

Análise Prática de Séries Temporais

Predição com Estatística e
Aprendizado de Máquina

Aileen Nielsen



Análise Prática de Séries Temporais

Copyright © 2021 da Starlin Alta Editora e Consultoria Eireli.
ISBN: 978-85-5081-562-6

Translated from original Practical Time Series Analysis. Copyright © 2020 by Aileen Nielsen. ISBN 978-1-492-04165-8. This translation is published and sold by permission of O'Reilly Media, Inc., the owner of all rights to publish and sell the same. PORTUGUESE language edition published by Starlin Alta Editora e Consultoria Eireli, Copyright © 2021 by Starlin Alta Editora e Consultoria Eireli.

Todos os direitos estão reservados e protegidos por Lei. Nenhuma parte deste livro, sem autorização prévia por escrito da editora, poderá ser reproduzida ou transmitida. A violação dos Direitos Autorais é crime estabelecido na Lei nº 9.610/98 e com punição de acordo com o artigo 184 do Código Penal.

A editora não se responsabiliza pelo conteúdo da obra, formulada exclusivamente pelo(s) autor(es).

Marcas Registradas: Todos os termos mencionados e reconhecidos como Marca Registrada e/ou Comercial são de responsabilidade de seus proprietários. A editora informa não estar associada a nenhum produto e/ou fornecedor apresentado no livro.

Impresso no Brasil — 1ª Edição, 2021 — Edição revisada conforme o Acordo Ortográfico da Língua Portuguesa de 2009.

Erratas e arquivos de apoio: No site da editora relatamos, com a devida correção, qualquer erro encontrado em nossos livros, bem como disponibilizamos arquivos de apoio se aplicáveis à obra em questão.

Acesse o site www.altabooks.com.br e procure pelo título do livro desejado para ter acesso às erratas, aos arquivos de apoio e/ou a outros conteúdos aplicáveis à obra.

Suporte Técnico: A obra é comercializada na forma em que está, sem direito a suporte técnico ou orientação pessoal/exclusiva ao leitor.

A editora não se responsabiliza pela manutenção, atualização e idioma dos sites referidos pelos autores nesta obra.

Dados Internacionais de Catalogação na Publicação (CIP) de acordo com ISBD

N669a	Nielsen, Aileen
Análise Prática de Séries Temporais: Predição com estatística e aprendizado de máquina / Aileen Nielsen ; traduzido por Cibelle Ravaglia. - Rio de Janeiro : Alta Books, 2021. 480 p. ; 17cm x 24cm.	
Tradução de: Practical Time Series Analysis Inclui índice. ISBN: 978-85-5081-562-6	
1. Ciência de dados. 2. Análise de dados. 3. Séries temporais. I. Ravaglia, Cibelle. II. Título.	
2021-3232	CDD 005.13 CDU 004.62

Elaborado por Odílio Hilário Moreira Junior - CRB-8/9949



Rua Viúva Cláudio, 291 — Bairro Industrial do Jacaré
CEP: 20.970-031 — Rio de Janeiro (RJ)
Tels.: (21) 3278-8069 / 3278-8419
www.altabooks.com.br — altabooks@altabooks.com.br

Produção Editorial
Editora Alta Books

Gerência Comercial
Daniele Fonseca

Editor de Aquisição
José Rugeri
acquisition@altabooks.com.br

Produtores Editoriais
Illysabelle Trajano
Maria de Lourdes Borges
Thales Silva
Thié Alves

Marketing Editorial
Livia Carvalho
Gabriela Carvalho
Thiago Brito
marketing@altabooks.com.br

Equipe de Design
Larissa Lima
Marcelli Ferreira
Paulo Gomes

Tradução
Cibelle Ravaglia

Copidesque
Mathheus Araújo

Diretor Editorial
Anderson Vieira

Coordenação Financeira
Solange Souza

Assistente Editorial
Caroline David

Equipe Ass. Editorial
Brenda Rodrigues
Luana Rodrigues
Mariana Portugal
Raquel Porto

Equipe Comercial
Adriana Baricelli
Daiana Costa
Fillipe Amorim
Kaïque Luiz
Victor Hugo Moraes
Viviane Paiva

Atuaram na edição desta obra:

