



Deutsches Forschungsinstitut
für öffentliche Verwaltung



Institut für
Gesetzesfolgenabschätzung
und Evaluation

Lernende Algorithmen im Evaluationskontext: Das Beispiel „Boosted Regression Trees“

Christoph E. Müller

[mueller@foev-speyer.de]

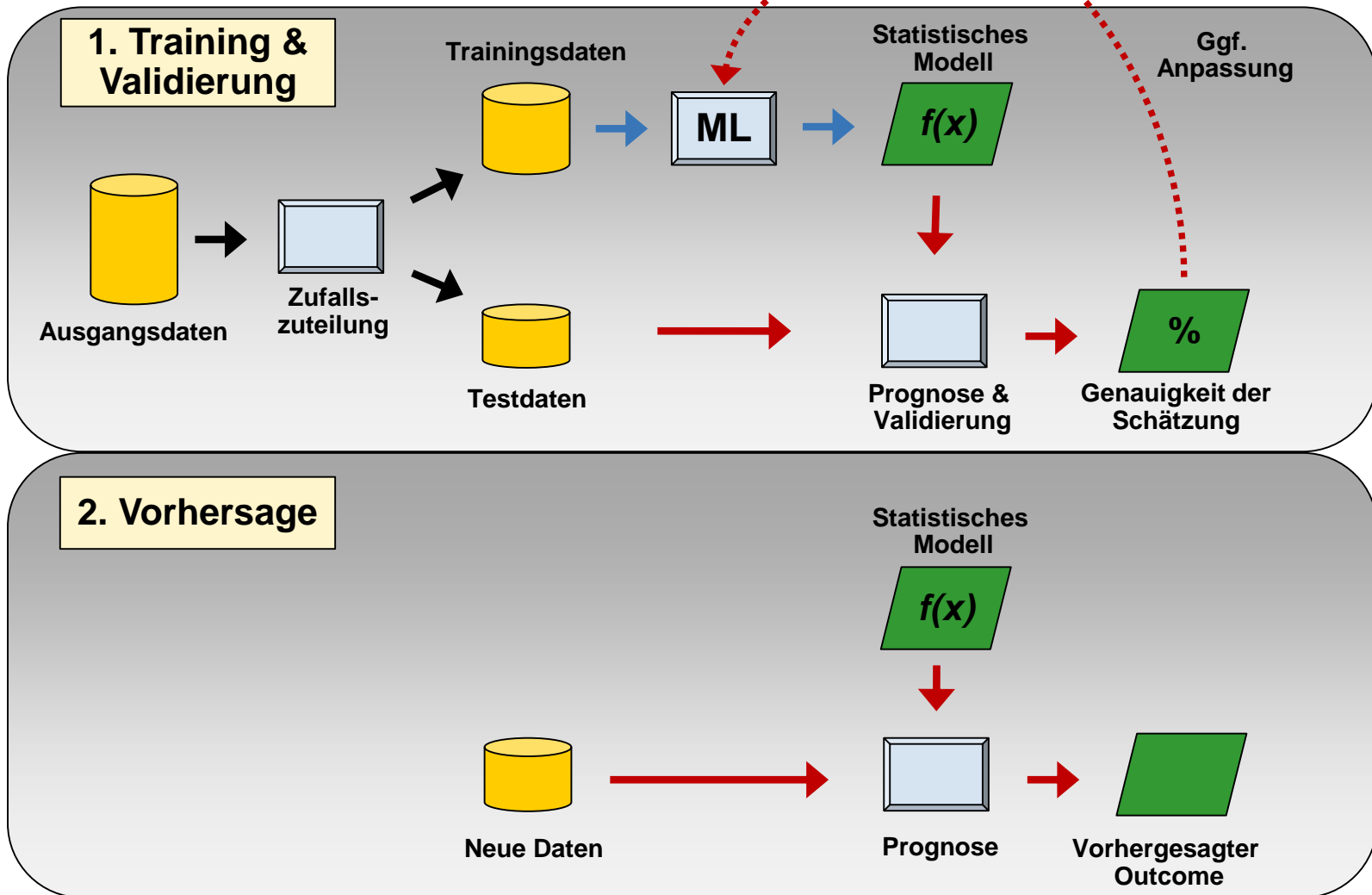
DeGEval-Jahrestagung 2017

Session „Big Data: Große Datenmengen – Großes Potenzial für die Evaluation?“

Was ist Machine Learning (ML)?

