

## Clustering And Data Mining In R Introduction

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 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, 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.

This Second Edition brings readers thoroughly up to date with the emerging field of text mining, the application of techniques of machine learning in conjunction with natural language processing, information extraction, and algebraic/mathematical approaches to computational information retrieval. The book explores a broad range of issues, ranging from the development of new learning approaches to the parallelization of existing algorithms. Authors highlight open research questions in document categorization, clustering, and trend detection. In addition, the book describes new application problems in areas such as email surveillance and anomaly detection.

This volume describes new methods with special emphasis on classification and cluster analysis. These methods are applied to problems in information retrieval, phylogeny, medical diagnosis, microarrays, and other active research areas.

Often considered more of an art than a science, books on clustering have been dominated by learning through example with techniques chosen almost through trial and error. Even the two most popular, and most related, clustering methods-K-Means for partitioning and Ward’s method for hierarchical clustering-have lacked the theoretical underpinning req

A Study of Data Mining: Clustering Applications for Large, Multi-dimensional Data Sets

Grouping Multidimensional Data

Recent Advances in Clustering

Proceedings of the Meeting of the International Federation of Classification Societies (IFCS), Illinois Institute of Technology, Chicago, 15–18 July 2004

Constrained Clustering

Text Mining

There are many invaluable books available on data mining theory and applications. However, in compiling a volume titled “DATA MINING: Foundations and Intelligent Paradigms: Volume I: Clustering, Association and Classification” we wish to introduce some of the latest developments to a broad audience of both specialists and non-specialists in this field.

Data mining has witnessed substantial advances in recent decades. New research questions and practical challenges have arisen from emerging areas and applications within the various fields closely related to human daily life, e.g. social media and social networking. This book aims to bridge the gap between traditional data mining and the latest advances in newly emerging information services. It explores the extension of well-studied algorithms and approaches into these new research arenas.

As technology and technological advancements become a more prevalent and essential aspect of daily and business life, educational institutions must keep pace in order to maintain relevance and retain their ability to adequately prepare students for their lives beyond education. Such institutions and their leaders are seeking relevant strategies for the implementation and effective use of new and upcoming technologies and leadership strategies to best serve students and educators within educational settings. As traditional education methods become more outdated, strategies to supplement and bolster them through technology and effective management become essential to the success of institutions and programs. The Handbook of Research on Modern Educational Technologies, Applications, and Management is an all-encompassing two-volume scholarly reference comprised of 58 original and previously unpublished research articles that provide cutting-edge, multidisciplinary research and expert insights on advancing technologies used in educational settings as well as current strategies for administrative and leadership roles in education. Covering a wide range of topics including but not limited to community engagement, educational games, data management, and mobile learning, this publication provides insights into technological advancements with educational applications and examines forthcoming implementation strategies. These strategies are ideal for teachers, instructional designers, curriculum developers, educational software developers, and information technology specialists looking to promote effective learning in the classroom through cutting-edge learning technologies, new learning theories, and successful leadership tactics. Administrators, educational leaders, educational policymakers, and other education professionals will also benefit from this publication by utilizing the extensive research on managing educational institutions and providing valuable training and professional development initiatives as well as implementing the latest administrative technologies. Additionally, academicians, researchers, and students in areas that include but are not limited to educational technology, academic leadership, mentorship, learning environments, and educational support systems will benefit from the extensive research compiled within this publication.

Clustering is an essential reference source that synthesizes the analytic principles of clustering and machine learning to big data and provides an interface between the main disciplines of engineering/technology and the organizational, administrative, and planning abilities of management. Featuring research on topics such as project management, contextual data modeling, and business information systems, this book is ideally designed for engineers, economists, finance officers, marketers, decision makers, business professionals, industry practitioners, academicians, students, and researchers seeking coverage on the implementation of big data and machine learning within specific professional fields.

Advances in Clustering: Concepts and Techniques. Application areas include pattern recognition, artificial intelligence, information technology, image processing, biology, psychology, and marketing. Readers also learn how to perform cluster analysis with the C/C++ and MATLAB programming languages.

