

## Python Data Science Essentials Second Edition Become An Efficient Data Science Practitioner By Understanding Python's Key Concepts

Learn data science by doing data science! Data Science Using Python and R will get you plugged into the world's two most widespread open-source platforms for data science: Python and R. Data science is hot. Bloomberg called data scientist "the hottest job in America." Python and R are the top two open-source data science tools in the world. In Data Science Using Python and R, you will learn step-by-step how to produce hands-on solutions to real-world business problems, using state-of-the-art techniques. Data Science Using Python and R is written for the general reader with no previous analytics or programming experience. An entire chapter is dedicated to learning the basics of Python and R. Then, each chapter presents step-by-step instructions and walkthroughs for solving data science problems using Python and R. Those with analytics experience will appreciate having a one-stop shop for learning how to do data science using Python and R. Topics covered include data preparation, exploratory data analysis, preparing to model the data, decision trees, model evaluation, misclassification costs, naïve Bayes classification, neural networks, clustering, regression modeling, dimension reduction, and association rules mining. Further, exciting new topics such as random forests and general linear models are also included. The book emphasizes data-driven error costs to enhance profitability, which avoids the common pitfalls that may cost a company millions of dollars. Data Science Using Python and R provides exercises at the end of every chapter, totaling over 500 exercises in the book. Readers will therefore have plenty of opportunity to test their newfound data science skills and expertise. In the Hands-on Analysis exercises, readers are challenged to solve interesting business problems using real-world data sets.

Discover how data science can help you gain in-depth insight into your business - the easy way! Jobs in data science abound, but few people have the data science skills needed to fill these increasingly important roles. Data Science For Dummies is the perfect starting point for IT professionals and students who want a quick primer on all areas of the expansive data science space. With a focus on business cases, the book explores topics in big data, data science, and data engineering, and how these three areas are combined to produce tremendous value. If you want to pick-up the skills you need to begin a new career or initiate a new project, reading this book will help you understand what technologies, programming languages, and mathematical methods on which to focus. While this book serves as a wildly fantastic guide through the broad, sometimes intimidating field of big data and data science, it is not an instruction manual for hands-on implementation. Here's what to expect: Provides a background in big data and data engineering before moving on to data science and how it's applied to generate value Includes coverage of big data frameworks like Hadoop, MapReduce, Spark, MPP platforms, and NoSQL Explains machine learning and many of its algorithms as well as artificial intelligence and the evolution of the Internet of Things Details data visualization techniques that can be used to showcase, summarize, and communicate the data insights you generate It's a big, big data world out there—let Data Science For Dummies help you harness its power and gain a competitive edge for your organization.

Data Science and Big Data Analytics is about harnessing the power of data for new insights. The book covers the breadth of activities and methods and tools that Data Scientists use. The content focuses on concepts, principles and practical applications that are applicable to any industry and technology environment, and the learning is supported and explained with examples that you can replicate using open-source software. This book will help you: Become a contributor on a data science team Deploy a structured lifecycle approach to data analytics problems Apply appropriate analytic techniques and tools to analyzing big data Learn how to tell a compelling story with data to drive business action Prepare for EMC Proven Professional Data Science Certification Corresponding data sets are available from the book's page at Wiley which you can find on the Wiley site by searching for the ISBN 9781118876138. Get started discovering, analyzing, visualizing, and presenting data in a meaningful way today!

Providing a unique approach to machine learning, this text contains fresh and intuitive, yet rigorous, descriptions of all fundamental concepts necessary to conduct research, build products, tinker, and play. By prioritizing geometric intuition, algorithmic thinking, and practical real world applications in disciplines including computer vision, natural language processing, economics, neuroscience, recommender systems, physics, and biology, this text provides readers with both a lucid understanding of foundational material as well as the practical tools needed to solve real-world problems. With in-depth Python and MATLAB/OCTAVE-based computational exercises and a complete treatment of cutting edge numerical optimization techniques, this is an essential resource for students and an ideal reference for researchers and practitioners working in machine learning, computer science, electrical engineering, signal processing, and numerical optimization.

Collect - Organize - Explore - Predict - Value

Practical Data Science Cookbook

Twisted Network Programming Essentials

Python Data Science Essentials

Data Science Using Python and R

Fundamentals of Data Visualization

Become an efficient data science practitioner by understanding Python's key concepts About This Book - Quickly get familiar with data science using Python 3.5 - Save time (and effort) with all the essential tools explained - Create effective data science projects and avoid common pitfalls with the help of examples and hints dictated by experience Who This Book Is For If you are an aspiring data scientist and you have at least a working knowledge of data analysis and Python, this book will get you started in data science. Data analysts with experience of R or MATLAB will also find the book to be a comprehensive reference to enhance their data manipulation and machine learning skills. What You Will Learn - Set up your data science toolbox using a Python scientific environment on Windows, Mac, and Linux - Get data ready for your data science project - Manipulate, fix, and explore data in order to solve data science problems - Set up an experimental pipeline to test your data science hypotheses - Choose the most effective and scalable learning algorithm for your data science tasks - Optimize your machine learning models to get the best performance - Explore and cluster graphs, taking advantage of interconnections and links in your data In Detail Fully expanded and upgraded, the second edition of Python Data Science Essentials takes you through all you need to know to succeed in data science using Python. Get modern insight into the core of Python data, including the latest versions of Jupyter notebooks, NumPy, pandas and scikit-learn. Look beyond the fundamentals with beautiful data visualizations with Seaborn and ggplot, web development with Bottle, and even the new frontiers of deep learning with Theano and TensorFlow. Dive into building your essential Python 3.5 data science toolbox, using a single-source approach that will allow to to work with Python 2.7 as well. Get to grips fast with data munging and preprocessing, and all the techniques you need to load, analyse, and process your data. Finally, get a complete overview of principal machine learning algorithms, graph analysis techniques, and all the visualization and deployment instruments that make it easier to present your results to an audience of both data science experts and business users. Style and approach The book is structured as a data science project. You will always benefit from clear code and simplified examples to help you understand the underlying mechanics and real-world datasets.

Summary Introducing Data Science teaches you how to accomplish the fundamental tasks that occupy data scientists. Using the Python language and common Python libraries, you'll experience firsthand the challenges of dealing with data at scale and gain a solid foundation in data science. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology Many companies need developers with data science skills to work on projects ranging from social media marketing to machine learning. Discovering what you need to learn to begin a career as a data scientist can seem bewildering. This book is designed to help you get started. About the Book Introducing Data Science Introducing Data Science explains vital data science concepts and teaches you how to accomplish the fundamental tasks that occupy data scientists. You'll explore data visualization, graph databases, the use of NoSQL, and the data science process. You'll use the Python language and common Python libraries as you experience firsthand the challenges of dealing with data at scale. Discover how Python allows you to gain insights from data sets so big that they need to be stored on multiple machines, or from data moving so quickly that no single machine can handle it. This book gives you hands-on experience with the most popular Python data science libraries, Scikit-learn and StatsModels. After reading this book, you'll have the solid foundation you need to start a career in data science. What's Inside Handling large data Introduction to machine learning Using Python to work with data Writing data science algorithms About the Reader This book assumes you're comfortable reading code in Python or a similar language, such as C, Ruby, or JavaScript. No prior experience with data science is required. About the Authors Davy Cielen, Arno D. B. Meysman, and Mohamed Ali are the founders and managing partners of Optimately and Maiton, where they focus on developing data science projects and solutions in various sectors. Table of Contents Data science in a big data world The data science process Machine learning Handling large data on a single computer First steps in big data Join the NoSQL movement The rise of graph databases Text mining and text analytics Data visualization to the end user

