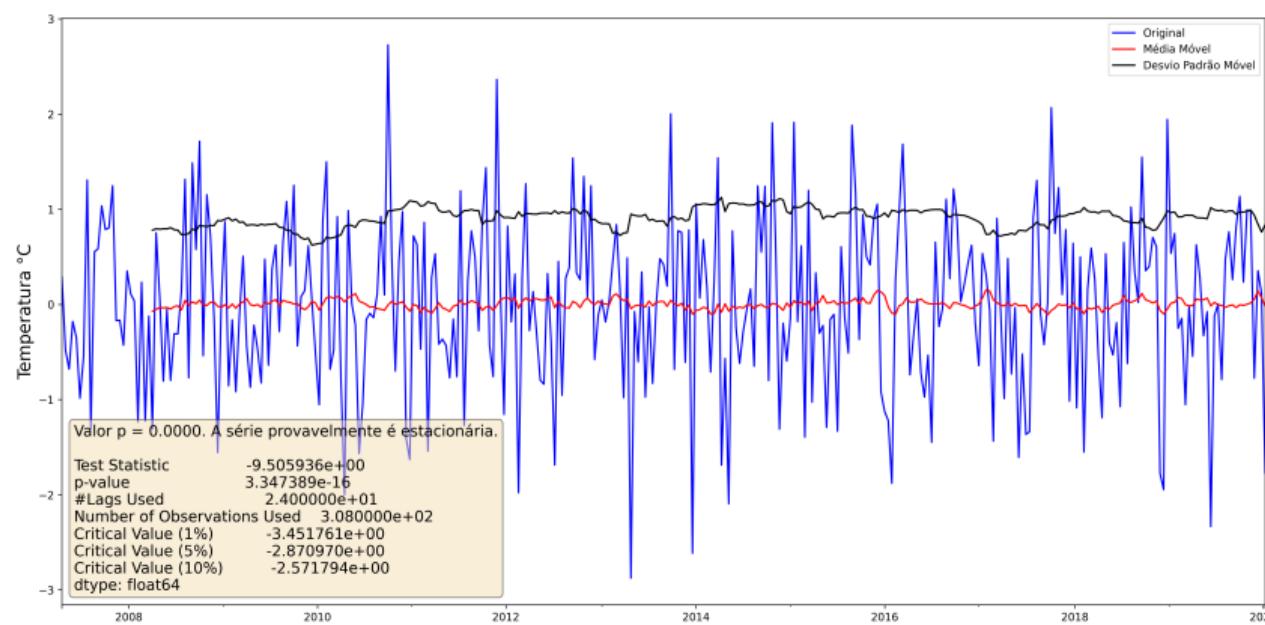
Modelo ARIMA

Figura 11: Exemplo de teste de estacionariedade à série temporal original.



Modelo ARIMA

Figura 12: Exemplo de teste de estacionariedade aplicado à série temporal diferenciada.



Modelo ARIMA

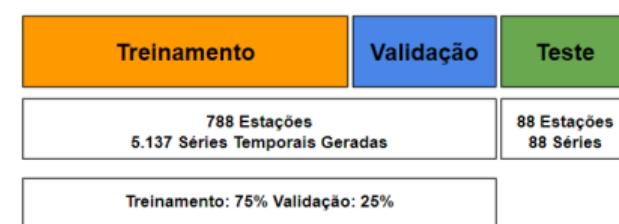
Figura 13: Parametrização do modelo auto arima e predição do ano de 2019 para as 88 séries temporais selecionadas de teste.

```
arima_model = auto_arima(train_data_diff,
                          start_p=1, start_q=1,
                          test='adf',
                          max_p=3, max_q=3, m=period,
                          start_P=0, seasonal=True,
                          d=None, D=1,
                          stepwise=True,
                          trace=True,
                          random_state=20, n_fits=10
)
```

