

1. Record Nr.	UNISA996503470803316
Autore	Nguyen Ngoc Thanh
Titolo	Intelligent Information and Database Systems : 14th Asian Conference, ACIIDS 2022, Ho Chi Minh City, Vietnam, November 28-30, 2022, Proceedings, Part II
Pubbl/distr/stampa	Cham : , : Springer, , 2023 ©2022
ISBN	3-031-21967-8
Descrizione fisica	1 online resource (766 pages)
Collana	Lecture Notes in Computer Science ; ; v.13758
Altri autori (Persone)	TranTien Khoa TukayevUalsher HongTzung-Pei TrawiskiBogdan SzczerbickiEdward
Disciplina	006.3
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di contenuto	Intro -- Preface -- Organization -- Contents - Part II -- Contents - Part I -- Machine Learning and Data Mining -- Machine Learning or Lexicon Based Sentiment Analysis Techniques on Social Media Posts -- 1 Introduction -- 2 Literature Review -- 3 Methodology -- 3.1 Comparison Between Sentiment Analysis Methods -- 4 Results and Discussion -- 5 Conclusion -- References -- A Comparative Study of Classification and Clustering Methods from Text of Books -- 1 Introduction -- 2 Related Works -- 3 Project Gutenberg -- 4 Natural Language Processing -- 4.1 Word Weighting Measures -- 5 Machine Learning Methods -- 5.1 Algorithms for Classification -- 5.2 Algorithm for Clustering -- 5.3 Measures of the Quality -- 6 Proposed Approach -- 7 Experiments -- 7.1 Experimental Design and Data Set -- 7.2 Results of Experiments -- 8 Conclusions -- References -- A Lightweight and Efficient GA-Based Model-Agnostic Feature Selection Scheme for Time Series Forecasting -- 1 Introduction -- 2 Related Works -- 2.1 Feature Selection Methods -- 2.2 GA-Based Feature Selection -- 3 GA-Based Model-Agnostic Feature Selection -- 3.1