Learn data science with Python by building five real-world projects! Experiment with card game predictions, tracking disease outbreaks, and more, as you build a flexible and intuitive understanding of data science. In Data Science Bookcamp you will learn: - Techniques for computing and plotting probabilities - Statistical analysis using Scipy - How to organize datasets with clustering algorithms - How to visualize complex multi-variable datasets - How to train a decision tree machine learning algorithm In Data Science Bookcamp you'll test and build your knowledge of Python with the kind of open-ended problems that professional data scientists work on every day. Downloadable data sets and thoroughly-explained solutions help you lock in what you've learned, building your confidence and making you ready for an exciting new data science career. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the technology A data science project has a lot of moving parts, and it takes practice and skill to get all the code, algorithms, datasets, formats, and visualizations working together harmoniously. This unique book guides you through five realistic projects, including tracking disease outbreaks from news headlines, analyzing social networks, and finding relevant patterns in ad click data. About the book Data Science Bookcamp doesn't stop with surface-level theory and toy examples. As you work through each project, you'll learn how to troubleshoot common problems like missing data, messy data, and algorithms that don't quite fit the model you're building. You'll appreciate the detailed setup instructions and the fully explained solutions that highlight common failure points. In the end, you'll be confident in your skills because you can see the results. What's inside - Web scraping - Organize datasets with clustering algorithms - Visualize complex multi-variable datasets - Train a decision tree machine learning algorithm About the reader For readers who know the basics of Python. No prior data science or machine learning skills required. About the author Leonard Apeltsin is the Head of Data Science at Anomaly, where his team applies advanced analytics to uncover healthcare fraud, waste, and abuse. Table of Contents CASE STUDY 1 FINDING THE WINNING STRATEGY IN A CARD GAME 1 Computing probabilities using Python 2 Plotting probabilities using Matplotlib 3 Running random simulations in NumPy 4 Case study 1 solution CASE STUDY 2 ASSESSING ONLINE AD CLICKS FOR SIGNIFICANCE 5 Basic probability and statistical analysis using SciPy 6 Making predictions using the central limit theorem and SciPy 7 Statistical hypothesis testing 8 Analyzing tables using Pandas 9 Case study 2 solution CASE STUDY 3 TRACKING DISEASE OUTBREAKS USING NEWS HEADLINES 10 Clustering data into groups 11 Geographic location visualization and analysis 12 Case study 3 solution CASE STUDY 4 USING ONLINE JOB POSTINGS TO IMPROVE YOUR DATA SCIENCE RESUME 13 Measuring text similarities 14 Dimension reduction of matrix data 15 NLP analysis of large text datasets 16 Extracting text from web pages 17 Case study 4 solution CASE STUDY 5 PREDICTING FUTURE FRIENDSHIPS FROM SOCIAL NETWORK DATA 18 An introduction to graph theory and network analysis 19 Dynamic graph theory techniques for node ranking and social network analysis 20 Network-driven supervised machine learning 21 Training linear classifiers with logistic regression 22 Training nonlinear classifiers with decision tree techniques 23 Case study 5 solution

This book aim to equip the reader with Java Programming, Text Mining and Natural Language Processing basics. There will be many examples and explanations that are lucid and straight to the point. You will be walked through various projects and develop your own text mining application. This book will show you how to use Stanford NLP libraries also. Asides, we have also uploaded some of our own softwares at:

http://DSTK.Tech Content Covered: IntroductionGetting Started (Installing IDE, ...)Language Essentials I (variables, data types, ...)Language Essentials II (loops, if... else..., methods)Object Essentials (classes, inheritance, polymorphism, encapsulation, ...)Text Mining Essentials (Import Text Files, Text Transformation (lowercase, stopwords), Text Understanding (Stanford NLP), Text Classification (Stanford Classifier) )ISBN: 978-1-63535-546-8

Opportunities and Options

Data Science For Dummies

Python for Scientists

Foundations of Data Science

Introducing Data Science

Machine Learning Refined

Explore the ever-growing world of genetic algorithms to solve search, optimization, and AI-related tasks, and improve machine learning models using Python libraries such as DEAP, scikit-learn, and NumPy Key Features Explore the ins and outs of genetic algorithms with this fast-paced guide Implement tasks such as feature selection, search optimization, and cluster analysis using Python Solve combinatorial problems, optimize functions, and enhance the performance of artificial intelligence applications Book Description Genetic algorithms are a family of search, optimization, and learning algorithms inspired by the principles of natural evolution. By imitating the evolutionary process, genetic algorithms can overcome hurdles encountered in traditional search algorithms and provide high-quality solutions for a variety of problems. This book will help you get to grips with a powerful yet simple approach to applying genetic algorithms to a wide range of tasks using Python, covering the latest developments in artificial intelligence. After introducing you to genetic algorithms and their principles of operation, you'll understand how they differ from traditional algorithms and what types of problems they can solve. You'll then discover how they can be applied to search and optimization problems, such as planning, scheduling, gaming, and analytics. As you advance, you'll also learn how to use genetic algorithms to improve your machine learning and deep learning models, solve reinforcement learning tasks, and perform image reconstruction. Finally, you'll cover several related technologies that can open up new possibilities for future applications. By the end of this book, you'll have hands-on experience of applying genetic algorithms in artificial intelligence as well as in numerous other domains. What you will learn Understand how to use state-of-the-art Python tools to create genetic algorithm-based applications Use genetic algorithms to optimize functions and solve planning and scheduling problems Enhance the performance of machine learning models and optimize deep learning network architecture Apply genetic algorithms to reinforcement learning tasks using OpenAI Gym Explore how images can be reconstructed using a set of semi-transparent shapes Discover other bio-inspired techniques, such as genetic programming and particle swarm optimization Who this book is for This book is for software developers, data scientists, and AI enthusiasts who want to use genetic algorithms to carry out intelligent tasks in their applications. Working knowledge of Python and basic knowledge of mathematics and computer science will help you get the most out of this book.

A Tour of Data Science: Learn R and Python in Parallel covers the fundamentals of data science, including programming, statistics, optimization, and machine learning in a single short book. It does not cover everything, but rather, teaches the key concepts and topics in Data Science. It also covers two of the most popular programming languages used in Data Science, R and Python, in one source. Key features: Allows you to learn R and Python in parallel Cover statistics, programming, optimization and predictive modelling, and the popular data manipulation tools - data.table and pandas Provides a concise and accessible presentation Includes machine learning algorithms implemented from scratch, linear regression, lasso, ridge, logistic regression, gradient boosting trees, etc. Appealing to data scientists, statisticians, quantitative analysts, and others who want to learn programming with R and Python from a data science perspective.

