Text Mining Classification Clustering And Applications

Since the end of the Cold War, the threat of large-scale wars has been substituted by new threats: terrorism, organised crime, trafficking, smuggling, proliferation of weapons of mass destruction. To react to them, a security strategy is necessary, but in order to be effective it requires several instruments, including technological tools. Consequently, research and development in the field of security is proving to be an ever-expanding field all over the world. Data mining is seen more and more not only as a key technology in business, engineering and science but as one of the key features in security. To stress that all these technologies must be seen as a way to improve not only the security of citizens but also their freedom, special attention will be given to data protection research issues. The 10th International Conference on Data Mining is part of the successful series and the topics include: Text mining and text analytics; Data mining applications; Data mining methods.

Placed at the intersection of the digital humanities and Enlightenment studies, this collection is an interdisciplinary effort that showcases the significant digital work done in the field of eighteenth-century studies and its potential to transform our disciplinary practices. By addressing essential period-related themes – from issues of canonicity, intellectual history, and book trade practices to novel ways of exploring canonical authors and texts, gender roles, and public sphere dynamics – this collection also makes a broader argument about the necessity of expanding the very notion of “Enlightenment” not only spatially but also conceptually, by revisiting its very tenets in light of new data. The essays included here demonstrate that, by translating these new findings in suggestive visualizations, we can unveil unforeseen patterns, trends, connections, or networks of influence that could potentially revise existing master narratives about the period and the ideological structures at the core of the Enlightenment. Chapters 1, 3, 8 and 10 are available open access under a Creative Commons Attribution 4.0 International License via link.springer.com.

Leverage Natural Language Processing (NLP) in Python and learn how to set up your own robust environment for performing text analytics. This second edition has gone through a major revamp and introduces several significant changes and new topics based on the recent trends in NLP. You’ll see how to use the latest state-of-the-art frameworks in NLP, coupled with machine learning and deep learning models for supervised sentiment analysis powered by Python to solve actual case studies. Start by reviewing Python for NLP fundamentals on strings and text data and move on to engineering representation methods for text data, including both traditional statistical models and newer deep learning-based embedding models. Improved techniques and new methods around parsing and processing text are discussed as well. Text summarization and topic models have been overhauled so the book showcases how to build, tune, and interpret topic models in the context of an interest dataset on NIPS conference papers. Additionally, the book covers text similarity techniques with a real-world example of movie recommendations, along with sentiment analysis using supervised and unsupervised techniques. There is also a chapter dedicated to semantic analysis where you’ll see how to build your own named entity recognition (NER) system from scratch. While the overall structure of the book remains the same, the entire code base, modules, and chapters has been updated to the latest Python 3.x release. What You’ll Learn • Understand NLP and text syntax, semantics and structure • Discover text cleaning and feature engineering • Review text classification and text clustering • Assess text summarization and topic models • Study deep learning for NLP • Who This Book Is For IT professionals, data analysts, developers, linguistic experts, data scientists and engineers and basically anyone with a keen interest in linguistics, analytics and generating insights from textual data.

This book introduces text analytics as a valuable method for deriving insights from text data. Unlike other text analytics publications, Practical Text Analytics: Maximizing the Value of Text Data makes technical concepts accessible to those without extensive experience in the field. Using text analytics, organizations can derive insights from content such as emails, documents, and social media. Practical Text Analytics is divided into five parts. The first part introduces text analytics, discusses the relationship with content analysis, and provides a general overview of text mining methodology. In the second part, the authors discuss the practice of text analytics, including data preparation and the overall planning process. The third part covers text analytics techniques such as cluster analysis, topic models, and machine learning. In the fourth part of the book, readers learn about techniques used to communicate insights from text analysis, including data storytelling. The final part of Practical Text Analytics offers examples of the application of software programs for text analytics, enabling readers to mine their own text data to uncover information.

