

Prof. Dr. Christia Beecks,
Dr. Simone Opel

Modul 64511

Einführung in Data Science

LESEPROBE

Fakultät für
**Mathematik und
Informatik**

Der Inhalt dieses Dokumentes darf ohne vorherige schriftliche Erlaubnis durch die FernUniversität in Hagen nicht (ganz oder teilweise) reproduziert, benutzt oder veröffentlicht werden. Das Copyright gilt für alle Formen der Speicherung und Reproduktion, in denen die vorliegenden Informationen eingeflossen sind, einschließlich und zwar ohne Begrenzung Magnetspeicher, Computerausdrucke und visuelle Anzeigen. Alle in diesem Dokument genannten Gebrauchsnamen, Handelsnamen und Warenbezeichnungen sind zumeist eingetragene Warenzeichen und urheberrechtlich geschützt. Warenzeichen, Patente oder Copyrights gelten gleich ohne ausdrückliche Nennung. In dieser Publikation enthaltene Informationen können ohne vorherige Ankündigung geändert werden.

Leseprobe Modul 64511 Einführung in Data Science

Im Modul „Einführung in Data Science“ möchten wir Ihnen einen Einstieg und einen fundierten Überblick über die vielfältigen Strukturen, Prinzipien und Methoden des faszinierenden Fachs **Data Science** geben. Wir spannen in diesem Kurs das große Feld der Data Science auf, so dass Sie das Fach in seiner ganzen Breite kennenlernen und Sie die verschiedenen Methoden, Verfahren und Ideen auch selbst ausprobieren können.

Eine klassische Leseprobe, wie Sie diese aus anderen Lehrveranstaltungen kennen, können wir Ihnen für das Modul „Einführung in Data Science“ nicht zur Verfügung stellen, da es sich um einen **reinen Online-Kurs** handelt. Das heißt, dass alle Studienmaterialien über die digitale Lernumgebung **Moodle** der FernUniversität in Hagen zur Verfügung gestellt werden. Im Folgenden haben wir Ihnen dennoch ein paar grundlegende Informationen zusammengestellt, damit Sie sich einen Überblick über den Inhalt und den Aufbau des Moduls verschaffen können.

Um Ihnen unterschiedliche Lernangebote zu machen, haben wir zahlreiche Materialien, Medien und Zugänge zu den verschiedenen Themen erarbeitet und zusammengestellt. Die wichtigsten Elemente wollen wir Ihnen hier kurz vorstellen.

Aufbau der Lektionen

Jede Lektion beginnt mit einem einführenden Video- oder Audiobeitrag, in dem erklärt wird, worum es geht, welche Kompetenzziele Sie erreichen sollen, welche weiteren Materialien vorhanden sind und an welchen Aktivitäten Sie teilnehmen werden.

Jede Lektion enthält mindestens einen Beitrag in Form von Videos oder Podcasts, sowie gegebenenfalls ergänzende Texte, Folien oder auch einmal einen notwendigen Readertext. An einigen Stellen finden Sie zusätzlich Hinweise auf optionale oder ergänzende Lektüre oder Verweise zu weiteren Quellen. Daran schließt sich immer eine wechselnde Anzahl von Übungsaufgaben und Aktivitäten an, die Ihnen helfen, das Gelernte zu festigen und zu vertiefen.

Jede Lektion enthält verschiedene Aufgaben zum Üben und zur Selbstkontrolle, zudem finden Sie an vielen Stellen auch Aktivitäten, die Sie mit einem oder mehreren Partnerinnen und Partnern erledigen sollen. Diese Aufgabenstellungen der Aktivitäten haben in der Regel Fertigstellungstermine, um zum Beispiel die Ergebnisse zusammenzustellen oder online zu präsentieren. Intensives gemeinsames Arbeiten an einer Fragestellung ist für das Verständnis der jeweiligen Thematik sehr gewinnbringend.

Kompetenzziele des Moduls

Am Ende des Kurses können Sie ...

