Learning To Classify Text Using Support Vector Machines The Springer International Series In Engineering And Computer Science

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 techniques and looks at the multilingual 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. Learning to Classify Text Using Support Vector MachinesSpringer Science & Business Media Recent work has pointed to the need for a detection-based approach to transfer capable of discovering elusive crosslinguistic effects through the use of human judges and computer classifiers that can learn to predict learners' language backgrounds based on their patterns of language use. This book addresses that need. It details the nature of the detection-based approach, discusses how this approach fits into the overall scope of transfer research, and discusses the few previous studies that have laid the groundwork for this approach. The core of the book consists of five empirical studies that use computer classifiers to detect the native-language affiliations of texts written by foreign language learners of English. The results highlight combinations of language features that are the most reliable predictors of learners' language backgrounds.

The pharmaceutical industry relies on numerous well-designed experiments involving high-throughput techniques and in silico approaches to analyze potential drug targets. These in silico methods are often predictive, yielding faster and less expensive analyses than traditional in vivo or in vitro procedures. In Silico Technologies in Drug Target Identification and Validation addresses the challenge of testing a growing number of new potential targets and reviews currently available in silico approaches for identifying and validating these targets. The book emphasizes computational tools, public and commercial databases, mathematical methods, and software for interpreting complex experimental data. The book describes how these tools are used to visualize a target structure, identify binding sites, and predict behavior. World-renowned researchers cover many topics not typically found in most informatics books, including functional annotation, siRNA design, pathways, text mining, ontologies, systems biology, database management, data pipelining, and pharmacogenomics. Covering issues that range from prescreening target selection to genetic modeling and valuable data integration, In Silico Technologies in Drug Target Identification and Validation is a self-contained and practical guide to the various computational tools that can accelerate the identification and validation stages of drug target discovery and determine the biological functionality of potential targets more effectively. Daniel E. Levy, editor of the Drug Discovery Series, is the founder of DEL BioPharma, a consulting service for drug discovery programs. He also maintains a blog

that explores organic chemistry.

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. Serving also 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.

Based on ideas from Support Vector Machines (SVMs), Learning To Classify Text Using Support Vector Machines presents a new approach to generating text classifiers from examples. The approach combines high performance and efficiency with theoretical understanding and improved robustness. In particular, it is highly effective without greedy heuristic components. The SVM approach is computationally efficient in training and classification, and it comes with a learning theory that can guide real-world applications. Learning To Classify Text Using Support Vector Machines gives a complete and detailed description of the SVM approach to learning text classifiers, including training algorithms, transductive text classification, efficient performance estimation, and a statistical learning model of text classification. In addition, it includes an overview of the field of text classification, making it self-contained even for newcomers to the field. This book gives a concise introduction to SVMs for pattern recognition, and it includes a detailed description of how to formulate text-classification tasks for machine learning.

"We live in the age of data. In the last few years, the methodology of extracting insights from data or "data science" has emerged as a discipline in its own right. The R programming language has become one-stop solution for all types of data analysis. The growing popularity of R is due its statistical roots and a vast open source package library. The goal of "Beginning Data Science with R" is to introduce the readers to some of the useful data science techniques and their implementation with the R programming language. The book attempts to strike a balance between the how: specific processes and methodologies, and understanding the why: going over the intuition behind how a particular technique works, so that the reader can apply it to the problem at hand. This book will be useful for readers who are not familiar with statistics and the R programming language.

***** BUY NOW (will soon return to 24.97 $) ***** MONEY BACK GUARANTEE BY AMAZON (See Below FAQ) ***** *** Free eBook for customers who purchase the print book from Amazon *** Are you thinking of learning more Natural Language Processing (NLP) using TensorFlow? This book is for you. It would seek to explain common terms and algorithms in an intuitive way. The authors used a progressive approach whereby we start out slowly and improve on the complexity of our solutions.This book and the accompanying examples, you would be well suited to tackle problems which pique your interests using ?NLP. From AI Sciences Publisher Our books may be the best one for beginners; it's a step-by-step guide for any person who wants to start learning Artificial Intelligence and Data Science from scratch. It will help you in preparing a solid foundation and learn any other high-level courses.To get the most out of the concepts that would be covered, readers are advised to adopt a hands on approach which would lead to better mental representations. Target Users The book designed for a variety of

