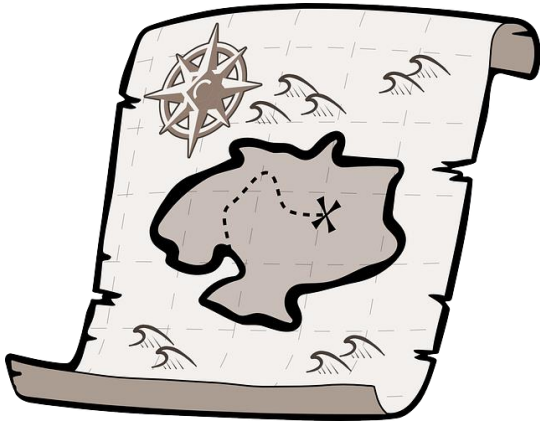
Modellgültigkeit

Generalisierungsorientierte Modellierungskultur

Themenfelder

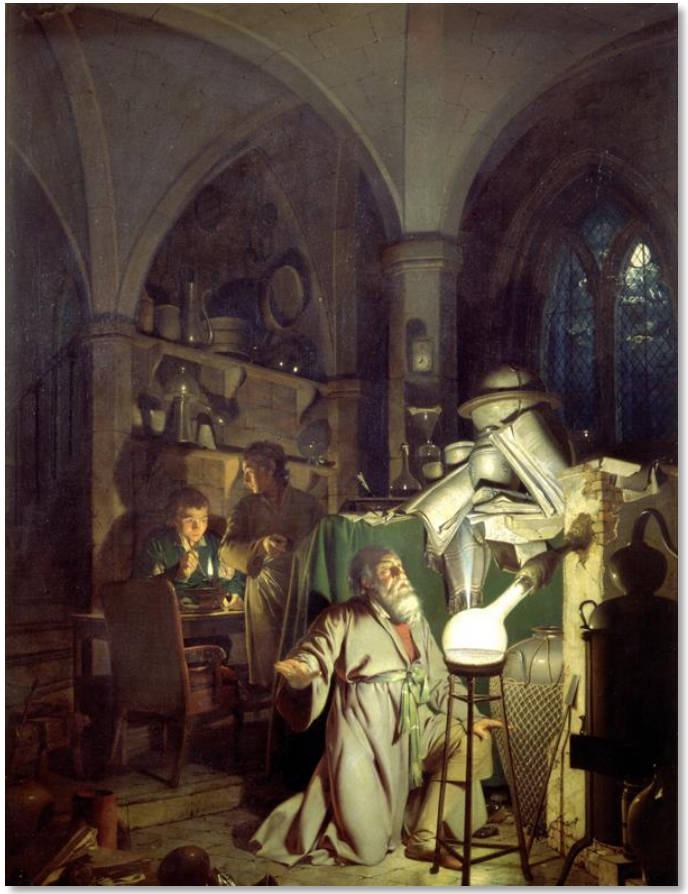
- Ziele einer empirischen Wissenschaft
- Kleiner Überblick Machine Learning
- Vergleich moderner und klassischer statistischer Methoden in der Forschung
- Prediction Performance: Gültigkeit für neue Daten
- Concept Drift: Verlust von Modellgültigkeit über Zeit
- Zusammenfassung und Perspektive



Ziele einer empirischen Wissenschaft

1. **Beschreibung**
   - **Deskriptive Statistik:** Zusammenfassende Maße und Grafiken, um die Daten verständlich zu machen
2. **Erklärung**
   - **Inferenzstatistik:** Schätzen von Modellparametern, welche Zusammensetzung und –hänge der Daten modellieren
3. **Vorhersage**
   - **Machine Learning:** Vorhersage neuer Daten, nachdem ein Modell durch Resampling und einen Lernalgorithmus trainiert wurde

➢Verallgemeinerung über **Raum** und **Zeit** ist Ziel



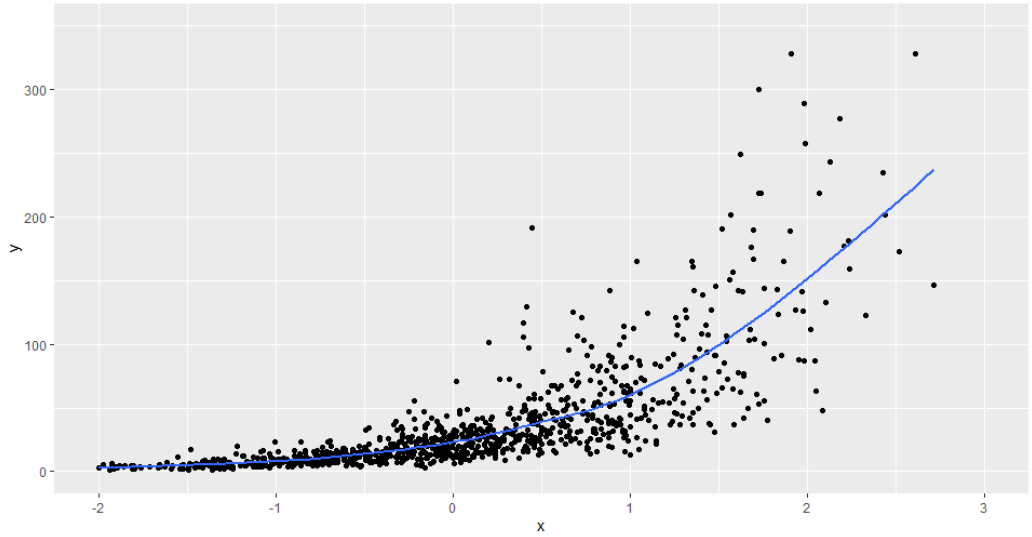
https://eclecticlight.co/2016/02/17/the-story-in-paintings-enlightened-by-science/

