

1. Record Nr.	UNISA996546854503316
Autore	Wrembel Robert
Titolo	Big Data Analytics and Knowledge Discovery [[electronic resource]] : 25th International Conference, DaWaK 2023, Penang, Malaysia, August 28–30, 2023, Proceedings // edited by Robert Wrembel, Johann Gamper, Gabriele Kotsis, A Min Tjoa, Ismail Khalil
Pubbl/distr/stampa	Cham : , : Springer Nature Switzerland : , : Imprint : Springer, , 2023
ISBN	3-031-39831-9
Edizione	[1st ed. 2023.]
Descrizione fisica	1 online resource (407 pages)
Collana	Lecture Notes in Computer Science, , 1611-3349 ; ; 14148
Altri autori (Persone)	GamperJohann KotsisGabriele TjoaA. Min KhalilIsmail
Disciplina	001.422 005.7
Soggetti	Quantitative research Data mining Application software Artificial intelligence Data Analysis and Big Data Data Mining and Knowledge Discovery Computer and Information Systems Applications Artificial Intelligence
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di contenuto	Intro -- Preface -- Organization -- From an Interpretable Predictive Model to a Model Agnostic Explanation (Abstract of Keynote Talk) -- Contents -- Data Quality -- Using Ontologies as Context for Data Warehouse Quality Assessment -- 1 Introduction -- 2 Related Work -- 3 Preliminaries -- 3.1 Running Example -- 3.2 Data Warehouse Formal Specification -- 3.3 Context Formal Specification -- 4 Data Warehouse to Ontology Mapping -- 5 Context-Based Data Quality Rules -- 6 Experimentation -- 6.1 Implementation -- 6.2 Validation -- 7 Conclusions and Future Work -- References -- Preventing Technical

Errors in Data Lake Analyses with Type Theory -- 1 Introduction -- 2 Related Works -- 3 Type-Theoretical Framework -- 4 Conclusion -- References -- EXOS: Explaining Outliers in Data Streams -- 1 Introduction -- 2 Related Work -- 3 Preliminaries -- 4 The Proposed Algorithm: EXOS -- 4.1 Estimator -- 4.2 Temporal Neighbor Clustering -- 4.3 Outlying Attribute Generators -- 5 Evaluation -- 5.1 Experimental Setup -- 5.2 Results and Analysis -- 6 Conclusions -- References -- Motif Alignment for Time Series Data Augmentation -- 1 Introduction -- 2 Preliminaries -- 2.1 Matrix Profile -- 2.2 Pan-Matrix Profile -- 2.3 DTW Alignment for Time Series Data Augmentation -- 3 Proposed Method -- 3.1 Motif Mapping -- 3.2 Time Series Augmentation -- 4 Experimental Evaluation -- 4.1 Setup -- 4.2 Aligning Time Series Using MotifDTW -- 4.3 Performance Gain -- 5 Conclusion -- References -- State-Transition-Aware Anomaly Detection Under Concept Drifts -- 1 Introduction -- 2 Related Works -- 3 Problem Definition -- 3.1 Terminology -- 3.2 Problem Statement -- 4 State-Transition-Aware Anomaly Detection -- 4.1 Reconstruction and Latent Representation Learning -- 4.2 Drift Detection in the Latent Space -- 4.3 State Transition Model -- 5 Experiment -- 5.1 Experiment Setup -- 5.2 Performance. 6 Conclusion -- References -- Anomaly Detection in Financial Transactions Via Graph-Based Feature Aggregations -- 1 Introduction -- 2 Related Work -- 2.1 Graph Embedding -- 2.2 Anomaly Detection -- 3 Problem Formalization -- 4 Proposed Method -- 4.1 PFA: Proximal Feature Aggregation -- 4.2 AFA: Anomaly Feature Aggregation -- 5 Experiment -- 5.1 Experimental Setup -- 5.2 Effectiveness Evaluation -- 5.3 Scalability Evaluation -- 6 Conclusion -- References -- The Synergies of Context and Data Aging in Recommendations -- 1 Introduction -- 2 ALBA: Adding Aging to LookBack Apriori -- 3 Context Modeling -- 4 Evaluation -- 4.1 Contexts -- 4.2 Methodology -- 4.3 Fitbit Validation -- 4.4 Auditel Validation -- 5 Conclusions and Future Work -- References -- Advanced Analytics and Pattern Discovery -- Hypergraph Embedding Based on Random Walk with Adjusted Transition Probabilities -- 1 Introduction -- 2 Related Work -- 3 Preliminaries -- 3.1 Notation -- 3.2 Hypergraph Projection -- 3.3 Random Walk and Stationary Distribution -- 3.4 Skip-Gram -- 4 Proposed Method -- 4.1 Random Walk -- 5 Experiment -- 5.1 Transition Probabilities in Steady State -- 5.2 Node Label Estimation -- 5.3 Parameter Dependence of F1 Score -- 6 Conclusion -- References -- Contextual Shift Method (CSM) -- 1 Introduction -- 2 Contextual Shifts -- 3 Contextual Shift Method -- 4 Experiments -- 5 Conclusion -- References -- Utility-Oriented Gradual Itemsets Mining Using High Utility Itemsets Mining -- 1 Introduction -- 2 Preliminary Definitions -- 3 High Utility Gradual Itemsets Mining -- 3.1 Database Encoding -- 3.2 High Utility Gradual Itemsets Extraction -- 4 Experimental Study -- 5 Conclusion -- References -- Discovery of Contrast Itemset with Statistical Background Between Two Continuous Variables -- 1 Introduction -- 2 Contrast ItemSB -- 3 Experimental Results -- 4 Conclusions -- References. DBGAN: A Data Balancing Generative Adversarial Network for Mobility Pattern Recognition -- 1 Introduction -- 2 Related Work -- 3 Background -- 3.1 Reproducing Kernel Hilbert Space Embeddings -- 3.2 Attention Mechanism -- 3.3 Generative Adversarial Network -- 4 DBGAN Mobility Pattern Classification Model -- 4.1 Attributes of Travel Trajectories Utilized for Classification -- 4.2 Sequences to Images with Kernel Embedding -- 4.3 Classification Using Self Attention-Based Generative Adversarial Network -- 5 Evaluation -- 6 Conclusion -- References -- Bitwise Vertical Mining of Minimal Rare Patterns -- 1

