

Data Science With Java Practical Methods For Scientists And Engineers

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

The book provides a one-stop solution for getting into data science with Python and teaches how to extract insights from data.

This practical text contains fairly "traditional" coverage of data structures with a clear and complete use of algorithm analysis, and some emphasis on file processing techniques as relevant to modern programmers. It fully integrates OO programming with these topics, as part of the detailed presentation of OO programming itself. Chapter topics include lists, stacks, and queues; binary and general trees; graphs; file processing and external sorting; searching; indexing; and limits to computation. For programmers who need a good reference on data structures.

Whether you are a software developer, systems architect, data analyst, or business analyst, if you want to take advantage of data mining in the development of advanced analytic applications, Java Data Mining, JDM, the new standard now implemented in core DBMS and data mining/analysis software, is a key solution component. This book is the essential guide to the usage of the JDM standard interface, written by contributors to the JDM standard. Data mining introduction - an overview of data mining and the problems it can address across industries; JDM's place in strategic solutions to data mining-related problems JDM essentials - concepts, design approach and design issues, with detailed code examples in Java; a Web Services interface to enable JDM functionality in an SOA environment; and illustration of JDM XML Schema for JDM objects JDM in practice - the use of JDM from vendor implementations and approaches to customer applications, integration, and usage; impact of data mining on IT infrastructure; a how-to guide for building applications that use the JDM API Free, downloadable KJDM source code referenced in the book available here Data Science is booming thanks to R and Python, but Java brings the robustness, convenience, and ability to scale

critical to today's data science applications. With this practical book, Java software engineers looking to add data science skills will take a logical journey through the data science pipeline. Author Michael Brzustowicz explains the basic math theory behind each step of the data science process, as well as how to apply these concepts with Java. You'll learn the critical roles that data IO, linear algebra, statistics, data operations, learning and prediction, and Hadoop MapReduce play in the process. Throughout this book, you'll find code examples you can use in your applications. Examine methods for obtaining, cleaning, and arranging data into its purest form Understand the matrix structure that your data should take Learn basic concepts for testing the origin and validity of data Transform your data into stable and usable numerical values Understand supervised and unsupervised learning algorithms, and methods for evaluating their success Get up and running with MapReduce, using customized components suitable for data science algorithms Examine the techniques and Java tools supporting the growing field of data science About This Book Your entry ticket to the world of data science with the stability and power of Java Explore, analyse, and visualize your data effectively using easy-to-follow examples Make your Java applications more capable using machine learning Who This Book Is For This book is for Java developers who are comfortable developing applications in Java. Those who now want to enter the world of data science or wish to build intelligent applications will find this book ideal. Aspiring data scientists will also find this book very helpful. What You Will Learn Understand the nature and key concepts used in the field of data science Grasp how data is collected, cleaned, and processed Become comfortable with key data analysis techniques See specialized analysis techniques centered on machine learning Master the effective visualization of your data Work with the Java APIs and techniques used to perform data analysis In Detail Data science is concerned with extracting knowledge and insights from a wide variety of data sources to analyse patterns or predict future behaviour. It draws from a wide array of disciplines including statistics, computer science, mathematics, machine learning, and data mining. In this book, we cover the important data science concepts and how they are supported by Java, as well as the often statistically challenging techniques, to provide you with an understanding of their purpose and application. The book starts with an introduction of data science, followed by the basic data science tasks of data collection, data cleaning, data analysis, and data visualization. This is followed by a discussion of statistical techniques and more advanced topics including machine learning, neural networks, and deep learning. The next section examines the major categories of data analysis including text, visual, and audio data, followed by a discussion of resources that support parallel implementation. The final chapter illustrates an in-depth data science problem and provides a comprehensive, Java-based solution. Due to the nature of the topic, simple examples of techniques are presented early followed by a more detailed treatment later in the book. This permits a more natural introduction to the techniques and concepts presented in the book. Style and approach This

book follows a tutorial approach, providing examples of each of the major concepts covered. With a step-by-step instructional style, this book covers various facets of data science and will get you up and running quickly.

Tackle the real-world complexities of modern machine learning with innovative, cutting-edge, techniques About This Book Fully-coded working examples using a wide range of machine learning libraries and tools, including Python, R, Julia, and Spark Comprehensive practical solutions taking you into the future of machine learning Go a step further and integrate your machine learning projects with Hadoop Who This Book Is For This book has been created for data scientists who want to see machine learning in action and explore its real-world application. With guidance on everything from the fundamentals of machine learning and predictive analytics to the latest innovations set to lead the big data revolution into the future, this is an unmissable resource for anyone dedicated to tackling current big data challenges. Knowledge of programming (Python and R) and mathematics is advisable if you want to get started immediately. What You Will Learn Implement a wide range of algorithms and techniques for tackling complex data Get to grips with some of the most powerful languages in data science, including R, Python, and Julia Harness the capabilities of Spark and Hadoop to manage and process data successfully Apply the appropriate machine learning technique to address real-world problems Get acquainted with Deep learning and find out how neural networks are being used at the cutting-edge of machine learning Explore the future of machine learning and dive deeper into polyglot persistence, semantic data, and more In Detail Finding meaning in increasingly larger and more complex datasets is a growing demand of the modern world. Machine learning and predictive analytics have become the most important approaches to uncover data gold mines. Machine learning uses complex algorithms to make improved predictions of outcomes based on historical patterns and the behaviour of data sets. Machine learning can deliver dynamic insights into trends, patterns, and relationships within data, immensely valuable to business growth and development. This book explores an extensive range of machine learning techniques uncovering hidden tricks and tips for several types of data using practical and real-world examples. While machine learning can be highly theoretical, this book offers a refreshing hands-on approach without losing sight of the underlying principles. Inside, a full exploration of the various algorithms gives you high-quality guidance so you can begin to see just how effective machine learning is at tackling contemporary challenges of big data. This is the only book you need to implement a whole suite of open source tools, frameworks, and languages in machine learning. We will cover the leading data science languages, Python and R, and the underrated but powerful Julia, as well as a range of other big data platforms including Spark, Hadoop, and Mahout. Practical Machine Learning is an essential resource for the modern data scientists who want to get to grips with its real-world application. With this book, you will not only learn the fundamentals of machine learning but dive deep into the complexities of real world data before moving on to using