target audiences. The most suitable users would include: Anyone who is intrigued by how algorithms arrive at predictions but has no previous knowledge of the field. Software developers and engineers with a strong programming background but seeking to break into the field of Data Science and NLP. Seasoned professionals in the field of artificial intelligence and machine learning who desire a bird's eye view of current techniques and approaches. What's Inside This Book? Introduction to Natural Language Processing What is Natural Language Processing Perspectivizing NLP: Areas of AI and Their Interdependencies Purpose of Natural Language Processing Text Manipulation Tokenization Stemming Lemmatization Normalization Accessing Text Corpora and Lexical Resources Processing Raw Text Categorizing and Tagging Words NLP Applications Text Classification Sentiment Classification Topic Modelling Question Answering Speech Recognition Machine Translation Word Representation Bag of Words One-Hot Encoding Word Vectors Representation Word2Vec and GloVe Learning to Classify Text Supervised Classification Decision Trees Naive Bayes Classifiers Maximum Entropy Classifiers Deep Learning for NLP What is Deep Learning Feed Forward Neural Networks Recurrent Neural Networks Gated Recurrent Unit Long Short Term Memory Language Processing and Python using NLTK Introduction to TensorFlow Text Classification Frequently Asked Questions Q: Is this book for me and do I need programming experience?A: If you want to smash NLP from scratch, this book is for you. If you already wrote a few lines of code and recognize basic programming statements, you'll be OK.Q: Does this book include everything I need to become a NLP expert?A: Unfortunately, no. This book is designed for readers taking their first steps in NLP and further learning will be required beyond this book to master all aspects of NLP.Q: Can I have a refund if this book doesn't fit for me?A: Yes, Amazon refund you if you aren't satisfied, for more information about the amazon refund service please go to the amazon help platform. We will also be happy to help you if you send us an email at contact@aisciences.net.

Automatic Text Classification has always been given importance in the filed of computer since the beginning of digital documents. Considering the large amounts of documents online and the speed with which the digital information is being produced, automating the task of text classification has a great practical use. Given the task of automation, the documents can be classified based on the genre of the articles, for instance : politics, sports, religion etc. The digital documents are available in the form of news feeds, online news article, journal papers etc. Text classification is a task of classifying a document into a predefined category. If we have a document d in a set of document D, and we have predefined classes $c_1, c_2, c_3 ... c_N$, the document d will be classified and be associated with a class c_i, based on what it contains. Text classification is done based on the readily available statistical algorithms, these algorithms need to be trained with a set of labeled documents and a set of test document are classified with the these algorithms. The accuracy with which the test documents are classifies gives us a measure of how well the algorithm can perform and thus can be used to categorize unlabeled documents. I aim to develop the Bayesian Classifier in java and train the algorithm with a certain test data and calculate the accuracy of the classifier and how well it fairs when applied to a testing data which is already labeled.

This volume presents theories, models, algorithms, and applications in clustering,

classification, and visualization. It also includes applications of clustering, classification, and visualization in various fields such as marketing, recommendation system, biology, sociology, and social survey. The contributions give insight into new models and concepts and show the variety of research in clustering, classification, and visualization. Comprehensive Coverage of the Entire Area of Classification Research on the problem of classification tends to be fragmented across such areas as pattern recognition, database, data mining, and machine learning. Addressing the work of these different communities in a unified way, Data Classification: Algorithms and Applications explores the underlying algorithms of classification as well as applications of classification in a variety of problem domains, including text, multimedia, social network, and biological data. This comprehensive book focuses on three primary aspects of data classification: Methods: The book first describes common techniques used for classification, including probabilistic methods, decision trees, rule-based methods, instance-based methods, support vector machine methods, and neural networks. Domains: The book then examines specific methods used for data domains such as multimedia, text, time-series, network, discrete sequence, and uncertain data. It also covers large data sets and data streams due to the recent importance of the big data paradigm. Variations: The book concludes with insight on variations of the classification process. It discusses ensembles, rare-class learning, distance function learning, active learning, visual learning, transfer learning, and semi-supervised learning as well as evaluation aspects of classifiers.

