Data Science From Scratch First Principles With Python

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

This is a first-principles-based, practical introduction to the fundamentals of data science aimed at the mathematically-comfortable reader with some programming skills. The book covers: The important parts of Python to know The important parts of Math / Probability / Statistics to know The basics of data science How commonly-used data science techniques work (learning by implementing them) What is Map-Reduce and how to do it in Python Other applications such as NLP, Network Analysis, and more 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.

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.

An introductory textbook offering a low barrier entry to data science; the hands-on approach will appeal to students from a range of disciplines.

A comprehensive overview of data science covering the analytics, programming, and business skills necessary to master the discipline Finding a good data scientist has been likened to hunting for a unicorn: the required combination of technical skills is simply very hard to find in one person. In addition, good data science is not just rote application of trainable skill sets; it requires the ability to think flexibly about all these areas and understand the connections between them. This book provides a crash course in data science, combining all the necessary skills into a unified discipline. Unlike many analytics books, computer science and software engineering are given extensive coverage since they play such a central role in the daily work of a data scientist. The author also describes classic machine learning algorithms, from their mathematical foundations to real-world applications. Visualization tools are reviewed, and their central importance in data science is highlighted. Classical statistics is addressed to help readers think critically about the interpretation of data and its common pitfalls. The clear communication of technical results, which is perhaps the most undertrained of data science skills, is given its own chapter, and all topics are explained in the context of solving real-world data problems. The book also features: • Extensive sample code and tutorials using Python™ along with its technical libraries • Core technologies of "Big Data," including their strengths and limitations and how they can be used to solve real-world problems • Coverage of the practical realities of the tools, keeping theory to a minimum; however, when theory is presented, it is done in an intuitive way to encourage critical thinking and creativity • A wide variety of case studies from industry • Practical advice on the realities of being a data scientist today, including the overall workflow, where time is spent, the types of datasets worked on, and the skill sets needed The Data Science Handbook is an ideal resource for data analysis methodology and big data software tools. The book is appropriate for people who want to practice data science, but lack the required skill sets. This includes software professionals who need to better understand analytics and statisticians who need to understand software. Modern data science is a unified discipline, and it is presented as such. This book is also an appropriate reference for researchers and entry-level graduate students who need to learn real-world analytics and expand their skill set. FIELD CADY is the data scientist at the Allen Institute for Artificial Intelligence, where he develops tools that use machine learning to mine scientific literature. He has also worked at Google and several Big Data startups. He has a BS in physics and math from Stanford University, and an MS in computer science from Carnegie Mellon.

? 55% OFF for Bookstores! Now at $49.95 instead of $59.95! ? Your Customers Will Never Stop To Use This Complete Guide! Did you know that according to Harvard Business Review the Data Scientist is the sexiest job of the 21st century? And for a reason! If "sexy" means having rare qualities that are much in demand, data scientists are already there. They are expensive to hire and, given the very competitive market for their services, difficult to retain. There simply aren't a lot of people with their combination of scientific background and computational and analytical skills. Data Science is all about transforming data into business value using

math and algorithms. And needless to say, Python is the must-know programming language of the 21st century. If you are interested in coding and Data Science, then you must know Python to succeed in these industries! Data Science for Beginners is the perfect place to start learning everything you need to succeed. Contained within these four essential books are the methods, concepts, and important practical examples to help build your foundation for excelling at the discipline that is shaping the modern word. This bundle is perfect for programmers, software engineers, project managers and those who just want to keep up with technology. With these books in your hands, you will: ? Learn Python from scratch including the basic operations, how to install it, data structures and functions, and conditional loops ? Build upon the fundamentals with advanced techniques like Object-Oriented Programming (OOP), Inheritance, and Polymorphism ? Discover the importance of Data Science and how to use it in real-world situations ? Learn the 5 steps of Data Analysis so you can comprehend and analyze data sitting right in front of you ? Increase your income by learning a new, valuable skill that only a select handful of people take the time to learn ? Discover how companies can improve their business through practical examples and explanations ? And Much More! This bundle is essential for anyone who wants to study Data Science and learn how the world is moving to an open-source platform. Whether you are a software engineer or a project manager, jump to the next level by developing a data-driven approach and learning how to define a data-driven vision of your business! Order Your Copy of the Bundle and Let Your Customers Start Their New Career Path Today! Learn Data Science NOW. Stop asking yourself where and how to start. Keep reading and find out how this book can help you with your journey. Are you afraid not to understand the technical language of data science? If so, let me tell you something. We all have to start somewhere. Approaching data science can be overwhelming, not if you have in your hands the right tools since day one. Once you start, I can guarantee you, you will want to learn more and more. Data science is an interdisciplinary subject that brings together three different fields of study. All three fields lie at the intersection of business intelligence and big data. More specifically this book will take you through: Which specific tools and analysis you need to know Various aspects involved in Data Mining Types, Quality and Data Preprocessing Things you must know for machine learning to be successful Utilizations and Procedure of Data Science How to exactly set up the appropriate environment for your machine learning needs....and much more!! Even if you never approached Data Science before, you now have the chance to deeply understand every concept and become more confident in what you want to achieve next. Data Science from Scratch has been written thinking of your needs and how to help you get started. The more you wait, the harder it gets. What are you waiting for? Scroll to the top and select on the right the BUY NOW with 1-Clickbutton.

