

Process Mining Data Science In Action

This book constitutes revised selected papers from the International Workshops held at the Second International Conference on Process Mining, ICPM 2020, which took place during October 4-9, 2020. The conference was planned to take place in Padua, Italy, but had to be held online due to the COVID-19 pandemic. The conference focuses on the area of process mining research and practice, including theory, algorithmic challenges, and applications. The co-located workshops provided a forum for novel research ideas. The 29 papers included in this volume were carefully reviewed and selected from 59 submissions. They stem from the following workshops: 1st International Workshop on Event Data and Behavioral Analytics (EDBA) 1st International Workshop on Leveraging Machine Learning in Process Mining (ML4PM) 1st International Workshop on Streaming Analytics for Process Mining (SA4PM'20) 5th International Workshop on Process Querying, Manipulation, and Intelligence (PQMI) 3rd International Workshop on Process-Oriented Data Science for Healthcare (PODS4H) 1st International Workshop on Trust and Privacy in Process Analytics (TPPA)

This book provides a practically applicable guide to the methodologies and technologies for the application of interactive process mining paradigm. Case studies are presented where this paradigm has been successfully applied in emergency medicine, surgery processes, human behavior modelling, strokes and outpatients' services, enabling the reader to develop a deep understanding of how to apply process mining technologies in healthcare to support them in inferring new knowledge from past actions, and providing accurate and personalized knowledge to improve their future clinical decision-making. Interactive Process Mining in Healthcare comprehensively covers how machine learning algorithms can be utilized to create real scientific evidence to improve daily healthcare protocols, and is a valuable resource for a variety of health professionals seeking to develop new methods to improve their clinical decision-making. Advancements in the technology and availability of data sources have led to the 'Big Data' era. Working with large data offers the potential to uncover more fine-grained patterns and take timely and accurate decisions, but it also creates a lot of challenges such as slow training and scalability of machine learning models. One of the major challenges in machine learning is to develop efficient and scalable learning algorithms, i.e., optimization techniques to solve large scale learning problems. Stochastic Optimization for Large-scale Machine Learning identifies different areas of improvement and recent research directions to tackle the challenge. Developed optimisation techniques are also explored to improve machine learning algorithms based on data access and on first and second order optimisation methods. Key Features: Bridges machine learning and Optimisation. Bridges theory and practice in machine learning. Identifies key research areas and recent research directions to solve large-scale machine learning problems. Develops optimisation techniques to improve machine learning algorithms for big data problems. The book will be a valuable reference to practitioners and researchers as well as students in the field of machine learning.

Praise for the Second Edition: "The authors present an intuitive and easy-to-read book. ... accompanied by many examples, proposed exercises, good references, and comprehensive appendices that initiate the reader unfamiliar with MATLAB." —Adolfo Alvarez Pinto, International Statistical Review "Practitioners of EDA who use MATLAB will want a copy of this book. ... The authors have done a great service by bringing together so many EDA routines, but their main accomplishment in this dynamic text is providing the understanding and tools to do EDA. —David A Huckaby, MAA Reviews Exploratory Data Analysis (EDA) is an important part of the data analysis process. The methods presented in this text are ones that should be in the toolkit of every data scientist. As computational sophistication has increased and data sets have grown in size and complexity, EDA has become an even more important process for visualizing and summarizing data before making assumptions to generate hypotheses and models. Exploratory Data Analysis with MATLAB, Third Edition presents EDA methods from a computational perspective and uses numerous examples and applications to show how the methods are used in practice. The authors use MATLAB code, pseudo-code, and algorithm descriptions to illustrate the concepts. The MATLAB code for examples, data sets, and the EDA Toolbox are available for download on the book's website. New to the Third Edition Random projections and estimating local intrinsic dimensionality Deep learning autoencoders and stochastic neighbor embedding Minimum spanning tree and additional cluster validity indices Kernel density estimation Plots for visualizing data distributions, such as beanplots and violin plots A chapter on visualizing categorical data

Big Data is a growing business trend, but there little advice available on how to use it practically. Written by a data mining expert with over 30 years of experience, this book uses case studies to help marketers, brand managers and IT professionals understand how to capture and measure data for marketing purposes.

Chandrika Kamath describes how techniques from the multi-disciplinary field of data mining can be used to address the modern problem of data overload in science and engineering domains. Starting with a survey of analysis problems in different applications, it identifies the common themes across these domains.

