AGH
Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie
Wydział Fizyki i Informatyki Stosowanej

Praca doktorska

Przemysław Furman

Identyfikacja stopnia narażenia społeczności na wybrane ksenobiotyki z wykorzystaniem analizy zanieczyszczenia powietrza oraz modeli transportu zanieczyszczeń i algorytmów uczenia maszynowego

Pierwszy promotor rozprawy: **dr hab. inż. Mirosław Zimnoch, prof. AGH**

Drugi promotor rozprawy: **dr hab. inż. Katarzyna Styszko, prof. AGH**
Wydział Energetyki i Paliw AGH

Kraków, 2023

Oświadczenie autora rozprawy:
Oświadczam, świadomy odpowiedzialności karnej za poświadczenie nieprawdy, że niniejszą pracę doktorską wykonałem osobiście i samodzielnie i nie korzystałem ze źródeł innych niż wymienione w pracy.

data, podpis autora

Oświadczenie promotorów rozprawy:
Niniejsza rozprawa jest gotowa do oceny przez recenzentów.

data, podpis promotorów rozprawy

Niniejsza rozprawa doktorska została wykonana w ramach Programu Operacyjnego Wiedza Edukacja Rozwój, nr projektu POWR.03.02.00-00-I004/16, współfinansowanego ze środków Unii Europejskiej.

W trakcie realizacji rozprawy doktorskiej Przemysław Furman korzystał ze wsparcia finansowego w ramach Akcji COST „COLOSSAL”(nr CA16109), realizowanej w ramach Europejskiego Programu Współpracy w Dziedzinie Badań Naukowo-Technicznych (COST).

Składam serdeczne podziękowania:

*Moim Promotorom **Panu Profesorowi Mirosławowi Zimnochowi** oraz **Pani Profesor Katarzynie Styszko** za cierpliwość, zaangażowanie i nieocenioną pomoc otrzymaną podczas realizacji niniejszej pracy.*

*Wszystkim **pracownikom i doktorantom** Zespołu Fizyki Środowiska Wydziału Fizyki i Informatyki Stosowanej AGH oraz Zespołu Badań Współczesnych Zagrożeń Środowiska Wydziału Energetyki i Paliw AGH, a w szczególności:*

***Pani Doktor Alicji Skibie** oraz **Pani Magister Justynie Pamule** za współpracę, zaangażowanie w nasze badania naukowe, gotowość do pomocy i dzielenie się wiedzą, a także za wspólne dyskusje, staże i konferencje.*

Dziękuję również:

***Panu Doktorowi Dariuszowi Widełowi** z Uniwersytetu Jana Kochanowskiego w Kielcach za możliwość odbycia stażu naukowego i przeprowadzenia badań, które przyczyniły się do rozwoju osobistego i naukowego.*

*Wszystkim **pracownikom i doktorantom** z Uniwersytetu Technicznego w Wiedniu, a przede wszystkim:*

***Panu Profesorowi Erwinowi Rosenbergowi**, **Pani Profesor Anne Kasper-Giebl** oraz **Pani Doktor Magdalenie Kistler** za okazję do współpracy przy inspirujących projektach badawczych oraz za możliwość odbycia stażu naukowego.*

***Pani Doktor Bernadette Kirchsteiger**, **Pani Magister Vanessie Nuernberger** oraz **Pani Magister Danieli Kau** za owocną współpracę i pomoc przy realizacji badań naukowych podczas stażu.*

*Pragnę również podziękować mojej wspaniałej **Rodzinie:***

***Mojej Żonie Klaudii** za wiarę i cierpliwość. Dzięki Twojemu wsparciu przekroczyłem własne granice i osiągnąłem cele, które wydawały się nieosiągalne.*

***Mojej Córce Monice** za to, że była, jest i będzie moją największą motywacją w życiu.*

***Moim Rodzicom Stefanii i Dariuszowi**, dzięki którym miałem możliwość kształcić się i zdobywać cenną wiedzę, którzy stale mnie mobilizowali i wspierali przez cały okres trwania studiów na Akademii Górniczo-Hutniczej.*

Spis Treści

Streszczenie w języku Polskim

Celem niniejszej rozprawy doktorskiej było opracowanie nowych modeli identyfikacji stopnia narażenia społeczności na wybrane zanieczyszczenia powietrza w oparciu o analizy stężenia oraz składu chemicznego frakcji pyłu zawieszonego PM_{10}, modele transportu zanieczyszczeń oraz algorytmy uczenia maszynowego. Próbki pyłu zawieszonego PM_{10} zostały pobrane na terenie Wadowic w 2017 oraz Krakowa w latach 2020-2021.

Wykonano badania składu chemicznego PM_{10} pod względem zawartości wielopierścieniowych węglowodorów aromatycznych WWA za pomocą chromatografii gazowej sprzężonej ze spektrometrem mas GC-MS. Badaniom poddano łącznie 252 próbki pyłu PM_{10}. Analizy obejmowały 16 podstawowych WWA, określonych przez US EPA (*United States Environmental Protection Agency*) za najbardziej szkodliwe: Acenaften (Ace), Acenaftylen (Acy), Antracen (An), Benzo[*a*]antracen (B[*a*]A), Benzo[a]piren (B[*a*]P), Benzo[*b*]fluoranten (B[*b*]F), Benzo[*ghi*]perylen (B[*ghi*]P), Benzo[*k*]fluoranten (B[*b*]K), Chryzen (Ch), Dibenzo[*a,h*]antracen (DBA), Fenantren (Fen), Fluoranten (Flu), Fluoren (Fl), Indeno[*1,2,3-cd*]piren (IP), , Piren (Pir) i Naftalen (Na). Otrzymane informacje dotyczące stężeń WWA zostały wykorzystane do określenia profili źródeł zanieczyszczeń, profili narażenia oraz wartości wskaźników ekwiwalentu toksyczności wielopierścieniowych węglowodorów aromatycznych zalecanych przez EPA: ekwiwalentu działania mutagennego względem B[*a*]P (ang. *mutagenic equivalent, MEQ*), ekwiwalentu działania toksycznego względem B[*a*]P (ang. *toxic equivalent, TEQ*) i ekwiwalentu działania kancerogennego względem 2,3,7,8-tetrachlorodienzo-p-dioksyny (ang. *carcinogenic equivalent, CEQ*).

W niniejszej pracy wykonano analizy częstotliwości występujących kierunków napływu mas powietrza w celu uzyskania informacji na temat możliwości transportu zanieczyszczeń z wybranych obszarów z okolic badanych miejsc. Analizy przeprowadzono za pomocą modelu HYSPLIT NOAA Air Resources Laboratory (*Hybrid Single-Particle Lagrangian Integrated Trajectory model*) opracowanego przez NOAA Air Resources Laboratory (*National Oceanic and Atmospheric Administration*). Interpretacja otrzymanych wyników trajektorii pozwoliła oszacować charakter źródeł zanieczyszczeń powietrza.

Dane dotyczące wskaźników ekwiwalentu toksyczności WWA (*MEQ, CEQ, TEQ*) zostały wykorzystane do opracowania modeli identyfikacji stopnia narażenia społeczności. Obliczenia zostały przeprowadzone w środowisku Azure Machine Learning. W modelach wykorzystano cztery algorytmy uczenia maszynowego: wektory nośne, regresje logistyczną, lasy losowe oraz sztuczne sieci neuronowe.

W oparciu o powyższe badania, najwyższą dokładność wykazano dla modelu sieci neuronowych oraz lasów losowych. Modele bazujące na algorytmach lasów losowych wykazały nadmierne dopasowanie. Najniższą dokładnością charakteryzowały się algorytmy wektorów nośnych oraz regresji logistycznej. Ocena istotności wpływu parametrów na dokładność modelu wykazała największy wpływ miejsca pobierania próbek na dokładność modelu. Przeprowadzone badania mogą stanowić podstawę do rozwoju systemów wczesnego ostrzegania bazujących na algorytmach sztucznych sieci neuronowych.

Abstract

The aim of this doctoral dissertation was to develop new models for identifying the level of community exposure to selected air pollutants based on analysis of particulate matter fractions PM_{10}, pollution transport models and machine learning algorithms. Samples of PM_{10} were collected in Wadowice in 2017 and Krakow in 2020-2021.

The chemical composition of PM_{10} in terms of the content of polycyclic aromatic hydrocarbons (PAHs) was carried out using gas chromatography-mass spectrometry (GC-MS). A total of 252 samples of particulate matter PM_{10} were analyzed. The analyzes contained 16 basic PAHs identified by the US EPA as the most harmful: Acenaphtene (Acn), Acenaphthylene (Acy), Anthracene (Ant), Benzo[*b*]fluoranthene (B[*b*]F), Benzo[*a*]anthracene (B[*a*]A), Benzo[*a*]pyrene (B[*a*]P), Benzo[*ghi*]perylene (B[*ghi*]P), Benzo[*k*]fluoranthene (B[*k*]F), Chrysene (Chry), Dibenzo[*ah*]anthracene (D[*ah*]A), Fluoranthene (Flt), Fluorene (Flu), Indeno[*1,2,3-cd*]pyrene (IP), Naphthalene (Nap), and Phenanthrene (Phen) and Pyrene (Pyr). The obtained information on the concentrations of PAHs was used to determine the profiles of pollution sources, exposure profiles and the values of polycyclic aromatic hydrocarbons equivalent toxicity indicators recommended by the EPA: mutagenic equivalent to B[*a*]P (ang. *mutagenic equivalent, MEQ*), toxic equivalent to B[*a*]P (ang. *toxic equivalent, TEQ*) and carcinogenic equivalent to 2,3,7,8-tetrachlorodienzo-p-dioxin (ang. *carcinogenic equivalent, CEQ*).

In this work, air trajectory frequency analysis were performed in order to obtain information on the possibility of transporting pollutants from selected areas in the vicinity of the studied sites. The analyzes were performed using the NOAA Air Resources Laboratory's HYSPLIT model (*Hybrid Single-Particle Lagrangian Integrated Trajectory model*) developed by the NOAA Air Resources Laboratory (*National Oceanic and Atmospheric Administration*). The interpretation of trajectory results provided insights into the nature of air pollution sources.

Data of polycyclic aromatic hydrocarbons equivalent toxicity indicators (*MEQ, CEQ, TEQ*) were used to develop models for identifying the level of community exposure. The calculations were performed using Azure Machine Learning. Four machine learning algorithms were used in the models: support vectors machine, logistic regressions, decision forest and neural network.

Based on the analyses, the models of neural networks and decision forests demonstrated the highest accuracy. However, the models based on decision forest algorithms showed signs of overfitting, while the support vector machine and logistic regression algorithms had lower accuracy. The assessment of parameter significance on model accuracy indicated that the sampling site had the greatest influence. This research can serve as a basis for the development of early warning systems using neural network algorithms.

Wprowadzenie i cele pracy

Liczne badania prowadzone w dziedzinie jakości powietrza wykazały, że zanieczyszczenie powietrza pyłem zawieszonym (ang. *Particulate Matter,* PM) ma globalny wpływ na klimat i zdrowie ludzi. Długotrwałe przebywanie w atmosferze silnie zanieczyszczonej pyłem może wywoływać choroby układu oddechowego, problemy z układem odpornościowym, chorobę Alzheimera, może przyczyniać się do powstawania nowotworów. Z tych względów pyłowe zanieczyszczenia powietrza są jednym z zanieczyszczeń, które wymagają stałego monitoringu. Aby zaprojektować skuteczne strategie monitorowania lub redukcji PM należy szczegółowo poznać ich stężenia, źródła, skład chemiczny pod względem zawartości związków nieorganicznych i organicznych z różnych regionów świata.

Charakterystyka chemiczna frakcji pyłów zawieszonych jest istotną informacją płynącą z badań dotyczących zanieczyszczenia powietrza, w odniesieniu do możliwych negatywnych skutków zdrowotnych. Z powodu narażenia na duże stężenia PM, Polska jest jednym z krajów europejskich borykającym się z tym problemem od dawna. Ze względu na ukształtowanie terenu, utrudnione jest rozpraszanie zanieczyszczeń, szczególnie w południowej części kraju, gdzie dominuje krajobraz górski i wyżynny. Dodatkowo charakteryzuje się dobrze rozwiniętym przemysłem oraz wysokim poziomem zaludnienia. Jest to obszar najbardziej dotknięty problemem niskiej jakości powietrza. W nawiązaniu do danych WHO (*World Health Organization, Światowa Organizacja Zdrowia*) z 2016 roku, liczba zgonów związana z utrzymującą się niską jakością powietrza została wyznaczona dla Polski na poziomie 68,9 na 100 000 mieszańców, co dało 28 pozycję wśród krajów europejskich. Pierwsze miejsce w rankingu zajmuje Szwecja (0,4 zgony na 100 000 mieszkańców). WHO szacuje, że na świecie w wyniku złej jakości powietrza umiera 7 000 000 osób każdego roku.

Ocena danych z chwilowych lub długoterminowych badań monitoringu powietrza i znajomość składu chemicznego pyłu zawieszonego może pomóc w pełniejszym zrozumieniu wpływu jakości powietrza na środowisko i zdrowie ludzi. W wielu badaniach powietrza atmosferycznego prowadzonych dla miast południowej Polski obserwuje się znacznie przekroczone dopuszczalne normy średniego dobowego stężenia pyłu PM_{10} wynoszące 50 $\mu g/m^3$. Ocena danych dotyczących stężeń pyłów jest istotna, jednak przy ocenie wpływu na zdrowie ważna jest też znajomość składu chemicznego, w szczególności zawartości związków silnie rakotwórczych jak wielopierścieniowe węglowodory aromatyczne (WWA). Rozszerzenie wiedzy o skład chemiczny może pomóc w pełniejszym zrozumieniu charakterystyki zanieczyszczeń i zagrożenia, jakie jest z nimi związane. Otrzymane informacje ułatwiają określenie profili narażenia oraz źródeł emisji WWA, a także pozwalają oszacować wartości wskaźników ekwiwalentu toksyczności wielopierścieniowych węglowodorów aromatycznych zalecanych przez US EPA (*United States Environmental Protection Agency*): ekwiwalentu działania mutagennego względem B[*a*]P (ang. *mutagenic equivalent, MEQ*), ekwiwalentu działania toksycznego względem B[*a*]P (ang. *toxic equivalent, TEQ*) i ekwiwalentu działania kancerogennego względem 2,3,7,8-tetrachlorodienzo-p-dioksyny (ang. *carcinogenic equivalent, CEQ*). W celu uzupełnienia informacji na temat możliwych źródeł emisji zanieczyszczeń, często stosowane są prognozy dyspersji zanieczyszczeń metodami deterministycznymi. Przykładem tego typu badań są analizy częstotliwości występujących kierunków napływu mas powietrza za pomocą modelu HYSPLIT NOAA Air Resources Laboratory (*Hybrid Single-Particle Lagrangian Integrated Trajectory model*) opracowanego przez NOAA Air Resources Laboratory (*National Oceanic and Atmospheric Administration*).

Interpretacja otrzymanych wyników trajektorii wstecznych pomaga oszacować charakter źródeł emisji – lokalny lub napływowy.

W zarządzaniu ryzykiem zdrowotnym i środowiskowym dużą rolę odgrywa system prognozowania stanu jakości powietrza. Prognozy pozwalają na wczesne ostrzeganie i alarmowanie społeczeństwa. Z pomocą przychodzi dział uczenia maszynowego i rozwiązania w postaci modeli predykcji. Modele bazują na odpowiednich algorytmach (np.: algorytmy wektorów nośnych, regresji logistycznej, lasów losowych oraz sieci neuronowych), których zadaniem jest taka interpretacja danych historycznych (np.: warunki meteorologiczne, stężenia zanieczyszczeń) oraz historycznych rezultatów (np.: przekroczone dopuszczalne poziomy stężeń pyłów zawieszonych), tak aby w przyszłości (w oparciu o nowe dane) model mógł wygenerować poprawną prognozę. Modele umożliwiają odwzorowywanie złożonych zależności przyczynowo skutkowych w identyfikacji stopnia narażenia społeczeństwa na wybrane zanieczyszczenia powietrza.

Jednym z tego typu algorytmów są sztuczne sieci neuronowe powstałe na wzór procesów, które przebiegają w mózgu istot żywych. Podobnie do swoich biologicznych pierwowzorów są zdolne do automatycznej adaptacji składników posiadanej struktury w celu znalezienia najlepszego rozwiązania złożonych zależności bazując na konkretnych danych wejściowych i przewidywanych danych wyjściowych modelu. Sieci neuronowe pozwalają rozwiązać zawiłe problemy nieliniowe, wielowymiarowe, wieloklasowe, często bardzo trudne do rozwiązania w sposób konwencjonalny. Predykcje te opierają się na znajomości stężeń frakcji pyłu zawieszonego PM_1, $PM_{2.5}$, PM_{10}, warunków meteorologicznych z uwzględnieniem innych zanieczyszczeń powietrza: tlenki siarki, tlenki azotu, tlenki węgla, ozon, benzen. Pomijane są natomiast często w tego typu układach parametry stężeń dla pozostałych zanieczyszczeń, takich jak wielopierścieniowe węglowodory aromatyczne.