Create, deploy, and test your Python applications, analyses, and models with ease using Streamlit Key Features Learn how to showcase machine learning models in a Streamlit application effectively and efficiently Become an expert Streamlit creator by getting hands-on with complex application creation Discover how Streamlit enables you to create and deploy apps effortlessly Book Description Streamlit shortens the development time for the creation of data-focused web applications, allowing data scientists to create web app prototypes using Python in hours instead of days. Getting Started with Streamlit for Data Science takes a hands-on approach to helping you learn the tips and tricks that will have you up and running with Streamlit in no time. You'll start with the fundamentals of Streamlit by creating a basic app and gradually build on the foundation by producing high-quality graphics with data visualization and testing machine learning models. As you advance through the chapters, you'll walk through practical examples of both personal data projects and work-related data-focused web applications, and get to grips with more challenging topics such as using Streamlit Components, beautifying your apps, and quick deployment of your new apps. By the end of this book, you'll be able to create dynamic web apps in Streamlit quickly and effortlessly using the power of Python. What you will learn Set up your first development environment and create a basic Streamlit app from scratch Explore methods for uploading, downloading, and manipulating data in Streamlit apps Create dynamic visualizations in Streamlit using built-in and imported Python libraries Discover strategies for creating and deploying machine learning models in Streamlit Use Streamlit sharing for one-click deployment Beautify Streamlit apps using themes, Streamlit Components, and Streamlit sidebar Implement best practices for prototyping your data science work with Streamlit Who this book is for This book is for data scientists and machine learning enthusiasts who want to create web apps using Streamlit. Whether you're a junior data scientist looking to deploy your first machine learning project in Python to improve your resume or a senior data scientist who wants to use Streamlit to make convincing and dynamic data analyses, this book will help you get there! Prior knowledge of Python programming will assist with understanding the concepts covered.

Now that people are aware that data can make the difference in an election or a business model, data science as an occupation is gaining ground. But how can you get started working in a wide-ranging, interdisciplinary field that's so clouded in hype? This insightful book, based on Columbia University's Introduction to Data Science class, tells you what you need to know. In many of these chapter-long lectures, data scientists from companies such as Google, Microsoft, and eBay share new algorithms, methods, and models by presenting case studies and the code they use. If you're familiar with linear algebra, probability, and statistics, and have programming experience, this book is an ideal introduction to data science. Topics include: Statistical inference, exploratory data analysis, and the data science process Algorithms Spam filters, Naive Bayes, and data wrangling Logistic regression Financial modeling Recommendation engines and causality Data visualization Social networks and data journalism Data engineering, MapReduce, Pregel, and Hadoop Doing Data Science is collaboration between course instructor Rachel Schutt, Senior VP of Data Science at News Corp, and data science consultant Cathy O'Neil, a senior data scientist at Johnson Research Labs, who attended and blogged about the course.

A surprisingly simple way for students to master any subject--based on one of the world's most popular online courses and the bestselling book A Mind for Numbers A Mind for Numbers and its wildly popular online companion course "Learning How to Learn" have empowered more than two million learners of all ages from around the world to master subjects that they once struggled with. Fans often wish they'd discovered these learning strategies earlier and ask how they can help their kids master these skills as well. Now in this new book for kids and teens, the authors reveal how to make the most of time spent studying. We all have the tools to learn what might not seem to come naturally to us at first--the secret is to understand how the brain works so we can unlock its power. This book explains: • Why sometimes letting your mind wander is an important part of the learning process • How to avoid "rut think" in order to think outside the box • Why having a poor memory can be a good thing • The value of metaphors in developing understanding • A simple, yet powerful, way to stop procrastinating Filled with illustrations, application questions, and exercises, this book makes learning easy and fun.

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 and Dask 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 data pipeline 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 that extend 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 meeting 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 Datashader 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 and the PyData 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 Why 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 Datashader PART 3 - Extending and deploying Dask Working with Bags and Arrays Machine learning with Dask-ML Scaling and deploying Dask

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.

