

SYT6 3xHIT 22/23 / Labor Übungen / GKL671 Einführung Data Science

GKL671 Einführung Data Science

✓ Done: Receive a grade ✓ Done: Receive a passing grade

Data Science - ein erster Einblick Einführung

Die folgende Einführung ist zum Selbststudium gedacht und soll selbstständig durch ein erstes Datenanalyseprojekt durchführen. Fertige ein Protokoll an, in dem du die wichtigsten Inhalte zusammenfasst.

Fangen wir als Erstes damit an, was eigentlich die Begriffe "Data Science", "Machine Learning" und "Künstliche Intelligenz" bedeuten.

Bei allen dreien geht es um das Lernen aus Daten und Strukturen. Künstliche Intelligenz ist dabei ein Überbegriff, der jegliche Form von Erlernen von Strukturen, Konzepten und Inhalten durch Computer beinhaltet, etwa durch künstliche neuronale Netzwerke. Ein Teilfeld davon ist "Machine Learning", das sich gezielt mit der algorithmischen Beschreibung von Strukturen beschäftigt, also etwa Mustererkennung, um Betrugsfälle aufzudecken oder automatisiert Schriftzeichen zu erkennen.

Die Erklärung durch ChatGPT selbst, kannst du dir



ansehen oder selbst austesten, indem du die Begriffe von OpenAI erklären lässt und die gegebenen Antworten verifizierst.

"Data Science" ist die am meisten den Menschen Involvierende Disziplin, da es hier nicht nur um das Lernen des Computers sondern auch des Menschen aus Daten geht. Daher sind Visualisierungen, Zusammenfassungen, Erklärungen und Schlussfolgerungen, also "Geschichtenerzählen auf Basis von Daten und Fakten" ein relevanter Aspekt. Sieh dir dazu das folgende Video an, um zu verstehen,



Praxisaufgabe

Arbeite im Folgenden die Aufgabenstellungen mithilfe der [Shiny App](#) zu den Galaxie Daten durch und dokumentiere deine Vorgehensweise in einem Protokoll. Diese ist im Rahmen eines Praktikums mit dem Wissen aus Data Science aus dem 4. Jahrgang von einem Schüler entwickelt worden und gibt euch auch einen Ausblick, welche Fertigkeiten in Datenanalyse und -interpretation (Data Science), maschinellem Lernen (Clustering) und App Entwicklung euch in einer Spezialisierung "Data-Science" ab dem 4. Jahrgang offen stehen.

Nun wirst du selbst erste Schritte eines Data Science Projekts beschreiben. Dabei ist das Kennenlernen der Daten durch unterschiedlichen graphische Darstellungen ein relevanter Schritt, um zu verstehen, welche Fragestellungen oder Schlussfolgerungen damit sinnvoll oder möglich zu beantworten sind. Selbst das beste neuronale Netzwerk kann keine Outputs zu Gesichtern von Menschen liefern, wenn es ausschließlich mit Bildern von Tieren trainiert wurde.

Die Aufgabenstellungen, die du dafür beantworten sollst, sind im Folgenden angegeben. Beschreibe jeweils deine Vorgehensweise, aus welcher graphischen Darstellung oder Zusammenfassung, du welche Informationen ermittelst:

- Ermittle die größte und kleinste Geschwindigkeit in der beobachteten Galaxie und ihre minimale und maximale radiale Position.
- Ermittle, bei welchem Winkel die extremsten Geschwindigkeiten angenommen werden.
- Beschreibe, welche Information sich in einer Punktwolke darstellen lässt, wenn die Nord-Süd-Richtung gegen die Ost-West-Richtung aufgetragen wird.
- Überlege, welche Information sichtbar wird, wenn der Winkel gegen die Location Number aufgetragen wird.

Nach der ersten Exploration der Daten ist der nächste Schritt die Anpassung eines Modells. Hier passen wir einen Clusteralgorithmus an, der Teilbereiche der Daten, die Cluster, automatisiert erkennt. Die Anzahl dieser Cluster kannst du dabei mit dem Schieberegler selbst einstellen. Diese Methodik stellt eine der Optionen zur Erkennung unbekannter Strukturen in Daten zu finden, ohne ihre Anzahl im Voraus zu kennen oder ein Labelling bereits vorliegen zu haben, durch das in vorgegebene Kategorien, (z.B. Tierarten, ...), eingeteilt werden kann.

- Vergleiche, ob die Anpassung von 2, 3, 4 oder 5 Clustern am geeignetsten ist und woran man das erkennen kann. Betrachte dabei auch die Darstellung der Cluster als Überlagerung der zugrundeliegenden Aufnahme des Teleskops.

Ziele

Das Ziel ist es grundlegende Methoden von Data Science und des maschinellen Lernens kennen zu lernen und interpretieren zu können.

Abgabe

Das Protokoll ist als PDF-Dokument abzugeben vorzulegen.

Bewertung

Gruppengröße: 1 Person

Anforderungen überwiegend erfüllt

- grundlegende Beschreibung und Verwendung der im Unterricht angeführten Begriffe: Data Science, Machine Learning, Künstliche Intelligenz; Aufgabengebiet eines Data Scientist; Datenexploration; Clustering
- Bearbeitung der Aufgabenstellungen der App

Anforderungen zur Gänze erfüllt

- Verbale Beschreibung und Erklärung aller angeführter Begriffe und deren Anwendung in konkreten Beispielen in vollständigen deutschen Sätzen

Submission status

Attempt number	This is attempt 1.
Submission status	Submitted for grading
	This assignment is not accepting submissions
Grading status	Graded
Last modified	Monday, 20 March 2023, 10:59 PM
File submissions	<div> GKL-671_kbohaczyk.pdf 20 March 2023, 10:59 PM </div>
Submission comments	<div> Comments (0) </div>

Feedback

Grade	GK vollständig
Graded on	Sunday, 26 March 2023, 5:03 PM
Graded by	<div>LE</div> List (Admin) Erhard

Feedback comments

+

Grundlegende Beschreibung und Verwendung der im Unterricht angeführten Begriffe

Data Science, Machine Learning, Künstliche Intelligenz; ...