

1. Record Nr.	UNINA9910872194103321
Autore	Longo Luca
Titolo	Explainable Artificial Intelligence : Second World Conference, XAI 2024, Valletta, Malta, July 17-19, 2024, Proceedings, Part II
Pubbl/distr/stampa	Cham : , : Springer International Publishing AG, , 2024 ©2024
ISBN	3-031-63797-6
Edizione	[1st ed.]
Descrizione fisica	1 online resource (0 pages)
Collana	Communications in Computer and Information Science Series ; ; v.2154
Altri autori (Persone)	LapuschkinSebastian SeifertChristin
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di contenuto	Intro -- Preface -- Organization -- Contents - Part II -- XAI for Graphs and Computer Vision -- Model-Agnostic Knowledge Graph Embedding Explanations for Recommender Systems -- 1 Introduction -- 1.1 Problem Setting and Objective -- 1.2 Overview of Main Findings and Contributions -- 2 Related Work -- 3 Methodology -- 3.1 Baselines -- 3.2 Metrics -- 3.3 Multi-Attribute Utility Theory (MAUT) -- 4 Embedding Proposal Algorithm -- 5 Results -- 6 Future Work, Limitations, and Open Directions -- 7 Conclusion -- References -- Graph-Based Interface for Explanations by Examples in Recommender Systems: A User Study -- 1 Introduction -- 2 Related Work -- 3 Generating Graph-Based Explanations-by-Examples -- 3.1 Obtaining the Explanation Links -- 3.2 Visualising Explanations Through Interactive Graphs -- 4 User Study -- 4.1 Methodology -- 4.2 User Study Results -- 5 Conclusions -- References -- Explainable AI for Mixed Data Clustering -- 1 Introduction -- 2 Related Work -- 3 Entropy-Based Cluster Explanations for Mixed Data -- 4 Evaluation -- 5 Discussion -- 6 Conclusion -- References -- Explaining Graph Classifiers by Unsupervised Node Relevance Attribution -- 1 Introduction -- 2 Preliminaries -- 2.1 Deep Graph Networks -- 2.2 XAI Methods for Graphs -- 2.3 Benchmarking XAI Methods for Graphs -- 2.4 Assessing XAI Methods for Graphs in Real-World Contexts -- 3 Methods -- 3.1 Unsupervised Attribution Strategies for the Realistic

Setting -- 3.2 A Measure of Quality for the Relevance Assignment -- 4 Experiments -- 4.1 Objective -- 4.2 Experimental Details -- 5 Results -- 6 Conclusion -- References -- Explaining Clustering of Ecological Momentary Assessment Data Through Temporal and Feature Attention -- 1 Introduction -- 2 Related Work -- 2.1 Clusters' Descriptive Representation -- 2.2 Explanations on TS Clustering -- 3 Review on Challenges of Explaining MTS Data.