This book provides an introduction to the mathematical and algorithmic foundations of data science, including machine learning, high-dimensional geometry, and analysis of large networks. Topics include the counterintuitive nature of data in high dimensions, important linear algebraic techniques such as singular value decomposition, the theory of random walks and Markov chains, the fundamentals of and important algorithms for machine learning, algorithms and analysis for clustering, probabilistic models for large networks, representation learning including topic modelling and non-negative matrix factorization, wavelets and compressed sensing. Important probabilistic techniques are developed including the law of large numbers, tail inequalities, analysis of random projections, generalization guarantees in machine learning, and moment methods for analysis of phase transitions in large random graphs. Additionally, important structural and complexity measures are discussed such as matrix norms and VC-dimension. This book is suitable for both undergraduate and graduate courses in the design and analysis of algorithms for data.

Become an efficient data science practitioner by understanding Python's key concepts About This Book Quickly get familiar with data science using Python 3.5 Save time (and effort) with all the essential tools explained Create effective data science projects and avoid common pitfalls with the help of examples and hints dictated by experience Who This Book Is For If you are an aspiring data scientist and you have at least a working knowledge of data analysis and Python, this book will get you started in data science. Data analysts with experience of R or MATLAB will also find the book to be a comprehensive reference to enhance their data manipulation and machine learning skills. What You Will Learn Set up your data science toolbox using a Python scientific environment on Windows, Mac, and Linux Get data ready for your data science project Manipulate, fix, and explore data in order to solve data science problems Set up an experimental pipeline to test your data science hypotheses Choose the most effective and scalable learning algorithm for your data science tasks Optimize your machine learning models to get the best performance Explore and cluster graphs, taking advantage of interconnections and links in your data In Detail Fully expanded and upgraded, the second edition of Python Data Science Essentials takes you through all you need to know to succeed in data science using Python. Get modern insight into the core of Python data, including the latest versions of Jupyter notebooks, NumPy, pandas and scikit-learn. Look beyond the fundamentals with beautiful data visualizations with Seaborn and ggplot, web development with Bottle, and even the new frontiers of deep learning with Theano and TensorFlow. Dive into building your essential Python 3.5 data science toolbox, using a single-source approach that will allow to to work with Python 2.7 as well. Get to grips fast with data munging and preprocessing, and all the techniques you need to load, analyse, and process your data. Finally, get a complete overview of principal machine learning algorithms, graph analysis techniques, and all the visualization and deployment instruments that make it easier to present your results to an audience of both data science experts and business users. Style and approach The book is structured as a data science project. You will always benefit from clear code and simplified examples to help you understand the underlying mechanics and real-world datasets.

Python Testing with pytest

Practical Statistics for Data Scientists

Python Data Science Essentials - Second Edition

Discovering, Analyzing, Visualizing and Presenting Data

Build a Career in Data Science

Learn By Examples - A Quick Guide to Java Programming for Text Mining and NLP

Learn how to use R to turn raw data into insight, knowledge, and understanding. This book introduces you to R, RStudio, and the tidyverse, a collection of R packages designed to work together to make data science fast, fluent, and fun. Suitable for readers with no previous programming experience, R for Data Science is designed to get you doing data science as quickly as possible. Authors Hadley Wickham and Garrett Grolemund guide you through the steps of importing, wrangling, exploring, and modeling your data and communicating the results. You 'll get a complete, big-picture understanding of the data science cycle, along with basic tools you need to manage the details. Each section of the book is paired with exercises to help you practice what you 've learned along the way. You 'll learn how to: Wrangle—transform your datasets into a form convenient for analysis Program—learn powerful R tools for solving data problems with greater clarity and ease Explore—examine your data, generate hypotheses, and quickly test them Model—provide a low-dimensional summary that captures true "signals" in your dataset Communicate—learn R Markdown for integrating prose, code, and results

In this book, you will create two desktop applications using Python GUI and MariaDB. This book is mariadb-based python programming Intentionally designed for various levels of interest and ability of learners, this book is suitable for students, engineers, and even researchers in a variety of disciplines. No advanced programming experience is needed, and only a few school-level programming skill are needed. In the first chapter, you will learn to use several widgets in PyQt5: Display a welcome message; Use the Radio Button widget; Grouping radio buttons; Displays options in the form of a check box; and Display two groups of check boxes. In chapter two, you will learn to use the following topics: Using Signal / Slot Editor; Copy and place text from one Line Edit widget to another; Convert data types and make a simple calculator; Use the Spin Box widget; Use scrollbars and sliders; Using the Widget List; Select a number of list items from one Widget List and display them on another Widget List widget; Add items to the Widget List; Use the Combo Box widget; Displays data selected by the user from the Calendar Widget; Creating a hotel reservation application; and Display tabular data using Table Widgets. In third chapter, you will learn: How to create the initial three tables project in the School database: Teacher, Class, and Subject tables; How to create database configuration files; How to create a Python GUI for inserting and editing tables; How to create a Python GUI to join and query the three tables. In fourth chapter, you will learn how to: Create a main form to connect all forms; Create a project will add three more tables to the school database: Student, Parent, and Tuition tables; Create a Python GUI for inserting and editing tables; Create a Python GUI to join and query over the three tables. In chapter five, you will join the six classes, Teacher, TClass, Subject, Student, Parent, and Tuition and make queries over those tables. In chapter six, you will create dan configure database. In this chapter, you will create Suspect table in crime database. This table has eleven columns: suspect\_id (primary key), suspect\_name, birth\_date, case\_date, report\_date, suspect\_status, arrest\_date, mother\_name, address, telephone, and photo. You will also create GUI to display, edit, insert, and delete for this table. In chapter seven, you will create a table with the name Feature\_Extraction, which has eight columns: feature\_id (primary key), suspect\_id (foreign key), feature1, feature2, feature3, feature4, feature5, and feature6. The six fields (except keys) will have a VARCHAR data type (200). You will also create GUI to display, edit, insert, and delete for this table. In chapter eight, you will create two tables, Police and Investigator. The Police table has six columns: police\_id (primary key), province, city, address, telephone, and photo. The Investigator table has eight columns: investigator\_id (primary key), investigator\_name, rank, birth\_date, gender, address, telephone, and photo. You will also create GUI to display, edit, insert, and delete for both tables. In chapter nine, you will create two tables, Victim and Case\_File. The Victim table has nine columns: victim\_id (primary key), victim\_name, crime\_type, birth\_date, crime\_date, gender, address, telephone, and photo. The Case\_File table has seven columns: case\_file\_id (primary key), suspect\_id (foreign key), police\_id (foreign key), investigator\_id (foreign key), victim\_id (foreign key), status, and description. You will create GUI to display, edit, insert, and delete for both tables as well.

The Essentials of Data Science: Knowledge Discovery Using R presents the concepts of data science through a hands-on approach using free and open source software. It systematically drives an accessible journey through data analysis and machine learning to discover and share knowledge from data. Building on over thirty years' experience in teaching and practising data science, the author encourages a programming-by-example approach to ensure students and practitioners attune to the practise of data science while building their data skills. Proven frameworks are provided as reusable templates. Real world case studies then provide insight for the data scientist to swiftly adapt the templates to new tasks and datasets. The book begins by introducing data science. It then reviews R's capabilities for analysing data by writing computer programs. These programs are developed and explained step by step. From analysing and visualising data, the framework moves on to tried and tested machine learning techniques for predictive modelling and knowledge discovery. Literate programming and a consistent style are a focus throughout the book.