Badania podjęte w niniejszej pracy opierały się na interdyscyplinarnym podejściu łączącym analizy chemiczne, modele deterministyczne oraz dział uczenia maszynowego. Część chemiczna skupiała się na analizie składu pyłu zawieszonego PM_{10} pobranego w Wadowicach w 2017 roku oraz w Krakowie w latach 2020/2021 pod względem zawartości wielopierścieniowych węglowodorów aromatycznych. Za pomocą chromatografii gazowej sprzężonej ze spektrometrem mas GC-MS, analizie poddano łącznie 252 próbki PM_{10}. Badania obejmowały 16 podstawowych WWA, określonych przez US EPA (*United States Environmental Protection Agency*) za najbardziej szkodliwe: Acenaften, Acenaftylen, Antracen, Benzo[*a*]antracen, Benzo[*a*]piren, Benzo[*b*]fluoranten, Benzo[*ghi*]perylen, Benzo[*k*]fluoranten, Chryzen, Dibenzo[*ah*]antracen, Fenantren, Fluoranten, Fluoren, Indeno[*1,2,3-cd*]piren, Piren i Naftalen. Otrzymane informacje dotyczące stężeń WWA zostały wykorzystane do określenia profili źródeł zanieczyszczeń, profili narażenia oraz wartości wskaźników *MEQ*, *TEQ* i *CEQ*. Dla próbek charakteryzujących się najwyższymi stężeniami PM_{10} wykonano analizy częstotliwości występujących kierunków napływu mas powietrza za pomocą modelu HYSPLIT NOAA Air Resources Laboratory. Dane dotyczące wskaźników ekwiwalentu toksyczności WWA (*MEQ, CEQ, TEQ*) zostały wykorzystane do opracowania modeli identyfikacji stopnia narażenia społeczności. Obliczenia zostały przeprowadzone w środowisku Azure Machine Learning. W modelach wykorzystano cztery algorytmy uczenia maszynowego: wektory nośne, regresje logistyczną, lasy losowe oraz sztuczne sieci neuronowe. Finalny model opierał się na sztucznych sieciach neuronowych i został poddany implementacji w ogólnodostępnym środowisku *Web Service*.

W ramach rozprawy doktorskiej testowane były następujące tezy badawcze:

1) Stężenia pyłu zawieszanego PM_{10} oraz wielopierścieniowych węglowodorów aromatycznych wykazują wzajemne zależności, które są charakterystyczne dla poszczególnych miast oraz panujących warunków synoptycznych

2) Modele deterministyczne transportu zanieczyszczeń pomagają w określeniu lokalizacji ich źródeł (lokalne, napływowe)

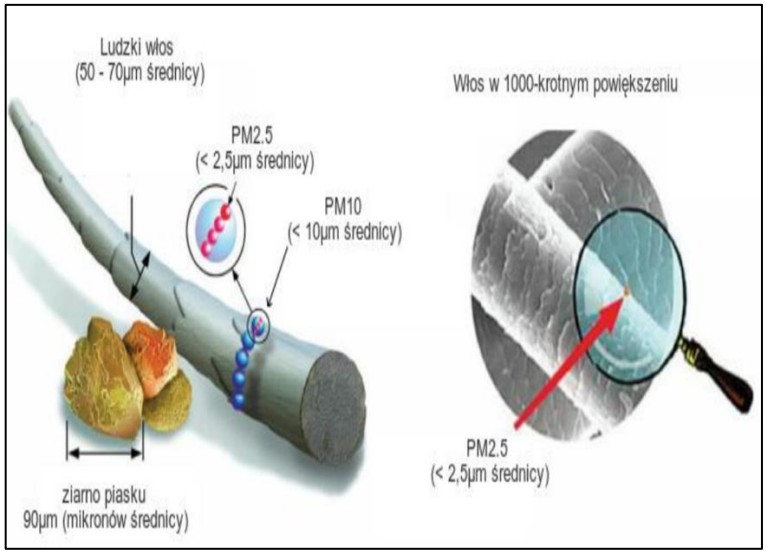
3) Uczenie maszynowe pozwala określić stopień narażenia społeczności na wybrane zanieczyszczenia powietrza (wielopierścieniowe węglowodory aromatyczne) z wykorzystaniem algorytmów sztucznych sieci neuronowych oraz wskaźników mutagenności (*MEQ*), kancerogenności (*CEQ*) i ekwiwalentu toksycznego działania (*TEQ*) pozwalając na przewidywanie poziomu wskaźników narażenia wymagających informacji o stężeniach całej gamy związków WWA jedynie na podstawie danych meteorologicznych, stężenia frakcji PM_{10} pyłu oraz stężenia B[a]P.

1. Część teoretyczna

1.1. Zanieczyszczenia powietrza

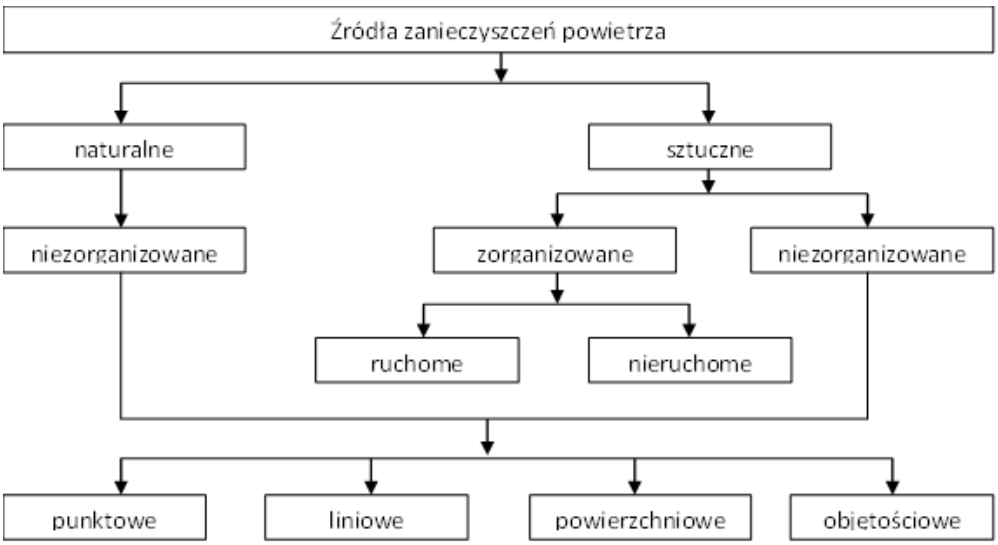
Stale rozwijająca się działalność człowieka obok korzyści gospodarczych, społecznych i ekonomicznych przynosi także skutki niepożądane. W większości tych przypadków nie są to działania bezpieczne – zagrażają one życiu i zdrowiu człowieka, wpływają bezpowrotnie na biosferę i ekosystem niszcząc ich środowiskową dynamikę i stabilizację, zwiększają ryzyko powstania katastrof ekologicznych, wypadków, pożarów. Wszystko to może przekładać się na poważne szkody społeczne i materialne. Dodatkowym skutkiem działalności człowieka jest emisja do środowiska naturalnego olbrzymich ilości zanieczyszczeń powietrza. Z powodu ich szybkiego rozprzestrzeniania się, reakcji fizyko-chemicznych, akumulacji w środowisku oraz organizmach żywych, opóźnionych reakcji biologicznych – niebezpieczeństwa płynące z obecności tych szkodliwych substancji nabierają charakteru globalnego o dalekosiężnych skutkach [Furman i in. 2021, Li i in. 2017].

Do najbardziej znanych i szczególnie niebezpiecznych zanieczyszczeń zaliczamy pyły zawieszone (PM). Są to wszelkie cząstki stałe oraz ciekłe znajdujące się w powietrzu tworzące mieszaniny dwufazową, gdzie cząstki PM stanowią fazę rozproszoną, a powietrze fazę rozpraszającą. Dyrektywa Parlamentu Europejskiego i Rady 2008/50/WE z dnia 21 maja 2008 r. w sprawie jakości powietrza i czystszego powietrza dla Europy pył zawieszony definiuje jako *„pył przechodzący przez otwór sortujący, zdefiniowany w referencyjnej metodzie pobierania próbek i pomiaru PM_{10}/$PM_{2.5}$, PN-EN 12341:2014, przy 50% granicy sprawności dla średnicy aerodynamicznej do 10µm/2.5µm”* [Dyrektywa Parlamentu Europejskiego i Rady 2008]. Pył zawieszony został podzielony na 6 klas uwzględniających średnice cząstek stałych w mikrometrach *µm*: PM_{10}, $PM_{2.5}$-PM_{10}, $PM_{2.5}$, PM_1, $PM_{0.1}$ oraz TSP jako całkowity pył zawieszony [Główny Inspektorat Ochrony Środowiska 2016, Szramowiat-Sala i in. 2014]. Poniżej (rysunek 1) zaprezentowano rozmiary PM w porównaniu do ludzkiego włosa oraz ziarna piasku:



Rysunek 1. Rozmiary cząstek pyłu zawieszonego [Li I in. 2017]

Wyróżniono 2 główne źródła emisji pyłów do powietrza. Pierwszym z nich są źródła naturalne. Do tej grupy zaliczamy emisje z pożarów lasów, wybuchów wulkanów, naturalne procesy erozyjne gleb i skał itd. Do drugiej grupy zaliczamy wspomnianą wcześniej działalność człowieka – źródła antropogeniczne (sztuczne): transport, paleniska domowe, przemysł (m.in. koksownie, przemysł metalurgiczny), ciepłownie, produkcja energii elektrycznej. Dodatkowym podziałem źródeł PM jest podział ze względu na miejsce i charakter występowania (rysunek 2)



Rysunek 2. Podział źródeł emisji pyłów zawieszonych [Janka 2022]

Najnowsze rozporządzenie Światowej Organizacji Zdrowia z 2021r. określa maksymalne dopuszczalne stężenia średnioroczne $PM_{2.5}$ oraz PM_{10} odpowiednio 5 $\mu g/m^3$ i 15 $\mu g/m^3$. Natomiast dopuszczalne średnie dobowe stężenie dla $PM_{2.5}$ zostało zmniejszone z 25 $\mu g/m^3$ do 15 $\mu g/m^3$, dla PM_{10} z 50 $\mu g/m^3$ do 45 $\mu g/m^3$ [World Health Organization 2021]. Dodatkowo nowelizacja Rozporządzenia Ministra Środowiska z dnia 8 października 2019r. w sprawie poziomów niektórych substancji w powietrzu zmieniła poziom informowania i poziom alarmowy odnoszący się do stężeń PM_{10}. Zgodnie z rozporządzeniem poziom alarmowy ogłaszany jest przy przekroczeniu średniej dobowej wartości 150 $\mu g/m^3$ PM_{10} (wcześniej 300 $\mu g/m^3$) oraz poziom informowania 100 $\mu g/m^3$ (wcześniej 200 $\mu g/m^3$) [Sejmik Województwa Małopolskiego 2020].

Parametrem kształtującym właściwości pyłów zawieszonych jest ich skład chemiczny. Pył emitowany prosto do atmosfery ze źródła określany jest jako pył pierwotny, natomiast pył wtórny to taki, który uległ procesom sorpcyjnym w powietrzu. Dla organizmów żywych najbardziej niebezpieczne są pyły zawieszone z zaadsorbowanymi na swojej powierzchni metalami ciężkimi (ołów, rtęć, kadm, cynk, chrom, nikiel, żelazo) oraz ksenobiotykami (wielopierścieniowe węglowodory aromatyczne WWA) [Borchers i in 2019, Jo i in. 2017, Tsai i in 2019]. Związki te zaliczane są do grupy substancji silnie kancerogennych i mutagennych, których właściwości fizykochemiczne i oddziaływanie na organizm ludzki nie jest do końca znane. Mimo stosunkowo niewielkiego udziału masowego związków z grupy WWA w pyłach zawieszonych są one najczęściej badanymi składnikami organicznymi [Dyrektywa Parlamentu Europejskiego i Rady 2008, Kozielska i in. 2016, Szramowiat i in. 2016].

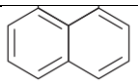
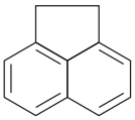
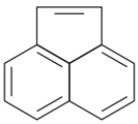
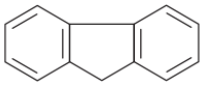
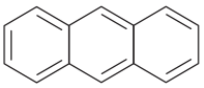
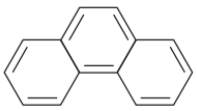
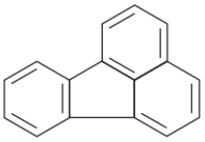
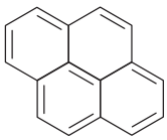
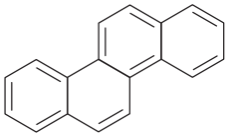
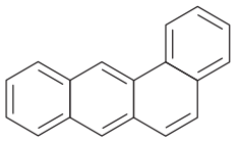
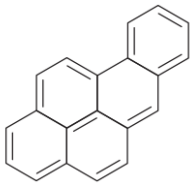
1.2. Ksenobiotyki

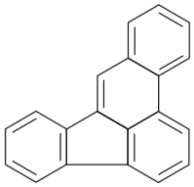
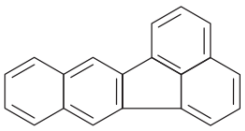
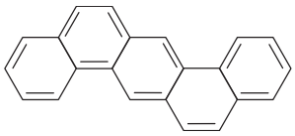
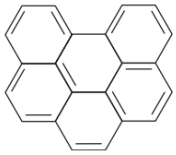
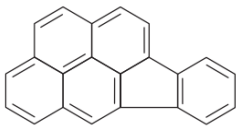
Ksenobiotyki stanowią grupę związków chemicznych o wysokiej aktywności biologicznej, obcych dla organizmów żywych. Są to głównie substancje wprowadzone sztucznie do środowiska, takie jak: leki, pestycydy, czynniki narażenia zawodowego oraz zanieczyszczenia środowiska pochodzenia organicznego i nieorganicznego (np.: WWA) czyli związki chemiczne otrzymane przez człowieka o strukturze chemicznej niewystępującej w przyrodzie, do których organizmy nie przystosowały się na drodze wcześniejszej ewolucji. Ksenobiotyki przedostają się do organizmu celowo (np.: prowadzenie terapii) lub w sposób niekontrolowany, gdzie później ulegają procesom zwanym metabolizmem ksenobiotyków (dystrybucja, biotransformacja, wydalanie).

1.2.1. Wielopierścieniowe węglowodory aromatyczne WWA

Do grupy ksenobiotyków zalicza się wielopierścieniowe węglowodory aromatyczne (WWA) (ang. *polycycic aromatic hydrocarbons PAHs*). Grupa związków organicznych stanowiących blisko 10 000 związków, które zawsze występowały w środowisku naturalnym, lecz działalność człowieka i stały rozwój cywilizacyjny doprowadził do wzrostu stężeń tych związków. Amerykański Narodowy Instytut Standaryzacji i Technologii stworzył listę ponad 660 wzorów strukturalnych WWA. Struktury tego typu związków charakteryzują się tym, że atomy wodoru i węgla tworzą minimum dwa sprzężone pierścienie aromatyczne bez żadnych innych elementów (bez heteroatomu/podstawnika). W środowisku naturalnym rozpowszechnionych jest 16 najbardziej znanych WWA, które stanowią podstawę większości badań (tabela 1) [Kozielska 2018, Kubiak 2013].

Tabela 1. Najczęściej oznaczane w środowisku wielopierścieniowe węglowodory aromatyczne [Kubiak 2013]

Związek WWA	Wzór sumaryczny	Wzór strukturalny	Masa cząsteczkowa	Temperatura wrzenia [°C]
Naftalen	$C_{10}H_8$		128,2	218,0
Acenaften	$C_{12}H_{10}$		154,2	96,2
Acenaftylen	$C_{12}H_8$		154,2	265,0-275,0
Fluoren	$C_{13}H_{10}$		166,2	295,0
Antracen	$C_{14}H_{10}$		178,2	342,0
Fenantren	$C_{14}H_{10}$		178,2	340,0
Fluoranten	$C_{16}H_{10}$		202,3	375,0
Piren	$C_{16}H_{10}$		202,3	404,0
Chryzen	$C_{18}H_{12}$		228,3	448,0
Benzo[a]antracen	$C_{18}H_{12}$		228,3	437,5
Benzo[a]piren	$C_{20}H_{12}$		252,3	495,0

Benzo[*b*]fluoranten	$C_{20}H_{12}$		252,3	481,2
Benzo[*k*]fluoranten	$C_{20}H_{12}$		252,3	480,0
Dibenzo[*ah*]antracen	$C_{22}H_{14}$		278,4	269,0-270,0
Benzo[*ghi*]perylen	$C_{22}H_{12}$		276,3	500,0
Indeno[*1,2,3-cd*]piren	$C_{22}H_{12}$		276,3	530,0

Wielopierścieniowe węglowodory aromatyczne w środowisku występują w postaci mieszaniny związków hetero- i homocyklicznych. Do pierwszej grupy należą związki posiadające w swojej budowie atomy tlenu, azotu lub/i siarki, natomiast do drugiej należą WWA z węglowodorami zbudowanymi z węgla i wodoru oraz podstawników alkilowych, grup aminowych lub/i nitrowych. Związki posiadające w swojej strukturze 5 i więcej pierścieni aromatycznych określane są jako „ciężkie", pozostałe jako „lekkie". WWA to związki hydrofobowe, litofilne, charakteryzują się słabą lipofilowością, są niepolarne. „Lekkie" WWA są mniej stabilne oraz wykazują mniejsze właściwości kancero i mutagenne. Wielopierścieniowe węglowodory aromatyczne w czystej postaci to ciała stałe, krystaliczne o specyficznej białej lub żółto-zielonej barwie oraz wysokiej temperaturze topnienia i niskiej prężności pary. Wszystkie wymienione cechy WWA są zależne od ich masy cząsteczkowej, co przekłada się w znacznym stopniu na zróżnicowane pojawianie się i egzystencję w środowisku naturalnym (woda, powietrze, gleba, organizmy żywe). WWA bardzo słabo rozpuszczają się w wodzie, natomiast dobrze w rozpuszczalnikach organicznych. Występowanie związków organicznych w wodzie zwiększa ich rozpuszczalność. Powyższe właściwości sprzyjają sorpcji WWA na powierzchniach ciał stałych. Z tego powodu zakłada się, że zdecydowana większość węglowodorów aromatycznych występuje w środowisku naturalnym w postaci zaadsorbowanej niezależnie od stanu skupienia fazy rozpraszanej (woda, powietrze). Główne reakcje, którym ulegają WWA w powietrzu lub wodzie to reakcje podstawiania i przyłączania, podczas których następuje rozerwanie wiązań nienasyconych między atomami węgla w pierścieniu aromatycznym. Czynnikami sprzyjającymi reakcjom jest światło, obecność tlenu, ozonu i wielu innych utleniaczy [Furman i in. 2021, Kozielska 2018, Kubiak 2013, Skiba i in. 2018].