Hadoop and its wider ecosystem of tools to process and manage your structured and unstructured data. You will explore different machine learning techniques for both supervised and unsupervised learning; from decision trees to Naive Bayes classifiers and linear and clustering methods, you will learn strategies for a truly advanced approach to the statistical analysis of data. The book also explores the cutting-edge advancements in machine learning, with worked examples and guidance on deep learning and reinforcement learning, providing you with practical demonstrations and samples that help take the theory—and mystery—out of even the most advanced machine learning methodologies. Style and approach A practical data science tutorial designed to give you an insight into the practical application of machine learning, this book takes you through complex concepts and tasks in an accessible way. Featuring information on a wide range of data science techniques, Practical Machine Learning is a comprehensive data science resource.

Get the most out of the popular Java libraries and tools to perform efficient data analysis About This Book Get your basics right for data analysis with Java and make sense of your data through effective visualizations. Use various Java APIs and tools such as Rapidminer and WEKA for effective data analysis and machine learning. This is your companion to understanding and implementing a solid data analysis solution using Java Who This Book Is For If you are a student or Java developer or a budding data scientist who wishes to learn the fundamentals of data analysis and learn to perform data analysis with Java, this book is for you. Some familiarity with elementary statistics and relational databases will be helpful but is not mandatory, to get the most out of this book. A firm understanding of Java is required. What You Will Learn Develop Java programs that analyze data sets of nearly any size, including text Implement important machine learning algorithms such as regression, classification, and clustering Interface with and apply standard open source Java libraries and APIs to analyze and visualize data Process data from both relational and non-relational databases and from time-series data Employ Java tools to visualize data in various forms Understand multimedia data analysis algorithms and implement them in Java. In Detail Data analysis is a process of inspecting, cleansing, transforming, and modeling data with the aim of discovering useful information. Java is one of the most popular languages to perform your data analysis tasks. This book will help you learn the tools and techniques in Java to conduct data analysis without any hassle. After getting a quick overview of what data science is and the steps involved in the process, you'll learn the statistical data analysis techniques and implement them using the popular Java APIs and libraries. Through practical examples, you will also learn the machine learning concepts such as classification and regression. In the process, you'll familiarize yourself with tools such as Rapidminer and WEKA and see how these Java-based tools can be used effectively for analysis. You will also learn how to analyze text and other types of multimedia. Learn to work with relational, NoSQL, and time-series data. This book will also show you how you can utilize different Java-based libraries to

create insightful and easy to understand plots and graphs. By the end of this book, you will have a solid understanding of the various data analysis techniques, and how to implement them using Java. Style and approach The book takes a very comprehensive approach to enhance your understanding of data analysis. Sufficient real-world examples and use cases are included to help you grasp the concepts quickly and apply them easily in your day-to-day work. Packed with clear, easy-to-follow examples, this book will turn you into an ace data analyst in no time.

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 to 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

Scientific Data Analysis using Jython Scripting and Java presents practical approaches for data analysis using Java scripting based on Jython, a Java implementation of the Python language. The chapters essentially cover all aspects of data analysis, from arrays and histograms to clustering analysis, curve fitting, metadata and neural networks. A comprehensive coverage of data visualisation tools implemented in Java is also included. Written by the primary developer of the jHepWork data-analysis framework, the book provides a reliable and complete reference source laying the foundation for data-analysis applications using Java scripting. More than 250 code snippets (of around 10-20 lines each) written in Jython and Java, plus several real-life examples help the reader develop a genuine feeling for data analysis techniques and their programming implementation. This is the first data-analysis and data-mining book which is completely based on the Jython language, and opens doors to scripting using a fully multi-platform and multi-threaded approach. Graduate students and researchers will benefit from the information presented in this book.

This book provides a complete and modern guide to web scraping, using Python as the programming language, without glossing over important details or best practices. Written with a data science audience in mind, the book explores both

scraping and the larger context of web technologies in which it operates, to ensure full understanding. The authors recommend web scraping as a powerful tool for any data scientist's arsenal, as many data science projects start by obtaining an appropriate data set. Starting with a brief overview on scraping and real-life use cases, the authors explore the core concepts of HTTP, HTML, and CSS to provide a solid foundation. Along with a quick Python primer, they cover Selenium for JavaScript-heavy sites, and web crawling in detail. The book finishes with a recap of best practices and a collection of examples that bring together everything you've learned and illustrate various data science use cases. What You'll Learn Leverage well-established best practices and commonly-used Python packages Handle today's web, including JavaScript, cookies, and common web scraping mitigation techniques Understand the managerial and legal concerns regarding web scraping Who This Book is For A data science oriented audience that is probably already familiar with Python or another programming language or analytical toolkit (R, SAS, SPSS, etc). Students or instructors in university courses may also benefit. Readers unfamiliar with Python will appreciate a quick Python primer in chapter 1 to catch up with the basics and provide pointers to other guides as well.

