

Prof. Dr. Pascal Kerschke,  
Jonathan Heins, M. Sc. & Felix Rauschert, M. Sc.

TU Dresden, Lehrstuhl für Big Data Analytics in Transportation

# Spezifische Aspekte von Data Analytics

Wintersemester 2022/2023

# Kursorganisation

---



Prof. Dr. Pascal Kerschke

pascal.kerschke@  
tu-dresden.de



Jonathan Heins, M. Sc.

jonathan.heins@  
tu-dresden.de



Felix Rauschert, M. Sc.

felix.rauschert@  
tu-dresden.de

- dieser Kurs findet regulär im 5. *Fachsemester* des B.Sc. Verkehrswirtschaft statt
- Sie sollten den Kurs *Data Analytics Grundlagen* bereits (erfolgreich) absolviert haben
- grundlegende Programmierkenntnisse in R (oder Python) werden vorausgesetzt

- dieser Kurs findet regulär im 5. *Fachsemester* des B.Sc. Verkehrswirtschaft statt
- Sie sollten den Kurs *Data Analytics Grundlagen* bereits (erfolgreich) absolviert haben
- grundlegende Programmierkenntnisse in R (oder Python) werden vorausgesetzt
- **Achtung:**  
Sollten Sie diese **Voraussetzungen nicht erfüllen** und dennoch an dem Kurs teilnehmen wollen, **melden Sie sich bitte umgehend** bei uns und erklären Sie kurz Ihre Situation!

# Allgemeiner Ablauf

- Kurs ist (sowohl inhaltlich als auch terminlich) mit der begleitenden Veranstaltung *Vertiefung in Data Analytics* eng verzahnt und abgestimmt
- jedes Seminar-Projekt besteht aus einem *methodischen* und einem *praktischen* Teil

# Allgemeiner Ablauf

- Kurs ist (sowohl inhaltlich als auch terminlich) mit der begleitenden Veranstaltung *Vertiefung in Data Analytics* eng verzahnt und abgestimmt
- jedes Seminar-Projekt besteht aus einem *methodischen* und einem *praktischen* Teil
- insgesamt bieten wir 3 Themen, die (ausgehend von der aktuellen Anzahl an Anmeldungen) entweder einzeln oder in 2er-Gruppen bearbeitet werden sollen
- methodischer Teil: Clusteralgorithmen  
(die die Inhalte der o.g. Vorlesung komplementieren)

# Allgemeiner Ablauf

- Kurs ist (sowohl inhaltlich als auch terminlich) mit der begleitenden Veranstaltung *Vertiefung in Data Analytics* eng verzahnt und abgestimmt
- jedes Seminar-Projekt besteht aus einem *methodischen* und einem *praktischen* Teil
- insgesamt bieten wir 3 Themen, die (ausgehend von der aktuellen Anzahl an Anmeldungen) entweder einzeln oder in 2er-Gruppen bearbeitet werden sollen
- methodischer Teil: Clusteralgorithmen (die die Inhalte der o.g. Vorlesung komplementieren)
- Benotung basierend auf:
  - Präsentation eines Clusteranalyse-Verfahrens,
  - Präsentation der Projektergebnisse (zum praktischen Teil), sowie
  - Seminarbericht

- von jetzt bis **Freitag, den 14. Oktober**
  1. Gruppenbildung (sofern gewünscht)
  2. individ. Prioritätenlisten für methodischen Teil festlegen (Prio 1 = höchstes Interesse)
  3. gemeinsame Prioritätenlisten für praktischen Teil festlegen (Prio 1 = höchstes Interesse)
  4. Prioritäten per E-Mail an die drei Kursorganisatoren schicken