Data Science from ScratchO'Reilly Media

Data Science from scratch for Beginners with Hands-On Projects Are you looking for a hands-on approach to learn Data Science fast? Do you need to start learning Machine Learning Fundamentals? This book is for you. This book is written for beginners and novices who want to develop fundamental data science skills and learn how to build models that learn useful information from data. This book will prepare the learner for a career or further learning that involves more advanced topics. It contains introduction and very basic concepts used in data science. The learner is not required to have any prior knowledge but some basic knowledge of mathematics is required. The working of each algorithm is traced back to its origin in probability, statistics or linear algebra which helps learner to understand the topics better. The concepts of probability and statistics are defined and explained at rudimentary level to make things simple and easy to comprehend. For intuitive understanding, algorithms have been explained through proper visualizations and various examples. While we will focus more on the techniques normally used in Data Science, we will also explain, in-details, all the Python libraries used in any data science project. What this book offers... You will learn all about data Science starting from Python Coding, Data Manipulation and Preprocessing, Data Visualization then data modeling using Python. All the modules will contain hands-on projects using real-world datasets. Clear and Easy to Understand Solutions All solutions in this book are extensively tested by a group of beta readers. The solutions provided are simplified as much as possible so that they can serve as examples for you to refer to when you are learning a new skill. What this book aims to do... This book is written with one goal in mind - to help beginners overcome their initial obstacles to learn Data Science from Scratch. A lot of times, newbies tend to feel intimidated by Data Science Models and Coding. The goal of this book is to isolate the different concepts so that beginners can gradually gain competency in the fundamentals and keys concepts before working on a project at the end of each chapter. Beginners in Data Science does not have to be scary or frustrating when you take one step at a time. Ready to start practicing and start learning Data Science from Scratch? Click the BUY button now to download this book Topics Covered: Preliminary to Understand Data Science Overview of Python and Data Processing Statistics and Probability Supervised Learning Techniques Unsupervised Learning Techniques Neural Networks and Deep Learning Reinforcement Learning Techniques ..and more... Click the BUY button and download the book now to start learning Data Science. ** MONEY BACK GUARANTEE BY AMAZON ** If you aren't satisfied, for more information about the amazon refund service please go to the amazon help platform or contact us by sending an email at contact@aispublishing.net. **GET YOUR COPY NOW, the price will be 22.99$ soon**

******Free eBook for customers who purchase the print book from Amazon****** Are you thinking of becoming a data analyst using Python? If you are looking for a complete guide to data analysis using Python language and its library that will help you to become an effective data scientist, this book is for you. From AI Sciences Publisher Our books may be the best one for beginners; it's a step-by-step guide for any person who wants to start learning Artificial Intelligence and Data Science from scratch. It will help you in preparing a solid foundation and learn any other high-level courses. To get the most out of the concepts that would be covered, readers are advised to adopt hands on approach, which would lead to better mental representations. Step By Step Guide and Visual Illustrations and Examples The Book give complete instructions for manipulating, processing, cleaning, modeling and crunching datasets in Python. This is a hands-on guide with practical case studies of data analysis problems effectively. You will learn pandas, NumPy, IPython, and Jupiter in the Process. Target Users This book is a practical introduction to data science tools in Python. It is ideal for analyst's beginners to Python and for Python programmers new to data science and computer science. Instead of tough math formulas, this book contains several graphs and images. What's Inside This Book? Introduction Why Choose Python for Data Science & Machine Learning Prerequisites & Reminders Python Quick Review Overview & Objectives A Quick Example Getting & Processing Data Data Visualization Supervised & Unsupervised Learning Regression Simple Linear Regression Multiple Linear Regression Decision Tree Random Forest Classification Logistic Regression K-Nearest Neighbors

Decision Tree Classification Random Forest Classification Clustering Goals & Uses of Clustering K-Means Clustering Anomaly Detection Association Rule Learning Explanation Apriori Reinforcement Learning What is Reinforcement Learning Comparison with Supervised & Unsupervised Learning Applying Reinforcement Learning Neural Networks An Idea of How the Brain Works Potential & Constraints Here's an Example Natural Language Processing Analyzing Words & Sentiments Using NLTK Model Selection & Improving Performance Sources & References Frequently Asked Questions Q: Is this book for me and do I need programming experience? A: if you want to smash Python for data analysis, this book is for you. Little programming experience is required. If you already wrote a few lines of code and recognize basic programming statements, you'll be OK. Q: Does this book include everything I need to become a data science expert? A: Unfortunately, no. This book is designed for readers taking their first steps in data analysis and further learning will be required beyond this book to master all aspects. Q: Can I have a refund if this book is not fitted for me? A: Yes, Amazon refund you if you aren't satisfied, for more information about the amazon refund service please go to the amazon help platform. We will also be happy to help you if you send us an email at contact@aisciences.net. AI Sciences Company offers you a free eBooks at http: //aisciences.net/free/