Learn the basics of analytics on big data using Java, machine learning and other big data tools About This Book Acquire real-world set of tools for building enterprise level data science applications Surpasses the barrier of other languages in data science and learn create useful object-oriented codes Extensive use of Java compliant big data tools like apache spark, Hadoop, etc. Who This Book Is For This book is for Java developers who are looking to perform data analysis in production environment. Those who wish to implement data analysis in their Big data applications will find this book helpful. What You Will Learn Start from simple analytic tasks on big data Get into more complex tasks with predictive analytics on big data using machine learning Learn real time analytic tasks Understand the concepts with examples and case studies Prepare and refine data for analysis Create charts in order to understand the data See various real-world datasets In Detail This book covers case studies such as sentiment analysis on a tweet dataset, recommendations on a movielens dataset, customer segmentation on an ecommerce dataset, and graph analysis on actual flights dataset. This book is an end-to-end guide to implement analytics on big data with Java. Java is the de facto language for major big data environments, including Hadoop. This book will teach you how to perform analytics on big data with production-friendly Java. This book basically divided into two sections. The first part is an introduction that will help the readers get acquainted with big data environments, whereas the second part will contain a hardcore discussion on all the concepts in analytics on big data. It will take you from data analysis and data visualization to the core concepts and advantages of machine learning, real-life usage of regression and classification using Naive Bayes, a deep discussion on the concepts of clustering, and a review of simple neural networks on big data using deepLearning4j or plain Java Spark code. This

book is a must-have book for Java developers who want to start learning big data analytics and want to use it in the real world. Style and approach The approach of book is to deliver practical learning modules in manageable content. Each chapter is a self-contained unit of a concept in big data analytics. Book will step by step builds the competency in the area of big data analytics. Examples using real world case studies to give ideas of real applications and how to use the techniques mentioned. The examples and case studies will be shown using both theory and code.

This book offers a thorough grounding in machine learning concepts combined with practical advice on applying machine learning tools and techniques in real-world data mining situations. Clearly written and effectively illustrated, this book is ideal for anyone involved at any level in the work of extracting usable knowledge from large collections of data.

Complementing the book's instruction is fully functional machine learning software.

Data Science is booming thanks to R and Python, but Java brings the robustness, convenience, and ability to scale critical to today's data science applications. With this practical book, Java software engineers looking to add data science skills will take a logical journey through the data science pipeline. Author Michael Brzustowicz explains the basic math theory behind each step of the data science process, as well as how to apply these concepts with Java. You'll learn the critical roles that data IO, linear algebra, statistics, data operations, learning and prediction, and Hadoop MapReduce play in the process. Throughout this book, you'll find code examples you can use in your applications. Examine methods for obtaining, cleaning, and arranging data into its purest form Understand the matrix structure that your data should take Learn basic concepts for testing the origin and validity of data Transform your data into stable and usable numerical values Understand supervised and unsupervised learning algorithms, and methods for evaluating their success Get up and running with MapReduce, using customized components suitable for data science algorithms

Implement numerical algorithms in Java using NM Dev, an object-oriented and high-performance programming library for mathematics. You'll see how it can help you easily create a solution for your complex engineering problem by quickly putting together classes. Numerical Methods Using Java covers a wide range of topics, including chapters on linear algebra, root finding, curve fitting, differentiation and integration, solving differential equations, random numbers and simulation, a whole suite of unconstrained and constrained optimization algorithms, statistics, regression and time series analysis. The mathematical concepts behind the algorithms are clearly explained, with plenty of code examples and illustrations to help even beginners get started. What You Will Learn Program in Java using a high-performance numerical library Learn the mathematics for a wide range of numerical computing algorithms Convert ideas and equations into code Put together algorithms and classes to build your own engineering solution Build solvers for industrial optimization problems Do data analysis using basic and advanced statistics Who This Book Is For Programmers, data

scientists, and analysts with prior experience with programming in any language, especially Java.

