

# ENBIS Spring Meeting 2024



## Report of Contributions

Contribution ID: 37

Type: **not specified**

# Anomaly Detection in Ordinal Quality-Related Processes by Control Charts

*Wednesday, 15 May 2024 13:20 (20 minutes)*

Control charts are a well-known approach for quality improvement and anomaly detection. They are applied to quality-related processes (e.g., metrics of product quality from a monitored manufacturing process) and allow to detect “deviations from normality”, i.e., if the process turns from its specified in-control state into an out-of-control state. In this study, we focus on ordinal data generating processes, where the monitored quality metrics are measured on an ordered qualitative scale. A survey of control charts for the sample-based monitoring of independent and identically distributed ordinal data is provided together with critical comparisons of the control statistics, for memory-less Shewhart-type and for memory-utilizing exponentially weighted moving average (EWMA) and cumulative-sum types of control charts. New results and proposals are also provided for process monitoring. Using some real-world quality scenarios from the literature, a simulation study for performance comparisons is conducted, covering sixteen different types of control chart. It is shown that demerit-type charts used in combination with EWMA smoothing generally perform better than the other charts, although the latter may rely on quite sophisticated derivations. A real-world data example for monitoring flashes in electric toothbrush manufacturing is discussed to illustrate the application and interpretation of the control charts in the study.

Reference:

Ottenstreu, S., Weiß, C.H., Testik, M.C. (2023)  
A Review and Comparison of Control Charts for Ordinal Samples.  
Journal of Quality Technology 55(4), 422-441.  
Open Access: <https://doi.org/10.1080/00224065.2023.2170839>

## Type of presentation

Contributed Talk

**Primary authors:** OTTENSTREUER, Sebastian (Helmut Schmidt University); Prof. WEISS, Christian (Helmut Schmidt University); Prof. TESTIK, Murat Caner (Hacettepe University)

**Presenter:** Prof. WEISS, Christian (Helmut Schmidt University)

**Session Classification:** Contributed session

**Track Classification:** Spring Meeting

Contribution ID: 38

Type: **not specified**

# Dataset Bias in Human Activity Recognition for Industrial Processes

*Thursday, 16 May 2024 14:15 (25 minutes)*

When creating multi-channel time-series datasets for Human Activity Recognition (HAR), researchers are faced with the issue of subject selection criteria. It is unknown what physical characteristics and/or soft biometrics, such as age, height, and weight, must be considered to train a classifier to achieve robustness toward heterogeneous populations in the training and testing data. This contribution statistically curates the training data to assess to what degree the physical characteristics of humans influence HAR performance. We evaluate the performance of three neural networks on four HAR datasets that vary in the sensors, activities, and recording for time-series HAR. The training data is intentionally biased to human characteristics to determine the features that impact motion behavior. The evaluations brought forth the impact of the subjects' characteristics on HAR. Thus providing insights regarding the robustness of the classifier with respect to heterogeneous populations. The study is a step forward in the direction of fair and trustworthy artificial intelligence by attempting to quantify representation bias in multi-channel time series HAR data.

## Type of presentation

Invited Talk

**Primary authors:** SCHMID, Lena (TU Dortmund University); RAVI NAIR, Nilah (<https://www.tu-dortmund.de/>)**Co-author:** REINING, Christopher (TU Dortmund University)**Presenter:** RAVI NAIR, Nilah (<https://www.tu-dortmund.de/>)**Session Classification:** Invited session**Track Classification:** Spring Meeting

Contribution ID: 39

Type: **not specified**

## Methods for Quantifying Dataset Similarity: a Review, Taxonomy and Comparison

*Wednesday, 15 May 2024 15:30 (20 minutes)*

Quantifying the similarity between datasets has widespread applications in statistics and machine learning. The performance of a predictive model on novel datasets, referred to as generalizability, depends on how similar the training and evaluation datasets are. Exploiting or transferring insights between similar datasets is a key aspect of meta-learning and transfer-learning. In simulation studies, the similarity between distributions of simulated datasets and real datasets, for which the performance of methods is assessed, is crucial. In two- or  $k$ -sample testing, it is checked, whether the underlying distributions of two or more datasets coincide.

Extremely many approaches for quantifying dataset similarity have been proposed in the literature. We examine more than 100 methods and provide a taxonomy, classifying them into ten classes, including (i) comparison of cumulative distribution functions, density functions, or characteristic functions; (ii) methods based on multivariate ranks; (iii) discrepancy measures for distributions; (iv) graph-based methods; (v) methods based on inter-point distances; (vi) kernel-based methods; (vii) methods based on binary classification; (viii) distance and similarity measures for datasets; (ix) comparison based on summary statistics; and (x) testing approaches. In an extensive review of these methods the main underlying ideas, formal definitions, and important properties were introduced. The main ideas of the classes are presented here.

We compare the more than 100 methods in terms of their applicability, interpretability, and theoretical properties, in order to provide recommendations for selecting an appropriate dataset similarity measure based on the specific goal of the dataset comparison and on the properties of the datasets at hand. An online tool facilitates the choice of the appropriate dataset similarity measure.

### Type of presentation

Contributed Talk

**Primary authors:** STOLTE, Marieke (Department of Statistics, TU Dortmund University); Dr KAP-PENBERG, Franziska (Department of Statistics, TU Dortmund University); Prof. RAHNENFÜHRER, Jörg (Department of Statistics, TU Dortmund University); Dr BOMMERT, Andrea (Department of Statistics, TU Dortmund University)

**Presenter:** STOLTE, Marieke (Department of Statistics, TU Dortmund University)

**Session Classification:** Contributed session

**Track Classification:** Spring Meeting

Contribution ID: 41

Type: **not specified**

## Mitigating Emergency Stop Collisions in AGV Fleets in Case of Control Failure

*Thursday, 16 May 2024 14:40 (25 minutes)*

Automated guided vehicles (AGVs) are an essential area of research for the industry to enable dynamic transport operations. Furthermore, AGV-based multi-robot systems (MRS) are being utilized in various applications, e.g. in production or in logistics. Most research today focuses on ensuring that the system is operational, which is not always achieved. In daily use, faults and failures in an AGV-based MRS are most likely inevitable. So industrial systems must support some safety methods, e.g. an emergency stop function. Although emergency stop functions are designed to prevent larger issues, their usage leads to a failure of the control system, since the affected systems are typically shut down immediately. Depending on the AGV type, an uncontrolled behaviour can occur. In case of control failure, this behaviour can lead to collisions and associated high costs. In this paper, we present and compare three approaches for avoiding collisions in an intralogistics scenario in the case of control failure. In the said scenario, the trajectory planning is being adapted to minimize or avoid collisions. The first approach calculates the next collisionfree time slot when an emergency stop occurs and continues the planned trajectories until this time. The second approach calculates several alternative trajectories with the existing trajectory planning without considering emergency stop collisions. The third approach aims to completely avoid emergency stop collisions by extending trajectory planning to include the detection of possible emergency stop collisions. We evaluate and compare the approaches by employing multiple metrics to assess their performance, including runtime, number of collisions that occur, and the system's throughput. We thoroughly discuss and analyse the results, offering insights into the strengths and weaknesses of the approaches.

