

Hands-on Time Series Analysis with Python

**From Basics to Bleeding
Edge Techniques**

**B V Vishwas
Ashish Patel**

Apress®

Hands-on Time Series Analysis with Python: From Basics to Bleeding Edge Techniques

B V Vishwas
Infosys
Bengaluru, India

Ashish Patel
Cygnets Infotech Pvt Ltd
Ahmedabad, India

ISBN-13 (pbk): 978-1-4842-5991-7
<https://doi.org/10.1007/978-1-4842-5992-4>

ISBN-13 (electronic): 978-1-4842-5992-4

Copyright © 2020 by B V Vishwas and Ashish Patel

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

Trademarked names, logos, and images may appear in this Book. Rather than use a trademark symbol with every occurrence of a trademarked name, logo, or image we use the names, logos, and images only in an editorial fashion and to the benefit of the trademark owner, with no intention of infringement of the trademark.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

While the advice and information in this Book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Managing Director, Apress Media LLC: Welmoed Spahr
Acquisitions Editor: Celestin Suresh John
Development Editor: James Markham
Coordinating Editor: Aditee Mirashi

Cover designed by eStudioCalamar

Cover image designed by Pixabay

Distributed to the book trade worldwide by Springer Science+Business Media New York, 233 Spring Street, 6th Floor, New York, NY 10013. Phone 1-800-SPRINGER, fax (201) 348-4505, e-mail orders-ny@springer-sbm.com, or visit www.springeronline.com. Apress Media, LLC is a California LLC and the sole member (owner) is Springer Science + Business Media Finance Inc (SSBM Finance Inc). SSBM Finance Inc is a **Delaware** corporation.

For information on translations, please e-mail booktranslations@springernature.com; for reprint, paperback, or audio rights, please e-mail bookpermissions@springernature.com.

Apress titles may be purchased in bulk for academic, corporate, or promotional use. eBook versions and licenses are also available for most titles. For more information, reference our Print and eBook Bulk Sales web page at www.apress.com/bulk-sales.

Any source code or other supplementary material referenced by the author in this Book is available to readers on GitHub via the Book's product page, located at www.apress.com/978-1-4842-5991-7. For more detailed information, please visit www.apress.com/source-code.

Printed on acid-free paper

Table of Contents

About the Authors.....	xi
About the Technical Reviewer	xiii
Acknowledgments	xv
Introduction	xvii
Chapter 1: Time-Series Characteristics.....	1
Types of Data	2
Time-Series Data.....	2
Cross-Section Data.....	4
Panel Data/Longitudinal Data	4
Trend	6
Detecting Trend Using a Hodrick-Prescott Filter.....	6
Detrending a Time Series	7
Seasonality	11
Multiple Box Plots.....	12
Autocorrelation Plot.....	13
Deseasoning of Time-Series Data	14
Seasonal Decomposition	15
Cyclic Variations.....	16
Detecting Cyclical Variations	17
Errors, Unexpected Variations, and Residuals.....	18
Decomposing a Time Series into Its Components.....	18
Summary.....	21

TABLE OF CONTENTS

Chapter 2: Data Wrangling and Preparation for Time Series.....23

 Loading Data into Pandas 24

 Loading Data Using CSV 24

 Loading Data Using Excel 25

 Loading Data Using JSON..... 25

 Loading Data from a URL..... 26

 Exploring Pandasql and Pandas Side by Side 27

 Selecting the Top Five Records 27

 Applying a Filter..... 28

 Distinct (Unique) 29

 IN 30

 NOT IN..... 32

 Ascending Data Order 33

 Descending Data Order 34

 Aggregation..... 35

 GROUP BY 36

 GROUP BY with Aggregation 38

 Join (Merge)..... 39

 INNER JOIN 41

 LEFT JOIN 43

 RIGHT JOIN 46

 OUTER JOIN 48

 Summary of the DataFrame 51

 Resampling 52

 Resampling by Month..... 53

 Resampling by Quarter 53

 Resampling by Year 53

Resampling by Week	54
Resampling on a Semimonthly Basis	55
Windowing Function	55
Rolling Window	56
Expanding Window	57
Exponentially Weighted Moving Window	57
Shifting.....	58
Handling Missing Data	60
BFILL.....	62
FFILL	62
FILLNA	63
INTERPOLATE	64
Summary.....	64
Chapter 3: Smoothing Methods	65
Introduction to Simple Exponential Smoothing.....	66
Simple Exponential Smoothing in Action	68
Introduction to Double Exponential Smoothing.....	76
Double Exponential Smoothing in Action.....	78
Introduction to Triple Exponential Smoothing	86
Triple Exponential Smoothing in Action	87
Summary.....	97
Chapter 4: Regression Extension Techniques for Time-Series Data ...	99
Types of Stationary Behavior in a Time Series.....	99
Making Data Stationery	101
Using Plots.....	101
Using Summary Statistics	101
Using Statistics Unit Root Tests	102

