



WESTFÄLISCHE  
WILHELMS-UNIVERSITÄT  
MÜNSTER

# Data Science @ eScience WWU

Arne Scheffer

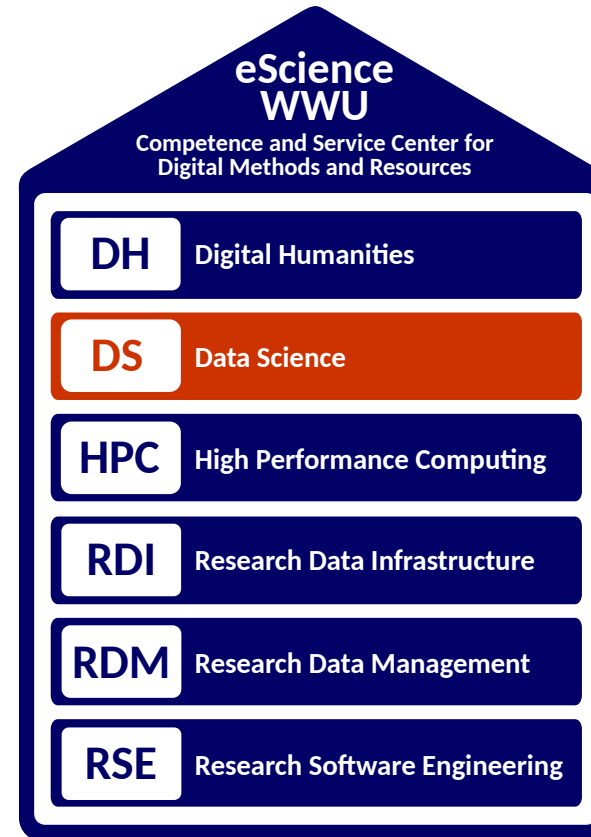


wissen.leben  
WWU Münster

# Data Science @ eScience WWU

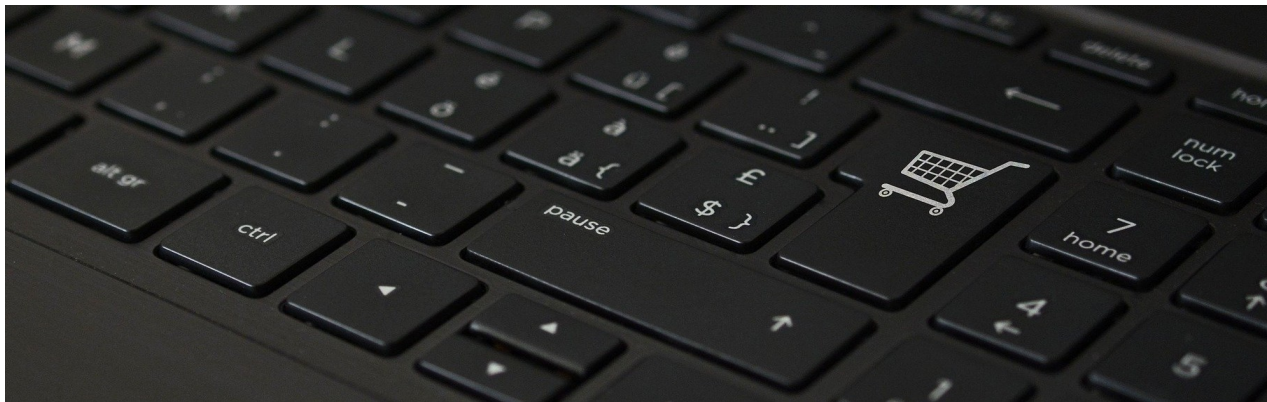
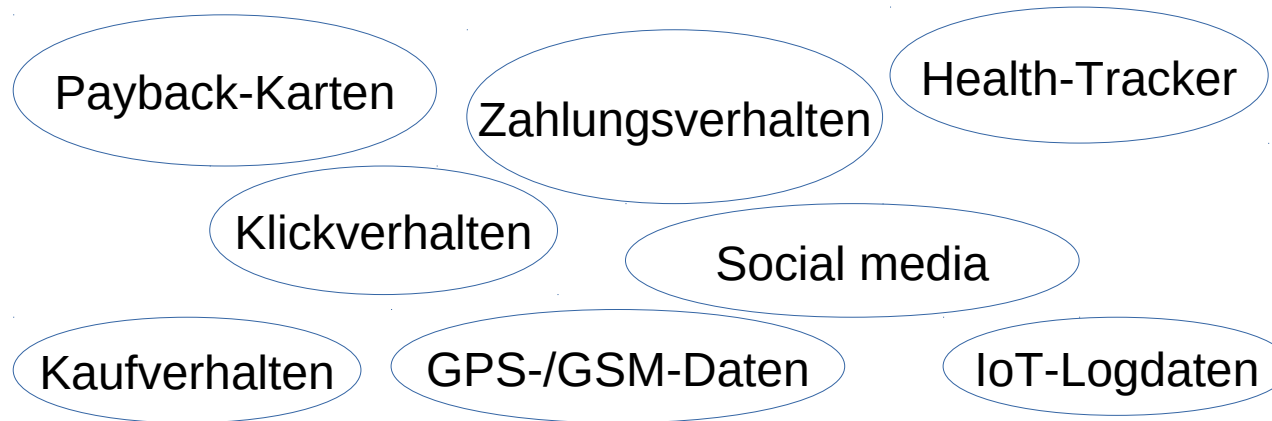
Ansprechpartner: Arne Scheffer