For many researchers, Python is a first-class tool mainly because of its libraries for storing, manipulating, and gaining insight from data. Several resources exist for individual pieces of this data science stack, but only with the Python Data Science Handbook do you get them all—IPython, NumPy, Pandas, Matplotlib, Scikit-Learn, and other related tools. Working scientists and data crunchers familiar with reading and writing Python code will find this comprehensive desk reference ideal for tackling day-to-day issues: manipulating, transforming, and cleaning data; visualizing different types of data; and using data to build statistical or machine learning models. Quite simply, this is the must-have reference for scientific computing in Python. With this handbook, you'll learn how to use: IPython and Jupyter; provide computational environments for data scientists using Python NumPy; includes the ndarray for efficient storage and manipulation of dense data arrays in Python Pandas; features the DataFrame for efficient storage and manipulation of labeled/columnar data in Python Matplotlib; includes capabilities for a flexible range of data visualizations in Python Scikit-Learn; for efficient and clean Python implementations of the most important and established machine learning algorithms

Applying genetic algorithms to solve real-world deep learning and artificial intelligence problems

How to Lead in Data Science

Simple, Rapid, Effective, and Scalable

Data Science with Python and Dask

Python for Data Analysis

Mastering Python for Data Science

"This textbook is a well-rounded, rigorous, and informative work presenting the mathematics behind modern machine learning techniques. It hits all the right notes: the choice of topics is up-to-date and perfect for a course on data science for mathematics students at the advanced to early graduate level. This book fills a sorely-needed gap in the existing literature by not sacrificing depth for breadth, presenting proofs of major theorems and subsequent derivations, as well as providing a copious amount of Python code. I only wish a book like this had been around for my journey!" -Nicholas Hoell, University of Toronto "This is a well-written book that provides a deeper dive into data-scientific methods than many introductory texts. The writing is clear, and the text logically builds up regularization, classification, and decision trees. Compared to its predecessors, it carves out a unique niche. -Adam Loy, Carleton College The purpose of Data Science and Machine Learning: Mathematical and Statistical Methods is to provide an accessible, yet comprehensive textbook intended for students interested in gaining a better understanding of mathematics and statistics that underpin the rich variety of ideas and machine learning algorithms in data science. Key Features: Focuses on mathematical understanding. Presentation is self-contained, accessible, and comprehensive. Extensive list of exercises and worked-out example concrete algorithms with Python code. Full color throughout. The Authors: Dirk P. Kroese, PhD, is a Professor of Mathematics and Statistics at The University of Queensland. He has published over 120 articles and five books in a wide range of areas in mathematics, statistics, data science, learning, and Monte Carlo methods. He is a pioneer of the well-known Cross-Entropy method—an adaptive Monte Carlo technique, which is being used around the world to help solve difficult estimation and optimization problems in science, engineering, and finance. Zdravko Botev, PhD, Australian Mathematical Sciences Institute Lecturer in Data Science and Machine Learning with an appointment at the University of New South Wales in Sydney, Australia. He is the recipient of the 2018 Christopher Heyde Medal of the Australian Academy of Science for distinguished research in Mathematical Sciences. Thomas Taimre, PhD, is a Senior Lecturer of Mathematics and Statistics at The University of Queensland. His research interests range from applied probability and Monte Carlo methods to applied physics and the remarkably universal self-mixing effect in lasers. He has published over 100 articles, holds a patent, and is the coauthor of Handbook of Monte Carlo Methods (Wiley). Radislav Vaisman, PhD, is a Lecturer of Mathematics and Statistics at The University of Queensland. His research interests lie at the intersection of applied probability, machine learning, and computer science. He has published over 20 articles and two books.

Summary Dask is a native parallel analytics tool designed to integrate seamlessly with the libraries you're already using, including Pandas, NumPy, and Scikit-Learn. With Dask you can crunch and work with huge datasets, using the tools you already have. And Data Science with Python is your guide to using Dask for your data projects without changing the way you work! Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. You'll find registration instructions inside the print book. About the Technology An efficient means everything for the success of a data science project. Dask is a flexible library for parallel computing in Python that makes it easy to build intuitive workflows for ingesting and analyzing large, distributed datasets. Dask provides dynamic task scheduling and parallel collections through the functionality of NumPy, Pandas, and Scikit-learn, enabling users to scale their code from a single laptop to a cluster of hundreds of machines with ease. About the Book Data Science with Python and Dask teaches you to build scalable projects that can handle massive datasets. After the Dask framework, you'll analyze data in the NYC Parking Ticket database and use DataFrames to streamline your process. Then, you'll create machine learning models using Dask-ML, build interactive visualizations, and build clusters using AWS and Docker. What's Inside Working with large structured and unstructured datasets Visualization with Seaborn and Dasher Implementing your own algorithms Building distributed apps with Dask Distributed Packaging and deploying Dask apps About the Reader For data scientists and developers with experience using Python or a similar language stack. About the Author Jesse Daniel is an experienced Python developer. He taught Python for Data Science at the University of Denver and leads a team of data scientists at a Denver-based media technology company. Table of Contents PART 1 - The Building Blocks of scalable computing scalable computing matters Introducing Dask PART 2 - Working with Structured Data using Dask DataFrames Introducing Dask DataFrames Loading data into DataFrames Cleaning and transforming DataFrames Summarizing and analyzing DataFrames Visualizing DataFrames with Seaborn

Visualizing location data with Dasher PART 3 - Extending and deploying Dask Working with Bags and Arrays Machine learning with Dask-ML Scaling and deploying Dask This invaluable addition to any data scientist's library shows you how to apply the R programming language and useful statistical techniques to everyday business situations as well as how to effectively present results to audiences of all levels. To answer the ever-increasing demand for machine learning and analysis, this new edition boasts additional R tools, modeling techniques, and more. Practical Data Science with R, Second Edition takes a practice-oriented approach to explaining basic principles in the ever-expanding field of data science. You'll jump right to real-world use cases and apply the R programming language and statistical analysis techniques to carefully explained examples based in marketing, business intelligence, and decision support. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications.

