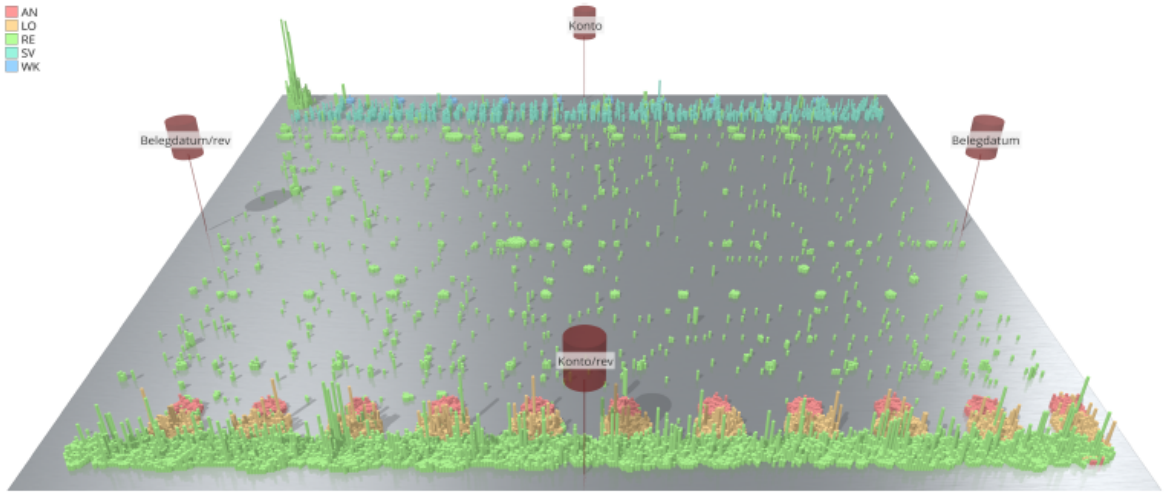
Master‘s Project WT 2022/23

Exploring AI Models Using Visualization



Background

Artificial intelligence techniques (AI) are used in many domains. They usually achieve high test and train accuracy but they cannot explain why they made a particular prediction. Often, they are considered as a black box. This leads to hardly acceptable uncertainty when using them in sensitive domains. The research area of explainable AI (XAI) has therefore emerged to develop methods for post-hoc analysis of trained models based on model-agnostic properties. In this master project we aim to develop and implement visual analytics techniques that will allow us to explore XAI techniques. We will consider several use-cases, such as finance or text data.

Description

Specific objectives will be discussed and detailed with the participants. We have the following tentative goals:

- Study existing post-hoc explanations for AI.
- Conceptualize and implement a 2.5D Visual Analytics approach combining several features for achieving deeper explanations of the model.
- Compare our approach with existing techniques for different domains.

The master’s project refers to current research and software projects of the Computer Graphics Systems group. It is especially suited for further research in the context of a master's thesis or a future doctoral thesis. Further, the master’s project can lay a foundation for working as a student assistant. Depending on the course of the project, the publication of a paper at an international conference is possible.

Contact

Daniel Atzberger (daniel.atzberger@hpi.de), Tim Cech (tim.cech@hpi.de), Willy Scheibel (willy.scheibel@hpi.de)