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

An authorised reissue of the long out of print classic textbook, Advanced Calculus by the late Dr Lynn Loomis and Dr Shlomo Sternberg both of Harvard University has been a revered but hard to find textbook for the advanced calculus course for decades. This book is based on an honors course in advanced calculus that the authors gave in the 1960's. The foundational material, presented in the unstarred sections of Chapters 1 through 11, was normally covered, but different applications of this basic material were stressed from year to year, and the book therefore contains more material than was covered in any one year. It can accordingly be used (with omissions) as a text for a year's course in advanced calculus, or as a text for a three-semester introduction to analysis. The prerequisites are a good grounding in the calculus of one variable from a mathematically rigorous point of view, together with some acquaintance with linear algebra. The reader should be familiar with limit and continuity type arguments and have a certain amount of mathematical sophistication. As possible introductory texts, we mention Differential and Integral Calculus by R Courant, Calculus by T Apostol, Calculus by M Spivak, and Pure Mathematics by G Hardy. The reader should also have some experience with partial derivatives. In overall plan the book divides roughly into a first half which develops the calculus (principally the differential calculus) in the setting of normed vector spaces, and a second half which deals with the calculus of differentiable manifolds.

"Turn yourself into a Data Head. You'll become a more valuable employee and make your organization more successful." Thomas H. Davenport, Research Fellow, Author of Competing on Analytics, Big Data @ Work, and The AI Advantage You've heard the hype around data—now get the facts. In Becoming a Data Head: How to Think, Speak, and Understand Data Science, Statistics, and Machine Learning, award-winning data scientists Alex Gutman and Jordan Goldmeier pull back the curtain on data science and give you the language and tools necessary to talk and think critically about it. You'll learn how to: Think statistically and understand the role variation plays in your life and decision making Speak intelligently and ask the right questions about the statistics and results you encounter in the workplace Understand what's really going on with machine learning, text analytics, deep learning, and artificial intelligence Avoid common pitfalls when working with and interpreting data Becoming a Data Head is a complete guide for data science in the workplace: covering everything from the personalities you'll work with to the math behind the algorithms. The authors have spent years in data trenches and sought to create a fun, approachable, and eminently readable book. Anyone can become a Data Head—an active participant in data science, statistics, and machine learning. Whether you're a business professional, engineer, executive, or aspiring data scientist, this book is for you.

The funniest book ever written about why your religion is false!Whether you're a Christian or a Jew, a Muslim or a Hindu, a Rasta or a Jain, an Environmentalist or a Cheondoist, a Scientologist or a Giant Stone Head Worshipper, your religion is false.But don't feel bad -- so is everyone else's! When you want to know what not to believe, this is the only book you need.In addition, you'll learn* Why "god" doesn't exist* Why there's no such thing as a "soul"* How to find "meaning" in a religion-less world* Which of your religious heroes are pedophiles* Why "religious tolerance" is a terrible ideaAnd, as a

bonus, the greatest religious joke ever told. You can't afford not to read this book!

The Hands-On, Example-Rich Introduction to Pandas Data Analysis in Python Today, analysts must manage data characterized by extraordinary variety, velocity, and volume. Using the open source Pandas library, you can use Python to rapidly automate and perform virtually any data analysis task, no matter how large or complex. Pandas can help you ensure the veracity of your data, visualize it for effective decision-making, and reliably reproduce analyses across multiple datasets. Pandas for Everyone brings together practical knowledge and insight for solving real problems with Pandas, even if you're new to Python data analysis. Daniel Y. Chen introduces key concepts through simple but practical examples, incrementally building on them to solve more difficult, real-world problems. Chen gives you a jumpstart on using Pandas with a realistic dataset and covers combining datasets, handling missing data, and structuring datasets for easier analysis and visualization. He demonstrates powerful data cleaning techniques, from basic string manipulation to applying functions simultaneously across dataframes. Once your data is ready, Chen guides you through fitting models for prediction, clustering, inference, and exploration. He provides tips on performance and scalability, and introduces you to the wider Python data analysis ecosystem. Work with DataFrames and Series, and import or export data Create plots with matplotlib, seaborn, and pandas Combine datasets and handle missing data Reshape, tidy, and clean datasets so they're easier to work with Convert data types and manipulate text strings Apply functions to scale data manipulations Aggregate, transform, and filter large datasets with groupby Leverage Pandas' advanced date and time capabilities Fit linear models using statsmodels and scikit-learn libraries Use generalized linear modeling to fit models with different response variables Compare multiple models to select the "best" Regularize to overcome overfitting and improve performance Use clustering in unsupervised machine learning