Build and run intelligent applications by leveraging key Java machine learning libraries About This Book Develop a sound strategy to solve predictive modelling problems using the most popular machine learning Java libraries. Explore a broad variety of data processing, machine learning, and natural language processing through diagrams, source code, and real-world applications This step-by-step guide will help you solve real-world problems and links neural network theory to their application Who This Book Is For This course is intended for data scientists and Java developers who want to dive into the exciting world of deep learning. It will get you up and running quickly and provide you with the skills you need to successfully create, customize, and deploy machine learning applications in real life. What You Will Learn Get a practical deep dive into machine learning and deep learning algorithms Explore neural networks using some of the most popular Deep Learning frameworks Dive into Deep Belief Nets and Stacked Denoising Autoencoders algorithms Apply machine learning to fraud, anomaly, and outlier detection Experiment with deep learning concepts, algorithms, and the toolbox for deep learning Select and split data sets into training, test, and validation, and explore validation strategies Apply the code generated in practical examples, including weather forecasting and pattern recognition In Detail Machine learning applications are everywhere, from self-driving cars, spam detection, document search, and trading strategies, to speech recognition Starting with an introduction to basic machine learning algorithms, this course takes you further into this vital world of stunning predictive insights and remarkable machine intelligence. This course helps you solve challenging problems in image processing, speech recognition, language modeling. You will discover how to detect anomalies and fraud, and ways to perform activity recognition, image recognition, and text. You will also work with examples such as weather forecasting, disease diagnosis, customer profiling, generalization, extreme machine learning and more. By the end of this course, you will have all the knowledge you need to perform deep learning on your system with varying complexity levels, to apply them to your daily work. The course provides you with highly practical content explaining deep learning with Java, from the following Packt books: Java Deep Learning Essentials Machine Learning in Java Neural Network Programming with Java, Second Edition Style and approach This course aims to create a smooth learning path that will teach you how to effectively use deep learning with Java with other de facto components to get the most out of it. Through this comprehensive course, you'll learn the basics of predictive modelling and progress to solve real-world problems and links neural network theory to their application

Data Mining: Practical Machine Learning Tools and Techniques, Third Edition, offers a thorough grounding in machine learning concepts as well as practical advice on applying machine learning tools and techniques in real-world data mining situations. This highly anticipated third edition of the most acclaimed work on data mining and machine learning will teach

you everything you need to know about preparing inputs, interpreting outputs, evaluating results, and the algorithmic methods at the heart of successful data mining. Thorough updates reflect the technical changes and modernizations that have taken place in the field since the last edition, including new material on Data Transformations, Ensemble Learning, Massive Data Sets, Multi-instance Learning, plus a new version of the popular Weka machine learning software developed by the authors. Witten, Frank, and Hall include both tried-and-true techniques of today as well as methods at the leading edge of contemporary research. The book is targeted at information systems practitioners, programmers, consultants, developers, information technology managers, specification writers, data analysts, data modelers, database R&D professionals, data warehouse engineers, data mining professionals. The book will also be useful for professors and students of upper-level undergraduate and graduate-level data mining and machine learning courses who want to incorporate data mining as part of their data management knowledge base and expertise. Provides a thorough grounding in machine learning concepts as well as practical advice on applying the tools and techniques to your data mining projects Offers concrete tips and techniques for performance improvement that work by transforming the input or output in machine learning methods Includes downloadable Weka software toolkit, a collection of machine learning algorithms for data mining tasks—in an updated, interactive interface. Algorithms in toolkit cover: data pre-processing, classification, regression, clustering, association rules, visualization

Machine learning has finally come of age. With H2O software, you can perform machine learning and data analysis using a simple open source framework that's easy to use, has a wide range of OS and language support, and scales for big data. This hands-on guide teaches you how to use H2O with only minimal math and theory behind the learning algorithms. If you're familiar with R or Python, know a bit of statistics, and have some experience manipulating data, author Darren Cook will take you through H2O basics and help you conduct machine-learning experiments on different sample data sets. You'll explore several modern machine-learning techniques such as deep learning, random forests, unsupervised learning, and ensemble learning. Learn how to import, manipulate, and export data with H2O Explore key machine-learning concepts, such as cross-validation and validation data sets Work with three diverse data sets, including a regression, a multinomial classification, and a binomial classification Use H2O to analyze each sample data set with four supervised machine-learning algorithms Understand how cluster analysis and other unsupervised machine-learning algorithms work

Become an advanced practitioner with this progressive set of master classes on application-oriented machine learning About This Book Comprehensive coverage of key topics in machine learning with an emphasis on both the theoretical and practical aspects More than 15 open source Java tools in a wide range of techniques, with code and practical usage.

More than 10 real-world case studies in machine learning highlighting techniques ranging from data ingestion up to analyzing the results of experiments, all preparing the user for the practical, real-world use of tools and data analysis.

Who This Book Is For This book will appeal to anyone with a serious interest in topics in Data Science or those already working in related areas: ideally, intermediate-level data analysts and data scientists with experience in Java. Preferably, you will have experience with the fundamentals of machine learning and now have a desire to explore the area further, are up to grappling with the mathematical complexities of its algorithms, and you wish to learn the complete ins and outs of practical machine learning.

What You Will Learn Master key Java machine learning libraries, and what kind of problem each can solve, with theory and practical guidance. Explore powerful techniques in each major category of machine learning such as classification, clustering, anomaly detection, graph modeling, and text mining. Apply machine learning to real-world data with methodologies, processes, applications, and analysis. Techniques and experiments developed around the latest specializations in machine learning, such as deep learning, stream data mining, and active and semi-supervised learning. Build high-performing, real-time, adaptive predictive models for batch- and stream-based big data learning using the latest tools and methodologies. Get a deeper understanding of technologies leading towards a more powerful AI applicable in various domains such as Security, Financial Crime, Internet of Things, social networking, and so on.

