

Aufgabenstellung Referat

Durchzuführende Arbeiten je Zweiergruppe

- a) Erstellung des Programmcodes.
- b) Erstellung einer kurzen Präsentation (PPT oder Word), die das erstellte Programm, die durchgeführte Analyse und die Ergebnisse kurz beschreibt. Zeigen Sie auch kurz mögliche Schwachpunkte des Modells oder des Programms auf und diskutieren Sie Alternativen oder Erweiterungsmöglichkeiten. (etwa 10 Seiten)
- c) Vorstellung (25-Minuten) des Programmcodes und der Präsentation via Teams. (mit Zwischenfragen)

Hinweise zur praktischen Durchführung:

- a) Die einzelnen Aufgaben (Programm, Präsentation, Vorstellung) sollen jeweils zu etwa gleichen Teilen von den beiden Personen in der Gruppe bearbeitet werden und die jeweiligen Beiträge auch namentlich gekennzeichnet werden.
- b) Das Programm und die Präsentation sollen in ILIAS hochgeladen werden. Der Abgabetermin ist der 13. Juli.
- c) Die Präsentation/Vorstellung kann in Deutsch oder Englisch erfolgen.

Aufgabenstellung (Regressionsanalyse - CO2 Emissionen - Temperaturabweichung)

Es soll der Zusammenhang zwischen der Temperaturabweichung und den CO2 Emissionen mit Hilfe einer Regressionsanalyse untersucht werden. Als Datenquelle sollen hier die beiden beigefügten Dateien `fossil-fuel-co2-emissions-by-nation.csv` und `TemperaturZeitreihe.json` verwendet werden. In der ersten Datei sind in der Spalte `Total` die CO2 Emissionen enthalten und in der zweiten Datei unter dem Attribut `Mean` die Temperatur. Laden Sie hierzu die beiden Dateien jeweils in Python und bereiten Sie die Daten dann in Python auf. In der ersten Datei benutzen wir nur CO2 Emissionen für Deutschland in der Spalte `Country`, und aus der zweiten Datei benutzen wir nur Temperaturabweichung vom Typ `GISTEMP` unter dem Attribut `Source`. Wandeln Sie bitte die monatlichen Temperaturabweichungen in jährliche Durchschnittswerte um (arithmetische Mittel). Fügen Sie dann in beiden Datensätzen eine Spalte für das jeweilige Jahr (`Year`) ein, und fügen dann beide Datensätze entlang dieser Spalte zusammen (`merge`). Führen Sie dann eine lineare Regressionsanalyse durch, wobei die Temperaturabweichung die abhängige Variable sein soll, und die CO2 Emissionen die unabhängige Variable sein soll. Verwenden Sie hierbei ein paar Varianten, z.B. mit/ohne Regressionskonstante (β_0), unterschiedliche Teil-Zeiträume,

mit/ohne Validierungssample. Schreiben Sie die Ergebnisse der Regressions-Analyse (Parameter (β_0 und β_1) und Bestimmtheitsmaß R^2) in eine CSV-Datei.

Zusätzliche Anforderungen:

- a) Das Programm soll aus mindestens drei Modulen bestehen, z.B. Dashboard, Config und LoadCleanData.
- b) Das Programm soll bei jedem Lauf ein neues Run-Directory erstellen, und alle Ausgabedateien sollen in dieses Run-Directory geschrieben werden.
- c) Konfigurationsparameter sollen von einer JSON-Datei eingelesen werden und dann im Programmcode benutzt werden, z.B. die Namen der relevanten Spalten der Eingabedateien, oder Standardwerte.
- d) Die beiden bereitgestellten Dateien sollen direkt und unbearbeitet in Python geladen werden. D.h. die Aufbereitung der Daten, z.B. Entfernung von NAs, das Selektieren oder Umbenennen von Spalten, soll nicht manuell in Excel erfolgen, sondern komplett in Python.
- e) Nach dem Zusammenfügen der Tabellen, d.h. kurz bevor die Daten in die Regressionsanalyse gehen, soll die Tabelle zu Validierungszwecken in eine externe Datei geschrieben werden. Stellen Sie zusätzlich noch die CO2 Emissionen und die Temperaturabweichung in einem Diagramm dar (z.B. matplotlib) und speichern auch das Diagramm zu Validierungszwecken in einer Bilddatei.
- f) Implementieren Sie Datenqualitäts-Kontrollen, um eventuelle Ausreißer in den Zeitreihen zu entdecken und zu korrigieren.

Hinweise zum Programm:

- a) Die Daten können Flats (z.B. gleicher Wert wiederholt sich mehrmals) und Spikes (extrem hohe Werte) enthalten, die beide unrealistisch sind.
- b) Für den Zeitraum 1945-1990 liegen die CO2 Emissionen separat vor (FORMER GERMAN DEMOCRATIC REPUBLIC und FEDERAL REPUBLIC OF GERMANY), hierzu kann man die CO2-Emissionen einfach addieren.

Datenquellen Die Daten wurden aus folgenden Quellen geladen, aber manuell noch etwas angepasst: <https://datahub.io/core/global-temp/r/monthly.json>
<https://datahub.io/core/co2-fossil-by-nation#data>