GENERATING COUNTERFACTUALS

UMFELD

Investigating and Developing approaches to generate realistic and human-interpretable counterfactuals to improve the explainability and performance of "black-box" algorithms. Counterfactuals are a promising field of research in the area of explainable machine learning. Their ability to visualize decision processes of "black-box" algorithms by only taking the input and output into account make them applicable to any type of machine learning algorithm. However, the current approaches are limited in their ability to generate human-interpretable counterfactual for more than one algorithm/use case at the time.

The purpose of this thesis is to develop a new, interactive generating algorithm for human-interpretable counterfactuals. In this context he/she develops a type of generative algorithm, that takes the human feedback into account while generating counterfactuals. Hereby, the student acquires state-of-the-art knowledge of Explainable and Interactive Machine Learning and provides the research community with an interactive approach for generating realistic image counterfactuals.

AUFGABEN

- In-depth literature review about data generation techniques for counterfactuals and interactive machine learning
- Development of an interactive generation process for counterfactuals
- Experimental comparison of state-of-the-art generation processes with the newly developed approach

WIR BIETEN

- Continuous and thorough mentoring of the student
- A highly motivated and fun team
- Constructive teamwork

WIR ERWARTEN

- Basic knowledge in python programming and machine learning
- Ability to plan and work independently
- Very good knowledge of German or English
- High level of motivation and enthusiasm
BEWERBUNG

We are looking forward to your application to Patrick Philipp (philipp@fzi.de) or Jacqueline Höllig (hoellig@fzi.de).

Please provide us with the following information:

- Transcript of Records
- CV

WEITERE INFORMATIONEN

- Start: Asap
- Responsible institute at KIT: AIFB | Prof. Dr. York Sure-Vetter
- Themen-Schwerpunkt: Maschinelles Lernen, Wissen und Informationsdienste
- Studiengänge: Informationswirtschaft, Verwandte Studiengänge, Wirtschaftsinformatik, Wirtschaftsingenieurwesen