

Data Mining With Ibm Spss Through Examples Format

This textbook explores the different aspects of data mining from the fundamentals to the complex data types and their applications, capturing the wide diversity of problem domains for data mining issues. It goes beyond the traditional focus on data mining problems to introduce advanced data types such as text, time series, discrete sequences, spatial data, graph data, and social networks. Until now, no single book has addressed all these topics in a comprehensive and integrated way. The chapters of this book fall into one of three categories: Fundamental chapters: Data mining has four main problems, which correspond to clustering, classification, association pattern mining, and outlier analysis. These chapters comprehensively discuss a wide variety of methods for these problems. Domain chapters: These chapters discuss the specific methods used for different domains of data such as text data, time-series data, sequence data, graph data, and spatial data. Application chapters: These chapters study important applications such as stream mining, Web mining, ranking, recommendations, social networks, and privacy preservation. The domain chapters also have an applied flavor. Appropriate for both introductory and advanced data mining courses, *Data Mining: The Textbook* balances mathematical details and intuition. It contains the necessary mathematical details for professors and researchers, but it is presented in a simple and intuitive style to improve accessibility for students and industrial practitioners (including those with a limited mathematical background). Numerous illustrations, examples, and exercises are included, with an emphasis on semantically interpretable examples. Praise for *Data Mining: The Textbook* - "As I read through this book, I have already decided to use it in my classes. This is a book written by an outstanding researcher who has made fundamental contributions to data mining, in a way that is both accessible and up to date. The book is complete with theory and practical use cases. It's a must-have for students and professors alike!" -- Qiang Yang, Chair of Computer Science and Engineering at Hong Kong University of Science and Technology "This is the most amazing and comprehensive text book on data mining. It covers not only the fundamental problems, such as clustering, classification, outliers and frequent patterns, and different data types, including text, time series, sequences, spatial data and graphs, but also various applications, such as recommenders, Web, social network and privacy. It is a great book for graduate students and researchers as well as practitioners." -- Philip S. Yu, UIC Distinguished Professor and Wexler Chair in Information Technology at University of Illinois at Chicago

Dive deeper into SPSS Statistics for more efficient, accurate, and sophisticated data analysis and visualization *SPSS Statistics for Data Analysis and Visualization* goes beyond the basics of SPSS Statistics to show you advanced techniques that exploit the full capabilities of SPSS. The authors explain when and why to use each technique, and then walk you through the execution with a pragmatic, nuts and bolts example. Coverage includes extensive, in-depth discussion of advanced statistical techniques, data visualization, predictive analytics, and SPSS programming, including automation and integration with other languages like R and Python. You'll learn the best methods to power through an analysis, with more efficient, elegant, and accurate code. IBM SPSS Statistics is complex: true mastery requires a deep understanding of statistical theory, the user interface, and programming. Most users don't encounter all of the methods SPSS offers, leaving many little-known modules undiscovered. This book walks you through tools you may have never noticed, and shows you how they can be used to streamline your workflow and enable you to produce more accurate results. Conduct a more efficient and accurate analysis Display complex relationships and create better visualizations Model complex interactions and master predictive analytics Integrate R and Python with SPSS Statistics for more efficient, more powerful code These "hidden tools" can help you produce charts that simply wouldn't be possible any other way, and the support for other programming languages gives you better options for solving complex problems. If you're ready to take advantage of everything this powerful software package has to offer, *SPSS Statistics for Data Analysis and Visualization* is the expert-led training you need.

This book presents the most common techniques used in data mining in a simple and easy to understand through one of the most common software solutions from among those existing in the market, in particular, IBM SPSS CLEMENTINE whose current name is IBM SPSS MODELER. Pursued as initial aim clarifying the applications concerning methods traditionally rated as difficult or dull. It seeks to present applications in data mining without having to manage high mathematical developments or complicated theoretical algorithms, which is the most common reason for the difficulties in understanding and implementation of this matter. Today data mining is used in different fields of science. Noteworthy applications in banking, and financial analysis of markets and trade, insurance and private health, in education, in industrial processes, in medicine, biology and bioengineering, telecommunications and in many other areas. Essentials to get started in data mining, regardless of the field in which it is applied, is the understanding of own concepts, task that does not require nor much less the domain of scientific apparatus involved in the matter. Later, when either necessary operative advanced, computer programs allow the results without having to decipher the mathematical development of the algorithms that are under the procedures. This book describes the simplest possible data mining concepts, so that they are understandable by readers with different training. The chapters begin describing the techniques in affordable language and then presenting the way to treat them through practical applications. An important part of each chapter are case studies completely resolved, including the interpretation of the results, which is precisely the most important thing in any matter with which they work. The book begins with an introduction to mining data and its phases. In successive chapters develop the initial phases (selection of information, data exploration, data cleansing, transformation of data, etc.). Subsequently elaborates on specific data mining, both predictive and descriptive techniques. Predictive techniques covers all models of regression, discriminant analysis, decision trees, neural networks and other techniques based on models. The descriptive techniques vary dimension reduction techniques, techniques of classification and segmentation (clustering), and exploratory data analysis techniques.