"This book is for students or anyone, with limited or no prior programming, statistics, and data analytics knowledge. This short guide is ideal for absolute beginners, or anyone who wants to acquire a basic working knowledge of data science. It is an excellent guide if you want to learn about the principals of data science from scratch, in just a few hours. The author discussed everything that you need to know about data science. First, you are guided to learn the meaning of data science. The history of data science has been discussed to help you know how people came to realize that data is a rich source of knowledge and intelligence. The theories underlying data science have been discussed. Examples include decision and estimation theories. The author discussed the various machine learning algorithms used in data science and the various steps one has to undergo when performing data science tasks, from data collection to data presentation and visualization. The author helps you to know the various ways through which you can apply data science in your business for increased profits. A simple language has been used to ensure ease of understanding, especially for beginners." --

Through a series of recent breakthroughs, deep learning has boosted the entire field of machine learning. Now, even programmers who know close to nothing about this technology can use simple, efficient tools to implement programs capable of learning from data. This practical book shows you how. By using concrete examples, minimal theory, and two production-ready Python frameworks—Scikit-Learn and TensorFlow—author Aurélien Géron helps you gain an intuitive understanding of the concepts and tools for building intelligent systems. You'll learn a range of techniques, starting with simple linear regression and progressing to deep neural networks. With exercises in each chapter to help you apply what you've learned, all you need is programming experience to get started. Explore the machine learning landscape, particularly neural nets Use Scikit-Learn to track an example machine-learning project end-to-end Explore several training models, including support vector machines, decision trees, random forests, and ensemble methods Use the TensorFlow library to build and train neural nets Dive into neural net architectures, including convolutional nets, recurrent nets, and deep reinforcement learning Learn techniques for training and scaling deep neural nets

With the resurgence of neural networks in the 2010s, deep learning has become essential for machine learning practitioners and even many software engineers. This book provides a comprehensive introduction for data scientists and software engineers with machine learning experience. You'll start with deep learning basics and move quickly to the details of important advanced architectures, implementing everything from scratch along the way. Author Seth Weidman shows you how neural networks work using a first principles approach. You'll learn how to apply multilayer neural networks, convolutional neural networks, and recurrent neural networks from the ground up. With a thorough understanding of how neural networks work mathematically, computationally, and conceptually, you'll be set up for success on all future deep learning projects. This book provides: Extremely clear and thorough mental models—accompanied by working code examples and mathematical explanations—for understanding neural networks Methods for implementing multilayer neural networks from scratch, using an easy-to-understand object-oriented framework Working implementations and clear-cut explanations of convolutional and recurrent neural networks Implementation of these neural network concepts using the popular PyTorch framework

Explore fundamental to advanced Python 3 topics in six steps, all designed to make you a worthy practitioner. This updated version's approach is based on the "six degrees of separation" theory, which states that everyone and everything is a maximum of six steps away and presents each topic in two parts: theoretical concepts and practical implementation using suitable Python 3 packages. You'll start with the fundamentals of Python 3 programming language, machine learning history, evolution, and the system development frameworks. Key data mining/analysis concepts, such as exploratory analysis, feature dimension reduction, regressions, time series forecasting and their efficient implementation in Scikit-learn are covered as well. You'll also learn commonly used model diagnostic and tuning techniques. These include optimal probability cutoff point for class creation, variance, bias, bagging, boosting, ensemble voting, grid search, random search, Bayesian optimization, and the noise reduction technique for IoT data. Finally, you'll review advanced text mining techniques, recommender systems, neural networks, deep learning, reinforcement learning

techniques and their implementation. All the code presented in the book will be available in the form of iPython notebooks to enable you to try out these examples and extend them to your advantage. What You'll Learn Understand machine learning development and frameworks Assess model diagnosis and tuning in machine learning Examine text mining, natuarl language processing (NLP), and recommender systems Review reinforcement learning and CNN Who This Book Is For Python developers, data engineers, and machine learning engineers looking to expand their knowledge or career into machine learning area.

"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"--

We're living in a digital world. Most of our global economy is digital and the sheer volume of data is stupendous. It's 2020 and we're living in the future. Data Scientist is one of the hottest job on the market right now. Demand for data science is huge and will only grow, and it seems like it will grow much faster than the actual number of data scientists. So if you want to make a career change and become a data scientist, now is the time. This book will guide you through the process. From my experience of working with multiple companies as a project manager, a data science consultant or a CTO, I was able to see the process of hiring data scientists and building data science teams. I know what's important to land your first job as a data scientist, what skills you should acquire, what you should show during a job interview.