Data Mining - a search for knowledge

Classification, Clustering, and Data Mining Applications

Survey of Text Mining II

Cluster Analysis and Data Mining

Temporal Data Mining via Unsupervised Ensemble Learning

Data Mining: Concepts and Techniques

Shows how Galileo, Newton, and Einstein tried to explain gravity. Discusses the concept of microgravity and NASA’s research on gravity and microgravity.

Reference and compendium of algorithms for pattern recognition, data mining and statistical computing.

Cluster analysis is used in data mining and is a common technique for statistical data analysis used in many fields of study, such as the medical & life sciences, behavioral & social sciences, engineering, and in computer science. Designed for training industry professionals or for a course on clustering and classification, it can also be used as a companion text for applied statistics. No previous experience in clustering or data mining is assumed. Informal algorithms for clustering data and interpreting results are emphasized. In order to evaluate the results of clustering and to explore data, graphical methods and data structures are used for representing data. Throughout the text, examples and references are provided, in order to enable the material to be comprehensible for a diverse audience. A companion disc includes numerous appendices with programs, data, charts, solutions, etc. eBook Customers: Companion files are available for downloading with order number/proof of purchase by writing to the publisher at info@merclearning.com. FEATURES \*Places emphasis on illustrating the underlying logic in making decisions during the cluster analysis \*Discusses the related applications of statistic, e.g., Ward ’s method (ANOVA), JAN (regression analysis & correlational analysis), cluster validation (hypothesis testing, goodness-of-fit, Monte Carlo simulation, etc.) \*Contains separate chapters on JAN and the clustering of categorical data \*Includes a companion disc with solutions to exercises, programs, data sets, charts, etc.

As organizations continue to develop, there is an increasing need for technological methods that can keep up with the rising amount of data and information that is being generated. Machine learning is a tool that has become powerful due to its ability to analyze large amounts of data quickly. Machine learning is one of many technological advancements that is being implemented into a multitude of specialized fields. An extensive study on the execution of these advancements within professional industries is necessary. The Handbook of Research on Big Data Clustering and Machine Learning is an essential reference source that synthesizes the analytic principles of clustering and machine learning to big data and provides an interface between the main disciplines of engineering/technology and the organizational, administrative, and planning abilities of management. Featuring research on topics such as project management, contextual data modeling, and business information systems, this book is ideally designed for engineers, economists, finance officers, marketers, decision makers, business professionals, industry practitioners, academicians, students, and researchers seeking coverage on the implementation of big data and machine learning within specific professional fields.

Data Classification and Incremental Clustering in Data Mining and Machine Learning

Handbook of Research on Modern Educational Technologies, Applications, and Management

New and Classical Approaches

Algorithms and Applications

Data Mining and Knowledge Discovery Handbook

A Data Recovery Approach, Second Edition

Research on the problem of clustering tends to be fragmented across the pattern recognition, database, data mining, and machine learning communities. Addressing this problem in a unified way, Data Clustering: Algorithms and Applications provides complete coverage of the entire area of clustering, from basic methods to more refined and complex data clustering approaches. It pays special attention to recent issues in graphs, social networks, and other domains. The book focuses on three primary aspects of data clustering: Methods, describing key techniques commonly used for clustering, such as feature selection, agglomerative clustering, partitional clustering, density-based clustering, probabilistic clustering, grid-based clustering, spectral clustering, and nonnegative matrix factorization Domains, covering methods used for clustering of domain data, such as categorical data, text data, multimedia data, graph data, biological data, stream data, uncertain data, time series clustering, high-dimensional clustering, and big data Variations and Insights, discussing important variations of the clustering process, such as semisupervised clustering, interactive clustering, multiview clustering, cluster ensembles, and cluster validation In this book, top researchers from around the world explore the characteristics of clustering problems in a variety of application areas. They also explain how to glean detailed insight from the clustering process—including how to verify the quality of the underlying clusters—through supervision, human intervention, or the automated generation of alternative clusters.

The Definitive Resource on Text Mining Theory and Applications from Foremost Researchers in the Field Giving a broad perspective of the field from numerous vantage points, Text Mining: Classification, Clustering, and Applications focuses on statistical methods for text mining and analysis. It examines methods to automatically cluster and classify text documents and applies these methods in a variety of areas, including adaptive information filtering, information distillation, and text search. The book begins with chapters on the classification of documents into predefined categories. It presents state-of-the-art algorithms and their use in practice. The next chapters describe novel methods for clustering documents into groups that are not predefined. These methods seek to automatically determine topical structures that may exist in a document corpus. The book concludes by discussing various text mining applications that have significant implications for future research and industrial use. There is no doubt that text mining will continue to play a critical role in the development of future information systems and advances in research will be instrumental to their success. This book captures the technical depth and immense practical potential of text mining, guiding readers to a sound appreciation of this burgeoning field.

