

1. Record Nr.	UNINA9910299357903321
Autore	Fernández Alberto (Computer scientist)
Titolo	Learning from Imbalanced Data Sets // by Alberto Fernández, Salvador García, Mikel Galar, Ronaldo C. Prati, Bartosz Krawczyk, Francisco Herrera
Pubbl/distr/stampa	Cham : , : Springer International Publishing : , : Imprint : Springer, , 2018
ISBN	3-319-98074-2
Edizione	[1st ed. 2018.]
Descrizione fisica	1 online resource (385 pages) : illustrations (some color)
Disciplina	006.3
Soggetti	Artificial intelligence Computers Computer communication systems Artificial Intelligence Information Systems and Communication Service Computer Communication Networks
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di bibliografia	Includes bibliographical references.
Nota di contenuto	1 Introduction to KDD and Data Science -- 2 Foundations on Imbalanced Classification -- 3 Performance measures -- 4 Cost-sensitive Learning -- 5 Data Level Preprocessing Methods -- 6 Algorithm-level Approaches -- 7 Ensemble Learning -- 8 Imbalanced Classification with Multiple Classes -- 9 Dimensionality Reduction for Imbalanced Learning -- 10 Data Intrinsic Characteristics -- 11 Learning from Imbalanced Data Streams -- 12 Non-Classical Imbalanced Classification Problems -- 13 Imbalanced Classification for Big Data -- 14 Software and Libraries for Imbalanced Classification. .
Sommario/riassunto	This book provides a general and comprehensible overview of imbalanced learning. It contains a formal description of a problem, and focuses on its main features, and the most relevant proposed solutions. Additionally, it considers the different scenarios in Data Science for which the imbalanced classification can create a real challenge. This book stresses the gap with standard classification tasks by reviewing the case studies and ad-hoc performance metrics that are applied in

this area. It also covers the different approaches that have been traditionally applied to address the binary skewed class distribution. Specifically, it reviews cost-sensitive learning, data-level preprocessing methods and algorithm-level solutions, taking also into account those ensemble-learning solutions that embed any of the former alternatives. Furthermore, it focuses on the extension of the problem for multi-class problems, where the former classical methods are no longer to be applied in a straightforward way. This book also focuses on the data intrinsic characteristics that are the main causes which, added to the uneven class distribution, truly hinders the performance of classification algorithms in this scenario. Then, some notes on data reduction are provided in order to understand the advantages related to the use of this type of approaches. Finally this book introduces some novel areas of study that are gathering a deeper attention on the imbalanced data issue. Specifically, it considers the classification of data streams, non-classical classification problems, and the scalability related to Big Data. Examples of software libraries and modules to address imbalanced classification are provided. This book is highly suitable for technical professionals, senior undergraduate and graduate students in the areas of data science, computer science and engineering. It will also be useful for scientists and researchers to gain insight on the current developments in this area of study, as well as future research directions. .
