TU Clausthal

Kolloquium zur Bachelorarbeit

**Tim Röthig**

# „Investigation of model agnostic methods for interpretation of AI regression models“

Recently, powerful hardware and huge data sources have facilitated the development of artificial Intelligence and machine learning methods and models. More accurate but less explainable AI methods need to be explained for users and developers. In this thesis, we focused on the explanation of AI-based regression models. At first, we proved the hypothesis that the state of the art black-box machine learning models (e.g. random forest, deep learning) produce better results than conventional explainable ones (linear regression, decision trees), by fitting different models on well-known benchmark datasets like the King County house sales and the New York Taxi trips. To get representative comparison results we experimented with different model parameters. The conducted experiments show that less explainable state-of-the art machine learning methods like random forest and deep learning produced more accurate results. Secondly, we investigated how model-agnostic interpretation methods can explain machine learning models and certain predictions obtained in the first part. Finally, we can conclude that the disadvantage of less explainable but more accurate “black-box” models can be compensated by applying model-agnostic methods.

Dienstag, 07.01.2020, 11:00 Uhr,
Besprechungsraum 106, (D3) Julius-Albert-Str. 4

Technische Universität Clausthal • Institut für Informatik • Julius-Albert-Straße 4 • 38678 Clausthal-Zellerfeld
Tel +49 (0)53 23 72-7114• Fax +49 (0)5323 72-7199 • http://www.in.tu-clausthal.de • info@in.tu-clausthal.de