https://smartcomputing.unifi.it/courses.html

eXplainable Artificial Intelligence and real-world applications

Code	SC_39003
Year	2024
Instructor	Antonio Luca Alfeo (University of Pisa)
Location	Department of Information Engineering, Largo Lucio Lazzarino 1, University of Pisa
ECTS Credits	5

Schedule

Day	Time	Room
Mon 15 Jan 2024	09:00 - 13:00	meeting room (6th floor)
Mon 15 Jan 2024	14:30 - 17:30	meeting room (6th floor)
Tue 16 Jan 2024	09:00 - 13:00	meeting room (6th floor)
Tue 16 Jan 2024	14:30 - 17:30	meeting room (6th floor)
Wed 17 Jan 2024	09:00 - 13:00	meeting room (6th floor)
Wed 17 Jan 2024	14:30 - 16:30	meeting room (6th floor)

Abstract

Despite the great classification performances provided by nowadays machine learning (ML) models, it can be difficult to employ their predictions in real-world decision-making processes. Indeed, recognition performances of the ML model can be affected by spurious correlations between input features, bias in the data, and imperfect convergence during the training process. These issues may be undetectable just by looking at the recognition performance [1]. Specifically, the ML model can work as a black box thus preventing decision- makers from validating and trusting their results. In this context, XAI approaches can provide some insights to be sure that the ML model is providing the right outcome for the right reason [2]. XAI is also important for the end-user i.e., who is going to be affected by the decision taken with the support of a ML algorithm. Indeed, recent regulations on personal data processing (i.e., the General Data Protection Regulation [3]), do include some level of explainability as a requirement to be met to employ AI in real-world decision-making processes. For the end user, it would be essential to know why a specific result has been obtained (i.e. why a bank does not allow for a loan) and how this result can be modified [4]. Finally, an explanation can also support researchers and domain experts while investigating the data, verifying the assumptions made upon their investigation, and obtaining new insights from the ML model. To address these issues, some insights into the reasoning behind ML models outcomes can be provided via different types of explanations. This course focuses on showing what these types are and when it is more convenient to apply one of them. Each lecture includes a little workshop, and the final assessment consists of the presentation and interpretation of the results obtained in each workshop.