Historically, the term quality was used to measure performance in the context of products, processes and systems. With rapid growth in data and its usage, data quality is becoming quite important. It is important to connect these two aspects of quality to ensure better performance. This book provides a strong connection between the concepts in data science and process engineering that is necessary to ensure better quality levels and takes you through a systematic approach to measure holistic quality with several case studies. Features: Integrates data science, analytics and process engineering concepts Discusses how to create value by considering data, analytics and processes Examines metrics management technique that will help evaluate performance levels of processes, systems and models, including AI and machine learning approaches Reviews a structured approach for analytics execution

More and more information about business processes is recorded by information systems in the form of so-called "event logs". Despite the omnipresence of such data, most organizations diagnose problems based on fiction rather than facts. Process mining is an emerging discipline based on process model-driven approaches and data mining. It not only allows organizations to fully benefit from the information stored in their systems, but it can also be used to check the conformance of processes, detect bottlenecks, and predict execution problems. Wil van der Aalst delivers the first book on process mining. It aims to be self-contained while covering the entire process mining spectrum from process discovery to operational support. In Part I, the author provides the basics of business process modeling and data mining necessary to understand the remainder of the book. Part II focuses on process discovery as the most important process mining task. Part III moves beyond discovering the control flow of processes and highlights conformance checking, and organizational and time perspectives. Part IV guides the reader in

successfully applying process mining in practice, including an introduction to the widely used open-source tool ProM. Finally, Part V takes a step back, reflecting on the material presented and the key open challenges. Overall, this book provides a comprehensive overview of the state of the art in process mining. It is intended for business process analysts, business consultants, process managers, graduate students, and BPM researchers.

Process mining techniques can be used to discover, analyze and improve real processes, by extracting models from observed behavior. The aim of this book is conformance checking, one of the main areas of process mining. In conformance checking, existing process models are compared with actual observations of the process in order to assess their quality. Conformance checking techniques are a way to visualize the differences between assumed process represented in the model and the real process in the event log, pinpointing possible problems to address, and the business process management results that rely on these models. This book combines both application and research perspectives. It provides concrete use cases that illustrate the problems addressed by the techniques in the book, but at the same time, it contains complete conceptualization and formalization of the problem and the techniques, and through evaluations on the quality and the performance of the proposed techniques. Hence, this book brings the opportunity for business analysts willing to improve their organization processes, and also data scientists interested on the topic of process-oriented data science.

This book introduces readers to the field of conformance checking as a whole and outlines the fundamental relation between modelled and recorded behaviour. Conformance checking interrelates the modelled and recorded behaviour of a given process and provides techniques and methods for comparing and analysing observed instances of a process in the presence of a model, independent of the model's origin. Its goal is to provide an overview of the essential techniques and methods in this field at an intuitive level, together with precise formalisations of its underlying principles. The book is divided into three parts, that are meant to cover different perspectives of the field of conformance checking. Part I presents a comprehensive yet accessible overview of the essential concepts used to interrelate modelled and recorded behaviour. It also serves as a reference for assessing how conformance checking efforts could be applied in specific domains. Next, Part II provides readers with detailed insights into algorithms for conformance checking, including the most commonly used formal notions and their instantiation for specific analysis questions. Lastly, Part III highlights applications that help to make sense of conformance checking results, thereby providing a necessary next step to increase the value of a given process model. They help to interpret the outcomes of conformance checking and incorporate them by means of enhancement and repair techniques. Providing the core building blocks of conformance checking and describing its main applications, this book mainly addresses students specializing in business process management, researchers entering process mining and conformance checking for the first time, and advanced professionals whose work involves process evaluation, modelling and optimization.