This book constitutes the refereed proceedings of the 11th Portuguese Conference on Artificial Intelligence, EPIA 2003, held in Beja, Portugal in December 2003. The 29 revised full papers and 20 revised short papers presented were carefully reviewed and selected from a total of 119 submissions. In accordance with the five constituting workshops, the papers are organized in topical sections on artificial life and evolutionary algorithms, constraint and logic programming systems, extraction of knowledge from databases, multi-agent systems and AI for the Internet, and natural language processing and text retrieval.

This accessible introduction shows the reader how to understand, implement, adapt, and apply Learning Classifier Systems (LCSs) to interesting and difficult problems. The text builds an understanding from basic ideas and concepts. The authors first explore learning through environment interaction, and then walk through the components of LCS that form this rule-based evolutionary algorithm. The applicability and adaptability of these methods is highlighted by providing descriptions of common methodological alternatives for different components that are suited to different types of problems from data mining to autonomous robotics. The authors have also paired exercises and a simple educational LCS (eLCS) algorithm (implemented in Python) with this book. It is suitable for courses or self-study by advanced undergraduate and postgraduate students in subjects such as Computer Science, Engineering, Bioinformatics, and Cybernetics, and by researchers, data analysts, and machine learning practitioners. With the ever-growing power to generate, transmit and collect huge amounts of data, information overload is now an imminent problem to mankind. The overwhelming demand for information processing is not just about a better - derstanding of data, but also a better usage of data in a timely fashion. Data mining, or knowledge discovery from databases, is proposed to gain insight into aspects of dataand to help

peoplemakeinformed, sensible, andbetter decisions. At present, growing attention has been paid to the study, development and - plication of data mining. As a result there is an urgent need for sophisticated techniques and tools that can handle new ?elds of data mining, e.g., spatialdata mining, biomedical data mining, and mining on high-speed and time-variant data streams. The knowledge of data mining should also be expanded to new applications.

The1stInternationalConferenceonAdvancedDataMiningandApplications (ADMA 2005) aimed to bring together the experts on data mining throughout the world. It provided a leading international forum for the dissemination of original research results in advanced data mining techniques, applications, al- rithms, software and systems, and di?erent applied disciplines. The conference attracted 539 online submissions and 63 mailing submissions from 25 di?erent countriesandareas.Allfullpaperswerepeer reviewedbyatleastthreemembers of the Program Committee composed of international experts in data mining ?elds. A total number of 100 papers were accepted for the conference. Amongst them 25 papers were selected as regular papers and 75 papers were selected as short papers, yielding a combined acceptance rate of 17%.

Text classification is becoming a crucial task to analysts in different areas. In the last few decades, the production of textual documents in digital form has increased exponentially. Their applications range from web pages to scientific documents, including emails, news and books. Despite the widespread use of digital texts, handling them is inherently difficult - the large amount of data necessary to represent them and the subjectivity of classification complicate matters. This book gives a concise view on how to use kernel approaches for inductive inference in large scale text classification; it presents a series of new techniques to enhance, scale and distribute text classification tasks. It is not intended to be a comprehensive survey of the state-of-the-art of the whole field of text classification. Its purpose is less ambitious and more practical: to explain and illustrate some of the important methods used in this field, in particular kernel approaches and techniques.

Abstract: Text classification is one of the classical problems in computer science, which is primarily used for categorizing data, spam detection, anonymization, information extraction, text summarization etc. Given the large amounts of data involved in the above applications, automated and accurate training models and approaches to classify data efficiently are needed. In this thesis, an extensive study of the interaction between natural language processing, information retrieval and text classification has been performed. A case study named "keyword extraction" that deals with 'identifying keywords and tags from millions of text questions' is used as a reference. Different classifiers are implemented using MapReduce paradigm on the case study and the experimental results are recorded using two newly built distributed computing Hadoop clusters. The main aim is to enhance the prediction accuracy, to examine the role of text pre-processing for noise elimination and to reduce the computation time and resource utilization on the clusters.