Clustering is the unsupervised classification of patterns into groups. In non-fuzzy or hard clustering, data is divided into crisp clusters, where each data point belongs to exactly one cluster. In fuzzy clustering, the data points can belong to more than one cluster, and associated with each of the points are membership grades which indicate the degree to which the data points belong to the different clusters. Different clustering algorithms produce clusters with different characteristics. In this book specifies the Clustering techniques using decision theory.

Nearly everyone knows K-means algorithm in the fields of data mining and business intelligence. But the ever-emerging data with extremely complicated characteristics bring new challenges to this "old" algorithm. This book addresses these challenges and makes novel contributions in establishing theoretical frameworks for K-means distances and K-means based consensus clustering, identifying the "dangerous" uniform effect and zero-value dilemma of K-means, adapting right measures for cluster validity, and integrating K-means with SVMs for rare class analysis. This book not only enriches the clustering and optimization theories, but also provides good guidance for the practical use of K-means, especially for important tasks such as network intrusion detection and credit fraud prediction. The thesis on which this book is based has won the "2019 National Excellent Doctoral Dissertation Award", the highest honor for not more than 100 PhD theses per year in China.

Data Analysis and Applications I

Data Mining and Predictive Analytics

Lecture Notes in Data Mining

Advances in Algorithms, Theory, and Applications

Fuzzy C-mean Clustering using Data Mining

Clustering for Data Mining

**Data clustering is an unsupervised classification method aims at creating groups of objects, or clusters, in such a way that objects in the same cluster are very similar and objects in different clusters are quite distinct. K-means is an iterative algorithm in which the number of clusters must be determined before the execution.In this book an efficient k-means algorithm is proposed. Since, in each iteration, the k-means algorithm computes the distances between data point and all centers, this is computationally very expensive especially for huge data sets. For each data point, we can keep the distance to the nearest cluster. At the next iteration, we compute the distance to the previous nearest cluster. If the new distance is less than or equal to the previous distance, the point stays in its cluster, and there is no need to compute its distances to the other cluster centers. This saves the time required to compute distances to k 1 clusters. Experimental results show the accuracy and effectiveness of the proposed method.**

**Data clustering is a highly interdisciplinary field, the goal of which is to divide a set of objects into homogeneous groups such that objects in the same group are similar and objects in different groups are quite distinct. Thousands of theoretical papers and a number of books on data clustering have been published over the past 50 years. However,**

**The recent explosive growth of our ability to generate and store data has created a need for new, scalable and efficient, tools for data analysis. The main focus of the discipline of knowledge discovery in databases is to address this need. Knowledge discovery in databases is the fusion of many areas that are concerned with different aspects of data handling and data analysis, including databases, machine learning, statistics, and algorithms. Each of these areas addresses a different part of the problem, and places different emphasis on different requirements. For example, database techniques are designed to efficiently handle relatively simple queries on large amounts of data stored in external (disk) storage. Machine learning techniques typically consider smaller data sets, and the emphasis is on the accuracy ofa relatively complicated analysis task such as classification. The analysis of large data sets requires the design of new tools that not only combine and generalize techniques from different areas, but also require the design and development ofaltogether new scalable techniques.**

**Since the initial work on constrained clustering, there have been numerous advances in methods, applications, and our understanding of the theoretical properties of constraints and constrained clustering algorithms. Bringing these developments together, Constrained Clustering: Advances in Algorithms, Theory, and Applications presents an extensive collection of the latest innovations in clustering data analysis methods that use background knowledge encoded as constraints. Algorithms The first five chapters of this volume investigate advances in the use of instance-level, pairwise constraints for partitional and hierarchical clustering. The book then explores other types of constraints for clustering, including cluster size balancing, minimum cluster size, and cluster-level relational constraints. Theory It also describes variations of the traditional clustering under constraints problem as well as approximation algorithms with helpful performance guarantees. Applications The book ends by applying clustering with constraints to relational data, privacy-preserving data publishing, and video surveillance data. It discusses an interactive visual clustering approach, a distance metric learning approach, existential constraints, and automatically generated constraints. With contributions from industrial researchers and leading academic experts who pioneered the field, this volume delivers thorough coverage of the capabilities and limitations of constrained clustering methods as well as introduces new types of constraints and clustering algorithms.**

