It's not a Bug, it's a Feature: How Visual Model Evaluation can help to incorporate Human Domain Knowledge in Data Science

Completed Research Paper

Dennis Eilers
Leibniz Universität Hannover
Königsworther Platz 1
30167 Hannover, Germany
eilers@iwi.uni-hannover.de

Cornelius Köpp
edicos Group
Hindenburgstraße 28/29
30175 Hannover, Germany
cornelius.koepp@edicos.de

Christoph Gleue
Leibniz Universität Hannover
Königsworther Platz 1
30167 Hannover, Germany
gleue@iwi.uni-hannover.de

Michael H. Breitner
Leibniz Universität Hannover
Königsworther Platz 1
30167 Hannover, Germany
breitner@iwi.uni-hannover.de

Abstract

The question of how to incorporate human domain knowledge in practical data science projects is still a major challenge. While machine learning tasks are usually carried out by technically skilled data scientists, these analysts do not necessarily have the required domain knowledge concerning a particular business problem to explain certain phenomena. In real-world data science applications, this may result in models that do not adequately reflect relationships in the data. We address this issue by introducing a heat map based technique for model error visualization to facilitate discussions of the results between data scientists and domain experts. By discussing model errors with domain experts during the iterative analysis process, the generated insights can be used for engineering new features (explanatory variables) which better represent the problem and therefore improve the results. We demonstrate the visualization approach based on artificial data and in the context of a real-world industry example.