Human Language Technology (HLT) and Natural Language Processing (NLP) systems have typically focused on the "factual" aspect of content analysis. Other aspects, including pragmatics, opinion, and style, have received much less attention. However, to achieve an adequate understanding of a text, these aspects cannot be ignored. The chapters in this book address the aspect of subjective opinion, which includes identifying different points of view, identifying different emotive dimensions, and classifying text by opinion. Various conceptual models and computational methods are presented. The models explored in this book include the following: distinguishing attitudes from simple factual assertions; distinguishing between the author's reports from reports of other people's opinions; and distinguishing between explicitly and implicitly stated attitudes. In addition, many applications are described that

promise to benefit from the ability to understand attitudes and affect, including indexing and retrieval of documents by opinion; automatic question answering about opinions; analysis of sentiment in the media and in discussion groups about consumer products, political issues, etc. ; brand and reputation management; discovering and predicting consumer and voting trends; analyzing client discourse in therapy and counseling; determining relations between scientific texts by finding reasons for citations; generating more appropriate texts and making agents more believable; and creating writers' aids. The studies reported here are carried out on different languages such as English, French, Japanese, and Portuguese. Difficult challenges remain, however. It can be argued that analyzing attitude and affect in text is an "NLP"-complete problem.

Nowadays, documents are increasingly being associated with multi-level category hierarchies rather than a flat category scheme. To access these documents in real time, we need fast automatic methods to navigate these hierarchies. Today's vast data repositories such as the web also contain many broad domains of data which are quite distinct from each other e.g. medicine, education, sports and politics. Each domain constitutes a subspace of the data within which the documents are similar to each other but quite distinct from the documents in another subspace. The data within these domains is frequently further divided into many subcategories. Subspace Learning is a technique popular with non-text domains such as image recognition to increase speed and accuracy. Subspace analysis lends itself naturally to the idea of hybrid classifiers. Each subspace can be processed by a classifier best suited to the characteristics of that particular subspace. Instead of using the complete set of full space feature dimensions, classifier performances can be boosted by using only a subset of the dimensions. This thesis presents a novel hybrid parallel architecture using separate classifiers trained on separate subspaces to improve two-level text classification. The classifier to be used on a particular input and the relevant feature subset to be extracted is determined dynamically by using a novel method based on the maximum significance value. A novel vector representation which enhances the distinction between classes within the subspace is also developed. This novel system, the Hybrid Parallel Classifier, was compared against the baselines of several single classifiers such as the Multilayer Perceptron and was found to be faster and have higher two-level classification accuracies. The improvement in performance achieved was even higher when dealing with more complex category hierarchies.

This book provides a comprehensive text on Web data mining. Key topics of structure mining, content mining, and usage mining are covered. The book brings together all the essential concepts and algorithms from related areas such as data mining, machine learning, and text processing to form an authoritative and coherent text. The book offers a rich blend of theory and practice. It is suitable for students, researchers and practitioners interested in Web mining. Lecturers can readily use it for classes on data mining, Web mining, and Web search. Internet support with lecture slides and project problems is available online.

Focus on implementing end-to-end projects using Python and leverage state-of-the-art algorithms. This book teaches you to efficiently use a wide range of natural language processing (NLP) packages to: implement text classification, identify parts of speech, utilize topic modeling, text summarization, sentiment analysis, information retrieval, and many more applications of NLP. The book begins with text data collection, web scraping, and the different types of data sources. It explains how to clean and pre-process text data, and offers ways to analyze data with advanced algorithms. You then explore semantic and syntactic analysis of the text. Complex NLP solutions that involve text normalization are covered along with advanced pre-processing methods, POS tagging, parsing, text summarization, sentiment analysis, word2vec, seq2seq, and much more. The book presents the fundamentals necessary for applications of machine learning and deep learning in NLP. This second edition goes over advanced techniques to convert text to features such as Glove, Elmo, Bert, etc. It also includes

an understanding of how transformers work, taking sentence BERT and GPT as examples. The final chapters explain advanced industrial applications of NLP with solution implementation and leveraging the power of deep learning techniques for NLP problems. It also employs state-of-the-art advanced RNNs, such as long short-term memory, to solve complex text generation tasks. After reading this book, you will have a clear understanding of the challenges faced by different industries and you will have worked on multiple examples of implementing NLP in the real world. What You Will Learn Know the core concepts of implementing NLP and various approaches to natural language processing (NLP), including NLP using Python libraries such as NLTK, textblob, SpaCy, Standford CoreNLP, and more Implement text pre-processing and feature engineering in NLP, including advanced methods of feature engineering Understand and implement the concepts of information retrieval, text summarization, sentiment analysis, text classification, and other advanced NLP techniques leveraging machine learning and deep learning Who This Book Is For Data scientists who want to refresh and learn various concepts of natural language processing (NLP) through coding exercises

The contributions in this volume represent the latest research results in the field of Classification, Clustering, and Data Analysis. Besides the theoretical analysis, papers focus on various application fields as Archaeology, Astronomy, Bio-Sciences, Business, Electronic Data and Web, Finance and Insurance, Library Science and Linguistics, Marketing, Music Science, and Quality Assurance.