Text Mining Classification, Clustering, and Applications CRC Press

Natural Language Processing and Text Mining not only discusses applications of Natural Language Processing techniques to certain Text Mining tasks, but also the converse, the use of Text Mining to assist NLP. It assembles a diverse views from internationally recognized researchers and emphasizes caveats in the attempt to apply Natural Language Processing to text mining. This state-of-the-art survey is a must-have for advanced students, professionals, and researchers.

This book provides a perspective on the application of machine learning-based methods in knowledge discovery from natural languages texts. By analysing various data sets, conclusions which are not normally evident, emerge and can be used for various purposes and applications. The book provides explanations of principles of time-proven machine learning algorithms applied in text mining together with step-by-step demonstrations of how to reveal the semantic contents in real-world datasets using the popular R-language with its implemented machine learning algorithms. The book is not only aimed at IT specialists, but is meant for a wider audience that needs to process big sets of text documents and has basic knowledge of the subject, e.g. e-mail service providers, online shoppers, librarians, etc. The book starts with an introduction to text-based natural language data processing and its goals and problems. It focuses on machine learning, presenting various algorithms with their use and possibilities, and reviews the positives and negatives. Beginning with the initial data pre-processing, a reader can follow the steps provided in the R-language including the subsuming of various available plug-ins into the resulting software tool. A big advantage is that R also contains many libraries implementing machine learning algorithms, so a reader can concentrate on the principal target without the need to implement the details of the algorithms her- or himself. To make sense of the results, the book also provides explanations of the algorithms, which supports the final evaluation and interpretation of the results. The examples are demonstrated using realworld data from commonly
This book constitutes the refereed proceedings of the 13th International Conference on Machine Learning and Data Mining in Pattern Recognition, MLDM 2017, held in New York, NY, USA in July/August 2017. The 31 full papers presented in this book were carefully reviewed and selected from 150 submissions. The topics range from theoretical topics for classification, clustering, association rule and pattern mining to specific data mining methods for different multi-media data types such as image mining, text mining, video mining, and Web mining. This book constitutes the refereed proceedings of the First International Conference on Advanced Data Mining and Applications, ADMA 2005, held in Wuhan, China in July 2005. The conference was focused on topics such as data mining and applications that can handle new fields of data mining, e.g. spatial data mining, biomedical data mining, and mining on high-speed and time-variant data streams; an expansion of data mining to new applications is also strived for. The 25 revised full papers and 75 revised short papers presented were carefully peer-reviewed and selected from over 600 submissions. The papers are organized in topical sections on association rules, classification, clustering, novel algorithms, text mining, multimedia mining, sequential data mining and time series mining, web mining, biomedical mining, advanced applications, security and privacy issues, spatial data mining, and streaming data mining. This book introduces the reader to methods of data mining on the web, including uncovering patterns in web content (classification, clustering, language processing), structure (graphs, hubs, metrics), and usage (modeling, sequence analysis, performance).

In an age where customer opinion and feedback can have an immediate, major effect upon the success of a business or organization, marketers must have the ability to analyze unstructured data in everything from social media and internet reviews to customer surveys and phone logs. Practical Text Analytics is an essential daily reference resource, providing real-world guidance on the effective application of text analytics. The book presents the analysis process so that it is immediately understood by the marketing professionals who must use it, so they can apply proven concepts and methods correctly and with confidence. By decoding industry terminology and demonstrating practical application of data models once reserved for experts, Practical Text Analytics shows marketers how to frame the right questions, identify key themes and find hidden meaning from unstructured data. Readers will learn to develop powerful new marketing strategies to elevate customer experience, solidify brand value and elevate reputation. Online resources include self-test questions, chapter review Q&A and an Instructor's Manual with text sources and instructions.