# Nächste Schritte

- von jetzt bis **Freitag, den 14. Oktober**
  1. Gruppenbildung (sofern gewünscht)
  2. individ. Prioritätenlisten für methodischen Teil festlegen (Prio 1 = höchstes Interesse)
  3. gemeinsame Prioritätenlisten für praktischen Teil festlegen (Prio 1 = höchstes Interesse)
  4. Prioritäten per E-Mail an die drei Kursorganisatoren schicken
- im Anschluss informieren wir Sie über die Ihnen zugeordneten Projekte (spätestens bis zum Vormittag des 17. Oktober)
- Anmeldung des Kurses beim Prüfungsamt (nachdem die Anmeldefrist abgelaufen ist, können Sie sich auch nicht mehr abmelden; ferner zählen nicht abgegebene Projekte dann als nicht bestanden)

- Präsentation der Clusteralgorithmen (Individualleistung) im Stile einer Vorlesung am 15. und 17. November in den Slots des Kurses *Vertiefung Data Analytics*

- Präsentation der Clusteralgorithmen (Individuelleistung) im Stile einer Vorlesung am 15. und 17. November in den Slots des Kurses *Vertiefung Data Analytics*
- kurzes Status-Update zum Stand der Bearbeitung des anwendungsorientierten Projekts: Mittwoch, 14. Dezember (FAL 001, 11:10 - 12:40 Uhr)

- Präsentation der Clusteralgorithmen (Individuelleistung) im Stile einer Vorlesung am 15. und 17. November in den Slots des Kurses *Vertiefung Data Analytics*
- kurzes Status-Update zum Stand der Bearbeitung des anwendungsorientierten Projekts: Mittwoch, 14. Dezember (FAL 001, 11:10 - 12:40 Uhr)
- Abgabe der Folien für die Abschlusspräsentation: Sonntag, 29. Januar (23:59 Uhr)
- Abschlusspräsentation: Dienstag, 31. Januar (im Rahmen der Veranstaltung *Vertiefung Data Analytics*)

- Präsentation der Clusteralgorithmen (Individuelleistung) im Stile einer Vorlesung am 15. und 17. November in den Slots des Kurses *Vertiefung Data Analytics*
- kurzes Status-Update zum Stand der Bearbeitung des anwendungsorientierten Projekts: Mittwoch, 14. Dezember (FAL 001, 11:10 - 12:40 Uhr)
- Abgabe der Folien für die Abschlusspräsentation: Sonntag, 29. Januar (23:59 Uhr)
- Abschlusspräsentation: Dienstag, 31. Januar (im Rahmen der Veranstaltung *Vertiefung Data Analytics*)
- Abgabe der Seminarberichte: Sonntag, 12. Februar (23:59 Uhr)

# Methodische Themen

---

1. Clustern von Trajektorien
2. Clustern großer Datensätze

# Methodische Themen

---

Clustern von Trajektorien

# Algorithmen zum Clustern von Trajektorien

- Clustern von Trajektorien (hier Radfahrten) schwierig, da die Distanz zwischen Trips nicht einfach ermittelbar ist.
- Algorithmen zum Clustern von Trajektorien:
  1. TRACCLUS [4],
  2. klcluster [1],
  3. k-means for trajectories [5]
- Sollten Sie einen (vermeintlich) vielversprechenderen Algorithmus entdecken (oder kennen), sprechen Sie uns gerne an!



# Methodische Themen

---

Clustern großer Datensätze



- Viele Clusteralgorithmen operieren auf den paarweisen Distanzen zwischen Beobachtungen; die Berechnung der Distanzmatrix hängt quadratisch von der Anzahl an Beobachtungen ab und kann somit sehr ineffizient sein.
- Algorithmen zum Clustern großer Datensätze:
  1. CURE [2],
  2. BIRCH [6],
  3. DENCLUE [3]
- Sollten Sie einen (vermeintlich) vielversprechenderen Algorithmus entdecken (oder kennen), sprechen Sie uns gerne an!

