
Apache Spark Machine Learning Blueprints

Big Data Concepts, Theories, and Applications
 Learning Apache Spark 2
 Big Data 2.0 Processing Systems
 IBM Data Engine for Hadoop and Spark
 Apache Spark 2.x Machine Learning Cookbook
 Mastering Machine Learning on AWS
 High Performance Spark
 The Human Element of Big Data
 Apache Spark Deep Learning Cookbook
 Learning Spark SQL
 Learning Spark
 Stream Processing with Apache Spark
 Scalable Uncertainty Management
 Beginning Apache Spark 2
 Automated Machine Learning for Business
 Big Data Processing with Apache Spark
 Natural Language Processing with Spark NLP
 Database Systems for Advanced Applications
 Hands-On Big Data Analytics with PySpark
 Learning from Imbalanced Data Sets
 Theory and Applications of Dependable Computer Systems
 Research Anthology on Machine Learning Techniques, Methods, and Applications
 Spark: The Definitive Guide
 Apache Spark Machine Learning Blueprints
 Disruptive Analytics
 Apache Spark for the Enterprise: Setting the Business Free
 Blueprints for Text Analytics Using Python
 Transforming Healthcare with Big Data and AI
 Hadoop Blueprints
 Information Systems
 Unreal Engine 4 Virtual Reality Projects
 Hands-On Data Science and Python Machine Learning
 Smart Healthcare for Sustainable Urban Development
 Next-Generation Machine Learning with Spark
 Machine Learning with Apache Spark Quick Start Guide
 The AI Marketing Canvas
 Hands-On Deep Learning with Apache Spark
 Apache Spark 2: Data Processing and Real-Time Analytics
 Mastering Apache Spark 2.x

Apache Spark Machine Learning Blueprints

Downloaded from archive.imba.com by guest

JACOBY ISAIAH

Big Data Concepts, Theories, and Applications Springer Nature

This IBM® Redbooks® publication provides topics to help the technical community take advantage of the resilience, scalability, and performance of the IBM Power Systems™ platform to implement or integrate an IBM Data Engine for Hadoop and Spark solution for analytics solutions to access, manage, and analyze data sets to improve business outcomes. This book documents topics to demonstrate and take advantage of the analytics strengths of the IBM POWER8® platform, the IBM analytics software portfolio, and selected third-party tools to help solve customer's data analytic workload requirements. This book describes how to plan, prepare, install, integrate, manage, and show how to use the IBM Data Engine for Hadoop and Spark solution to run analytic workloads on IBM POWER8. In addition, this publication delivers documentation to complement available IBM analytics solutions to help your data analytic needs. This publication strengthens the position of IBM analytics and big data

solutions with a well-defined and documented deployment model within an IBM POWER8 virtualized environment so that customers have a planned foundation for security, scaling, capacity, resilience, and optimization for analytics workloads. This book is targeted at technical professionals (analytics consultants, technical support staff, IT Architects, and IT Specialists) that are responsible for delivering analytics solutions and support on IBM Power Systems.

[Learning Apache Spark 2](#) IBM Redbooks

Combine advanced analytics including Machine Learning, Deep Learning Neural Networks and Natural Language Processing with modern scalable technologies including Apache Spark to derive actionable insights from Big Data in real-time Key Features Make a hands-on start in the fields of Big Data, Distributed Technologies and Machine Learning Learn how to design, develop and interpret the results of common Machine Learning algorithms Uncover hidden patterns in your data in order to derive real actionable insights and business value Book Description Every person and every organization in the world manages data, whether they realize it or not. Data is used to describe the world around us and can be used for almost any purpose, from analyzing consumer habits to fighting disease and serious

organized crime. Ultimately, we manage data in order to derive value from it, and many organizations around the world have traditionally invested in technology to help process their data faster and more efficiently. But we now live in an interconnected world driven by mass data creation and consumption where data is no longer rows and columns restricted to a spreadsheet, but an organic and evolving asset in its own right. With this realization comes major challenges for organizations: how do we manage the sheer size of data being created every second (think not only spreadsheets and databases, but also social media posts, images, videos, music, blogs and so on)? And once we can manage all of this data, how do we derive real value from it? The focus of *Machine Learning with Apache Spark* is to help us answer these questions in a hands-on manner. We introduce the latest scalable technologies to help us manage and process big data. We then introduce advanced analytical algorithms applied to real-world use cases in order to uncover patterns, derive actionable insights, and learn from this big data. What you will learn