### Type of presentation

Invited Talk

**Primary author:** OGORELYSHEVA, Natalia (Fraunhofer IML)**Presenter:** OGORELYSHEVA, Natalia (Fraunhofer IML)**Session Classification:** Invited session

**Track Classification:** Spring Meeting

Contribution ID: 43

Type: **not specified**

# Cooperative Game Theory and Importance Quantification

*Thursday, 16 May 2024 11:00 (25 minutes)*

Being able to quantify the importance of random inputs of an input-output black-box model is at the cornerstone of the fields of sensitivity analysis (SA) and explainable artificial intelligence (XAI). To perform this task, methods such as Shapley effects and SHAP have received a lot of attention. The former offers a solution for output variance decomposition with non-independent inputs, and the latter proposes a way to decompose predictions of predictive models. Both of these methods are based upon the Shapley values, an allocation mechanism from cooperative game theory.

This presentation aims to shed light on the underlying mechanism behind the paradigm of cooperative games for input importance quantification. To that extent, a link is drawn with the Möbius inversion formula to boolean lattices leading to coalitional decompositions of quantities of interest. Allocations can be seen as aggregations of such decomposition, leading to a more general view of the importance quantification problem.

This generalization is leveraged in order to solve a problem in the context of global SA with dependent inputs. The Shapley effects are known not to be able to detect exogenous inputs (i.e., variables not in the model). Using a different allocation, namely the proportional values, leads to interpretable importance indices with the ability to identify such inputs.

These indices are illustrated on a classical problem of surrogate modeling of a costly numerical model: the transmittance performance of an optical filter. It allows for clear and interpretable decision rules for feature selection and dimension reduction.

## Type of presentation

Invited Talk

**Primary author:** IL IDRISSE, Marouane (Institut de Mathématiques de Toulouse)**Co-authors:** IOOSS, Bertrand (EDFR&D); GAMBOA, Fabrice (Institut de Mathématiques de Toulouse); LOUBES, Jean-Michel (Institut de Mathématiques de Toulouse); Mr BOUSQUET, Nicolas (EDF R&D)**Presenter:** IL IDRISSE, Marouane (Institut de Mathématiques de Toulouse)**Session Classification:** Invited session**Track Classification:** Spring Meeting

Contribution ID: 44

Type: **not specified**

## Application of Sign Depth to Point Process Models

*Wednesday, 15 May 2024 13:40 (20 minutes)*

An important task in reliability studies is the lifetime testing of systems consisting of dependent or interacting components. Since the fatigue of a composite material is largely determined by the failure of its components, its risk of breakdown can be linked to the observable component failure times. These failure times form a simple point process that has numerous applications also in econometrics and epidemiology, among others.

A powerful tool for modeling simple point processes is the stochastic intensity, which can be thought of as the instantaneous average rate for the occurrence of an event. Here, this event represents the failure of a system component. Under a random time change based on the cumulative intensity, any such point process can be transformed into a unit-rate Poisson process with exponential interarrival times. If we consider a parametric model for the stochastic intensity, we can perform this transformation for each parameter to obtain so-called hazard transforms. As soon as the parameter deviates from its true value, these transforms will generally no longer follow a standard exponential distribution. At this point, familiar goodness-of-fit tests such as the Kolmogorov-Smirnov test can be applied.

However, viewing the transforms as “residuals”, data depth approaches commonly encountered in the regression context can be considered as an alternative. In particular, the consistent 3-sign depth test provides a much more powerful generalization of the classical sign test. The major benefit of data depth methods lies in their inherent robustness, for instance in the presence of contaminated data due to measurement errors or unexpected external influences. This robustness often entails a drop in power of the associated test. In a simulation study, we therefore compare the 3-sign depth test with competing approaches in terms of power and robustness, and find that satisfactory results can still be achieved even if almost half of the data is contaminated.

Finally, we apply our depth-based method to real data from a civil engineering experiment conducted at TU Dortmund University. We assess whether these robust approaches are suitable for predicting the lifetimes of prestressed concrete beams exposed to different cyclic loading and investigate if there is evidence of a statistically significant accumulation of damage over the course of the experiment.

### Type of presentation

Contributed Talk

**Primary author:** JAKUBZIK, Mirko (TU Dortmund University)**Presenter:** JAKUBZIK, Mirko (TU Dortmund University)**Session Classification:** Contributed session**Track Classification:** Spring Meeting



Contribution ID: 45

Type: **not specified**

## Explanation Groves –Analyzing the Trade-off between Complexity and Adequacy of an Explanation

*Wednesday, 15 May 2024 15:50 (20 minutes)*

The increasing popularity of machine learning in many application fields has led to an increasing demand in methods of explainable machine learning as they are e.g. provided by the R packages DALEX (Biecek, 2018) and iml (Molnar, 2018). A general process to ensure the development of transparent and auditable machine learning models in industry (TAX4CS) is given in Bücke et al. (2021).

In turn, comparatively few research has been dedicated to the limits of explaining complex machine learning models (cf. e.g. Rudin, 2019, Szepannek and Lübke, 2023). In the presentation, explanation groves (Szepannek and von Holt, 2024) will be introduced. Explanation groves extract a set of understandable rules in order to explain arbitrary machine learning models. In addition, the degree of complexity of the resulting explanation can be defined by the user. In consequence, they provide a useful tool to analyze the trade off between the complexity of a given explanation on one hand and how well it represents the original model on the other hand.

After presenting the method some results on real world data will be shown. A corresponding R package xgrove is available on CRAN (Szepannek, 2023) and will be briefly demonstrated.

Biecek P (2018). DALEX: Explainers for Complex Predictive Models in R. *Journal of Machine Learning Research*, 19(84), 1-5. <https://jmlr.org/papers/v19/18-416.html>.

Bücke, M.; Szepannek, G., Gosiewska, A. Biecek, P. (2021): Transparency, Auditability and explainability of Machine Learning Models in Credit Scoring, *Journal of the Operation Research Society*, DOI: 10.1080/01605682.2021.1922098.

Molnar C, Bischl B, Casalicchio G (2018). “iml: An R package for Interpretable Machine Learning.” *JOSS*, 3(26), 786. DOI:10.21105/joss.00786.

