1. Record Nr. UNINA9910627258703321

Titolo	Explainable AI : foundations, methodologies and applications / / Mayuri Mehta, Vasile Palade, Indranath Chatterjee, editors
Pubbl/distr/stampa	Cham, Switzerland : , : Springer, , [2023] ©2023
ISBN	3-031-12807-9
Descrizione fisica	1 online resource (273 pages)
Collana	Intelligent systems reference library ; ; 232
Disciplina	006.301
Soggetti	Artificial intelligence - Philosophy
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di bibliografia	Includes bibliographical references.
Nota di contenuto	Intro -- Preface -- Contents -- Contributors -- Abbreviations -- 1 Black Box Models for eXplainable Artificial Intelligence -- 1.1 Introduction to Machine Learning -- 1.1.1 Motivation -- 1.1.2 Scope of the Paper -- 1.2 Importance of Cyber Security in eXplainable Artificial Intelligence -- 1.2.1 Importance of Trustworthiness -- 1.3 Deep Learning (DL) Methods Contribute to XAI -- 1.4 Intrusion Detection System -- 1.4.1 Classification of Intrusion Detection System -- 1.5 Applications of Cyber Security and XAI -- 1.6 Comparison of XAI Using Black Box Methods -- 1.7 Conclusion -- References -- 2 Fundamental Fallacies in Definitions of Explainable AI: Explainable to Whom and Why? -- 2.1 Introduction -- 2.1.1 A Short History of Explainable AI -- 2.1.2 Diversity of Motives for Creating Explainable AI -- 2.1.3 Internal Inconsistency of Motives for Creating XAI -- 2.1.4 The Contradiction Between the Motives for Creating Explainable AI -- 2.1.5 Paradigm Shift of Explainable Artificial Intelligence -- 2.2 Proposed AI Model -- 2.2.1 The Best Way to Optimize the Interaction Between Human and AI -- 2.2.2 Forecasts Are not Necessarily Useful Information -- 2.2.3 Criteria for Evaluating Explanations -- 2.2.4 Explainable to Whom and Why? -- 2.3 Proposed Architecture -- 2.3.1 Fitness Function for Explainable AI -- 2.3.2 Deep Neural Network is Great for Explainable AI -- 2.3.3 The More Multitasking the Better -- 2.3.4 How to Collect Multitasking Datasets -- 2.3.5 Proposed Neural Network Architecture -- 2.4 Conclusions -- References -- 3

An Overview of Explainable AI Methods, Forms and Frameworks -- 3.1 Introduction -- 3.2 XAI Methods and Their Classifications -- 3.2.1 Based on the Scope of Explainability -- 3.2.2 Based on Implementation -- 3.2.3 Based on Applicability -- 3.2.4 Based on Explanation Level -- 3.3 Forms of Explanation -- 3.3.1 Analytical Explanation.
3.3.2 Visual Explanation -- 3.3.3 Rule-Based Explanation -- 3.3.4 Textual Explanation -- 3.4 Frameworks for Model Interpretability and Explanation -- 3.4.1 Explain like I'm 5 -- 3.4.2 Skater -- 3.4.3 Local Interpretable Model-Agnostic Explanations -- 3.4.4 Shapley Additive Explanations -- 3.4.5 Anchors -- 3.4.6 Deep Learning Important Features -- 3.5 Conclusion and Future Directions -- References -- 4 Methods and Metrics for Explaining Artificial Intelligence Models: A Review -- 4.1 Introduction -- 4.1.1 Bringing Explainability to AI Decision-Need for Explainable AI -- 4.2 Taxonomy of Explaining AI Decisions -- 4.3 Methods of Explainable Artificial Intelligence -- 4.3.1 Techniques of Explainable AI -- 4.3.2 Stages of AI Explainability -- 4.3.3 Types of Post-model Explaination Methods -- 4.4 Metrics for Explainable Artificial Intelligence -- 4.4.1 Evaluation Metrics for Explaining AI Decisions -- 4.5 Use-Case: Explaining Deep Learning Models Using Grad-CAM -- 4.6 Challenges and Future Directions -- 4.7 Conclusion -- References -- 5 Evaluation Measures and Applications for Explainable AI -- 5.1 Introduction -- 5.2 Literature Review -- 5.3 Basics Related to XAI -- 5.3.1 Understanding -- 5.3.2 Explicability -- 5.3.3 Explainability -- 5.3.4 Transparency -- 5.3.5 Explaining -- 5.3.6 Interpretability -- 5.3.7 Correctability -- 5.3.8 Interactivity -- 5.3.9 Comprehensibility -- 5.4 What is Explainable AI? -- 5.4.1 Fairness -- 5.4.2 Causality -- 5.4.3 Safety -- 5.4.4 Bias -- 5.4.5 Transparency -- 5.5 Need for Transparency and Trust in AI -- 5.6 The Black Box Deep Learning Models -- 5.7 Classification of XAI Methods -- 5.7.1 Global Methods Versus Local Methods -- 5.7.2 Surrogate Methods Versus Visualization Methods -- 5.7.3 Model Specific Versus Model Agnostic -- 5.7.4 Pre-Model Versus In-Model Versus Post-Model -- 5.8 XAI's Evaluation Methods.
5.8.1 Mental Model -- 5.8.2 Explanation Usefulness and Satisfaction -- 5.8.3 User Trust and Reliance -- 5.8.4 Human-AI Task Performance -- 5.8.5 Computational Measures -- 5.9 XAI's Explanation Methods -- 5.9.1 Lime -- 5.9.2 Sp-Lime -- 5.9.3 DeepLIFT -- 5.9.4 Layer-Wise Relevance Propagation -- 5.9.5 Characteristic Value Evaluation -- 5.9.6 Reasoning from Examples -- 5.9.7 Latent Space Traversal -- 5.10 Explainable AI Stakeholders -- 5.10.1 Developers -- 5.10.2 Theorists -- 5.10.3 Ethicists -- 5.10.4 Users -- 5.11 Applications -- 5.11.1 XAI for Training and Tutoring -- 5.11.2 XAI for 6G -- 5.11.3 XAI for Network Intrusion Detection -- 5.11.4 XAI Planning as a Service -- 5.11.5 XAI for Prediction of Non-Communicable Diseases -- 5.11.6 XAI for Scanning Patients for COVID-19 Signs -- 5.12 Possible Research Ideology and Discussions -- 5.13 Conclusion -- References -- 6 Explainable AI and Its Applications in Healthcare -- 6.1 Introduction -- 6.2 The Multidisciplinary Nature of Explainable AI in Healthcare -- 6.2.1 Technological Outlook -- 6.2.2 Legal Outlook -- 6.2.3 Medical Outlook -- 6.2.4 Ethical Outlook -- 6.2.5 Patient Outlook -- 6.3 Different XAI Techniques Used in Healthcare -- 6.3.1 Methods to Explain Deep Learning Models -- 6.3.2 Explainability by Using White-Box Models -- 6.3.3 Explainability Methods to Increase Fairness in Machine Learning Models -- 6.3.4 Explainability Methods to Analyze Sensitivity of a Model -- 6.4 Application of XAI in Healthcare -- 6.4.1 Medical Diagnostics -- 6.4.2 Medical Imaging -- 6.4.3 Surgery -- 6.4.4 Detection of COVID-19 -- 6.5 Conclusion -- References -- 7 Explainable AI Driven Applications for Patient Care and Treatment --

