

Machine Learning with PySpark

Master the new features in PySpark 3.1 to develop data-driven, intelligent applications. This updated edition covers topics ranging from building scalable machine learning models, to natural language processing, to recommender systems.

Machine Learning with PySpark, Second Edition begins with the fundamentals of Apache Spark, including the latest updates to the framework. Next, you will learn the full spectrum of traditional machine learning algorithm implementations, along with natural language processing and recommender systems. You'll gain familiarity with the critical process of selecting machine learning algorithms, data ingestion, and data processing to solve business problems. You'll see a demonstration of how to build supervised machine learning models such as linear regression, logistic regression, decision trees, and random forests. You'll also learn how to automate the steps using Spark pipelines, followed by unsupervised models such as K-means and hierarchical clustering. A section on Natural Language Processing (NLP) covers text processing, text mining, and embeddings for classification. This new edition also introduces Koalas in Spark and how to automate data workflow using Airflow and PySpark's latest ML library.

After completing this book, you will understand how to use PySpark's machine learning library to build and train various machine learning models, along with related components such as data ingestion, processing and visualization to develop data-driven intelligent applications

What you will learn:

- Build a spectrum of supervised and unsupervised machine learning algorithms
- Use PySpark's machine learning library to implement machine learning and recommender systems
- Leverage the new features in PySpark's machine learning library
- Understand data processing using Koalas in Spark
- Handle issues around feature engineering, class balance, bias and variance, and cross validation to build optimally fit models

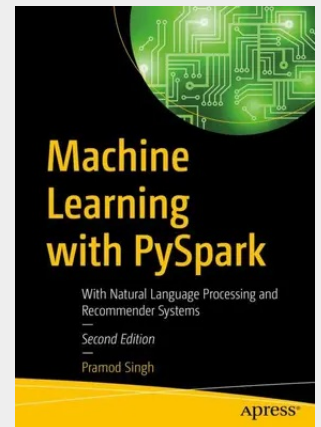
Who This Book Is For

Data science and machine learning professionals.

Master the new features in PySpark 3.1 to develop data-driven, intelligent applications. This updated edition covers topics ranging from building scalable machine learning models, to natural language processing, to recommender systems. *Machine Learning with PySpark, Second Edition* begins with the fundamentals of Apache Spark, including the latest updates to the framework. Next, you will learn the full spectrum of traditional machine learning algorithm implementations, along with natural language processing and recommender systems. You'll gain familiarity with the critical process of selecting machine learning algorithms, data ingestion, and data processing to solve business problems. You'll see a demonstration of how to build supervised machine learning models such as linear regression, logistic regression, decision trees, and random forests. You'll also learn how to automate the steps using Spark pipelines, followed by unsupervised models such as K-means and hierarchical clustering. A section on Natural Language Processing (NLP) covers text processing, text mining, and embeddings for classification. This new edition also introduces Koalas in Spark and how to automate data workflow using Airflow and PySpark's latest ML library. After completing this book, you will understand how to use PySpark's machine learning library to build and train various machine learning models, along with related components such as data ingestion, processing and visualization to develop data-driven intelligent applications

What you will learn:

- Build a spectrum of supervised and unsupervised machine learning algorithms
- Use PySpark's machine learning library to implement machine learning and recommender systems



64,19 €

59,99 € (zzgl. MwSt.)

Lieferfrist: bis zu 10 Tage

Artikelnummer: 9781484277768

Medium: Buch

ISBN: 978-1-4842-7776-8

Verlag: Apress

Erscheinungstermin: 09.12.2021

Sprache(n): Englisch

Auflage: 2. Auflage 2021

Produktform: Kartoniert

Gewicht: 460 g

Seiten: 220

Format (B x H): 178 x 254 mm

