

# Prognoza ilości PM10 w Krakowie - dokumentacja

Weronika Niedźwiedz

## 1. Opis programu

Program zwraca prognozę ilości PM10 w powietrzu, podaną w  $\mu\text{g}/\text{m}^3$ , dla jednej w 8 wybranych stacji na najbliższe 70 godzin. Prognoza jest drukowana w formie dokładnych wartości z odpowiadającą im godziną oraz wykresu. Wybór stacji oraz drukowanie wyników odbywa się w konsoli.

## 2. Uruchamianie programu

Główna funkcja programu to **main()**, znajdująca się w pliku **przewid.py**. Aby uruchomić program, należy uruchomić ten plik. Korzysta on z funkcji znajdujących się w plikach **stacje.py** oraz **pogodateraz.py**. Pozostałe pliki służą do pobierania i przetwarzania danych, na podstawie których obliczana jest prognoza. Dla danych o PM10 z roku 2017 te czynności zostały już wykonane, nie ma więc potrzeby używania tych plików w celu innym, niż rozszerzanie programu o kolejne zestawy danych.

## 3. Dane archiwalne używane przez program

Prognoza bazuje na pobranych uprzednio danych: ilości PM10 w powietrzu oraz danych pogodowych (temperatura, wilgotność, zachmurzenie, opady, kierunek wiatru, prędkość wiatru) z całości roku 2017, zbieranych w odstępach godzinnych. Dane pogodowe są wspólne dla całego miasta, natomiast dane o zanieczyszczeniu są indywidualne dla 8 stacji pomiarowych.

Dane zostały przetworzone w następujący sposób: dla każdego z czynników pogodowych oraz godzin powstał podział odpowiadających im odczytów zanieczyszczenia na podstawie wartości tych czynników. Na przykład: dla czynnika temperatury zbiór wszystkich pomiarów został podzielony na grupy, z których każda odpowiadała pomiarom wykonanym w konkretnej temperaturze.

Podział taki został wykonany osobno na danych z każdej stacji pomiarowej. W przypadku niektórych czynników konieczny był podział na segmenty odpowiadające przedziałom wartości, nie konkretnym wartościom. Na przykład: pomiary zależne od prędkości wiatru zostały podzielone na segmenty; są więc grupy pomiarów przy wietrze 0-4km/h, 5-9 km/h, 10-14 km/h, itd.

Z każdej grupy pomiarów została wyciągnięta wartość średnia oraz odchylenie standardowe wartości pomiaru. To właśnie te dane zostały zapisane w folderze dane/analiza. Znane są więc średnia wartość oraz odchylenie standardowe PM10 dla warunków pogodowych o takiej samej wartości danego czynnika, np. wartość przy temperaturze -23°C.

## 4. Źródła danych archiwalnych

Dane pogodowe: <https://pogoda.interia.pl/>

Dane zostały pobrane metodą web scrapingu.

Dane o zanieczyszczeniu: <https://powietrze.gios.gov.pl/pjp/archives>

Dane zostały ręcznie pobrane w formie arkuszy kalkulacyjnych.

## 5. Zasady działania

Prognoza dla każdej stacji bazuje na aktualnym, znanym stanie zanieczyszczenia oraz opisanych wyżej danych.

Pierwszym krokiem do otrzymania prognozy zanieczyszczeń jest pobranie aktualnej prognozy pogody - używane są dane dla najbliższych 70 godzin od momentu uruchomienia funkcji. Dla każdej godziny obliczana jest szacowana ilość zanieczyszczenia na podstawie danych opisywanych wcześniej. Dla każdego czynnika jest wyliczana poniższa wartość:

$$x = \text{średnia} * \frac{S-s}{(S*l)-1}$$

*S – suma odchyleń standardowych wszystkich czynników*

*s – odchylenie standardowe branego pod uwagę czynnika*

*l – ilość czynników*

*średnia – średnia wartość PM10 dla branego pod uwagę czynnika*

Szacunkową ilością smogu jest suma tych wartości obliczonych dla każdego czynnika.

Jak można zauważyć, program zakłada, że mniejsze odchylenie standardowe dla czynnika oznacza, że czynnik bardziej wpływa na stan zanieczyszczeń. Powodem tego jest to, że można założyć, że jeżeli wartości zanieczyszczenia mają bardziej stałe wielkości przy konkretnych warunkach, to warunki te na nie w jakiś sposób wpływają. Natomiast gdyby wartości przypisane danemu czynnikowi były całkowicie losowe, ich odchylenie standardowe byłoby większe i w rezultacie czynnik miałby mniejszy wpływ na szacowaną wartość.