1.2.2. Powstawanie oraz źródła WWA

Tak jak w przypadku pyłów zawieszonych, tak i dla WWA wyróżnia się 2 grupy źródeł pochodzenia: antropogeniczne i naturalne. Do naturalnych źródeł emisji zalicza się aktywność głównie aktywność wulkaniczną i pożary lasów. Jako antropogeniczne źródła, w pierwszej kolejności należy wymienić spalanie paliw kopalnianych, przetwórstwo ropy naftowej, transport, spalanie odpadów komunalnych. Wielkość emisji WWA do atmosfery oraz ich wzajemne relacje zależą od sprawności wykorzystywanych urządzeń ochrony środowiska (filtry itp.). Największe ilości WWA emitowane są podczas niecałkowitych i niekontrolowanych procesów spalania węgla lub materii organicznej (drewno), a także w spalinach samochodowych, czy dymie papierosowym. Uwolnione do atmosfery WWA i zaadsorbowane na powierzchniach pyłów mogą być transportowane na spore odległości, co przekłada się na kumulacje tych zanieczyszczeń na obszarach słabo zindustrializowanych lub typowo rolniczych, przez co transport pyłów zawieszonych traktowany jest jako osobne źródło emisji WWA. Opadające lub wymywane z atmosfery pyły z WWA kondensują się na glebie lub w wodzie i mogą w łatwy sposób przedostawać się do surowców rolno-spożywczych. Z tego powodu Amerykańska Agencja Ochrony Środowiska (US EPA) podjęła decyzję o konieczności kontrolowania zawartości wielopierścieniowych węglowodorów aromatycznych we wszystkich elementach środowiska naturalnego (woda, powietrze, gleba, rośliny). W celu dokładniejszej identyfikacji źródeł emisji, WWA mogą być traktowane jako markery ów źródeł – pojawienie się konkretnych WWA może być powiązane z konkretnymi źródłami pochodzenia. Dla przykładu:

- Fluoren, Piren, Fluoranten są traktowane głównie jako markery spalania paliw w silnikach benzynowych i diesla oraz spalania odpadów, oleju,
- Chryzen, Benzo[*k*]fluoranten, Benzo*[b*]fluoranten odpowiadają za emisję zanieczyszczeń w przemyśle hutniczym, koksowniach, traktowane są jako marker spalania drewna,
- Indeno[*1,2,3-cd*]piren, Benzo[*ghi*]perylene to markery emisji szerokiego zakresu począwszy od transportu (silniki diesla i benzynowe) do spalania węgla, oleju lub gazu ziemnego.

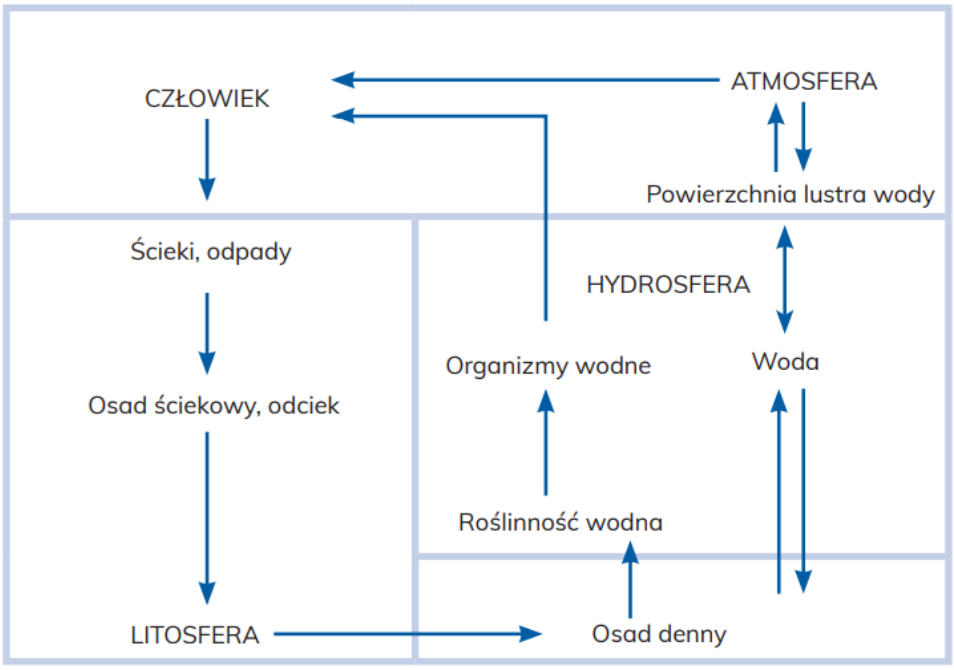
Sama identyfikacja WWA nie pozwala w jednoznaczny sposób stwierdzić, o które źródło emisji chodzi. W tym celu sporządzono stosunki danych węglowodorów określane mianem wskaźników diagnostycznych. Ich zastosowanie opiera się na założeniu, że konkretne wartości stosunków stężeń WWA są charakterystyczne i ściśle zależne dla jednego źródła emisji. W tabeli 2 zaprezentowano przykładowe i najczęściej stosowane wskaźniki diagnostyczne emisji wielopierścieniowych węglowodorów aromatycznych [Pohl 2019]:

Tabela 2. Wskaźniki diagnostyczne WWA [Kulshrestha i in. 2019, Simoneit 2015, Yunker i in. 2002]

Wskaźnik	Wartość	Źródło
$\frac{Fluoren}{Fluoren + Piren}$	< 0.50	Silniki benzynowe
	> 0.50	Silniki diesla
$\frac{Fluoranten}{Fluoranten + Piren}$	< 0.50	Spalanie paliwa (benzyna/diesel)
	> 0.50	Spalanie węgla/drewna
$\frac{Benzo[b]fluoraten}{Benzo[k]fluoraten}$	~ 0.92	Spalanie drewna
	~ 1.26	Silniki benzynowe
	2.50 - 2.90	Przemysł (huty)
	3.50 - 3.90	Węgiel/koks
$\frac{Piren}{Benzo[a]piren}$	0.90 ± 0.40	Silniki benzynowe
	0.80 ± 0.90	Silniki diesla
	0.70	Spalanie drewna
$\frac{Benzo[a]piren}{Benzo[a]piren + Chryzen}$	0.08-0.39	Spalanie drewna
	< 0.50	Ogrzewanie domów (drewno/węgiel)
	> 0.50	Transport
$\frac{Indeno[1,2,3-cd]piren}{Indeno[1,2,3-cd]piren + Benzo[ghi]perylen}$	~ 0.18	Transport
	~ 0.37	Silniki diesla
	~ 0.32	Silniki benzynowe
	~ 0.32	Gaz ziemny
	~ 0.36	Spalanie oleju
	~ 0.56	Spalanie węgla
	~ 0.64	Spalanie drewna
$\frac{Benzo[a]antracen}{Benzo[a]antracen + Chryzen}$	~ 0.50	Transport
	~ 0.73	Silniki diesla i benzynowe

1.2.3. Przemiany WWA oraz transport w różnych elementach środowiska

Człowiek narażony jest na WWA ze wszystkich stron środowiska naturalnego. Emitowane węglowodory w procesach spalania w pierwszej kolejności trafiają do atmosfery, gdzie mogą w łatwy sposób przedostawać się na dalekie odległości. W zależności od warunków meteorologicznych są wymywane do wód powierzchniowych i gleb. W wodzie transport WWA zależny jest od ich rozpuszczalności jednak jest on zazwyczaj ograniczony przez niską rozpuszczalność większości WWA. Po związaniu się z osadami dennymi związki te ulegają procesom fotochemicznym i biologicznym, co przyczynia się do ich znacznej kumulacji w osadach. W stale zmieniających się warunkach wodnych, WWA są ponownie mobilizowane i najczęściej dochodzi do ponownego zanieczyszczenia wody przez węglowodory. Jest to nierozłączne zjawisko procesu krążenia materii w ekosystemach wodnych. Poniżej (rysunek 3) zaprezentowano obieg wielopierścieniowych węglowodorów aromatycznych w środowisku [Pohl 2019].



Rysunek 3. Obieg wielopierścieniowych węglowodorów aromatycznych w środowisku naturalnym [Pohl 2019]

1.2.4. Oddziaływanie WWA na środowisko i człowieka

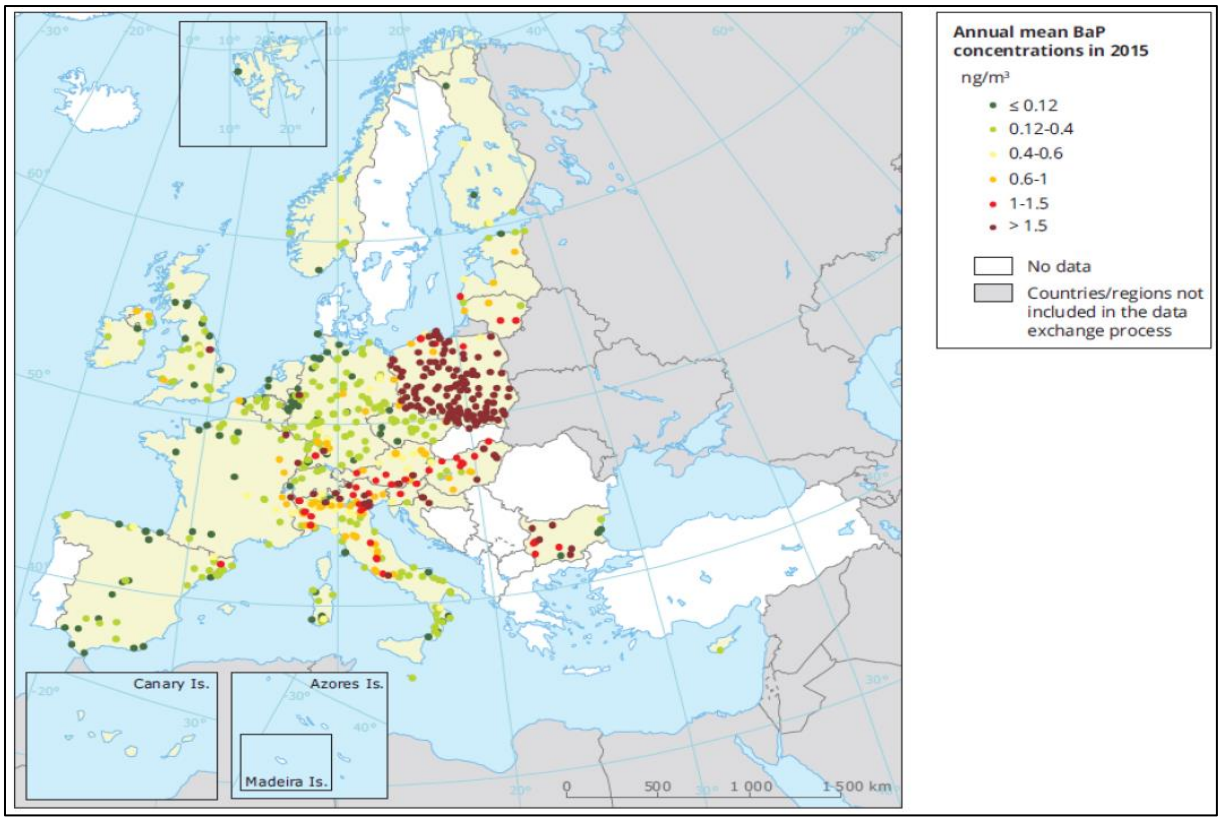
Wszystkie wielopierścieniowe węglowodory aromatyczne charakteryzują się właściwościami kancerogennymi i mutagennymi. W wyniku długotrwałego oddziaływania związków na organizmy żywe wykazano również ich silny wpływ na liczne mutacje biologiczne. WWA o mniejszych właściwościach kancerogennych w znacznym stopniu wzmagają efekt synergiczny. Agencja Ochrony Środowiska US EPA sporządziła listę 16 najbardziej szkodliwych WWA uwzględniając ich właściwości fizykochemiczne oraz oddziaływanie na środowisko i człowieka. Do tej grupy zalicza się: Acenaften, Acenaftylen, Antracen, Benzo[*a*]antracen, Benzo[*a*]piren, Benzo[*b*]fluoranten, Benzo[g*hi*]perylen, Benzo[*k*]fluoranten, Chryzen, Dibenzo[*ah*]antracen, Fenantren, Fluoranten, Fluoren, Indeno[*1,2,3-cd*]piren, Piren i Naftalen [Furman i in. 2021, Kozielska 2018, Kubiak 2013].

Wielopierścieniowe węglowodory aromatyczne mogą przedostawać się do organizmu człowieka poprzez układ oddechowy, pokarmowy lub przez skórę. WWA w postaci pary lub pyłów z łatwością osiadają w drogach oddechowych i przedostają się do płuc. WWA posiadają zdolność do tworzenia wiązań kowalencyjnych z cząsteczkami RNA i DNA, co znacznie przyczynia się do mutacji komórek, a co za tym idzie do inicjowania procesów nowotworowych. WWA i ich pochodne wpływają negatywnie na procesy rozrodcze organizmów żywych. Wykazano, że większa część ciężkich WWA przyczynia się do nieprawidłowości w budowie serca i powstawania defektów rozwojowych powłok brzusznych. Zaobserwowano zaniżenie masy ciała płodu narażonego na działanie WWA. Dodatkowo, WWA wpływają na spadek jakości i ilości produkowanego nasienia oraz jego zdolności do penetracji i zapłodnienia komórki jajowej. Wszystko to przekłada się na zwiększone ryzyko wystąpienia wad rozwojowych lub zmian nowotworowych organizmów żywych. WWA i ich pochodne intensyfikują niebezpieczeństwo związane z przedwczesnymi porodami i zaburzeniami wzrostu płodu. WWA posiadają zdolność do wiązania się ze strukturą DNA łożyska, a w połączeniu z ich silnymi właściwościami mutagennymi może dojść do samoistnych poronień we wczesnym okresie ciąży. Z tego względu zaleca się kobietom w ciąży w miarę możliwości unikania przebywania w środowisku silnie zanieczyszczonym pyłami

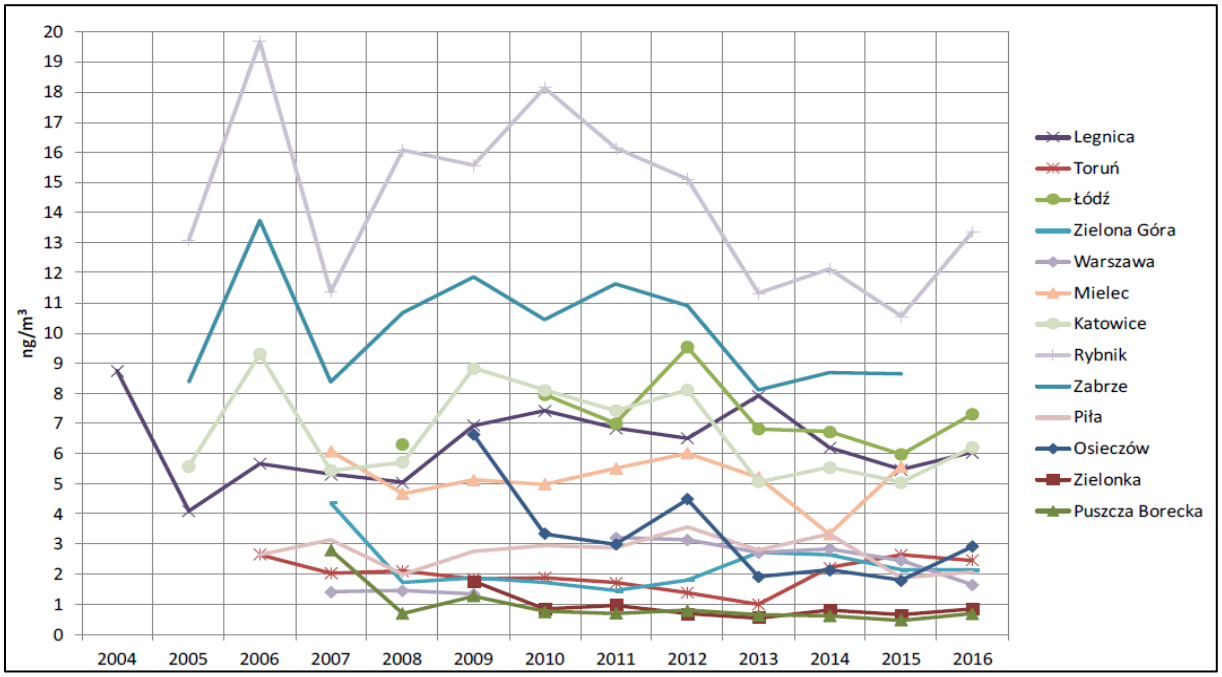
zawieszonymi. Wszystkie nitrowe pochodne WWA zostały zapisane przez Międzynarodową Agencję Badań nad Rakiem (IARC) [IARC 1987] oraz Agencję Ochrony Środowiska (US EPA) [U.S. EPA 2003] na listę związków wykazujących działania genotoksyczne, posiadające właściwości silnie rakotwórcze i silnie wpływające na DNA. Najsilniejszymi związkami kancerogennymi z grupy WWA jest Benzo[a]piren i Dibenzo[ah]antracen [Furman i in. 2021, Kozielska 2018, Kubiak 2013, Skiba i in. 2019, Weinmayr i in. 2018, Zaciera 2007].