3.1 Clustering Explanations -- 4 Framework for Clustering Explanations -- 4.1 EMA Data -- 4.2 EMA Clustering -- 4.3 Proposed Framework for Clustering Explanations -- 5 Analysis and Results -- 5.1 Performance Evaluation -- 5.2 Cluster-Level Explanations Through Temporal-Attention -- 5.3 Cluster-Level Explanations Through Feature-Attention -- 5.4 Individual-Level Explanations -- 6 Discussion -- 6.1 The Role of the Multi-aspect Attention -- 6.2 The Role of the Multi-level Analysis -- 6.3 The Impact of Utilizing a Pre-defined Clustering Result -- 7 Conclusion -- References -- Graph Edits for Counterfactual Explanations: A Comparative Study -- 1 Introduction -- 2 Related Work -- 3 Method -- 3.1 The Importance of Graph Machine Learning -- 4 Experiments -- 4.1 Experimental Setup -- 4.2 Quantitative Results -- 4.3 Qualitative Results -- 5 Conclusion -- References -- Model Guidance via Explanations Turns Image Classifiers into Segmentation Models -- 1 Introduction -- 1.1 Relation to Previous Works -- 1.2 Limitations -- 2 Unrolled Heatmap Architectures -- 2.1 LRP Basics -- 2.2 Unrolled LRP Architectures for Convolutional Classifiers -- 2.3 Losses and Training -- 2.4 Relation to Previous Formal Analyses and Standard Architectures -- 3 Unrolled Heatmap Architectures for Segmentation: Results -- 4 Conclusion -- References -- Understanding the Dependence of Perception Model Competency on Regions in an Image -- 1 Importance of Understanding Model Competency -- 2 Background and Related Work -- 2.1 Uncertainty Quantification -- 2.2 Out-of-Distribution (OOD) Detection -- 2.3 Explainable Image Classification -- 2.4 Explainable Competency Estimation -- 3 Approach for Understanding Model Competency -- 3.1 Estimating Model Competency -- 3.2 Identifying Regions Contributing to Low Competency -- 4 Method Evaluation and Analysis -- 4.1 Metrics for Comparison.

4.2 Dataset 1: Lunar Environment -- 4.3 Dataset 2: Speed Limit Signs -- 4.4 Dataset 3: Outdoor Park -- 4.5 Analysis of Results -- 5 Conclusions -- 6 Limitations and Future Work -- A Data Sources and Algorithm Parameters -- B Comparison to Class Activation Maps -- B.1 Dataset 1: Lunar Environment -- B.2 Dataset 2: Speed Limit Signs -- B.3 Dataset 3: Outdoor Park -- References -- A Guided Tour of Post-hoc XAI Techniques in Image Segmentation -- 1 Introduction -- 2 Categorization of XAI Techniques for Image Segmentation -- 3 Review of XAI Techniques for Image Segmentation -- 3.1 Local XAI -- 3.2 Evaluation of Local XAI Methods -- 3.3 A Comparative Analysis of Local XAI Methods -- 3.4 Global XAI -- 4 Tools for Practitioners -- 5 Discussion -- 6 Conclusion -- A Reviewed XAI Algorithms -- References -- Explainable Emotion Decoding for Human and Computer Vision -- 1 Introduction -- 2 Related Works -- 2.1 Explainable Computer Vision -- 2.2 Brain Decoding: Machine Learning on fMRI Data -- 2.3 Emotion Decoding for Human and Computer Vision -- 3 Experimental Setup -- 3.1 Frames, fMRI and Labels -- 3.2 Machine Learning on Movie Frames -- 3.3 Machine Learning on fMRI Data -- 3.4 XAI for Emotion Decoding -- 3.5 CNN-Humans Attentional Match: A Comparative Analysis -- 4 Experimental Results -- 4.1 Machine Learning on Movie Frames -- 4.2 Machine Learning on fMRI Data -- 4.3 Explainability for fMRI-Based Models -- 4.4 Comparative Analysis -- 5

Conclusions -- References -- Explainable Concept Mappings of MRI: Revealing the Mechanisms Underlying Deep Learning-Based Brain Disease Classification -- 1 Introduction -- 2 Methodology -- 2.1 Dataset -- 2.2 Preprocessing -- 2.3 Standard Classification Network -- 2.4 Relevance-Guided Classification Network -- 2.5 Training -- 2.6 Model Selection -- 2.7 Bootstrapping Analysis -- 2.8 Concepts Identification.