Understand how Spark fits in the context of the big data ecosystem
Understand how to deploy and configure a local development environment using Apache Spark
Understand how to design supervised and unsupervised learning models
Build models to perform NLP, deep learning, and cognitive services using Spark ML libraries
Design real-time machine learning pipelines in Apache Spark
Become familiar with advanced techniques for processing a large volume of data by applying machine learning algorithms

Who this book is for This book is aimed at Business Analysts, Data Analysts and Data Scientists who wish to make a hands-on start in order to take advantage of modern Big Data technologies combined with Advanced Analytics.

Big Data 2.0 Processing Systems Packt Publishing Ltd

Speed up the design and implementation of deep learning solutions using Apache Spark

Key Features
Explore the world of distributed deep learning with Apache Spark
Train neural networks with deep learning libraries such as BigDL and TensorFlow
Develop Spark deep learning applications to intelligently handle large and complex datasets

Book Description
Deep learning is a subset of machine learning where datasets with several layers of complexity can be processed. *Hands-On Deep Learning with Apache Spark* addresses the sheer complexity of technical and analytical parts and the speed at which deep learning solutions can be implemented on Apache Spark. The book starts with the fundamentals of Apache Spark and deep learning. You will set up Spark for deep learning, learn principles of distributed modeling, and understand different types of neural nets. You will then implement deep learning models, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and long short-term memory (LSTM) on Spark. As you progress through the book, you will gain hands-on experience of what it takes to understand the complex datasets you are dealing with. During the course of this book, you will use popular deep learning frameworks, such as TensorFlow, Deeplearning4j, and Keras to train your distributed models. By the end of this book, you'll have gained experience with the implementation of your models on a variety of use cases. What you will learn

Understand the basics of deep learning
Set up Apache Spark for deep learning
Understand the principles of distribution modeling and different types of neural networks
Obtain an understanding of deep learning algorithms
Discover textual analysis and deep learning with Spark
Use popular deep learning frameworks, such as Deeplearning4j, TensorFlow, and Keras
Explore popular deep learning algorithms

Who this book is for If you are a Scala developer, data scientist, or data analyst who wants to learn how to use Spark for implementing efficient deep learning models,

Hands-On Deep Learning with Apache Spark is for you.

Knowledge of the core machine learning concepts and some exposure to Spark will be helpful.

IBM Data Engine for Hadoop and Spark Oxford University Press

This book offers a direct, actionable plan CMOs can use to map out initiatives that are properly sequenced and designed for success—regardless of where their marketing organization is in the process. The authors pose the following critical questions to marketers: (1) How should modern marketers be thinking about artificial intelligence and machine learning? and (2) How should marketers be developing a strategy and plan to implement AI into their marketing toolkit? The opening chapters provide marketing leaders with an overview of what exactly AI is and how it is different than traditional computer science approaches.

Venkatesan and Lecinski, then, propose a best-practice, five-stage framework for implementing what they term the "AI Marketing Canvas." Their approach is based on research and interviews they conducted with leading marketers, and offers many tangible examples of what brands are doing at each stage of the AI Marketing Canvas. By way of guidance, Venkatesan and Lecinski provide examples of brands—including Google, Lyft, Ancestry.com, and Coca-Cola—that have successfully woven AI into their marketing strategies. The book concludes with a discussion of important implications for marketing leaders—for your team and culture.