Data science libraries, frameworks, modules, and toolkits are great for doing data science, but they're also a good way to dive into the discipline without actually understanding data science. In this book, you'll learn how many of the most fundamental data science tools and algorithms work by implementing them from scratch. If you have an aptitude for mathematics and some programming skills, author Joel Grus will help you get comfortable with the math and statistics at the core of data science, and with hacking skills you need to get started as a data scientist. Today's messy glut of data holds answers to questions no one's even thought to ask. This book provides you with the know-how to dig those answers out. Get a crash course in Python Learn the basics of linear algebra, statistics, and probability—and understand how and when they're used in data science Collect, explore, clean, munge, and manipulate data Dive into the fundamentals of machine learning Implement models such as k-nearest Neighbors, Naive Bayes, linear and logistic regression, decision trees, neural networks, and clustering Explore recommender systems, natural language processing, network analysis, MapReduce, and databases

"This book is so good. I wish I'd written it." -- Tim Hopper (@tdhopper)"Highly recommended: a grand tour of computer science theory and practical software engineering, explored through the lens of 10 Fizz Buzz solutions in Python. Outstanding." -- Paco Nathan (@pacoid)"I'd never have thought a book about Fizz Buzz would make me a better programmer, but I was wrong. Joel in the course of 10 chapters does a broad survey of core Python concepts, software design and testing, mathematics, and more (including deep learning) using Fizz Buzz as the guiding example. It's that rare technical book that remains engaging, entertaining, and accessible." -- Binal Patel (@binalkp91)More real Python tips than any "Python tricks" book! From a Python beginner to an experienced ML practitioner, you're bound to learn something about the language and its application to a progressive level of algorithmic applications. Recommended for the anyone looking to "level up" their Python or problem solving skills! -- Tom Marthaler (@tmarthal)Fizz Buzz is the following (simple) problem: Print the numbers from 1 to 100, except that if the number is divisible by 3, instead print "fizz"; if the number is divisible by 5, instead print "buzz"; and if the number is divisible by 15, instead print "fizzbuzz".It originated as a children's game, but has since taken on a new life as a lowest-common-denominator litmus test for assessing computer programmers.If you are an experienced programmer, it is an extremely easy problem to solve. Because of this, it has taken on a third life as the prototypical bad interview problem. Everyone knows that it's the question you ask people to make sure that they're not completely incompetent as programmers. Accordingly, if your interviewer asks you to solve it, he's suggesting he thinks it possible that you're completely incompetent as a programmer. You would not be wrong to feel insulted!My association with this problem began in 2016, when I wrote a blog post called Fizz Buzz in Tensorflow, the (possibly fictional) story of one such insulted programmer who decided to show up his interviewer by approaching Fizz Buzz as a deep learning problem. This post went modestly viral, and ever since then I have been seen as a thought leader in the Fizz Buzz space.Accordingly, over the years I have come up with and/or collected various other stupid and/or clever ways of solving Fizz Buzz. I have not blogged about them, as I am not the sort of person who beats a joke to death, but occasionally I will tweet about them, and recently in response someone suggested that I write a book on "100 Ways of Writing Fizz Buzz in Python."Now, I could probably come up with 100 ways of solving Fizz Buzz, but most of them would not be very interesting. Luckily for you, I was able to come up with 10 that are interesting in various ways, each of which turned out to be a good launching-off point for (sometimes meandering) discussions of various aspects of coding, Python, Fizz Buzz, mathematics, software design, technical interviewing, and various other topics.Hence "Ten Essays on Fizz Buzz".In many ways this is a strange book. Its goal is not to teach you a specific field or a specific technology. I hope you will learn a lot from reading it, but it's not really a book that you'd read in order to learn anything in particular. Most technical books are about specific technical topics; this one sort of isn't.Nonetheless, it is a technical book. Each essay contains code that implements a different solution of Fizz Buzz. Each essay uses code to illustrate its ideas. Each essay represents my current best thinking about how to solve problems using code. If you have a coding job, you should feel no reluctance to expense this book to your employer Python's simplicity lets you become productive quickly, but this often means you aren't using everything it has to offer. With this hands-on guide, you'll learn how to write effective, idiomatic Python code by leveraging its best—and possibly most neglected—features. Author Luciano Ramalho takes you through Python's core language features and libraries, and shows you how to make your code shorter, faster, and more readable at the same time. Many experienced programmers try to bend Python to fit patterns they learned from other languages, and never discover Python features outside of their

experience. With this book, those Python programmers will thoroughly learn how to become proficient in Python 3. This book covers: Python data model: understand how special methods are the key to the consistent behavior of objects Data structures: take full advantage of built-in types, and understand the text vs bytes duality in the Unicode age Functions as objects: view Python functions as first-class objects, and understand how this affects popular design patterns Object-oriented idioms: build classes by learning about references, mutability, interfaces, operator overloading, and multiple inheritance Control flow: leverage context managers, generators, coroutines, and concurrency with the concurrent.futures and asyncio packages Metaprogramming: understand how properties, attribute descriptors, class decorators, and metaclasses work