Rudin, C (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell* 1, 206–215. DOI:10.1038/s42256-019-0048-x

Szepannek G (2023). xgrove: Explanation Groves. R package version 0-1-7. <https://CRAN.R-project.org/package=xgrove>.

Szepannek, G., von Holt, B.-H. (2024): Can't See the Forest for the Trees –Analyzing Groves for Random Forest Explanation, *Behaviormetrika*, DOI: 10.1007/s41237-023-00205-2.

Szepannek, G., Luebke, K. (2022): Explaining Artificial Intelligence with Care – Analyzing the Explainability of Black Box Multiclass Machine Learning Models in Forensics, *Künstliche Intelligenz*, DOI : 10.1007/s13218-022-00764-8.

### Type of presentation

Contributed Talk

**Primary author:** SZE PANNEK, Gero (Stralsund University of Applied Sciences)

**Presenter:** SZEPAÑNEK, Gero (Stralsund University of Applied Sciences)

**Session Classification:** Contributed session

**Track Classification:** Spring Meeting

Contribution ID: 46

Type: **not specified**

## Quality aspects of machine learning in official statistics

*Wednesday, 15 May 2024 16:10 (20 minutes)*

Advanced statistical and machine learning models methods are becoming increasingly important in applied data science. At the same time, their trustworthiness is critical for the progress and adoption of data science applications in various fields, including official statistics.

„Bad quality reduces trust very, very fast.“ Taking up this dictum, official statistics in Germany have considered what a quality concept for the use of machine learning could look like. Six quality dimensions (including explainability and robustness) and two cross-sectional aspects (including fairness) were developed, as well as concrete guidelines on how compliance with these can be measured. All this with the aim of deriving a binding standard for official statistics.

The talk will highlight motivation, genesis and first practical implementations of this concept of quality.

Main references:

- Yung W, Tam S-M, Buelens B, Chipman H, Dumpert F, Ascari G, Rocci F, Burger J, Choi I (2022) A quality framework for statistical algorithms. *Stat J IAOS* 38(1):291–308. <https://doi.org/10.3233/SJI-210875>
- Saidani Y, Dumpert F, Borgs C et al (2023) Qualitätsdimensionen maschinellen Lernens in der amtlichen Statistik. *AStA Wirtsch Sozialstat Arch*. <https://doi.org/10.1007/s11943-023-00329-7>

### Type of presentation

Contributed Talk

**Primary authors:** DUMPERT, Florian (Federal Statistical Office of Germany); SAIDANI, Younes (Federal Statistical Office of Germany)

**Presenter:** DUMPERT, Florian (Federal Statistical Office of Germany)

**Session Classification:** Contributed session

**Track Classification:** Spring Meeting

Contribution ID: 47

Type: **not specified**

## Why Order Matters for Ordinal Prediction with Random Forests

*Thursday, 16 May 2024 10:20 (20 minutes)*

Traditionally, ordinal response data have been modeled through parametric models such as the proportional odds model. More recently, popular machine learning methods such as random forest (RF) have been extended for ordinal prediction. As RF does not inherently support ordinal response data, a common approach is assigning numeric scores to the ordinal response categories and learning a regression RF model on the numeric scores instead. However, this requires the pre-specification of said numeric scores. While some approaches simply use an integer representation of the  $k$  ordinal response categories (i.e., 1, 2, ...,  $k$ ), other methods such as Ordinal Forest (OF; Hornung, 2019) and the Ordinal Score Optimization Algorithm (OSOA; Buczak et al., 2024) have been proposed which both internally optimize the numeric scores w.r.t. the predictive performance achieved when using them. For predicting unseen observations, both OF and OSOA rely on a Transform-First-Aggregate-After (TFAA) procedure, where for each new observation numeric score predictions are generated at the tree level and transformed back into the ordinal response category. In a second step, an aggregated prediction is then obtained via majority voting. In this work, we propose a novel prediction approach, where the numeric score predictions are first aggregated into a single, combined numeric score prediction which in turn is transformed back into a categorical prediction (i.e., Aggregate-First-Transform-After; AFTA). We show that AFTA prediction can notably enhance the predictive performance of OF and OSOA. Further, we propose Border Ranger (BR), a novel RF method for ordinal prediction that reaches similar predictive performance as the AFTA prediction enhanced OF while avoiding the computationally intensive optimization procedure. We evaluate all methods on simulation and real data.

### Type of presentation

Contributed Talk

**Primary author:** BUCZAK, Philip (TU Dortmund University)**Presenter:** BUCZAK, Philip (TU Dortmund University)**Session Classification:** Contributed session**Track Classification:** Spring Meeting

Contribution ID: 48

Type: **not specified**

# Global Sensitivity Analysis Reporting Tool for Easily Detecting Variable Impact and Interaction

*Thursday, 16 May 2024 11:25 (25 minutes)*

Explainable Artificial Intelligence (XAI) stands as a crucial area of research essential for advancing AI applications in real-world contexts. Within XAI, Global Sensitivity Analysis (GSA) methods assume significance, offering insights into the influential impact of individual or grouped parameters on the predictions of machine learning models, as well as the outcomes of simulators and real-world processes.

One area where GSA proves particularly valuable is in black-box optimization. When setting up an optimization problem, it is crucial to meticulously select parameters that play an important role and are therefore suitable as variables. This choice significantly influences the outcome of the optimization procedure; if the wrong variables are chosen, the effectiveness of the optimization process is compromised. Additionally, the performance of many black-box optimizers is influenced by the dimensionality of the problem. To obtain reliable results, the amount of sampled data needs to grow exponentially with the dimensionality. Therefore, choosing to optimize only the significant parameters allows for reducing the dimensionality of the problem and, consequently, the number of queries to the objective function.

In this talk, we present GSAREport [1,2], an open-source software recently developed by our team. GSAREport generates elaborate reports that describe the global sensitivities of input parameters using a variety of GSA methods. These reports allow users to inspect which features are crucial for a given real-world function/simulator, or model.

To provide a usage example, we evaluate the tool's performance on a real-world test case in engineering design. We examine the impact of parameters of different nature on the performance of a crash box, for varying dimensions and sample sizes. Our findings underscore the relevance of our tool as (1) an instrument for gaining a deeper understanding of the features contributing to component performance and (2) a preliminary step to reduce dimensionality in optimization, thereby enhancing algorithm efficiency while maintaining flexibility. Despite the specific use-case considered, we show the potential of GSAREport as an interdisciplinary and user-friendly tool capable of expediting design processes and enhancing the overall understanding across diverse application areas.

References:

- [1] B. Van Stein, E. Raponi, Z. Sadeghi, N. Bouman, R. C. H. J. Van Ham, and T. Bäck, A Comparison of Global Sensitivity Analysis Methods for Explainable AI with an Application in Genomic Prediction. *IEEE Access* (2022).
- [2] B. Van Stein, E. Raponi, GSAREport: easy to use global sensitivity reporting. *Journal of Open Source Software* (2022), 7 (78), 4721.

