

Vertiefungsthemen Data Science

Sommersemester 2021

Dozenten: Prof. Dr. B. Mitschang, Dr. H. Schwarz
Ansprechpartner: Holger Schwarz, holger.schwarz@ipvs.uni-stuttgart.de
Sprache: Deutsch/German

Beschreibung

Große Datenmengen zu erfassen, zu verwalten, zu verarbeiten und zu analysieren sind zentrale Anforderungen an Informationssysteme, häufig zusammengefasst unter dem Begriff „Big Data“. Zu den wesentlichen Treibern in diesem Bereich gehört die große Menge web-basierter und mobiler Anwendungen, wie E-Mail, Blogs, soziale Medien und E-Commerce ebenso wie die stark wachsende Menge an Sensoren im Kontext von Industrie 4.0 und dem Internet der Dinge. Während die Sammlung und Speicherung dieser Daten mittlerweile durch Fortschritte in der Speichertechnologie weitgehend problemlos möglich ist, gestaltet sich die Gewinnung von Wissen aus diesen (Roh-)Daten als sehr komplexer Aufgabenbereich.



Data Scientists bewegen sich in diesem Aufgabenbereich und operieren an der Schnittstelle zwischen spezifischer Domänenexpertise, technischer und mathematischer Expertise. Zur Unterstützung dieser anspruchsvollen Tätigkeit stehen vielfältige Methoden und Werkzeuge zur Verfügung, die im Rahmen von Analysepipelines und Vorgehensmodellen zum Einsatz kommen. Diese unterstützen Data Scientists bei der Auswahl und Vorverarbeitung relevanter Daten, bei der Durchführung und Anpassung der einzelnen Analyseschritte sowie bei der Darstellung und Interpretation der Analyseergebnisse. In diesem Seminar werden unter anderem Themen in folgenden Bereichen angeboten:

- Methoden und Werkzeuge zur Datenexploration und Datenvorverarbeitung
- Deskriptive, prädiktive und präskriptive Analyseverfahren
- Data-Mining-Verfahren und maschinelles Lernen
- Infrastrukturen für effiziente Datenanalyse (z.B. NoSQL, Data Lake)
- ...

Anmerkungen

- **Die Seminarthemen werden in einer ersten Besprechung voraussichtlich in KW6 vorgestellt. Der Termin dieser obligatorischen Besprechung wird per E-Mail bekannt gegeben.**
- Im Rahmen des Seminars bearbeitet jede(r) Studierende eines der angebotenen Themen. Einstiegsliteratur hierzu wird von den Betreuer(inne)n bereitgestellt. Die Ergebnisse der Arbeit werden in einer ca. 20-seitigen Ausarbeitung zusammengefasst und im Rahmen eines ca. 30-minütigen Vortrags präsentiert.

Voraussetzungen

Kenntnisse aus "Einführung Data Science" und "Modellierung" werden vorausgesetzt.