Apache Spark 2.x Machine Learning Cookbook Apress

This book constitutes selected papers from the 18th European, Mediterranean, and Middle Eastern Conference, EMCIS 2021, which took place during December 8-9, 2021. The conference was initially planned to take place in Dubai, UAE, but had to change to an online event due to the COVID-19 pandemic. EMCIS covers technical, organizational, business, and social issues in the application of information technology and is dedicated to the definition and establishment of Information Systems (IS) as a discipline of high impact for IS professionals and practitioners. It focuses on approaches that facilitate the identification of innovative research of significant relevance to the IS discipline following sound research methodologies that lead to results of measurable impact. The 54 full papers presented in this volume were carefully reviewed and selected from a total of 155 submissions. They were organized in topical sections named: Big Data and Analytics; Blockchain Technology and Applications; Cloud Computing; Digital Governance; Digital Services and Social Media; Emerging Computing Technologies and Trends for Business Process Management; Healthcare Information Systems; Information Systems security and Information Privacy Protection; Innovative Research Projects; IT Governance and Alignment; and Management and Organisational Issues in Information Systems.

Mastering Machine Learning on AWS O'Reilly Media

Build efficient data flow and machine learning programs with this flexible, multi-functional open-source cluster-computing framework

Key Features
Master the art of real-time big data processing and machine learning
Explore a wide range of use-cases to analyze large data
Discover ways to optimize your work by using many features of Spark 2.x and Scala

Book Description
Apache Spark is an in-memory, cluster-based data processing system that provides a wide range of functionalities such as big data processing, analytics, machine learning, and more. With this Learning Path, you can take your knowledge of Apache Spark to the next level by learning how to expand Spark's functionality and building your own data flow and machine learning programs on this platform. You will work with the different modules in Apache Spark, such as interactive querying with Spark SQL, using DataFrames and datasets, implementing streaming analytics with

Spark Streaming, and applying machine learning and deep learning techniques on Spark using MLlib and various external tools. By the end of this elaborately designed Learning Path, you will have all the knowledge you need to master Apache Spark, and build your own big data processing and analytics pipeline quickly and without any hassle. This Learning Path includes content from the following Packt products: Mastering Apache Spark 2.x by Romeo Kienzler Scala and Spark for Big Data Analytics by Md. Rezaul Karim, Sridhar Alla Apache Spark 2.x Machine Learning Cookbook by Siamak Amirghodsi, Meenakshi Rajendran, Broderick Hall, Shuen Mei Cookbook What you will learn Get to grips with all the features of Apache Spark 2.x Perform highly optimized real-time big data processing Use ML and DL techniques with Spark MLlib and third-party tools Analyze structured and unstructured data using SparkSQL and GraphX Understand tuning, debugging, and monitoring of big data applications Build scalable and fault-tolerant streaming applications Develop scalable recommendation engines Who this book is for If you are an intermediate-level Spark developer looking to master the advanced capabilities and use-cases of Apache Spark 2.x, this Learning Path is ideal for you. Big data professionals who want to learn how to integrate and use the features of Apache Spark and build a strong big data pipeline will also find this Learning Path useful. To grasp the concepts explained in this Learning Path, you must know the fundamentals of Apache Spark and Scala.

[High Performance Spark](#) Springer

Apache Spark Machine Learning Blueprints Packt Publishing Ltd

The Human Element of Big Data Springer Nature

Machine learning continues to have myriad applications across industries and fields. To ensure this technology is utilized appropriately and to its full potential, organizations must better understand exactly how and where it can be adapted. Further study on the applications of machine learning is required to discover its best practices, challenges, and strategies. The Research Anthology on Machine Learning Techniques, Methods, and Applications provides a thorough consideration of the innovative and emerging research within the area of machine learning. The book discusses how the technology has been used in the past as well as potential ways it can be used in the future to ensure industries continue to develop and grow. Covering a range of topics such as artificial intelligence, deep learning, cybersecurity, and robotics, this major reference work is ideal for computer scientists, managers, researchers, scholars, practitioners, academicians, instructors, and students.

[Apache Spark Deep Learning Cookbook](#) Lulu.com

Apache Spark is amazing when everything clicks. But if you haven't seen the performance improvements you expected, or still don't feel confident enough to use Spark in production, this practical book is for you. Authors Holden Karau and Rachel Warren demonstrate performance optimizations to help your Spark queries run faster and handle larger data sizes, while using fewer resources. Ideal for software engineers, data engineers, developers, and system administrators working with large-scale data applications, this book describes techniques that can reduce data infrastructure costs and developer hours. Not only will you gain a more comprehensive understanding of Spark, you'll also learn how to make it sing. With this book, you'll explore: How Spark SQL's new interfaces improve performance over SQL's RDD data structure The choice between data joins in Core Spark and Spark SQL Techniques for getting the most out of standard RDD transformations How to work around performance issues in Spark's key/value pair paradigm Writing high-performance Spark code without Scala or the JVM How to test for functionality and performance when applying suggested improvements Using