- ML-Algorithmen: Befehlsanordnungen für Rechensysteme, mit denen meist große Datenmengen („big data“) analysiert werden.
- ML-Algorithmen sind i.d.R. Prognoseverfahren, können aber auch zur Exploration, Erklärung oder Identifikation genutzt werden.
- ML-Algorithmen erlernen Muster/Zusammenhänge in Daten selbständig, ohne im Vorfeld bereits weitreichende Annahmen zu machen (im Gegensatz zu vielen „klassischen Verfahren“).
- Zahlreiche Anwendungsfelder: Text-/Spracherkennung, Medizin, personalisierte Werbung, Gesichtserkennung, Kriminalistik, etc.
- Für die Evaluation vor allem interessant: Alternative zu „klassischen Verfahren“ der quantitativen Sozialforschung

Ablauf Machine Learning



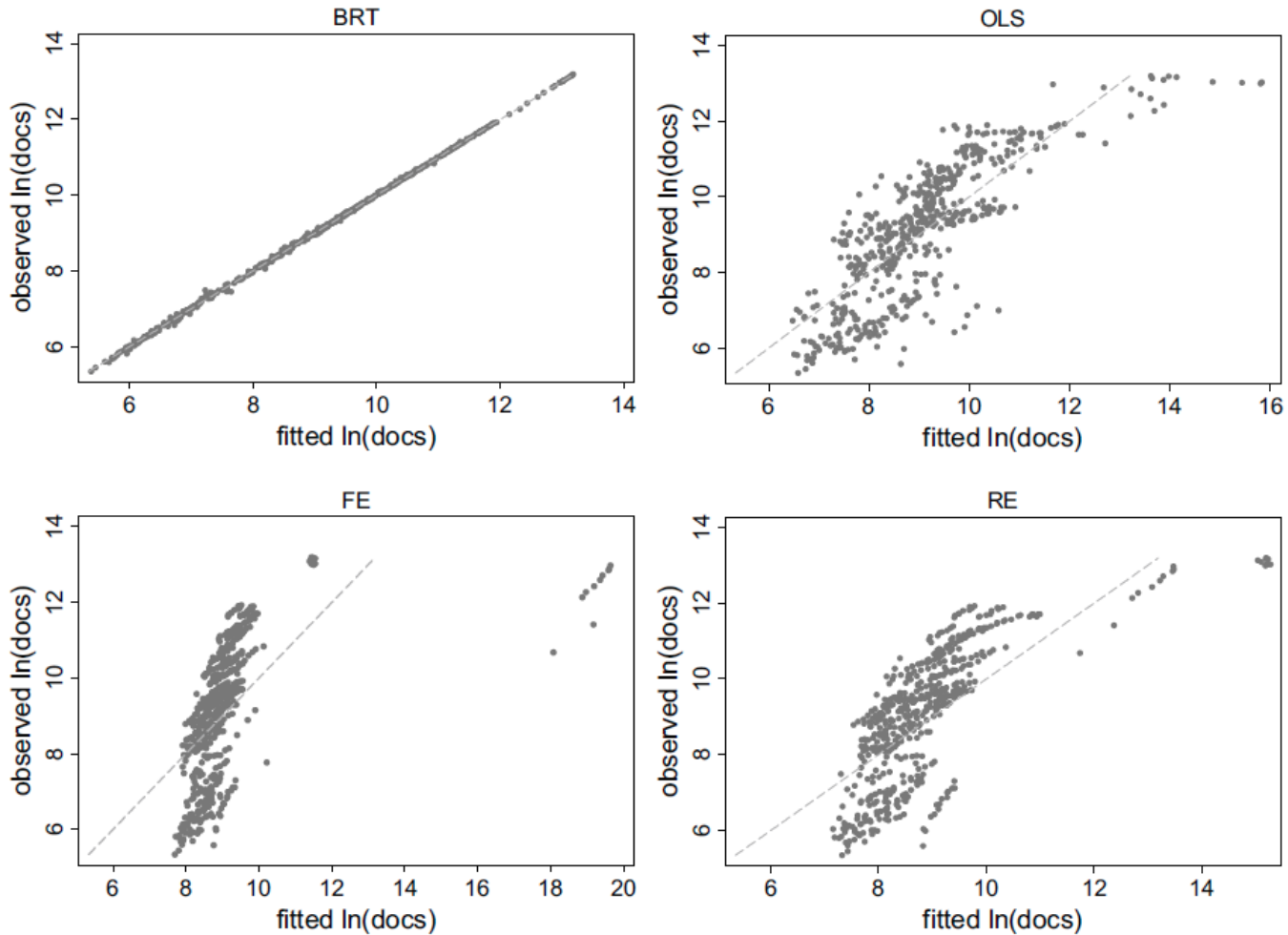
Nachbildung von: <https://www.edureka.co/blog/introduction-to-supervised-learning/>