Revisão Gramatical
Thamiris Leiroza
Hellen Suzuki

Diagramação
Joyce Matos

 **Ouvidoria:** ouvidoria@altabooks.com.br

Editora afiliada à:



abdr
ASSOCIAÇÃO BRASILEIRA DE
DIREITOS REPROGRÁFICOS

ASSOCIADO



Sumário

Prefácio	xi
Quem Deveria Ler Este Livro	xii
Conhecimento Esperado	xii
Por que Escrevi Este Livro	xiii
Roteiro do Livro	xiv
Recursos Online	xv
Convenções Usadas Neste Livro	xvi
Usando Exemplos de Códigos	xvii
Agradecimentos	xvii
1. Séries Temporais: Panorama e Breve História	1
A História das Séries Temporais em Diversas Aplicabilidades	1
Medicina Como um Problema de Séries Temporais	2
Previsões Meteorológicas	6
Previsão de Crescimento Econômico	7
Astronomia	9
A Análise de Séries Temporais Decola	10
As Origens da Análise Estatística de Séries Temporais	12
As Origens da Análise de Séries Temporais de Aprendizado de Máquina	13
Leituras e Recursos Adicionais	13
2. Encontrando Dados e Usando o Data Wrangling com Séries Temporais	15
Onde Encontrar Dados de Séries Temporais	16
Conjuntos de Dados Preparados	16
Séries Temporais Encontradas	22
Readaptando uma Coleção de Dados de Séries Temporais	
a Partir de um Conjunto de Tabelas	24
Um Exemplo Prático: Reunindo uma Coleção de Dados de Séries Temporais	24
Construindo uma Série Temporal Encontrada	30

Problemas com Timestamps	33
De Quem É Esse Tal de Timestamp?	33
Guesstimate de Timestamps para Compreender os Dados	34
O que É uma Escala de Tempo Significativa?	37
Limpe Seus Dados	38
Lidando com Dados Ausentes	38
Upsampling e Downsampling	49
Suavização de Dados	52
Dados Sazonais	57
Fusos Horários	60
Prevenindo o Lookahead	65
Leituras e Recursos Adicionais	66
3. Análise Exploratória de Dados para Séries Temporais	69
Métodos Conhecidos	69
Gráficos e Plotagens	70
Histogramas	73
Gráficos de Dispersão	74
Métodos Exploratórios Específicos de Séries Temporais	77
Compreendendo a Estacionariedade	78
Usando Funções de Janela	82
Compreendendo e Identificando a Self-correlation	87
Correlações Espúrias	98
Algumas Visualizações Úteis	100
Visualizações 1D	100
Visualizações 2D	101
Visualizações 3D	110
Leituras e Recursos Adicionais	113
4. Simulando Dados de Séries Temporais	115
Por que a Simulação de Séries Temporais É Especial?	116
Simulação versus Previsão	117
Simulações no Código	117
Faça Você Mesmo	118
Construindo um Universo de Simulação Autogerido	124
Simulação das Leis da Física	130
Apontamentos Finais sobre as Simulações	136
Simulações Estatísticas	137
Simulações de Aprendizado Profundo	137
Leituras e Recursos Adicionais	138

5. Armazenando Dados Temporais	139
Definindo os Requisitos	141
Live Data versus Dados Armazenados	142
Soluções de Banco de Dados	145
SQL versus NoSQL	145
Soluções Populares de Banco de Dados e de Arquivos de Séries Temporais	149
Soluções de Arquivo	153
NumPy	154
Pandas	155
Equivalentes Padrões ao R	155
Xarray	155
Leituras e Recursos Adicionais	156
6. Modelos Estatísticos para Séries Temporais.....	159
Por que Não Usar uma Regressão Linear?	159
Métodos Estatísticos Desenvolvidos para Séries Temporais	161
Modelos Autorregressivos	162
Modelos de Média Móvel	176
O modelo	177
Modelos Autorregressivos Integrados de Média Móvel	181
Autorregressão Vetorial	191
Variações em Modelos Estatísticos	196
Vantagens e Desvantagens dos Métodos Estatísticos para Séries Temporais	197
Leituras e Recursos Adicionais	198
7. Modelos de Espaço de Estados para Séries Temporais	201
Modelos de Espaço de Estados: Prós e Contras	203
O Filtro de Kalman	204
Visão Geral	204
Código para o Filtro Kalman	206
Modelos Ocultos de Markov	211
Como o Modelo Funciona	212
Como Ajustamos o Modelo	214
Ajustando um HMM no Código	218
Séries Temporais Estruturais Bayesianas	223
Código para o bsts	224
Leituras e Recursos Adicionais	228