Pomimo tego, że większa część związków organicznych przedostaje się do organizmu człowieka drogą oddechową to większe szkody mogą wywoływać WWA przedostające się drogą pokarmową, z pożywieniem. Spożywana żywność może być zanieczyszczona związkami WWA przez powietrze, wodę oraz glebę. Większa część WWA zostaje zatrzymana na owocach i warzywach na ich woskowej powierzchni. Warzywa, nasiona i ziarno spożywane przez organizmy żywe jest najczęściej zanieczyszczona WWA zawartymi w powietrzu i wodzie. Warzywa mają styczność z WWA zawartymi w glebie i posiadają zdolność ich metabolizowania. Żywność produkowana na obszarach z rozwiniętym transportem narażona jest na sorpcję dużych ilości WWA i znacznie większych ilości pochodn-WWA. Produkty takie jak ryby i owoce morza pochłaniają szkodliwe związki zawarte w wodzie i osadach dennych jednak stężenia w tych produktach jest zależna od zdolności do metabolizmu tych związków. Do układu pokarmowego przedostają się również WWA powstałe podczas obróbki żywności: wędzenie, ogrzewanie, smażenie, grillowanie, pieczenie, palenie kawy, ekstrakcja oleju. Badania udowodniły znaczne zwiększenie stężeń WWA po obróbce termicznej. Dodatkowo, im wyższa temperatura i czas takiej obróbki tym więcej szkodliwych związków powstaje [Rogula-Kozłowska i in. 2017].

Benzo[a]piren to jedyny związek WWA objęty normami stężeniowymi, które wynoszą 1 ng/m^3 w skali rocznej [Dyrektywa Parlamentu Europejskiego i Rady 2008]. Benzo[a]piren jest powszechnie oznaczany w analizach środowiskowych jako marker całkowitej zawartości wielopierścieniowych węglowodorów aromatycznych. Badania przeprowadzone dla wielu miast w Polsce dowodzą, że stężenia tego związku znacznie przewyższają dopuszczalne stężenia. Dla Wadowic [Furman i in. 2021] i badań przeprowadzonych w 2017r. średnie stężenie Benzo[a]pirenu w okresie grzewczym wynosiło 4,98 ng/m^3, a w okresie poza grzewczym było równe 1,10 ng/m^3. Podobnych obserwacji dokonano w innych miastach w Polsce, gdzie dominuje transport jako główne źródło zanieczyszczeń [Kozielska i in. 2016]. Wysokie średnie dzienne stężenia Benzo[a]pirenu zaobserwowano w badaniach [Kozielska i in. 2013] przeprowadzonych na terenie Górnego Śląska, gdzie na ruchliwych ulicach o wysokim natężeniu ruchu w godzinach szczytu odnotowano zakres stężeń B[a]P od 7,90 ng/m^3 do 11,10 ng/m^3. Niestety duże zagrożenie płynące z obecności Benzo[a]pirenu występuje także w pomieszczeniach zamkniętych. Prace [Rogula-Kozłowska i in. 2017] przeprowadzone w Warszawie i Gliwicach wykazały znaczne stężenia tego związku wewnątrz pomieszczeń. Średnie wartości stężeń B[a]P wynosiły odpowiednio 1,11 ng/m^3 i 3,27 ng/m^3. Na mapie (rysunek 4) przedstawiono rozkład średniego rocznego stężenia Benzo[a]pirenu w Europie w 2015r.



Rysunek 4. Stężenia średnie roczne Benzo[a]pirenu w Europie w 2015r [European Environment Agency 2017]



Rysunek 5. Stężenia średnie roczne Benzo[a]pirenu w latach 2004-2016 w wybranych miastach w Polsce [Główny Inspektorat Ochrony Środowiska 2017]

W Polsce stężenia B[a]P utrzymują się stale na wysokim poziomie, znacznie przekraczając wartości dopuszczalne (rysunek 5). Powszechność występowania wielopierścieniowych węglowodorów aromatycznych, a przede wszystkim ich wpływ na organizm człowieka, sprawiają, że dąży się do unowocześniania dostępnych metod ograniczania ich wpływu na środowisko.

1.3. Wskaźniki narażenia na WWA

Z punktu widzenia zdrowia ludzkiego ważnym jest oszacowanie przybliżonego stopnia narażenia na wielopierścieniowe węglowodory aromatyczne. Związki te zaliczane są do grupy jednych z najbardziej niebezpiecznych substancji znajdujących się w powietrzu. Jest to spowodowane ich właściwościami kancerogennymi i mutagennymi. Najniebezpieczniejsze są węglowodory posiadające w swojej budowie od 4 do 6 pierścieni aromatycznych. Z tego powodu tworzone są szerokie profile uwzględniające procentowy udział cięższych WWA w pyle zawieszonym [Furman i in. 2021, Holtzer i Grabowska 2010, Skiba i in. 2019]. Kolejnym sposobem przedstawienia zagrożenia płynącego ze słabej jakości powietrza jest profil odzwierciedlający zawartość najbardziej kancerogennych WWA w stosunku do wszystkich zidentyfikowanych WWA. Do grupy WWA o najsilniejszych właściwościach kancerogennych zaliczamy: Benzo[a]antracen, Benzo[a]piren, Benzo[b]fluoranten, Benzo[k]fluoranten, Indeno[*1,2,3-cd*]piren, Dibenzo[ah]antracen i Chryzen [Kozielska i in. 2016]. Narażenie na WWA określa się z wykorzystaniem następujących wskaźników ekwiwalentu toksyczności wielopierścieniowych węglowodorów aromatycznych zalecanych przez EPA: ekwiwalentu działania mutagennego względem B[a]P (*MEQ*), ekwiwalentu działania toksycznego względem B[a]P (*TEQ*) i ekwiwalentu działania kancerogennego względem 2,3,7,8-tetrachlorodienzo-p-dioksyny (*CEQ*). Wartości tych wskaźników szacowane są przy użyciu wzorów z konkretnymi współczynnikami przypisanymi do danego związku, które biorą pod uwagę ich siłę działania kancerogennego i mutagennego. Poniżej wzory dla poszczególnych wskaźników (w nawiasach przedstawione są odpowiednie stężenia WWA):

Równanie 1. Równoważnik mutagenności MEQ [Rogula-Kozłowska i in. 2013]

$$MEQ = 0{,}00056 * [Acy] + 0{,}082 * [BaA] + 0{,}017 * [Ch] + 0{,}25 * [BbF] + 0{,}11 * [BkF] + 1 * [BaP] + 0{,}31 * [IP] + 0{,}29 * [DBA] + 0{,}19[BghiP]$$

Równanie 2. Równoważnik kancerogenności CEQ [Rogula-Kozłowska i in. 2013]

$$CEQ = 0{,}001 * ([Acy] + [Ace] + [Flu] + [Fen] + [Fl] + [Pir]) + 0{,}01 * ([An] + [Ch] + BghiP]) + 0{,}1 * ([BaA] + BbF + [BkF] + [IP]) + 1 * [BaP] + 5 * [DBA]$$

Równanie 3. Ekwiwalent toksycznego działania TEQ [Rogula-Kozłowska i in. 2013]

$$TEQ = 0{,}000025 * [BaA] + 0{,}0002 * [Ch] + 0{,}000354 * [BaP] + 0{,}0011 * [IP] + 0{,}00203 * [DBA] + 0{,}00253 * [BbF] + 0{,}00487 * [BkF]$$

Wartości współczynników w równaniach 1-3 pochodzą z Nisbet i LaGoy 1992, Durant i in. 1996 oraz Willett i in. 1997. Największy wpływ na wskaźnik mutagenności *MEQ* posiada Benzo[a]piren, dla *CEQ* jest to Dibenzo[ah]antracen i Benzo[a]piren, dla *TEQ* to Benzo[k]fluoranten. Jest to jeden z dodatkowych powodów uwzględnienia głównie B[a]P w przewidywaniu wskaźnika narażenia społeczności z wykorzystaniem wskaźników narażenia.

1.4. Ocena i prognozowanie jakości powietrza

Narzędziem, który w znacznym stopniu wspomaga ocenę jakości powietrza może być zautomatyzowany model predykcji wskaźnika średniego narażenia AEI (*Average Exposure Indicator*). Tego typu model powinien uwzględniać, obok stężenia pyłów zawieszonych i parametrów meteorologicznych (temperatura powietrza, prędkość i kierunek wiatru itp.), także stężenia groźnych dla zdrowia związków, takich jak wielopierścieniowe węglowodory aromatyczne. Modele tego typu są w dużym stopniu rozpowszechnione, jednak ich rozbudowa o dodatkowe parametry wejściowe, takie jak informacje o stężeniach WWA, ułatwi określenie stopnia narażenia społeczności na zanieczyszczenia powietrza. Stworzenie takich systemów

predykcji jest zadaniem trudnym. Wszystkie procesy zachodzące w atmosferze przebiegają pod wpływem wielu czynników i uwarunkowań (przemiany fizykochemiczne związków, wymywanie w wyniku opadów atmosferycznych, reakcje chemiczne z ich udziałem itp.). Jednak dostęp do archiwalnych baz danych dla stężeń pyłu zawieszonego i warunków meteorologicznych pozwala na wstępne oszacowanie stanu zagrożenia płynącego ze złej jakości powietrza. Należy jednak pamiętać, że procesy kształtujące warunki powietrza i stężenia pyłów przebiegają dynamicznie, natomiast ich wzajemne oddziaływania są skomplikowane i nie posiadają charakteru liniowego – z tego względu niezbędne w systemach są uproszczenia oraz zastosowanie schematów predykcji nieliniowych. Pomimo nieustannie rozwijających się zastosowań modeli deterministycznych w dziedzinie jakości powietrza, warto zwrócić uwagę na nowy, obiecujący i również coraz częściej wykorzystywany kierunek rozwoju – algorytmy uczenia maszynowego [Poloczek i in. 2021, Skrzypski i Jach-Szakiel 2008].

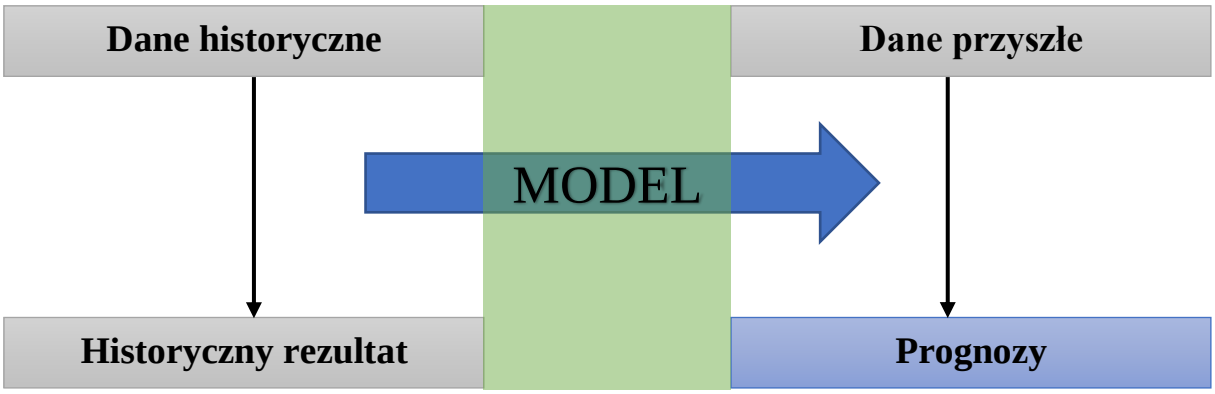
1.5. Uczenie maszynowe

Uczenie maszynowe to dział informatyki zajmujący się tworzeniem systemów komputerowych, które zamiast sztywno zakodowanych reguł, dostarczają wnioski oparte o analizę danych – takie systemy samodzielnie się uczą przy ograniczonej lub bez ingerencji człowieka. Określenie uczenia maszynowego bardzo często stosowane jest zamiennie z *Data Science*, czyli nauką o danych, które jest znacznie szerszym pojęciem, w którego skład wchodzą: algorytmika, automatyzacja procesów i decyzji, wizualizacja danych, metodyka pracy, integracja itd. Wszystko to jest zawarte w dodatkowym pojęciu – sztucznej inteligencji. Sztuczna inteligencja jest określeniem silnie interdyscyplinarnym, w którego skład wchodzi fizyka, chemia, medycyna, socjologia, psychologia, filozofia, psychologia i wiele innych dziedzin nauki.

Za początek uczenia maszynowego można przyjąć rok 1950, w którym to Alan Turing publikuje swój słynny Test Turinga [Łupkowski 2010] – test mający za zadanie weryfikację płynności granicy między sztuczną, a naturalną inteligencją. Test polegał na prowadzeniu konwersacji przez sędziego (człowieka) z żywą osobą oraz maszyną. W momencie, w którym sędzia nie jest w stanie wiarygodnie ocenić z kim rozmawia, wtedy test określa się za zaliczony – maszyna wygrywa. W 1957 pojawia się Perceptron, czyli pierwsze częściowo mechaniczne urządzenie, którego zasada działania opierała się na odzwierciedlaniu pracy mózgu człowieka. Jest to podstawa do dalszego rozwoju sieci neuronowych. W latach 1952-1962 stale rozwijał się również projekt Arthura Samuela z firmy IBM, opierający się na szkolnictwie zawodników szachowych przy użyciu programu wykorzystującego uczenie maszynowe. Rozwój sieci neuronowych intensyfikuje się w latach 80. Powstaje wtedy program AM (Automated Mathematician) autorstwa Douglasa Lenata, służący do wyszukiwania nowych praw matematycznych [Henderson 2007]. Lata 90 to znaczny rozwój nowych algorytmów, takich jak lasy losowe, maszyna wektorów nośnych. Prace Geralda Tesauro [Tesauro 1994] przyczyniają się do powstania programu TD-Gammon, który potrafi konkurować z mistrzami świata w grze Backgammon. Metoda opierała się na nauce strategii z ponad miliona gier z żywym przeciwnikiem. W 1997 roku Deep Blue, czyli program stworzony przez IBM wygrywa partię szachów z wielokrotnym mistrzem świata – Garriem Kasparowem. Deep Blue niezwykle dokładnie przeanalizował i opracował wszystkie dotychczasowe rozgrywki Kasparowa tworząc w ten sposób odpowiedni algorytm programu [ChessBasse 2013]. Lata 90 to również znaczny wzrost wykorzystania uczenia maszynowego w rozwoju Internetu, a dokładniej wyszukiwarek internetowych, takich jak Google, Bing czy Yahoo. W 2006 roku wiele zespołów naukowych

zostało zachęconych do ulepszenia algorytmu rekomendacyjnego Netflixa. W tym samym czasie program szachowy Fritz 10 pokonuje aktualnego wtedy mistrza świata w szachach, Władimira Kramnika. W 2010 roku następuje premiera platformy Kaggle bazującej na uczeniu maszynowym, która stanowiła jedno z podstawowych źródeł informacji dla środowiska *Data Science*. Rok później to okres narodzin Siri – pierwszy z wielu osobistych asystentów posługujących się sztuczną inteligencją [Kaplan i Haenlein 2019]. 2014 rok to czas, w którym modele rozpoznawające twarze przewyższają dokładnością możliwości człowieka. Już rok później, prawie identyczny model pokonuje człowieka w kwestii rozpoznawania obiektów znajdujących się na zdjęciach. W 2016 następuje wygrana programu AlphaGo, który wygrał z człowiekiem w grze Go – najbardziej skomplikowana, klasyczna gra planszowa [Zylinski 2018]. W tym samym roku Skype wykorzystuje uczenie maszynowe do tłumaczenia niektórych języków w czasie rzeczywistym [Osowski 2000, Zylinski 2018]. Dział uczenia maszynowego ma coraz większy wpływ na wiele dziedzin gospodarki i naszego życia codziennego – przemysł, handel, finanse, medycyna, rozrywka. Wiele nowych rozwiązań technologicznych bazuje na wykorzystaniu algorytmów uczenia maszynowego. Zastosowanie ich w sektorze środowiskowym – jakości powietrza – w dużym stopniu przyczyni się do pozyskania coraz to dokładniejszych modeli przewidywań w zakresie stopnia narażenia społeczności na zanieczyszczenia powietrza.