Model LSTM

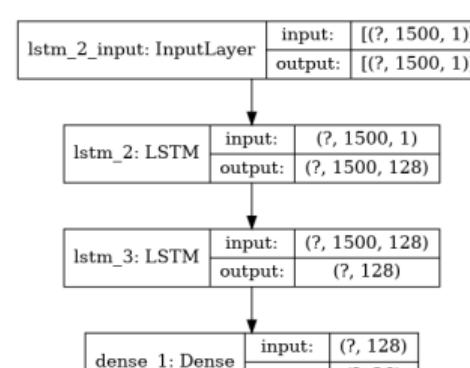
Período para obter o dado de entrada (X)	Período para obter a saída esperada (Y)
todos os dados anteriores à 01/01/2019	01/01/2019 - 31/12/2019
todos os dados anteriores à 01/01/2018	01/01/2018 - 31/12/2018
—	— — —
todos os dados anteriores à 01/01/2010	01/01/2010 - 31/12/2010

Figura 14: Dados de Treinamento, Validação e Teste.



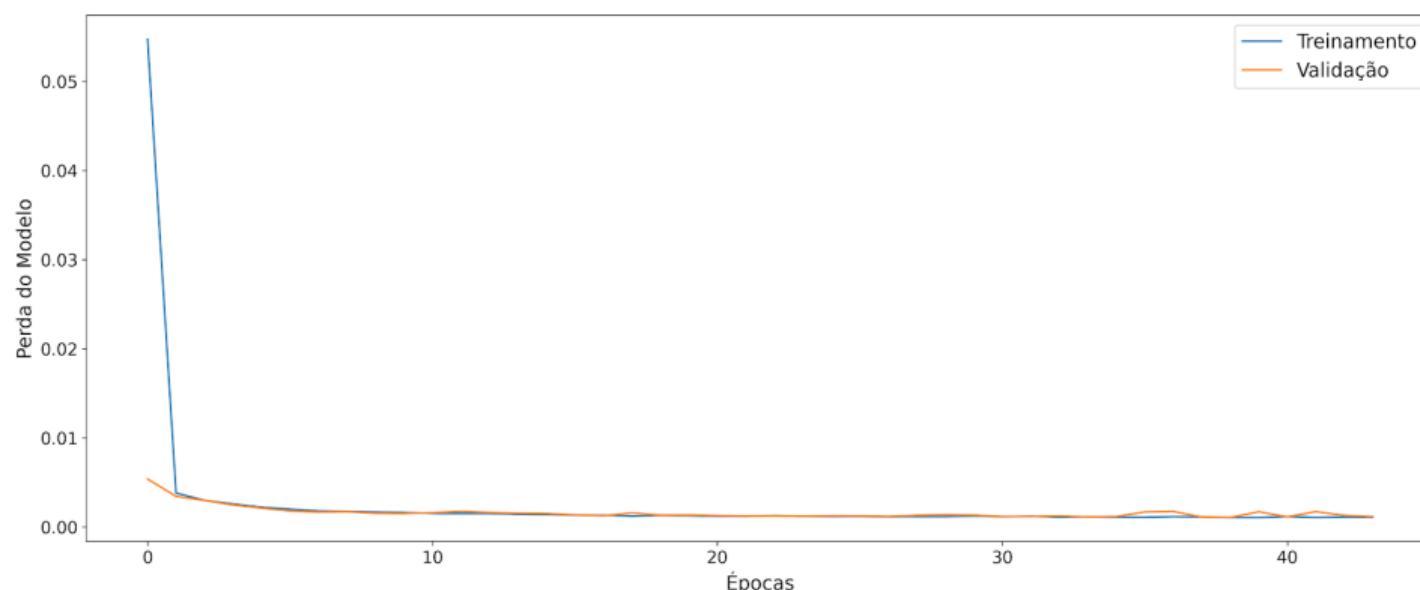
Model LSTM

Figura 15: Arquitetura LSTM desenvolvida.