Erklärung und Vorhersage

Universität Regensburg

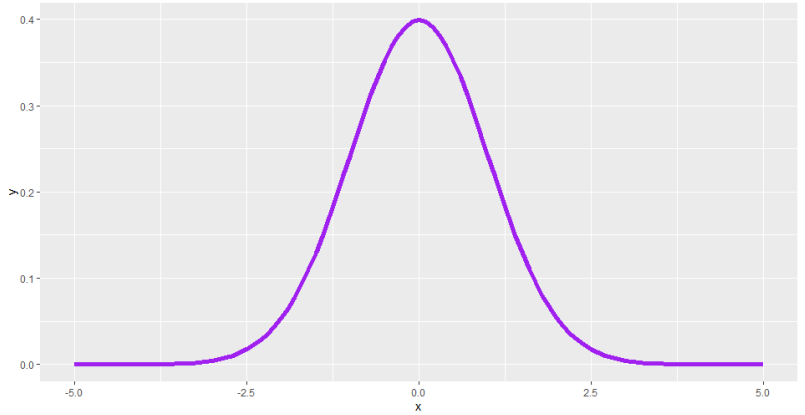
Leo Breiman: Zwei Kulturen der Statistik

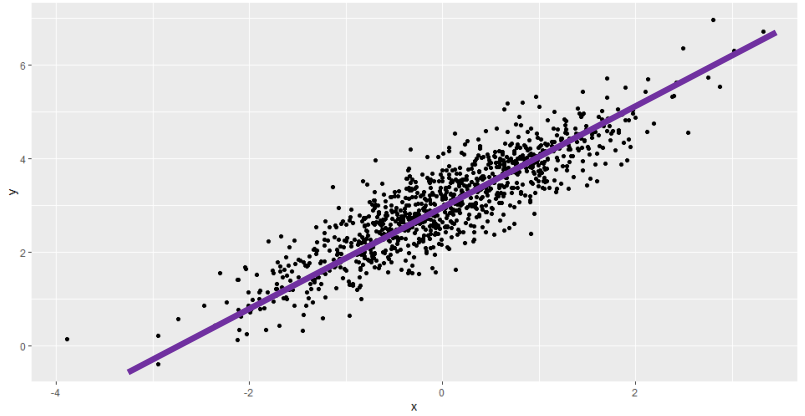
- Starke **theoretische Vorannahmen** über den datengenerierenden Prozess
  - z.B.: Exponentieller Zusammenhang
  - Klassische Statistik
    - Fokus auf Erklärung und Modellannahmen
    - *p*-Werte für Inferenz
- Annahme eines **unbekannten** datengenerierenden Prozesses
  - Machine Learning
    - Fokus auf **Vorhersagegüte**
    - Schätzung des Generalisierungsfehlers



Annahmen klassischer Modelle

- **Allgemeines Lineares Modell**
  - Normalverteilung der Fehler
    $$\varepsilon \sim N(0; \sigma^2)$$
  - Lineare Zusammenhänge
    $$y = \beta x + \varepsilon$$
- **Verallgemeinertes Lineares Modell**
  $$y = g(\beta x + \varepsilon)$$





Drei Phasen nach Efron & Hastie

Die Phasen der statistischen Modelle im 20. und 21. Jahrhundert (grob eingeteilt):

1. Klassische Inferenzstatistik
   - ALM, GLM etc.

2. Computationale Methoden
   - Bayesianische Statistik, Bootstrap

3. Computerintensive Methoden
   - Machine Learning

Kleiner Überblick Machine Learning

Modellarten

- Baumbasierte Modelle
  - Random Forest, Boosting
- Kernelbasierte Modelle
  - Support Vector Machines
- Deep Learning
  - Neuronale Netzwerkmodelle