Introduction -- 2 Background and Related Works -- 3 Our RP-VIPER Algorithm -- 4 Evaluation -- 5 Conclusions -- References -- Inter-item Time Intervals in Sequential Patterns -- 1 Introduction -- 2 Related Work -- 3 Representing Time in Sequences -- 3.1 Preliminaries -- 3.2 Integrating Intervals in Sequences -- 4 Experiments -- 4.1 Datasets and Models -- 4.2 Results -- 5 Conclusion -- References -- Fair-DSP: Fair Dynamic Survival Prediction on Longitudinal Electronic Health Record -- 1 Introduction -- 2 Related Work -- 3 Methodology -- 3.1 Fair Dynamic Survival Model -- 3.2 Individual Fairness -- 3.3 Group Fairness -- 4 Experiments -- 4.1 Quantitative Analysis -- 4.2 Sensitivity Study -- 5 Conclusions -- References -- Machine Learning -- DAT@Z21: A Comprehensive Multimodal Dataset for Rumor Classification in Microblogs -- 1 Introduction -- 2 Related Works -- 2.1 Fake Health News Datasets -- 2.2 Fake News Datasets -- 3 Data Collection -- 3.1 News Articles and Ground Truth Collection -- 3.2 Preparing the Tweets Collection -- 3.3 Tweets Collection -- 4 Rumor Classification Using DAT@Z21 -- 4.1 Baselines -- 4.2 Experiment Settings -- 4.3 Experimental Results -- 5 Conclusion and Perspectives -- References.