Addresses the impacts of data mining on education and reviews applications in educational research teaching, and learning This book discusses the insights, challenges, issues, expectations, and practical implementation of data mining (DM) within educational mandates. Initial series of chapters offer a general overview of DM, Learning Analytics (LA), and data collection models in the context of educational research, while also defining and discussing data mining's four guiding principles— prediction, clustering, rule association, and outlier detection. The next series of chapters showcase the pedagogical applications of Educational Data Mining (EDM) and feature case studies drawn from Business, Humanities, Health Sciences, Linguistics, and Physical Sciences education that serve to highlight the successes and some of the limitations of data mining research applications in educational settings. The remaining chapters focus exclusively on EDM's emerging role in helping to advance educational research—from identifying at-risk students and closing socioeconomic gaps in achievement to aiding in teacher evaluation and facilitating peer conferencing. This book features contributions from international experts in a variety of fields. Includes case studies where data mining techniques have been effectively applied to advance teaching and learning Addresses applications of data mining in educational research, including: social networking and education; policy and legislation in the classroom; and identification of at-risk students Explores Massive Open Online Courses (MOOCs) to study the effectiveness of online networks in promoting learning and understanding the communication patterns among users and students Features supplementary resources including a primer on foundational aspects of educational mining and learning analytics Data Mining and Learning Analytics: Applications in Educational Research is written for both scientists in EDM and educators interested in using and integrating DM and LA to improve education and advance educational research.

This open access book comprehensively covers the fundamentals of clinical data science, focusing on data collection, modelling and clinical applications. Topics covered in the first section on data collection include: data sources, data at scale (big data), data stewardship (FAIR data) and related privacy concerns. Aspects of predictive modelling using techniques such as classification, regression or clustering, and prediction model validation will be covered in the second section. The third section covers aspects of (mobile) clinical decision support systems, operational excellence and value-based healthcare. Fundamentals of Clinical Data Science is an essential resource for healthcare professionals and IT consultants intending to develop and refine their skills in personalized medicine, using solutions based on large datasets from electronic health records or telemonitoring programmes. The book's promise is “no math, no code”and will explain the topics in a style that is optimized for a healthcare audience.

This book describes process mining use cases and business impact along the value chain, from corporate to local applications, representing the state of the art in domain know-how. Providing a set of industrial case studies and best practices, it complements academic publications on the topic. Further the book reveals the challenges and failures in order to offer readers practical insights and guidance on how to avoid the pitfalls and ensure successful operational deployment. The book is divided into three parts: Part I provides an introduction to the topic from fundamental principles to key success factors, and an overview of operational use cases. As a holistic description of process mining in a business environment, this part is particularly useful for readers not yet familiar with the topic. Part II presents detailed use cases written by contributors from a variety of functions and industries. Lastly, Part III provides a brief overview of the future of process mining, both from academic and operational perspectives. Based on a solid academic foundation, process mining has received increasing interest from operational businesses, with many companies already reaping the benefits. As the first book to present an overview of successful industrial applications, it is of particular interest to professionals who want to learn more about the possibilities and opportunities this new technology offers. It is also a valuable resource for researchers looking for empirical results when considering

requirements for enhancements and further developments.

"This book provides an overview of data mining techniques under an ethical lens, investigating developments in research best practices and examining experimental cases to identify potential ethical dilemmas in the information and communications technology sector"--Provided by publisher.

The main goal of this book is to explain the core ideas of process mining, and to demonstrate how they can be implemented using just some basic tools that are available to any computer scientist or data scientist. It describes how to analyze event logs in order to discover the behavior of real-world business processes. The end result can often be visualized as a graph, and the book explains how to use Python and Graphviz to render these graphs intuitively. Overall, it enables the reader to implement process mining techniques on his or her own, independently of any specific process mining tool. An introduction to two popular process mining tools, namely Disco and ProM, is also provided. In this second edition the code snippets have been updated to Python 3, and some smaller errors have been corrected. The book will be especially valuable for self-study or as a precursor to a more advanced text. Practitioners and students will be able to follow along on their own, even if they have no prior knowledge of the topic. After reading this book, they will be able to more confidently proceed to the research literature if needed.

Handbook of Statistical Analysis and Data Mining Applications, Second Edition, is a comprehensive professional reference book that guides business analysts, scientists, engineers and researchers, both academic and industrial, through all stages of data analysis, model building and implementation. The handbook helps users discern technical and business problems, understand the strengths and weaknesses of modern data mining algorithms and employ the right statistical methods for practical application. This book is an ideal reference for users who want to address massive and complex datasets with novel statistical approaches and be able to objectively evaluate analyses and solutions. It has clear, intuitive explanations of the principles and tools for solving problems using modern analytic techniques and discusses their application to real problems in ways accessible and beneficial to practitioners across several areas—from science and engineering, to medicine, academia and commerce. Includes input by practitioners for practitioners Includes tutorials in numerous fields of study that provide step-by-step instruction on how to use supplied tools to build models Contains practical advice from successful real-world implementations Brings together, in a single resource, all the information a beginner needs to understand the tools and issues in data mining to build successful data mining solutions Features clear, intuitive explanations of novel analytical tools and techniques, and their practical applications