Data Mining for Business Analytics: Concepts, Techniques, and Applications in Python presents an applied approach to data mining concepts and methods, using Python software for illustration. Readers will learn how to implement a variety of popular data mining algorithms in Python (a free and open-source software) to tackle business problems and opportunities. This is the sixth version of this successful, first using Python. It covers both statistical and machine learning algorithms for prediction, classification, visualization, dimension reduction, recommender systems, clustering, text mining and network analysis. It also includes: A new co-author, Peter Gedeck, who brings both experience teaching business analytics courses using Python, and expertise in the application of machine learning methods to the drug-discovery process A new section on ethical issues in data mining Updates and new material based on feedback from instructors teaching MBA, undergraduate, diploma and executive courses, and from their students More than a dozen case studies demonstrating applications for the data mining techniques described End-of-chapter exercises that help readers gauge and expand their comprehension and competency of the material presented A companion website with more than two dozen data sets, and instructor materials including exercise solutions, PowerPoint slides, and case solutions Data Mining for Business Analytics: Concepts, Techniques, and Applications in Python is an ideal textbook for undergraduate and upper-undergraduate level courses in data mining, predictive analytics, and business analytics. This new edition is also an excellent reference for analysts, researchers, and practitioners working with quantitative methods in the fields of finance, business, marketing, computer science, and information technology. “This book has by far the most comprehensive review of business analytics methods that I have ever seen, covering everything from classical approaches such as linear and logistic regression, through to modern methods like neural networks, bagging and boosting, and even much more business specific procedures such as social network analysis and text mining. If not the bible, it is at the least a definitive manual on the subject.” —Gareth M. James, University of Southern California and co-author (with Witten, Hastie and Tibshirani) of the best-selling book An Introduction to Statistical Learning, with Applications in R

This book presents the concepts, implementation of text mining with real life examples implemented using Python libraries. You will find ideas how to use texts for extracting valuable and applicable information. The book is designed for academicians, students, researchers and those who are working as data scientist in sector. The book not only defines but also gives Python examples of Information Retrieval, Concept Extraction, Classification, Clustering, Sentiment Analysis, Topic Extraction, Text Summarization, and Web Mining. In the book you will also find a practical example how to use Genetic Algorithms, Naive Bayes and Artificial Neural Networks for text mining. Table of ContentsForewordAbout the AuthorAcknowledgementsCHAPTER I Concepts of Text Mining1)HISTORY of TEXT MINING2)DEFINITION of TEXT MINING3) COMPONENTS of TEXT MINING4) PRACTICAL APPLICATIONS of TEXT MININGCHAPTER II Text Mining Algorithms and Examples1)INFORMATION RETRIEVAL(i) Similarity(ii) Vectorization(iii) Calculating Term Weighting and Frequency(iv) Measuring the quality of IR2)INFORMATION EXTRACTION(i) Lexical Analysis(ii) Tokenization(iii) Filtering: Stop-words(iv) Lemmatization(v) Bag of Words(vi) N-Gram(vii) Tagging/Annotation, XML3) BASIC TASKS FOR TEXT MINING(i) Text Categorization(ii) Data Mining Techniques: Link Analysis and Association Analysis, Visualization, And Predictive Analytics(iii) Pattern Recognition(iv) Text Clustering And Word Clouding(v) Natural Language Processing (NLP) (vi) Sentiment Analysis4) AUTOMATIC DOCUMENT SUMMARIZATION(i) Extraction-based summarization(ii) Abstraction-based summarization(iii) Aided SummarizationCHAPTER III Text Mining With Python1) STARTING A TEXT MINING IMPLEMENTATION2) PYTHON ENVIRONMENT3) Examples with Python

The world of text mining is simultaneously a minefield and a gold mine. It is an exciting application field and an area of scientific research that is currently under rapid development. It uses techniques from well-established scientific fields (e.g. data mining, machine learning, information retrieval, natural language processing, case based reasoning, statistics and knowledge management) in an effort to help people gain insight, understand and interpret large quantities of (usually) semi-structured and unstructured data. Despite the advances made during the last few years, many issues remain unmet. Proper co-ordination activities, dissemination of current trends and standardisation of the procedures have been identified, as key needs. There are many questions still unanswered, especially to the potential users; what is the scope of Text Mining, who uses it and for what purpose, what constitutes the leading trends in the field of Text Mining -especially in relation to IT- and whether there still remain areas to be covered.