## Type of presentation

Invited Talk

**Primary author:** RAPONI, Elena (LIACS, Leiden University)

**Co-authors:** Dr VAN STEIN, Niki (LIACS, Leiden University); Prof. BÄCK, Thomas (LIACS, Leiden University)

**Presenter:** RAPONI, Elena (LIACS, Leiden University)

**Session Classification:** Invited session

**Track Classification:** Spring Meeting

Contribution ID: 49

Type: **not specified**

# Security evaluation of Deep Neural Networks and Gradient Boosting Decision Trees in an insurance context

*Thursday, 16 May 2024 13:50 (20 minutes)*

Machine learning (ML) will play an increasingly important role in many processes of insurance companies in the future [1]. However, ML models are at risk of being attacked and manipulated [2]. In this work, the robustness of Gradient Boosted Decision Tree (GBDT) models and Deep Neural Networks (DNN) in an insurance context is evaluated. It is analyzed how vulnerable each model is against label-flipping, backdoor, and adversarial example (AE) attacks. Therefore, two GBDT models and two DNNs were trained on two different tabular datasets from an insurance context. The ML tasks performed on these datasets are claim prediction (regression) and fraud detection (binary classification).

Label-flipping attacks do not present a high threat in the scenarios of this work, as the obstacles to a successful attack are particularly high in relation to the potential gain for an adversary. Nevertheless, a small fraction of flipped labels can reduce the general performance of the models drastically. In the case of backdoor attacks manipulated samples were added to the training data. It was shown that these attacks can be highly successful, even with just a few added samples. This indicates that a potentially large threat through these attacks exists. However, the success of backdoor attacks also heavily depends on the underlying training data set, illustrating the need for further research examining the factors that contribute to the success of this kind of attack. Lastly, a modified version of the Feature Importance Guided Attack [3] was used for the AE attacks. These attacks can also be very successful against both model types. Modifications of just one or few features can have a strong effect. The threat level of this attack depends on how easily those features can be manipulated in a real-world case. Additionally, this attack can be executed by an attacker with little knowledge about the ML based application.

The research shows that overall, DNNs and GBDT models are clearly vulnerable against different attacks. Past research in this domain mainly focused on homogenous data [4, 5]. Therefore, this work provides important implications regarding the vulnerability of ML models in a setting with tabular (insurance) data. Hence, depending on the application context potential vulnerabilities of the models need to be evaluated and mitigated.

## Type of presentation

Contributed Talk

**Primary author:** KÜHLEM, Robin (Debeka)**Co-authors:** Prof. MAUTHE, Andreas (University of Koblenz); Dr OTTEN, Daniel (Debeka); Dr LUDWIG, Daniel (Debeka); ROSENBAUM, Alexander (University of Koblenz); Prof. HUDDE, Anselm (Debeka and Koblenz University of Applied Sciences)**Presenter:** KÜHLEM, Robin (Debeka)**Session Classification:** Contributed session

**Track Classification:** Spring Meeting



Contribution ID: 50

Type: **not specified**

## Automated Supervision and Validation of On-Site Consumption Data at TU Dortmund University

*Wednesday, 15 May 2024 14:40 (3 minutes)*

We present an automated script, which controls the meter readings for electricity, water, heat and cold at TU Dortmund University, Germany. The script combines historic and current consumption data and calculates individual forecasts for every meter. These one-step-ahead forecasts are compared with true values afterwards to identify deviation from the regular energy consumption or anomalies. The script also detects missing values, which indicate that the meter is not working (correctly), or the connection is lost and some other special cases.

The goal of our script is to detect errors in the meter system automatically and faster than by hand. We compare the daily values with our forecasts to identify anomalies, i.e. meters with high electricity, water, heat or cold consumption. This will help to reduce the workload for the facility management and the waste of energy and thus makes the university more sustainable and efficient.

To this end, we compared multiple methods with respect to forecast accuracy and computing effort: a basic linear model, generalized additive model with spline based smoothing, SARIMA model with a seasonal period of seven, Holt-Winters additive-seasonal model, Regression with SARIMA errors (SARIMAX), and Random Forest. Holt-Winters and SARIMAX turned out to be the 'best' models and were implemented. In addition to monitoring and prediction, we thereby also identify important features effecting the consumption.

### Type of presentation

Poster

**Primary author:** FESCA, Felix (TU Dortmund University)**Co-author:** PAULY, Markus (TU Dortmund University)**Presenter:** FESCA, Felix (TU Dortmund University)**Session Classification:** Poster**Track Classification:** Spring Meeting

Contribution ID: 51

Type: **not specified**

# Training Gradient Boosted Decision Trees on Tabular Data Containing Label Noise for Classification Tasks

*Thursday, 16 May 2024 10:00 (20 minutes)*

Label noise, the mislabeling of instances in a dataset, is harmful to classifier performance, increases model complexity, and impairs adequate feature selection. It is frequent in large scale datasets and naturally occurs when human experts are involved. While extensive research has focused on mitigating label noise in image and text datasets through deep neural networks, there exists a notable gap in addressing these issues within Gradient Boosted Decision Trees (GBDTs) and tabular datasets.

This study aims to bridge this gap by adapting two noise detection methods, originally developed for deep learning, to enhance the robustness of GBDTs. Through this adaptation, we aim to augment the resilience of GBDTs against label noise, thereby improving their performance and reliability in real-world applications. The algorithms' effectiveness is rigorously tested against several benchmark datasets that have been intentionally polluted with various amounts and types of noise.

One of the devised algorithms achieves with state-of-the-art noise detection performance on the Adult dataset, showcasing its potential to effectively identify and mitigate label noise.

The investigation extends to analyzing the overarching effects of label noise on the performance of GBDTs the challenges of different types of noise, and the effectiveness of various noise treatment strategies.

The insights derived from this study not only enhance our understanding of the detrimental effects of label noise on the accuracy and reliability of GBDTs but also inform practical guidelines for handling such noise. Through rigorous analysis, the study proposes a direction for future research in enhancing GBDTs' resilience to label noise and ensuring their continued success in tabular data classification tasks.