Eigenschaften

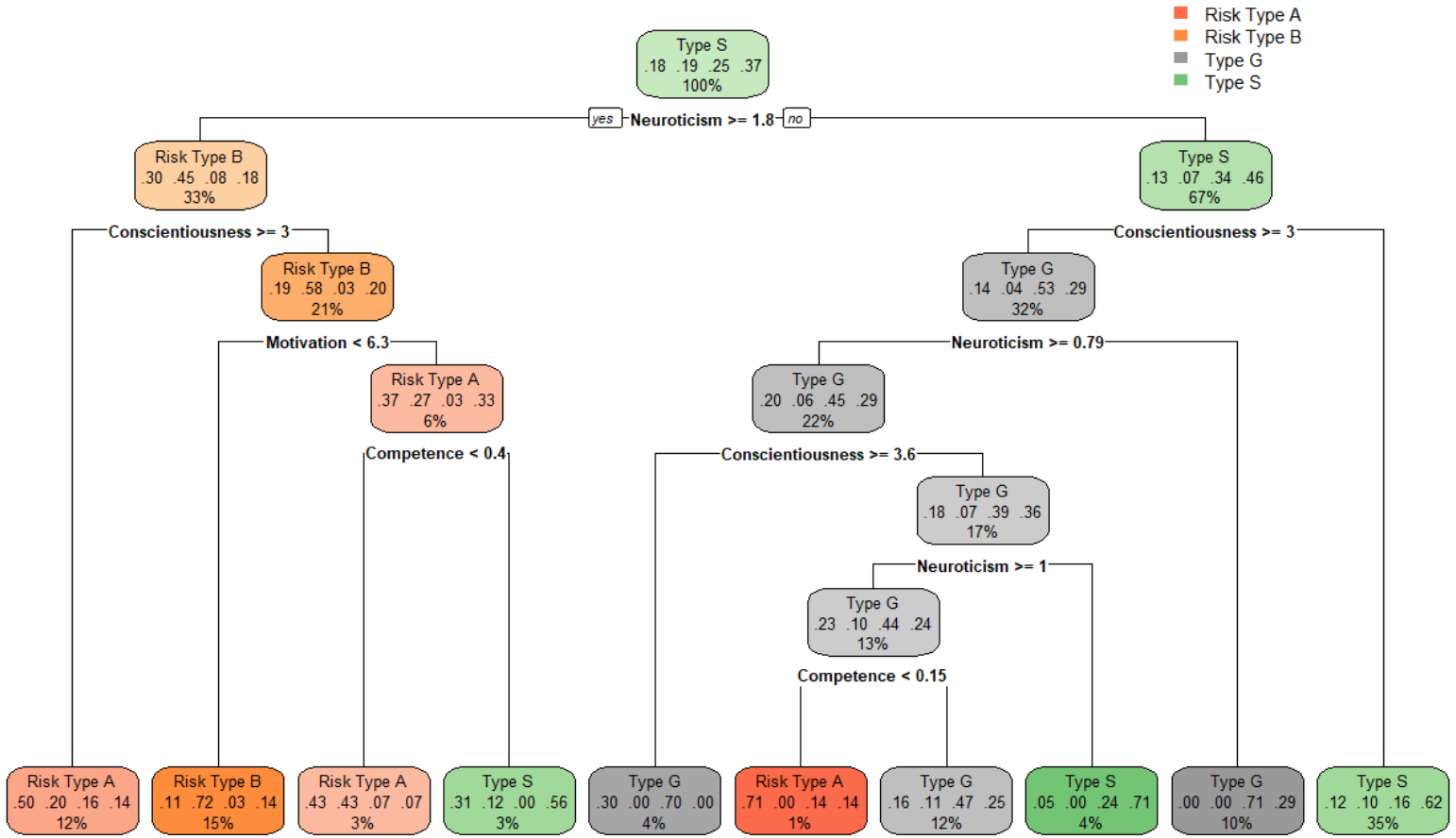
- Nutzen von **Resampling**
- **Optimiert auf Vorhersage** neuer Daten
- Weitgehend ohne direkt interpretierbare Parameter
- Hohe Funktionalität mit vielen Variablen und Daten