Problem Formulation -- 3.2 Overview -- 3.3 GA-Based Feature Selector -- 3.4 Training Data Generator -- 4 Performance Evaluation -- 4.1 Evaluation Settings -- 4.2 Impact of GA-Based Feature Selector -- 4.3 Impact of Training Data Generator -- 5 Conclusion -- References -- Machine Learning Approach to Predict Metastasis in Lung Cancer Based on Radiomic Features -- 1 Background -- 2 Materials and Methods -- 2.1 Data -- 2.2 Radiomics Features -- 2.3 Classification Workflow -- 3 Feature Selection Challenges -- 3.1 Multiple ROIs from the Same Patient -- 3.2 Response Variable Type -- 3.3 Small Differences Between Classes -- 4 Results -- 5 Discussion and Future Work -- References -- Covariance Controlled Bayesian Rose Trees -- 1 Introduction -- 2 Algorithm. 2.1 Hierarchical Clustering -- 2.2 Bayesian Rose Trees -- 2.3 Constraining BRT Hierarchies -- 2.4 Parameterisation -- 2.5 Depth Level as a Function of the Likelihood -- 2.6 Hierarchy Outside of Defined Clusters -- 3 Method Comparison -- 4 Conclusions -- References -- Potential of Radiomics Features for Predicting Time to Metastasis in NSCLC -- 1 Background -- 2 Materials and Methods -- 2.1 Data -- 2.2 Radiomics Features -- 2.3 Data Pre-processing and Unsupervised Analysis -- 2.4 Modeling of Metastasis Free Survival -- 3 Results -- 4 Discussion and Future Work -- References -- A Survey of Network Features for Machine Learning Algorithms to Detect Network Attacks -- 1 Introduction -- 2 Background Study -- 3 Literature Survey -- 4 Shortcoming of Existing Literature -- 5 Recommendations -- References -- The Quality of Clustering Data Containing Outliers -- 1 Introduction -- 1.1 The Structure of the Paper -- 2 State of Art -- 3 Clustering Data Containing Outliers -- 3.1 Clustering Algorithms: Hierarchical AHC vs Partitional K-Means -- 3.2 Clustering Quality Indices -- 3.3 Outlier Definition -- 3.4 Outlier Detection Algorithms -- 4 Experiments -- 4.1 Data Description -- 4.2 Methodology -- 4.3 Experimental Environment -- 4.4 Results -- 4.5 Discussion -- 5 Summary -- References -- Aggregated Performance Measures for Multi-class Classification -- 1 Introduction -- 2 Method -- 2.1 Classification of a Single Data Point -- 2.2 Aggregation Over Classes and Thresholds -- 2.3 Normalisation -- 2.4 The Case of Specificity -- 2.5 The Compound Measure of Accuracy -- 3 Discussion -- References -- Prediction of Lung Cancer Survival Based on Multiomic Data -- 1 Introduction -- 2 Materials and Methods -- 2.1 Data Used in the Study -- 2.2 Feature Definition and Pre-selection -- 2.3 Variable Importance Study -- 2.4 Classification of Data -- 3 Results. 3.1 Aggregation and Dimensionality Reduction -- 3.2 Predictive Potential of Various -Omics Datasets -- 3.3 Variable Importance Study in a Multiomic Dataset -- 4 Discussion -- References -- Graph Neural Networks-Based Multilabel Classification of Citation Network -- 1 Introduction -- 2 Related Works -- 3 Dataset Description -- 4 Experiments -- 5 Multilabel Classification Approach -- 6 Conclusion and Future Works -- References -- Towards Efficient Discovery of Partial Periodic Patterns in Columnar Temporal Databases -- 1 Introduction -- 2 Related Work -- 3 The Model of Partial Periodic Pattern -- 4 Proposed Algorithm -- 4.1 3P-ECLAT Algorithm -- 5 Experimental Results -- 5.1 Evaluation of Algorithms by Varying minPS -- 5.2 Evaluation of Algorithms by Varying Per -- 5.3 Scalability Test -- 5.4 A Case Study: Finding Areas Where People Have Been Regularly Exposed to Hazardous Levels of PM2.5 Pollutant -- 6 Conclusions and Future Work -- References -- Avoiding Time Series Prediction Disbelief with Ensemble Classifiers in Multi-class Problem Spaces -- 1 Introduction -- 2 Time Series Analysis Life-Cycle -- 3 Prediction Disbelief in Acceptance Tests of Forecasting Models -- 4 Discussion --

5 Conclusions -- References -- Speeding Up Recommender Systems Using Association Rules -- 1 Introduction -- 2 Preliminaries -- 2.1 Factorization Machines -- 2.2 Association Rules -- 2.3 Related Works -- 3 FMAR Recommender System -- 3.1 Problem Definition -- 3.2 Factorization Machine Apriori Based Model -- 3.3 Factorization Machine FP-Growth Based Model -- 4 Evaluation for FMAR -- 4.1 Performance Comparison and Analysis -- 5 Conclusions and Future Work -- References -- An Empirical Experiment on Feature Extractions Based for Speech Emotion Recognition -- 1 Introduction -- 2 Literature Review -- 3 Dataset -- 4 Feature Extraction -- 5 Methodology -- 5.1 Input Preparation.

