1.	Record Nr.	UNINA9910796534103321
	Titolo	Apache spark 2.x machine learning cookbook : over 100 recipes to simplify machine learning model implementations with Spark / / Siamak Amirghodsi [and three others]
	Pubbl/distr/stampa	Birmingham, England : , : Packt Publishing, , 2017 ©2017
	Edizione	[1st edition]
	Descrizione fisica	1 online resource (1 volume) : illustrations
	Disciplina	006.754
	Soggetti	Data mining - Computer programs
	Lingua di pubblicazione	Inglese
	Formato	Materiale a stampa
	Livello bibliografico	Monografia
	Note generali	Includes index.
	Sommario/riassunto	Simplify machine learning model implementations with Spark About This Book Solve the day-to-day problems of data science with Spark This unique cookbook consists of exciting and intuitive numerical recipes Optimize your work by acquiring, cleaning, analyzing, predicting, and visualizing your data Who This Book Is For This book is for Scala developers with a fairly good exposure to and understanding of machine learning techniques, but lack practical implementations with Spark. A solid knowledge of machine learning algorithms is assumed, as well as hands-on experience of implementing ML algorithms with Scala. However, you do not need to be acquainted with the Spark ML libraries and ecosystem. What You Will Learn Get to know how Scala and Spark go hand-in-hand for developers when developing ML systems with Spark Build a recommendation engine that scales with Spark Find out how to build unsupervised clustering systems to classify data in Spark Build machine learning systems with the Decision Tree and Ensemble models in Spark Deal with the curse of high-dimensionality in big data using Spark Implement Text analytics for Search Engines in Spark Streaming Machine Learning System implementation using Spark In Detail Machine learning aims to extract knowledge from data, relying on fundamental concepts in computer

science, statistics, probability, and optimization. Learning about algorithms enables a wide range of applications, from everyday tasks such as product recommendations and spam filtering to cutting edge applications such as self-driving cars and personalized medicine. You will gain hands-on experience of applying these principles using Apache Spark, a resilient cluster computing system well suited for large-scale machine learning tasks. This book begins with a quick overview of setting up the necessary IDEs to facilitate the execution of code examples that will be covered in various chapters. It also highlights some key issues developers face while working with machine learning algorithms on the Spark platform. We progress by uncovering the various Spark APIs and the implementation of ML algorithms with developing classification systems, recommendation engines, text analytics, clustering, and learning systems. Toward the final chapters, we'll focus on building high-end applications and explain various unsupervised methodologies and challenges to tackle when implementing with big data ML systems. Style and approach This book is packed with intu...