IBM SPSS Modeler is a set of data mining tools that enable you to quickly develop predictive models using business expertise and deploy them into business operations to improve decision making. Designed around the industry-standard CRISP-DM model, IBM SPSS Modeler supports the entire data mining process, from data to better business results. IBM SPSS Modeler offers a variety of modeling methods taken from machine learning, artificial intelligence, and statistics. The methods available on the Modeling palette allow you to derive new information from your data and to develop predictive models. Each method has certain strengths and is best suited for particular types of problems. SPSS Modeler can be purchased as a standalone product, or used as a client in combination with SPSS Modeler Server. A number of additional options are also available, as summarized in the following sections. SPSS Modeler is a functionally complete version of the product that you install and run on your personal computer. You can run SPSS Modeler in local mode as a standalone product, or use it in distributed mode along with IBM SPSS Modeler Server for improved performance on large data sets. With SPSS Modeler, you can build accurate predictive models quickly and intuitively, without programming. Using the unique visual interface, you can easily visualize the data mining process. With the support of the advanced analytics embedded in the product, you can discover previously hidden patterns and trends in your data. You can model outcomes and understand the factors that influence them, enabling you to take advantage of business opportunities and mitigate risks.

Payment fraud can be defined as an intentional deception or misrepresentation that is designed to result in an unauthorized benefit. Fraud schemes are becoming more complex and difficult to identify. It is estimated that industries lose nearly \$1 trillion USD annually because of fraud. The ideal solution is where you avoid making fraudulent payments without slowing down legitimate payments. This solution requires that you adopt a comprehensive fraud business architecture that applies predictive analytics. This IBM® Redbooks® publication begins with the business process flows of several industries, such as banking, property/casualty insurance, and tax revenue, where payment fraud is a significant problem. This book then shows how to incorporate technological advancements that help you move from a post-payment to pre-payment fraud detection architecture. Subsequent chapters describe a solution that is specific to the banking industry that can be easily extrapolated to other industries. This book describes the benefits of doing fraud detection on IBM System z®. This book is intended for financial decisionmakers, consultants, and architects, in addition to IT administrators.

This book presents the most common techniques used in data mining in a simple and easy to understand through one of the most common software solutions from among those existing in the market, in particular, IBM SPSS. Pursued as initial aim clarifying the applications concerning methods traditionally rated as difficult or dull. It seeks to present applications in data mining without having to manage high mathematical developments or complicated theoretical algorithms, which is the most common reason for the difficulties in understanding and implementation of this matter. Today data mining is used in different fields of science. Noteworthy applications in banking, and financial analysis of markets and trade, insurance and private health, in education, in industrial processes, in medicine, biology and bioengineering, telecommunications and in many other areas. Essentials to get started in data mining, regardless of the field in which it is applied, is the understanding of own concepts, task that does not require nor much less the domain of scientific apparatus involved in the matter. Later, when either necessary operative advanced, computer programs allow the results without having to decipher the mathematical development of the algorithms that are under the procedures. This book describes the simplest possible data mining concepts, so that they are understandable by readers with different training. The chapters begin describing the techniques in affordable language and then presenting the way to treat them through practical applications. An important part of each chapter are case studies completely resolved, including the interpretation of the results, which is precisely the most important thing in any matter with which they work. The book begins with an introduction to mining data and its phases. In successive chapters develop the initial phases (selection of information, data exploration, data cleansing, transformation of data, etc.). Subsequently elaborates on specific data mining, both predictive and descriptive techniques. Predictive techniques covers all models of regression, discriminant analysis, decision trees, neural networks and other techniques based on models. The descriptive techniques vary dimension reduction techniques, techniques of classification and segmentation (clustering), and exploratory data analysis techniques.

The fun and friendly guide to mastering IBM's Statistical Package for the Social Sciences Written by an author team with a combined 55 years of experience using SPSS, this updated guide takes the guesswork out of the subject and helps you get the most out of using the leader in predictive analysis. Covering the latest release and updates to SPSS 27.0, and including more than 150 pages of basic statistical theory, it helps you understand the mechanics behind the calculations, perform predictive analysis, produce informative graphs, and more. You'll even dabble in programming as you expand SPSS functionality to suit your specific needs. Master the fundamental mechanics of SPSS Learn how to get data into and out of the program Graph and analyze your data more accurately and efficiently Program SPSS with Command Syntax Get ready to start handling data like a pro—with step-by-step instruction and expert advice!