## Type of presentation

Contributed Talk

**Primary author:** EISENBÜRGER, Anita (Debeka)**Co-authors:** Dr OTTEN, Daniel (Debeka); Prof. HOPFGARTNER, Frank (Universität Koblenz)**Presenter:** EISENBÜRGER, Anita (Debeka)**Session Classification:** Contributed session**Track Classification:** Spring Meeting

Contribution ID: 52

Type: **not specified**

## Adaptive Conformal Prediction by Reweighting Nonconformity Score

*Thursday, 16 May 2024 11:50 (25 minutes)*

Despite attractive theoretical guarantees and practical successes, Predictive Interval (PI) given by Conformal Prediction (CP) may not reflect the uncertainty of a given model. This limitation arises from CP methods using a constant correction for all test points, disregarding their individual epistemic uncertainties, to ensure coverage properties. To address this issue, we propose using a Quantile Regression Forest (QRF) to learn the distribution of nonconformity scores and utilizing the QRF's weights to assign more importance to samples with residuals similar to the test point. This approach results in PI lengths that are more aligned with the model's uncertainty or the epistemic uncertainty. Further, the weights learnt by the QRF provide a partition of the features space, allowing for more efficient computations and improved adaptiveness of the PI through groupwise calibration. Our approach enjoys an assumption-free finite-sample marginal and training-conditional or PAC coverage, and under suitable assumptions, it also ensures asymptotic conditional coverage. Our methods work for any nonconformity score and are available as a Python package. We conduct experiments on simulated and real-world data that demonstrate significant improvements compared to existing methods.

### Type of presentation

Invited Talk

**Primary author:** I. AMOUKOU, Salim (University Paris Saclay)**Co-author:** Mr BRUNEL, Nicolas (LaMME (Laboratoire de Modélisation et Mathématiques d'EVRY), EDMH)**Presenter:** I. AMOUKOU, Salim (University Paris Saclay)**Session Classification:** Invited session**Track Classification:** Spring Meeting

Contribution ID: 53

Type: **not specified**

## On the trustworthiness of simulation based uncertainty evaluations for industrial measurement instruments like CMMs

Wednesday, 15 May 2024 14:20 (20 minutes)

In metrology, the science of measurement, as well in industrial applications in which measurement accuracy is of importance, it is required to evaluate the uncertainty of each measurement. For complex instruments like an industrial work horse as the coordinate measurement machine (CMM), evaluating the uncertainty can be a similarly complex task. To this purpose a simulation model, often called a virtual experiment, virtual instrument or digital metrological twin, is created, with the help of which a task specific measurement uncertainty can be determined [1]. The main metrological guidance documents that can be used in these circumstances are the Guide to the Evaluation of Uncertainty (GUM) [2] and its first supplement [3]. Various implementation can be thought of as being in line with the ideas of this document. In earlier papers some aspects related to sensitivity to input values [4] and GUM-conformity [5] were considered.

In this contribution we will analyse how different ways of performing the computer instrument lead to different values for the measurement uncertainty. This will be mainly done by means of an experimental numerical study involving a virtual CMM. In this simplified two-dimensional numerical model, the scale errors of the axes of the CMM as well as their deviation from orthogonality are modelled, together with fully random instrument noise. The object of interest is an imperfect circle of which the radius and non-circularity is to be determined.

By varying the input to the virtual experiment we will study the robustness of the outcome. The trustworthiness will be assessed in terms of frequentist long-run success rates of the calculated coverage intervals against the ground truth given by the virtual experiment. Although the GUM is not based on frequentist statistics, and long-run success rates are not mentioned as the way of validating an estimator with an uncertainty, we will argue that this is nevertheless a very useful way of validating uncertainty statements, as it is also not so clear how Bayesian statistical methods could be applied in a straightforward manner.

### References

- [1] B. van Dorp, H. Haitjema, and P. Schellekens, The virtual CMM method for three-dimensional coordinate machines, *Positions*, 2, 2002, pp. 634
- [2] Joint Committee for Guides in Metrology, Evaluation of measurement data –Guide to the expression of uncertainty in measurement, Sèvres, France: International Bureau of Weights and Measures (BIPM), 2008.
- [3] Joint Committee for Guides in Metrology, Evaluation of measurement data –Supplement 1 to the ‘Guide to the expression of uncertainty in measurement’–Propagation of distributions using a Monte Carlo method, Sèvres, France: International Bureau of Weights and Measures (BIPM), 2008.
- [4] G. Kok, G. Wübbeler, C. Elster, Impact of Imperfect Artefacts and the Modus Operandi on Uncertainty Quantification Using Virtual Instruments, *Metrology*, vol. 2, n°. 2, 2022, pp. 311-319.
- [5] G. Wübbeler, M. Marschall, K. Kniel, D. Heißelmann, F. Härtig (+ another 1 author), GUM-Compliant Uncertainty Evaluation Using Virtual Experiments, *Metrology*, vol. 2, n°. 1, 2021, pp. 114-127.

### Type of presentation

Contributed Talk

**Primary author:** KOK, Gertjan (VSL)

**Co-authors:** MARSCHALL, Manuel (Physikalisch-Technische Bundesanstalt); Mr VAN DIJK, Marcel (VSL)

**Presenter:** KOK, Gertjan (VSL)

**Session Classification:** Contributed session

**Track Classification:** Spring Meeting

Contribution ID: 54

Type: **not specified**

## Machine learning and interpretability methods to investigate Alzheimer's disease

*Wednesday, 15 May 2024 11:45 (25 minutes)*

The use of machine learning methods in clinical settings is increasing. One reason for this is the availability of more complex models that promise more accurate predictive performance, especially for the study of heterogeneous diseases with multimodal data, such as Alzheimer's disease. However, as machine learning models become more complex, their interpretability decreases. The reduced interpretability of so-called black-box models can have a negative impact on clinicians' and patients' trust in the decisions of a system. For this reason, several methods to overcome this problem have been developed. These methods are summarised under the terms of interpretable machine learning and explainable artificial intelligence.

The presented research investigates how methods from the domain of interpretable machine learning and explainable artificial intelligence can be used for the early detection of Alzheimer's disease. To this end, a systematic comparison of different machine learning and explanation methods is presented. The comparison includes black-box deep learning convolutional neural networks, classical machine learning models as well as models that are interpretable by design. For all models, several reasonable explanation methods such as SHAP and GradCAM were applied. Common problems such as model calibration and feature correlations, which often occur when working with explainability methods, were addressed during the investigations. It is validated whether the activated brain regions and learned connections in different models are biologically plausible. This validation can increase the trust in the decision of machine learning models. In addition, it is investigated whether the models have learned new, biologically plausible connections that can help in the development of new biomarkers.

### Type of presentation

Invited Talk

**Primary author:** BLOCH, Louise (Department of Computer Science, University of Applied Sciences and Arts Dortmund)

**Presenter:** BLOCH, Louise (Department of Computer Science, University of Applied Sciences and Arts Dortmund)

**Session Classification:** Invited session

**Track Classification:** Spring Meeting

Contribution ID: 55

Type: **not specified**

## Discrimination certain group of people through Artificial Intelligence.

*Wednesday, 15 May 2024 14:44 (3 minutes)*

Artificial intelligence plays an important role today. It facilitates or completely takes over office tasks such as writing, formatting, and correcting texts, or in the medical field, enabling early detection and diagnosis of diseases.