This volume contains a selection of revised papers that were presented at the Software Aspects of Robotic Systems, SARS 2011 Workshop and the Machine Learning for System Construction, MLSC 2011 Workshop, held during October 17-18 in Vienna, Austria, under the auspices of the International Symposium Series on Leveraging Applications of Formal Methods, Verification, and Validation, ISoLA. The topics covered by the papers of the SARS and the MLSC workshop demonstrate the breadth and the richness of the respective fields of the two workshops stretching from robot programming to languages and compilation techniques, to real-time and fault tolerance, to dependability, software architectures, computer vision, cognitive robotics, multi-robot-coordination, and simulation to bio-inspired algorithms, and from machine learning for anomaly detection, to model construction in software product lines to classification of web service interfaces. In addition the SARS workshop hosted a special session on the recently launched KOROS project on collaborating robot systems that is borne by a consortium of researchers of the faculties of architecture and planning, computer science, electrical engineering and information technology, and mechanical and industrial engineering at the Vienna University of Technology. The four papers devoted to this session highlight important research directions pursued in this interdisciplinary research project. 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. The field of pattern recognition has seen enormous progress since its beginnings

almost 50 years ago. A large number of different approaches have been proposed. Hybrid methods aim at combining the advantages of different paradigms within a single system. Hybrid Methods in Pattern Recognition is a collection of articles describing recent progress in this emerging field. It covers topics such as the combination of neural nets with fuzzy systems or hidden Markov models, neural networks for the processing of symbolic data structures, hybrid methods in data mining, the combination of symbolic and subsymbolic learning, and so on. Also included is recent work on multiple classifier systems. Furthermore, the book deals with applications in on-line and off-line handwriting recognition, remotely sensed image interpretation, fingerprint identification, and automatic text categorization.

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.

Here are the refereed proceedings of the 9th International Conference on Text, Speech and Dialogue, TSD 2006. The book presents 87 revised full papers together with 2 invited papers reviewing state-of-the-art research in the field of natural language processing. Coverage ranges from theoretical and methodological issues to applications with special focus on corpora, texts and transcription, speech analysis, recognition and synthesis, as well as their intertwining within NL dialogue systems.

A guide on the use of SVMs in pattern classification, including a rigorous performance comparison of classifiers and regressors. The book presents architectures for multiclass classification and function approximation problems, as well as evaluation criteria for classifiers and regressors. Features: Clarifies the characteristics of two-class SVMs; Discusses kernel methods for improving the generalization ability of neural networks and fuzzy systems; Contains ample illustrations and examples; Includes performance evaluation using publicly available data sets; Examines Mahalanobis kernels, empirical feature space, and the effect of model selection by cross-validation; Covers sparse SVMs, learning using privileged information, semi-supervised learning, multiple classifier

systems, and multiple kernel learning; Explores incremental training based batch training and active-set training methods, and decomposition techniques for linear programming SVMs; Discusses variable selection for support vector regressors. Automatic text document classification is a fundamental problem in machine learning. Given the dynamic nature and the exponential growth of the World Wide Web, one needs the ability to classify not only a massive number of documents, but also documents that belong to wide variety of domains. Some examples of the domains are e-mails, blogs, Wikipedia articles, news articles, newsgroups, online chats, etc. It is the difference in the writing style that differentiates these domains. Text documents are usually classified using supervised learning algorithms that require large set of pre-labeled data. This requirement, of labeled data, poses a challenge in classifying documents that belong to different domains. Our goal is to classify text documents in the testing domain without requiring any labeled documents from the same domain. Our research develops specialized cross-domain learning algorithms based the distributions over words obtained from a collection of text documents by topic models such as Latent Dirichlet Allocation (LDA). Our major contributions include (1) empirically showing that conventional supervised learning algorithms fail to generalize their learned models across different domains and (2) development of novel and specialized cross-domain classification algorithms that show an appreciable improvement over conventional methods used for cross-domain classification that is consistent for different datasets. Our research addresses many real-world needs. Since massive number of new types of text documents is generated daily, it is crucial to have the ability to transfer learned information from one domain to another domain. Cross-domain classification lets us leverage information learned from one domain for use in the classification of documents in a new domain.