Uczenie maszynowe dzieli się na 2 główne typy: nadzorowane i nienadzorowane. Ogólny schemat pierwszego z nich przedstawiono na rysunku 6:



Rysunek 6. Ogólny schemat uczenia nadzorowanego

Nadzorowane modele uczenia składają się z odpowiednio przygotowanego zbioru informacji - dane historyczne, historyczny rezultat, przyszłe dane. Zadaniem modelu jest taka interpretacja danych historycznych, aby rezultaty zgadzały się z odpowiadającymi im historycznymi rezultatami, tak aby w przyszłości (w oparciu o nowe dane) model mógł wygenerować poprawną prognozę. Górna część powyższego schematu odpowiada danym wejściowym opisującym rzeczywistość, dolna na oczekiwanych wynikach modelu. Nienadzorowane modele uczenia opierają się wyłącznie na danych historycznych. Brak tutaj zmiennej zależnej – historycznego rezultatu. Metody uczenia nienadzorowanego wykorzystują algorytmy samorzutnie wyszukujące zależności danych wejściowych. Przykładem metody uczenia nienadzorowanego jest analiza skupień, czyli grupowanie składników danych w oparciu o zaobserwowane wzorce. Często spotykanym schematem jest również zastosowanie uczenia nienadzorowanego przed modelem wykorzystującym uczenie nadzorowane. Taki układ stosuje się w sytuacjach, gdzie do dyspozycji jest bardzo duża ilość danych lub gdy informacje/zależności między danymi są na tyle zróżnicowane, że niełatwo jest przyporządkować odpowiedzi do każdej z nich. Model sam analizuje dane wejściowe,

wyszukuje wzorce i zależności, a następnie proponuje odpowiedzi, tworząc ogólne wzorce wykorzystywane w kolejnych etapach uczenia. Przykładem zastosowania tego typu uczenia są systemy rozpoznawania mowy i obrazu [Osowski 2000, Poloczek i in. 2021].

Analizy prognostyczne stanowią jeden z najbardziej rozpowszechnionych obszarów zastosowania algorytmów uczenia maszynowego. Modele tego typu potrafią wyszukać nietrywialne zależności między składowymi danych i stworzyć prognozy dokładniejsze niż mógłby to zrobić człowiek. Do najbardziej znanych algorytmów w tym przypadku należy regresja liniowa, a także coraz częściej stosowane algorytmy sztucznych sieci neuronowych. Przykładem zastosowań analiz prognostycznych są przewidywania dotyczące łańcuchów dostaw w przemyśle oraz wskazanie obszarów tych łańcuchów, gdzie odpowiednie algorytmy uczące mają szczególne znaczenie [Maternowska 2019]. Sztuczne sieci neuronowe z powodzeniem znajdują zastosowanie w badaniach dotyczących predykcji zanieczyszczeń powietrza. Model bazując na historycznych danych dotyczących stężenia pyłu oraz na warunkach meteo, jest w stanie z wysoką dokładnością, z uwzględnieniem prognozy dnia następnego, przewidzieć wartości stężeń konkretnych zanieczyszczeń [Wilkosz i in. 2021] Kolejnym ważnym zastosowaniem jest wyszukiwanie i śledzenie wzorców, a co za tym idzie, wyszukiwanie odchyleń (anomalii), które mogą być bardzo ważne np.: z punktu widzenia biznesowego – wykrywanie fałszerstw podatkowych, pranie brudnych pieniędzy, czy identyfikacja ataków hackerskich [Sobczyk i in. 2019]. W tym przypadku zastosowanie znalazły metody takie jak klasyfikacja lub analiza skupień. W analizach środowiskowych, analiza skupień pozwala znaleźć zależności pomiędzy konkretnymi parametrami modelu. W badaniach jakości powietrza dla Krakowa [Szlachtowska i in. 2016], stosując algorytmy analizy skupień, wykazano korelacje pomiędzy stężeniami dwutlenku azotu oraz porą dnia. Następnym zastosowaniem uczenia maszynowego są silniki rekomendacyjne, najczęściej spotykane w życiu codziennym [Rutkowski 2020]. Wszystkie komercyjne działalności bazują na zdobyciu oraz utrzymaniu zaufania klienta. Najnowsze algorytmy pozwalają stworzyć spersonalizowaną ofertę pod konkretną osobę począwszy od odpowiedniego doboru produktu, przez wskazówki odnośnie preferencji muzycznych, po selekcję artykułów i reklam na stronach internetowych [Osowski 2000].

Algorytm regresji liniowej dzięki swojej prostocie, stanowi bazę dla scenariuszy regresyjnych. Jest mało wymagający obliczeniowo, znakomicie sprawdza się przy dużych zbiorach danych. Dzięki swojej prostocie łatwo da się interpretować uzyskane prognozy, wykazując podstawowe zależności pomiędzy danymi wejściowymi, a wyjściowymi. Ideą tego algorytmu jest liniowa kombinacja zmiennych i parametrów tak aby w najlepszy możliwy sposób dopasować model do wszystkich dostępnych danych. Stworzona w ten sposób linia/prosta regresji przedstawia obliczoną/prognozowaną wartość konkretnej zmiennej z uwzględnieniem danych wartości innej zmiennej/zmiennych – ukazany zostaje trend dostępnych danych. Najprostszy wzór takiej zależności przedstawia równanie 4:

Równanie 4. Równanie prostej regresji liniowej [Tomski 2022]

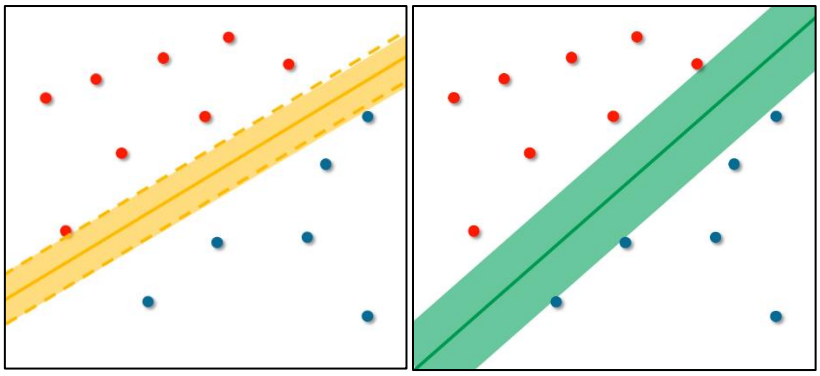
$$y = ax + b$$

gdzie: *y* określane jest zmienną objaśnianą lub zmienną zależną, *x* to zmienna objaśniająca lub niezależna, a – współczynnik kierunkowy, który decyduje o kącie nachylenia otrzymanej prostej względem osi *x* w układzie *x* i *y*, b – wyraz wolny określający punkt przecięcia prostej z osią *y* w układzie *x* i *y*. Przedstawiony przykład dotyczy wyłącznie jednej zmiennej. Z perspektywy modelu, im większa wartość bezwzględna współczynnika kierunkowego, tym większy wpływ zmiennych niezależnych na zmienne zależne. W realnych scenariuszach

niemożliwym jest wyznaczenie takiej prostej regresji liniowej, aby odpowiadała ona wszystkim danym wejściowym. Zadaniem algorytmu jest więc zminimalizowanie funkcji błędu. Jedną z propozycji rozwiązania tego problemu jest zastosowanie metody najmniejszych kwadratów. Algorytm regresji liniowej – jak sama nazwa wskazuje – już na wstępie zakłada, że zależność pomiędzy elementami zestawu danych wejściowych i wyjściowych jest liniowa. W przypadku nieliniowej zależności dużo lepszym rozwiązaniem jest zastosowanie innych algorytmów – sztucznych sieci neuronowych, lasów losowych. Modele regresji liniowej mogą być stosowane również dla zmiennych kategorycznych [Osowski 2000].

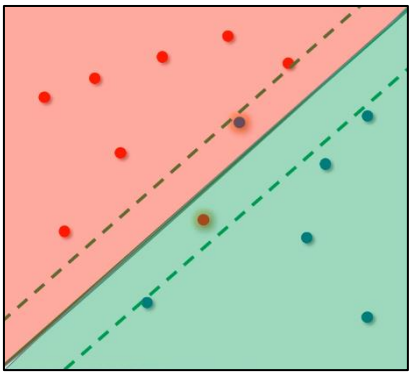
Drugim, dobrze znanym algorytmem jest regresja logistyczna będąca rozwinięciem koncepcji regresji liniowej. W najczęściej spotykanym przykładzie, regresja logistyczna służy jako klasyfikator binarny – metoda opierająca się na klasyfikowaniu danych do jednej z dwóch klas z odpowiednim prawdopodobieństwem. Regresja logistyczna, w porównaniu do regresji liniowej, oparta jest o funkcję logistyczną, odpowiedzialną za zmianę dowolnej wartości liczbowej na liczbę z przedziału 0 – 1. Dla tego algorytmu zmienna objaśniana nie informuje nas o występowaniu lub braku występowania danej predykcji lecz pozwala obliczyć prawdopodobieństwo wystąpienia danego zdarzenia. Regresja logistyczna to algorytm działający porównywalnie szybko do regresji liniowej. Znakomicie sobie radzi z dużymi zbiorami danych. Otrzymane prognozy modelu opierającego się na tym algorytmie w łatwy sposób można zinterpretować. Funkcje logistyczne wykorzystywane są do danych zawierających kategorie. W tym celu stosuje się jedną z odmian algorytmu – wielomianową/wieloklasową regresję logistyczną [Tomski 2022].

Do jednych z najbardziej rozwiniętych i najczęściej stosowanych algorytmów klasyfikujących jest algorytm wektorów nośnych (maszyna wektorów nośnych – SVM). Najczęściej stosowany jest przy kategoryzacji tekstu lub klasyfikowaniu obrazów. Podstawowy algorytm bazuje na kategoryzacji do 2 klas. Rozwinięciem w przypadku większej liczby klas stanowi zastosowanie dodatkowego modułu w modelu (strategia jeden przeciw wszystkim) lub odmiany algorytmu – Support Vector Clustering (SVC). W najprostszym przykładzie (rysunek 7) (łatwo identyfikowalnych klas – czerwone i niebieskie kropki odpowiadają konkretnej obserwacji), algorytm SVM tworzy w układzie danych prostą pozwalającą na zdefiniowanie podziału pomiędzy zbiorami oraz margines (obszar żółty/zielony na rysunku 7), stworzony na styku pierwszych punktów każdej z klas. Dla żółtej/zielonej linii margines z jednej strony styka się z pierwszą obserwacją niebieską, z drugiej strony z pierwszą obserwacją czerwoną (rysunek 7). Szerokość powstałego marginesu wynika z rozkładu danych i determinuje, czy model z łatwością będzie się uczył.



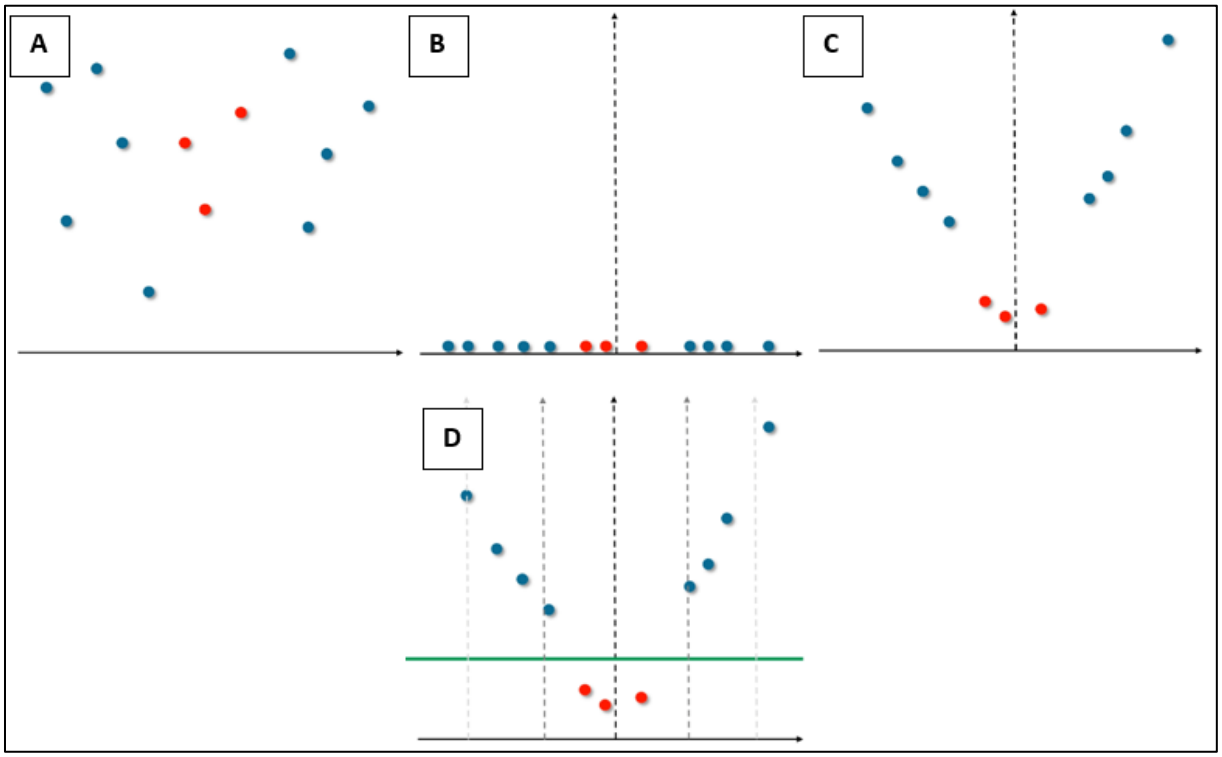
Rysunek 7. Przykład marginesu (kolor żółty/zielony) dla pierwszych zmiennych każdej klasy algorytmu SVM [Zylinski 2018]

W przypadku złożoności danych – trudno identyfikowalnych klas (rysunek 8)– algorytm SVM radzi sobie poprzez zastosowanie, zamiast sztywnej granicy, opcji „płynnych marginesów” (linia przerywana – rysunek 8). Dopuszczamy w ten sposób możliwość wymieszania danych, które nie należą do danej klasy, tylko do klasy przeciwnej. Zmniejszone zostaje w ten sposób ryzyko nadmiernego dopasowania.



Rysunek 8. Przykład "płynnych marginesów" (linia przerywana) algorytmu SVM [Zylinski 2018]

Dla danych wejściowych, dla których niemożliwa jest liniowa separacja (rysunek 9A) stosuje się transformację danych – wprowadzenie dodatkowych wymiarów/zmiennych. Przykładem tego może być rzutowanie całego układu danych (w układzie *x* i *y*) na jedną oś zmiennych (rysunek 9B) oraz zastosowanie funkcji kwadratowej (rysunek 9C), co pozwoli w łatwy sposób odseparować klasy danych. Cały proces również jest zautomatyzowany za pomocą funkcji jądrowych (kernel functions) (rysunek 9D).



Rysunek 9. Przykład zastosowania algorytmu SVM bez liniowej separacji danych (czerwone i niebieskie kropki odpowiadają obserwacją) [Zylinski 2018]

Drzewa decyzyjne to kolejny rodzaj algorytmu często wykorzystywanego w modelach uczenia maszynowego. W modelach uczenia maszynowego drzewo zostaje w sposób automatyczny zbudowane bazując na zbiorach danych treningowych. Najczęściej stosuje się drzewa binarne – wybór 1 lub 0. Algorytm powstaje poprzez wybranie konkretnej cechy oraz jej wartości, która daje największy zysk informacyjny w całym zestawie danych wejściowych. Dla przykładu: jeśli dopuszczalne średnie dobowe stężenie pyłu wynosi 50 $\mu g/m^3$, a wszystkie niższe wartości określane będą klasą „*niskie zagrożenie*", a wyższe wartości klasą „*wysokie zagrożenie*", to w łatwy sposób można uzyskać informację dla każdej wartości – idealna separacja klas, wysoki zysk informacyjny. Gotowe modele/programy szybko dostarczają informacje/prognozy. Dużą zaletą tego algorytmu jest możliwość wykorzystania ich w zbiorze danych z nieliniową zależnością. Do wad zalicza się wysokie prawdopodobieństwo nadmiernego dopasowania oraz wysoką wrażliwość na wszelkie zmiany w zbiorze treningowym – nawet niewielka zmiana parametrów wejściowych w znaczny sposób może wpłynąć na dokładność modelu. Odmianą algorytmu drzew decyzyjnych są lasy losowe. Modele te składają się z wielu poziomów drzew, a każde z nich oparte jest o część zbioru testowego i część predyktorów. Podobnie jak dla drzew decyzyjnych, tak i tutaj można zastosować lasy losowe w przypadku klasyfikacji i regresji danych. Model lasów losowych generuje klasę będącą dominantą klas (dla danych klasyfikowanych) lub średnią poszczególnych drzew (dla danych regresyjnych). Zastosowanie lasów losowych zwiększa możliwości obliczeniowe modelu, przez co model szybciej wykonuje obliczenia.