This is a practical cookbook with intermediate-advanced recipes for SPSS Modeler data analysts. It is loaded with step-by-step examples explaining the process followed by the experts. If you have had some hands-on experience with IBM SPSS Modeler and now want to go deeper and take more control over your data mining process, this is the guide for you. It is ideal for practitioners who want to break into advanced analytics.

Modern statistical software provides the ability to compute statistics in a timely, orderly fashion. This introductory statistics textbook presents clear explanations of basic statistical concepts and introduces students to the IBM SPSS program to demonstrate how to conduct statistical analyses via the popular point-and-click and the "syntax file" methods. The focal point is to show students how easy it is to analyse data using SPSS once they have learned the basics. Provides clear explanation of basic statistical concepts

that provides the foundation for the beginner students' statistical journey. Introduces the SPSS software program. Gives clear explanation of the purpose of specific statistical procedures (e.g., frequency distributions, measures of central tendencies, measures of variability, etc.). Avoids the conventional cookbook approach that contributes very little to students' understanding of the rationale of how the correct results were obtained. The advantage of learning the IBM SPSS software package at the introductory class level is that most social sciences students will employ this program in their later years of study. This is because SPSS is one of the most popular of the many statistical packages currently available. Learning how to use this program at the very start not only familiarizes students with the utility of this program but also provides them with the experience to employ the program to conduct more complex analyses in their later years.

This latest edition has been fully updated to accommodate the needs of users of SPSS Releases 17, 18 and 19 while still being applicable to users of SPSS Releases 15 and 16. As with previous editions, Alan Bryman and Duncan Cramer continue to offer a comprehensive and user-friendly introduction to the widely used IBM SPSS Statistics. The simple, non-technical approach to quantitative data analysis enables the reader to quickly become familiar with SPSS and with the tests available to them. No previous experience of statistics or computing is required as this book provides a step-by-step guide to statistical techniques, including: Non-parametric tests Correlation Simple and multiple regression Analysis of variance and covariance Factor analysis. This book comes equipped with a comprehensive range of exercises for further practice, and it covers key issues such as sampling, statistical inference, conceptualization and measurement and selection of appropriate tests. The authors have also included a helpful glossary of key terms. The data sets used in Quantitative Data Analysis with IBM SPSS 17, 18 and 19 are available online at http://www.routledgetextbooks.com/textbooks/_author/bryman-9780415579193/; in addition, a set of multiple-choice questions and a chapter-by-chapter PowerPoint lecture course are available free of charge to lecturers who adopt the book. Using the same accessible, hands-on approach as its best-selling predecessor, the Handbook of Univariate and Multivariate Data Analysis with IBM SPSS, Second Edition explains how to apply statistical tests to experimental findings, identify the assumptions underlying the tests, and interpret the findings. This second edition now covers more topics

The world contains an unimaginably vast amount of digital information which is getting ever vaster ever more rapidly. This makes it possible to do many things that previously could not be done: spot business trends, prevent diseases, combat crime and so on. Managed well, the textual data can be used to unlock new sources of economic value, provide fresh insights into science and hold governments to account. As the Internet expands and our natural capacity to process the unstructured text that it contains diminishes, the value of text mining for information retrieval and search will increase dramatically. This comprehensive professional reference brings together all the information, tools and methods a professional will need to efficiently use text mining applications and statistical analysis. The Handbook of Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications presents a comprehensive how-to reference that shows the user how to conduct text mining and statistically analyze results. In addition to providing an in-depth examination of core text mining and link detection tools, methods and operations, the book examines advanced preprocessing techniques, knowledge representation considerations, and visualization approaches. Finally, the book explores current real-world, mission-critical applications of text mining and link detection using real world example tutorials in such varied fields as corporate, finance, business intelligence, genomics research, and counterterrorism activities. -Extensive case studies, most in a tutorial format, allow the reader to 'click through' the example using a software program, thus learning to conduct text mining analyses in the most rapid manner of learning possible -Numerous examples, tutorials, power points and datasets available via companion website on Elsevierdirect.com -Glossary of text mining terms provided in the appendix

