**Počítání indexu kvality ovzduší**

**Cíl programu:** Vytvořit model sloužící k předpovědi indexu kvality ovzduší (IKO) z historických dat pomocí multilineární regrese. Následně tento model použít k předpovědi IKO na základě předpovědi počasí.

**Použité knihovny:** Pandas, numpy, datetime, math, beautiful soup, scipy, openpyxl, requests, unicodedata

**Soubory:**

CHMU\_pocasi\_all.xlsx – soubor stažený ze webu ČHMÚ ([www.chmi.cz](http://www.chmi.cz)), obsahuje historické záznamy o počasí na měřící stanici Praha-Libuš

Skodliviny\_2018.xlsx - osobně vytvořený soubor se záznamy o škodlivých látkách (SO2, NO2, PM10 a O3) naměřených na stanici Praha-Libuš, data získána z webu ČHMÚ ([www.chmi.cz](http://www.chmi.cz))

Index\_kvality\_ovzdusi.pdf - dokument vydaný ČHMÚ ([www.chmi.cz](http://www.chmi.cz)), obsahuje informace o metodách počítání IKO. POZOR: Je zde uvedený vzorec výpočtu IKO, nicméně pokud porovnáme IKO získané pomocí tohoto vzorce a IKO na webu ČHMÚ vychází rozdílné hodnoty. Proto je v programu použit jiný vzorec, o kterém se domnívám, že s ním ČHMÚ ve skutečnosti počítá IKO.

Index\_kvality\_ovzdusi\_vysvetleni.pdf - dokument vydaný ČHMÚ, obsahuje vysvětlení hodnot IKO.

data\_2018.xlsx - soubor, který vytvoří Zpracovani\_dat.py, obsahuje data z CHMU\_počasí\_all.xlsx a vypočítané IKO ve vhodnějším formátu pro použití v programu.

**Části programu:**

Zpracovani\_dat.py - Počítá IKO (pomocí vzorce jiného, než je uveden v Index\_kvality\_ovzdusi.pdf) a vytváří soubor data\_2018.xlsx.

Tvorba\_modelu.py - Vytváří a upravuje modely pro počítání IKO. Funkce zde definované využívají vzorce vysvětlené na této stránce: <https://reliawiki.org/index.php/Multiple_Linear_Regression_Analysis#Example_2>

Atributy modelu:

* Coef – koeficienty modelu u jednotlivých veličin (teplota, tlak atd.)
* Data – vhodná data pro kalibraci modelu, data\_2018 ořezaná o sloupce, které nejsou v modelu používány
* Results – vektor s IKO odpovídajícím datům
* Head - názvy veličin, které jsou použity pro počítání IKO
* P\_values – p hodnoty odpovídající koeficientům
* R2adj – R2adj (adjustovaný index determinace) modelu

Výsledkem jsou 3 modely:

Model\_1 počítající se všemi veličinami

Model\_2 s odstraněnými veličinami, které mají vysokou p hodnotu, ale s ponechanou konstantou

Model\_3 - stejný jako model\_2, ale bez konstanty, špatně počítá R2adj

Scraping.py - Získává data z předpovědi počasí na <https://meteobox.cz/> na základě zadaných parametrů (datum, čas, místo).

Main – main