

Kurs-Name: Data Analytics & Big Data  
WiSe 2024/25 – Prof. Dr. Frank Krickel  
Aufgabenblatt zu 03 KLASSIFIKATION

*Bei der Erstellung dieses Aufgabenblatts wurde ChatGPT eingesetzt.*

Aufgabe 1: Lineare Regression (Wdh. aus Vorlesung)

- Lade die Datei *Einkommen.xlsx* in KNIME
- Erstelle ein Linear Regression-Modell, in dem du das Einkommen aus folgenden Parametern prognostizierst:
  - IQ
  - Alter
  - Höchster Abschluss
  - Größe
  - Gender
- Versuche Kombinationen aus diesen Prädiktoren
- Welche Kombination hat den höchsten r-squared bzw. adjusted r-squared Wert?

Aufgabe 2: Lineare Regression (Transfer)

- Lade die Datei *FuelConsumptionCo2.csv* in KNIME
- Erstelle ein Linear Regression-Modell, in dem du die CO2-Emissionen aus folgenden Parametern prognostizierst:
  - Fuel Consumption
- Versuche weitere Prädiktoren
- Welcher Prädiktor hat den höchsten r-squared bzw. adjusted r-squared Wert?
- 

Aufgabe 3: Lineare Regression (Vertiefung)

- Finde Daten zu den deutschen Bundesländern. Z.B. Fläche, Einwohnerzahl, BIP, uvm. Versuche herauszufinden, ob es „gute“ lineare Korrelationen zwischen diesen Merkmalen gibt.

Aufgabe 4: Lineare Regression: "Untersuchung der Beziehungen zwischen Klima, Wirtschaft und Lebensqualität in europäischen Ländern" (Vorsicht: Es ist nicht sicher, dass bei dieser Übung gute Zusammenhänge entdeckt werden können)

**Ziel der Übung:**

Finde heraus, ob es Korrelationen zwischen Klimaindikatoren (z.B. Durchschnittstemperatur), Wirtschaftsdaten (z.B. BIP pro Kopf) und Lebensqualitätsindikatoren (z.B. Lebenserwartung) in europäischen Ländern gibt. Untersuche, ob klimatische Faktoren die Wirtschaftskraft und Lebensqualität beeinflussen können.

**Datenbeschaffung:**

Sammle Daten aus verschiedenen Quellen (z.B. Eurostat, World Bank, OECD, OpenWeather). Folgende Daten könnten untersucht werden:

1. **Durchschnittstemperatur (°C)**
2. **Niederschlag (mm/Jahr)**
3. **BIP pro Kopf (Euro)**
4. **Arbeitslosenquote (%)**
5. **Lebenserwartung (Jahre)**
6. **Bildungsniveau (z.B. Prozent der Bevölkerung mit Hochschulabschluss)**
7. **CO2-Emissionen pro Kopf (Tonnen)**

**Aufgabenstellung:**

1. **Daten sammeln:** Finde aktuelle Daten für eine Auswahl europäischer Länder (z.B. 10 bis 15 Länder).
2. **Korrelationen untersuchen:** Berechne die Korrelationen zwischen verschiedenen Variablen. Beispiele für Fragen, die untersucht werden könnten:
  - Gibt es eine Korrelation zwischen der Durchschnittstemperatur und dem BIP pro Kopf?
  - Wie stark ist die Beziehung zwischen der Lebenserwartung und dem Bildungsniveau?
  - Gibt es eine Korrelation zwischen Niederschlag und Arbeitslosenquote?
3. **Visualisierung:** Erstelle Scatterplots, um die Beziehung zwischen zwei Variablen zu visualisieren (z.B. Temperatur vs. BIP, BIP vs. Lebenserwartung). Beschreibe, ob eine lineare Beziehung vorliegt.
4. **Regressionsanalyse:** Führe eine lineare Regression durch, um zu untersuchen, ob die Temperatur die Lebenserwartung vorhersagen kann oder ob das BIP pro Kopf die CO2-Emissionen beeinflusst.
5. **Interpretation:** Diskutiere die Ergebnisse. Welche Variablen scheinen miteinander korreliert zu sein? Was könnte die Ursache der Korrelationen sein, und gibt es möglicherweise eine Kausalität?

#### Aufgabe 5: Logistische Regression

- Lade die Datei *Einkommen2.csv* in KNIME
- Denke an die Normalisierung
- Erstelle ein Logistic Regression-Modell, in dem du die Einkommenshöhe (niedrig, mittel, hoch) prognostizierst:

#### Aufgabe 6: Logistische Regression

- Lade die Datei *Einkommen3.csv* in KNIME
- Denke an die Normalisierung
- Erstelle ein Logistic Regression-Modell, in dem du das Income prognostizierst:
- 

#### Aufgabe 7: kNN:

- Bearbeite die Aufgaben 5 und 6 mit einem knn-Modell