http://thecontextofthings.com/2016/04/18/transparency-in-business/

Beispiel Entscheidungsbaum

Universität Regensburg

Klassifizierung von vier Arbeitstypen

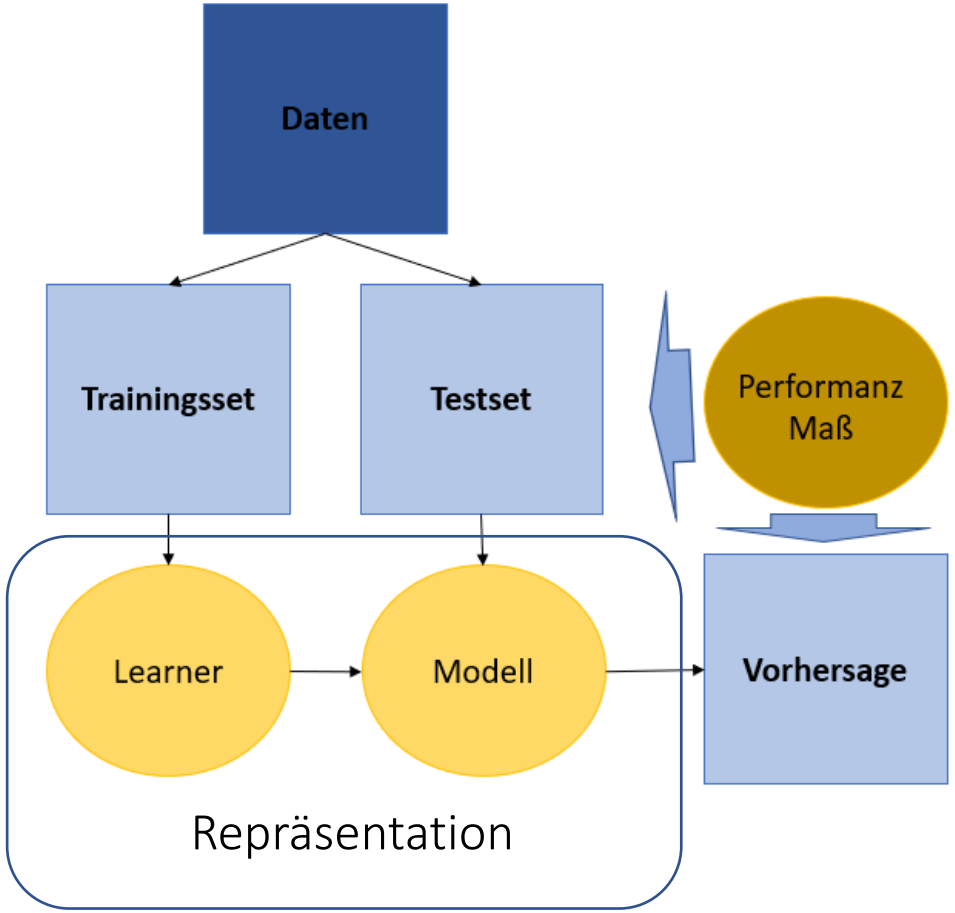


Grafik: Efron, B., & Hastie, T. (2016). *Computer age statistical inference*(Vol. 5). Cambridge University Press., S. 125

Resampling

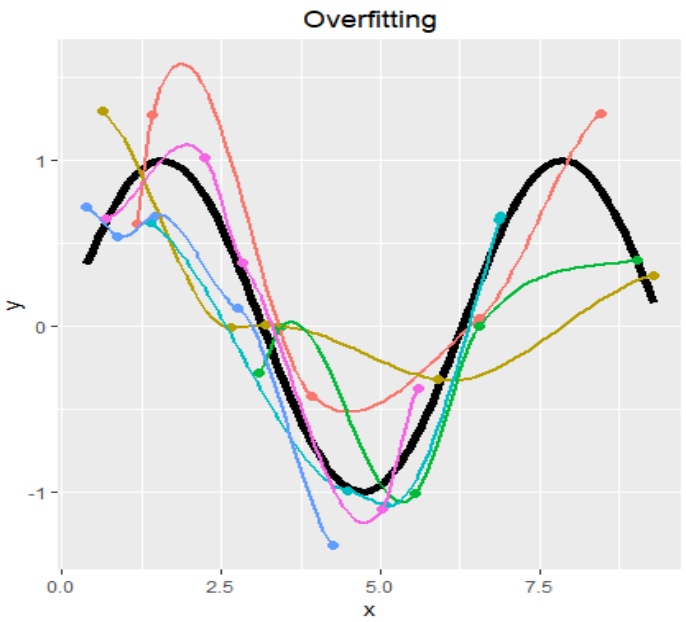
- **Nutzen von Trainings- und Testsets**
  - Durch Resampling wird ein Learner zum Modell trainiert
- **Performanzmaß für den Generalisierungsfehler**
  - Vergleich verschiedener Modelltypen

➢ Modell mit **akkuratester Vorhersage** wird genutzt



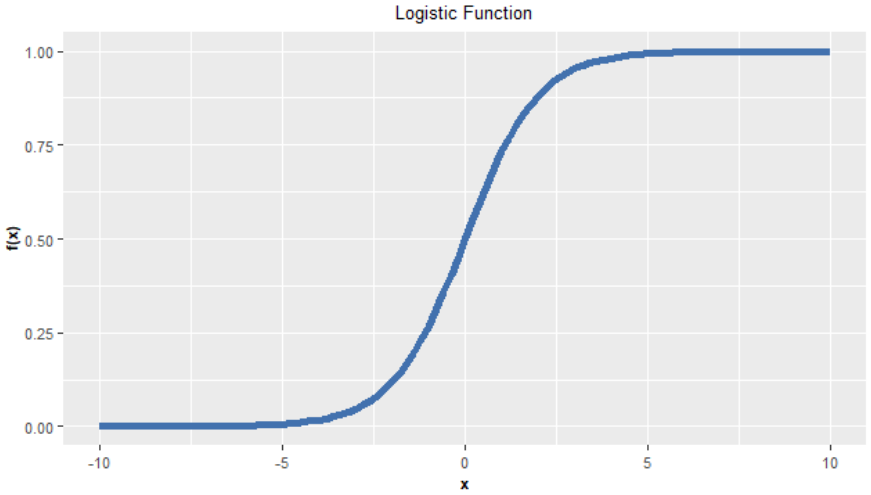
Over- und Underfit

- **Overfitting**
  - Zu starke Anpassung an die Stichprobendaten
- **Underfitting**
  - Zu geringe Anpassung an die Stichprobendaten