Put Predictive Analytics into Action Learn the basics of Predictive Analysis and Data Mining through an easy to understand conceptual framework and immediately practice the concepts learned using the open source RapidMiner tool. Whether you are brand new to Data Mining or working on your tenth project, this book will show you how to analyze data, uncover hidden patterns and relationships to aid important decisions and predictions. Data Mining has become an essential tool for any enterprise that collects, stores and processes data as part of its operations. This book is ideal for business users, data analysts, business analysts, business intelligence and data warehousing professionals and for anyone who wants to learn Data Mining. You'll be able to: 1. Gain the necessary knowledge of different data mining techniques, so that you can select the right technique for a given data problem and create a general purpose analytics process. 2. Get up and running fast with more than two dozen commonly used powerful algorithms for predictive analytics using practical use cases. 3. Implement a simple step-by-step process for predicting an outcome or discovering hidden relationships from the data using RapidMiner, an open source GUI based data mining tool Predictive analytics and Data Mining techniques covered: Exploratory Data Analysis, Visualization, Decision trees, Rule induction, k-Nearest Neighbors, Naïve Bayesian, Artificial Neural Networks, Support Vector machines, Ensemble models, Bagging, Boosting, Random Forests, Linear regression, Logistic regression, Association analysis using Apriori and FP Growth, K-Means clustering, Density based clustering, Self Organizing Maps, Text Mining, Time series forecasting, Anomaly detection and Feature selection. Implementation files can be downloaded from the book companion site at www.LearnPredictiveAnalytics.com Demystifies data mining concepts with easy to understand language Shows how to get up and running fast with 20 commonly used powerful techniques for predictive analysis Explains the process of using open source RapidMiner tools Discusses a simple 5 step process for implementing algorithms that can be used for performing predictive analytics Includes practical use cases and examples

Data Analytics Applied to the Mining Industry describes the key challenges facing the mining sector as it transforms into a digital industry able to fully exploit process automation, remote operation centers, autonomous equipment and the opportunities offered by the industrial internet of things. It provides guidelines on how data needs to be collected, stored and managed to enable the different advanced data analytics methods to be applied effectively in practice, through use of case studies, and worked examples. Aimed at graduate students, researchers, and professionals in the industry of mining engineering, this book: Explains how to implement advanced data analytics through case studies and examples in mining engineering Provides approaches and methods to improve data-driven decision making Explains a concise overview of the state of the art for Mining Executives and Managers Highlights and describes critical opportunity areas for mining optimization Brings experience and learning in digital transformation from adjacent sectors

Written by renowned data science experts Foster Provost and Tom Fawcett, Data Science for Business introduces the fundamental principles of data science, and walks you through the "data-analytic thinking" necessary for extracting useful knowledge and business value from the data you collect. This guide also helps you understand the many data-mining techniques in use today. Based on an MBA course Provost has taught at New York University over the past ten years, Data Science for Business provides examples of real-world business problems to illustrate these principles. You'll not only learn how to improve communication between business stakeholders and data scientists, but also how participate intelligently in your company's data science projects. You'll also discover how to think data-analytically, and fully appreciate how data science methods can support business decision-making. Understand how data science fits in your organization—and how you can use it for competitive advantage Treat data as a business asset that requires careful investment if you're to gain real value Approach business problems data-analytically, using the data-mining process to gather good data in the most appropriate way Learn general concepts for actually extracting knowledge from data Apply data science principles when interviewing data science job candidates

"This book introduces you to R, RStudio, and the tidyverse, a collection of R packages designed to work together to make data science fast, fluent, and fun. Suitable for readers with no

previous programming experience"--

What are the possibilities for process mining in hospitals? In this book the authors provide an answer to this question by presenting a healthcare reference model that outlines all the different classes of data that are potentially available for process mining in healthcare and the relationships between them. Subsequently, based on this reference model, they explain the application opportunities for process mining in this domain and discuss the various kinds of analyses that can be performed. They focus on organizational healthcare processes rather than medical treatment processes. The combination of event data and process mining techniques allows them to analyze the operational processes within a hospital based on facts, thus providing a solid basis for managing and improving processes within hospitals. To this end, they also explicitly elaborate on data quality issues that are relevant for the data aspects of the healthcare reference model. This book mainly targets advanced professionals involved in areas related to business process management, business intelligence, data mining, and business process redesign for healthcare systems as well as graduate students specializing in healthcare information systems and process analysis.