However, data provided by algorithms can significantly disadvantage certain individuals. The results of such discrimination are often noticed later and can lead to misunderstandings or even legal violations.

This literature review provides a general overview of the discrimination of certain groups of people by artificial intelligence. The search strategy is based on English and German word pairs and well-known databases, conducted using the PRISMA method. The literature found primarily focuses on various digital platforms (such as delivery platforms, job portals, or social platforms). Selected examples of injustice in delivery, advertising, and biometrics aim to demonstrate the diverse and unpredictable nature of AI discrimination.

So it is demonstrated how self-service booking makes it easier to assign tasks to drivers, but inadvertently results in negative impacts on their statistics and reputation due to factors like cancellations and participation in strikes. Additionally, Facebook Ad Platform assists in crafting personalized ads; however, the algorithm's use of gender-based images can result in gender bias. Biometric data provides valuable information for identifying individuals, but assigning demographic labels during analysis may also lead to discriminatory outcomes.

It is expected that the use of AI algorithms will increase in the future, and they will be tested before deployment to ensure that the algorithm is fair.

### Type of presentation

Poster

**Primary author:** KOVALSKAYA, Evgenia (FH Dortmund)

**Presenter:** KOVALSKAYA, Evgenia (FH Dortmund)

**Session Classification:** Poster

**Track Classification:** Spring Meeting

Contribution ID: 56

Type: **not specified**

## Trustworthy Federated Learning in Healthcare

*Wednesday, 15 May 2024 10:30 (25 minutes)*

Machine learning has transformed many industries, being employed not only on large centralized datasets, but increasingly on data generated by a multitude of networked, complex devices such as mobile phones, autonomous vehicles or industrial machines. However, data-privacy and security concerns often prevent the centralization of this data, most prominently in healthcare. Federated learning allows to train machine learning models in-situ, i.e., on the data-generating devices, without sharing data. For federated learning to be applicable in critical applications, such as healthcare, however, it must be trustworthy. That is, we need theoretical sound guarantees on data privacy and model performance. I will present the main challenges for achieving trustworthiness in federated learning, as well as recent approaches that address these challenges.

### Type of presentation

Invited Talk

**Primary author:** KAMP, Michael (Institute for AI in Medicine (IKIM), UK Essen)**Presenter:** KAMP, Michael (Institute for AI in Medicine (IKIM), UK Essen)**Session Classification:** Invited session**Track Classification:** Spring Meeting



Contribution ID: 57

Type: **not specified**

## On Trustworthiness of Large Language Models

*Thursday, 16 May 2024 15:45 (1 hour)*

Past few years have witnessed significant leaps in capabilities of Large Language Models (LLMs). LLMs of today can perform a variety of tasks such as summarization, information retrieval and even mathematical reasoning with impressive accuracy. What is even more impressive is LLMs' ability to follow natural language instructions without needing dedicated training datasets. However, issues like bias, hallucinations and lack of transparency pose a major impediment to wide adoption of these models. In this talk, I will review how we got from "traditional NLP" to today's LLMs, and some of the reasons behind trustworthiness issues surrounding LLMs. I will then focus on a single issue —hallucinations in factual question answering—and show how artifacts associated with model generations can provide hints that the generation contains a hallucination.

### Type of presentation

Invited Talk

**Primary author:** ZAFAR, Muhammad Bilal (Ruhr University Bochum and Research Center for Trustworthy Data Science and Security)

**Presenter:** ZAFAR, Muhammad Bilal (Ruhr University Bochum and Research Center for Trustworthy Data Science and Security)

**Session Classification:** Keynote

**Track Classification:** Spring Meeting

Contribution ID: 58

Type: **not specified**

## Benchmarking Trust: A Metric for Trustworthy Machine

*Thursday, 16 May 2024 15:05 (25 minutes)*

In the evolving landscape of machine learning research, the concept of trustworthiness receives critical consideration, both concerning data and models. However, the lack of a universally agreed upon definition of the very concept of trustworthiness presents a considerable challenge. The lack of such a definition impedes meaningful exchange and comparison of results when it comes to assessing trust. To make matters worse, coming up with a quantifiable metric is currently hardly possible. In consequence, the machine learning community cannot operationalize the term, beyond its current state as a hardly graspable concept.

In this talk, a first step towards such an operationalization of the notion of is presented – The FRIES Trust Score, a novel metric designed to evaluate the trustworthiness of machine learning models and datasets. Grounded in five foundational pillars – fairness, robustness, integrity, explainability, and safety – this approach provides a holistic framework for trust assessment based on quality assurance methods. This talk further aims to shed light on the critical importance of trustworthiness in machine learning and showcases the potential of the implementation of a human-in-the-loop trust score to facilitate objective evaluations in the dynamic and interdisciplinary field of trustworthy AI.

### Type of presentation

Invited Talk

**Primary author:** RUTINOWSKI, Jérôme (TU Dortmund University)**Presenter:** RUTINOWSKI, Jérôme (TU Dortmund University)**Session Classification:** Invited session**Track Classification:** Spring Meeting

Contribution ID: 59

Type: **not specified**

## Towards Compliance of AI Algorithms : Bias Analysis and Robustness.

*Thursday, 16 May 2024 09:00 (1 hour)*

AI is about to revolutionize many sectors of society and will soon have significant economic, legal, social and regulatory consequences.

In the world of production, transport, human resource management, and health, to name but a few, a growing share of diagnostic and planning processes is operated by AI-based systems.

Controlling the risks of deploying these AI for high risk systems is therefore becoming a considerable challenge for a variety of actors in society. The authorities in charge of regulation, in the first place, but also the producers of these systems themselves as well as the companies that buy and operate them.

While their performance gains are no longer in doubt, their interpretability and the transparency of their decisions is now considered crucial for their wider deployment. Auditing these processes based on artificial intelligence, if we refer to the DSA, DMA and AI texts, will become a necessity in the same way as the quality standards that govern the production of goods and market services. However, auditing a complex algorithm, based on AI-based components –to the point of identifying and measuring its risks and biases–is a scientifically and technically delicate task.

We will explain how mathematical methods based on the theory of optimal transport can provide a natural framework able to handle algorithmic biases, from their quantification and measures, to the mitigation process. Audit methods based on such tools can be used to certify the presence of disloyal behavior that may lead to possible discrimination in numerous fields but in an industrial context can also be used to identify deviations in the data that may lead to a loss of performance.