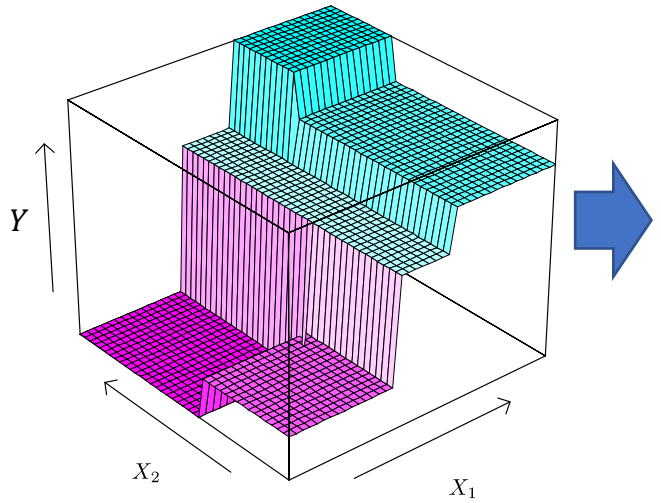


Vergleich Klassifizierungsmodelle

- Klassischerweise wird die **logistische Regression** zur (dichotomen) Kategorisierung genutzt
  - Interpretierbare Parameter, allerdings wenig flexibel beim Datenfit
- **Baumbasierte Machine Learning Modelle** sind flexibler
  - Allerdings Interpretierbarkeit oft schwierig und Extrapolation begrenzt