- Zentrum für Informationsverarbeitung
- Betreuung der Gruppen 3.2 und 3.3
- Schwerpunktthemen:  
Datenbanken & Anwendungsinfrastruktur
- Mathematiker (Informatik) und ...



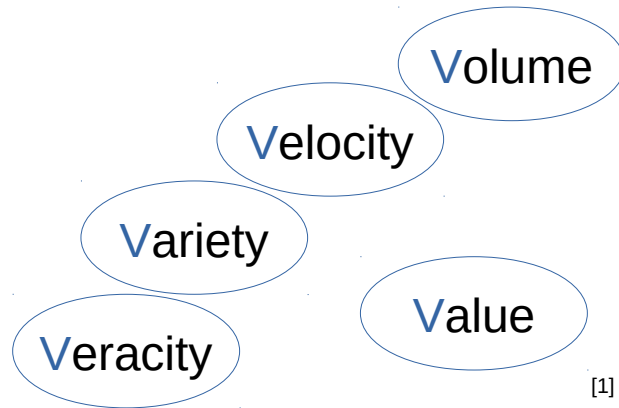
# Daten heute: Social Media, Mehrwertdienste, ...

- Persönliche (Profil-)Daten als Wertgegenstand im Geschäftsmodell





# Big Data: 5 Voraussetzungen



[1]