Ostatnim krokiem jest wzięcie pod uwagę aktualnego stanu zanieczyszczenia. W tym celu tworzona jest nowa lista, zawierająca różnice między wartościami oszacowanymi dla każdej godziny. W ten sposób dowiadujemy się, jak zwiększa lub zmniejsza się zanieczyszczenie pod wpływem godziny oraz prognozowanych zmian pogody. Różnice te aplikowane są po kolei do aktualnej wartości zanieczyszczenia - w ten sposób powstaje prognoza dla każdej godziny.

## 6. Źródła danych aktualnych

Dane pogodowe: <https://pogoda.interia.pl/prognoza-szczegolowa-krakow.cld,4970>

Dane zostały pobrane metodą web scrapingu.

Dane o zanieczyszczeniu: <https://airly.eu/pl/>

Dane zostały pobrane poprzez API.

Airly nie udostępnia danych z dokładnie tych samych czujników, dla których istnieją dane archiwalne. Z tego powodu do pobierania aktualnych danych zostały użyte stacje znajdujące się najbliżej tych wcześniej badanych.

## 7. Zawartość plików

Używane do danych archiwalnych:

- **pogoda.py** - zawiera funkcję **pogodaD(dzien,miesiac,rok)** pobierającą dane pogodowe dla podanej daty
- **zapis.py** - zawiera funkcje zapisujące oraz wczytujące dane do i z pliku
- **testdanych.py** - służył do pobrania danych pogodowych z całego roku
- **dzielenie.py** - używany do dzielenia pomiarów w grupy zależne od danego czynnika oraz wyliczania ich średniej i odchylenia standardowego

Używane do danych aktualnych:

- **stacje.py** - pobiera dane o stacjach badających zanieczyszczenie oraz czyta pochodzące z nich pomiary z arkuszy kalkulacyjnych
- **pogodateraz.py** - służył do pobierania aktualnej prognozy pogody; zmodyfikowana wersja pliku **pogoda.py**
- **przewid.py** - główny plik programu; pobiera zapisane wcześniej dane, wyświetla menu dla użytkownika i oblicza prognozę zanieczyszczeń, po czym drukuje wyniki

Dane:

- **MetadaneKrk** - dane o lokalizacji stacji pomiarowych
- **2017\_PM10\_1g** i podobne - dane o zanieczyszczeniu z danego roku; w prognozie został użyty jedynie **2017\_PM10\_1g**
- **pogoda2017** - dane pogodowe z roku 2017 do odczytania przez program
- **folder analiza** - zawiera pliki, w których zapisano odchylenie standardowe, średnią odczytów dla danego czynnika oraz przedziały, dla których zostały one obliczone; do odczytania przez program

## 8. Dokładność modelu

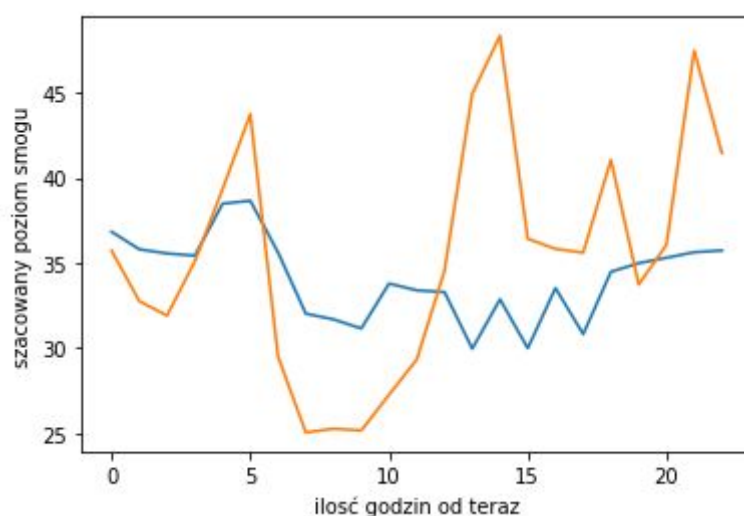
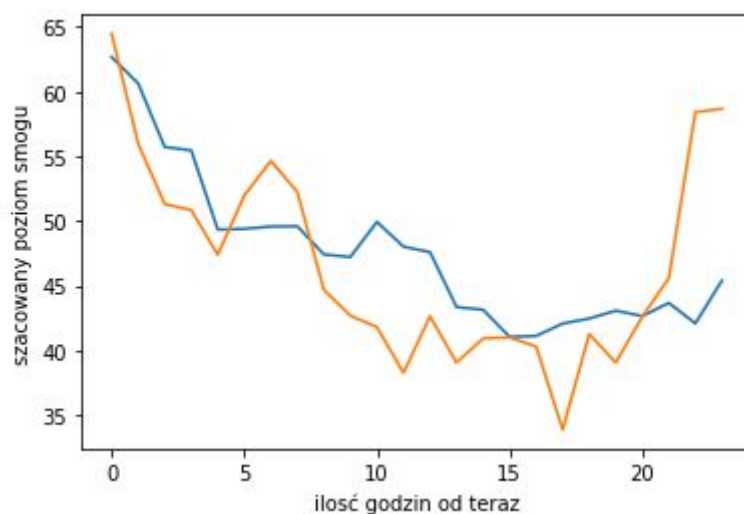
Model, jak wiele prognoz, nie pokrywa się dokładnie z rzeczywistością. Smog nie jest zjawiskiem łatwym do prognozowania, ponieważ zależy od wielu czynników, z których wielu nie da się zmierzyć lub nie są dostępne dane o ich pomiarach. Z tego powodu prognozy nie zawsze dobrze odzwierciedlają rzeczywistość, zwłaszcza w dużych przedziałach czasowych.