### Type of presentation

Invited Talk

**Primary author:** LOUBES, Jean-Michel (Université Toulouse)

**Presenter:** LOUBES, Jean-Michel (Université Toulouse)

**Session Classification:** Keynote

**Track Classification:** Spring Meeting

Contribution ID: 60

Type: **not specified**

## Statistical Inference for Trustworthy AI: Cases for xAI and Uncertainty Quantification

*Wednesday, 15 May 2024 09:15 (1 hour)*

Trustworthy AI is dedicated to the development of methodologies and proofs that demonstrate the “proper” behavior of Artificial Intelligence algorithms, in order to favor their acceptance by users and organizations. By considering explainable AI and Uncertainty Quantification, we will show that defining consistent inferential procedures give systematic, reliable and arguable information for users. Starting from the Shapley Values for Importance Attributions, whose standard computations and interpretations have important limitations, we introduce the concept of Same Decision Probability that permits to identify import local and regional variables, and for which we can derive statistical consistency. Hence, regional measures of importance for variables appears to be a good scale for deriving consistent explanations, such as sufficient explanations, instead of local measures. In a second part, we will discuss the usefulness and potential of conformal prediction for deriving prediction intervals with guaranteed coverage rate. We will insist on the genericity and flexibility of the algorithms that permits to develop distribution-free inference for a large set of AI tasks, hence providing reliability measures useful for interacting with users.

### Type of presentation

Invited Talk

**Primary author:** BRUNEL, Nicolas (LaMME (Laboratoire de Modélisation et Mathématiques d’EVRY), EDMH)

**Presenter:** BRUNEL, Nicolas (LaMME (Laboratoire de Modélisation et Mathématiques d’EVRY), EDMH)

**Session Classification:** Keynote

**Track Classification:** Spring Meeting

Contribution ID: 61

Type: **not specified**

# A Gradient-Enhanced Neural Network as a Surrogate for Elastoplastic Finite Element Analysis

*Wednesday, 15 May 2024 14:00 (20 minutes)*

When monitoring complex manufacturing processes, various methods, for instance the optimization of observed systems or quantification of their uncertainty, are applied to support and improve the processes. These methods necessitate repeated evaluations of the systems associated responses. While complex numerical models such as finite element models are capable of this, their solutions come with high computational cost with increasing complexity. In certain cases, artificial neural networks are suitable as surrogate models with less computational cost while maintaining a certain degree of accuracy. In general supervised learning, an artificial neural network trains on data, whereby a data loss evaluates the difference between available data and computed model response. For such a surrogate model, a computationally expensive, but accurate preliminary numerical model evaluates its system according to inputs to produce its training data. Often, these numerical models can provide further data, such as sensitivities, through computationally cheap adjoint methods. By including these sensitivities with respect to the inputs, the performance of the training convergence is improved. This Sobolev training adjusts the defined data loss of the neural network in order to consider the sensitivities. Instead of additional outputs for the neural network model, these sensitivities are computed by the derivatives of the model output through its layers. This expansion of the data loss leads to the consideration of appropriate weighting of each individual response and sensitivity loss. Specifically in this work, the goal is to define a second, parallel optimization process during training, with which to determine the optimal weighting of all individual losses, for optimal convergence performance. A finite element model is prepared, which evaluates various output variables for a given mechanical system of linear or nonlinear behavior, to generate a small dataset. Then a neural network is Sobolev-trained with this dataset. During training, a parallel optimization process occurs for the weighting of each individual loss. We explore this by applying a set of residual weights to the Sobolev loss function and then optimizing a predefined target function in regards to the loss for the set of residual weights. The results demonstrate that applying certain residual weight optimization methods improve convergence performance and not only reduce the total range of accuracy among trained models, but also shift the range to a better accuracy.

## Type of presentation

Contributed Talk

**Primary authors:** KILICSOY, Ali Osman Mert (TU Dortmund, CRE); BARTHOLD, F.-J.; LIEDMANN, J.; VALDEBENITO, M.A.; FAES, M.G.R

**Presenter:** KILICSOY, Ali Osman Mert (TU Dortmund, CRE)

**Session Classification:** Contributed session

**Track Classification:** Spring Meeting

Contribution ID: 62

Type: **not specified**

## Explainable AI and clinical real-world data for personalized cancer treatment

*Wednesday, 15 May 2024 10:55 (25 minutes)*

Although a large amount of data is collected on each patient during cancer care, clinical decisions are mostly based on limited parameters and expert knowledge. This is mainly due to insufficient data infrastructure and a lack of tools to comprehensively integrate diverse clinical data. At University Hospital Essen, medical data is stored in FHIR format, enabling cutting-edge analyses of real-world patient journeys. Based on the multimodal data from more than 15,000 cancer patients, explainable AI (xAI) can model individual patient outcomes, integrating clinical records, image-derived body compositions, and genetic data. xAI makes it possible to assess the prognostic contribution of each parameter at both the patient and cohort level and provide AI-derived (AID) markers for clinical decision support. This demonstrates how efficient hospital data management, combined with AI techniques, can fundamentally transform cancer care.

### Type of presentation

Invited Talk

**Primary author:** KEYL, Julius (UK Essen)**Presenter:** KEYL, Julius (UK Essen)**Session Classification:** Invited session**Track Classification:** Spring Meeting

Contribution ID: 63

Type: **not specified**

# Interpretable AI in Medicine: Generating Radiological Reports with Panoptic Scene Graphs

*Wednesday, 15 May 2024 11:20 (25 minutes)*

We explore the integration of panoptic scene graphs in the field of chest radiographs, to enhance explainable medical report generation. Panoptic scene graphs require a model to generate a more comprehensive scene graph representation based on panoptic segmentations rather than rigid bounding boxes and thus present a more holistic image representation. These graphs facilitate accurate report generation by structuring the diagnostic process in an interpretable data format and enabling a precise mapping between relevant findings and visual regions. Through the utilization of the generated graph structures, large language models can produce reliable medical reports, for which medical practitioners through the underlying datastructure gain insights into AI-generated decisions, thus fostering a deeper understanding of the underlying rationale. This advancement marks a significant step towards enhancing transparency and interpretability in medical AI systems, ultimately improving patient care and clinical decision-making.

## Type of presentation

Invited Talk

**Primary author:** SEIBOLD, Constantin (University Klinik Essen)

**Presenter:** SEIBOLD, Constantin (University Klinik Essen)

**Session Classification:** Invited session

**Track Classification:** Spring Meeting

Contribution ID: 64

Type: **not specified**

# Global Sensitivity Analysis for Assessing Fairness of Artificial Intelligence for Medical Data

*Wednesday, 15 May 2024 14:48 (3 minutes)*

The utilisation of Artificial Intelligence (AI) in medical practice is on the rise due to various factors. Firstly, they can process large datasets and recognise complex relationships that may be difficult for humans to discern in the enormous amount of medical data. Therefore AI systems can enhance the efficiency and accuracy of medical processes, thus saving resources.