This research scholarly illustrated book has more than 250 illustrations. The simple models of supervised machine learning with Gaussian Naive Bayes, Naive Bayes, decision trees, classification rule learners, linear regression, logistic regression, local polynomial regression, regression trees, K-nearest neighbors, and support vector machines lay a more excellent foundation for statistics. The author of the book Dr. Ganaopathi Pulipaka, a top influencer of machine learning in the US, has created this as a reference book for universities. This book contains an incredible text machine learning and engineering beyond a compact manual. The author goes to extraordinary lengths to make academic machine learning and deep learning literature comprehensible to create a new body of knowledge. The book aims at readership from university students, enterprise data science beginners, machine learning and deep learning engineers at scale for high-performance computing environments. A Greater Foundation of Machine Learning Engineering covers a broad range of classical linear algebra and calculus with program implementations in PyTorch, TensorFlow, and Python with in-depth coverage. The author does not hesitate to go into math equations for each algorithm at length that usually many foundational machine learning books lack leveraging the JupyterLab environment. Newcomers can leverage the book from University or people from data science or software lives to the advanced practitioners of machine learning and deep learning. Though the book title suggests machine learning, there are several implementations of deep learning algorithms, including deep reinforcement learning. The book's mission is to help build a foundation for machine learning and deep learning engineers with all the algorithms, processors to train and deploy into production for enterprise-wide machine learning implementations. This book also introduces all the concepts of natural language processing required for machine learning algorithms in Python. The book covers Bayesian statistics without assuming high-level mathematics or statistics experience from the readers. It delivers the core concepts and implementations required with R code with open datasets. The book also covers unsupervised machine learning with association rules and k-means clustering, metal-learning algorithms, bagging, boosting, random forests, and ensemble methods. The book delves into the origins of deep learning in a scholarly way covering neural networks, restricted Boltzmann machines, deep belief networks, and deep Boltzmann machines, LSTM, and natural language processing techniques with deep learning algorithms and math equations. It leverages the NLTK library of Python with PyTorch, Python, and TensorFlow's installation steps, then demonstrates how to build neural networks with TensorFlow. Deploying machine learning algorithms require a blend of cloud computing platforms, SQL databases, and NoSQL databases. Any data scientist with a statistics background that looks to transition into a machine learning engineer role requires an in-depth understanding of machine learning implementations on Amazon, Google, or Microsoft Azure cloud computing platforms. The book provides real-world client projects for understanding the complete implementation of machine learning algorithms. This book is a marvel that does not leave any application of machine learning algorithms. It sets a more excellent foundation for newcomers and expands the horizons for experienced deep learning practitioners. It is almost inevitable that there will be a series of more advanced algorithms follow-up books from the author in some shape or form after several perfect foundation for machine learning engineering.

An Introduction to Statistical Learning

A Greater Foundation for Machine Learning Engineering

50 Essential Concepts

Practical Data Science with R

Big Data, Machine Learning, and More, Using Python Tools

with Applications in R

Summary Introducing Data Science teaches you how to accomplish the fundamental tasks that occupy data scientists. Using the Python language and common Python libraries, you'll experience firsthand the challenges of dealing with data at scale and gain a solid foundation in data science. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology Many companies need developers with data science skills to work on projects ranging from social media marketing to machine learning. Discovering what you need to learn to begin a career as a data scientist can seem bewildering. This book is designed to help you get started. About the Book Introducing Data ScienceIntroducing Data Science explains vital data science concepts and teaches you how to accomplish the fundamental tasks that occupy data scientists. You'll explore data visualization, graph databases, the use of NoSQL, and the data science process. You'll use the Python language and common Python libraries as you experience firsthand the challenges of dealing with data at scale. Discover how Python allows you to gain insights from data sets so big that they need to be stored on multiple machines, or from data moving so quickly that no single machine can handle it. This book gives you hands-on experience with the most popular Python data science libraries, Scikit-learn and StatsModels. After reading this book, you'll have the solid foundation you need to start a career in data science. What's Inside Handling large data Introduction to machine learning Using Python to work with data Writing data science algorithms About the Reader This book assumes you're comfortable reading code in Python or a similar language, such as C, Ruby, or JavaScript. No prior experience with data science is required. About the Authors Davy Claien, Arno D. B. Meysman, and Mohamed Ali are the founders and managing partners of Optimately and Maiton, where they focus on developing data science projects and solutions in various sectors. Table of Contents Data science in a big data world The data science process Machine learning Handling large data on a single computer First steps in big data Join the NoSQL movement The rise of graph databases Text mining and text analytics Data visualization to the end user

Get complete instructions for manipulating, processing, cleaning, and crunching datasets in Python. Updated for Python 3.6, the second edition of this hands-on guide is packed with practical case studies that show you how to solve a broad set of data analysis problems effectively. You'll learn the latest versions of pandas, NumPy, IPython, and Jupyter in the process. Written by Wes McKinney, the creator of the Python pandas project, this book is a practical, modern introduction to data science tools in Python. It's ideal for analysts new to Python and for Python programmers new to data science and scientific computing. Data files and related material are available on GitHub. Use the IPython shell and Jupyter notebook for exploratory computing Learn basic and advanced features in NumPy (Numerical Python) Get started with data analysis tools in the pandas library Use flexible tools to load, clean, transform, merge, and reshape data Create informative visualizations with matplotlib Apply the pandas groupby facility to slice, dice, and summarize datasets Analyze and manipulate regular and irregular time series data Learn how to solve real-world data analysis problems with thorough, detailed examples

Scientific Python is a significant public domain alternative to expensive proprietary software packages. This book teaches from scratch everything the working scientist needs to know using copious, downloadable, useful and adaptable code snippets. Readers will discover how easy it is to implement and test non-trivial mathematical algorithms and will be guided through the many freely available add-on modules. A range of examples, relevant to many different fields, illustrate the language's capabilities. The author also shows how to use pre-existing legacy code (usually in Fortran77) within the Python environment, thus avoiding the need to master the original code. In this new edition, several chapters have been re-written to reflect the IPython notebook style. With an extended index, an entirely new chapter discussing SymPy and a substantial increase in the number of code snippets, researchers and research students will be able to quickly acquire all the skills needed for using Python effectively.

Summary You are going to need more than technical knowledge to succeed as a data scientist. Build a Career in Data Science teaches you what school leaves out, from how to land your first job to the lifecycle of a data science project, and even how to become a manager. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the technology What are the keys to a data scientist's long-term success? Blending your technical know-how with the right "soft skills" turns out to be a central ingredient of a rewarding career. About the book Build a Career in Data Science is your guide to landing your first data science job and developing into a valued senior employee. By following clear and simple instructions, you'll learn to craft an amazing resume and ace your interviews. In this demanding, rapidly changing field, it can be challenging to keep projects on track, adapt to company needs, and manage tricky stakeholders. You'll love the insights on how to handle expectations, deal with failures, and plan your career path in the stories from seasoned data scientists included in the book. What's inside Creating a portfolio of data science projects Assessing and negotiating an offer Leaving gracefully and moving up the ladder Interviews with professional data scientists About the reader For readers who want to begin or advance a data science career. About the author Emily Robinson is a data scientist at Warby Parker. Jacqueline Nolis is a data science consultant and mentor. Table of Contents: PART 1 - GETTING STARTED WITH DATA SCIENCE 1. What is data science? 2. Data science companies 3. Getting the skills 4. Building a portfolio PART 2 - FINDING YOUR DATA SCIENCE JOB 5. The search: Identifying the right job for you 6. The application: Résumés and cover letters 7. The interview: What to expect and how to handle it 8. The offer: Knowing what to accept PART 3 - SETTLING INTO DATA SCIENCE 9. The first months on the job 10. Making an effective analysis 11. Deploying a model into production 12. Working with stakeholders PART 4 - GROWING IN YOUR DATA SCIENCE ROLE 13. When your data science project fails 14. Joining the data science community 15. Leaving your job gracefully 16. Moving up the ladder