Clustering Algorithms

A Data Mining Thinking

Theory, Algorithms, and Applications

Multiobjective Genetic Algorithms for Clustering

Applications in Data Mining and Bioinformatics

Data Mining and Machine Learning Applications

**DATA MINING AND MACHINE LEARNING APPLICATIONS The book elaborates in detail on the current needs of data mining and machine learning and promotes mutual understanding among research in different disciplines, thus facilitating research development and collaboration. Data, the latest currency of today’s world, is the new gold. In this new form of gold, the most beautiful jewels are data analytics and machine learning. Data mining and machine learning are considered interdisciplinary fields. Data mining is a subset of data analytics and machine learning involves the use of algorithms that automatically improve through experience based on data. Massive datasets can be classified and clustered to obtain accurate results. The most common technologies used include classification and clustering methods. Accuracy and error rates are calculated for regression and classification and clustering to find actual results through algorithms like support vector machines and neural networks with forward and backward propagation. Applications include fraud detection, image processing, medical diagnosis, weather prediction, e-commerce and so forth. The book features: A review of the state-of-the-art in data mining and machine learning, A review and description of the learning methods in human-computer interaction, Implementation strategies and future research directions used to meet the design and application requirements of several modern and real-time applications for a long time, The scope and implementation of a majority of data mining and machine learning strategies. A discussion of real-time problems. Audience Industry and academic researchers, scientists, and engineers in information technology, data science and machine and deep learning, as well as artificial intelligence more broadly.**

**This volume contains articles accepted for presentation during The Intelligent Information Systems Symposium IIS’2002 which was held in Sopot, Poland, on June 3–6, 2002. This is eleventh, in the order, symposium organized by the Institute of Computer Science of Polish Academy of Sciences and devoted to new trends in (broadly understood) ArtificialIntelligence. The meetings started back to 1992. With small initial audience, workshops in the series grew to an important meeting of Polish and foreign scientists working at the universities in Europe, Asia and the Northern America. Over years, the workshops transformed into regular symposia devoted to latest trends in such fields like Machine Learning, Knowledge Discovery, Natural Language Processing, Knowledge Based Systems and Reasoning, and Soft Computing (i.e. Fuzzy and Rough sets, Bayesian Networks, and Evolutionary Algorithms). At present, about 50-60 papers are accepted each year. Besides, for several years now, the symposia are accompanied by a number of tutorials, given by the outstanding scientists in their domain. The main topics of this year symposium included: • decision trees and other classifier systems • neural network and biologically motivated systems • clustering methods • handling imprecision and uncertainty • deductive, distributed and agent-based systems We were pleased to see the continuation of the last year trend towards an increase in the number of co-operative contributions and in the number and diversity of practical applications of theoretical research.**

**The amount and the complexity of the data gathered by current enterprises are increasing at an exponential rate. Consequently, the analysis of Big Data is nowadays a central challenge in Computer Science, especially for complex data. For example, given a satellite image database containing tens of Terabytes, how can we find regions aiming at identifying native rainforests, deforestation or reforestation? Can it be made automatically? Based on the work discussed in this book, the answers to both questions are a sound “yes”, and the results can be obtained in just minutes. In fact, results that used to require days or weeks of hard work from human specialists can now be obtained in minutes with high precision. Data Mining in Large Sets of Complex Data discusses new algorithms that take steps forward from traditional data mining (especially for clustering) by considering large, complex datasets. Usually, other works focus in one aspect, either data size or complexity. This work considers both: it enables mining complex data from high impact applications, such as breast cancer diagnosis, region classification in satellite images, assistance to climate change forecast, recommendation systems for the Web and social networks; the data are large in the Terabyte-scale, not in Giga as usual; and very accurate results are found in just minutes. Thus, it provides a crucial and well timed contribution for allowing the creation of real time applications that deal with Big Data of high complexity in which mining on the fly can make an immeasurable difference, such as supporting cancer diagnosis or detecting deforestation.**