Boosted Regression Trees (BRT)

- BRT ist eine „ensemble method“ → Kombination von Algorithmen
 - Classification and Regression Trees (CART)
 - Boosting
- BRT ist ein Prognoseverfahren analog zu linearer, logistischer oder Count-Data Regression (z.B. Poisson)
- BRT liefert keine Koeffizienten und p -Werte als Ergebnis, sondern relative Einflüsse von Variablen und grafische Zusammenhänge
- Deutlich geringere Anforderungen an die Datenstruktur, z.B.
 - Keine funktionalen Annahmen (z.B. Linearität)
 - Skalenniveau der unabh. Variablen spielt keine Rolle
 - Anzahl der Prädiktoren spielt keine Rolle
 - (Relevante) Interaktionen werden automatisch berücksichtigt

Boosted Regression Trees (GBRT)

Model Performance



Quelle: Mueller, C. E. (2016). Accurate Forecast of Countries' Research Output by Macro-Level Indicators. *Scientometrics*, 109 (2), 1307-1328.

BRT: Anwendungsbeispiel

- Evaluation einer Intervention (Webseite) zur Aufklärung von Verbrauchern über den Klimaschutz
- Methode:
 - 1. Experimentelles Design zur Prüfung von Wirkungen
 - 2. Nachgelagerte Überprüfung der wirkfördernden Faktoren
- **Einsatz von BRT** zur Identifikation von Faktoren, die Wirkungen fördern.
- Abh. Variable: Lerneffekt durch Webseitenbesuch
- Unabh. Variablen:
 - Informationsqualität (INFOQUAL)
 - Qualität der Webseite (WEBQUAL)
 - Emotionen (EMO)
 - Persönliche Themenrelevanz (TR)
 - Geschlecht & Alter

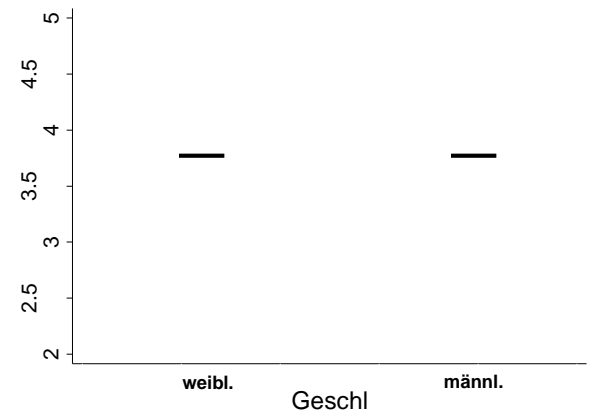
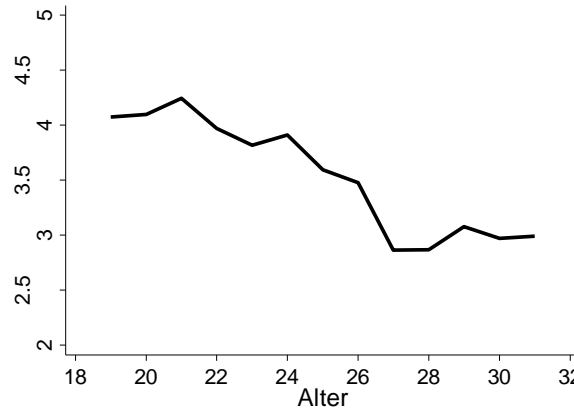
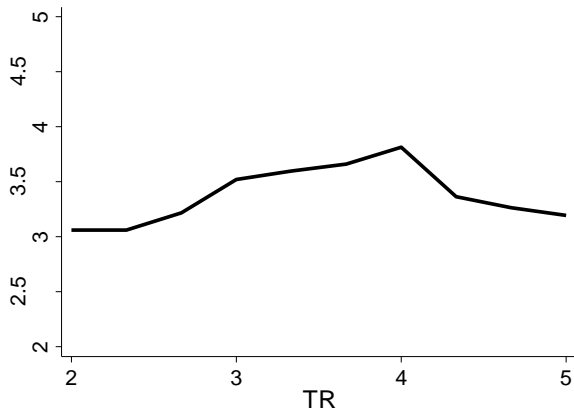
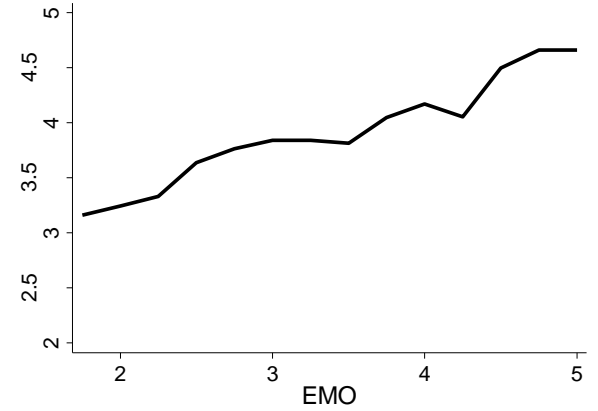
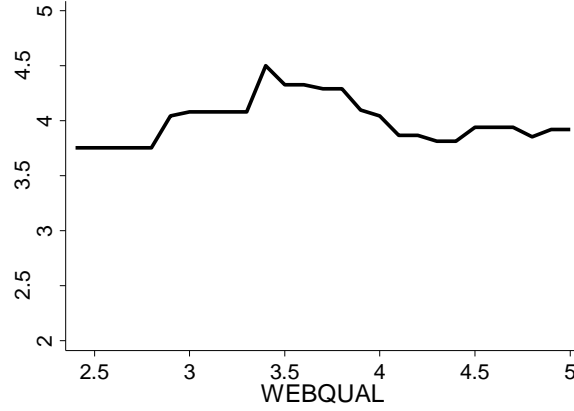
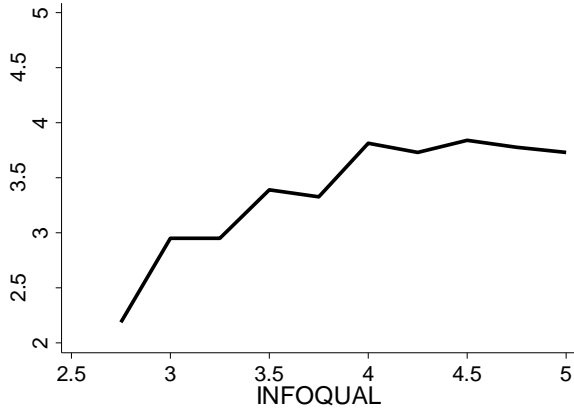
```
Influence of each variable (Percent):  
41.758961 INFOQUAL  
11.912615 WEBQUAL  
22.093289 EMO  
13.83456 TR  
1.721334 Geschl  
8.6792415 Alter
```