R and Data Mining introduces researchers, post-graduate students, and analysts to data mining using R, a free software environment for statistical computing and graphics. The book provides practical methods for using R in applications from academia to industry to extract knowledge from vast amounts of data. Readers will find this book a valuable guide to the use of R in tasks such as classification and prediction, clustering, outlier detection, association rules, sequence analysis, text mining, social network analysis, sentiment analysis, and more. Data mining techniques are growing in popularity in a broad range of areas, from banking to insurance, retail, telecom, medicine, research, and government. This book focuses on the modeling phase of the data mining process, also addressing data exploration and
model evaluation. With three in-depth case studies, a quick reference guide, bibliography, and links to a wealth of online resources, R and Data Mining is a valuable, practical guide to a powerful method of analysis. Presents an introduction into using R for data mining applications, covering most popular data mining techniques Provides code examples and data so that readers can easily learn the techniques Features case studies in real-world applications to help readers apply the techniques in their work The two-volume set LNAI 8346 and 8347 constitutes the thoroughly refereed proceedings of the 9th International Conference on Advanced Data Mining and Applications, ADMA 2013, held in Hangzhou, China, in December 2013. The 32 regular papers and 64 short papers presented in these two volumes were carefully reviewed and selected from 222 submissions. The papers included in these two volumes cover the following topics: opinion mining, behavior mining, data stream mining, sequential data mining, web mining, image mining, text mining, social network mining, classification, clustering, association rule mining, pattern mining, regression, predication, feature extraction, identification, privacy preservation, applications, and machine learning. 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. Text Mining: Applications and Theory presents the state-of-the-art algorithms for text mining from both the academic and industrial perspectives. The contributors span several countries and scientific domains: universities, industrial corporations, and government laboratories, and demonstrate the use of techniques from machine learning, knowledge discovery, natural language processing and information retrieval to design computational models for automated text analysis and mining. This volume demonstrates how advancements in the fields of applied mathematics, computer science, machine learning, and natural language processing can collectively capture, classify, and interpret words and their contexts. As suggested in the preface, text mining is needed when “words are not enough.” This book: Provides state-of-the-art algorithms and techniques for critical tasks in text mining applications, such as clustering, classification, anomaly and trend detection, and stream analysis. Presents a survey of text visualization and visualization tools and looks at a practical text classification problem. Discusses the issue of cybercrime associated with chatrooms. Features advances in visual analytics and machine learning along with illustrative examples. Is accompanied by a supporting website featuring datasets. Applied mathematicians, statisticians, practitioners and students in computer science, bioinformatics and engineering will find this book extremely useful. This two-volume set LNAI 10934 and LNAI 10935 constitutes the refereed proceedings of the 14th International Conference on Machine Learning and Data Mining in Pattern Recognition, MLDM 2018, held in New York, NY, USA in July 2018. The 92 regular papers presented in this two-volume set were carefully reviewed and selected from 298 submissions. The topics range from theoretical topics for classification, clustering, association rule and pattern mining to specific data mining methods for the different multi-media data types such as image mining, text mining, video mining, and Web mining. Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications brings together all the information, tools and methods a professional will need to efficiently use text mining applications and statistical analysis. Winner of a 2012 PROSE Award in Computing and Information Sciences from the Association of American Publishers, this book 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. The world contains an unimaginably vast amount of digital information which is getting even 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 data that it contains diminishes, the value of text mining for information retrieval and search will increase dramatically. 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. First title ever present soft computing approaches and their application in data mining along with the traditional hard-computing approaches Addresses the principles of multimedia data compression techniques (for image, video, text) and their role in data mining Discusses principles and classical algorithms on string matching and their role in data mining Extracting content from text continues to be an important research problem for information processing and management. Approaches to capture the semantics of text-based document collections may be based on Bayesian models, probability theory, vector space models, statistical models, or even graph theory. As the volume of digitized textual media continues to grow, so does the need for designing robust, scalable indexing and search strategies (software) to meet a variety of user needs. Knowledge extraction or creation from text requires systematic yet reliable processing that can be codified and adapted for changing needs and environments. This book will draw upon experts in both academia and industry to recommend practical approaches to the purification, indexing, and mining of textual information. It will address document identification, clustering and categorizing documents, cleaning text, and visualizing semantic models of text. This book includes a selection of papers from the 2018 World Conference on Information Systems and Technologies (WorldCIST'18), held in Naples, Italy on March27-29, 2018. WorldCIST is a global forum for researchers and practitioners to present and discuss recent results and innovations, current trends, professional experiences and the challenges of modern information systems and technologies research together with their technological development and applications. The main topics covered are: A) Information and Knowledge Management; B) Organizational Models and Information Systems; C) Software and Systems Modeling; D) Software Systems, Architectures, Applications and Tools; E) Multimedia Systems and Applications; F) Computer Networks, Mobility and Pervasive Systems; G) Intelligent and Decision Support Systems; H) Big Data Analytics and Applications; I) Human–Computer Interaction; J) Ethics, Computers & Security; K) Health Informatics; L) Information Technologies in Education; M) Information Technologies in Radiocommunications; N) Technologies for Biomedical Applications. This successful textbook on predictive text mining offers a unified perspective on a rapidly evolving field, integrating topics spanning the varied disciplines of data science, machine learning, databases, and computational linguistics. Also serves as a practical guide, this unique book provides helpful advice illustrated by examples and case studies. This highly anticipated second edition has been thoroughly revised and expanded with new material on deep learning, graph models, mining social media, errors and pitfalls in big data evaluation, Twitter sentiment analysis, and dependency parsing discussion. The fully updated content also features in-depth discussions on issues of document classification, information retrieval, clustering and organizing documents, information extraction, web-based data-sourcing, and prediction and evaluation. Features includes chapter summaries and exercises; explores the application of each method; provides several case studies; contains links to free text-mining software. 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.