An essential guide to two burgeoning topics in machine learning – classification trees and ensemble learning Ensemble Classification Methods with Applications in R introduces the concepts and principles of ensemble classifiers methods and includes a review of the most commonly used techniques. This important resource shows how ensemble classification has become an extension of the individual classifiers. The text puts the emphasis on two areas of machine learning: classification trees and ensemble learning. The authors explore ensemble classification methods' basic characteristics and explain the types of problems that can emerge in its application. Written by a team of noted experts in the field, the text is divided into two main sections. The first section outlines the theoretical underpinnings of the topic and the second section is designed to include examples of practical applications. The book contains a wealth of illustrative cases of business failure prediction, zoology, ecology and others. This vital guide: Offers an important text that has been tested both in the classroom and at tutorials at conferences Contains authoritative information written by leading experts in the field Presents a comprehensive text that can be applied to courses in machine learning, data mining and artificial intelligence Combines in one volume two of the most intriguing topics in machine learning: ensemble learning and classification trees Written for researchers from many fields such as biostatistics, economics, environment, zoology, as well as students of data mining and machine learning, Ensemble Classification Methods with Applications in R puts the focus on two topics in machine learning: classification trees and ensemble learning.

This thesis seeks to establish if the use of negation in Inductive Rule Learning (IRL) for text

classification is effective. Text classification is a widely research topic in the domain of data mining. There have been many techniques directed at text classification; one of them is IRL, widely chosen because of its simplicity, comprehensibility and interpretability by humans. IRL is a process whereby rules in the form of $antecedent -> conclusion$ are learnt to build a classifier. Thus, the learnt classifier comprises a set of rules, which are used to perform classification. To learn a rule, words from pre-labelled documents, known as features, are selected to be used as conjunctions in the rule antecedent. These rules typically do not include any negated features in their antecedent; although in some cases, as demonstrated in this thesis, the inclusion of negation is required and beneficial for the text classification task. With respect to the use of negation in IRL, two issues need to be addressed: (i) the identification of the features to be negated and (ii) the improvisation of rule refinement strategies to generate rules both with and without negation. To address the first issue, feature space division is proposed, whereby the feature space containing features to be used for rule refinement is divided into three sub-spaces to facilitate the identification of the features which can be advantageously negated. To address the second issue, eight rule refinement strategies are proposed, which are able to generate both rules with and without negation. Typically, single keywords which are deemed significant to differentiate between classes are selected to be used in the text representation in the text classification task. Phrases have also been proposed because they are considered to be semantically richer than single keywords. Therefore, with respect to the work conducted in this thesis, three different types of phrases (n-gram phrases, keyphrases and fuzzy phrases) are extracted to be used as the text representation in addition to the use of single keywords. To establish the effectiveness of the use of negation in IRL, the eight proposed rule refinement strategies are compared with one another, using keywords and the three different types of phrases as the text representation, to determine whether the best strategy is one which generates rules with negation or without negation. Two types of classification tasks are conducted; binary classification and multi-class classification. The best strategy in the proposed IRL mechanism is compared to five existing text classification techniques with respect to binary classification: (i) the Sequential Minimal Optimization (SMO) algorithm, (ii) Naive Bayes (NB), (iii) JRip, (iv) OlexGreedy and (v) OlexGA from the Waikato Environment for Knowledge Analysis (WEKA) machine learning workbench. In the multi-class classification task, the proposed IRL mechanism is compared to the Total From Partial Classification (TFPC) algorithm. The datasets used in the experiments include three text datasets: 20 Newsgroups, Reuters-21578 and Small Animal Veterinary Surveillance Network (SAVSNET) datasets and five UCI Machine Learning Repository tabular datasets. The results obtained from the experiments showed that the strategies which generated rules with negation were more effective when the keyword representation was used and less prominent when the phrase representations were used. Strategies which generated rules with negation also performed better with respect to binary classification compared to multi-class classification. In comparison with the other machine learning techniques selected, the proposed IRL mechanism was shown to generally outperform all the compared techniques and was competitive with SMO.