7.1 General -- 7.2 Benefits of Technology and AI in Healthcare Sector -- 7.3 Most Common AI-Based Healthcare Applications -- 7.4 Issues/Concerns of Using AI in Health Care.
7.5 Why Explainable AI? -- 7.6 History of XAI -- 7.7 Explainable AI's Benefits in Healthcare -- 7.8 XAI Has Proposed Applications for Patient Treatment and Care -- 7.9 Future Prospects of XAI in Medical Care -- 7.10 Case Study on Explainable AI -- 7.11 Framework for Explainable AI -- 7.12 Conclusion -- References -- 8 Explainable Machine Learning for Autonomous Vehicle Positioning Using SHAP -- 8.1 Introduction -- 8.1.1 Global Navigation Satellite System (GNSS) and Autonomous Vehicles -- 8.1.2 Navigation Using Inertial Measurement Sensors -- 8.1.3 Inertial Positioning Using Wheel Encoder Sensors -- 8.1.4 Motivation for Explainability in AV Positioning -- 8.2 eXplainable Artificial Intelligence (XAI): Background and Current Challenges -- 8.2.1 Why XAI in Autonomous Driving? -- 8.2.2 What is XAI? -- 8.2.3 Types of XAI -- 8.3 XAI in Autonomous Vehicle and Localisation -- 8.4 Methodology -- 8.4.1 Dataset: IO-VNBD (Inertial and Odometry Vehicle Navigation Benchmark Dataset) -- 8.4.2 Mathematical Formulation of the Learning Problem -- 8.4.3 WhONet's Learning Scheme -- 8.4.4 Performance Evaluation Metrics -- 8.4.5 Training of the WhONet Models -- 8.4.6 WhONet's Evaluation -- 8.4.7 SHapley Additive exPlanations (SHAP) Method -- 8.5 Results and Discussions -- 8.6 Conclusions -- References -- 9 A Smart System for the Assessment of Genuineness or Trustworthiness of the Tip-Off Using Audio Signals: An Explainable AI Approach -- 9.1 Introduction -- 9.2 Background -- 9.3 Proposed Methodology -- 9.3.1 Dataset Used -- 9.3.2 Pre-processing -- 9.3.3 Feature Extracted -- 9.3.4 Feature Selected -- 9.3.5 Machine Learning in SER -- 9.3.6 Performance Index -- 9.4 Results and Discussion -- 9.5 Conclusion -- References -- 10 Face Mask Detection Based Entry Control Using XAI and IoT -- 10.1 Introduction -- 10.2 Literature Review -- 10.3 Methodology.
10.3.1 Web Application Execution -- 10.3.2 Implementation -- 10.3.3 Activation Functions -- 10.3.4 Raspberry Pi Webserver -- 10.4 Results -- 10.4.1 Dataset -- 10.4.2 Model Summary -- 10.4.3 Model Evaluation -- 10.5 Conclusion -- References -- 11 Human-AI Interfaces are a Central Component of Trustworthy AI -- 11.1 Introduction -- 11.2 Regulatory Requirements for Trustworthy AI -- 11.3 Explicability-An Ethical Principle for Trustworthy AI -- 11.4 User-Centered Approach to Trustworthy AI -- 11.4.1 Stakeholder Analysis and Personas for AI -- 11.4.2 User-Testing for AI -- 11.5 An Example Use Case: Computational Pathology -- 11.5.1 AI in Computational Pathology -- 11.5.2 Stakeholder Analysis for Computational Pathology -- 11.5.3 Human-AI Interface in Computational Pathology -- 11.6 Conclusion -- 11.7 List of Abbreviations -- References.