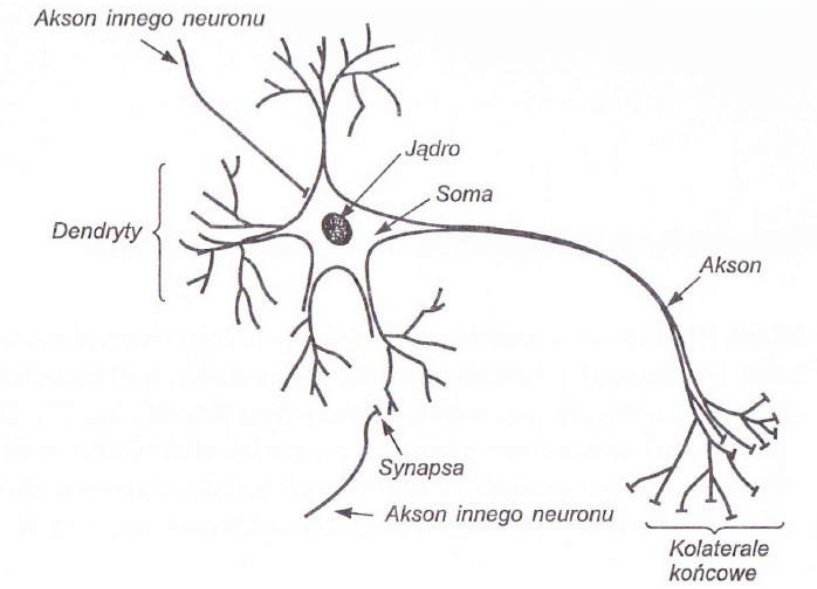
1.6. Sztuczne sieci neuronowe

Algorytm sztucznych sieci neuronowych znajduje zastosowanie w wielu aspektach badań jakości powietrza. Badania [Makar i in. 2022] dotyczące PM_{10} pozwoliły stworzyć wysokiej dokładności model predykcji smogu. Zależność danych rzeczywistych i prognozowanych była równa 0,99. Algorytmy te wykorzystano również do identyfikacji mechanizmów dominujących w rozprzestrzenianiu zanieczyszczeń [Rogula i Żeliński 2004] oraz do modelowania całkowitej emisji związków z konkretnego źródła [Niżewski i in. 2007]. W wielu przypadkach, sztuczne sieci neuronowe stosuje się do przewidywania stężeń pyłów zawieszonych, prognozowania klas stanu jakości powietrza w wysoką dokładnością [Łozowicka-Stupnicka i Talarczyk 2005, Skrzypski i Jach-Szakiel 2008, Suleiman i in. 2019]. Tematyka sztucznych sieci neuronowych to interdyscyplinarny obszar nauki powiązany z wieloma dziedzinami, począwszy od informatyki, przez medycynę, automatykę, elektronikę, matematykę, kończąc na socjologii, filozofii i psychologii [Maternowska 2019, Matuszko i Piotrowicz 2015, Rutkowski 2020, Sobczyk i in. 2019]. Podstawą wiedzy o sieciach neuronowych jest żywy organizm, a dokładniej wszelkie procesy zachodzące podczas pracy mózgu. Procesy te stanowią fundament w poszukiwaniu nowatorskich rozwiązań technologicznych ułatwiających codzienne życie oraz przyśpieszających rozwój cywilizacyjny. Przeprowadzone badania w niniejszej pracy wykazały wysoką dokładność modeli opartych na sztucznych sieciach neuronowych w porównaniu do wcześniej opisanych algorytmów.

1.6.1. Podstawy biologiczne działania sztucznych sieci neuronowych

Podstawowym budulcem systemu nerwowego organizmów są komórki nerwowe – zwane neuronami. Wiedza na temat wzajemnych relacji, współistnienia oraz pracy neuronów jest niezbędna do pozyskania informacji dotyczących otrzymywania, przesyłu i przetwarzania danych zachodzących w sieci. Neuron (rysunek 10) zbudowany jest z jądra osadzonego w somie, z której wyrastają liczne wypustki. Wypustki te odpowiadają za połączenia z innymi

elementami sieci. Informacje docierają i wychodzą z komórki za pomocą, odpowiednio synaps i aksonu. Wyjściowe informacje przechodzą przez układ kolaterali trafiając do kolejnym som i dendrytów znajdujących się w innych neuronach, tworząc w ten sposób kolejne synapsy. Przepływ informacji pomiędzy neuronami to proces chemiczno-elektryczny. Podczas takiego procesu przekazywanie danych następuje w wyniku zadziałania na synapsy bodźcem zewnętrznym, wydzieleniu odpowiednich substancji chemicznych (neuromediatory) i zmiany potencjału elektrycznego błony komórkowej. W ciele komórki wszystkie impulsy elektryczne są sumowane i następuje generacja sygnału wyjściowego, który dociera do kolejnej komórki. Szacuje się, że ludzki mózg posiada około 100 miliardów neuronów. Tak duża liczba komórek powoduje, że pojawiający się pojedynczy błąd w przesyle informacji ginie w tle wszystkich procesów. Jest to główna cecha układu nerwowego sprawiająca, że sieć jest odporna na zakłócenia. Kolejną ważną cechą jest szybkość przesyłania danych – ich przetwarzanie odbywa się w sposób równoległy przez dużą ilość połączeń międzyneuronowych, co daje znaczną przewagę w porównaniu do standardowych systemów elektronicznych. Dzięki takiemu układowi sieci, operacje zachodzące w mózgu (rozpoznawanie obrazów, dźwięków, podejmowanie decyzji, myślenie, zapamiętywanie itp.) odbywają się w ciągu milisekund. W obecnych czasach, nawet przy dostępie najnowocześniejszej technologii, nie jest możliwe stworzenie sztucznej sieci neuronowej idealnie odpowiadającej sieci neuronów znajdujących się w mózgu [Osowski 2000, Poloczek i in. 2021, Tadeusiewicz i Szaleniec 2015].



Rysunek 10. Uproszczony model rzeczywisty komórki nerwowej [Osowski 2000]

1.6.2. Pierwsze modele sieci neuronowej

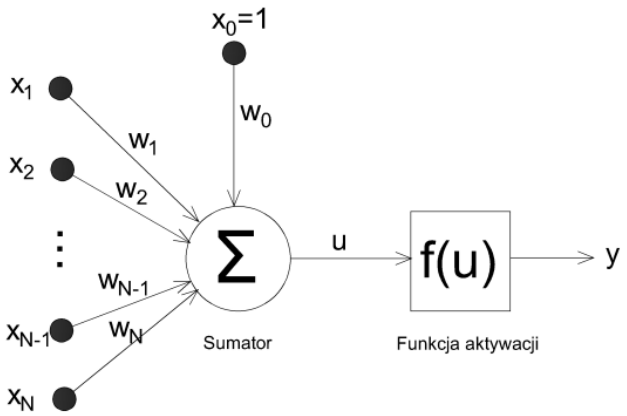
Pojęcie sztucznych sieci neuronowych to określenie wszelkich matematycznych struktur i ich programowych/sprzętowych modeli odpowiedzialnych za przetwarzanie lub obliczenia sygnałów za pomocą rzędów elementów przetwarzających informacje (sztucznych neuronów). Pod tym pojęciem kryje się także termin sztucznej inteligencji – „*zdolności systemu do prawidłowego interpretowania danych pochodzących z zewnętrznych źródeł, nauki na ich podstawie oraz wykorzystania tej wiedzy, aby wykonywać określone zadania i osiągnąć cele poprzez elastyczne dostosowanie*"[Kaplan i Haenlein 2019]. Pierwsze modele opracowane przez McCullocha-Pittsa w 1943r opierały się na postaci binarnej neuronu. Zakładano, że sygnał wyjściowy neuronu przyjmuje wartość 0 lub 1. Sygnały wejściowe x_j (j = 1,2, …, N)

sumowane są z przypisanymi im wagami synaptycznymi w_{ij} (gdzie i i j to kierunek przepływu informacji od węzła i do węzła j). Wyjściowe sygnały określane były równaniem 5:

Równanie 5. Zasada działania modelu matematycznego McCullocha-Pittsa [Osowski 2000]

$$y_i = f\left(\sum_{j=1}^{N} w_{ij}\, x_j(t) + w_{i0}\right)$$

Zmienną powyższej funkcji jest sygnał sumacyjny x_i. Gdy $w_{ij} > 0$ to synapsa jest synapsą pobudzającą, wartość ujemna świadczy o obecności synapsy hamującej, a wartość 0 wskazuje na brak połączenia między węzłami i i j neuronu. [Osowski 2000, Wilkosz i in. 2021]. Poniżej przestawiono matematyczny model neuronu McCullocha-Pittsa, w którym x_0 stanowi próg aktywacji:



Rysunek 11. Matematyczny model neuronu według McCullocha-Pittsa [Osowski 2000]

W 1949r światło dzienne ujrzała publikacja Donalda O. Hebba „*Organizacja zachowań. Teoria neuropsychologii*". Autor połączył dotychczasową wiedzę na temat inżynierii z pracą mózgu. Pomimo dość znikomej wiedzy na temat procesów zachodzących w sieci neuronowej, nie przeszkadzało to Hebbowi na stworzenie teorii „*Hebbowskiego uczenia się*". Teoria wywodzi się z pamięci asocjacyjnych – wytwarzania całości informacji z niewielkiego fragmentu pamięci danych. Zakłada przypisanie odpowiednich wag w_{ij} neutronu, gdzie wzrost wartości Δw_{ij} w modelu uczącym się w określonej liczbie cykli k jest proporcjonalny do iloczynu sygnałów wyjściowych y_i i y_j z uwzględnieniem współczynnika uczenia η:

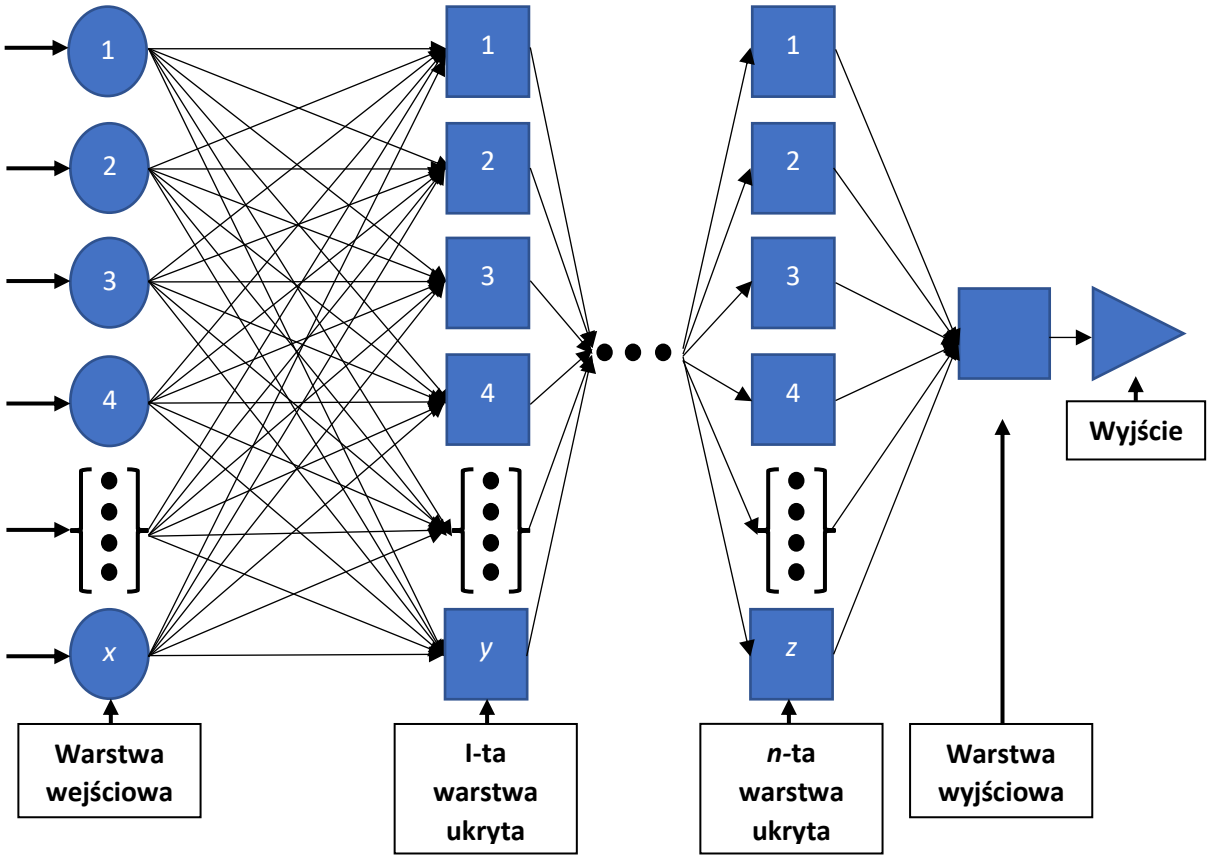
Równanie 6. Matematyczny opis modelu neuronu według Hebba [Osowski 2000]

$$w_{ij}(k+1) = w_{ij}(k) + \eta y_i(k) y_j(k)$$

Lata 60 były kolejnym okresem rozwoju sieci neuronowych. Bernard Widrow przyczynił się intensyfikacji dziedziny elektrotechniki poprzez implementacje techniczne adaptacyjnych układów odpowiedzialnych za przetwarzanie sygnałów. B. Widrow był głównym pomysłodawcą podstaw teoretycznych pracy sieci neuronowych [Widrow i Hoff 1960]. Mniej więcej w tym samym czasie została opublikowana książka autorstwa Franka Rosenblatta „*Principle of neurodynamics*" [Rosenblatt 1992]. Autor bazując na modelu McCullocha-Pittsa (wartościach binarnych funkcji), przedstawił pracę układu neuronowego jako model perceptronowy. Pomysł ten został skrytykowany, co znacznie spowolniło rozwój dziedziny badań nad sieciami neuronowymi. Dopiero lata 80 i pojawienie się układów wielkiej skali integracji (VLSI) zachęciły naukowców do ponownych badań nad tematyką sztucznej

inteligencji. Spore nakłady finansowe przeznaczone na badania nie tylko pozwoliły na znaczne poszerzenie wiedzy teoretycznej, ale również umożliwiły zainteresowanie i dynamiczny rozwój aplikacji (neurokomputery) [Osowski 2000, Poloczek i in. 2021].

Obecnie, do najbardziej rozpowszechnionych typów sztucznej sieci neuronowej zaliczany jest perceptron – zbudowany najczęściej z pojedynczej warstwy wejściowej, kilku warstw ukrytych i pojedynczej warstwy wyjściowej. Rysunek 12 przedstawia schemat sieci zbudowany z warstwy wejściowej składającej się z *x* neuronów wejściowych, *I* do *n* warstw ukrytych zbudowanych, odpowiednio z *y* i *z* neuronów. Schemat zamyka warstwa wyjściowa zbudowana, w tym przypadku, z 1 neuronu.



Rysunek 12. Przykładowy schemat modelu sieci neuronowej [Kobylec 2003]

Z racji budowy takiego układu, perceptron często określany jest jako sieć w pełni połączona, wszystkie wejścia kolejnej warstwy połączone są z wyjściami warstwy poprzedniej. Wadą takiego modelu jest konieczność dobrania odpowiedniej liczby warstw ukrytych, a także liczby neutronów stanowiących podstawę każdej warstwy. Jednym ze sposobów uczenia modelu sieci neuronowej jest technika uczenia nadzorowanego, gdzie proces optymalizacji polega na odpowiednim doborze parametrów oraz danych wejściowych i odpowiadającym im rezultatów. Jakość modelu sieci neuronowej wyznaczana jest poprzez porównanie uzyskanych rezultatów na wyjściu sieci z oczekiwanymi danymi wyjściowymi [Poloczek i in. 2021].

