

Pratikum Blatt 5 (5 Punkte)

mit der Datei 'rawdata_luftqualitaet.csv'

Datenanalyse und Einführung in Maschinelles Lernen
WS 2025/26

Datenanalyse mit Python
Abgabe am 3./4.Dezember 2025

Dozentin: Grit Behrens
mailto: grit.behrens@hsbi.de



Studiengang Informatik Fachbereich Campus Minden

Aufgabe 1 (5 Punkte)

Laden Sie den Datensatz 'rawdata_luftqualitaet.csv'. Er wurde im Forschungsprojekt ENVIRON erstellt und enthält Messdaten aus Wohnzimmern in Bielefelder Wohnungen zur Luftfeuchtigkeit, zur Temperatur und zum CO₂ – Gehalt der Raumluft. Zusätzlich wurden die Temperatur des Heizkörpers und die Temperatur einer Außenwand von innen gemessen. Die Daten wurden nach der Luftqualität bewertet im Attribut state_air_quality wobei 0-gute Qualität, 1-neutral empfundene Qualität und 2- schlechte Luftqualität bedeuten. In dieser Aufgabe sollen die Messwerte als Eingangswerte für einen MLP-Klassifikator dienen, um die Luftqualitätsbewertungen zu trainieren. Nutzen Sie die in der Vorlesung eingeführten Bibliotheken `numpy`, `pandas`, `seaborn`, `matplotlib` und `sklearn`. Aus `sklearn` nutzen sie bitte die Funktionen `preprocessing`, `train_test_split` aus `sklearn.model_selection`, `MLPClassifier` aus `sklearn.neural_network` und `accuracy_score` aus `sklearn.metrics`.

- Visualisieren Sie die Daten in geeigneter Weise als Datentabelle (erste Zeilen), Liniendiagramm (ggf. Ausschnitt wählen), Heatmap und Scattermatrix. Geben Sie auch Minimalwerte, Maximalwerte, Standardabweichung, Mittelwert und die Anzahl der Messwerte pro Spalte an. Analysieren Sie die Darstellungen und erklären Sie Ihre Beobachtungen. (1P)
- Trennen Sie die Daten in zufälliger Auswahl in zwei disjunkte Datensätze, den Trainings- und den Testdatensatz im Verhältnis 80% zu 20%! (1P)
- Normalisieren Sie die Messdaten auf Wertebereiche zwischen 0 und 1 und visualisieren Sie sie als Kontrolle in geeigneter Weise. (1P)
- Trainieren Sie einen MLP-Klassifikator mit dem Trainingsdatensatz und geben Sie die Prognose für die Bewertungen der Luftqualität als Ergebnis aus. Sie können gern mit den Hyperparametern experimentieren, um bessere oder schnellere Ergebnisse im Training zu erzielen. (1P)
- Evaluieren Sie das erzielte Ergebnis auf dem Testdatensatz, indem Sie die prognostizierte Bewertung mit der Originalbewertung vergleichen und die Genauigkeit der Klassifikation (`accuracy_score`) berechnen. (1P)