

1. Record Nr.	UNISA996464448303316
Titolo	Explainable AI within the digital transformation and cyber physical systems : XAI methods and applications // Moamar Sayed-Mouchaweh, editor
Pubbl/distr/stampa	Cham, Switzerland : , : Springer, , [2021] ©2021
ISBN	3-030-76409-5
Descrizione fisica	1 online resource (201 pages)
Disciplina	006.3
Soggetti	Artificial intelligence
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di contenuto	Intro -- Preface -- Contents -- About the Editor -- 1 Prologue: Introduction to Explainable Artificial Intelligence -- 1.1 Explainable Machine Learning -- 1.2 Beyond State-of-the-Art: Contents of the Book -- 1.2.1 Chapter 2: Principles of Explainable Artificial Intelligence -- 1.2.2 Chapter 3: Science of Data: A New Ladder for Causation -- 1.2.3 Chapter 4: Explainable Artificial Intelligence for Predictive Analytics on Customer Turnover -- 1.2.4 Chapter 5: An Efficient Explainable Artificial Intelligence Model of Automatically Generated Summaries Evaluation -- 1.2.5 Chapter 6: On the Transparent Predictive Models for Ecological Momentary Assessment Data -- 1.2.6 Chapter 7: Mitigating the Class Overlap Problem in Discriminative Localization: COVID-19 and Pneumonia Case Study -- 1.2.7 Chapter 8: A Critical Study on the Importance of Feature Selection for Diagnosing Cyber-Attacks in Water Critical Infrastructures -- 1.2.8 Chapter 9: A Study on the Effect of Dimensionality Reduction on Cyber-Attack Identification in Water Storage Tank SCADA Systems -- References -- 2 Principles of Explainable Artificial Intelligence -- 2.1 Introduction -- 2.2 Motivations for XAI -- 2.3 Dimensions of XAI -- 2.4 Explanations and Explanators -- 2.4.1 Single Tree Approximation -- 2.4.2 Rules List and Rules Set -- 2.4.3 Partial Dependency -- 2.4.4 Local Rule-Based Explanation -- 2.4.5 Feature Importance -- 2.4.6 Saliency Maps -- 2.4.7 Prototype-Based Explanations -- 2.4.8 Counterfactual

Explanations -- 2.5 Conclusions -- References -- 3 Science of Data: A New Ladder for Causation -- 3.1 Introduction -- 3.2 Related Works -- 3.2.1 Cognitive Architectures -- 3.2.2 Inferential Logic -- 3.2.3 Probabilistic Fuzzy Logic (PFL) -- 3.2.4 Neural Networks (NN) -- 3.2.4.1 Microscopic Neural Network (NN) -- 3.2.4.2 Macroscopic Neural Network (NN) -- 3.2.4.3 Graph Neural Networks. 3.2.4.4 Hybrid Neural Networks for Reasoning -- 3.3 Cognitive Architecture Equipped with PFL and GNNs -- 3.4 Conclusion -- References -- 4 Explainable Artificial Intelligence for Predictive Analytics on Customer Turnover: A User-Friendly Interface for Non-expert Users -- 4.1 Introduction -- 4.2 Background -- 4.2.1 Shapley Values -- 4.2.2 Types of Explanation Techniques -- 4.3 Related Works -- 4.3.1 Shapley Additive Explanations -- 4.3.2 Contrastive Explanations -- 4.3.3 XAI User Interfaces -- 4.4 Our Explainable AI Web Interface -- 4.4.1 Back-End Component -- 4.4.2 Front-End Component -- 4.4.2.1 Home and Expected Loss -- 4.4.2.2 Local Feature Importance -- 4.4.2.3 Global Feature Importance -- 4.4.2.4 Model Recommendation -- 4.5 Evaluation -- 4.6 Conclusions -- References -- 5 An Efficient Explainable Artificial Intelligence Model of Automatically Generated Summaries Evaluation: A Use Case of Bridging Cognitive Psychology and Computational Linguistics -- 5.1 Introduction -- 5.1.1 Automatic Text Summarization -- 5.1.2 Evaluation Protocols of Automatically Generated Text Summaries -- 5.1.3 Cognitive Psychology Models for Text Comprehension -- 5.1.3.1 The Resonance Model -- 5.1.3.2 The Landscape Model -- 5.1.3.3 The Langston and Trabasso Model -- 5.1.3.4 The Construction-Integration Model -- 5.1.3.5 The Predication Model -- 5.1.3.6 The Gestalt Models -- 5.1.3.7 The Golden and Rumelhart Model -- 5.1.3.8 The Distributed Situation Space Model -- 5.1.3.9 The Structure Building Model -- 5.1.4 Originality of Our Work -- 5.2 CATSE: A Cognitive Automatic Text Summarization Evaluation Protocol -- 5.2.1 The Main Idea -- 5.2.2 Levels of Representation -- 5.2.2.1 The Surface Level -- 5.2.2.2 The Intermediate Level: The Textbase -- 5.2.2.3 The Cognitive Level: The Situation Model -- 5.2.3 The CATSE Protocol -- 5.2.3.1 The Construction Phase. 5.2.3.2 The Integration Phase -- 5.3 Experiments and Results -- 5.3.1 Datasets -- 5.3.2 Experimental Results -- 5.4 Conclusion -- References -- 6 On the Transparent Predictive Models for Ecological Momentary Assessment Data -- 6.1 Introduction -- 6.1.1 Ecological Momentary Assessment (EMA) -- 6.1.2 Classification of EMA Data -- 6.1.3 Model Transparency -- 6.2 Dataset -- 6.3 Analysis Methods -- 6.3.1 Classification Settings -- 6.3.2 Tools -- 6.3.2.1 Own Pipeline: Model Training and Testing -- 6.3.3 Experiments -- 6.3.4 Model Interpretation and Analysis -- 6.3.4.1 Analysis of Categorical Features -- 6.3.4.2 Analysis of Continuous Features -- 6.3.4.3 Interpretation of the Resulting Values -- 6.3.4.4 Model Comparison -- 6.4 Results -- 6.4.1 Experiments with KNIME Analytics Platform -- 6.4.2 Experiments with Own Pipeline -- 6.4.3 Model Interpretation and Analysis -- 6.5 Discussion -- 6.5.1 Is Personalization Always Necessary? -- 6.5.2 Model Similarity -- 6.5.3 Model Validation -- 6.5.4 EMA with Other Features -- 6.5.5 Personalization of Class Labels -- 6.5.6 Method Limitations -- 6.6 Conclusion -- References -- 7 Mitigating the Class Overlap Problem in Discriminative Localization: COVID-19 and Pneumonia Case Study -- 7.1 Introduction -- 7.2 Related Work -- 7.3 Discriminative Localization -- 7.3.1 Class Activation Maps -- 7.3.2 Saliency Maps with Backpropagation -- 7.3.3 Amplified Directed Divergence with Ensembles -- 7.3.4 Scaled Directed Divergence (SDD) -- 7.4 Experiments -- 7.4.1 Method -- 7.4.2 COVID-19 and

