

Luftqualität Analyse in Abhängung mit Wetter und Verkehr

In dem Project planen wir die Luftqualität unterschiedlicher Länder miteinander zu vergleichen und mittels Machine-learning die Qualität in den nächsten Monaten bis Jahren vorherzusagen.

Hierzu werden daten von offiziellen Seiten mittels API gesammelt. Daten die nicht offiziell verfügbar sind, werden mittels web-scraping zusammengeschrieben.

Members:

- Daniel Gebhart (if23b168)
- Stefan Werner (if23b227)

Planned Data Sources:

Luftqualitäts- und Gesundheitsdaten:

- <https://www.stateofglobalair.org/data/>
- <https://www.healthdata.org/research-analysis/library/state-global-air-2025>
- <https://knowledge.unicef.org/CEED/resource/state-global-air-report-2024>
- Kaggle Datensätze:
 - <https://www.kaggle.com/datasets/waqi786/global-air-quality-dataset>
 - <https://www.kaggle.com/datasets/fedesoriano/air-quality-data-set>
 - <https://www.kaggle.com/search>

Wetterdaten:

- OpenWeatherMap API
- NOAA Climate Data
- Meteostat API

Verkehrsdaten:

- Google Maps Traffic API
- TomTom Traffic API
- Open Traffic Data

Planned Data Storage:

Es wird eine Kombination aus Docker und MongoDB geben, um die Daten zu speichern. Ob die daten tabellarisch gespeichert werden oder als JSON ist jetzt noch nicht geklärt.

Planned Procedure:

Es wird ein normales Maschine-learning Procederes verfahren benutzt werden, um die daten zu verarbeiten mit einem 80/20 Split von Training und Testdaten

Datenbereinigung und -standardisierung:

- Pre-Processing
- verschiedene Messsysteme/Einheiten harmonisieren

Feature Engineering:

- Kombination der Luftqualitätsmessungen mit Wetterdaten und Verkehrskennzahlen
- Erstellen zusätzlicher Features wie gleitende Durchschnitte, Trends, saisonale Muster
- Train/Test-Split (80/20)

Modellauswahl und Modeltraining:

- noch nicht entschieden, ob lineare Modelle oder komplexere Varianten genutzt werden (z. B. Random Forest, XGBoost, LSTM-Netzwerke für Zeitreihen)

Evaluierung:

- Metriken wie RMSE, MAE oder R^2

Expected Output:

Am Ende des Projekts soll ein Vorhersagemodell entstehen, das basierend auf historischen Luftqualitätswerten, Wetterdaten und ggf. Verkehrsdaten die Luftqualität für die kommenden Monate oder Jahre möglichst zuverlässig schätzen kann.