**Often considered more as an art than a science, the field of clustering has been dominated by learning through examples and by techniques chosen almost through trial-and-error. Even the most popular clustering methods—K-Means for partitioning the data set and Ward’s method for hierarchical clustering—have lacked the theoretical attention that wou**

Introduction to Data Mining

Data Clustering in C++

Volume I: Clustering, Association and Classification

A Data Recovery Approach

Data Mining: Foundations and Intelligent Paradigms

Rough Fuzzy Clustering Using Decision Theory

**This is the first book primarily dedicated to clustering using multiobjective genetic algorithms with extensive real-life applications in data mining and bioinformatics. The authors first offer detailed introductions to the relevant techniques – genetic algorithms, multiobjective optimization, soft computing, data mining and bioinformatics. They then demonstrate systematic applications of these techniques to real-world problems in the areas of data mining, bioinformatics and geoscience. The authors offer detailed theoretical and statistical notes, guides to future research, and chapter summaries. The book can be used as a textbook and as a reference book by graduate students and academic and industrial researchers in the areas of soft computing, data mining, bioinformatics and geoscience.**

**The aim of this book is to illustrate that advanced fuzzy clustering algorithms can be used not only for partitioning of the data. It can also be used for visualization, regression, classification and time-series analysis, hence fuzzy cluster analysis is a good approach to solve complex data mining and system identification problems. This book is oriented to undergraduate and postgraduate and is well suited for teaching purposes.**

**This series of books collects a diverse array of work that provides the reader with theoretical and applied information on data analysis methods, models, and techniques, along with appropriate applications. Volume 1 begins with an introductory chapter by Gilbert Saporta, a leading expert in the field, who summarizes the developments in data analysis over the last 50 years. The book is then divided into three parts: Part 1 presents clustering and regression cases; Part 2 examines grouping and decomposition, GARCH and threshold models, structural equations, and SME modeling; and Part 3 presents symbolic data analysis, time series and multiple choice models, modeling in demography, and data mining.**

Publisher Description

Clustering, Classification, and Retrieval

Cluster Analysis for Data Mining and System Identification

Intelligent Information Systems 2002

Handbook of Research on Big Data Clustering and Machine Learning

Generalized Density Based Clustering for Spatial Data Mining

An Introduction

Data Mining: Concepts and Techniques provides the concepts and techniques in processing gathered data or information, which will be used in various applications. Specifically, it explains data mining and the tools used in discovering knowledge from the collected data. This book is referred as the knowledge discovery from data (KDD). It focuses on the feasibility, usefulness, effectiveness, and scalability of techniques of large data sets. After describing data mining, this edition explains the methods of knowing, preprocessing, processing, and warehousing data. It then presents information about data warehouses, online analytical processing (OLAP), and data cube technology. Then, the methods involved in mining frequent patterns, associations, and correlations for large data sets are described. The book details the methods for data classification and introduces the concepts and methods for data clustering. The remaining chapters discuss the outlier detection and the trends, applications, and research frontiers in data mining. This book is intended for Computer Science students, application developers, business professionals, and researchers who seek information on data mining. Presents dozens of algorithms and implementation examples, all in pseudo-code and suitable for use in real-world, large-scale data mining projects Addresses advanced topics such as mining object-relational databases, spatial databases, multimedia databases, time-series databases, text databases, the World Wide Web, and applications in several fields Provides a comprehensive, practical look at the concepts and techniques you need to get the most out of your data

The goal of traditional clustering is to assign each data point to one and only one cluster. In contrast, fuzzy clustering assigns different degrees of membership to each point. The membership of a point is thus shared among various clusters. This creates the concept of fuzzy boundaries which differs from the traditional concept of well-defined boundaries. In hard clustering, data is divided into distinct clusters, where each data element belongs to exactly one cluster. In fuzzy clustering (also referred to as soft clustering), data elements can belong to more than one cluster, and associated with each element is a set of membership levels. These indicate the strength of the association between that data element and a particular cluster. Fuzzy clustering is a process of assigning these membership levels, and then using them to assign data elements to one or more clusters. This algorithm uses the FCM traditional algorithm to locate the centers of clusters for a bulk of data points. The potential of all data points is being calculated with respect to specified centers. The availability of dividing the data set into large number of clusters will slow the processing time and needs more memory size for the program. Hence traditional clustering should device the data to four clusters and each data point should be located in one specified cluster. Imprecision in data and information gathered from and about our environment is either statistical(e.g., the outcome of a coin toss is a matter of chance) or no statistical (e.g., “apply the brakes pretty soon”). Many algorithms can be implemented to develop clustering of data sets. Fuzzy C-mean clustering (FCM) is efficient and common algorithm. We are tuning this algorithm to get a solution for the rest of data point which omitted because of its farness from all clusters. To develop a high performance algorithm that sort and group data set in variable number of clusters to use this data in control and managing of those clusters.