Pneumonia Data -- 7.4.3 COVID-19 AND Pneumonia Classifier -- 7.4.4 Scaled Directed Divergence with Natural Imagery -- 7.4.5 Scaled Directed Divergence Applied to Chest X-rays -- 7.5 Discussion -- 7.6 Conclusion -- References -- 8 A Critical Study on the Importance of Feature Selection for Diagnosing Cyber-Attacks in Water Critical Infrastructures.

8.1 Introduction -- 8.2 Background -- 8.2.1 Infinite Feature Selection -- 8.2.2 Infinite Latent Feature Selection -- 8.2.3 Evolutionary Computation Feature Selection -- 8.2.4 Relief Feature Selection -- 8.2.5 Mutual Information -- 8.2.6 Maximum Relevance and Minimum Redundancy -- 8.2.7 Feature Selection via Concave Minimization -- 8.2.8 Laplacian Score -- 8.2.9 Multi-Cluster Feature Selection -- 8.2.10 Recursive Feature Elimination -- 8.2.11 L0-Norm -- 8.2.12 Fisher Score -- 8.3 Design of Intrusion Detection System -- 8.3.1 Data Collection -- 8.3.2 Decision-Making -- 8.4 Experimental Results -- 8.4.1 Experimental Setting -- 8.4.2 Results Analysis -- 8.4.3 Feature Analysis -- 8.5 Conclusion -- References -- 9 A Study on the Effect of Dimensionality Reduction on Cyber-Attack Identification in Water Storage Tank SCADA Systems -- 9.1 Introduction -- 9.2 Background -- 9.2.1 Principal Component Analysis -- 9.2.2 Factor Analysis -- 9.2.3 Confirmatory Factor Analysis -- 9.2.4 Multidimensional Scaling -- 9.2.5 Linear Discriminant Analysis -- 9.2.6 Isomap -- 9.2.7 Semantic Mapping -- 9.2.8 Probabilistic Principal Component Analysis -- 9.2.9 Locally Linear Embedding -- 9.2.10 Laplacian Eigenmaps -- 9.2.11 Landmark Isomap -- 9.2.12 Hessian-based Locally Linear Embedding -- 9.2.13 Local Tangent Space Alignment -- 9.2.14 Kernel Principal Component Analysis -- 9.2.15 Generalized Discriminant Analysis -- 9.2.16 Neighborhood Preserving Embedding -- 9.2.17 Locality Preserving Projections -- 9.2.18 Diffusion Maps -- 9.2.19 Locally Linear Coordination -- 9.2.20 Manifold Charting -- 9.2.21 Large Margin Nearest Neighbor -- 9.2.22 Independent Component Analysis -- 9.3 Design of Intrusion Detection System -- 9.3.1 Data Collection -- 9.3.2 Decision Making -- 9.4 Experimental Results -- 9.4.1 Experiment Setting -- 9.4.2 Results Analysis -- 9.5 Conclusion -- References -- Index.

---