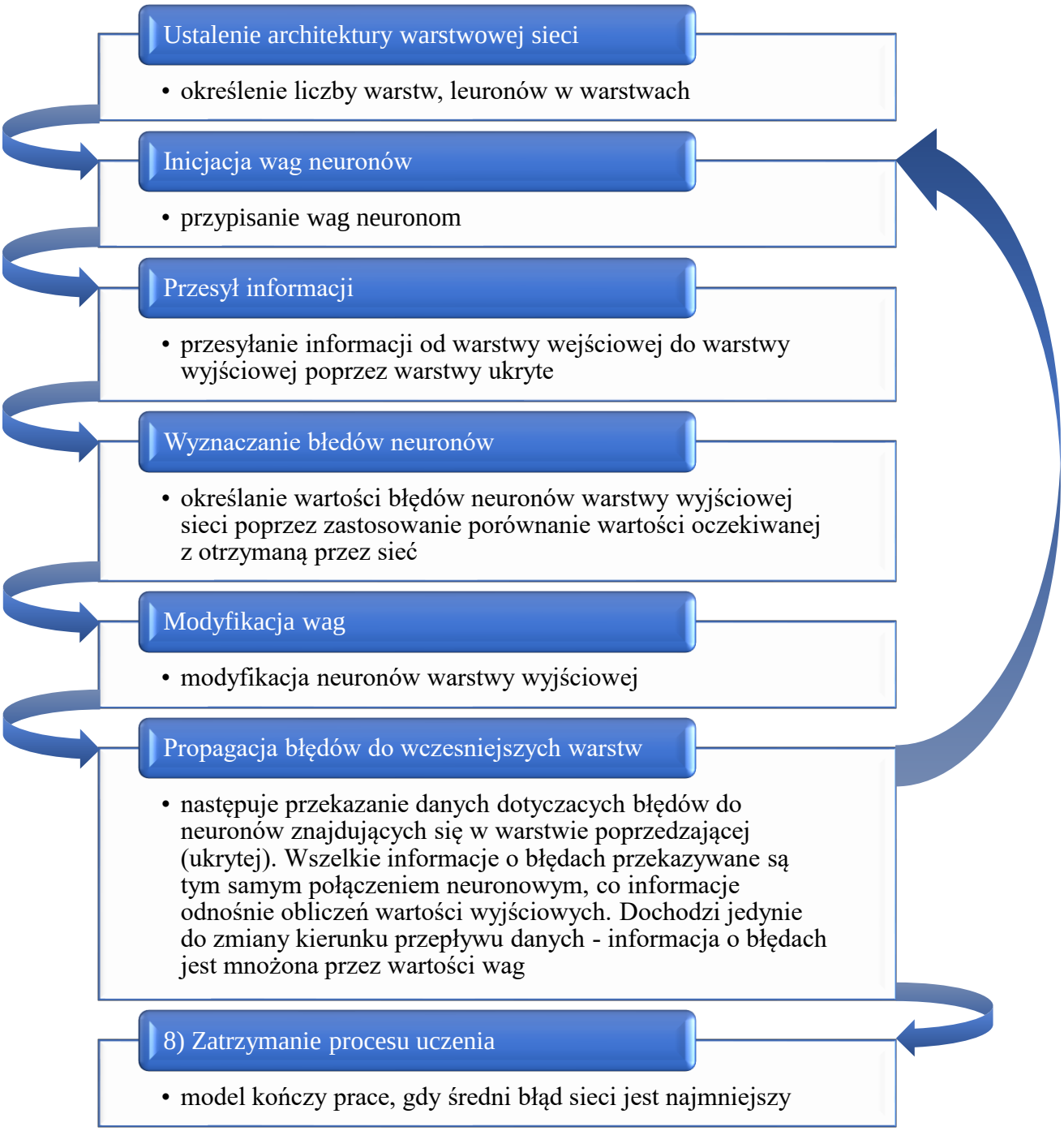
Parametrem neuronu, który decyduje o jego właściwościach i roli pełnionej w procesie uczenia, są jego wagi w_i. Z jednej strony wagi wyrażają stopień ważności danych przekazywanych pomiędzy jednym, a drugim neuronem sieci, z drugiej strony przedstawiają pamięć neuronu – zapamiętują związki/relacje zachodzące między sygnałami wejściowymi i wyjściowymi. Na dokładność modelu duży wpływ ma zakres wstępnych wartości wag sieci, z

którego losuje się wagi. Najczęściej wagi określane są w całym układzie poprzez odpowiedni algorytm uczenia. Ogół wartości wag określanych we wszystkich neuronach w trakcie uczenia warunkuje sposób działania neuronu. Najlepszym rozwiązaniem byłby start z wartościami wag, które zbliżone są do optymalnych. Jest to metoda pozwalająca znacznie przyśpieszyć proces uczenia. Niestety, nie istnieją metody pozwalające na taki dobór wag, który zapewni właściwy punkt startowy (bez względu na rodzaj zadania). Z tej przyczyny nauka sieci neuronowych bazuje na wyznaczaniu wartości wag wyznaczonych na podstawie zbiorów danych wejściowych (z przeszłości) i zbiorów danych wyjściowych tak, aby zminimalizować wskaźnik błędu działania sieci. Sieć neuronowa posiłkuje się cechami danych i tworzy zasadę najbardziej odpowiadającą dla danego zjawiska. Zasada ta określa parametry nieliniowej funkcji o wielu wejściach, czyli przestawia najdokładniej zależności pomiędzy danymi na wejściu i wyjściu. Proces ma charakter wieloetapowy. Jego celem jest minimalizacja wartości błędów sieci, będących miarą różnić wartości rzeczywistych danych wejściowych d_i i wartości obliczonych przez sieć y_i [Frydrychowicz i Szymańska 2008]. Formułą, która jest zazwyczaj stosowana do wyznaczenia błędu jest współczynnik korelacji lyb suma kwadratów różnić pomiędzy wspomnianymi wartościami:

Równanie 7. Wartość błędu sieci neuronowej [Frydrychowicz i Szymańska 2008].

$$E = \sum_{i=1}^{N} (d_i - y_i)^2$$

gdzie N to liczebność zbioru uczącego. Informacja odnośnie błędu przetwarzana jest przez algorytm nauki i wykorzystywana do zmiany współczynników wagowych neuronu. Kroki prowadzące do wyznaczenia kolejnych wartości wag określane są jako „epoki uczenia". Pojedyncza epoka obejmuje jednostkowy przekaz informacji dotyczący błędów oraz modyfikację parametrów sieci, a konkretnie wag połączeń. Zły dobór wag w układzie może doprowadzić do zjawiska przedwczesnego nasycenia neuronów – pojawienie się stałego błędu w procesie uczenia. W większości przypadków błąd ten pojawia się dla zbyt dużych wartości początkowych wag. Tylko część neuronów sieci może być nasycona. Pozostałe znajdują się w zakresie liniowym i dla takich neuronów informacje zwrotne, które odpowiadają za naukę neuronu przyjmują normalną postać i dla nich wagi zmieniają się w sposób normalny (następuje szybka redukcja błędu). W takim układzie, neurony wysycone nie uczestniczą w odwzorowaniu danych przez co następuje zmniejszenie liczby neuronów sieci. Nasycone neurony znacznie spowalniają uczenie się modelu, a spowolnienie to może trwać nieskończenie lub do momentu wyczerpania iteracji. Często spotykanym wyjściem z tego typu problemów jest zmniejszenie zakresu losowanych wag początkowych neuronów. Główną zaletą sztucznych sieci neuronowych jest automatyczne wytrenowanie wag neuronów sieci. Z pomocą przychodzi tutaj najbardziej elementarny algorytm uczenia określany jako wsteczna propagacja błędu. Jest to metoda, która pozwala zmodyfikować wagi w całej sieci modelu (we wszystkich jego warstwach). Założeniem wstecznej propagacji błędów jest minimalizacja sumy kwadratów błędów uczenia z zastosowaniem optymalizacyjnej metody największego spadku [Osowski 2000]. Na rysunku 13 zaprezentowano ogólny schemat procesu trenowania sieci neuronowej:



Rysunek 13. Etapy uczenia sieci neuronowej [Frydrychowicz i Szymańska 2008]

Wysoka dokładność predykcji modelu, możliwość zastosowania zautomatyzowanego procesu wstecznej propagacji błędu dla niedostępnych, głębszych warstw sieci, umiejętność dostosowywania się do zmiennych sytuacji, zdolność do klasyfikacji zjawisk i ich interpretacji, równoległe przetwarzanie dużej ilości danych w krótkim czasie czyni algorytm sztucznych sieci neuronowych bardzo uniwersalnym oraz jednym z najskuteczniejszych algorytmów uczenia maszynowego.

2. Część doświadczalna

2.1. Miejsca pobierania próbek

W niniejszej pracy przeprowadzono analizy jakościowe i ilościowe frakcji PM_{10} pyłu zawieszonego. Celem analiz było określenie stężeń pyłu oraz wielopierścieniowych węglowodorów aromatycznych. Kampanie pomiarowe przeprowadzono w Wadowicach w roku 2017 oraz Krakowie w latach 2020/2021. Miejsca pobierania próbek wybrano ze względu na znaczne różnice pod względem liczby ludności, powierzchni i stopnia uprzemysłowienia a także charakteru stosowanych źródeł ciepła.

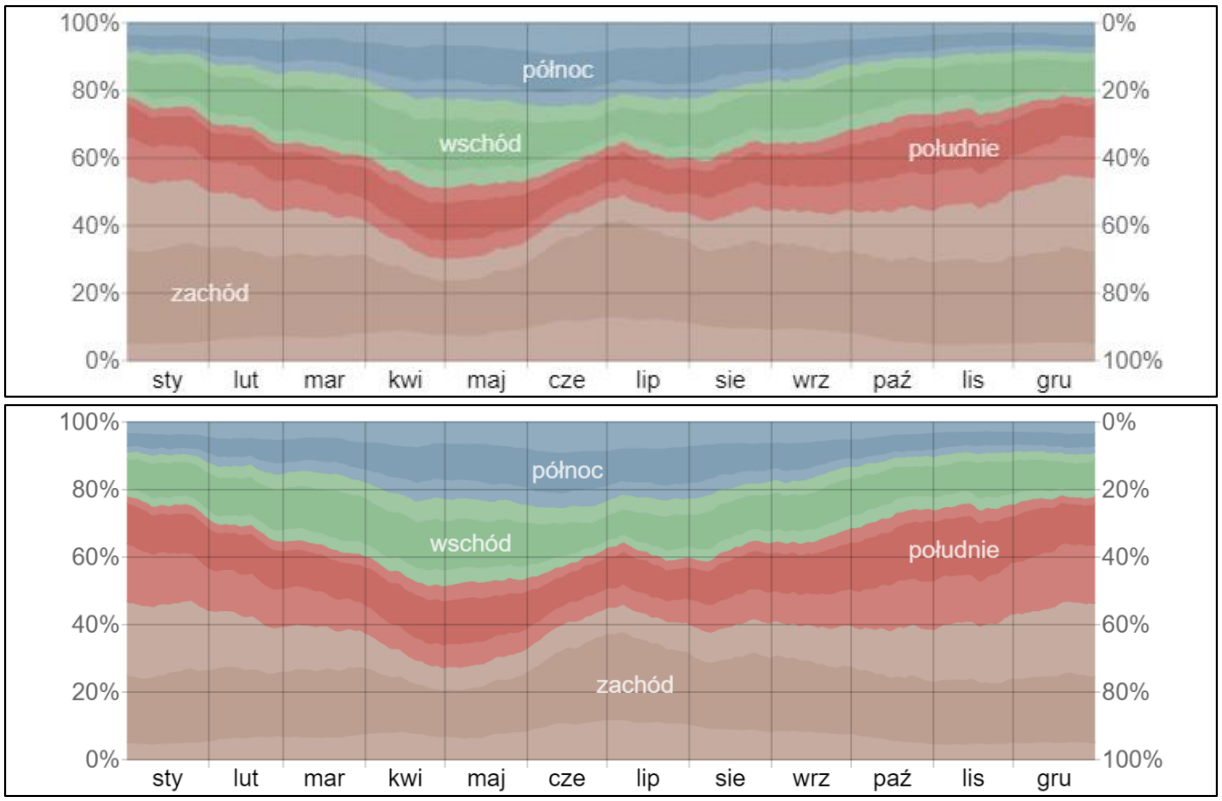
Miejscowość Wadowice – miasto o powierzchni 10,54 km2, liczbie mieszkańców 18230 i gęstości zaludnienia 1729,6 os./km2 (stan na 30.06.2021) [Główny Urząd Statystyczny 2023] charakteryzuje się znaczną różnicą w wysokości względnej terenu wahającą się w granicach od 263 m n.p.m. do 318 m n.p.m obniżającą się z zachodu na wschód w kierunku doliny Skawy. Wadowice znajdują się w Karpackim regionie klimatycznym. W ciągu roku temperatura waha się w granicach od -5 °C do 23 °C. Intensywne opady atmosferyczne przypadają na okres maj-sierpień. Przeważającym kierunkiem wiatru jest kierunek zachodni (rysunek 14), średnia roczna prędkość wiatru wynosi około 3,7 m/s. Na stan powietrza atmosferycznego mają tutaj wpływ głównie źródła, takie jak transport oraz liczne paleniska domowe. W 2016 roku, według raportu Światowej Organizacji Zdrowia, Wadowice zostały umieszone na 20 miejscu na liście najbardziej zanieczyszczonych miast w Unii Europejskiej [World Health Organization 2016].

Kraków o powierzchni 326,85 km^2 i liczbie ludności 802800 (2456 os./km^2), wysokości względnej 188–383 m n.p.m (stan na 30.06.2022) [Urząd Statystyczny w Krakowie 2023] położony jest w dolinie Wisły, na styku czterech krain geograficznych: Wyżyny Krakowsko-Częstochowskiej od północy, Kotliny Sandomierskiej od wschodu, Pogórza Zachodniobeskidzkiego od południa oraz Kotliny Oświęcimskiej od zachodu [Solon i in. 2018]. Miasto to charakteryzuje się klimatem kontynentalnym, w którym najniższe miesięczne temperatury notowane są w grudniu (ok. -3 °C), a najwyższe w lipcu (ok. 24 °C). Średnie roczne opady atmosferyczne wynoszą 730 mm i ze względu na efekty miasta są nieco wyższe niż na peryferiach (690 mm dla stacji w Balicach). Dominujący wiatr ma kierunek zachodni (19,7%) (rysunek 14). Częsta inwersja temperaturowa oraz gęsta zabudowa przyczyniają się do utrudnionej naturalnej wentylacji miasta, co sprzyja występowaniu wysokich stężeń zanieczyszczeń [Matuszko i Piotrowicz 2015]. Na mocy uchwały Nr XVII/243/16 Sejmiku Województwa Małopolskiego z dnia 15 stycznia 2016r, Kraków został objęty bezwzględnym zakazem spalania paliw stałych [Sejmik Województwa Małopolskiego 2016]. 24 kwietnia 2017 roku Sejmik Województwa Małopolskiego przegłosował uchwałę o dopuszczeniu do spalania paliw o odpowiedniej jakości [Sejmik Województwa Małopolskiego 2017a]. Z racji tego, że otaczające Kraków miejscowości nie mają zakazu palenia paliw stałych, powstało określenie „obwarzanka Krakowskiego" jako obszaru, z którego obserwuje się napływ zanieczyszczeń na wskutek spływów chłodnego powietrza i w konsekwencji kumulację zanieczyszczeń nad obszarem miasta [Sejmik Województwa Małopolskiego 2020].

W tabeli 3 przedstawiono szczegółowo archiwalne dane meteorologiczne dla Krakowa i Wadowic w oparciu o analizę statystyczną historycznych godzinowych raportów pogodowych udostępnianych na stronie WeatherSpark i rekonstrukcji modeli od 1 stycznia 1980 do 31 grudnia 2016 [WeatherSpark 2023]. Tabela zawiera średnie miesięczne wartości następujących parametrów meteorologicznych: temperatura, czas trwania opadów atmosferycznych, prędkość wiatru.

Tabela 3. Archiwalne dane meteorologiczne dla Krakowa i Wadowic (1980-2016) [WeatherSpark 2023]

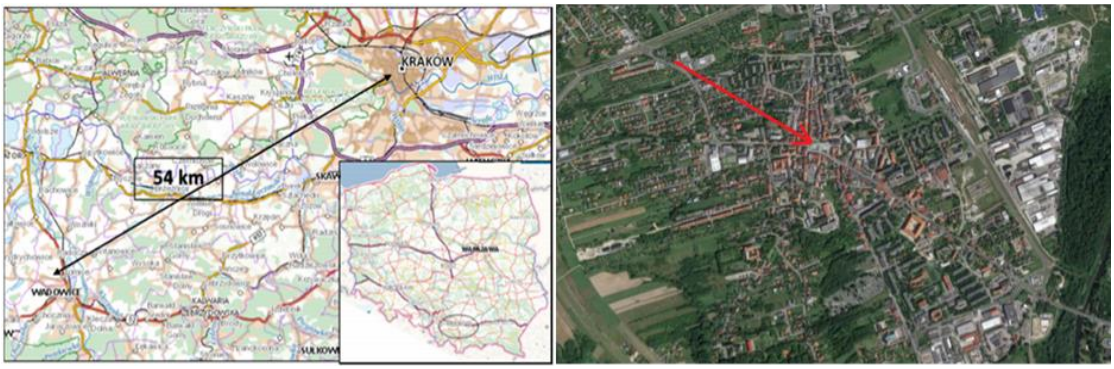
Średnia temperatura		**Miesiąc**											
		01	**02**	**03**	**04**	**05**	**06**	**07**	**08**	**09**	**10**	**11**	**12**
Kraków	Maks.	1 °C	3 °C	8 °C	14 °C	19 °C	22 °C	24 °C	24 °C	19 °C	14 °C	7 °C	3 °C
	Temperatura	-2 °C	-0 °C	3 °C	9 °C	14 °C	17 °C	19 °C	18 °C	14 °C	9 °C	4 °C	-0 °C
	Min.	-5 °C	-4 °C	-1 °C	4 °C	9 °C	12 °C	14 °C	13 °C	9 °C	4 °C	0 °C	-3 °C
Wadowice	Maks.	1 °C	3 °C	7 °C	13 °C	19 °C	21 °C	23 °C	23 °C	18 °C	13 °C	7 °C	2 °C
	Temperatura	-2 °C	-1 °C	3 °C	9 °C	14 °C	17 °C	19 °C	18 °C	14 °C	9 °C	4 °C	-1 °C
	Min.	-5 °C	-4 °C	-1 °C	4 °C	8 °C	11 °C	13 °C	12 °C	9 °C	5 °C	0 °C	-3 °C
Długość opadów [dni]		**Miesiąc**											
		01	**02**	**03**	**04**	**05**	**06**	**07**	**08**	**09**	**10**	**11**	**12**
Kraków	Deszcz	3,5 dn.	3,2 dn.	5,0 dn.	6,9 dn.	10,3 dn.	11,3 dn.	11,1 dn.	9,3 dn.	7,3 dn.	6,1 dn.	4,7 dn.	3,8 dn.
	Opad mieszany	0,7 dn.	0,9 dn.	0,8 dn.	0,3 dn.	0,0 dn.	0,0 dn.	0,0 dn.	0,0 dn.	0,0 dn.	0,1 dn.	0,5 dn.	1,0 dn.
	Śnieg	1,2 dn.	0,8 dn.	0,4 dn.	0,0 dn.	0,0 dn.	0,0 dn.	0,0 dn.	0,0 dn.	0,0 dn.	0,0 dn.	0,5 dn.	0,9 dn.
	Dowolny	5,4 dn.	5,0 dn.	6,3 dn.	7,3 dn.	10,3 dn.	11,3 dn.	11,1 dn.	9,3 dn.	7,3 dn.	6,3 dn.	5,7 dn.	5,7 dn.
Wadowice	Deszcz	3,8 dn.	3,8 dn.	5,8 dn.	7,4 dn.	10,6 dn.	11,7 dn.	11,6 dn.	9,7 dn.	7,7 dn.	6,4 dn.	5,0 dn.	4,1 dn.
	Opad mieszany	0,8 dn.	0,8 dn.	0,8 dn.	0,3 dn.	0,0 dn.	0,0 dn.	0,0 dn.	0,0 dn.	0,0 dn.	0,1 dn.	0,6 dn.	1,0 dn.
	Śnieg	1,5 dn.	1,2 dn.	0,6 dn.	0,1 dn.	0,0 dn.	0,0 dn.	0,0 dn.	0,0 dn.	0,0 dn.	0,0 dn.	0,7 dn.	1,2 dn.
	Dowolny	6,1 dn.	5,8 dn.	7,2 dn.	7,8 dn.	10,7 dn.	11,7 dn.	11,6 dn.	9,7 dn.	7,7 dn.	6,6 dn.	6,3 dn.	6,3 dn.
Prędkość wiatru [m/s]		**Miesiąc**											
		01	**02**	**03**	**04**	**05**	**06**	**07**	**08**	**09**	**10**	**11**	**12**
Kraków		4,8	4,7	4,5	3,9	3,5	3,5	3,4	3,4	3,7	3,9	4,1	4,6
Wadowice		4,5	4,4	4,2	3,6	3,3	3,2	3,1	3,1	3,4	3,6	3,9	4,3