5.2 Classification Models -- 6 Experimental Results -- 7 Conclusion and Discussion -- References -- Parameter Distribution Ensemble Learning for Sudden Concept Drift Detection -- 1 Introduction -- 2 Methods -- 2.1 BO-ERICS Phase -- 2.2 Ensemble Phase -- 3 Experiments and Discussion -- 3.1 Datasets -- 3.2 Evaluation -- 3.3 Results -- 3.4 Discussion -- 4 Conclusions -- References -- MLP-Mixer Approach for Corn Leaf Diseases Classification -- 1 Introduction -- 2 Related Work -- 2.1 Literature Review -- 2.2 MLP-Mixer -- 2.3 Deep Learning -- 3 Methods -- 3.1 Data Requirements, Collection and Preparation -- 3.2 Configure the Hyperparameters -- 3.3 Build a Classification Model -- 3.4 Define an Experiment and Data Augmentation -- 3.5 The MLP-Mixer Model Structure -- 3.6 Build, Train, and Evaluate the MLP-Mixer Model -- 4 Experiment and Result -- 4.1 Image Segmentation -- 4.2 Experiment Results (Train and Evaluate Model) -- 4.3 Discussion -- 5 Conclusion -- References -- A Novel Neural Network Training Method for Autonomous Driving Using Semi-Pseudo-Labels and 3D Data Augmentations -- 1 Introduction -- 2 Related Work -- 3 A Novel Training Method with Semi-Pseudo-Labeling and 3D Augmentations -- 3.1 Semi-Pseudo-Labeling -- 3.2 3D Augmentations -- 3.3 An Example of Training with Semi-Pseudo-Labeling and 3D Augmentations -- 4 Experiments -- 4.1 Argoverse -- 4.2 In-House Highway Dataset -- 5 Conclusion -- References -- Machine Learning Methods for BIM Data -- 1 Introduction -- 2 BIM Data - IFC Files -- 3 Machine Learning Techniques for BIM -- 3.1 Learning Semantic Information - Space Classification -- 3.2 Semantic Enrichment of BIM Models from Point Clouds -- 3.3 Building Condition Diagnosis -- 3.4 BIM Enhancement in the Facility Management Context -- 3.5 Knowledge Extraction from BIM -- 4 Conclusions -- References.

Self-Optimizing Neural Network in Classification of Real Valued Experimental Data -- 1 Introduction -- 2 Self Optimizing Neural Network -- 2.1 SONN Formalism -- 2.2 Fundamental Coefficient of Discrimination -- 2.3 Structure of the Network and the Weight Factor -- 2.4 Network Response -- 3 Experiment and Results -- 3.1 Dataset -- 3.2 Data Preparation -- 3.3 Classification -- 4 Conclusion -- References -- Analyzing the Effectiveness of the Gaussian Mixture Model Clustering Algorithm in Software Enhancement Effort Estimation -- 1 Introduction -- 2 Backgrounds -- 2.1 The FPA Overview -- 2.2 The Gaussian Mixture Model Clustering Algorithm -- 2.3 The k-means Clustering Algorithm -- 3 Research Methodology -- 3.1 Dataset Pre-processing -- 3.2 Determine the Number of Clusters -- 3.3 Evaluation Criteria -- 4 Results and Discussions -- 5 Conclusion -- References -- Graph Classification via Graph Structure Learning -- 1 Introduction -- 2 Related Works -- 3 Proposed Method: GC-GSL -- 3.1 Extracting Topological Attribute Vector -- 3.2 Rooted Subgraph Mining -- 3.3 Neural Network Graph Embedding -- 3.4 Computational Complexity -- 4 Experiments -- 4.1 Results -- 4.2 Discussions -- 5 Conclusion --

References -- Relearning Ensemble Selection Based on New Generated Features -- 1 Introduction -- 2 Related Works -- 3 The Proposed Framework -- 3.1 Generation of Diverse Base Classifiers -- 3.2 Relearning Base Classifiers -- 3.3 Feature Generation Based on Learned and Relearned Base Classifiers -- 3.4 Learning Second-Level Base Classifier Based on New Vector of the Features -- 3.5 Selection Base Classifiers Based on Second-Level Classification Result -- 4 Experiments -- 4.1 Experimental Setup -- 4.2 Results -- 5 Discussion -- 6 Conclusions -- References -- Random Forest in Whitelist-Based ATM Security -- 1 Introduction -- 2 Related Work -- 3 Test Procedure. 4 Data Pre-processing.

---

Sommario/riassunto

This book constitutes the refereed proceedings of the 14th Asian Conference on Intelligent Information and Database Systems, ACIIDS 2022, held Ho Chi Minh City, Vietnam in November 2022. The 113 full papers accepted for publication in these proceedings were carefully reviewed and selected from 406 submissions. The papers of the 2 volume-set are organized in the following topical sections: data mining and machine learning methods, advanced data mining techniques and applications, intelligent and contextual systems, natural language processing, network systems and applications, computational imaging and vision, decision support and control systems, and data modeling and processing for industry 4.0. The accepted and presented papers focus on new trends and challenges facing the intelligent information and database systems community.

---