SORS: Navigating the Landscape of Explainable AI

Objectives

Click here to download the presentation

Abstract: In today's technological landscape, the mechanics of Artificial Intelligence (AI) and Machine Learning (ML) can often feel obscured, functioning as elusive "black boxes". This opacity raises concerns about trust, reliability, and wider application. Explainable AI (XAI) seeks to address these challenges, working towards making AI models more comprehensible. This presentation offers an exploration of XAI, discussing various approaches, their strengths and weaknesses, and their application to different data types, including tabular, image, and text data. Moreover, we will delve into the potential of XAI to not only demystify these models, but also to offer a novel perspective on data analysis. This perspective can facilitate scientific insights, advancing our understanding of what is and isn't working in our models, thus allowing for more refined algorithmic decisions and the potential uncovering of new discoveries.

Short bio:

Ettore Mariotti is a Ph.D. student at the Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS). His research, rooted in explainable artificial intelligence, involves a comprehensive approach to understanding machine learning models. He both develops and utilizes post-hoc explanations for black box models and creates white box models that are inherently interpretable. His work aids in bridging the gap between complex AI systems and human comprehension.

Previously, at the University of Padua, Ettore applied deep learning to IACT astroparticle physics data analysis, an endeavor that sparked his interest in explainable AI. This experience marked the commencement of his journey at the intersection of advanced technology and scientific inquiry.

Recently, Ettore had the opportunity to intern at INDITEX. Here, he was able to apply his academic knowledge to real-life challenges, enriching his understanding of the practical application of theoretical concepts.

This talk is co organized together with the Master in Artificial Intelligence (UPC,UB,URV).

Speakers

Speaker: Ettore Mariotti, Ph.D. student at the Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS)
Host: Prof. Ulises Cortés, High Performance Artificial Intelligence Group Manager, CS, BSC and Coordinator of the Master in Artificial Intelligence (UPC,UB,URV)

Barcelona Supercomputing Center - Centro Nacional de Supercomputación

---