In Detail Java is one of the main languages used by practicing data scientists; much of the Hadoop ecosystem is Java-based, and it is certainly the language that most production systems in Data Science are written in. If you know Java, Mastering Machine Learning with Java is your next step on the path to becoming an advanced practitioner in Data Science. This book aims to introduce you to an array of advanced techniques in machine learning, including classification, clustering, anomaly detection, stream learning, active learning, semi-supervised learning, probabilistic graph modeling, text mining, deep learning, and big data batch and stream machine learning. Accompanying each chapter are illustrative examples and real-world case studies that show how to apply the newly learned techniques using sound methodologies and the best Java-based tools available today. On completing this book, you will have an understanding of the tools and techniques for building powerful machine learning models to solve data science problems in just about any domain.

Style and approach A practical guide to help you explore machine learning—and an array of Java-based tools and frameworks—with the help of practical examples and real-world use cases.

Build machine learning (ML) solutions for Java development. This book shows you that when designing ML apps, data is the key driver and must be considered throughout all phases of the project life cycle. Practical Java Machine Learning helps you understand the importance of data and how to organize it for use within your ML project. You will be introduced to tools which can help you identify and manage your data including JSON, visualization, NoSQL databases, and cloud

platforms including Google Cloud Platform and Amazon Web Services. Practical Java Machine Learning includes multiple projects, with particular focus on the Android mobile platform and features such as sensors, camera, and connectivity, each of which produce data that can power unique machine learning solutions. You will learn to build a variety of applications that demonstrate the capabilities of the Google Cloud Platform machine learning API, including data visualization for Java; document classification using the Weka ML environment; audio file classification for Android using ML with spectrogram voice data; and machine learning using device sensor data. After reading this book, you will come away with case study examples and projects that you can take away as templates for re-use and exploration for your own machine learning programming projects with Java. What You Will Learn Identify, organize, and architect the data required for ML projects Deploy ML solutions in conjunction with cloud providers such as Google and Amazon Determine which algorithm is the most appropriate for a specific ML problem Implement Java ML solutions on Android mobile devices Create Java ML solutions to work with sensor data Build Java streaming based solutions Who This Book Is For Experienced Java developers who have not implemented machine learning techniques before.

Data collection, processing, analysis, and more About This Book Your entry ticket to the world of data science with the stability and power of Java Explore, analyse, and visualize your data effectively using easy-to-follow examples A highly practical course covering a broad set of topics - from the basics of Machine Learning to Deep Learning and Big Data frameworks. Who This Book Is For This course is meant for Java developers who are comfortable developing applications in Java, and now want to enter the world of data science or wish to build intelligent applications. Aspiring data scientists with some understanding of the Java programming language will also find this book to be very helpful. If you are willing to build efficient data science applications and bring them in the enterprise environment without changing your existing Java stack, this book is for you! What You Will Learn Understand the key concepts of data science Explore the data science ecosystem available in Java Work with the Java APIs and techniques used to perform efficient data analysis Find out how to approach different machine learning problems with Java Process unstructured information such as natural language text or images, and create your own search Learn how to build deep neural networks with DeepLearning4j Build data science applications that scale and process large amounts of data Deploy data science models to production and evaluate their performance In Detail Data science is concerned with extracting knowledge and insights from a wide variety of data sources to analyse patterns or predict future behaviour. It draws from a wide array of disciplines including statistics, computer science, mathematics, machine learning, and data mining. In this course, we cover the basic as well as advanced data science concepts and how they are implemented using the popular Java tools and libraries. The course starts with an introduction of data science, followed by the basic data science tasks of data

collection, data cleaning, data analysis, and data visualization. This is followed by a discussion of statistical techniques and more advanced topics including machine learning, neural networks, and deep learning. You will examine the major categories of data analysis including text, visual, and audio data, followed by a discussion of resources that support parallel implementation. Throughout this course, the chapters will illustrate a challenging data science problem, and then go on to present a comprehensive, Java-based solution to tackle that problem. You will cover a wide range of topics – from classification and regression, to dimensionality reduction and clustering, deep learning and working with Big Data. Finally, you will see the different ways to deploy the model and evaluate it in production settings. By the end of this course, you will be up and running with various facets of data science using Java, in no time at all. This course contains premium content from two of our recently published popular titles: *Java for Data Science* and *Mastering Java for Data Science*. This course follows a tutorial approach, providing examples of each of the concepts covered. With a step-by-step instructional style, this book covers various facets of data science and will get you up and running quickly.

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

The data lake is a daring new approach for harnessing the power of big data technology and providing convenient self-service capabilities. But is it right for your company? This book is based on discussions with practitioners and executives from more than a hundred organizations, ranging from data-driven companies such as Google, LinkedIn, and Facebook, to governments and traditional corporate enterprises. You'll learn what a data lake is, why enterprises need one, and how to build one successfully with the best practices in this book. Alex Gorelik, CTO and founder of Waterline Data, explains why old systems and processes can no longer support data needs in the enterprise. Then, in a collection of

essays about data lake implementation, you'll examine data lake initiatives, analytic projects, experiences, and best practices from data experts working in various industries. Get a succinct introduction to data warehousing, big data, and data science Learn various paths enterprises take to build a data lake Explore how to build a self-service model and best practices for providing analysts access to the data Use different methods for architecting your data lake Discover ways to implement a data lake from experts in different industries

A good data scientist knows how to do something really well, but a great data scientist can do "something of everything." From raw data all the way to shining in front of C-level executives, a great data scientist has the skills to architect data systems, build applications, perform modeling and machine learning and wrap up the results in a clear (and quickly iterable) manner. From data models to ETL to databases to distributed algorithms and learning, this book has you covered. While many resources for Java (and data science) exist, none of them combine the two, and especially not at a level where sophisticated concepts are demonstrated clearly and in simplest terms. Data Science with Java marries the two in a practical way. Learn an extremely practical set of tools for creating enterprise grade data science applications Get past the intimidating barrier to machine learning and statistics—and learn how useful object-oriented code can be Are you thinking about learning how to use the Python programming language? Thinking about getting started on a programming career? Are you thinking about learning data science? This book is for you!