Written by renowned data science experts Foster Provost and Tom Fawcett, Data Science for Business introduces the fundamental principles of data science, and walks you through the "data-analytic thinking" necessary for extracting useful knowledge and business value from the data you collect. This guide also helps you understand the many data-mining techniques in use today. Based on an MBA course Provost has taught at New York University over the past ten years, Data Science for Business provides examples of real-world business problems to illustrate these principles. You'll not only learn how to improve communication between business stakeholders and data scientists, but also how participate intelligently in your company's data science projects. You'll also discover how to think data-analytically, and fully appreciate how data science methods can support business decision-making. Understand how data science fits in your organization—and how you can use it for competitive advantage Treat data as a business asset that requires careful investment if you're to gain real value Approach business problems data-analytically, using the data-mining process to gather good data in the most appropriate way Learn general concepts for actually extracting knowledge from data Apply data science principles when interviewing data science job candidates

Master Data Analytics Hands-On by Solving Fascinating Problems You'll Actually Enjoy! Harvard Business Review recently called data science "The Sexiest Job of the 21st Century." It's not just sexy: For millions of managers, analysts, and students who need to solve real business problems, it's indispensable. Unfortunately, there's been nothing easy about learning data science–until now. Getting Started with Data Science takes its inspiration from worldwide best-sellers like Freakonomics and Malcolm Gladwell's Outliers: It teaches through a powerful narrative packed with unforgettable stories. Murtaza Haider offers informative, jargon-free coverage of basic theory and technique, backed with plenty of vivid examples and hands-on practice opportunities. Everything's software and platform agnostic, so you can learn data science whether you work with R, Stata, SPSS, or SAS. Best of all, Haider teaches a crucial skillset most data science books ignore: how to tell powerful stories using graphics and tables. Every chapter is built around real research challenges, so you'll always know why you're doing what you're doing. You'll master data science by answering fascinating questions, such as: • Are religious individuals more or less likely to have extramarital affairs? • Do attractive professors get better teaching evaluations? • Does the higher price of cigarettes deter smoking? • What determines housing prices more: lot size or the number of bedrooms? • How do teenagers and older people differ in the way they use social media? • Who is more likely to use online dating services? • Why do some purchase iPhones and others Blackberry devices? • Does the presence of children influence a family's spending on alcohol? For each problem, you'll walk through defining your question and the answers you'll need; exploring how others have approached similar challenges; selecting your data and methods; generating your statistics; organizing your report; and telling your story. Throughout, the focus is squarely on what matters most: transforming data into insights that are clear, accurate, and can be acted upon.

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 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

If you are looking to start a new career that is in high demand, then you need to continue reading. Data scientists are changing the way big data is used in different institutions.

A concise introduction to the emerging field of data science, explaining its evolution, relation to machine learning, current uses, data infrastructure issues, and ethical challenges. The goal of data science is to improve decision making through the analysis of data. Today data science determines the ads we see online, the books and movies that are recommended to us online, which emails are filtered into our spam folders, and even how much we pay for health insurance. This volume in the MIT Press Essential Knowledge series offers a concise introduction to the emerging field of data science, explaining its evolution, current uses, data

infrastructure issues, and ethical challenges. It has never been easier for organizations to gather, store, and process data. Use of data science is driven by the rise of big data and social media, the development of high-performance computing, and the emergence of such powerful methods for data analysis and modeling as deep learning. Data science encompasses a set of principles, problem definitions, algorithms, and processes for extracting non-obvious and useful patterns from large datasets. It is closely related to the fields of data mining and machine learning, but broader in scope. This book offers a brief history of the field, introduces fundamental data concepts, and describes the stages in a data science project. It considers data infrastructure and the challenges posed by integrating data from multiple sources, introduces the basics of machine learning, and discusses how to link machine learning expertise with real-world problems. The book also reviews ethical and legal issues, developments in data regulation, and computational approaches to preserving privacy. Finally, it considers the future impact of data science and offers principles for success in data science projects.

Master how to use the Julia language to solve business critical data science challenges. After covering the importance of Julia to the data science community and several essential data science principles, we start with the basics including how to install Julia and its powerful libraries. Many examples are provided as we illustrate how to leverage each Julia command, dataset, and function. Specialized script packages are introduced and described. Hands-on problems representative of those commonly encountered throughout the data science pipeline are provided, and we guide you in the use of Julia in solving them using published datasets. Many of these scenarios make use of existing packages and built-in functions, as we cover: 1. 1. An overview of the data science pipeline along with an example illustrating the key points, implemented in Julia 2. 2. Options for Julia IDEs 3. 3. Programming structures and functions 4. 4. Engineering tasks, such as importing, cleaning, formatting and storing data, as well as performing data preprocessing 5. 5. Data visualization and some simple yet powerful statistics for data exploration purposes 6. 6. Dimensionality reduction and feature evaluation 7. 7. Machine learning methods, ranging from unsupervised (different types of clustering) to supervised ones (decision trees, random forests, basic neural networks, regression trees, and Extreme Learning Machines) 8. 8. Graph analysis including pinpointing the connections among the various entities and how they can be mined for useful insights. Each chapter concludes with a series of questions and exercises to reinforce what you learned. The last chapter of the book will guide you in creating a data science application from scratch using Julia.