# Praktische Themen

---

1. Untersuchung des Radverkehrs in Dresden  
(Kooperation mit der Professur Verkehrsökologie)
2. Analyse von Emissionsdaten  
(Kooperation mit der Professur Verkehrsökologie)
3. Überwachung radiologischer Daten  
(Kooperation mit AI & OR)

## Praktische Themen

---

Projekt 1: Untersuchung des  
Radverkehrs in Dresden

# Projekt 1: Untersuchung des Radverkehrs in Dresden

PROJEKT RADVERS – WELCHER RADFAHRTYP SIND SIE?  
VERKEHRSWISSENSCHAFTLER DER TU DRESDEN  
ERFORSCHEN DAS VERHALTEN VON RADFAHRERN



- RadVers-Projekt (2018): Untersuchung des Radverkehrs in Dresden
- über drei Wochen wurden die Fahrten von ca. 200 Personen aufgezeichnet
- im Zuge dieses Seminars soll das Fahrverhalten von Radfahrenden gezielt untersucht werden
- u.a. clustern der Fahrradfahrten mit Hilfe geeigneter Algorithmen
- Ziel: Erkenntnisse aus den Analysen ableiten und möglichst sinnvoll visualisieren

## Praktische Themen

---

### Projekt 2: Analyse von Emissionsdaten



- Professur für Verkehrsökologie verfügt über eine umfangreiche Datenbasis zur Schadstoffemission im Straßenverkehr
- Betrachtung der Daten über einen längeren Zeitraum und ggf. an verschiedenen Standorten
- im Zuge dieses Seminars sollen Muster in den Daten ermittelt und gezielt verkehrswissenschaftliche Fragestellungen untersucht werden

## Praktische Themen

---

Projekt 3: Überwachung radiologischer  
Daten

# Projekt 3: Überwachung radiologischer Daten

- Kooperation mit Praxispartner
- Einhaltung von Strahlenschutzwerten ist bei radiologischen Untersuchungen (CT, Röntgen,...) überwachungspflichtig
- Ziel der Seminararbeit: Exploration der Daten, Analyse mithilfe geeigneter Clusterverfahren, Visualisierung der Ergebnisse



## Methodische Themen (individuell):

- Trajektorien clustern:

1. TRACLUS [4]
2. klcluster [1]
3. k-means for trajectories [5]

- große Daten clustern:

1. CURE [2]
2. BIRCH [6]
3. DENCLUE [3]

## Praktische Themen (Gruppe):

1. Untersuchung des Radverkehrs in Dresden
2. Analyse von Emissionsdaten
3. Überwachung radiologischer Daten

Fragen, Anmerkungen und/oder Kommentare?

# References

---

- [1] Kevin Buchin, Anne Driemel, Natasja van de L'Isle, and André Nusser. kcluster: Center-based Clustering of Trajectories. In [Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems](#), pages 496 – 499. ACM, 2019. doi: 10.1145/3347146.3359111.
- [2] Sudipto Guha, Rajeev Rastogi, and Kyuseok Shim. CURE: An Efficient Clustering Algorithm for Large Databases. [ACM Sigmod Record](#), 27(2):73 – 84, 1998. doi: 10.1016/S0306-4379(01)00008-4.
- [3] Alexander Hinneburg and Daniel A. Keim. A General Approach to Clustering in Large Databases with Noise. [Knowledge and Information Systems](#), 5(4):387 – 415, 2003. doi: 10.1007/s10115-003-0086-9.
- [4] Jae-Gil Lee, Jiawei Han, and Kyu-Young Whang. Trajectory Clustering: A Partition-and-Group Framework. In [Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data](#), pages 593 – 604. ACM, 2007. doi: 10.1145/1247480.1247546.
- [5] Jana Seep and Jan Vahrenhold. K-means for Semantically Enriched Trajectories. In [Proceedings of the 1st ACM SIGSPATIAL International Workshop on Animal Movement Ecology and Human Mobility](#), pages 38 – 47. ACM, 2021. doi: 10.1145/3486637.3489495.
- [6] Tian Zhang, Raghu Ramakrishnan, and Miron Livny. BIRCH: An Efficient Data Clustering Method for Very Large Databases. [ACM Sigmod Record](#), 25(2):103 – 114, 1996. doi: 10.1145/235968.233324.