Nevertheless, the utilisation of AI in medical practice is not without risk. Automated decision-making processes may contain unconscious biases that could result in disparate treatment and discrimination against specific patient groups. The measurement of fairness in AI is an evolving field of research, with no standardised definition or method for assessing fairness.

This contribution discusses global sensitivity analysis as a tool for evaluating the fairness of AI deployment in medical practice. A concrete example is used to demonstrate how sensitivity analysis can be performed in R to identify potential discrimination. The results emphasise the need to implement mechanisms to ensure fairness in the use of AI in medical practice and raise awareness that human biases may also influence automated decision-making processes.

## Type of presentation

Poster

**Presenter:** KUHLMANN DE CANAVIRI, Lara (Fachhochschule Dortmund)**Session Classification:** Poster



Contribution ID: 65

Type: **not specified**

# A Robust Multivariate Profile Monitoring Framework

*Wednesday, 15 May 2024 14:51 (3 minutes)*

In classical statistical process monitoring (SPM) applications the Phase I sample is assumed to come from an in-control process, which is however not always valid, especially when the monitoring characteristic for each item/case is a vector of profiles, i.e., a multivariate profile.

The presence of untrustable observations, or, in general, of outliers, especially in high-dimensional settings, can significantly bias and reduce the power of the SPM framework in detecting anomalies.

In particular, when the dimensionality of the data is high, the fraction of perfectly observed cases can be very small, and outliers may occur more realistically in one or a few components only (componentwise), which may be difficult to identify, rather than in all components (casewise). On the other hand, in these cases, down-weighting or eliminating entire observations of multivariate profiles that are contaminated in one or a few components might be unacceptably wasteful.

This research introduces a novel monitoring framework for multivariate functional quality characteristics, named robust multivariate functional control chart (RoMFCC), that is robust to the influence of both functional casewise and componentwise outliers. The RoMFCC framework contains four main elements: a functional filter to detect functional componentwise outliers, a robust imputation of missing components in multivariate functional data, a robust dimension reduction that deals with functional casewise outliers, and a procedure for prospective process monitoring. The performance of the proposed framework is assessed through a wide Monte Carlo simulation also in comparison to competing monitoring schemes that have already appeared in the literature before. The practical applicability of the RoMFCC is demonstrated through a case study in the SPM of a resistance spot welding process in automotive body-in-white manufacturing.

The RoMFCC is implemented in the R package `funcharts`, openly available on CRAN.

## Acknowledgements

The research activity of A. Lepore and F. Centofanti was carried out within the MICS (Made in Italy –Circular and Sustainable) Extended Partnership and received funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) –MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.3 –D.D. 1551.11-10-2022, PE00000004). The research activity of B. Palumbo was carried out within the MOST - Sustainable Mobility National Research Center and received funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) –MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.4 –D.D. 1033.17-06-2022, CN00000023). This work reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

## Type of presentation

Poster

**Presenter:** Dr LEPORE, Antonio (Università degli Studi di Napoli Federico II - Dept. of Industrial Engineering)

**Session Classification:** Poster

Contribution ID: 66

Type: **not specified**

## Making data models better explainable and actionable using JMP Profilers

*Wednesday, 15 May 2024 16:35 (40 minutes)*

JMP software converts data into insights with no coding required, and is a leading solution for real-world problem-solving in many industries. Some users call the JMP Profiler, the key tool for any data modeler, “the coolest thing in JMP”. This presentation will demonstrate several Profilers and various use cases and discuss its value in both industrial settings as well as in teaching and learning.

Profilers are interactive visualizations of any model built in JMP, being tree-based, regression models, neural networks or other predictive models. The profiles are cross-section views of the response surface for any number of factors (Xs) and responses (Ys). All factors can be changed interactively to see the effects on the response(s) and on other profiles. Additional Profiler features help with the model understanding and interpretation, like confidence intervals, overlaid data points or interaction traces, sensitivity indicators and extrapolation warnings.

Based on the desirability representing the goals for each response, like maximize, minimize or match target, the Profiler can also find the best factor settings to optimize the response(s) for the system or process at hand. A built-in Monte-Carlo Simulation and Gaussian Process model helps to find more robust settings in the light of any stochastic variation of the factors.

Beside the Prediction Profiler, we will also demonstrate a Contour Profiler, an Interaction Profiler and a Design Space Profiler –all interactive and visual tools to get the most out of your models.

### Type of presentation

**Presenter:** KRAFT, Volker (JMP)

**Session Classification:** JMP

Contribution ID: 67

Type: **not specified**

## Trust in automated systems as a multidimensional psychological construct

*Thursday, 16 May 2024 13:30 (20 minutes)*

Machine learning models are often the basis of current automated systems. Trust in an automated system is typically justified only up to a certain degree: A moderately reliable system deserves less trust than a highly reliable one. Ideally, trust is calibrated, in the sense that a human interacting with a system neither over- nor undertrusts the system. To be able to relate objective measures of reliability like classification accuracy, fairness or robustness measures to perceived trust, the latter needs to be quantified. However, trust is no unidimensional construct, with several related facets determining trust. Existing psychometric questionnaires have several shortcomings. By building on existing theories from a range of fields, we present a theoretically well-founded questionnaire that includes 30 five point Likert scale items for six dimensions of trust: Global Trust, Integrity, Unbiasedness, Perceived Performance, Vigilance and Transparency. The Global Trust items are intended to be used as an economic short form. The questionnaire's performance has been evaluated in several studies, including an English and a German version. Here, we focus on the largest English language sample of  $N = 883$  that was used to derive the final TrustSix scale from a larger initial item pool. Perceived trust in three vignettes (fictional automated systems) is measured, i.e., systems for skin cancer detection, poisonous mushroom detection and automated driving, each based on machine learning models. Special emphasis has been placed to explore the exact factorial structure of the latent variables and check their stability across vignettes. A Global Trust factor could be discovered with the help of a bifactor rotation, with five additional factors for the more specific trust dimensions. Reliability of each 5 item subscale is satisfactory ( $\alpha = .76 - .96$ ), and with satisfactory overall reliability for the main factor (hierarchical McDonald's  $\omega = .75 - .80$ , total McDonald's  $\omega = .97-.98$ ), and correlations with adjacent constructs indicating sufficient discriminant validity.

### Type of presentation

**Primary authors:** WISCHNEWSKI, Magdalena (TU Dortmund University); BEISEMANN, Marie (TU Dortmund University); KRÄMER, Nicole (University of Duisburg-Essen); DOEBLER, Philipp (TU Dortmund - Fakultät Statistik)

**Presenter:** DOEBLER, Philipp (TU Dortmund - Fakultät Statistik)

**Session Classification:** Contributed session