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

An introduction to a broad range of topics in deep learning, covering mathematical and conceptual background, deep learning techniques used in industry, and research perspectives. "Written by three experts in the field, Deep Learning is the only comprehensive book on the subject." —Elon Musk, cochair of OpenAI; cofounder and CEO of Tesla and SpaceX Deep learning is a form of machine learning that enables computers to learn from experience and understand the world in terms of a hierarchy of concepts. Because the computer gathers knowledge from experience, there is no need for a human computer operator to formally specify all the knowledge that the computer needs. The hierarchy of concepts allows the computer to learn complicated concepts by building them out of simpler ones; a graph of these hierarchies would be many layers deep. This book introduces a broad range of topics in deep learning. The text offers mathematical and conceptual background, covering relevant concepts in linear algebra, probability theory and information theory, numerical computation, and machine learning. It describes deep learning techniques used by practitioners in industry, including deep feedforward networks, regularization, optimization algorithms, convolutional networks, sequence modeling, and practical methodology; and it surveys such applications as natural language processing, speech recognition, computer vision, online recommendation systems, bioinformatics, and videogames. Finally, the book offers research perspectives, covering such theoretical topics as linear factor models, autoencoders, representation learning, structured probabilistic models, Monte Carlo methods, the partition function, approximate inference, and deep generative models. Deep Learning can be used by undergraduate or graduate students planning careers in either industry or research, and by software engineers who want to begin using deep learning in their products or platforms. A website offers supplementary material for both readers and instructors.

Statistical methods are a key part of 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

Gain hands-on experience with industry-standard data analysis and machine learning tools in Python Key Features Learn techniques to use data to identify the exact problem to be solved Visualize data using different graphs Identify how to select an appropriate algorithm for data extraction Book Description Data Science Projects with Python is designed to give you practical guidance on industry-standard data analysis and machine learning tools in Python, with the help of realistic data. The book will help you understand how you can use pandas and Matplotlib to critically examine a dataset with summary statistics and graphs, and extract the insights you seek to derive. You will continue to build on your knowledge as you learn how to prepare data and feed it to machine learning algorithms, such as regularized logistic regression and random forest, using the scikit-learn package.

You'll discover how to tune the algorithms to provide the best predictions on new and, unseen data. As you delve into later chapters, you'll be able to understand the working and output of these algorithms and gain insight into not only the predictive capabilities of the models but also their reasons for making these predictions. By the end of this book, you will have the skills you need to confidently use various machine learning algorithms to perform detailed data analysis and extract meaningful insights from unstructured data. What you will learn Install the required packages to set up a data science coding environment Load data into a Jupyter Notebook running Python Use Matplotlib to create data visualizations Fit a model using scikit-learn Use lasso and ridge regression to reduce overfitting Fit and tune a random forest model and compare performance with logistic regression Create visuals using the output of the Jupyter Notebook Who this book is for If you are a data analyst, data scientist, or a business analyst who wants to get started with using Python and machine learning techniques to analyze data and predict outcomes, this book is for you. Basic knowledge of computer programming and data analytics is a must. Familiarity with mathematical concepts such as algebra and basic statistics will be useful.

Copyright: 4870cf119448b259c91c5358116473c7

You'll discover how to tune the algorithms to provide the best predictions on new and, unseen data. As you delve into later chapters, you'll be able to understand the working and output of these algorithms and gain insight into not only the predictive capabilities of the models but also their reasons for making these predictions. By the end of this book, you will have the skills you need to confidently use various machine learning algorithms to perform detailed data analysis and extract meaningful insights from unstructured data. What you will learn Install the required packages to set up a data science coding environment Load data into a Jupyter Notebook running Python Use Matplotlib to create data visualizations Fit a model using scikit-learn Use lasso and ridge regression to reduce overfitting Fit and tune a random forest model and compare performance with logistic regression Create visuals using the output of the Jupyter Notebook Who this book is for If you are a data analyst, data scientist, or a business analyst who wants to get started with using Python and machine learning techniques to analyze data and predict outcomes, this book is for you. Basic knowledge of computer programming and data analytics is a must. Familiarity with mathematical concepts such as algebra and basic statistics will be useful.