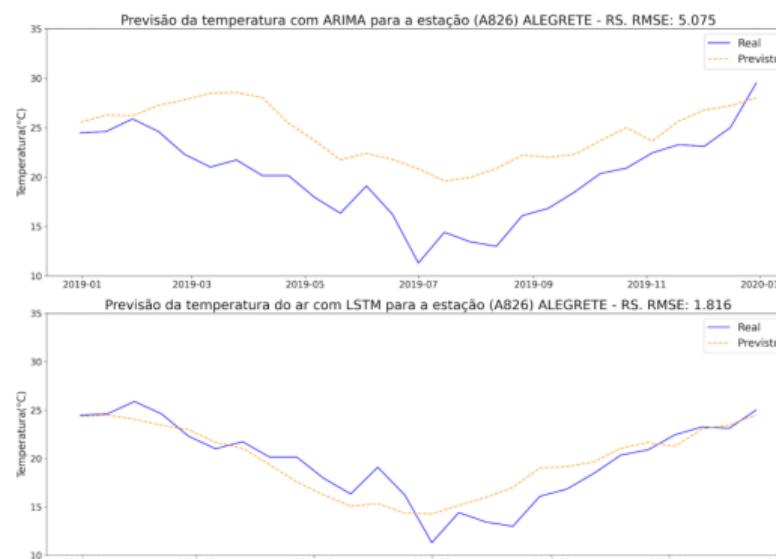
Model LSTM

Figura 16: Treinamento do Modelo.



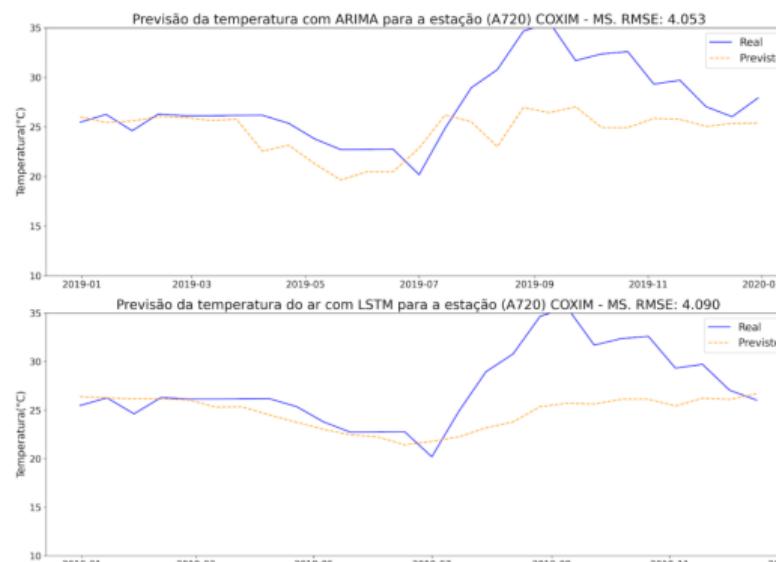
Apresentação dos Resultados

Figura 17: Resultado da previsão da temperatura do ar para o ano de 2019 na estação meteorológica localizada no município de Alegrete, no estado do Rio Grande do Sul.



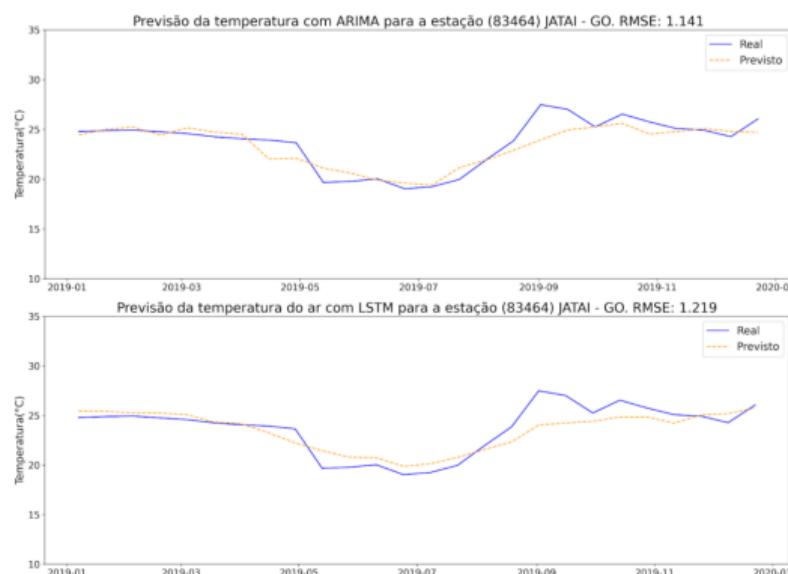
Apresentação dos Resultados

Figura 18: Resultado da previsão da temperatura do ar para o ano de 2019 na estação meteorológica localizada no município de Coxim, no estado do Mato Grosso do Sul.