The Complete Guide to Data Science with Hadoop—For Technical Professionals, Businesspeople, and Students Demand is soaring for professionals who can solve real data science problems with Hadoop and Spark. Practical Data Science with Hadoop® and Spark is your complete guide to doing just that. Drawing on immense experience with Hadoop and big data, three leading experts bring together everything you need: high-level concepts, deep-dive techniques, real-world use cases, practical applications, and hands-on tutorials. The authors introduce the essentials of data science and the modern Hadoop ecosystem, explaining how Hadoop and Spark have evolved into an effective platform for solving data science problems at scale. In addition to comprehensive application coverage, the authors also provide useful guidance on the important steps of data ingestion, data munging, and visualization. Once the groundwork is in place, the authors focus on specific applications, including machine learning, predictive modeling for sentiment analysis, clustering for document analysis, anomaly detection, and natural language processing (NLP). This guide provides a strong technical foundation for those who want to do practical data science, and also presents business-driven guidance on how to apply Hadoop and Spark to optimize ROI of data science initiatives. Learn What data science is, how it has evolved, and how to plan a data science career How data volume, variety, and velocity shape data science use cases Hadoop and its ecosystem, including HDFS, MapReduce, YARN, and Spark Data importation with Hive and Spark Data quality, preprocessing, preparation, and modeling Visualization: surfacing insights from huge data sets Machine learning: classification, regression, clustering, and anomaly detection Algorithms and Hadoop tools for predictive modeling Cluster

analysis and similarity functions Large-scale anomaly detection NLP: applying data science to human language

Learn how to process and analysis data using Python KEY FEATURES - The book has theories explained elaborately along with Python code and corresponding output to support the theoretical explanations. The Python codes are provided with step-by-step comments to explain each instruction of the code. - The book is not just dealing with the background mathematics alone or only the programs but beautifully correlates the background mathematics to the theory and then finally translating it into the programs. - A rich set of chapter-end exercises are provided, consisting of both short-answer questions and long-answer questions.

DESCRIPTION This book introduces the fundamental concepts of Data Science, which has proved to be a major game-changer in business solving problems. Topics covered in the book include fundamentals of Data Science, data preprocessing, data plotting and visualization, statistical data analysis, machine learning for data analysis, time-series analysis, deep learning for Data Science, social media analytics, business analytics, and Big Data analytics. The content of the book describes the fundamentals of each of the Data Science related topics together with illustrative examples as to how various data analysis techniques can be implemented using different tools and libraries of Python programming language. Each chapter contains numerous examples and illustrative output to explain the important basic concepts. An appropriate number of questions is presented at the end of each chapter for self-assessing the conceptual understanding. The references presented at the end of every chapter will help the readers to explore more on a given topic.

WHAT WILL YOU LEARN Perform processing on data for making it ready for visual plot and understand the pattern in data over time. Understand what machine learning is and how learning can be incorporated into a program. Know how tools can be used to perform analysis on big data using python and other standard tools. Perform social media analytics, business analytics, and data analytics on any data of a company or organization.

WHO THIS BOOK IS FOR The book is for readers with basic programming and mathematical skills. The book is for any engineering graduates that wish to apply data science in their projects or wish to build a career in this direction. The book can be read by anyone who has an interest in data analysis and would like to explore more out of interest or to apply it to certain real-life problems.

TABLE OF CONTENTS 1. Fundamentals of Data Science 2. Data Preprocessing 3. Data Plotting and Visualization 4. Statistical Data Analysis 5. Machine Learning for Data Science 6. Time-Series Analysis 7. Deep Learning for Data Science 8. Social Media Analytics 9. Business Analytics 10. Big Data Analytics

JavaScript is the native language of the Internet. Originally created to make web pages more dynamic, it is now used for software projects of all kinds, including scientific visualization and data services. However, most data scientists have little or no experience with JavaScript, and most introductions to the language are written for people who want to build shopping carts rather than share maps of coral reefs. This book will introduce you to JavaScript's power and idiosyncrasies and guide you through the key features of the language and its tools and libraries. The book places equal focus on client- and server-side programming, and shows readers how to create interactive web content, build and test data services, and visualize data in the browser. Topics include: The core features of modern JavaScript Creating templated web pages Making those pages interactive using React Data visualization

using Vega-Lite Using Data-Forge to wrangle tabular data Building a data service with Express Unit testing with Mocha All of the material is covered by the Creative Commons Attribution-Noncommercial 4.0 International license (CC-BY-NC-4.0) and is included in the book's companion website at <http://js4ds.org> . Maya Gans is a freelance data scientist and front-end developer by way of quantitative biology. Toby Hodges is a bioinformatician turned community coordinator who works at the European Molecular Biology Laboratory. Greg Wilson co-founded Software Carpentry, and is now part of the education team at RStudio

