DSAI2023/24

Blatt 1–Data Science Grundlagen, Ethik

Aufgabe 1–Was versteht man unter Data Science?

Data Science ist ein interdisziplinäres Wissenschaftsfeld, welches wissenschaftlich fundierte Methoden, Prozesse, Algorithmen und Systeme zur Extraktion von Erkenntnissen, Mustern und Schlüssen sowohl aus strukturierten als auch unstrukturierten Daten ermöglicht.  
Es setzt sich zusammen aus: der Fähigkeit zu hacken, dem Wissen über Mathematik und Statistik sowie substanziellem Expertenwissen.

Aufgabe 2 –Welche Gefahren verbergen sich hinter AI, insbesondere hinsichtlich Ethik? Finden Sie ein Beispiel aus dem Internet und referenzieren Sie es.

Arbeitslosigkeit  
Es gibt Bereiche, in denen KI-Methoden deutlich besser funktionieren als der Mensch, daher läuft jetzt schon vieles ohne menschliches Zutun ab.

Quelle: https://www.nextpit.de/kuenstliche-intelligenz-gefahren

Aufgabe 3 -Was ist die Datenschutzgrundverordnung (DSGVO)? Was darf diesbezüglich die Künstliche Intelligenz?

Es ist eine Reihe von Gesetzen, die vom 25. Mai 2018 an anwendbar sind. Sie befassen sich damit, wie Unternehmen die privaten Daten von Einzelpersonen erfassen und handhaben.

KI darf z.B. nur persönliche Daten nur anonymisiert auswerten und verarbeiten.

Aufgabe 4

Erklärung der Daten Histogramm:

* Mittelwert (loc): 35 Jahre (Durchschnittsalter der Gruppe).
* Standardabweichung (scale): 10 Jahre (Streuung der Altersdaten um den Durchschnitt).
* Anzahl der Datenpunkte (size): 500 Personen.

Histogramm-Interpretation:

* Die X-Achse repräsentiert die Altersgruppen.
* Die Y-Achse repräsentiert die Häufigkeit (Anzahl der Personen) in jeder Altersgruppe.
* Das Histogramm zeigt, wie sich das Alter in der Gruppe verteilt.

Erklärung der Daten Scatterplot:

* Stunden Sport pro Woche: Die Anzahl der Stunden, die eine Person pro Woche für Sportaktivitäten aufwendet.
* Fitness-Score: Ein fiktiver Wert, der den Fitness-Level repräsentiert, berechnet auf Basis der Sportstunden und einer normalverteilten Störung.

Scatterplot-Interpretation:

* Die X-Achse repräsentiert die Anzahl der Stunden Sport pro Woche.
* Die Y-Achse repräsentiert den Fitness-Score.
* Jeder Punkt im Scatterplot repräsentiert eine Person in der Stichprobe.