Spark MLlib and Spark ML machine learning libraries Spark's Streaming components and external community packages **Learning Spark SQL** Apache Spark Machine Learning Blueprints Advanced analytics on your Big Data with latest Apache Spark 2.x About This Book An advanced guide with a combination of instructions and practical examples to extend the most up-to-date Spark functionalities. Extend your data processing capabilities to process huge chunk of data in minimum time using advanced concepts in Spark. Master the art of real-time processing with the help of Apache Spark 2.x Who This Book Is For If you are a developer with some experience with Spark and want to strengthen your knowledge of how to get around in the world of Spark, then this book is ideal for you. Basic knowledge of Linux, Hadoop and Spark is assumed. Reasonable knowledge of Scala is expected. What You Will Learn Examine Advanced Machine Learning and DeepLearning with MLlib, SparkML, SystemML, H2O and DeepLearning4j Study highly optimised unified batch and real-time data processing using SparkSQL and Structured Streaming Evaluate large-scale Graph Processing and Analysis using GraphX and GraphFrames Apply Apache Spark in Elastic deployments using Jupyter and Zeppelin Notebooks, Docker, Kubernetes and the IBM Cloud Understand internal details of cost based optimizers used in Catalyst, SystemML and GraphFrames Learn how specific parameter settings affect overall performance of an Apache Spark cluster Leverage Scala, R and python for your data science projects In Detail Apache Spark is an in-memory cluster-based parallel processing system that provides a wide range of functionalities such as graph processing, machine learning, stream processing, and SQL. This book aims to take your knowledge of Spark to the next level by teaching you how to expand Spark's functionality and implement your data flows and machine/deep learning programs on top of the platform. The book commences with an overview of the Spark ecosystem. It will introduce you to Project Tungsten and Catalyst, two of the major advancements of Apache Spark 2.x. You will understand how memory management and binary processing, cache-aware computation, and code generation are used to speed things up dramatically. The book extends to show how to incorporate H2O, SystemML, and DeepLearning4j for machine learning, and Jupyter Notebooks and Kubernetes/Docker for cloud-based Spark. During the course of the book, you will learn about the latest enhancements to Apache Spark 2.x, such as interactive querying of live data and unifying DataFrames and Datasets. You will also learn about the updates on the APIs and how DataFrames and Datasets affect SQL, machine learning, graph processing, and streaming. You will learn to use Spark as a big data operating system, understand how to implement advanced analytics on the new APIs, and explore how easy it is to use Spark in day-to-day tasks. Style and approach This book is an extensive guide to Apache Spark modules and tools and shows how Spark's functionality can be extended for real-time processing and storage with worked examples.

[Learning Spark](#) Packt Publishing Ltd

This book covers three major parts of Big Data: concepts, theories and applications. Written by world-renowned leaders in Big Data, this book explores the problems, possible solutions and directions for Big Data in research and practice. It also focuses on high level concepts such as definitions of Big Data from different angles; surveys in research and applications; and existing tools, mechanisms, and systems in practice. Each chapter is independent from the other chapters, allowing users to read any chapter directly. After examining the practical side of Big Data, this book presents theoretical perspectives. The theoretical research ranges from Big Data representation, modeling and topology to distribution and dimension reducing. Chapters also

investigate the many disciplines that involve Big Data, such as statistics, data mining, machine learning, networking, algorithms, security and differential geometry. The last section of this book introduces Big Data applications from different communities, such as business, engineering and science. Big Data Concepts, Theories and Applications is designed as a reference for researchers and advanced level students in computer science, electrical engineering and mathematics. Practitioners who focus on information systems, big data, data mining, business analysis and other related fields will also find this material valuable.