# Big Data: Umgang mit **Volume** und **Velocity**

MapReduce

Object storage API

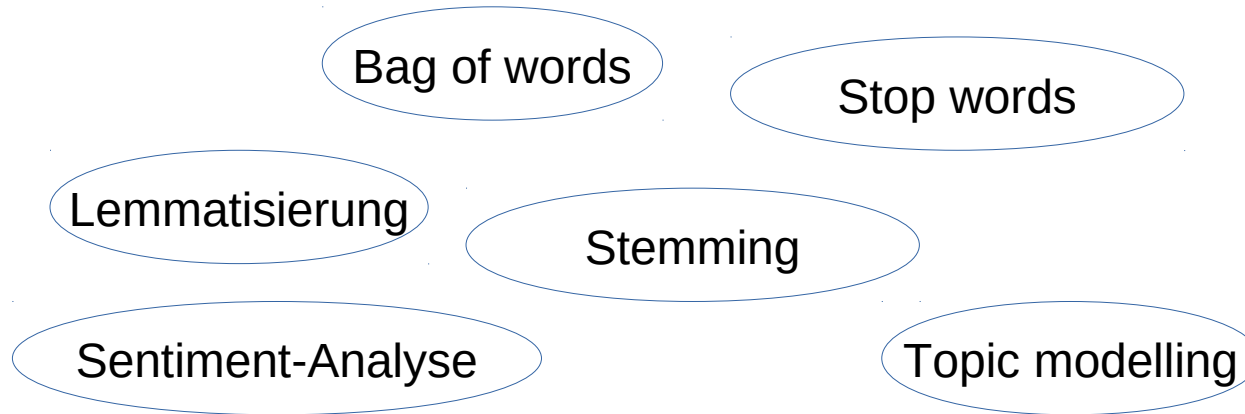
Apache Hadoop

Apache Spark

Massive parallel processing



# Big Data: Umgang mit **Variety** - Beispiel Text Mining







Value

# generieren: Angewendete Methoden

Kategorisierung

Regression

Clustering, Ähnlichkeiten

Ereigniszeitanalyse

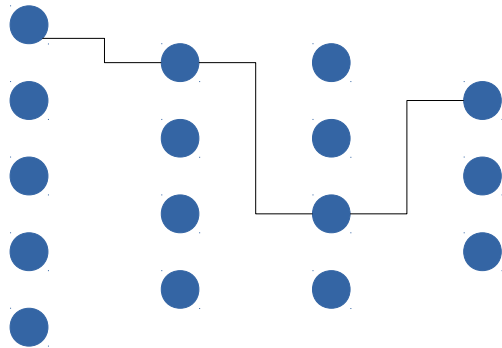


# Maschinelles Lernen Unüberwacht und Überwacht

k-means/-median

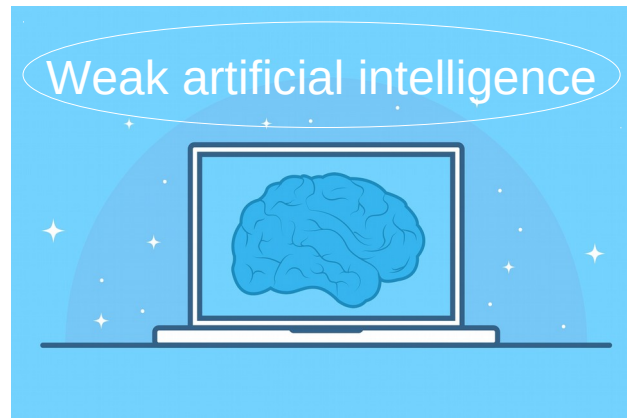
Random forest

Neuronales Netz



Entscheidungsbäume

Support-vector machine







5 V!

Welches V fehlte?



...Veracity

Vielen Dank fürs Zuhören!

# Quellenangaben

- [1] Textverweise zu den 5V:
  - ursprünglich 3V: vgl. „3D Data Management: Controlling Data Volume, Velocity, and Variety“
    - Doug Laney
    - 06.02.2001
  - Ergänzt um Veracity: vgl. „Harness the Power of Big Data The IBM Big Data Platform“, S. 14ff
    - Paul Zikopoulos, Dirk deRoos, Krishnan Parasuraman, Thomas Deutsch, James Giles, David Corrigan
    - 08.11.2012
  - Ergänzt um Value: vgl. „Transforming Big Data into Smart Data: Deriving value via harnessing Volume, Variety, and Velocity using semantic techniques and technologies“
    - Amit Sheth
    - 2014 IEEE 30th International Conference on Data Engineering
- Bilder jeweils Pixabay
  - matrix-1735640: Jonny Lindner
  - matrix-4646234: Gerd Altmann
  - analytics-3088958: xresch
  - cloud-3805852: Gerd Altmann
  - watch-4638673: George Dyulgerow
  - board-3700116: athree23
  - ecommerce-3546296: Pete Linforth
  - work-731198: Free-Photos
  - artificial-4082314: Mohammed Hassan
  - chart-2785902: Goumbik