

POLITECHNIKA WROCŁAWSKA  
WYDZIAŁ INFORMATYKI I TELEKOMUNIKACJI



---

# Prognoza jakości powietrza na podstawie danych z projektu Grow Green

---

Sprawozdanie z laboratorium

AUTOR

**Krzysztof Maciejewski**

nr albumu: **260449**

kierunek: **Informatyka**

*10 czerwiec 2022*

### Streszczenie

Praca przedstawia system przewidywania jakości powietrza na podstawie daty. Dataset został pobrany ze strony z zasobów Wrocław Otwarte Dane. Pobrane dane zostały następnie oczyszczone oraz użyto metody regresji liniowej do predykcji jakości powietrza we Wrocławiu na podstawie daty. Jakość otrzymanych wyników została sprawdzona za pomocą błędu średniokwadratowego.

## 1 Wstęp – sformułowanie problemu

Autor potrzebuje przewidzieć zmianę jakości powietrza, która wydarzy się w najbliższej przyszłości. Pozwoli mu to na podjęcie decyzji o tym czy bezpiecznie jest wychodzić danego dnia i o danej godzinie z domu oraz czy na przykład należy skorzystać z maski ochronnej.

## 2 Opis danych

Wielkość datasetu to 8760 wierszy. Kolumny to: Day, Month, Year, Time, Air, Yest, temp, rhum. Wartości "Air" określają stężenie pyłu PM 2.5 [ $\mu\text{g}/\text{m}^3$ ] (typ float). Kolumna "temp" oznacza temperaturę w stopniach Celsjusza, a kolumna "rhum" oznacza wilgotność powietrza w procentach, kolumna "Yest" oznacza odczyt jakości z poprzedniego dnia. Pozostałe kolumny (typ int) służą do określenia daty i czasu w którym zostały wykonane pomiary. Dane były pobierane co godzinę od 1.01.2017 do 31.12.2017.

## 3 Opis rozwiązania

Dane zostały pobrane ze strony [www.wroclaw.pl/open-data/](http://www.wroclaw.pl/open-data/) w formacie pliku csv. Baza została zapisana w postaci ramki danych biblioteki Pandas. Zawiera ona informacje o 8760 odczytach jakości powietrza. Dane opisujące temperaturę i wilgotność powietrza zostały pobrane za pomocą API z meteostat.

Używając metody *regresji liniowej* na danych uzyskano model pozwalający na określenie jakości powietrza na podstawie dnia, miesiąca, godziny, temperatury, wilgotności i jakości powietrza dnia wcześniejszego.

## 4 Rezultaty obliczeń

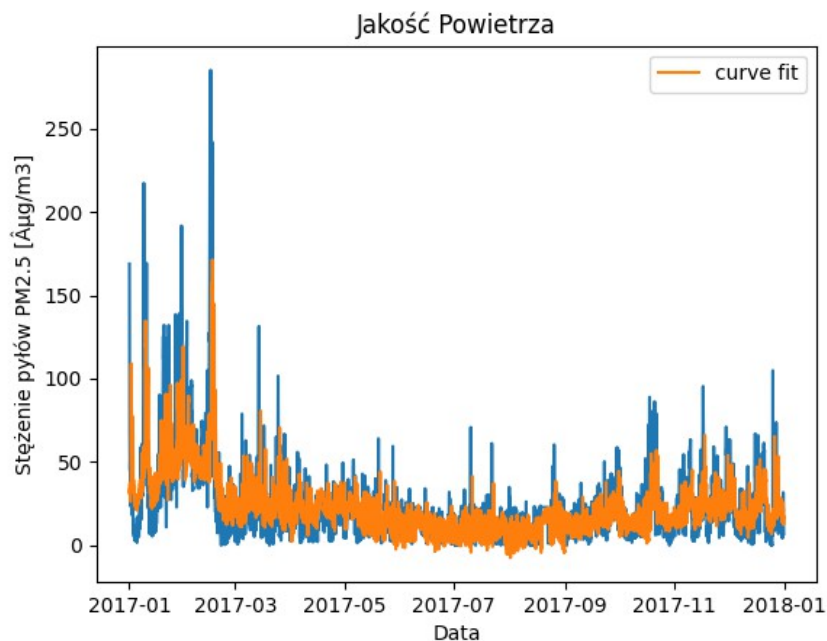
### 4.1 Plan badań

Na zbiorze danych zostanie użyty model regresji liniowej, a wyniki jakości powietrza obliczone za jej pomocą, zostaną użyte do obliczenia błędu średniokwadratowego.

### 4.2 Wyniki obliczeń

Model oceny jakości powietrza można przedstawić następującym wzorem:

$$\text{Score} = a * \text{month} + b * \text{day} + c * \text{time} + d * \text{temp} + e * \text{rhum} + f * \text{Yest} + g \quad (1)$$



Rysunek 1: Jakość powietrza z naniesioną krzywą regresji

Na rys. 1 pokazany jest wykres obrazujący dopasowanie krzywej. Można z niego wywnioskować, że czas pomiaru nie jest istotny, a najważniejszym elementem w ocenie jakości powietrza jest miesiąc i odczyt jakości z dnia wcześniejszego.

Podsumowując, symulacje ujawniają, że wg. (1) ocena jakości powietrza powinna być bardzo prosta, jednakże wyniki otrzymane ze wzoru cechują się mniejszą zmianą w trakcie roku, niż rzeczywiste dane pomiarowe. Do oceny modelu został użyty błąd średniokwadratowy, którego wartość wyniosła 305,45.

## 5 Wnioski

Moim zdaniem regresja liniowa nie była najlepszym możliwym modelem przyjętym do analizy jakości powietrza w okresie jednego roku. Na wykresie widać, że dane nie układają się liniowo. Być może zmiana okresu czasu w jakim analizowałem pomiary pomogłaby w lepszym dopasowaniu modelu do danych i otrzymaniu bardziej wartościowych predykcji.

## A Dodatek

Kody źródłowe zostały umieszczone w repozytorium github:

<https://github.com/kdmaciejewski/Smog-Prediction>.