

1. Record Nr.	UNINA9910148730003321
Autore	Karim Rezaul
Titolo	Large scale machine learning with Spark : discover everything you need to build robust machine learning applications with Spark 2.0 / / Md. Rezaul Karim, Md. Mahedi Kaysar
Pubbl/distr/stampa	Birmingham, England ; ; Mumbai, India : , : Packt Publishing, , 2016 ©2016
ISBN	1-78588-371-2
Edizione	[1st edition]
Descrizione fisica	1 online resource (472 pages) : illustrations
Disciplina	006.31
Soggetti	Machine learning
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Note generali	Includes index.
Sommario/riassunto	Discover everything you need to build robust machine learning applications with Spark 2.0 About This Book Get the most up-to-date book on the market that focuses on design, engineering, and scalable solutions in machine learning with Spark 2.0.0 Use Spark's machine learning library in a big data environment You will learn how to develop high-value applications at scale with ease and a develop a personalized design Who This Book Is For This book is for data science engineers and scientists who work with large and complex data sets. You should be familiar with the basics of machine learning concepts, statistics, and computational mathematics. Knowledge of Scala and Java is advisable. What You Will Learn Get solid theoretical understandings of ML algorithms Configure Spark on cluster and cloud infrastructure to develop applications using Scala, Java, Python, and R Scale up ML applications on large cluster or cloud infrastructures Use Spark ML and MLlib to develop ML pipelines with recommendation system, classification, regression, clustering, sentiment analysis, and dimensionality reduction Handle large texts for developing ML applications with strong focus on feature engineering Use Spark Streaming to develop ML applications for real-time streaming Tune ML models with cross-validation, hyperparameters tuning and train split

Enhance ML models to make them adaptable for new data in dynamic and incremental environments In Detail Data processing, implementing related algorithms, tuning, scaling up and finally deploying are some crucial steps in the process of optimising any application. Spark is capable of handling large-scale batch and streaming data to figure out when to cache data in memory and processing them up to 100 times faster than Hadoop-based MapReduce. This means predictive analytics can be applied to streaming and batch to develop complete machine learning (ML) applications a lot quicker, making Spark an ideal candidate for large data-intensive applications. This book focuses on design engineering and scalable solutions using ML with Spark. First, you will learn how to install Spark with all new features from the latest Spark 2.0 release. Moving on, you'll explore important concepts such as advanced feature engineering with RDD and Datasets. After studying developing and deploying applications, you will see how to use external libraries with Spark. In summary, you will be able to develop complete and personalised ML applications from da...

---