Python Data Science Handbook

Big data, machine learning, and more, using Python tools

Data Science Essentials in Python

Import, Tidy, Transform, Visualize, and Model Data

The Essentials of Data Science: Knowledge Discovery Using R

Data Science for Undergraduates

Explore the world of data science through Python and learn how to make sense of data About This Book Master data science methods using Python and its libraries Create data visualizations and mine for patterns Advanced techniques for the four fundamentals of Data Science with Python - data mining, data analysis, data visualization, and machine learning Who This Book Is For If you are a Python developer who wants to master the world of data science then this book is for you. Some knowledge of data science is assumed. What You Will Learn Manage data and perform linear algebra in Python Derive inferences from the analysis by performing inferential statistics Solve data science problems in Python Create high-end visualizations using Python Evaluate and apply the linear regression technique to estimate the relationships among variables. Build recommendation engines with the various collaborative filtering algorithms Apply the ensemble methods to improve your predictions Work with big data technologies to handle data at scale In Detail Data science is a relatively new knowledge domain which is used by various organizations to make data driven decisions. Data scientists have to wear various hats to work with data and to derive value from it. The Python programming language, beyond having conquered the scientific community in the last decade, is now an indispensable tool for the data science practitioner and a must-know tool for every aspiring data scientist. Using Python will offer you a fast, reliable, cross-platform, and mature environment for data analysis, machine learning, and algorithmic problem solving. This comprehensive guide helps you move beyond the hype and transcend the theory by providing you with a hands-on, advanced study of data science. Beginning with the essentials of Python in data science, you will learn to manage data and perform linear algebra in Python. You will move on to deriving inferences from the analysis by performing inferential statistics, and mining data to reveal hidden patterns and trends. You will use the matplotlib library to create high-end visualizations in Python and uncover the fundamentals of machine learning. Next, you will apply the linear regression technique and also learn to apply the logistic regression technique to your applications, before creating recommendation engines with various collaborative filtering algorithms and improving your predictions by applying the ensemble methods. Finally, you will perform K-means clustering, along with an analysis of unstructured data with different text mining techniques and leveraging the power of Python in big data analytics. Style and approach This book is an easy-to-follow, comprehensive guide on data science using Python. The topics covered in the book can all be used in real world scenarios.

Effective visualization is the best way to communicate information from the increasingly large and complex datasets in the natural and social sciences. But with the increasing power of visualization software today, scientists, engineers, and business analysts often have to navigate a bewildering array of visualization choices and options. This practical book takes you through many commonly encountered visualization problems, and it provides guidelines on how to turn large datasets into clear and compelling figures. What visualization type is best for the story you want to tell? How do you make informative figures that are visually pleasing? Author Claus O. Wilke teaches you the elements most critical to successful data visualization. Explore the basic concepts of color as a tool to highlight, distinguish, or represent a value Understand the importance of redundant coding to ensure you provide key information in multiple ways Use the book's visualizations directory, a graphical guide to commonly used types of data visualizations Get extensive examples of good and bad figures Learn how to use figures in a document or report and how employ them effectively to tell a compelling story

Learn how to fuse today's data science tools and techniques with your SAP enterprise resource planning (ERP) system. With this practical guide, SAP veterans Greg Foss and Paul Modderman demonstrate how to use several data analysis tools to solve interesting problems with your SAP data. Data engineers and scientists will explore ways to add SAP data to their analysis processes, while SAP business analysts will learn practical methods for answering questions about the business. By focusing on grounded explanations of both SAP processes and data science tools, this book gives data scientists and business analysts powerful methods for discovering deep data truths. You'll explore: Examples of how data analysis can help you solve several SAP challenges Natural language processing for unlocking the secrets in text Data science techniques for data clustering and segmentation Methods for detecting anomalies in your SAP data Data visualization techniques for making your data come to life

Statistical methods are a key part of data science, yet very few data scientists have any formal statistics training. Courses and books on basic statistics rarely cover the topic from a data science perspective. This practical guide explains how to apply various statistical methods to data science, tells you how to avoid their misuse, and gives you advice on what's important and what's not. Many data science resources incorporate statistical methods but lack a deeper statistical perspective. If you're familiar with the R programming language, and have some exposure to statistics, this quick reference bridges the gap in an accessible, readable format. With this book, you'll learn: Why exploratory data analysis is a key preliminary step in data science How random sampling can reduce bias and yield a higher quality dataset, even with big data How the principles of experimental design yield definitive answers to questions How to use regression to estimate outcomes and detect anomalies Key classification techniques for predicting which categories a record belongs to Statistical machine learning methods that "learn" from data Unsupervised learning methods for extracting meaning from unlabeled data

Data Analysis and Prediction Algorithms with R

A Primer on Making Informative and Compelling Figures

Introduction to Data Science

Mathematical and Statistical Methods

Data Science and Big Data Analytics

Building Two Desktop Applications Using Python GUI and MariaDB

**Data is bigger, arrives faster, and comes in a variety of formats—and it all needs to be processed at scale for analytics or machine learning. But how can you process such varied workloads efficiently? Enter Apache Spark. Updated to include Spark 3.0, this second edition shows data engineers and data scientists why structure and unification in Spark matters. Specifically, this book explains how to perform simple and complex data analytics and employ machine learning algorithms. Through step-by-step walkthroughs, code snippets, and notebooks, you'll be able to: Learn Python, SQL, Scala, or Java high-level Structured APIs Understand Spark operations and SQL Engine Inspect, tune, and debug Spark operations with Spark configurations and Spark UI**

Connect to data sources: JSON, Parquet, CSV, Avro, ORC, Hive, S3, or Kafka Perform analytics on batch and streaming data using Structured Streaming Build reliable data pipelines with open source Delta Lake and Spark Develop machine learning pipelines with MLlib and productionize models using MLflow

Do less work when testing your Python code, but be just as expressive, just as elegant, and just as readable. The pytest testing framework helps you write tests quickly and keep them readable and maintainable - with no boilerplate code. Using a robust yet simple fixture model, it's just as easy to write small tests with pytest as it is to scale up to complex functional testing for applications, packages, and libraries. This book shows you how. For Python-based projects, pytest is the undeniable choice to test your code if you're looking for a full-featured, API-independent, flexible, and extensible testing framework. With a full-bodied fixture model that is unmatched in any other tool, the pytest framework gives you powerful features such as assert rewriting and plug-in capability - with no boilerplate code. With simple step-by-step instructions and sample code, this book gets you up to speed quickly on this easy-to-learn and robust tool. Write short, maintainable tests that elegantly express what you're testing. Add powerful testing features and still speed up test times by distributing tests across multiple processors and running tests in parallel. Use the built-in assert statements to reduce false test failures by separating setup and test failures. Test error conditions and corner cases with expected exception testing, and use one test to run many test cases with parameterized testing. Extend pytest with plugins, connect it to continuous integration systems, and use it in tandem with tox, mock, coverage, unittest, and doctest. Write simple, maintainable tests that elegantly express what you're testing and why. **What You Need:** The examples in this book are written using Python 3.6 and pytest 3.0. However, pytest 3.0 supports Python 2.6, 2.7, and Python 3.3-3.6.