TABLE OF CONTENTS

Interpreting the P-value	104
Augmented Dickey-Fuller Test	104
Kwiatkowski-Phillips-Schmidt-Shin Test	105
Using Stationary Data Techniques	106
Differencing	106
Random Walk.....	107
First-Order Differencing (Trend Differencing)	108
Second-Order Differencing (Trend Differencing)	109
Seasonal Differencing	110
Autoregressive Models	111
Moving Average.....	113
Autocorrelation and Partial Autocorrelation Functions	116
Introduction to ARMA	117
Autoregressive Model.....	118
Moving Average	119
Introduction to Autoregressive Integrated Moving Average	119
The Integration (I)	121
ARIMA in Action	122
Introduction to Seasonal ARIMA.....	129
SARIMA in Action	131
Introduction to SARIMAX.....	143
SARIMAX in Action	144
Introduction to Vector Autoregression.....	154
VAR in Action	155
Introduction to VARMA	172
VARMA in Action	172
Summary.....	184

Chapter 5: Bleeding-Edge Techniques	185
Introduction to Neural Networks.....	185
Perceptron	186
Activation Function	188
Binary Step Function	188
Linear Activation Function	189
Nonlinear Activation Function.....	190
Forward Propagation.....	195
Backward Propagation.....	196
Gradient Descent Optimization Algorithms.....	199
Learning Rate vs. Gradient Descent Optimizers	199
Recurrent Neural Networks.....	202
Feed-Forward Recurrent Neural Network	204
Backpropagation Through Time in RNN	206
Long Short-Term Memory	210
Step-by-Step Explanation of LSTM.....	212
Peephole LSTM.....	214
Peephole Convolutional LSTM	215
Gated Recurrent Units.....	216
Convolution Neural Networks	219
Generalized CNN Formula.....	222
One-Dimensional CNNs	223
Auto-encoders	224
Summary.....	226

TABLE OF CONTENTS

Chapter 6: Bleeding-Edge Techniques for Univariate Time Series....227

Single-Step Data Preparation for Time-Series Forecasting 227

Horizon-Style Data Preparation for Time-Series Forecasting 229

LSTM Univariate Single-Step Style in Action 230

LSTM Univariate Horizon Style in Action 242

Bidirectional LSTM Univariate Single-Step Style in Action 253

Bidirectional LSTM Univariate Horizon Style in Action 262

GRU Univariate Single-Step Style in Action..... 271

GRU Univariate Horizon Style in Action 279

Auto-encoder LSTM Univariate Single-Step Style in Action..... 288

Auto-encoder LSTM Univariate Horizon Style in Action 297

CNN Univariate Single-Step Style in Action 306

CNN Univariate Horizon Style in Action 315

Summary..... 324

Chapter 7: Bleeding-Edge Techniques for Multivariate Time Series .. 325

LSTM Multivariate Horizon Style in Action 325

Bidirectional LSTM Multivariate Horizon Style in Action 337

Auto-encoder LSTM Multivariate Horizon Style in Action..... 346

GRU Multivariate Horizon Style in Action 356

CNN Multivariate Horizon Style in Action 365

Summary..... 374

Chapter 8: Prophet.....	375
The Prophet Model.....	375
Implementing Prophet.....	376
Adding Log Transformation.....	381
Adding Built-in Country Holidays.....	386
Adding Exogenous variables using add_regressors(function)	389
Summary.....	394
Index.....	395

About the Authors



B V Vishwas is a Data Scientist, AI researcher and AI Consultant, Currently living in Bengaluru(INDIA). His highest qualification is Master of Technology in Software Engineering from Birla Institute of Technology & Science, Pilani, India and his primary focus and inspiration is Data Warehousing, Big Data, Data Science (Machine Learning, Deep Learning, Timeseries, Natural Language Processing, Reinforcement Learning, and Operation Research). He has over seven years of IT experience currently working at Infosys as Data Scientist & AI Consultant. He has also worked on Data Migration, Data Profiling, ETL & ELT, OWB, Python, PL/SQL, Unix Shell Scripting, Azure ML Studio, Azure Cognitive Services, and AWS.