Introducing the IBM SPSS Modeler, this book guides readers through data mining processes and presents relevant statistical methods. There is a special focus on step-by-step tutorials and well-documented examples that help demystify complex mathematical algorithms and computer programs. The variety of exercises and solutions as well as an accompanying website with data sets and SPSS Modeler streams are particularly valuable. While intended for students, the simplicity of the Modeler makes the book useful for anyone wishing to learn about basic and more advanced data mining, and put this knowledge into practice. A step-by-step guide to data mining applications in CRM. Following a handbook approach, this book bridges the gap between analytics and their use in everyday marketing, providing guidance on solving real business problems using data mining techniques. The book is organized into three parts. Part one provides a methodological roadmap, covering both the business and the technical aspects. The data mining process is presented in detail along with specific guidelines for the development of optimized acquisition, cross/ deep/ up selling and retention campaigns, as well as effective customer segmentation schemes. In part two, some of the most useful data mining algorithms are explained in a simple and comprehensive way for business users with no technical expertise. Part three is packed with real world case studies which employ the use of three leading data mining tools: IBM SPSS Modeler, RapidMiner and Data Mining for Excel. Case studies from industries including banking, retail and telecommunications are presented in detail so as to serve as templates for developing similar applications. Key Features: Includes numerous real-world case studies which are presented step by step, demystifying the usage of data mining models and clarifying all the methodological issues. Topics are presented with the use of three leading data mining tools: IBM SPSS Modeler, RapidMiner and Data Mining for Excel. Accompanied by a website featuring material from each case study, including datasets and relevant code. Combining data mining and business knowledge, this practical book provides all the necessary information for designing, setting up, executing and deploying data mining techniques in CRM. Effective CRM using Predictive Analytics will benefit data mining practitioners and consultants, data analysts, statisticians, and CRM officers. The book will also be useful to academics and students interested in applied data mining.

La minería de datos o Data Mining puede definirse inicialmente como un proceso de descubrimiento de nuevas y significativas relaciones, patrones y tendencias al examinar grandes cantidades de datos. La disponibilidad de grandes volúmenes de información y el uso generalizado de herramientas informáticas ha transformado el análisis de datos orientándolo hacia determinadas técnicas especializadas englobadas bajo el nombre de minería de datos o Data Mining. Las técnicas de minería de datos persiguen el descubrimiento automático del conocimiento contenido en la información almacenada de modo ordenado en grandes bases de datos. Estas técnicas tienen como objetivo descubrir patrones, perfiles y tendencias a través del análisis de los datos utilizando técnicas avanzadas como muestreo, análisis exploratorio de datos, técnicas de reducción de la dimensión, técnicas de modelización avanzada, clasificación, segmentación, predicción, reconocimiento de patrones y otras técnicas avanzadas de análisis de datos. Este libro trata la mayoría de estas técnicas desde el punto de vista práctico utilizando el software

IBM SPSS MODELER (IBM SPSS CLEMENTINE), uno de los más adecuados del mercado para estas tareas.

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

La minería de datos o Data Mining puede definirse inicialmente como un proceso de descubrimiento de nuevas y significativas relaciones, patrones y tendencias al examinar grandes cantidades de datos. La disponibilidad de grandes volúmenes de información y el uso generalizado de herramientas informáticas ha transformado el análisis de datos orientándolo hacia determinadas técnicas especializadas englobadas bajo el nombre de minería de datos o Data Mining. Las técnicas de minería de datos persiguen el descubrimiento automático del conocimiento contenido en la información almacenada de modo ordenado en grandes bases de datos. Estas técnicas tienen como objetivo descubrir patrones, perfiles y tendencias a través del análisis de los datos utilizando técnicas avanzadas como muestreo, análisis exploratorio de datos, técnicas de reducción de la dimensión, técnicas de modelización avanzada, clasificación, segmentación, predicción, reconocimiento de patrones y otras técnicas avanzadas de análisis de datos. Este libro trata la mayoría de estas técnicas desde el punto de vista práctico utilizando el software IBM SPSS, uno de los más adecuados del mercado para estas tareas.

Get to grips with the fundamentals of data mining and predictive analytics with IBM SPSS Modeler Key Features Get up—and-running with IBM SPSS Modeler without going into too much depth. Identify interesting relationships within your data and build effective data mining and predictive analytics solutions A quick, easy—to-follow guide to give you a fundamental understanding of SPSS Modeler, written by the best in the business Book Description IBM SPSS Modeler allows users to quickly and efficiently use predictive analytics and gain insights from your data. With almost 25 years of history, Modeler is the most established and comprehensive Data Mining workbench available. Since it is popular in corporate settings, widely available in university settings, and highly compatible with all the latest technologies, it is the perfect way to start your Data Science and Machine Learning journey. This book takes a detailed, step-by-step approach to introducing data mining using the de facto standard process, CRISP-DM, and Modeler's easy to learn "visual programming" style. You will learn how to read data into Modeler, assess data quality, prepare your data for modeling, find interesting patterns and relationships within your data, and export your predictions. Using a single case study throughout, this intentionally short and focused book sticks to the essentials. The authors have drawn upon their decades of teaching thousands of new users, to choose those aspects of Modeler that you should learn first, so that you get off to a good start using proven best practices. This book provides an overview of various popular data modeling techniques and presents a detailed case study of how to use CHAID, a decision tree model. Assessing a model's performance is as important as building it; this book will also show you how to do that. Finally, you will see how you can score new data and export your predictions. By the end of this book, you will have a firm understanding of the basics of data mining and how to effectively use Modeler to build predictive models. What you will learn Understand the basics of data mining and familiarize yourself with Modeler's visual programming interface Import data into Modeler and learn how to properly declare metadata Obtain summary statistics and audit the quality of your data Prepare data for modeling by selecting and sorting cases, identifying and removing duplicates, combining data files, and modifying and creating fields Assess simple relationships using various statistical and graphing techniques Get an overview of the different types of models available in Modeler Build a decision tree model and assess its results Score new data and export predictions Who this book is for This book is ideal for those who are new to SPSS Modeler and want to start using it as quickly as possible, without going into too much detail. An understanding of basic data mining concepts will be helpful, to get the best out of the book.