Stream Processing with Apache Spark Apress

A solution-based guide to put your deep learning models into production with the power of Apache Spark Key Features Discover practical recipes for distributed deep learning with Apache Spark Learn to use libraries such as Keras and TensorFlow Solve problems in order to train your deep learning models on Apache Spark Book Description With deep learning gaining rapid mainstream adoption in modern-day industries, organizations are looking for ways to unite popular big data tools with highly efficient deep learning libraries. As a result, this will help deep learning models train with higher efficiency and speed. With the help of the Apache Spark Deep Learning Cookbook, you'll work through specific recipes to generate outcomes for deep learning algorithms, without getting bogged down in theory. From setting up Apache Spark for deep learning to implementing types of neural net, this book tackles both common and not so common problems to perform deep learning on a distributed environment. In addition to this, you'll get access to deep learning code within Spark that can be reused to answer similar problems or tweaked to answer slightly different problems. You will also learn how to stream and cluster your data with Spark. Once you have got to grips with the basics, you'll explore how to implement and deploy deep learning models, such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) in Spark, using popular libraries such as TensorFlow and Keras. By the end of the book, you'll have the expertise to train and deploy efficient deep learning models on Apache Spark. What you will learn Set up a fully functional Spark environment Understand practical machine learning and deep learning concepts Apply built-in machine learning libraries within Spark Explore libraries that are compatible with TensorFlow and Keras Explore NLP models such as Word2vec and TF-IDF on Spark Organize dataframes for deep learning evaluation Apply testing and training modeling to ensure accuracy Access readily available code that may be reusable Who this book is for If you're looking for a practical and highly useful resource for implementing efficiently distributed deep learning models with Apache Spark, then the Apache Spark Deep Learning Cookbook is for you. Knowledge of the core machine learning concepts and a basic understanding of the Apache Spark framework is required to get the best out of this book. Additionally, some programming knowledge in Python is a plus.

Scalable Uncertainty Management Springer Nature

Before you can build analytics tools to gain quick insights, you first need to know how to process data in real time. With this practical guide, developers familiar with Apache Spark will learn how to put this in-memory framework to use for streaming data. You'll discover how Spark enables you to write streaming jobs in almost the same way you write batch jobs. Authors Gerard Maas and François Garillot help you explore the theoretical underpinnings of Apache Spark. This comprehensive guide features two sections that compare and contrast the streaming APIs Spark now supports: the original Spark Streaming library and the newer Structured Streaming API. Learn fundamental stream processing concepts and examine different streaming

architectures Explore Structured Streaming through practical examples; learn different aspects of stream processing in detail Create and operate streaming jobs and applications with Spark Streaming; integrate Spark Streaming with other Spark APIs Learn advanced Spark Streaming techniques, including approximation algorithms and machine learning algorithms Compare Apache Spark to other stream processing projects, including Apache Storm, Apache Flink, and Apache Kafka Streams [Beginning Apache Spark 2](#) CRC Press

Learn how to use, deploy, and maintain Apache Spark with this comprehensive guide, written by the creators of the open-source cluster-computing framework. With an emphasis on improvements and new features in Spark 2.0, authors Bill Chambers and Matei Zaharia break down Spark topics into distinct sections, each with unique goals. You'll explore the basic operations and common functions of Spark's structured APIs, as well as Structured Streaming, a new high-level API for building end-to-end streaming applications. Developers and system administrators will learn the fundamentals of monitoring, tuning, and debugging Spark, and explore machine learning techniques and scenarios for employing MLlib, Spark's scalable machine-learning library. Get a gentle overview of big data and Spark Learn about DataFrames, SQL, and Datasets—Spark's core APIs—through worked examples Dive into Spark's low-level APIs, RDDs, and execution of SQL and DataFrames Understand how Spark runs on a cluster Debug, monitor, and tune Spark clusters and applications Learn the power of Structured Streaming, Spark's stream-processing engine Learn how you can apply MLlib to a variety of problems, including classification or recommendation

[Automated Machine Learning for Business](#) "O'Reilly Media, Inc."