There is no royal road to science, and only those who do not dread the fatiguing climb of its steep paths have a chance of gaining its luminous summits. Karl Marx A Universal Genius of the 19th Century Many scientists from all over the world during the past two years since the MLDM 2007 have come along on the stony way to the sunny summit of science and have worked hard on new ideas and applications in the area of data mining in pattern r-ogntion. Our thanks go to all those who took part in this year's MLDM. We appre- ate their submissions and the ideas shared with the Program Committee. We received over 205 submissions from all over the world to the International Conference on -chine Learning and Data Mining, MLDM 2009. The Program Committee carefully selected the best papers for this year's program and gave detailed comments on each submitted paper. There were 63 papers selected for oral presentation and 17 papers for poster presentation. The topics range from theoretical topics for classification, clustering, association rule and pattern mining to specific data-mining methods for the different multimedia data types such as image mining, text mining, video mining and Web mining. Among these topics this year were special contributions to subtopics such as attribute discrete- zation and data preparation, novelty and outlier detection, and distances and simila-
ties.

This book constitutes the refereed proceedings of the 8th International Conference, MLDM 2012, held in Berlin, Germany in July 2012. The 51 revised full papers presented were carefully reviewed and selected from 212 submissions. The topics range from theoretical topics for classification, clustering, association rule and pattern mining to specific data mining methods for the different multimedia data types such as image mining, text mining, video mining and web mining. Organizations generate and collect large volumes of textual data. Unfortunately, many companies are unable to capitalize fully on the value of this data because information implicit within it is not easy to discern. Primarily intended for business analysts and statisticians across multiple industries, this book provides an introduction to the types of problems encountered and current available text mining solutions.