$$\frac{e^x}{1+e^x}$$

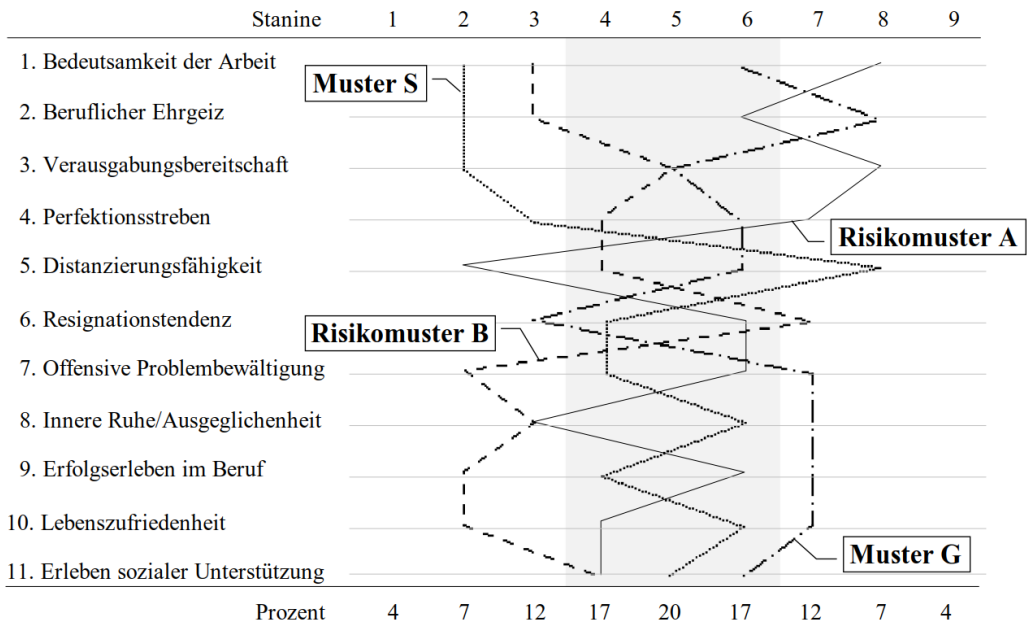


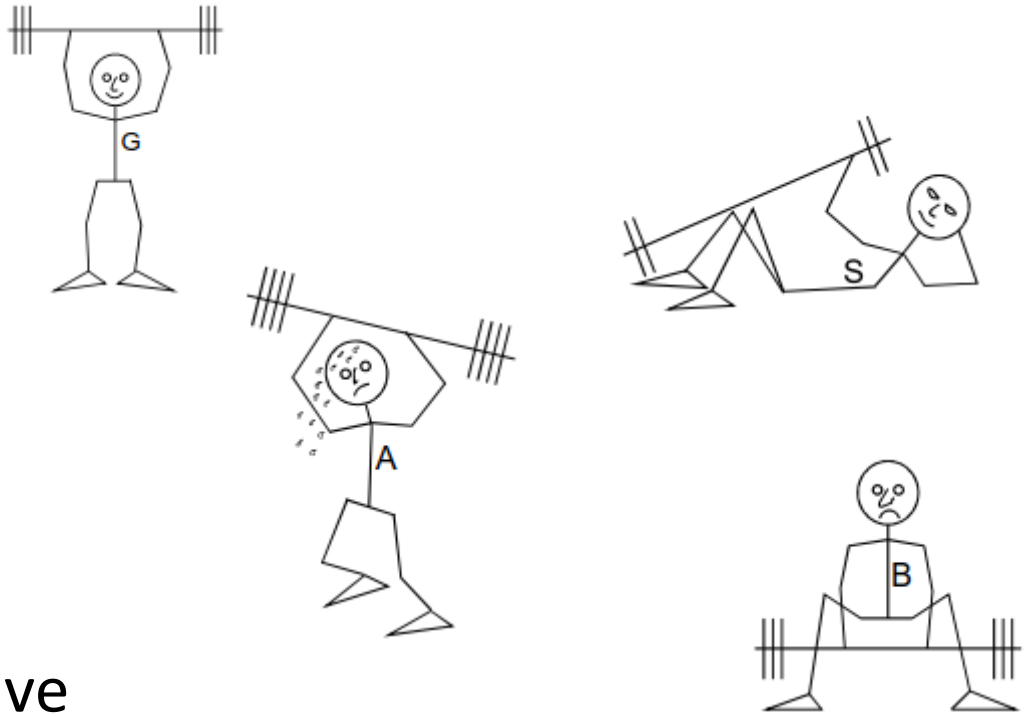
?

Grafik: Efron, B., & Hastie, T. (2016). *Computer age statistical inference*(Vol. 5). Cambridge University Press., S. 125

Exemplarische Studie Persönlichkeitstypen

- AVEM: arbeitsbezogene Verhaltens- und Erlebnismuster (Schaarschmidt & Fischer, 1996) Modell an *N* = 478 Lehrern erhoben





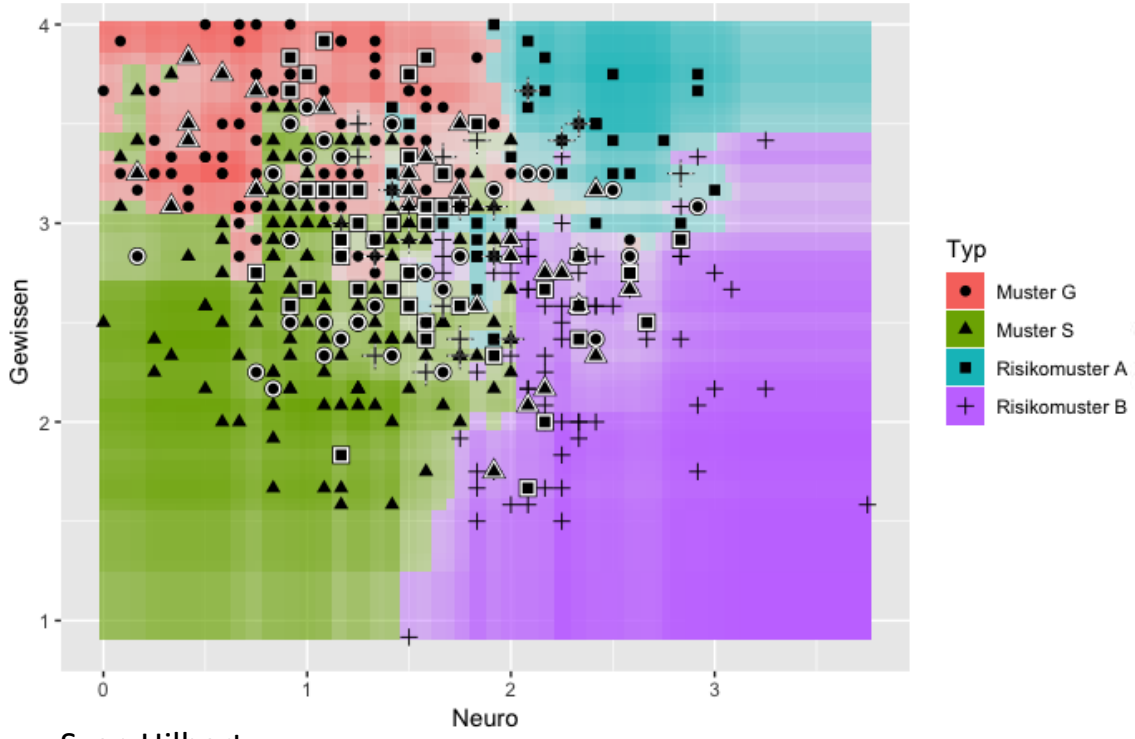
➢Versuch der Vorhersage durch die Big Five