This book demonstrates how to use multilevel and longitudinal modeling techniques available in the IBM SPSS mixed-effects program (MIXED). Annotated screen shots provide readers with a step-by-step understanding of each technique and navigating the program. Readers learn how to set up, run, and interpret a variety of models. Diagnostic tools, data management issues, and related graphics are introduced throughout. Annotated syntax is also available for those who prefer this approach. Extended examples illustrate the logic of model development to show readers the rationale of the research questions and the steps around which the analyses are structured. The data used in the text and syntax examples are available at www.routledge.com/9780415817110. Highlights of the new edition include: Updated throughout to reflect IBM SPSS Version 21. Further coverage of growth trajectories, coding time-related variables, covariance structures, individual change and longitudinal experimental designs (Ch.5). Extended discussion of other types of research designs for examining change (e.g., regression discontinuity, quasi-experimental) over time (Ch.6). New examples specifying multiple latent constructs and parallel growth processes (Ch. 7). Discussion of alternatives for dealing with missing data and the use of sample weights within multilevel data structures (Ch.1). The book opens with the conceptual and methodological issues associated with multilevel and longitudinal modeling, followed by a discussion of SPSS data management techniques which facilitate working with multilevel, longitudinal, and cross-classified data sets. Chapters 3 and 4 introduce the basics of multilevel modeling: developing a multilevel model, interpreting output, and trouble-shooting common programming and modeling problems. Models for investigating individual and organizational change are presented in chapters 5 and 6, followed by models with multivariate outcomes in chapter 7. Chapter 8 provides an illustration of multilevel models with cross-classified data structures. The book concludes with ways to expand on the various multilevel and longitudinal modeling techniques and issues when conducting multilevel analyses. Ideal as a supplementary text for graduate courses on multilevel and longitudinal modeling, multivariate statistics, and research design taught in education, psychology, business, and sociology, this book's practical approach also appeals to researchers in these fields. The book provides an excellent supplement to Heck & Thomas's An Introduction to Multilevel Modeling Techniques, 2nd Edition; however, it

can also be used with any multilevel and/or longitudinal modeling book or as a stand-alone text.

This is an applied handbook for the application of data mining techniques in the CRM framework. It combines a technical and a business perspective to cover the needs of business users who are looking for a practical guide on data mining. It focuses on Customer Segmentation and presents guidelines for the development of actionable segmentation schemes. By using non-technical language it guides readers through all the phases of the data mining process.

Get efficient in performing data mining and machine learning using IBM SPSS Modeler Key Features Learn how to apply machine learning techniques in the field of data science Understand when to use different data mining techniques, how to set up different analyses, and how to interpret the results A step-by-step approach to improving model development and performance Book Description Machine learning (ML) combined with data mining can give you amazing results in your data mining work by empowering you with several ways to look at data. This book will help you improve your data mining techniques by using smart modeling techniques. This book will teach you how to implement ML algorithms and techniques in your data mining work. It will enable you to pair the best algorithms with the right tools and processes. You will learn how to identify patterns and make predictions with minimal human intervention. You will build different types of ML models, such as the neural network, the Support Vector Machines (SVMs), and the Decision tree. You will see how all of these models works and what kind of data in the dataset they are suited for. You will learn how to combine the results of different models in order to improve accuracy. Topics such as removing noise and handling errors will give you an added edge in model building and optimization. By the end of this book, you will be able to build predictive models and extract information of interest from the dataset What you will learn Hone your model-building skills and create the most accurate models Understand how predictive machine learning models work Prepare your data to acquire the best possible results Combine models in order to suit the requirements of different types of data Analyze single and multiple models and understand their combined results Derive worthwhile insights from your data using histograms and graphs Who this book is for If you are a data scientist, data analyst, and data mining professional and are keen to achieve a 30% higher salary by adding machine learning to your skillset, then this is the ideal book for you. You will learn to apply machine learning techniques to various data mining challenges. No prior knowledge of machine learning is assumed.