Manufacturing companies are increasingly enhancing their web presence as a strategy for improving their visibility in the global market. The exponential growth of manufacturing websites has resulted in a drastic increase in the size and variety of unstructured manufacturing information available online. This poses both challenges and opportunities. The challenge is related to efficient information search and retrieval when dealing with a large volume of heterogeneous information. Traditional search methods, such as keyword search, can no longer meet the information retrieval and organization needs of the manufacturing cyberspace. At the same time, the textual data available online contains a wealth of technical knowledge that, if mined properly, may result in discovery of new patterns and trends that were otherwise unknown. There is an acute need for more advanced computational tools and techniques that can help search, organize, and summarize large archives of text pertaining to technological capabilities of manufacturing suppliers. In this research, three text mining techniques, namely, Classification, Clustering, and Topic Modeling are applied to analyzing manufacturing data. R programming package is used for implementation of the aforementioned techniques. The novelty of the proposed classification technique is in adopting concept-based method rather than term-based method that results in higher semantic relevance of the results. Also, clustering and topic modeling are serialized to improve the likelihood of discovering useful knowledge patterns.

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.

Text Mining and Visualization: Case Studies Using Open-Source Tools provides an introduction to text mining using some of the most popular and powerful open-source tools: KNIME, RapidMiner, Weka, R, and Python. The contributors—all highly experienced with text mining and open-source software—explain how text data are gathered and processed from a w

This book constitutes the refereed proceedings of the 11th International Conference on Machine Learning and Data Mining in Pattern Recognition, MLDM 2015, held in Hamburg, Germany in July 2015. The 41 full papers presented were carefully reviewed and selected from 123 submissions. The topics range from theoretical topics for classification, clustering, association rule and pattern mining to specific data mining methods for the different multimedia data types such as image mining, text mining, video mining and Web mining. Welcome to the proceedings of the Sixth International Conference on Management Science and Engineering Management (ICMSEM2012) held from November 11 to 14, 2012 at Quaid-i-Azam University,
Islamabad, Pakistan and supported by Sichuan University (Chengdu, China), Quaid-i-Azam University (Islamabad, Pakistan) and The National Natural Science Foundation of China. The International Conference on Management Science and Engineering Management is the annual conference organized by the International Society of Management Science and Engineering Management. The goals of the Conference are to foster international research collaborations in Management Science and Engineering Management as well as to provide a forum to present current research results. The papers are classified into 8 sections: Computer and Networks, Information Technology, Decision Support System, Industrial Engineering, Supply Chain Management, Project Management, Manufacturing and Ecological Engineering. The key issues of the sixth ICMSEM cover various areas in MSEM, such as Decision Support System, Computational Mathematics, Information Systems, Logistics and Supply Chain Management, Relationship Management, Scheduling and Control, Data Warehousing and Data Mining, Electronic Commerce, Neural Networks, Stochastic models and Simulation, Heuristics Algorithms, Risk Control, and Carbon Credits.

This book constitutes the refereed proceedings of the 10th International Conference on Machine Learning and Data Mining in Pattern Recognition, MLDM 2014, held in St. Petersburg, Russia in July 2014. The 40 full papers presented were carefully reviewed and selected from 128 submissions. The topics range from theoretical topics for classification, clustering, association rule and pattern mining to specific data mining methods for the different multimedia data types such as image mining, text mining, video mining and Web mining.

This volume collects revised versions of papers presented at the 29th Annual Conference of the Gesellschaft für Klassifikation, the German Classification Society, held at the Otto-von-Guericke-University of Magdeburg, Germany, in March 2005. In addition to traditional subjects like Classification, Clustering, and Data Analysis, converage extends to a wide range of topics relating to Computer Science: Text Mining, Web Mining, Fuzzy Data Analysis, IT Security, Adaptivity and Personalization, and Visualization.

This book discusses text mining and different ways this type of data mining can be used to find implicit knowledge from text collections. The author provides the guidelines for implementing text mining systems in Java, as well as concepts and approaches. The book starts by providing detailed text preprocessing techniques and then goes on to provide concepts, the techniques, the implementation, and the evaluation of text categorization. It then goes into more advanced topics including text summarization, text segmentation, topic mapping, and automatic text management.

Copyright: 48d5e440a2ab8c4391409763a362cd85

Page 5/5