Explore big data concepts, platforms, analytics, and their applications using the power of Hadoop 3 Key Features Learn Hadoop 3 to build effective big data analytics solutions on-premise and on cloud Integrate Hadoop with other big data tools such as R, Python, Apache Spark, and Apache Flink Exploit big data using Hadoop 3 with real-world examples Book Description Apache Hadoop is the most popular platform for big data processing, and can be combined with a host of other big data tools to build powerful analytics solutions. Big Data Analytics with Hadoop 3 shows you how to do just that, by providing insights into the software as well as its benefits with the help of practical examples. Once you have taken a tour of Hadoop 3's latest features, you will get an overview of HDFS, MapReduce, and YARN, and how they enable faster, more efficient big data processing. You will then move on to learning how to integrate Hadoop with the open source tools, such as Python and R, to analyze and visualize data and perform statistical computing on big data. As you get acquainted with all this, you will explore how to use Hadoop 3 with Apache Spark and Apache Flink for real-time data analytics and stream processing. In addition to this, you will understand how to use Hadoop to build analytics solutions on the cloud and an end-to-end pipeline to perform big data analysis using practical use cases. By the end of this book, you will be well-versed with the analytical capabilities of the Hadoop ecosystem. You will be able to build powerful solutions to perform big data analytics and get insight effortlessly. What you will learn Explore the new features of Hadoop 3 along with HDFS, YARN, and MapReduce Get well-versed with the analytical capabilities of Hadoop ecosystem using practical examples Integrate Hadoop with R and Python for more efficient big data processing Learn to use Hadoop with Apache Spark and Apache Flink for real-time data analytics Set up a Hadoop cluster on AWS cloud Perform big data analytics on AWS using Elastic Map Reduce Who this book is for Big Data Analytics with Hadoop 3 is for you if you are looking to build high-performance analytics solutions for your enterprise or business using Hadoop 3's powerful features, or you're new to big data analytics. A basic understanding of the Java programming language is required.

This book approaches big data, artificial intelligence, machine learning, and business intelligence through the lens of Data Science. We have grown accustomed to seeing these terms mentioned time and time again in the mainstream media. However, our understanding of what they actually mean often remains limited. This book provides a general overview of the terms and approaches used broadly in data science, and provides detailed information on the underlying theories, models, and application scenarios. Divided into three main parts, it addresses what data science is; how and where it is used; and how it can be implemented using modern open source software. The book offers an essential guide to modern data science for all students, practitioners, developers and managers seeking a deeper understanding of how various aspects of data science work, and of how

they can be employed to gain a competitive advantage.

Your logical, linear guide to the fundamentals of data science programming Data science is exploding—in a good way—with a forecast of 1.7 megabytes of new information created every second for each human being on the planet by 2020 and 11.5 million job openings by 2026. It clearly pays dividends to be in the know. This friendly guide charts a path through the fundamentals of data science and then delves into the actual work: linear regression, logical regression, machine learning, neural networks, recommender engines, and cross-validation of models. Data Science Programming All-In-One For Dummies is a compilation of the key data science, machine learning, and deep learning programming languages: Python and R. It helps you decide which programming languages are best for specific data science needs. It also gives you the guidelines to build your own projects to solve problems in real time. Get grounded: the ideal start for new data professionals What lies ahead: learn about specific areas that data is transforming Be meaningful: find out how to tell your data story See clearly: pick up the art of visualization Whether you're a beginning student or already mid-career, get your copy now and add even more meaning to your life—and everyone else's!

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

With this practical book, AI and machine learning practitioners will learn how to successfully build and deploy data science projects on Amazon Web Services. The Amazon AI and machine learning stack unifies data science, data engineering, and application development to help level up your skills. This guide shows you how to build and run pipelines in the cloud, then integrate the results into applications in minutes instead of days. Throughout the book, authors Chris Fregly and Antje Barth demonstrate how to reduce cost and improve performance. Apply the Amazon AI and ML stack to real-world use cases for natural language processing, computer vision, fraud detection, conversational devices, and more Use automated machine learning to implement a specific subset of use cases with SageMaker Autopilot Dive deep into the complete model development lifecycle for a BERT-based NLP use case including data ingestion, analysis, model training, and deployment Tie everything together into a repeatable

machine learning operations pipeline Explore real-time ML, anomaly detection, and streaming analytics on data streams with Amazon Kinesis and Managed Streaming for Apache Kafka Learn security best practices for data science projects and workflows including identity and access management, authentication, authorization, and more