- erläutern, was Data Science bedeutet, und exemplarische Anwendungsgebiete beschreiben.
- die wichtigsten Fragestellungen des Gebiets Data Science benennen.
- einen Data Science Life Cycle beschreiben und die damit verbundenen Schritten und Methoden anwenden.
- unterschiedliche Analysearten für Ihre Daten unterscheiden und anwenden.
- verschiedene Aspekte der Datensicherheit und Datenintegrität kritisch reflektieren und beurteilen.
- ethische Fragestellungen theoriebasiert auf den Data Science Life Cycle anwenden und diskutieren.
- die grundlegenden Werte- und Normensysteme ethischer und rechtlicher Grundlagen erläutern.
- einen vollständigen Data Science Life Cycle durchlaufen und damit eigene Fragestellungen entwickeln und beantworten.

Gliederung der Lektionen

1. Lektion: *Data Science - Was steckt hinter diesem Begriff? Eine Einführung in den Bereich und diesen Kurs*

In dieser Lektion führen wir Sie in die faszinierende Welt der Data Science ein. Neben organisatorischen und strukturellen Aspekten dieses Kurses werden Sie lernen, was Data Science bedeutet und welche Gemeinsamkeiten und Unterschiede diese vergleichsweise junge Wissenschaft mit anderen spannenden Bereichen, wie beispielsweise dem maschinellen Lernen oder der künstlichen Intelligenz, besitzt. Wir werden uns mit den grundlegenden Fragestellungen und Merkmalen der Data Science beschäftigen und den sogenannten Data Science Life Cycle auf einer konzeptionellen Ebene kennenlernen. Sie werden lernen, welches Wissen und Fähigkeiten ein Data Scientist prinzipiell beherrschen muss und dass wir Big Data nicht nur als Fluch, sondern als Segen betrachten sollen.

Die erste Lektion werden Sie mit einem „datenanalytischen“ Verständnis verlassen, das Ihnen hilft, zukünftige datenbasierte Herausforderungen aus unterschiedlichen Sichten zu bewerten und erste Lösungsstrategien zu entwickeln.

2. Lektion: *„With great power comes great responsibility“ - Welcome to the Peter Parker Principle*

In dieser Lektion lernen Sie, warum Datenethik wichtig ist – und warum „Ethics by Design“ ein wichtiges Konzept in der Data Science und allgemein beim Umgang mit Daten ist. Sie werden am Ende dieser Lektion theorie- und kriterienbasiert verschiedene Ansätze, Ideen und Konzepte aus dem Bereich Data Science diskutieren und einordnen können. Weiter

werden wir uns mit der Frage der Technikneutralität, den Kriterien für faire intelligente Systeme und der Kulturabhängigkeit der Betrachtung solcher Fragestellungen intensiv auseinandersetzen.

Da es hier an vielen Stellen um die Diskussion von Theorien und Lösungsansätzen geht, enthält diese Lektion relativ viele Elemente, in denen Sie mit anderen zusammenarbeiten sollen. Das bietet auch eine weitere gute Möglichkeit, Ihre Mitstudierenden besser kennenzulernen!

3. Lektion: *Es lebt! Der Data Science Life Cycle und Vorgehensmodelle*

Diese Lektion beschäftigt sich mit der strukturierten Herangehensweise an Data Science-Projekte. Dazu werden Sie die einzelnen Schritte des Data Science Life Cycles kennenlernen, um damit ein strukturiertes Vorgehen für konkrete Problemstellungen zu entwickeln. Wir werden uns mit den unterschiedlichen Vorgehensmodellen auseinandersetzen und damit wissenschaftliche, industrielle und projektorientierte Sichten auf den Data Science Life Cycle kennenlernen. Darüber hinaus diskutieren wir, wie sich Methoden des maschinellen Lernens in die Prozesse integrieren lassen und welche Anforderungen damit verbunden sind.

Nach Abschluss dieser Lektion sind Sie in der Lage, Data Science strukturiert anzuwenden und datenbasierte Lösungen zu entwickeln und zu bewerten. Sie können verstehen, dass Daten zentral, aber nicht der erste Einstiegspunkt in einem komplexen Data Science Projekt sein sollten.