Dealing with Data Bias in Classification: Can Generated Data Ensure Representation and Fairness? -- 1 Introduction -- 2 Related Work -- 3 Measuring Discrimination -- 4 Problem Formulation -- 5 Methodology -- 6 Evaluation -- 6.1 Comparing Pre-processors -- 6.2 Investigating the Fairness-Agnostic Property -- 7 Conclusion -- 8 Discussion and Future Work -- A Proof of Time Complexity -- References -- Random Hypergraph Model Preserving Two-Mode Clustering Coefficient -- 1 Introduction -- 2 Preliminaries -- 3 Extending the Hyper dK-Series to the Case of $d_v = 2.5+$ -- 4 Experiments -- 5 Conclusion -- References -- A Non-overlapping Community Detection Approach Based on -Structural Similarity -- 1 Introduction -- 2 Preliminaries -- 3 A Hierarchical Clustering Approach Based on -Structural Similarity -- 4 Experiments -- 5 Conclusion and Future Work -- A Appendix a -- B Appendix B -- References -- Improving Stochastic Gradient Descent Initializing with Data Summarization -- 1 Introduction -- 2 Definitions -- 2.1 Input Data Set -- 2.2 LR Model -- 3 System and Algorithms -- 3.1 Gamma Summarization () -- 3.2 Mini-batch SGD -- 3.3 Mini-batch SGD Initialization Using Gamma -- 4 Experiments -- 4.1 Experimental Setup -- 4.2 Experimental Results -- 5 Related Work -- 6 Conclusions -- References -- Feature Analysis of Regional Behavioral Facilitation Information Based on Source Location and Target People in Disaster -- 1 Introduction -- 2 Related Work -- 3 Basic Concept of RBF Tweet Classification -- 3.1 Extraction of BF Tweets -- 3.2 RBF Tweet Extraction and Classification -- 4 Analysis of RBF Tweets -- 4.1 Training and Test Data -- 4.2 Research Question -- 4.3 Results and Discussion of Research Questions -- 5 Conclusion -- References -- Exploring Dialog Act Recognition in Open Domain Conversational Agents -- 1 Introduction -- 2 Related Works.

3 Proposed Dialog Act Taxonomy -- 3.1 Data Sources -- 4 Proposed Dialog Act Classifier -- 4.1 Experimental Setup -- 4.2 Performance Evaluation -- 4.3 Generalizability of Model -- 5 Conclusion -- References -- UniCausal: Unified Benchmark and Repository for Causal Text Mining -- 1 Introduction -- 2 Related Work -- 2.1 Tasks -- 2.2 Datasets -- 2.3 Other Large Causal Resources -- 3 Methodology -- 3.1 Creation of UniCausal -- 3.2 Baseline Model -- 4 Experiments -- 4.1 Baseline Performance -- 4.2 Impact of Datasets -- 4.3 Adding CauseNet to Investigate the Importance of Linguistic Variation in Examples -- 5 Conclusion -- References -- Deep Learning -- Accounting for Imputation Uncertainty During Neural Network Training

-- 1 Introduction -- 2 Related Works -- 3 Contributions -- 3.1 Single-Hotpatching -- 3.2 Multiple-Hotpatching -- 4 Experiments -- 4.1 Experimental Protocol -- 4.2 Results -- 5 Discussion and Conclusion -- References -- Supervised Hybrid Model for Rumor Classification: A Comparative Study of Machine and Deep Learning Approaches -- 1 Introduction -- 2 Related Work -- 3 Datasets and Preprocessing -- 4 Implementation -- 4.1 Traditional ML Approaches -- 4.2 DL Approaches -- 4.3 The Ensemble Stack ML Model -- 4.4 The Hybrid ML-DL Model -- 5 Results and Analysis -- 6 Conclusion and Future Work -- References -- Attention-Based Counterfactual Explanation for Multivariate Time Series -- 1 Introduction -- 2 Related Work -- 3 Methodology -- 3.1 Notation -- 3.2 Proposed Method -- 4 Experiments -- 4.1 Datasets -- 4.2 Baseline Methods -- 4.3 Experimental Result -- 5 Conclusion -- References -- DRUM: A Real Time Detector for Regime Shifts in Data Streams via an Unsupervised, Multivariate Framework -- 1 Introduction -- 2 Related Work -- 3 DRUM -- 4 Evaluation -- 5 Conclusion -- References. Hierarchical Graph Neural Network with Cross-Attention for Cross-Device User Matching.

Sommario/riassunto

This book constitutes the proceedings of the 25th International Conference on Big Data Analytics and Knowledge Discovery, DaWaK 2023, which took place in Penang, Malaysia, during August 29-30, 2023. The 18 full papers presented together with 19 short papers were carefully reviewed and selected from a total of 83 submissions. They were organized in topical sections as follows: Data quality; advanced analytics and pattern discovery; machine learning; deep learning; and data management.