In this book, you will implement two data science projects using Scikit-Learn, Scipy, and other libraries with Python GUI. In Project 1, you will learn how to use Scikit-Learn, NumPy, Pandas, Seaborn, and other libraries to perform how to predict early stage diabetes using Early Stage Diabetes Risk Prediction Dataset provided by Kaggle (<https://www.kaggle.com/ishandutta/early-stage-diabetes-risk-prediction-dataset/download>). This dataset contains the sign and symptom data of newly diabetic or would be diabetic patient. This has been collected using direct questionnaires from the patients of Sylhet Diabetes Hospital in Sylhet, Bangladesh and approved by a doctor. You will develop a GUI using PyQt5 to plot distribution of features, feature importance, cross validation score, and prediced values versus true values. The machine learning models used in this project are Adaboost, Random Forest, Gradient Boosting, Logistic Regression, and Support Vector Machine. In Project 2, you will learn how to use Scikit-Learn, NumPy, Pandas, and other libraries to perform how to analyze and predict breast cancer using Breast Cancer Prediction Dataset provided by Kaggle (<https://www.kaggle.com/merishnasuwal/breast-cancer-prediction-dataset/download>). Worldwide, breast cancer is the most common type of cancer in women and the second highest in terms of mortality rates. Diagnosis of breast cancer is performed when an abnormal lump is found (from self-examination or x-ray) or a tiny speck of calcium is seen (on an x-ray). After a suspicious lump is found, the doctor will conduct a diagnosis to determine whether it is cancerous and, if so, whether it has spread to other parts of the body. This breast cancer dataset was obtained from the University of Wisconsin Hospitals, Madison from Dr. William H. Wolberg. You will develop a GUI using PyQt5 to plot distribution of features, pairwise relationship, test scores, prediced values versus true values, confusion matrix, and decision boundary. The machine learning models used in this project are K-Nearest Neighbor, Random Forest, Naive Bayes, Logistic Regression, Decision Tree, and Support Vector Machine. 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

If you work in data science, or are hoping to make it your career, Python is a must! It's no secret that many jobs will be eliminated

in the near future. While that reality may be difficult to digest, it's also believed that the most responsible party -- namely automation -- will lead the way in a golden age of opportunity. That's because there will be a tremendous need for people who are capable of programming all those new systems. And, the language of choice for machine learning software is Python which, as it happens, is the perfect first language to learn. Python has now toppled Java to become the most taught programming languages in Universities and Academia! Learning Python not only will make you a polyglot programmer but also gives you a powerful tool in your arsenal to write scripts, create a web application, and open the door to the exciting fields of data science and machine learning. In this book, you will learn the fundamentals of Python, and the benefits are: Python works on different platforms (Windows, Mac, Linux, Raspberry Pi, etc.). Python has a simple syntax similar to the English language. Python has syntax that allows developers to write programs with fewer lines than some other programming languages. Python runs on an interpreter system, meaning that code can be executed as soon as it is written. This means that prototyping can be very quick. Python can be treated in a procedural way, an object-orientated way or a functional way. It's important to know programming and coding in today's world and if you don't know coding you are missing something and Python is a great way to start learning to code. WHAT ARE YOU WAITING FOR? A FEW DOLLARS SPENT ARE THE VALUE OF YOUR SKILL IMPROVEMENT? It's definitely possible to learn Python on your own. To learn it steps by steps by pressing the BUY NOW BUTTOM!

Here's what you get in this book: - 300 practice questions and answers spanning the breadth of topics under the data science umbrella - Covers statistics, machine learning, SQL, NoSQL, Hadoop and bioinformatics - Emphasis on real-world application with a chapter on Python libraries for machine learning - Focus on the most frequently asked interview questions. Avoid information overload - Compact format: easy to read, easy to carry, so you can study on-the-go Now, you finally have what you need to crush your data science interview, and land that dream job. About The Author Zack Austin has been building large scale enterprise systems for clients in the media, telecom, financial services and publishing since 2001. He is based in New York City.

Use Java to create a diverse range of Data Science applications and bring Data Science into production About This Book An overview of modern Data Science and Machine Learning libraries available in Java Coverage of a broad set of topics, going from the basics of Machine Learning to Deep Learning and Big Data frameworks. Easy-to-follow illustrations and the running example of building a search engine. Who This Book Is For This book is intended for software engineers who are comfortable with developing Java applications and are familiar with the basic concepts of data science. Additionally, it will also be useful for data scientists who do not yet know Java but want or need to learn it. If you are willing to build efficient data science applications and bring them in the enterprise environment without changing the existing stack, this book is for you! What You Will Learn Get a solid understanding of the data processing toolbox available in Java Explore the data science ecosystem available in Java Find out how to approach different machine learning problems with Java Process unstructured information such as natural language text or images Create your own search engine Get state-of-the-art performance with XGBoost Learn how to build deep neural networks with DeepLearning4j Build applications that scale and process large amounts of data Deploy data science models to production

and evaluate their performance In Detail Java is the most popular programming language, according to the TIOBE index, and it is a typical choice for running production systems in many companies, both in the startup world and among large enterprises. Not surprisingly, it is also a common choice for creating data science applications: it is fast and has a great set of data processing tools, both built-in and external. What is more, choosing Java for data science allows you to easily integrate solutions with existing software, and bring data science into production with less effort. This book will teach you how to create data science applications with Java. First, we will revise the most important things when starting a data science application, and then brush up the basics of Java and machine learning before diving into more advanced topics. We start by going over the existing libraries for data processing and libraries with machine learning algorithms. After that, we cover topics such as classification and regression, dimensionality reduction and clustering, information retrieval and natural language processing, and deep learning and big data. Finally, we finish the book by talking about the ways to deploy the model and evaluate it in production settings. Style and approach This is a practical guide where all the important concepts such as classification, regression, and dimensionality reduction are explained with the help of examples.

Although traditional texts present isolated algorithms and data structures, they do not provide a unifying structure and offer little guidance on how to appropriately select among them. Furthermore, these texts furnish little, if any, source code and leave many of the more difficult aspects of the implementation as exercises. A fresh alternative to

[Copyright: 845cc661ad04f01a3c1d25b8b8032a9d](#)