

|  |  |
|--|--|
| <b>Georg-August-Universität Göttingen</b><br><b>Modul B.Inf.1131: Data Science: Grundlagen</b><br><i>English title: Data Science: Basics</i>   | 6 C<br>4 SWS   |
| <p><b>Lernziele/Kompetenzen:</b><br/>         Das Modul vermittelt grundlegende Kompetenzen im Umgang mit Daten und ihrer Analyse. Es gliedert sich in vier Teilbereiche</p> <p><b>Konzepte.</b> Nach erfolgreicher Teilnahme</p> <ul style="list-style-type: none"> <li>• kennen Studierende verschiedene Datentypen und können sie mit deskriptiven Statistiken beschreiben</li> <li>• kennen Studierende verschiedene Arten der Datenerhebung (experimentelles Design) und können deren Vorteile und Risiken benennen</li> <li>• kennen Studierende verschiedene Formen von Voreingenommenheit (Bias) in den Daten und die resultierenden Risiken, und können neue Kontexte hinsichtlich Bias bewerten</li> <li>• kennen Studierende Probleme der Fairness in Datenverarbeitung und Erhebung und können neue Kontexte hinsichtlich Fairness bewerten.</li> </ul> <p><b>Software Werkzeuge.</b> Erfolgreiche Teilnahme befähigt Studierende zum</p> <ul style="list-style-type: none"> <li>• benutzen einer Shell zur grundlegenden Datenvorverarbeitung</li> <li>• analysieren von Daten mit grundlegenden Softwarebibliotheken für Datenverarbeitung in Python (Pandas, Numpy, Scipy, Matplotlib, ...)</li> <li>• testen von Software und statischen Algorithmen auf Korrektheit</li> </ul> <p><b>Statistische Werkzeuge.</b> Erfolgreiche Teilnahme befähigt Studierende zum</p> <ul style="list-style-type: none"> <li>• unterscheiden zwischen statistischer Inferenz und deskriptiver Statistik</li> <li>• beherrschen der Grundlagen statistischer Inferenz (Fehler, p-Wert, Trennschärfe, Null-Hypothese, Konfidenzintervalle, ...) und vorhersagen welche Parameter diese beeinflussen</li> <li>• durchführen einfacher statistischer Tests mit Bootstrap- und Permutationstests</li> <li>• anwenden grundlegender Methoden des überwachten und unüberwachten Maschinellen Lernen (Klassifikation, Regression, Clustering).</li> </ul> <p><b>Stil.</b> Erfolgreiche Teilnahme befähigt Studierende zum</p> <ul style="list-style-type: none"> <li>• anwenden guter Praktiken von Visualisierung von Daten</li> <li>• verfassen aussagekräftiger Projektberichte</li> <li>• strukturieren von reproduzierbaren Daten- und Softwareprojekten</li> <li>• strukturieren von Software für Wiederverwendbarkeit</li> <li>• anwenden von Prinzipien guter Codestrukturierung und -praktiken</li> <li>• anwenden grundlegende Formen des Projekt- und Team-Managements</li> </ul> | <p><b>Arbeitsaufwand:</b><br/>         Präsenzzeit:<br/>         56 Stunden<br/>         Selbststudium:<br/>         124 Stunden</p> |
| <b>Lehrveranstaltung: Data Science: Grundlagen</b> (Vorlesung, Übung)  | 4 SWS  |
| <b>Prüfung: Klausurähnliche Hausarbeit (Bearbeitungszeit: 1 Woche)</b><br><b>Prüfungsanforderungen:</b>  | 6 C  |

- Fähigkeit grundlegende statistische Begrifflichkeiten und Konzepte anzuwenden (Statistiken, einfache Tests mit Permutationen oder Bootstrapping, Konfidenzintervalle, ...) und zu interpretieren
- Kenntnis verschiedener Datentypen, und die Fähigkeit sie mit deskriptiven Statistiken zu beschreiben und geeignet visuell darstellen
- Fertigkeit Daten mit geeigneten Softwarebibliotheken und Shell in Python zu verarbeiten
- Kenntnis verschiedener Arten der Datenerhebung und Fähigkeit zur Bewertung der Vorteile und Risiken
- Kenntnis verschiedener Formen von Voreingenommenheit (Bias) in den Daten und die resultierenden Risiken, und Fähigkeit zur Bewertung neuer Kontexte hinsichtlich Bias
- Fähigkeit zur Evaluation von Fairness in Datenverarbeitung und Erhebung in neuen Kontexten
- Kenntnis von Prinzipien guter Codestrukturierung und Fähigkeit diese auf Code anwenden
- Fähigkeit statistische Algorithmen zu testen und debuggen
- Fähigkeit grundlegende Methoden des überwachten und unüberwachten Maschinellen Lernen auf neue Probleme anzuwenden
- Kenntnis guter Praktiken von Berichtverfassung und Fähigkeit sie auf neue Projekte anwenden
- Fähigkeit Daten und Softwareprojekte reproduzierbar zu strukturieren

|  |   |
|--|---|
| <b>Zugangsvoraussetzungen:</b><br>keine  | <b>Empfohlene Vorkenntnisse:</b><br>Grundkenntnisse in Python |
| <b>Sprache:</b><br>Deutsch, Englisch   | <b>Modulverantwortliche[r]:</b><br>Prof. Dr. Fabian Sinz      |
| <b>Angebotshäufigkeit:</b><br>jedes Sommersemester   | <b>Dauer:</b><br>1 Semester                                   |
| <b>Wiederholbarkeit:</b><br>zweimalig  | <b>Empfohlenes Fachsemester:</b><br>2                         |
| <b>Maximale Studierendenzahl:</b><br>100   |   |
| <b>Bemerkungen:</b><br>Durch erfolgreiches Lösen und Erklären der Übungsaufgaben können Bonus-Prozent für die Klausur erworben werden. |   |