Powerful, Flexible Tools for a Data-Driven World As the data deluge continues in today's world, the need to master data mining, predictive analytics, and business analytics has never been greater. These techniques and tools provide unprecedented insights into data, enabling better decision making and forecasting, and ultimately the solution of increasingly complex problems. **Learn from the Creators of the RapidMiner Software** Written by leaders in the data mining community, including the developers of the RapidMiner software, *RapidMiner: Data Mining Use Cases and Business Analytics Applications* provides an in-depth introduction to the application of data mining and business analytics techniques and tools in scientific research, medicine, industry, commerce, and diverse other sectors. It presents the most powerful and flexible open source software solutions: RapidMiner and RapidAnalytics. The software and their extensions can be freely downloaded at www.RapidMiner.com. **Understand Each Stage of the Data Mining Process** The book and software tools cover all relevant steps of the data mining process, from data loading, transformation, integration, aggregation, and visualization to automated feature selection, automated parameter and process optimization, and integration with other tools, such as R packages or your IT infrastructure via web services. The book and software also extensively discuss the analysis of unstructured data, including text and image mining. **Easily Implement Analytics Approaches Using RapidMiner and RapidAnalytics** Each chapter describes an application, how to approach it with data mining methods, and how to implement it with RapidMiner and RapidAnalytics. These application-oriented chapters give you not only the necessary analytics to solve problems and tasks, but also reproducible, step-by-step descriptions of using RapidMiner and RapidAnalytics. The case studies serve as blueprints for your own data mining applications, enabling you to effectively solve similar problems.

An introduction to the modeling of business information systems, with processes formally modeled using Petri nets. This comprehensive introduction to modeling business-information systems focuses on business processes. It describes and demonstrates the formal modeling of processes in terms of Petri nets, using a well-established theory for capturing and analyzing models with concurrency. The precise semantics of this formal method offers a distinct advantage for modeling processes over the industrial modeling languages found in other books on the subject. Moreover, the simplicity and expressiveness of the Petri nets concept make it an ideal language for explaining foundational concepts and constructing exercises. After an overview of business information systems, the book introduces the modeling of processes in terms of classical Petri nets. This is then extended with data, time, and hierarchy to model all aspects of a process. Finally, the book explores analysis of Petri net models to detect design flaws and errors in the design process. The text, accessible to a broad audience of professionals and students, keeps technicalities to a minimum and offers numerous examples to illustrate the concepts covered. Exercises at different levels of difficulty make the book ideal for independent study or classroom use.

With big data analytics comes big insights into profitability Big data is big business. But having the data and the computational power to process it isn't nearly enough to produce meaningful results. *Big Data, Data Mining, and Machine Learning: Value Creation for Business Leaders and Practitioners* is a complete resource for technology and marketing executives looking to cut through the hype and produce real results that hit the bottom line. Providing an engaging, thorough overview of the current state of big data analytics and the growing trend toward high performance computing architectures, the book is a detail-driven look into how big data analytics can be leveraged to foster positive change and drive efficiency. With continued exponential growth in data and ever more competitive markets, businesses must adapt quickly to gain every competitive advantage available. Big data analytics can serve as the linchpin for initiatives that drive business, but only if the underlying technology and analysis is fully understood and appreciated by engaged stakeholders. This book provides a view into the topic that executives, managers, and practitioners require, and includes: A complete overview of big data and its notable characteristics Details on high performance computing architectures for analytics, massively parallel processing (MPP), and in-memory databases Comprehensive coverage of data mining, text analytics, and machine learning algorithms A discussion of explanatory and predictive modeling, and how they can be applied to decision-making processes *Big Data, Data Mining, and Machine Learning* provides technology and marketing executives with the complete resource that has been notably absent from the veritable libraries of published books on the topic. Take control of your organization's big data analytics to produce real results with a resource that is comprehensive in scope and light on hyperbole.