Data mining is the process of automatically searching large volumes of data for models and patterns using computational techniques from statistics, machine learning and information theory; it is the ideal tool for such an extraction of knowledge. Data mining is usually associated with a business or an organization's need to identify trends and profiles, allowing, for example, retailers to discover patterns on which to base marketing objectives. This book looks at both classical and recent techniques of data mining, such as clustering, discriminant analysis, logistic regression, generalized linear models, regularized regression, PLS regression, decision trees, neural networks, support vector machines, Vapnik theory, naive Bayesian classifier, ensemble learning and detection of association rules. They are discussed along with illustrative examples throughout the book to explain the theory of these methods, as well as their strengths and limitations. Key Features: Presents a comprehensive introduction to all techniques used in data mining and statistical learning, from classical to latest techniques. Starts from basic principles up to advanced concepts. Includes many step-by-step examples with the main software (R, SAS, IBM SPSS) as well as a thorough discussion and comparison of those software. Gives practical tips for data mining implementation to solve real world problems. Looks at a range of tools and applications, such as association rules, web mining and text mining, with a special focus on credit scoring. Supported by an accompanying website hosting datasets and user analysis. Statisticians and business intelligence analysts, students as well as computer science, biology, marketing and financial risk professionals in both commercial and government organizations across all business and industry sectors will benefit from this book.

Learn the art and science of predictive analytics — techniques that get results Predictive analytics is what translates big data into meaningful, usable business information. Written by a leading expert in the field, this guide examines the science of the underlying algorithms as well as the principles and best practices that govern the art of predictive analytics. It clearly explains the theory behind predictive analytics, teaches the methods, principles, and techniques for conducting predictive analytics projects, and offers tips and tricks that are essential for successful predictive modeling. Hands-on examples and case studies are included. The ability to successfully apply predictive analytics enables businesses to effectively interpret big data; essential for competition today This guide teaches not only the principles of predictive analytics, but also how to apply them to achieve real, pragmatic solutions Explains methods, principles, and techniques for conducting predictive analytics projects from start to finish Illustrates each technique with hands-on examples and includes as series of in-depth case studies that apply predictive analytics to common business scenarios A companion website provides all the data sets used to generate the examples as well as a free trial version of software Applied Predictive Analytics arms data and business analysts and business managers with the tools they need to interpret and capitalize on big data.

Uncovering and analyzing data associated with the current business environment is essential in maintaining a competitive edge. As such, making informed decisions based on this data is crucial to managers across industries. Integration of Data Mining in Business Intelligence Systems investigates the incorporation of data mining into business technologies used in the decision making process. Emphasizing cutting-edge research and relevant concepts in data discovery and analysis, this book is a comprehensive reference source for policymakers, academicians, researchers, students, technology developers, and professionals interested in the application of data mining techniques and practices in business information systems.

This IBM® Redpaper™ publication presents the process and steps that were taken to move from an R language forecasting solution to an IBM SPSS® Modeler solution. The paper identifies the key challenges that the team faced and the lessons they learned. It describes the journey from analysis through design to key actions that were taken during development to make the conversion successful. The solution approach is described in detail so that you can learn how the team broke the original R solution architecture into logical components in order to plan for the conversion project. You see key aspects of the conversion from R to IBM SPSS Modeler and how basic parts, such as data preparation, verification, pre-screening, and automating data quality checks, are accomplished. The paper consists of three chapters: Chapter 1 introduces the business background and the problem domain. Chapter 2 explains critical technical challenges that the team confronted and solved. Chapter 3 focuses on lessons that were learned during this process and ideas that might apply to your conversion project. This paper applies to various audiences: Decision makers and IT Architects who focus on the architecture, roadmap, software platform, and total cost of ownership. Solution development team members who are involved in creating statistical/analytics-based solutions and who are familiar with R and IBM SPSS Modeler.

Each chapter of Performing Data Analysis Using IBM SPSS covers a particular statistical procedure and offers the following: an example problem or analysis goal, together with a data set; IBM SPSS analysis with step-by-step analysis setup and accompanying screen shots; and

IBM SPSS output with screen shots and narrative on how to read or interpret the results of the analysis.

Now in its second edition, this textbook introduces readers to the IBM SPSS Modeler and guides them through data mining processes and relevant statistical methods. Focusing on step-by-step tutorials and well-documented examples that help demystify complex mathematical algorithms and computer programs, it also features a variety of exercises and solutions, as well as an accompanying website with data sets and SPSS Modeler streams. While intended for students, the simplicity of the Modeler makes the book useful for anyone wishing to learn about basic and more advanced data mining, and put this knowledge into practice. This revised and updated second edition includes a new chapter on imbalanced data and resampling techniques as well as an extensive case study on the cross-industry standard process for data mining.