An easy-to-understand guide to learn practical Machine Learning techniques with Mathematical foundations KEY FEATURES - A balanced combination of underlying mathematical theories & practical examples with Python code - Coverage of latest topics like multi-label classification, Text Mining, Doc2Vec, Word2Vec, XMeans clustering, unsupervised outlier detection, techniques to deploy ML models in production-grade systems with PMML, etc - Coverage of sufficient & relevant visualization techniques specific to any topic DESCRIPTION This book will be ideal for working professionals who want to learn Machine Learning from scratch. The first chapter will be an introductory chapter to make readers comfortable with the idea of Machine Learning and the required mathematical theories. There will be a balanced

combination of underlying mathematical theories corresponding to any Machine Learning topic and its implementation using Python. Most of the implementations will be based on 'scikit-learn,' but other Python libraries like 'Gensim' or 'PyTorch' will also be used for some topics like text analytics or deep learning. The book will be divided into chapters based on primary Machine Learning topics like Classification, Regression, Clustering, Deep Learning, Text Mining, etc. The book will also explain different techniques of putting Machine Learning models into production-grade systems using Big Data or Non-Big Data flavors and standards for exporting models. WHAT WILL YOU LEARN - Get familiar with practical concepts of Machine Learning from ground zero - Learn how to deploy Machine Learning models in production - Understand how to do "Data Science Storytelling" - Explore the latest topics in the current industry about Machine Learning WHO THIS BOOK IS FOR This book would be ideal for experienced Software Professionals who are trying to get into the field of Machine Learning. Anyone who wishes to Learn Machine Learning concepts and models in the production lifecycle. TABLE OF CONTENTS 1. Introduction to Machine Learning & Mathematical preliminaries 2. Classification 3. Regression 4. Clustering 5. Deep Learning & Neural Networks 6. Miscellaneous Unsupervised Learning 7. Text Mining 8. Machine Learning models in production 9. Case Studies & Data Science Storytelling

This book gathers the proceedings of the Seventh International Conference on Computational Science and Technology 2020 (ICCST 2020), held in Pattaya, Thailand, on 29–30 August 2020. The respective contributions offer practitioners and researchers a range of new computational techniques and solutions, identify emerging issues, and outline future research directions, while also showing them how to apply the latest large-scale, high-performance computational methods.

This book constitutes the refereed proceedings of the 8th International Conference on Advanced Data Mining and Applications, ADMA 2012, held in Nanjing, China, in December 2012. The 32 regular papers and 32 short papers presented in this volume were carefully reviewed and selected from 168 submissions. They are organized in topical sections named: social media mining; clustering; machine learning: algorithms and applications; classification; prediction, regression and recognition; optimization and approximation; mining time series and streaming data; Web mining and semantic analysis; data mining applications; search and retrieval; information recommendation and hiding; outlier detection; topic modeling; and data cube computing.

In the same way that it has become part of all our lives, computer technology is now integral to the work of the legal profession. The JURIX Foundation has been organizing annual international conferences in the area of computer science and law since 1988, and continues to support cutting-edge research and applications at the interface between law and computer technology. This book contains the 16 full papers and 6 short papers presented at the 26th International Conference on Legal Knowledge and Information Systems (JURIX 2013), held in December 2013 in Bologna, Italy. The papers cover a wide range of research topics and application areas concerning the advanced management of legal information and knowledge, including computational techniques for: classifying and extracting information from, and detecting conflicts in, regulatory texts; modeling legal argumentation and representing case narratives; improving the retrieval of legal information and extracting information from legal case texts; conducting e-discovery; and, applications involving intellectual property and IP licensing, online dispute resolution, delivering legal aid to the public and organizing the administration of local law and regulations. The book will be of interest to all those associated with the legal profession whose work involves the use of computer technology.

This book constitutes the refereed proceedings of the 12th International Conference of the Italian Association for Artificial Intelligence, AI*IA 2011, held in Palermo, Italy, in September 2011. The 31 revised full papers presented together with 3 invited talks and 13 posters were

carefully reviewed and selected from 58 submissions. The papers are organized in topical sections on machine learning; distributed AI: robotics and MAS; theoretical issues: knowledge representation and reasoning; planning, cognitive modeling; natural language processing; and AI applications.

Copyright: 18ebe5aec06e97b64e0f636296256252