Access real-world documentation and examples for the Spark platform for building large-scale, enterprise-grade machine learning applications. The past decade has seen an astonishing series of advances in machine learning. These breakthroughs are disrupting our everyday life and making an impact across every industry. Next-Generation Machine Learning with Spark provides a gentle introduction to Spark and Spark MLlib and advances to more powerful, third-party machine learning algorithms and libraries beyond what is available in the standard Spark MLlib library. By the end of this book, you will be able to apply your knowledge to real-world use cases through dozens of practical examples and insightful explanations. What You Will Learn Be introduced to machine learning, Spark, and Spark MLlib 2.4.x Achieve lightning-fast gradient boosting on Spark with the XGBoost4J-Spark and LightGBM libraries Detect anomalies with the Isolation Forest algorithm for Spark Use the Spark NLP and Stanford CoreNLP libraries that support multiple languages Optimize your ML workload with the Alluxio in-memory data accelerator for Spark Use GraphX and GraphFrames for Graph Analysis Perform image recognition using convolutional neural networks Utilize the Keras framework and distributed deep learning libraries with Spark Who This Book Is For Data scientists and machine learning engineers who want to take their knowledge to the next level and use Spark and more powerful, next-generation algorithms and libraries beyond what is available in the standard Spark MLlib library; also serves as a primer for aspiring data scientists and engineers who need an introduction to machine learning, Spark, and Spark MLlib.

Big Data Processing with Apache Spark Springer

Develop a range of cutting-edge machine learning projects with Apache Spark using this actionable guide About This Book Customize Apache Spark and R to fit your analytical needs in customer research, fraud detection, risk analytics, and recommendation engine development Develop a set of practical

Machine Learning applications that can be implemented in real-life projects A comprehensive, project-based guide to improve and refine your predictive models for practical implementation Who This Book Is For If you are a data scientist, a data analyst, or an R and SPSS user with a good understanding of machine learning concepts, algorithms, and techniques, then this is the book for you. Some basic understanding of Spark and its core elements and application is required. What You Will Learn Set up Apache Spark for machine learning and discover its impressive processing power Combine Spark and R to unlock detailed business insights essential for decision making Build machine learning systems with Spark that can detect fraud and analyze financial risks Build predictive models focusing on customer scoring and service ranking Build a recommendation systems using SPSS on Apache Spark Tackle parallel computing and find out how it can support your machine learning projects Turn open data and communication data into actionable insights by making use of various forms of machine learning In Detail There's a reason why Apache Spark has become one of the most popular tools in Machine Learning - its ability to handle huge datasets at an impressive speed means you can be much more responsive to the data at your disposal. This book shows you Spark at its very best, demonstrating how to connect it with R and unlock maximum value not only from the tool but also from your data. Packed with a range of project "blueprints" that demonstrate some of the most interesting challenges that Spark can help you tackle, you'll find out how to use Spark notebooks and access, clean, and join different datasets before putting your knowledge into practice with some real-world projects, in which you will see how Spark Machine Learning can help you with everything from fraud detection to analyzing customer attrition. You'll also find out how to build a recommendation engine using Spark's parallel computing powers. Style and approach This book offers a step-by-step approach to setting up Apache Spark, and use other analytical tools with it to process Big Data and build machine learning projects. The initial chapters focus more on the theory aspect of machine learning with Spark, while each of the later chapters focuses on building standalone projects using Spark. Springer Nature

The book presents a collection of peer-reviewed articles from the International Conference on Advances and Applications of Artificial Intelligence and Machine Learning - ICAAAIML 2020. The book covers research in artificial intelligence, machine learning, and deep learning applications in healthcare, agriculture, business, and security. This volume contains research papers from academicians, researchers as well as students. There are also papers on core concepts of computer networks, intelligent system design and deployment, real-time systems, wireless sensor networks, sensors and sensor nodes, software engineering, and image processing. This book will be a valuable resource for students, academics, and practitioners in the industry working on AI applications.

Natural Language Processing with Spark NLP Packt Publishing Ltd This book teaches the full process of how to conduct machine learning in an organizational setting. It develops the problem-solving mind-set needed for machine learning and takes the reader through several exercises using an automated machine learning tool. To build experience with machine learning, the book provides access to the industry-leading AutoML tool, DataRobot, and provides several data sets designed to build deep hands-on knowledge of machine learning.