A comprehensive guide to well-known workflow patterns: recurrent, generic business process constructs, described from the control-flow, data, and resource perspectives. The study of business processes has emerged as a highly effective approach to coordinating an organization's complex service- and knowledge-based activities. The growing field of business process management (BPM) focuses on methods and tools for designing, enacting, and analyzing business processes. This volume offers a definitive guide to the use of patterns, which synthesize the wide range of approaches to modeling business processes. It provides a unique and comprehensive introduction to the well-known workflow patterns collection—recurrent, generic constructs describing common business process modeling and execution scenarios, presented in the form of problem-solution dialectics. The underlying principles of the patterns approach ensure that they are independent of any specific enabling technology, representational formalism, or modeling approach, and thus broadly applicable across the business process modeling and business process technology domains. The authors, drawing on extensive research done by the Workflow Patterns Initiative, offer a detailed introduction to the fundamentals of business process modeling and management; describe three major pattern catalogs, presented from control-flow, data, and resource perspectives; and survey related BPM patterns. The book, a companion to the authoritative Workflow Patterns website, will be an essential resource for both academics and practitioners working in business process modeling and business process management.

Process Mining Data Science in Action Springer

Data Preprocessing for Data Mining addresses one of the most important issues within the well-known Knowledge Discovery from Data process. Data directly taken from the source will likely have inconsistencies, errors or most importantly, it is not ready to be considered for a data mining process. Furthermore, the increasing amount of data in recent science, industry and business applications, calls to the requirement of more complex tools to analyze it. Thanks to data preprocessing, it is possible to convert the impossible into possible, adapting the data to fulfill the input demands of each data mining algorithm. Data preprocessing includes the data reduction techniques, which aim at reducing the complexity of the data, detecting or removing irrelevant and noisy elements from the data. This book is intended to review the tasks that fill the gap between the data acquisition from the source and the data mining process. A comprehensive look from a practical point of view, including basic concepts and surveying the techniques proposed in the specialized literature, is given. Each chapter is a stand-alone guide to a particular data preprocessing topic, from basic concepts and detailed descriptions of classical algorithms, to an incursion of an exhaustive catalog of recent developments. The in-depth technical descriptions make this book suitable for technical professionals, researchers, senior undergraduate and graduate students in data science, computer science and engineering.

The main goal of this book is to explain the core ideas of process mining, and to demonstrate how they can be implemented using just some basic tools that are available to any computer scientist or data scientist. It describes how to analyze event logs in order to discover the behavior of real-world business processes. The end result can often be visualized as a graph, and the book explains how to use Python and Graphviz to render these graphs intuitively. Overall, it enables the reader to implement process mining techniques on his or her own, independently of any specific process mining tool. An introduction to two popular process mining tools, namely Disco and ProM, is also provided. The book will be especially valuable for self-study or as a precursor to a more advanced text. Practitioners and students will be able to follow along on their own, even if they have no prior knowledge of the topic. After reading this book, they will be able to more confidently proceed to the research literature if needed. After a brief presentation of the state of the art of process-mining techniques, Andrea Burratin proposes different scenarios for the deployment of process-mining projects, and in particular a characterization of companies in terms of their process awareness. The approaches proposed in this book belong to two different computational paradigms: first to classic "batch process mining," and second to more recent "online process mining." The book encompasses a revised version of the author's PhD thesis, which won the "Best Process Mining Dissertation Award" in 2014, awarded by the IEEE Task Force on Process Mining.

The Conference is aimed at giving the highest quality exposure to Engineering, Computer Science, Information System and Technology, Mathematics, Statistics and Actuarial Science & Risk Management. The event attracts practicing professionals, researchers, professors and University students from Pakistan and around the world.