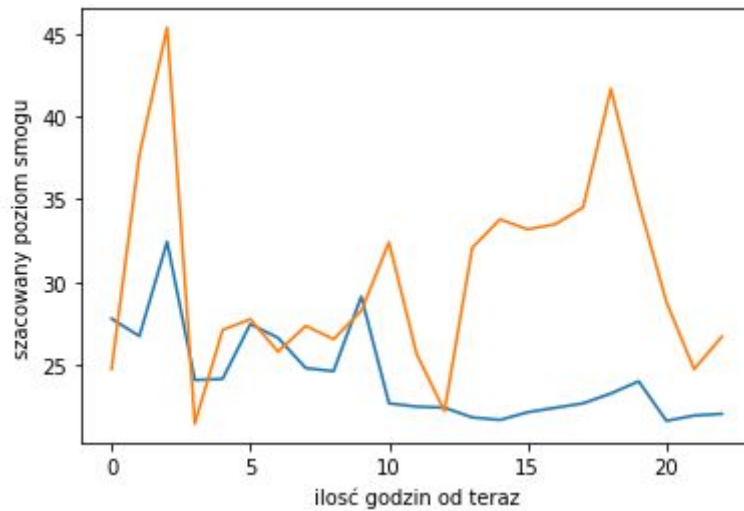
Błędy rzędu  $10 \mu\text{g}/\text{m}^3$  możemy uznać za akceptowalne - wiele oficjalnych prognoz podaje przewidywaną ilość w przedziałach takiej wielkości. W znacznej większości przypadków błędy prognoz na kilka najbliższych godzin są w granicach tej wartości, jednak błąd zazwyczaj rośnie wraz z czasem.

Model nie nadąża za nagłymi skokami w poziomie smogu - są one prawdopodobnie spowodowane losowymi czynnikami lub błędami w pomiarach, których nie jest w stanie przewidzieć.

Dokładność prognozy mogłaby potencjalnie zostać zwiększona przez użycie danych archiwalnych z większej ilości lat oraz uwzględnienie większej ilości czynników, nie pozwoliły jednak na to ramy czasowe projektu.

Poniżej przedstawiono kilka wykresów testujących zgodność prognozy z prawdą. Każdy z nich przedstawia 24 godziny losowo wybrane z danych z roku 2017. Linia pomarańczowa przedstawia rzeczywiste ilości PM10 z tego czasu, natomiast linia niebieska to prognoza bazująca na danych pogodowych oraz odczycie z pierwszej sprawdzanej godziny. Jak można zauważyć, błąd przekracza  $10 \mu\text{g}/\text{m}^3$  w sytuacji nagłych skoków pomiaru oraz w późniejszych częściach prognozy.





## 9. Bibliografia

Model został wykonany głównie metodą analizowania dostępnych danych i określania ich wagi dla wyniku, więc głównym źródłem były dane same w sobie. Oto treści, które pomogły w dobraniu danych i stworzeniu modelu:

- <https://www.waikatoregion.govt.nz/environment/natural-resources/air/weather-affects-air-quality/>
- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1616122/>
- <https://epa.tas.gov.au/epa/air/air-quality-in-tasmania/reducing-air-pollution/how-weather-affects-air-quality>