Temporal Data Mining via Unsupervised Ensemble Learning provides the principle knowledge of temporal data mining in association with unsupervised ensemble learning and the fundamental problems of temporal data clustering from different perspectives. By providing three proposed ensemble approaches of temporal data clustering, this book presents a practical focus of fundamental knowledge and techniques, along with a rich blend of theory and practice. Furthermore, the book includes illustrations of the proposed approaches based on data and simulation experiments to demonstrate all methodologies, and is a guide to the proper usage of these methods. As there is nothing universal that can solve all problems, it is important to understand the characteristics of both clustering algorithms and the target temporal data so the correct approach can be selected for a given clustering problem. Scientists, researchers, and data analysts working with machine learning and data mining will benefit from this innovative book, as will undergraduate and graduate students following courses in computer science, engineering, and statistics. Includes fundamental concepts and knowledge, covering all key tasks and techniques of temporal data mining, i.e., temporal data representations, similarity measure, and mining tasks Concentrates on temporal data clustering tasks from different perspectives, including major algorithms from clustering algorithms and ensemble learning approaches Presents a rich blend of theory and practice, addressing seminal research ideas and looking at the technology from a practical point-of-view Data Mining and Knowledge Discovery Handbook organizes all major concepts, theories, methodologies, trends, challenges and applications of data mining (DM) and knowledge discovery in databases (KDD) into a coherent and unified repository. This book first surveys, then provides comprehensive yet concise algorithmic descriptions of methods, including classic methods plus the extensions and novel methods developed recently. This volume concludes with in-depth descriptions of data mining applications in various interdisciplinary industries including finance, marketing, medicine, biology, engineering, telecommunications, software, and security. Data Mining and Knowledge Discovery Handbook is designed for research scientists and graduate-level students in computer science and engineering. This book is also suitable for professionals in fields such as computing applications, information systems management, and strategic research management.

Advances in K-means Clustering

Classification, Clustering, and Data Analysis

Uncertainty Handling and Quality Assessment in Data Mining

Data Mining in Large Sets of Complex Data

Applied Data Mining

Proceedings of the IIS ’ 2002 Symposium, Sopot, Poland, June 3 – 6, 2002

Essay from the year 2012 in the subject Computer Science - Theory, grade: B. Atlantic International University (School of Science and Engineering), course: Doctorate in Information Technology, language: English, abstract: Data mining is an independent science that based on advanced ways for information retrieval. Data mining is dealing with knowledge discovery in data warehouses without predefined hypotheses. So it is quite different from other applications such as decision support systems. OLAP and others which are looking for information on the factors and assumptions that we know it in advance. Data Mining supports multiple algorithms which have the ability to adopt automatic classification of historical data and predict future events. Data mining in the databases is designed to extract the hidden information, and it is a modern technology that imposed itself strongly in the information revolution, in the light of the great technological development and widespread use of data warehouses. Data mining techniques focus on building future forecasts and explore the behavior and trends, allowing a good estimation for right decisions that taken in a timely manner. This paper provides a general definition of data mining science and its most important techniques and algorithms used.

The book presents a long list of useful methods for classification, clustering and data analysis. By combining theoretical aspects with practical problems, it is designed for researchers as well as for applied statisticians and will support the fast transfer of new methodological advances to a wide range of applications.

This study presents current clustering algorithms to data mine large, highly dimensional and heterogeneous data sets.

A Data Mining Approach

Clustering

An Improved K-means Clustering Algorithm For Data Mining

Data Clustering

An Object-Oriented Approach

Classification, Clustering, and Applications