Database Systems for Advanced Applications IGI Global This book covers the fundamentals of machine learning with Python in a concise and dynamic manner. It covers data mining and large-scale machine learning using Apache Spark. About This

Book Take your first steps in the world of data science by understanding the tools and techniques of data analysis Train efficient Machine Learning models in Python using the supervised and unsupervised learning methods Learn how to use Apache Spark for processing Big Data efficiently Who This Book Is For If you are a budding data scientist or a data analyst who wants to analyze and gain actionable insights from data using Python, this book is for you. Programmers with some experience in Python who want to enter the lucrative world of Data Science will also find this book to be very useful, but you don't need to be an expert Python coder or mathematician to get the most from this book. What You Will Learn Learn how to clean your data and ready it for analysis Implement the popular clustering and regression methods in Python Train efficient machine learning models using decision trees and random forests Visualize the results of your analysis using Python's Matplotlib library Use Apache Spark's MLlib package to perform machine learning on large datasets In Detail Join Frank Kane, who worked on Amazon and IMDb's machine learning algorithms, as he guides you on your first steps into the world of data science. Hands-On Data Science and Python Machine Learning gives you the tools that you need to understand and explore the core topics in the field, and the confidence and practice to build and analyze your own machine learning models. With the help of interesting and easy-to-follow practical examples, Frank Kane explains potentially complex topics such as Bayesian methods and K-means clustering in a way that anybody can understand them. Based on Frank's successful data science course, Hands-On Data Science and Python Machine Learning empowers you to conduct data analysis and perform efficient machine learning using Python. Let Frank help you unearth the value in your data using the various data mining and data analysis techniques available in Python, and to develop efficient predictive models to predict future results. You will also learn how to perform large-scale machine learning on Big Data using Apache Spark. The book covers preparing your data for analysis, training machine learning models, and visualizing the final data analysis. Style and approach This comprehensive book is a perfect blend of theory and hands-on code examples in Python which can be used for your reference at any time.

Hands-On Big Data Analytics with PySpark Packt Publishing Ltd Use Hadoop to solve business problems by learning from a rich set of real-life case studies About This Book Solve real-world business problems using Hadoop and other Big Data technologies Build efficient data lakes in Hadoop, and develop systems for various business cases like improving marketing campaigns, fraud detection, and more Power packed with six case studies to get you going with Hadoop for Business Intelligence Who This Book Is For If you are interested in building efficient business solutions using Hadoop, this is the book for you This book assumes that you have basic knowledge of Hadoop, Java, and any scripting language. What You Will Learn Learn about the evolution of Hadoop as the big data platform Understand the basics of Hadoop architecture Build a 360 degree view of your customer using Sqoop and Hive Build and run classification models on Hadoop using BigML Use Spark and Hadoop to build a fraud detection system Develop a churn detection system using Java and MapReduce Build an IoT-based data collection and visualization system Get to grips with building a Hadoop-based Data Lake for large enterprises Learn about the coexistence of NoSQL and In-Memory databases in the Hadoop ecosystem In Detail If you have a basic understanding of Hadoop and want to put your knowledge to use to build fantastic Big Data solutions for business, then this book is for you. Build six real-life, end-to-end solutions using the tools in the Hadoop ecosystem, and take

your knowledge of Hadoop to the next level. Start off by understanding various business problems which can be solved using Hadoop. You will also get acquainted with the common architectural patterns which are used to build Hadoop-based solutions. Build a 360-degree view of the customer by working with different types of data, and build an efficient fraud detection system for a financial institution. You will also develop a system in Hadoop to improve the effectiveness of marketing campaigns. Build a churn detection system for a telecom company, develop an Internet of Things (IoT) system to monitor the environment in a factory, and build a data lake - all making use of the concepts

and techniques mentioned in this book. The book covers other technologies and frameworks like Apache Spark, Hive, Sqoop, and more, and how they can be used in conjunction with Hadoop. You will be able to try out the solutions explained in the book and use the knowledge gained to extend them further in your own problem space. **Style and approach** This is an example-driven book where each chapter covers a single business problem and describes its solution by explaining the structure of a dataset and tools required to process it. Every project is demonstrated with a step-by-step approach, and explained in a very easy-to-understand manner.

Related with Apache Spark Machine Learning Blueprints:

- Star Online Practice Test : [click here](#)