$$R^2_{\text{Training}} = 97\%$$

$$R^2_{\text{Test}} = 76\%$$

Eigene Berechnungen. Daten aus Gaus, H., & Müller, C.E. (2013). Das Internet als Instrument zur Klimaschutzklärung von Verbrauchern: Eine empirische Untersuchung zu Wirkung und Wirkungsweise eines Informationsportals. *Umweltpsychologie*, 17 (1), 36-59.

BRT: Anwendungsbeispiel



Eigene Darstellung. Daten aus Gaus, H., & Müller, C.E. (2013). Das Internet als Instrument zur Klimaschutzklärung von Verbrauchern: Eine empirische Untersuchung zu Wirkung und Wirkungsweise eines Informationsportals. *Umweltpsychologie*, 17 (1), 36-59.

- Explorative Datenanalyse
 - Unterstützung der Hypothesengenerierung
 - Identifikation nicht-linearer Zusammenhänge
- Prognose von Ereignissen/Outcomes
 - Z.B. um Ergebnisse von Pilotstudien zu generalisieren
 - Meist deutlich höherer Modellfit als „klassische“ Verfahren
- Hypothesenprüfung
 - Z.B. im Rahmen von theoriegeleiteter Evaluation
 - Weniger Annahmen als „klassische“ Verfahren
- Einsatz bei quasi-experimentellen Studien
 - Z.B. Bildung von Vergleichsgruppen

- Gefahr der Automatisierung/Enttheoretisierung von Evaluation
 - „Kitchen Sink Approach“: Verzicht auf theoretische Vorarbeit
- Keine Lösung der Kausalitätsproblematik
 - Sinnvolle Einbindung in entsprechende Untersuchungsdesigns ist erforderlich
- Anforderungen an die Größe von Datensätzen sind oft höher als bei herkömmlichen statistischen Verfahren
- Hohe statistische und technische Anforderungen an Evaluierende
 - Komplexe Algorithmen
 - Oft schwierig umzusetzen mit gängiger Statistiksoftware

Vielen Dank!