8. Gerando e Selecionando Características para uma Série Temporal	231
Exemplo Introdutório	232
Considerações Gerais ao Calcular Características	233
A Natureza das Séries Temporais	234
Conhecimento do Domínio	235
Considerações Externas	235
Um Catálogo de Lugares para Encontrar Características e Se Inspirar	236
Bibliotecas Open Source para Gerar Características de Séries Temporais	237
Exemplos de Características Específicas de Domínio	242
Como Selecionar Características Depois de Gerá-las	245
Reflexões Finais	248
Leituras e Recursos Adicionais	249
9. Aprendizado de Máquina para Séries Temporais	251
Classificação de Séries Temporais	252
Selecionando e Gerando Características	252
Métodos de Árvores de Decisão	256
Clusterização	263
Gerando Características a Partir de Dados	265
Métricas de Distância com Reconhecimento Temporal	272
Código de Clusterização	277
Leituras e Recursos Adicionais	279
10. Aprendizado Profundo para Séries Temporais	281
Conceitos de Aprendizado Profundo	284
Programando uma Rede Neural	286
Dados, Símbolos, Operações, Camadas e Grafos	287
Construindo um Pipeline de Treinamento	291
Inspecionando Nosso Conjunto de Dados	291
Etapas de um Pipeline de Treinamento	295
Redes Feedforward	310
Um Simples Exemplo	311
Usando Mecanismo de Atenção para Tornar as	
Redes Feedforward Conscientes do Tempo	314
CNNs	316
Modelo Convolutacional Simples	318
Modelos Convolucionais Alternativos	320
RNNs	322
Continuando Nosso Exemplo de Energia Elétrica	325
A Inovação dos Autoencoders	326

Arquiteturas Combinadas	328
Resumindo	332
Leituras e Recursos Adicionais	333
11. Medição de Erros	335
Conceitos Básicos: Como Testar Previsões	336
Considerações Específicas do Modelo para Backtesting	339
Quando Sua Previsão É Boa o Suficiente?	340
Estimando a Incerteza em Seu Modelo com uma Simulação	342
Predizendo Múltiplos Passos à Frente	345
Ajuste Diretamente o Horizonte de Interesse	345
Abordagem Recursiva para Horizontes Temporais Distantes	346
Aprendizado Multitarefa Aplicado a Séries Temporais	346
Armadilhas de Validação de Modelo	347
Leituras e Recursos Adicionais	347
12. Considerações de Desempenho em Ajustes e Disponibilização de Modelos de Séries Temporais	349
Trabalhando com Ferramentas Construídas para Casos de Usos Gerais	350
Modelos Construídos para Dados Transversais Não “Compartilham” Dados entre Amostras	350
Modelos que Não São Pré-calculados Geram Lag Desnecessário entre a Medição de Dados e a Realização de uma Previsão	352
Formatos de Armazenamento de Dados: Vantagens e Desvantagens	353
Armazene Seus Dados em Formato Binário	354
Pré-processe Seus Dados de Forma que Você Possa “Deslizar” sobre Eles	354
Modificando Sua Análise para Se Adequar às Considerações de Desempenho	355
Usar Todos os Seus Dados Não É Necessariamente Melhor	355
Modelos Complicados nem Sempre São Melhores	356
Breve Menção de Ferramentas Alternativas de Alto Desempenho	357
Leituras e Recursos Adicionais	358
13. Aplicações na Área de Assistência Médica	359
Predizendo a Gripe	359
Caso de Estudo sobre a Gripe em uma Área Metropolitana	359
Qual É a Tecnologia de Ponta na Previsão da Gripe?	375
Predizendo os Índices Glicêmicos no Sangue	376
Limpeza e Exploração de Dados	377
Gerando Características	381
Ajustando um Modelo	387
Leituras e Recursos Adicionais	392

14. Aplicações na Área Financeira	393
Obtendo e Explorando Dados Financeiros	394
Pré-processamento de Dados Financeiros para Aprendizado Profundo	400
Adicionando Quantidades de Interesse aos Nossos Valores Brutos	400
Escalonando Quantidades de Interesse sem Lookahead	401
Formatando Nossos Dados para uma Rede Neural	404
Construindo e Treinando uma RNN	406
Leituras e Recursos Adicionais	413
15. Usos Governamentais de Séries Temporais	415
Obtendo os Dados Governamentais	416
Explorando Grandes Dados de Séries Temporais	418
Recorra ao Upsampling e Agregue os Dados à Medida que Os Iteramos	421
Classificando os Dados	422
Análise Estatística Online de Dados de Séries Temporais	426
Questões Pendentes	436
Outras Melhorias	437
Leituras e Recursos Adicionais	437
16. Pacotes de Séries Temporais	439
Previsão em Escala	439
Previsão Industrial In-house do Google	440
Pacote Prophet Open Source do Facebook	442
Detecção de Anomalia	447
Pacote AnomalyDetection Open Source do Twitter	447
Outros Pacotes de Séries Temporais	450
Leituras e Recursos Adicionais	451
17. Previsões sobre como Prever	453
Previsão como Serviço	453
O Aprendizado Profundo Aumenta as Possibilidades Probabilísticas	454
Importância Crescente do Aprendizado de Máquina em	
Detrimento da Estatística	455
Combinação Crescente de Metodologias Estatísticas e	
Aprendizado de Máquina	455
Mais Previsões para a Vida Cotidiana	456
Índice	457