Learn the techniques and math you need to start making sense of your data About This Book Enhance your knowledge of coding with data science theory for practical insight into data science and analysis More than just a math class, learn how to perform real-world data science tasks with R and Python Create actionable insights and transform raw data into tangible value Who This Book Is For You should be fairly well acquainted with basic algebra and should feel comfortable reading snippets of R/Python as well as pseudo code. You should have the urge to learn and apply the techniques put forth in this book on either your own data sets or those provided to you. If you have the basic math skills but want to apply them in data science or you have good programming skills but lack math, then this book is for you. **What You Will Learn** Get to know the five most important steps of data science Use your data intelligently and learn how to handle it with care Bridge the gap between mathematics and programming Learn about probability, calculus, and how to use statistical models to control and clean your data and drive actionable results Build and evaluate baseline machine learning models Explore the most effective metrics to determine the success of your machine learning models Create data visualizations that communicate actionable insights Read and apply machine learning concepts to your problems and make actual predictions In Detail Need to turn your skills at programming into effective data science skills? Principles of Data Science is created to help you join the dots between mathematics, programming, and business analysis. With this book, you'll feel confident about asking—and answering—complex and sophisticated questions of your data to move from abstract and raw statistics to actionable ideas. With a unique approach that bridges the gap between mathematics and computer science, this book takes you through the entire data science pipeline. Beginning with cleaning and preparing data, and effective data mining strategies and techniques, you'll move on to build a comprehensive picture of how every piece of the data science puzzle fits together. Learn the fundamentals of computational mathematics and statistics, as well as some pseudocode being used today by data scientists and analysts. You'll get to grips with machine learning, discover the statistical models that help you take control and navigate even the densest datasets, and find out how to create powerful visualizations that communicate what your data means. **Style and approach** This is an easy-to-understand and accessible tutorial. It is a step-by-step guide with use cases, examples, and illustrations to get you well-versed with the concepts of data science. Along with explaining the fundamentals, the book will also introduce you to slightly advanced concepts later on and will help you implement these techniques in the real world.

Over 85 recipes to help you complete real-world data science projects in R and Python About This Book Tackle every step in the data science pipeline and use it to acquire, clean, analyze, and visualize your data Get beyond the theory and implement real-world projects in data science using R and Python Easy-to-follow recipes will help you understand and implement the numerical computing concepts Who This Book Is For If you are an aspiring data scientist who wants to learn data science and numerical programming concepts through hands-on, real-world project examples, this is the book for you. Whether you are brand new to data science or you are a seasoned expert, you will benefit from learning about the structure of real-world data science projects and the programming examples in R and Python. **What You Will Learn** Learn and understand the installation procedure and environment required for R and Python on various platforms Prepare data for analysis by implement various data science concepts such as acquisition, cleaning and munging through R and Python Build a predictive model and an exploratory model Analyze the results of your model and create reports on the acquired data Build various tree-based methods and Build random forest In Detail As increasing amounts of data are generated each year, the need to analyze and create value out of it is more important than ever. Companies that know what to do with their data and how to do it well will have a competitive advantage over companies that don't. Because of this, there will be an increasing demand for people that possess both the analytical and technical abilities to extract valuable insights from data and create valuable solutions that put those insights to use. Starting with the basics, this book covers how to set up your numerical programming environment, introduces you to the data science pipeline, and guides you through several data projects in a step-by-step format. By sequentially working through the steps in each chapter, you will quickly familiarize yourself with the process and learn how to apply it to a variety of situations with examples using the two most popular programming languages for data analysis—R and Python. **Style and approach** This step-by-step guide to data science is full of hands-on examples of real-world data science tasks. Each recipe focuses on a particular task involved in the data science pipeline, ranging from readying the dataset to analytics and visualization

Learning Spark

Five real-world Python projects

R for Data Science

Learn R and Python in Parallel

Practical Data Science with SAP

Data Wrangling with Pandas, NumPy, and IPython

*A field guide for the unique challenges of data science leadership, filled with transformative insights, personal experiences, and industry examples. In How To Lead in Data Science you will learn: Best practices for leading projects while balancing complex trade-offs Specifying, prioritizing, and planning projects from vague requirements Navigating structural challenges in your organization Working through project failures with positivity and tenacity Growing your team with coaching, mentoring, and advising Crafting technology roadmaps and championing successful projects Driving diversity, inclusion, and belonging within teams Architecting a long-term business strategy and data roadmap as an executive Delivering a data-driven culture and structuring productive data science organizations How to Lead in Data Science is full of techniques for leading data science at every seniority level—from heading up a single project to overseeing a whole company's data strategy. Authors Jike Chong and Yue Cathy Chang share hard-won advice that they've developed building data teams for LinkedIn, Acorns, Yiren Digital, large asset-management firms, Fortune 50 companies, and more. You'll find advice on plotting your long-term career advancement, as well as quick wins you can put into practice right away. Carefully crafted assessments and interview scenarios encourage introspection, reveal personal blind spots, and highlight development areas. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the technology Lead your data science teams and projects to success! To make a consistent, meaningful impact as a data science leader, you must articulate technology roadmaps, plan effective project strategies, support diversity, and create a positive environment for professional growth. This book delivers the wisdom and practical skills you need to thrive as a data science leader at all levels, from team member to the C-suite. About the book How to Lead in Data Science shares unique leadership techniques from high-performance data teams. It's filled with best practices for balancing project trade-offs and producing exceptional results, even when beginning with vague requirements or unclear expectations. You'll find a clearly presented modern leadership framework based on current case studies, with insights reaching all the way to Aristotle and Confucius. As you read, you'll build practical skills to grow and improve your team, your company's data culture, and yourself. What's inside How to coach and mentor team members Navigate an organization's structural challenges Secure commitments from other teams and partners Stay current with the technology landscape Advance your career About the reader For data science practitioners at all levels. About the author Dr. Jike Chong and Yue Cathy Chang build, lead, and grow high-performing data teams across industries in public and private companies, such as Acorns, LinkedIn, large asset-management firms, and Fortune 50 companies. Table of Contents 1 What makes a successful data scientist? PART 1 THE TECH LEAD: CULTIVATING LEADERSHIP 2 Capabilities for leading projects 3 Virtues for leading projects PART 2 THE MANAGER: NURTURING A TEAM 4 Capabilities for leading people 5 Virtues for leading people PART 3 THE DIRECTOR: GOVERNING A FUNCTION 6 Capabilities for leading a function 7 Virtues for leading a function PART 4 THE EXECUTIVE: INSPIRING AN INDUSTRY 8 Capabilities for leading a company 9 Virtues for leading a company PART 5 THE LOOP AND THE FUTURE 10 Landscape, organization, opportunity, and practice 11 Leading in data science and a future outlook*

*Introduction to Data Science: Data Analysis and Prediction Algorithms with R introduces concepts and skills that can help you tackle real-world data analysis challenges. It covers concepts from probability, statistical inference, linear regression, and machine learning. It also helps you develop skills such as R programming, data wrangling, data visualization, predictive algorithm building, file organization with UNIX/Linux shell, version control with Git and GitHub, and reproducible document preparation. This book is a textbook for a first course in data science. No previous knowledge of R is necessary, although some experience with programming may be helpful. The book is divided into six parts: R, data visualization, statistics with R, data wrangling, machine learning, and productivity tools. Each part has several chapters meant to be presented as one lecture. The author uses motivating case studies that realistically mimic a data scientist's experience. He starts by asking specific questions and answers these through data analysis so concepts are learned as a means to answering the questions. Examples of the case studies included are: US murder rates by state, self-reported student heights, trends in world health and economics, the impact of vaccines on infectious disease rates, the financial crisis of 2007-2008, election forecasting, building a baseball team, image processing of hand-written digits, and movie recommendation systems. The statistical concepts used to answer the case study questions are only briefly introduced, so complementing with a probability and statistics textbook is highly recommended for in-depth understanding of these concepts. If you read and understand the chapters and complete the exercises, you will be prepared to learn the more advanced concepts and skills needed to become an expert.*