Data Science and Analytics with Python is designed for practitioners in data science and data analytics in both academic and business environments. The aim is to present the reader with the main concepts used in data science using tools developed in Python, such as SciKit-learn, Pandas, Numpy, and others. The use of Python is of particular interest, given its recent popularity in the data science community. The book can be used by seasoned programmers and newcomers alike. The book is organized in a way that individual chapters are sufficiently independent from each other so that the reader is comfortable using the contents as a reference. The book discusses what data science and analytics are, from the point of view of the process and results obtained. Important features of Python are also covered, including a Python primer. The basic elements of machine learning, pattern recognition, and artificial intelligence that underpin the algorithms and implementations used in the rest of the book also appear in the first part of the book. Regression analysis using Python, clustering techniques, and classification algorithms are covered in the second part of the book. Hierarchical clustering, decision trees, and ensemble techniques are also explored, along with dimensionality reduction techniques and recommendation systems. The support vector machine algorithm and the Kernel trick are discussed in the last part of the book. About the Author Dr. Jesús Rogel-Salazar is a Lead Data scientist with experience in the field working for companies such as AKQA, IBM Data Science Studio, Dow Jones and others. He is a visiting researcher at the Department of Physics at Imperial College London, UK and a member of the School of Physics, Astronomy and Mathematics at the University of Hertfordshire, UK. He obtained his doctorate in physics at Imperial College London for work on quantum atom optics and ultra-cold matter. He has held a position as senior lecturer in mathematics as well as a consultant in the financial industry since 2006. He is the author of the book Essential Matlab and Octave, also published by CRC Press. His interests include mathematical modelling, data science, and optimization in a wide range of applications including optics, quantum mechanics, data journalism, and finance.

Business process management (BPM) constitutes one of the most exciting - search areas in computer science and the BPM Conference together with its workshops provides a distinct platform for presenting the latest research and showing future directions in this area. These proceedings contain the final versions of papers accepted for the workshops held in conjunction with the 7th International Conference on Business Process Management (BPM 2009). The BPM 2009 conference and workshops took place in Ulm, Germany. We received many interesting workshop proposals, eight of which were selected. Ultimately the workshops ran on September 7, 2009 featuring highly interesting keynotes, inspiring scientific presentations, and fruitful discussions. The history of five years of BPM workshops in a row proves the continued success of the workshop program. The workshop held in 2009 included one new workshop on empirical research in business process management and seven well-established workshops. First International Workshop on Empirical Research in Business Process Management (ER-BPM 2009). The ER-BPM 2009 workshop addressed the demand for empirical research methods such as experimental or case studies to BPM and invited fellow colleagues to investigate both the potential and the limitations of BPM methods and technologies in practice. The ER-BPM workshop aimed at closing the gap in knowledge on process management and at discussing empirical research in the space of BPM and associated phenomena. 12th International Workshop on Reference Modeling (RefMod 2009). Although conceptual models have proven to be a useful means to support information systems engineering in the past few years, creating and especially maintaining conceptual models can be quite challenging and costly. Getting your models into production is the fundamental challenge of machine learning. MLOps offers a set of proven principles aimed at solving this problem in a reliable and automated way. This insightful guide takes you through what MLOps is (and how it differs from DevOps) and shows you how to put it into practice to operationalize your machine learning models. Current and aspiring machine learning engineers--or anyone familiar with data science and Python--will build a foundation in MLOps tools and methods (along with AutoML and monitoring and logging), then learn how to implement them in AWS, Microsoft Azure, and Google Cloud. The faster you deliver a machine learning system that works, the faster you can focus on the business problems you're trying to crack. This book gives you a head start. You'll discover how to: Apply DevOps best practices to machine learning Build production machine learning systems and maintain them Monitor, instrument, load-test, and operationalize machine learning systems Choose the correct MLOps tools for a given machine learning task Run machine learning models on a variety of platforms and devices, including mobile phones and specialized hardware Fundamentals of Data Science is designed for students, academicians and practitioners with a complete walkthrough right from the foundational groundwork required to outlining

all the concepts, techniques and tools required to understand Data Science. Data Science is an umbrella term for the non-traditional techniques and technologies that are required to collect, aggregate, process, and gain insights from massive datasets. This book offers all the processes, methodologies, various steps like data acquisition, pre-process, mining, prediction, and visualization tools for extracting insights from vast amounts of data by the use of various scientific methods, algorithms, and processes. Readers will learn the steps necessary to create the application with SQL, NoSQL, Python, R, Matlab, Octave and Tableau. This book provides a stepwise approach to building solutions to data science applications right from understanding the fundamentals, performing data analytics to writing source code. All the concepts are discussed in simple English to help the community to become Data Scientist without much pre-requisite knowledge. Features : Simple strategies for developing statistical models that analyze data and detect patterns, trends, and relationships in data sets. Complete roadmap to Data Science approach with dedicated sections which includes Fundamentals, Methodology and Tools. Focussed approach for learning and practice various Data Science Tools with Sample code and examples for practice. Information is presented in an accessible way for students, researchers and academicians and professionals.