This book describes the CRISP-DM modelling process for data mining. SPSS (then ISL and now IBM-SPSS) had been providing Data Mining based services since 1990 and had launched the first commercial Data Mining tool Clementine in 1994. Now the principal product that implements the CRISP-DM Methodology is IBM SPSS Modeler. This book describes also the Control Language for Expression Manipulation (CLEM), which is a powerful tool used to analyze and manipulate the data used in IBM SPSS Modeler streams. You can use CLEM within nodes to perform tasks ranging from evaluating conditions or deriving values to inserting data into reports. CLEM expressions consist of values, field names, operators, and functions. Using the correct syntax, you can create a wide variety of powerful data operations. IBM SPSS Modeler is an integrated data mining tool that includes several data sources (ASCII, XLS, ODBC, etc.), a visual interface based on data processes/flows (streams), different data mining tools (correlation, association rules, regression, segmentation, classification, neural networks, decision rules and trees, etc.), data manipulation (pick & mix, sampling, combination and separation, etc.), model combination, data visualisation, model export to different languages (C, SPSS, SAS, etc.), integrated data export to other programs (XLS) and report generation.

This IBM® Redpaper™ publication updated technical overview provides essential details about the data processing steps, message flows, and analytical models that power IBM Predictive Maintenance and Quality (PMQ) Version 2.0. The new version of PMQ builds on the first one, released in 2013, to help companies efficiently monitor and maintain production assets and improve their overall availability, utilization, and performance. It analyzes various types of data to detect failure patterns and poor quality parts earlier than traditional quality control methods, with the goal of reducing unscheduled asset downtime and improving quality metrics. Version 2.0 includes an improved method of interacting with the solution's analytic data store using an API from the new Analytics Solution Foundation, a reusable, configurable, and extensible component that supports a number of the solution's analytic functions. The new version also changes the calculation of profiles and KPIs, which is now done using orchestrations that are defined in XML. This updated technical overview provides details about these new orchestration definitions.

Learn methods of data analysis and their application to real-world data sets This updated second edition serves as an introduction to data mining methods and models, including association rules, clustering, neural networks, logistic regression, and multivariate analysis. The authors apply a unified "white box" approach to data mining methods and models. This approach is designed to walk readers through the operations and nuances of the various methods, using small data sets, so readers can gain an insight into the inner workings of the method under review. Chapters provide readers with hands-on analysis problems, representing an opportunity for readers to apply their newly-acquired data mining expertise to solving real problems using large, real-world data sets. Data Mining and Predictive Analytics: Offers comprehensive coverage of association rules, clustering, neural networks, logistic regression, multivariate analysis, and R statistical programming language Features over 750 chapter exercises, allowing readers to assess their understanding of the new material Provides a detailed case study that brings together the lessons learned in the book Includes access to the companion website, www.dataminingconsultant.com, with exclusive password-protected instructor content Data Mining and Predictive Analytics will appeal to computer science and statistic students, as well as students in MBA programs, and chief executives.

Data mining is the art and science of intelligent data analysis. By building knowledge from information, data mining adds considerable value to the ever increasing stores of electronic data that abound today. In performing data mining many decisions need to be made regarding the choice of methodology, the choice of data, the choice of tools, and the choice of algorithms. Throughout this book the reader is introduced to the basic concepts and some of the more popular algorithms of data mining. With a focus on the hands-on end-to-end process for data mining, Williams guides the reader through various capabilities of the easy to use, free, and open source Rattle Data Mining Software built on the sophisticated R Statistical Software. The focus on doing data mining rather than just reading about data mining is refreshing. The book covers data understanding, data preparation, data refinement, model building, model evaluation, and practical deployment. The reader will learn to rapidly deliver a data mining project using software easily installed for free from the Internet. Coupling Rattle with R delivers a very sophisticated data mining environment with all the power, and more, of the many commercial offerings.

Designed to help students analyze and interpret research data using IBM SPSS, this user-friendly book, written in easy-to-understand language, shows readers how to choose the appropriate statistic based on the design, and to interpret outputs appropriately. The authors prepare readers for all of the steps in the research process: design, entering and checking data, testing assumptions, assessing reliability and validity, computing descriptive and inferential parametric and nonparametric statistics, and writing about outputs. Dialog windows and SPSS syntax, along with the output, are provided. Three realistic data sets, available on the Internet, are used to solve the chapter problems. The new edition features: Updated to IBM SPSS version 20 but the book can also be used with older and newer versions of SPSS. A new chapter (7) including an introduction to Cronbach's alpha and factor analysis. Updated Web Resources with PowerPoint slides, additional activities/suggestions, and the answers to even-numbered interpretation questions for the instructors, and chapter study guides and outlines and extra SPSS problems for the students. The web resource is located www.routledge.com/9781848729827. Students, instructors, and individual purchasers can access the data files to accompany the book at www.routledge.com/9781848729827. IBM SPSS for Introductory Statistics, Fifth Edition provides helpful teaching tools: All of the key IBM SPSS windows needed to perform the analyses. Complete outputs with call-out boxes to highlight key points. Flowcharts and tables to help select appropriate statistics and interpret effect sizes. Interpretation sections and questions help students better understand and interpret the output. Assignments organized the way students proceed when they conduct a research project. Examples of how to write about outputs and make tables in APA format. Helpful appendices on how to get started with SPSS and write research questions. An ideal supplement for courses in either statistics, research methods, or any course in which SPSS is used, such as in departments of psychology, education, and other social and health sciences. This book is also appreciated by researchers interested in using SPSS for their data analysis.