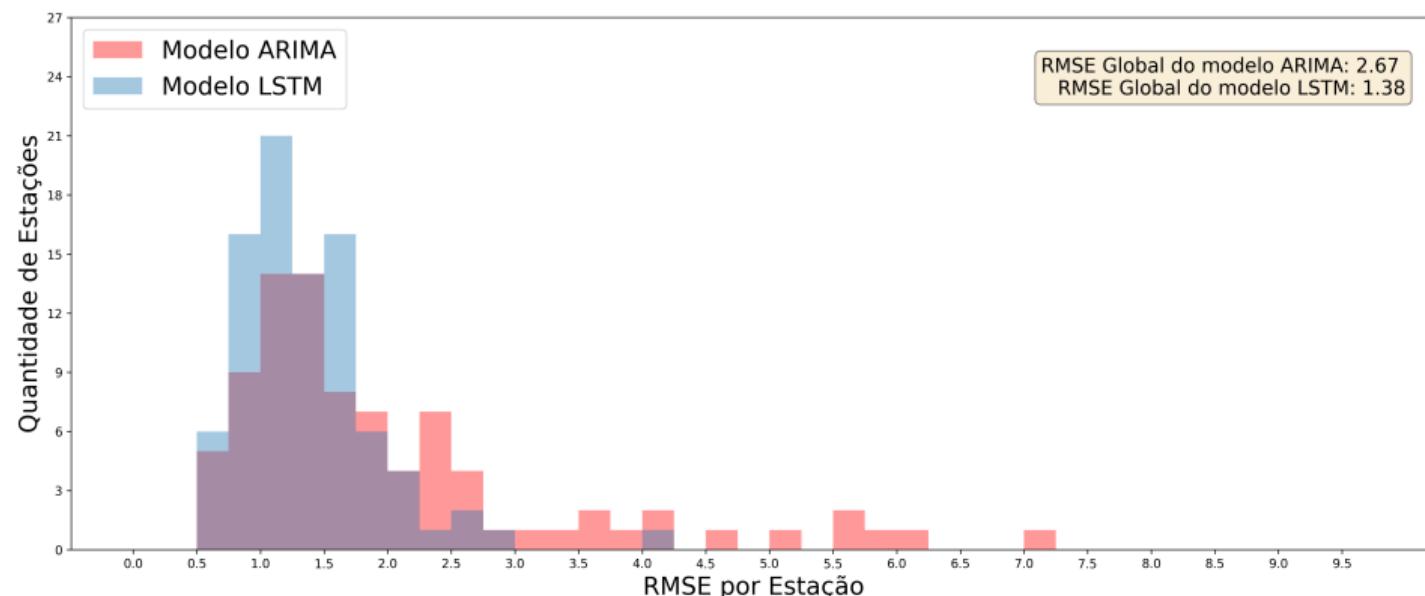
Apresentação dos Resultados

Figura 19: Resultado da previsão da temperatura do ar para o ano de 2019 na estação meteorológica localizada no município de Jataí, no estado de Goiás.



Apresentação dos Resultados

Figura 20: Comparação entre os valores obtidos de RMSE para as 88 estações avaliadas.



Conclusão e Trabalhos Futuros

- Modelos ARIMA são um dos métodos mais antigos e populares de previsão de séries temporais, conseguem obter bons resultados, mas possuem grande dependência em sua parametrização, o que dificulta sua operacionalização;
- Modelos baseados em RNNs, como a LSTM, possuem a capacidade de trabalhar com longas séries temporais e, em geral, possuem melhor generalização que modelos clássicos de ML, porém necessitam de grandes conjuntos de dados para o treinamento;
- A utilização de outras variáveis, como a precipitação, umidade relativa do ar, pressão atmosférica, radiação solar global e outras, podem ser utilizadas em trabalhos futuros para ajudar na previsão da temperatura do ar, já que muitas dessas variáveis são correlacionadas.



Obrigado!