4. Lektion: *Wie fangen wir jetzt an? Grundlegende Konzepte und Methoden*

Es wird methodisch. Nachdem Sie in den vorangegangenen Lektionen ein breites Verständnis von Data Science und den ethischen Herausforderungen im Umgang mit Daten entwickelt haben, schauen wir uns in dieser Lektion an, wie Sie Ihr strukturelles Vorgehen mit Leben füllen, um damit von der konzeptionellen Ebene auf die methodische Ebene überzugehen. Sie werden Ihr datenanalytisches Denken mit grundlegenden Konzepten der Data Science anreichern und erste Methoden kennenlernen. Anhand von einfachen Beispielen werden Sie lernen, wie Sie verschiedene Problemstellungen unterscheiden und die richtige Methodik auswählen.

Diese Lektion vermittelt Ihnen methodisches Wissen, um Ihre Data Science Erfahrung zu bereichern. Wir möchten Ihnen insbesondere zeigen, dass es oft nicht eine perfekte Lösung gibt, sondern dass viele existierende und zukünftige Methoden eingesetzt werden können. Die Anwendungen dieser Methoden mit geeigneten Tools und Anwendungen ist Gegenstand der nächsten Lektion.

5. Lektion: *Von der Theorie in die Praxis - Methoden, Tools und ihre Anwendungen*

In dieser Lektion werden wir einerseits weitere grundlegende Methoden vorstellen und andererseits wichtige Werkzeuge und Anwendungen kennenlernen. Sie lernen die zwei bedeutenden Programmiersprachen der Data Science kennen – Python und R – und werden darüber hinaus einen Einblick in gängige Analyse Plattformen erhalten. Auch werden wir uns mit Cloudbasierten Systemen beschäftigen, um Ihnen die vielfältigen Möglichkeiten aufzuzeigen.

Spätestens jetzt sollten Sie in der Lage sein, Ihre Lösung nicht nur konzeptionell und methodisch zu entwerfen, sondern auch eine zielorientierte Lösung zu entwickeln und zu implementieren. Doch damit Sie nicht über die rechtlichen Aspekte im Umgang mit den Daten stolpern, geben wir Ihnen in der nächsten Lektion einen Einblick in das Datenrecht.

6. Lektion: *Implementing Ethics into a Data Driven Culture*

Da Sie inzwischen nicht nur einen Data Science Life Cycle kennen, sondern auch die verschiedenen damit verbundenen Methoden anwenden können, ist es Zeit, diese Kompetenzen mit den Erkenntnissen aus Lektion 2 zu verbinden. Sie beschäftigen sich daher tiefer mit Data Privacy und Data Ethics und lernen die Grenzen der ethischen Ansätze und auch der vorhandenen Leitlinien kennen. Da es uns immer auch darum geht, nicht nur technisch zuverlässige und robuste, sondern insbesondere vertrauenswürdige Systeme zu konzipieren, analysieren wir unterschiedliche Quellen zum ethischen Umgang mit Daten und wenden diese auf unterschiedliche Fallbeispiele an. Am Ende dieser Lektion können Sie Ihr eigenes Forschungsdatenmanagement nach den erlernten Kriterien reflektieren und wissen, wie Sie Ihre Forschungstätigkeit planen können. Auch hier finden Sie die Gelegenheit, die erarbeiteten Inhalte in Übungen und Gruppenaufgaben umzusetzen.

7. Lektion: *Von der Kunst, die richtigen Informationen darzustellen - Visualisierung und Kommunikation*

Die Art und Weise, wie wir die aus den Daten gewonnene Erkenntnisse visuell kommunizieren ist von entscheidender Bedeutung für die Akzeptanz und auch den Erfolg eines Data Science Projekts. In dieser letzten Lektion beschäftigen wir uns daher mit den Themen Datenvisualisierung und Kommunikation. Dieser Teil stellt einen wichtigen Teil des Kurses dar, da die Visualisierung von Ergebnissen eine Kunst an sich ist und eigenen Regeln und Prinzipien folgt. Nach Abschluss dieser Lektion sollten Sie in der Lage sein, die grundlegenden Aspekte der Datenvisualisierung zu diskutieren und einfache, aber aussagekräftige Datenvisualisierungen zu erstellen.

Wir hoffen, Ihr Interesse an der faszinierenden Disziplin Data Science geweckt zu haben und freuen uns darauf, Sie im Modul Einführung in Data Science zu begrüßen!

Simone Opel & Christian Beecks