The field of Business Process Management (BPM) is marred by a seemingly endless sequence of (proposed) industry standards. Contrary to other fields (e.g., civil or electronic engineering), these standards are not the result of a widely supported consolidation of well-understood and well-established concepts and practices. In the BPM domain, it is frequently the case that BPM vendors opportunistically become involved in the creation of proposed standards to exert or maintain their influence and interests in the field. Despite the initial fervor associated with such standardization activities, it is no less frequent that vendors either choose to drop their support for standards that they earlier championed on an opportunistic basis or elect only to partially support them in their commercial offerings. Moreover, the results of the standardization processes themselves are a concern. BPM standards tend to deal with complex concepts, yet they are never properly defined and all-too-often not informed by established research. The result is a plethora of languages and tools, with no consensus on concepts and their implementation. They also fail to provide clear direction in the way in which BPM standards should evolve. One can also observe a dichotomy between the "business" side of BPM and its "technical" side. While it is clear that the application of BPM will fail if not placed in a proper business context, it is equally clear that its application will go nowhere if it remains merely a motivational exercise with schemas of business processes hanging on the wall gathering dust. Due to recent theoretical findings and advances in statistical computing, there has been a rapid development of techniques and applications in the area of missing data analysis. *Statistical Methods for Handling Incomplete Data* covers the most up-to-date statistical theories and computational methods for analyzing incomplete data. Features: Uses the mean score equation as a building block for developing the theory for missing data analysis. Provides comprehensive coverage of computational techniques for missing data analysis. Presents a rigorous treatment of imputation techniques, including multiple imputation, fractional imputation. Explores the most recent advances of the propensity score method and estimation techniques for nonignorable missing data. Describes a survey sampling application. Updated with a new chapter on Data Integration. Now includes a chapter on Advanced Topics, including kernel ridge regression imputation and neural network model imputation. The book is primarily aimed at researchers and graduate students from statistics, and could be used as a reference by applied researchers with a good quantitative background. It includes many real data examples and simulated

examples to help readers understand the methodologies.

moderation of the workshops, and the publication process. Delve into your data for the key to success. Data mining is quickly becoming integral to creating value and business momentum. The ability to detect unseen patterns hidden in the numbers exhaustively generated by day-to-day operations allows savvy decision-makers to exploit every tool at their disposal in the pursuit of better business. By creating models and testing whether patterns hold up, it is possible to discover new intelligence that could change your business's entire paradigm for a more successful outcome. *Data Mining for Dummies* shows you why it doesn't take a data scientist to gain this advantage, and empowers average business people to start shaping a process relevant to their business's needs. In this book, you'll learn the hows and whys of mining to the depths of your data, and how to make the case for heavier investment into data mining capabilities. The book explains the details of the knowledge discovery process including: Model creation, validity testing, and interpretation. Effective communication of findings. Available tools, both paid and open-source. Data selection, transformation, and evaluation. *Data Mining for Dummies* takes you step-by-step through a real-world data-mining project using open-source tools that allow you to get immediate hands-on experience working with large amounts of data. You'll gain the confidence you need to start making data mining practices a routine part of your successful business. If you're serious about doing everything you can to push your company to the top, *Data Mining for Dummies* is your ticket to effective data mining. Now in its second edition, this book focuses on practical algorithms for mining data from even the largest datasets.

This is the second edition of Wil van der Aalst's seminal book on process mining, which now discusses the field also in the broader context of data science and big data approaches. It includes several additions and updates, e.g. on inductive mining techniques, the notion of alignments, a considerably expanded section on software tools and a completely new chapter of process mining in the large. It is self-contained, while at the same time covering the entire process-mining spectrum from process discovery to predictive analytics. After a general introduction to data science and process mining in Part I, Part II provides the basics of business process modeling and data mining necessary to understand the remainder of the book. Next, Part III focuses on process discovery as the most important process mining task, while Part IV moves beyond discovering the control flow of processes, highlighting conformance checking, and organizational and time perspectives. Part V offers a guide to successfully applying process mining in practice,

including an introduction to the widely used open-source tool ProM and several commercial products. Lastly, Part VI takes a step back, reflecting on the material presented and the key open challenges. Overall, this book provides a comprehensive overview of the state of the art in process mining. It is intended for business process analysts, business consultants, process managers, graduate students, and BPM researchers.

[Copyright: 70c6479ea5592b08e93fb99f85d96d7f](#)