1. Record Nr. UNINA9910633935503321

Titolo	Data mining : 20th Australasian Conference, AusDM 2022, Western Sydney, Australia, December 12-15, 2022, proceedings / / edited by Laurence A. F. Park
Pubbl/distr/stampa	Singapore : , : Springer, , [2023] ©2023
ISBN	981-19-8746-7
Descrizione fisica	1 online resource (256 pages)
Collana	Communications in Computer and Information Science ; ; v.1741
Disciplina	006.312
Soggetti	Data mining
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Note generali	Includes index.
Nota di contenuto	Intro -- Preface -- Organization -- Contents -- Research Track -- Measuring Content Preservation in Textual Style Transfer -- 1 Introduction -- 2 Background and Motivation -- 2.1 Cosine Similarity -- 2.2 Disentanglement of Style and Content -- 2.3 The Style Invariant Embedding Assumption -- 3 Experiment -- 3.1 Dataset -- 3.2 Procedure -- 4 Results and Discussion -- 5 Conclusion -- References -- A Temperature-Modified Dynamic Embedded Topic Model -- 1 Introduction -- 2 Related Work -- 3 Methodology -- 3.1 The Dynamic Embedded Topic Model -- 3.2 The Proposed Approach: DETM-tau -- 4 Experiments and Results -- 4.1 Experimental Set-Up -- 4.2 Results -- 5 Conclusion -- References -- Measuring Difficulty of Learning Using Ensemble Methods -- 1 Introduction -- 2 Related Work -- 3 Instance Difficulty -- 3.1 Difficulty Measures -- 4 Experiments -- 5 Conclusion -- References -- Graph Embeddings for Non-IID Data Feature Representation Learning -- 1 Introduction -- 2 Background and Related Work -- 2.1 Classification Models and IID Assumption -- 2.2 Graph and Knowledge Graph Embeddings -- 2.3 Summary -- 3 Methodology -- 4 Dataset and Experiment Design -- 4.1 Dataset -- 4.2 Experiment Design -- 5 Results and Discussions -- 5.1 Imbalanced Data -- 5.2 Advantage of Using the Node2vec Embeddings -- 5.3 Evaluation and Discussion -- 6 Conclusions and Future Work -- 6.1 Traffic -- 6.2 Learn Feature Representation for Non-IID Data via Graph

Embeddings -- 6.3 Future Work -- References -- Enhancing Understandability of Omics Data with SHAP, Embedding Projections and Interactive Visualisations -- 1 Introduction -- 2 Framework for Using SHAP to Optimise UMAP and PCA Input Data -- 2.1 Initial Visualisations from the UMAP and PCA Projection Methods -- 2.2 Explainable Machine Learning SHAP for Important Feature Selection -- 2.3 Final Optimised Visualisations.
3 UMAP and PCA Visualisations -- 3.1 Datasets -- 3.2 How Do PCA, UMAP, and SHAP Work? -- 3.3 Similar Projection and Clustering Patterns Between PCA and UMAP -- 4 Rank and Select the Most Important Features with SHAP -- 5 Validation of the SHAP Results -- 6 Conclusion and Future Work -- References -- WinDrift: Early Detection of Concept Drift Using Corresponding and Hierarchical Time Windows -- 1 Introduction -- 2 Preliminaries -- 3 The WinDrift (WD) Method -- 3.1 Key Components -- 3.2 Step-by-Step Description -- 4 Experimental Results -- 4.1 Experimental Setup -- 4.2 Datasets -- 4.3 Numerical Results -- 5 Conclusion and Future Work -- References -- Investigation of Explainability Techniques for Multimodal Transformers -- 1 Introduction -- 2 Problem Definition -- 2.1 Quantifying Syntactic Grounding Through Label Attribution -- 2.2 Investigating Semantic Relationships Through Optimal Transport -- 3 Explainability Techniques -- 3.1 Label Attribution -- 3.2 Optimal Transport -- 4 A Case Study in VisualBERT Explainability -- 5 Conclusion -- References -- Effective Imbalance Learning Utilizing Informative Data -- 1 Introduction -- 2 Related Work -- 2.1 Sampling Method -- 2.2 Cost-Sensitive Methods -- 2.3 Ensemble Methods -- 2.4 Data Representation -- 3 Proposed Framework and Approach -- 3.1 Informative Samples Located -- 3.2 Extracting Information -- 3.3 Model Test -- 4 Experiments and Results -- 4.1 Results on General Test -- 4.2 Results on Different Reference Data -- 5 Conclusion -- References -- Interpretable Decisions Trees via Human-in-the-Loop-Learning -- 1 Introduction -- 2 Learning Classifiers Involving Dataset Visualisations -- 3 Experts Iteratively Construct Decision Trees -- 3.1 Using Parallel Coordinates -- 3.2 The Splits the User Shall Apply -- 3.3 Information that Supports Interaction -- 3.4 Visualising the Tree -- 3.5 Visualising Rules.
4 Design of the Usability Evaluation -- 5 Results -- 5.1 Validity Threats -- 6 Conclusion -- References -- Application Track -- A Comparative Look at the Resilience of Discriminative and Generative Classifiers to Missing Data in Longitudinal Datasets -- 1 Introduction -- 2 Background and Related Work -- 3 LoGAN: A GAN Based Longitudinal Classifier for Missing Data -- 3.1 The LoGAN Approach -- 4 Experiments -- 4.1 Dataset -- 4.2 Baseline Models -- 4.3 Experimental Setup and Model Design -- 4.4 Evaluation Criteria -- 5 Results and Discussion -- 5.1 Training Performance by Model Setting -- 5.2 Performance on Balanced Data -- 5.3 Performance on Imbalanced Data -- 6 Final Remarks -- 7 Conclusions -- References -- Hierarchical Topic Model Inference by Community Discovery on Word Co-occurrence Networks -- 1 Introduction -- 2 Related Work -- 3 Community Topic -- 3.1 Co-occurrence Network Construction -- 3.2 Community Mining -- 3.3 Topic Filtering and Term Ordering -- 3.4 Topic Hierarchy -- 4 Empirical Evaluation -- 4.1 Datasets -- 4.2 Preprocessing -- 4.3 Evaluation Metrics -- 5 Results -- 6 Conclusion -- References -- UMLS-Based Question-Answering Approach for Automatic Initial Frailty Assessment -- 1 Introduction -- 2 Related Work -- 3 The Proposed Approach -- 3.1 Discovery of UMLS Based Concepts -- 3.2 UMLS-Based Concept Selection Algorithm -- 3.3 Answering TFI Questionnaire Using UMLS-Based Concepts -- 3.4 Frailty