Vorhersage und Erklärung bei Persönlichkeitstypen

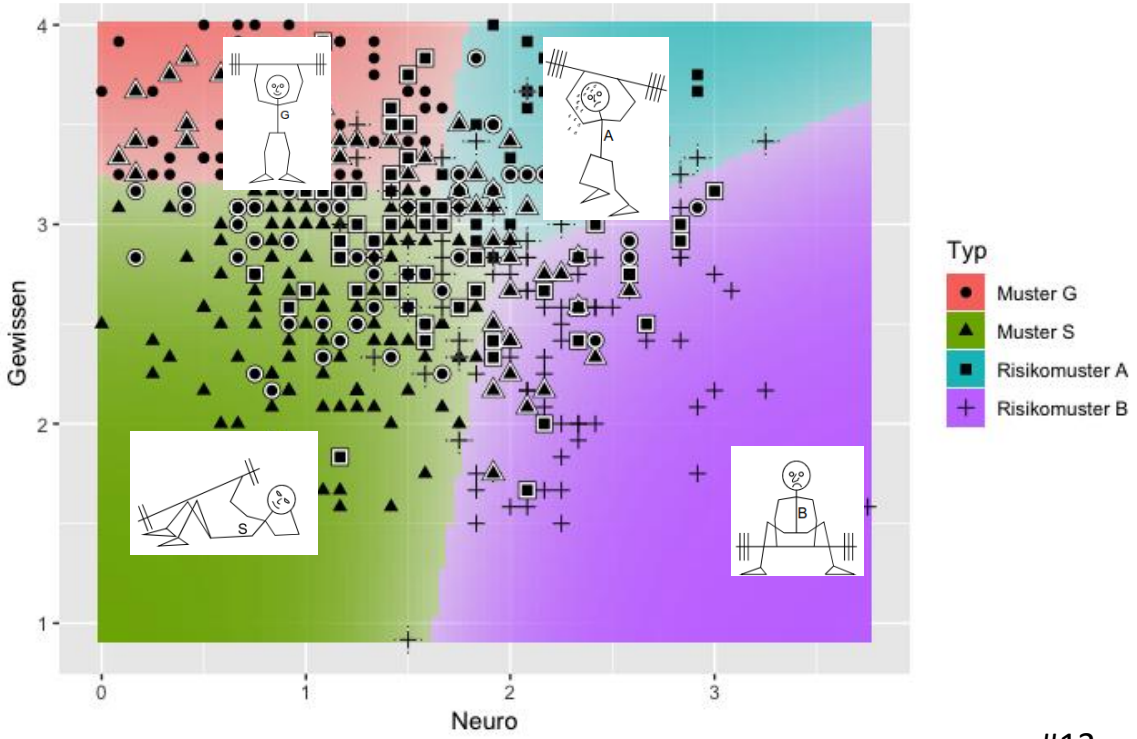
Universität Regensburg

Zwei Modellierungen: **Random Forest** und **Regressionsmodell**

Modell mit höchster Vorhersagekraft



Modell mit bester Interpretierbarkeit



Modellgültigkeit über Zeit

- Vorhersagegüte eines Modells gilt nur für den Zeitpunkt der Modellbildung
- Auch Modelle können Anachronismen werden
  - Gültigkeit von Modellen lässt in manchen Fällen nach
- Phänomen ist unter dem Namen **Concept Drift** bekannt