*Data science is emerging as a field that is revolutionizing science and industries alike. Work across nearly all domains is becoming more data driven, affecting both the jobs that are available and the skills that are required. As more data and ways of analyzing them become available, more aspects of the economy, society, and daily life will become dependent on data. It is imperative that educators, administrators, and students begin today to consider how to best prepare for and keep pace with this data-driven era of tomorrow. Undergraduate teaching, in particular, offers a critical link in offering more data science exposure to students and expanding the supply of data science talent. Data Science for Undergraduates: Opportunities and Options offers a vision for the emerging discipline of data science at the undergraduate level. This report outlines some considerations and approaches for academic institutions and others in the broader data science communities to help guide the ongoing transformation of this field.*

*An Introduction to Statistical Learning provides an accessible overview of the field of statistical learning, an essential toolset for making sense of the vast and complex data sets that have emerged in fields ranging from biology to finance to marketing to astrophysics in the past twenty years. This book presents some of the most important modeling and prediction techniques, along with relevant applications. Topics include linear regression, classification, resampling methods, shrinkage approaches, tree-based methods, support vector machines, clustering, and more. Color graphics and real-world examples are used to illustrate the methods presented. Since the goal of this textbook is to facilitate the use of these statistical learning techniques by practitioners in science, industry, and other fields, each chapter contains a tutorial on implementing the analyses and methods presented in R, an extremely popular open source statistical software platform. Two of the authors co-wrote The Elements of Statistical Learning (Hastie, Tibshirani and Friedman, 2nd edition 2009), a popular reference book for statistics and machine learning researchers. An Introduction to Statistical Learning covers many of the same topics, but at a level accessible to a much broader audience. This book is targeted at statisticians and non-statisticians alike who wish to use cutting-edge statistical learning techniques to analyze their data. The text assumes only a previous course in linear regression and no knowledge of matrix algebra.*

Data Science Bookcamp

Machine Learning Techniques for Enterprise Data

The Hallmarks of the Great Beyond in Pytorch, R, Tensorflow, and Python

Essential Tools for Working with Data

Data Science and Machine Learning

Foundational Python for Data Science

Data science and machine learning - two of the world's hottest fields - are attracting talent from a wide variety of technical, business, and liberal arts disciplines. Python, the world's #1 programming language, is also the most popular language for data science and machine learning. This is the first guide specifically designed to help students with widely diverse backgrounds learn foundational Python so they can use it for data science and machine learning. This book is catered to introductory-level college courses on data science. Leading data science instructor and practitioner Kennedy Behrman first walks through the process of learning to code for the first time with Python and Jupyter notebook, then introduces key libraries every Python data science programmer needs to master. Once students have learned these foundations, Behrman introduces intermediate and applied Python techniques for real-world problem-solving. Throughout, Foundational Python for Data Science presents hands-on exercises, learning assessments, case studies, and more - all created with colab (jupyter compatible) notebooks, so students can execute all coding examples interactively without installing or configuring any software.

Go from messy, unstructured artifacts stored in SQL and NoSQL databases to a neat, well-organized dataset with this quick reference for the busy data scientist. Understand text mining, machine learning, and network analysis; process numeric data with the NumPy and Pandas modules; describe and analyze data using statistical and network-theoretical methods; and see actual examples of data analysis at work. This one-stop solution covers the essential data science you need in Python. Data science is one of the fastest-growing disciplines in terms of academic research, student enrollment, and employment. Python, with its flexibility and scalability, is quickly overtaking the R language for data-scientific projects. Keep Python data-science concepts at your fingertips with this modular, quick reference to the tools used to acquire, clean, analyze, and store data. This one-stop solution covers essential Python, databases, network analysis, natural language processing, elements of machine learning, and visualization. Access structured and unstructured text and numeric data from local files, databases, and the Internet. Arrange, rearrange, and clean the data. Work with relational and non-relational databases, data visualization, and simple predictive analysis (regressions, clustering, and decision trees). See how typical data analysis problems are handled. And try your hand at your own solutions to a variety of medium-scale projects that are fun to work on and look good on your resume. Keep this handy quick guide at your side whether you're a student, an entry-level data science professional converting from R to Python, or a seasoned Python developer who doesn't want to memorize every function and option. **What You Need:** You need a decent distribution of Python 3.3 or above that includes at least NLTK, Pandas, NumPy, Matplotlib, Networkx, SciKit-Learn, and BeautifulSoup. A great distribution that meets the requirements is Anaconda, available for free from www.continuum.io. If you plan to set up your own database servers, you also need MySQL (www.mysql.com) and MongoDB (www.mongodb.com). Both packages are free and run on Windows, Linux, and Mac OS.

Written for developers who want build applications using Twisted, this book presents a task-oriented look at this open source, Python-based technology.

Optimize your marketing strategies through analytics and machine learning Key Features Understand how data science drives successful marketing campaigns Use machine learning for better customer engagement, retention, and product recommendations Extract insights from your data to optimize marketing strategies and increase profitability Book Description Regardless of company size, the adoption of data science and machine learning for marketing has been rising in the industry. With this book, you will learn to implement data science techniques to understand the drivers behind the successes and failures of marketing campaigns. This book is a comprehensive guide to help you understand and predict customer behaviors and create more effectively targeted and personalized marketing strategies. This is a practical guide to performing simple-to-advanced tasks, to extract hidden insights from the data and use them to make smart business decisions. You will understand what drives sales and increases customer engagements for your products. You will learn to implement machine learning to forecast which customers are more likely to engage with the products and have high lifetime value. This book will also show you how to use machine learning techniques to understand different customer segments and recommend the right products for each customer. Apart from learning to gain insights into consumer behavior using exploratory analysis, you will also learn the concept of A/B testing and implement it using Python and R. By the end of this book, you will be experienced enough with various data science and machine learning techniques to run and manage successful marketing campaigns for your business. What you will learn Learn how to compute and visualize marketing KPIs in Python and R Master what drives successful marketing campaigns with data science Use machine learning to predict customer engagement and lifetime value Make product recommendations that customers are most likely to buy Learn how to use A/B testing for better marketing decision making Implement machine learning to understand different customer segments Who this book is for If you are a marketing professional, data scientist, engineer, or a student keen to learn how to apply data science to marketing, this book is what you need! It will be beneficial to have some basic knowledge of either Python or R to work through the examples. This book will also be beneficial for beginners as it covers basic-to-advanced data science concepts and applications in marketing with real-life examples.

Foundations, Algorithms, and Applications

Improve your marketing strategies with machine learning using Python and R

A Tour of Data Science

Hands-On Data Science for Marketing

Principles of Data Science

Hands-On Genetic Algorithms with Python