"IBM SPSS Modeler enables you to explore data, identify important relationships that you can leverage, and build predictive models quickly, allowing your organization to base its decisions purely on the insights obtained from your data. With the help of this course, you'll follow the industry-standard data mining process, gaining new skills at each stage, from loading data to integrating results into everyday business practices. Get a handle on the most efficient ways of extracting data from your own sources, preparing it for exploration and modeling. You will be acquainted with the best methods for building models that will perform well in your workplace. Go beyond the basics and get the full power of your data mining workbench using IBM SPSS Modeler with this handy tutorial."--Resource description page.

This accessible, practice-oriented and compact text provides a hands-on introduction to market research. Using the market research process as a framework, it explains how to collect and describe data and presents the most important and frequently used quantitative analysis

techniques, such as ANOVA, regression analysis, factor analysis and cluster analysis. The book describes the theoretical choices a market researcher has to make with regard to each technique, discusses how these are converted into actions in IBM SPSS version 22 and how to interpret the output. Each chapter concludes with a case study that illustrates the process using real-world data. A comprehensive Web appendix includes additional analysis techniques, datasets, video files and case studies. Tags in the text allow readers to quickly access Web content with their mobile device. The new edition features: Stronger emphasis on the gathering and analysis of secondary data (e.g., internet and social networking data) New material on data description (e.g., outlier detection and missing value analysis) Improved use of educational elements such as learning objectives, keywords, self-assessment tests, case studies, and much more Streamlined and simplified coverage of the data analysis techniques with more rules-of-thumb Uses IBM SPSS version 22

Master the fundamentals of SPSS with this newly updated and instructive resource The newly and thoroughly revised Second Edition of SPSS Essentials delivers a comprehensive guide for students in the social sciences who wish to learn how to use the Statistical Package for the Social Sciences (SPSS) for the effective collection, management, and analysis of data. The accomplished researchers and authors provide readers with the practical nuts and bolts of SPSS usage and data entry, with a particular emphasis on managing and manipulating data. The book offers an introduction to SPSS, how to navigate it, and a discussion of how to understand the data the reader is working with. It also covers inferential statistics, including topics like hypothesis testing, one-sample Z-testing, T-testing, ANOVAs, correlations, and regression. Five unique appendices round out the text, providing readers with discussions of dealing with real-world data, troubleshooting, advanced data manipulations, and new workbook activities. SPSS Essentials offers a wide variety of features, including: A revised chapter order, designed to match the pacing and content of typical undergraduate statistics classes An explanation of when particular inferential statistics are appropriate for use, given the nature of the data being worked with Additional material on understanding your data sample, including discussions of SPSS output and how to find the most relevant information A companion website offering additional problem sets, complete with answers Perfect for undergraduate students of the social sciences who are just getting started with SPSS, SPSS Essentials also belongs on the bookshelves of advanced placement high school students and practitioners in social science who want to brush up on the fundamentals of this powerful and flexible software package.

Data Mining with SPSS Modeler Theory, Exercises and Solutions Springer Nature

IBM SPSS Statistics 26 Step by Step: A Simple Guide and Reference, sixteenth edition, takes a straightforward, step-by-step approach that makes SPSS software clear to beginners and experienced researchers alike. Extensive use of four-color screen shots, clear writing, and step-by-step boxes guide readers through the program. Output for each procedure is explained and illustrated, and every output term is defined. Exercises at the end of each chapter support students by providing additional opportunities to practice using SPSS. This book covers the basics of statistical analysis and addresses more advanced topics such as multi-dimensional scaling, factor analysis, discriminant analysis, measures of internal consistency, MANOVA (between- and within-subjects), cluster analysis, Log-linear models, logistic regression and a chapter describing residuals. Back matter includes a description of data files used in exercises, an exhaustive glossary, suggestions for further reading and a comprehensive index. IBM SPSS Statistics 26 Step by Step is distributed in 85 countries, has been an academic best seller through most of the earlier editions, and has proved invaluable aid to thousands of researchers and students. New to this edition: Screenshots, explanations, and step-by-step boxes have been fully updated to reflect SPSS 26 How to handle missing data has been revised and expanded and now includes a detailed explanation of how to create regression equations to replace missing data More explicit coverage of how to report APA style statistics; this primarily shows up in the Output sections of Chapters 6 through 16, though changes have been made throughout the text.

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.

[Copyright: a83de73fb5e0ba7bd681abca0a22a8ed](#)