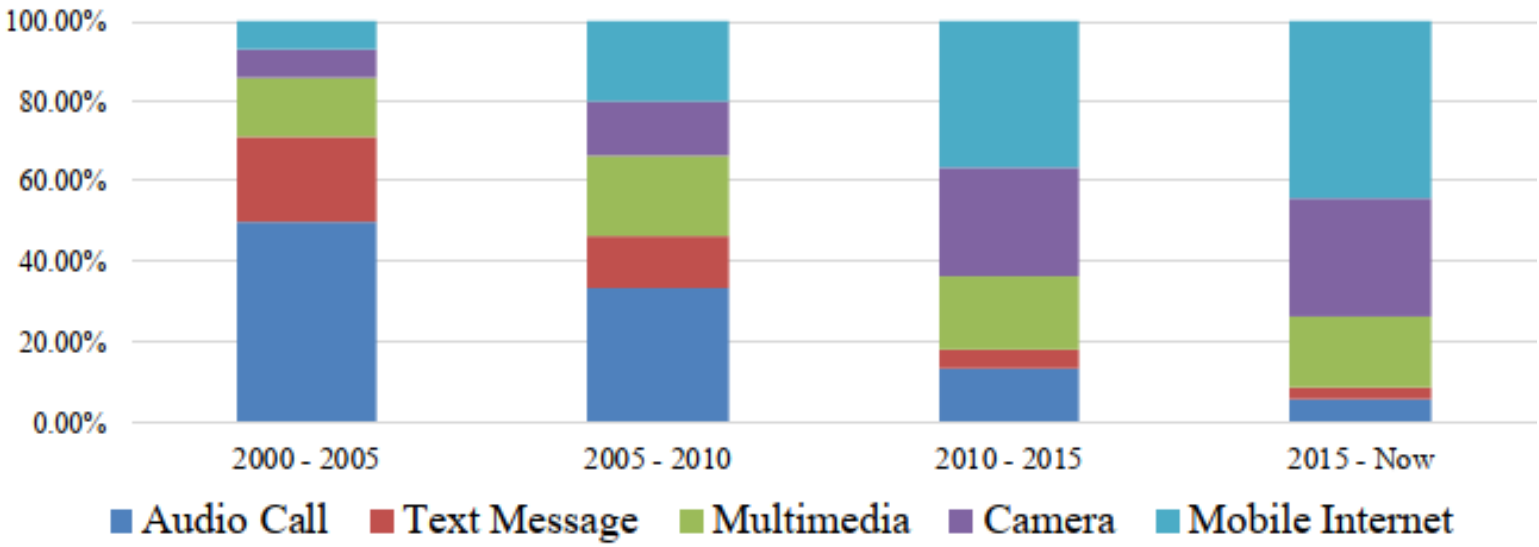
Gründe für Concept Drift

- Fragebogenitems können aus der Zeit fallen
  - **Anzahl von Fachbüchern** im Regal wird durch digitale Medien weniger aussagekräftig
- Es können plötzliche Veränderungen auftreten
  - Bildungsreformen
  - Lehrpläne können sich ändern
- Datenströme können sich ändern
  - Facebook Nutzung wird ein immer schlechterer Prädiktor für Prokrastination

Visualisierung Concept Drift

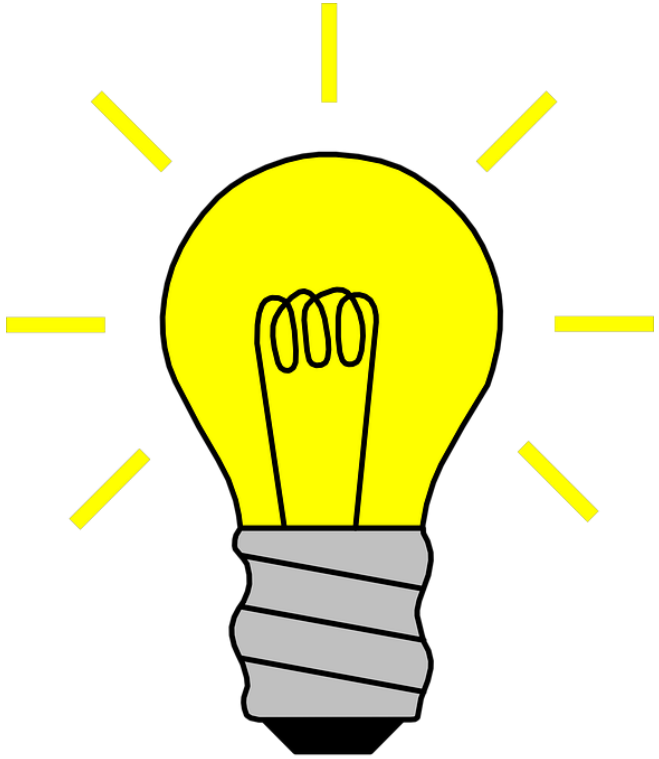
Studien zur Smartphone Nutzung

- Lu et al. (2018)



Methoden gegen den Concept Drift

- Zusammenfassen von Daten zu Kategorien
  - Nutzung von Kommunikationsmedien als Kategorie
- Technische Lösungsansätze
  - Beobachtung der Stabilität der Vorhersagegüte
    - ➢ Implementation in den Algorithmus
- Kontinuierliche Überprüfung mit Testsamples
  - Den bei Modellerstellung etablierten Generalisierungsfehler als Benchmark nutzen



Zusammenfassung: Erhöhung der Modellgültigkeit

Techniken zur **Erhöhung** der **Modellgültigkeit**

- Modellgültigkeit durch Vorhersagegüte **überprüfen**
  - **Overfitting** und Überinterpretation von *p*-Werten **vermeiden**
  - **Vorhersage** mit **Beschreibung** und **Erklärung** verbinden

- **Kontinuierliche Überprüfung** der Modelle
  - Generalisierungsfehler wiederholt schätzen

- Weiterer wichtiger Punkt: **Open Science**

https://www.britannica.com/biography/Wilhelm-Wundt

Open Science

- Open Science ist ein wichtiger Aspekt glaubwürdiger empirischer Forschung
  - ➢ Bereitstellung der Daten ist ein wichtiger Beitrag zu Überprüfbarkeit der Modelle
  - ➢ Möglichkeit zur Bildung neuer Modelle anhand von Daten
- Eine breite Datenbasis ist die wichtigste Grundlage für die Schätzung gültiger Modelle

www.osf.io

Vielen Dank

Universität Regensburg

Literatur

Efron, B., & Hastie, T. (2016). *Computer age statistical inference* (Vol. 5). Cambridge University Press.

Lu, J., Liu, A., Dong, F., Gu, F., Gama, J., & Zhang, G. (2018). Learning under concept drift: A review. IEEE Transactions on Knowledge and Data Engineering, 31(12), 2346-2363.

Schaarschmidt, U., & Fischer, A. (1996). *AVEM: arbeitsbezogene Verhaltens- und Erlebnismuster*. Swets Test Services.