Assessment -- 4 Experiment and Results -- 4.1 Dataset -- 4.2 Experiment Settings -- 4.3 Results -- 5 Discussion -- 6 Conclusion -- References -- Natural Language Query for Technical Knowledge Graph Navigation -- 1 Introduction -- 2 Related Work -- 3 Approach -- 4 Application -- 4.1 Overview of Maintenance KG -- 4.2 Neural Named Entity Recognition Ensemble -- 5 Results and Discussion -- 5.1 Ensemble NER Performance Analysis.
5.2 Question Types Discussion -- 6 Conclusions and Future Work -- References -- Decomposition of Service Level Encoding for Anomaly Detection -- 1 Introduction -- 2 Algorithm and Notations -- 2.1 Input/Output Spaces -- 2.2 Algorithm -- 2.3 Defining Interval Sub Extreme (SE) -- 3 Results -- 3.1 Physiotherapy Service Levels -- 3.2 General Practitioner Service Levels -- 3.3 Psychiatric Service Levels -- 3.4 Discipline Comparison -- 4 Summary and Further Work -- References -- Improving Ads-Profitability Using Traffic-Fingerprints -- 1 Introduction -- 2 Algorithm -- 2.1 Step 1 - Clustering of Domains -- 2.2 Step 2 - Creating Blocking Rules -- 2.3 Step 3 - Reassigning Domains to Clusters -- 3 Offline Experiments -- 4 Online Experiments -- 5 Conclusions -- References -- Attractiveness Analysis for Health Claims on Food Packages -- 1 Introduction -- 2 Related Work -- 3 Consumer Preference Prediction of Health Claims -- 3.1 Dataset Collection -- 3.2 Prediction Model -- 4 Evaluation and Results -- 5 Case Studies -- 5.1 Specialised Terminology Factors -- 5.2 Sentiment and Metaphoricity Factors -- 6 The Deployment of the Proposed Attractiveness Analysis Model -- 7 Conclusion -- References -- SchemaDB: A Dataset for Structures in Relational Data -- 1 Introduction -- 1.1 Existing Datasets -- 1.2 Challenges of Flat Data -- 2 Dataset Curation -- 2.1 Collection and Filtration -- 2.2 Graph Transform and Canonisation -- 2.3 Heuristic Augmentation -- 3 Analytics -- 3.1 Summary Statistics -- 4 Research Potential and Applications -- 5 Conclusion -- References -- Author Index.