Ashish Patel is a Senior Data Scientist, AI researcher, and AI Consultant with over seven years of experience in the field of AI, Currently living in Ahmedabad(INDIA). He has a Master of Engineering Degree from Gujarat Technological University and his keen interest and ambition to research in the following domains such as (Machine Learning, Deep Learning, Time series, Natural Language Processing, Reinforcement Learning, Audio Analytics, Signal Processing, Sensor

ABOUT THE AUTHORS

Technology, IoT, Computer Vision). He is currently working as Senior Data Scientist for Cynet infotech Pvt Ltd. He has published more than 15 + Research papers in the field of Data Science with Reputed Publications such as IEEE. He holds Rank 3 as a kernel master in Kaggle. Ashish has immense experience working on cross-domain projects involving a wide variety of data, platforms, and technologies.

About the Technical Reviewer



Over his nearly three-decade career, **Alexey Panchekha, PhD, CFA**, has spent 10 years in academia, where he focused on nonlinear and dynamic processes; 10 years in the technology industry, where he specialized in program design and development; and eight years in financial services. In the latter arena, he specialized in applying mathematical techniques and technology to risk management

and alpha generation. For example, Panchekha was involved in the equity derivative trading technology platform at Goldman Sachs, and he led the creation of the multi-asset multigeographies portfolio risk management system at Bloomberg. He also served as the head of research at Markov Process International, a leader in portfolio attribution and analytics. Most recently, Panchekha cofounded Turing Technology Associates, Inc., with Vadim Fishman. Turing is a technology and intellectual property company that sits at the intersection of mathematics, machine learning, and innovation. Its solutions typically service the financial technology (fintech) industry. Turing primarily focuses on enabling technology that supports the burgeoning ensemble active management (EAM). Prior to Turing, Panchekha was managing director at Incapital, and head of research at F-Squared Investments, where he designed innovative volatility-based risk-sensitive investment strategies. He is fluent in multiple computer programming languages and software and database programs and is certified in deep learning software. He earned a PhD from Kharkiv Polytechnic University with studies in physics and mathematics as well as an MS in physics. Panchekha is a CFA charterholder.

Acknowledgments

This being my first book, I found transforming idea and real-world experience into its current shape to be a strenuous task. I am grateful to almighty God for blessing and guiding me in all endeavors. I would like to thank my parents (Vijay Kumar and Rathnamma), brother (Shreyas), other family, and friends for helping me sail though the sea of life.

—B V Vishwas

First and foremost, praises and thanks to God, the almighty, for His showers of blessings throughout the book-writing process. I would like to express my deep and sincere gratitude to my parents (Dinesh Kumar N Patel and Javnika ben D Patel), my sister (Nisha Patel), and my family and friends (Shailesh Patel, Sanket Patel, Nikit Patel, Mansi Patel, Khushboo Shah) for their support and valuable prayers.

—Ashish Patel

Special thanks to Celestin Suresh John, Aditee Mirashi, James Markham, Alexey Panchekha, and the Apress team for bringing this book to life.

Introduction

This book explains the concepts of time series from traditional to bleeding-edge techniques with full-fledged examples.

The book begins by covering time-series fundamentals and their characteristics, Structure & Components of time series data, preprocessing, and ways of crafting features through data wrangling. Next, it covers traditional time-series techniques such as the smoothing methods ARMA, ARIMA, SARIMA, SARIMAX, VAR, and VARMA using trending frameworks such as [Statsmodels](#) and Pmdarima.

Further covers how to leverage advanced deep learning-based techniques such as ANN, CNN, RNN, LSTM, GRU, and Autoencoder to solve time-series problems using Tensorflow. It concludes by explaining how to use the popular framework fbprophet for modeling time-series analysis.

After completion of the book, the reader will have thorough knowledge of concepts and techniques to solve time-series problems. All the code presented in this book is available in Jupyter Notebooks; this allows readers to do hands-on experiments and enhance them in exciting ways.