2.9 Relevance-Weighted Concept Map Representation -- 2.10 R2* and Relevance Region of Interest Analysis -- 3 Results -- 4 Discussion -- 5 Conclusion -- References -- Logic, Reasoning, and Rule-Based Explainable AI -- Template Decision Diagrams for Meta Control and Explainability -- 1 Introduction -- 2 Related Work -- 3 Foundations -- 4 Template Decision Diagrams -- 4.1 Hierarchical Decision Diagrams by Templates -- 4.2 Standard Template Boxes -- 5 Templates for Self-adaptive Systems and Meta Control -- 5.1 An Overview of the Case Study -- 5.2 Modeling the Case Study with Template DDs -- 6 Improving Control Strategy Explanations -- 6.1 Explainability Metrics for Template DDs -- 6.2 Decision Diagram Refactoring -- 6.3 Implementation and Evaluation -- 7 Conclusion -- References -- A Logic of Weighted Reasons for Explainable Inference in AI -- 1 Introduction -- 2 Weighted Default Justification Logic -- 2.1 Justification Logic Preliminaries -- 2.2 Weighted Default Justification Logic -- 2.3 Example -- 3 Strong Weighted Default Justification Logic -- 3.1 Preliminaries -- 3.2 Strong Weighted Default Justification Logic -- 3.3 Example -- 3.4 Justification Default Graphs -- 4 WDJL and WDJL+ as Explainable Neuro-Symbolic Architectures -- 5 Related Work -- 5.1 Numeric Inference Graphs -- 5.2 Numeric Argumentation Frameworks -- 6 Conclusions and Future Work -- References -- On Explaining and Reasoning About Optical Fiber Link Problems -- 1 Introduction -- 2 Literature Review -- 3 Dataset Overview -- 4 Explanations Pipeline Architecture -- 4.1 Data Aggregation -- 4.2 Data Cleansing and Normalisation -- 4.3 Data Transformation -- 4.4 ML Training -- 4.5 AI Explainability -- 5 Experimental Results -- 5.1 Model Performance -- 5.2 Model Explainability -- 6 Conclusion -- A Appendix A -- References.

Construction of Artificial Most Representative Trees by Minimizing Tree-Based Distance Measures -- 1 Introduction -- 2 Methods -- 2.1 Random Forests -- 2.2 Variable Importance Measures (VIMs) -- 2.3 Selection of Most Representative Trees (MRTs) -- 2.4 Construction of Artificial Most Representative Trees (ARTs) -- 2.5 Simulation Design -- 2.6 Benchmarking Experiment -- 3 Results -- 3.1 Prediction Performance -- 3.2 Included Variables -- 3.3 Computation Time -- 3.4 Influence of Tuning Parameters -- 3.5 Tree Depth -- 3.6 Benchmarking -- 4 Discussion -- Appendix -- References -- Decision Predicate Graphs: Enhancing Interpretability in Tree Ensembles -- 1 Introduction -- 2 Literature Review -- 3 Decision Predicate Graphs -- 3.1 Definition -- 3.2 From Ensemble to a DPG -- 3.3 DPG Interpretability -- 4 Empirical Results and Discussion -- 4.1 DPG: Iris Insights -- 4.2 Comparing to the Graph-Based Solutions -- 4.3 Potential Improvements -- 5 Conclusion -- References -- Model-Agnostic and Statistical Methods for eXplainable AI -- Observation-Specific Explanations Through Scattered Data Approximation -- 1 Introduction -- 2 Methodology -- 2.1 Observation-Specific Explanations -- 2.2 Surrogate Models Using Scattered Data Approximation -- 2.3 Estimation of the Observation-Specific Explanations -- 3 Application -- 3.1 Simulated Studies -- 3.2 Real-World Application -- 4 Discussion -- References -- CNN-Based Explanation Ensembling for Dataset, Representation and Explanations Evaluation -- 1 Introduction -- 2

Related Works -- 3 The Concept of CNN-Based Ensembled Explanations -- 3.1 Experimental Setup for Training -- 3.2 Ablation Studies -- 3.3 Method Evaluation -- 4 Metrics for Representation, Dataset and Explanation Evaluation -- 4.1 Representation Evaluation -- 4.2 Dataset Evaluation -- 4.3 Explanations Evaluation -- 5 Conclusions and Future Works -- References.
Local List-Wise Explanations of LambdaMART.