Rysunek 14. Procentowy udział różnych kierunków wiatru uśrednionych miesięcznie okresu 1980-2016 dla Krakowa (góra) i Wadowic (dół) [WeatherSpark 2023]

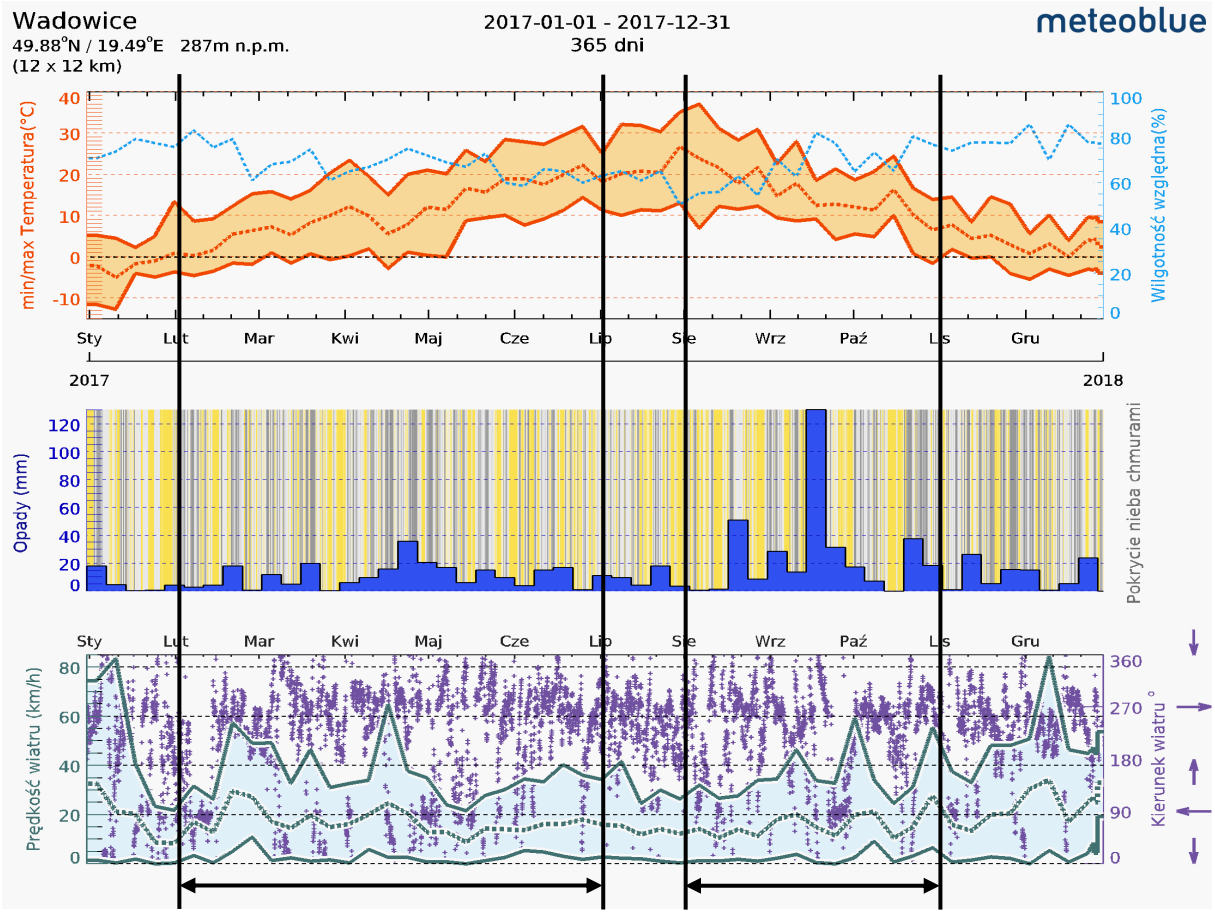
2.2. Kampanie pomiarowe pyłu zawieszonego PM_{10}

Pył zawieszony w Wadowicach (rysunek 15) pobierano w punkcie o współrzędnych 49°88'N, 19°49'E, 270 m n.p.m. za pomocą pobornika nisko-objętościowego PNS-15 (rysunek 16). Pobornik składa się z wysięgnika, na którym zamontowana była głowica separująca, gdzie zasysany poprzez pompę pył zawieszony był separowany na określone wielkości cząstki i następnie kierowany na filtr. Próbki pyłu pobierano podczas badania jakości powietrza w ramach akcji Krakowskiego Alarmu Smogowego od lutego do października 2017r. Proces pobierania wykonywany był zgodnie z procedurą zawartą w Dyrektywie Parlamentu Europejskiego i Rady 2008/50/WE z dnia 21 maja 2008r [Dyrektywa Parlamentu Europejskiego i Rady 2008]. W trakcie pobierania próbki, frakcja PM_{10} pyłu zawieszonego zatrzymywana jest na filtrze kwarcowym (Whatman QM-A 47mm) o średnicy aktywnej 44 mm. Filtry przed poborem były odpowiednio kondycjonowane zgodnie z normą PN-EN 12341:2006a. Następnie, pobrane próbki były kondycjonowane według normy PN-EN 14907:2006b i przechowywane w zamrażarce w temperaturze -20°C do czasu analizy. Rysunek 17 prezentuje archiwalne dane pogodowe dla Wadowic dla okresu pobierania próbek:



Rysunek 15. Miejsce pobierania próbek - Wadowice 2017 [Google Maps 2022]

Rysunek 16. Pobornik nisko-objętościowy PNS-15 [materiał własny]

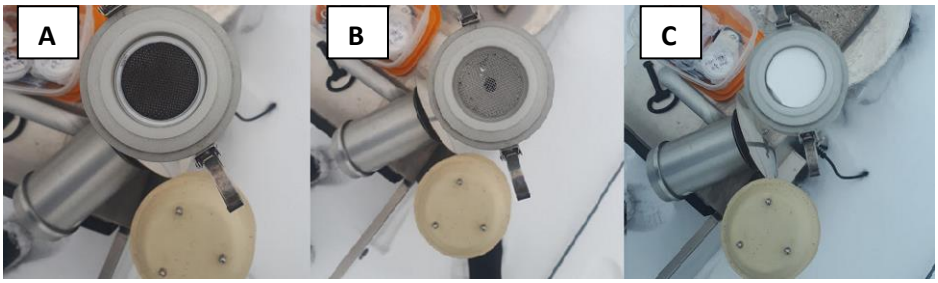


Rysunek 17. Dane meteorologiczne dla okresu kampanii pobierania próbek (z zaznaczonymi okresami pobierania) - Wadowice [Meteoblue 2023]

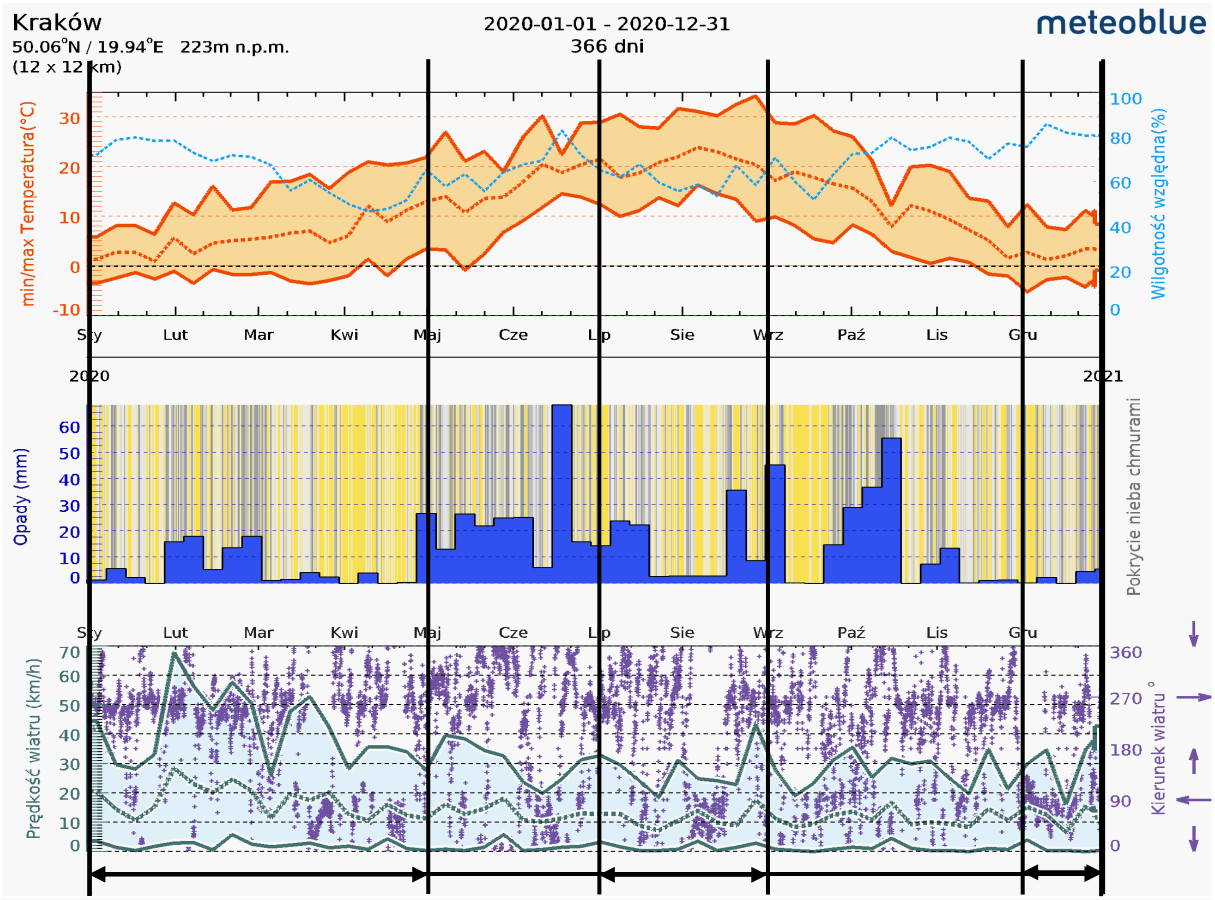
W Krakowie pobierano średniodobowe próbki pyłu zawieszonego PM_{10} niskoobjętościowym próbnikiem LV Leckel (rysunek 19 i 20) który działa analogicznie jak model wykorzystywany w Wadowicach (rysunek 16). Pobór odbywał się w kilku okresach: styczeń-kwiecień 2020, lipiec-sierpień 2020, grudzień 2020-luty 2021. Dla właściwej selekcji frakcji pyłu, pobornik wyposażony był w stabilizator przepływu powietrza utrzymujący go na poziomie 2,3 m^3/h i czasu pobierania. Próbnik był umieszczony na dachu trzy piętrowego budynku Wydziału Fizyki i Informatyki Stosowanej, Akademii Górniczo-Hutniczej w Krakowie (50°07'N, 19°91'E, 220 m n.p.m) (rysunek 18) położonym pomiędzy ulicami Kawiory, Nawojki, Czarnowiejskiej i Władysława Reymonta. Ulice te charakteryzują się wysokim natężeniem ruchu drogowego, co przekłada się na wysoką emisję spalin. Na rysunku 21 przedstawiono dane meteorologiczne występujące podczas kampanii pomiarowej:

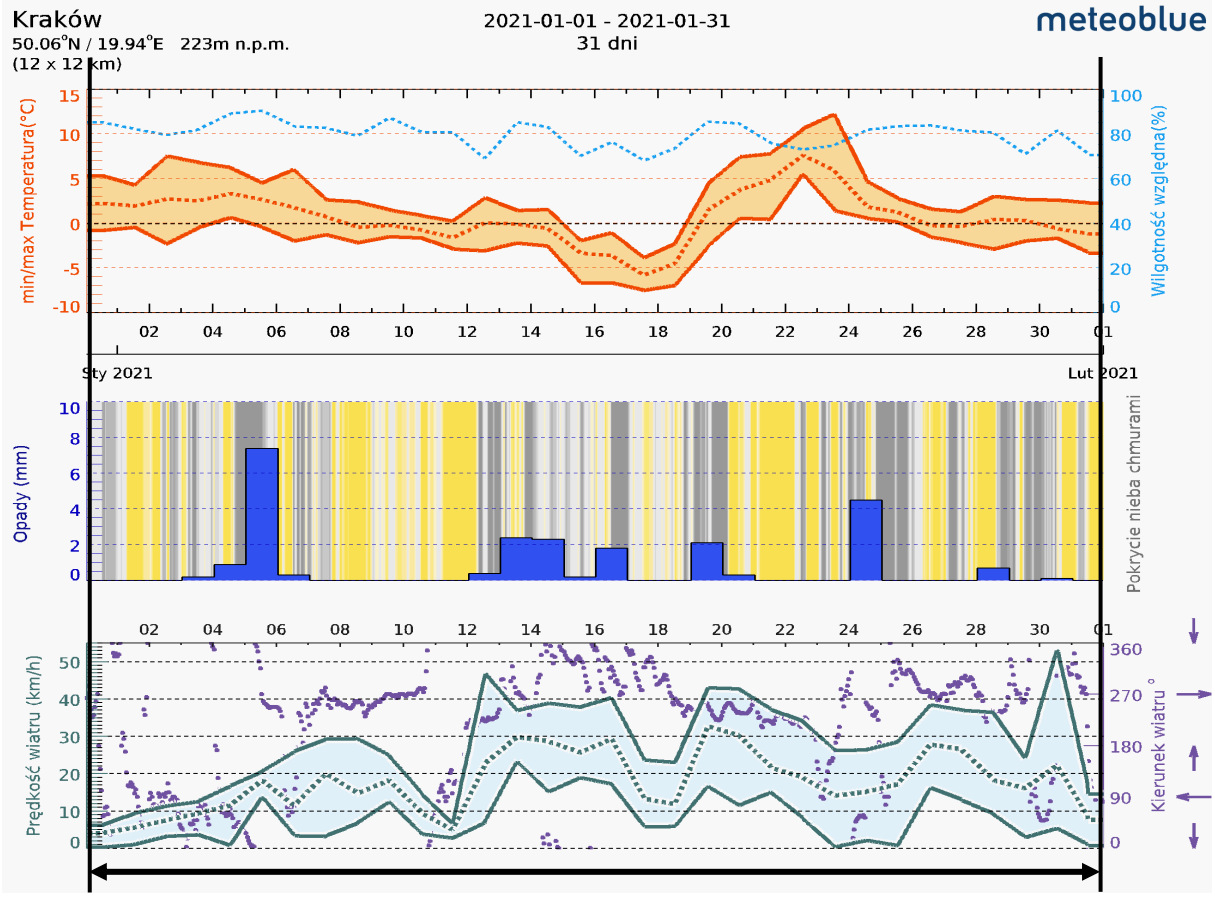
Rysunek 18. Miejsce pobierania próbek - Kraków 2020-2021 [Google Maps 2022]

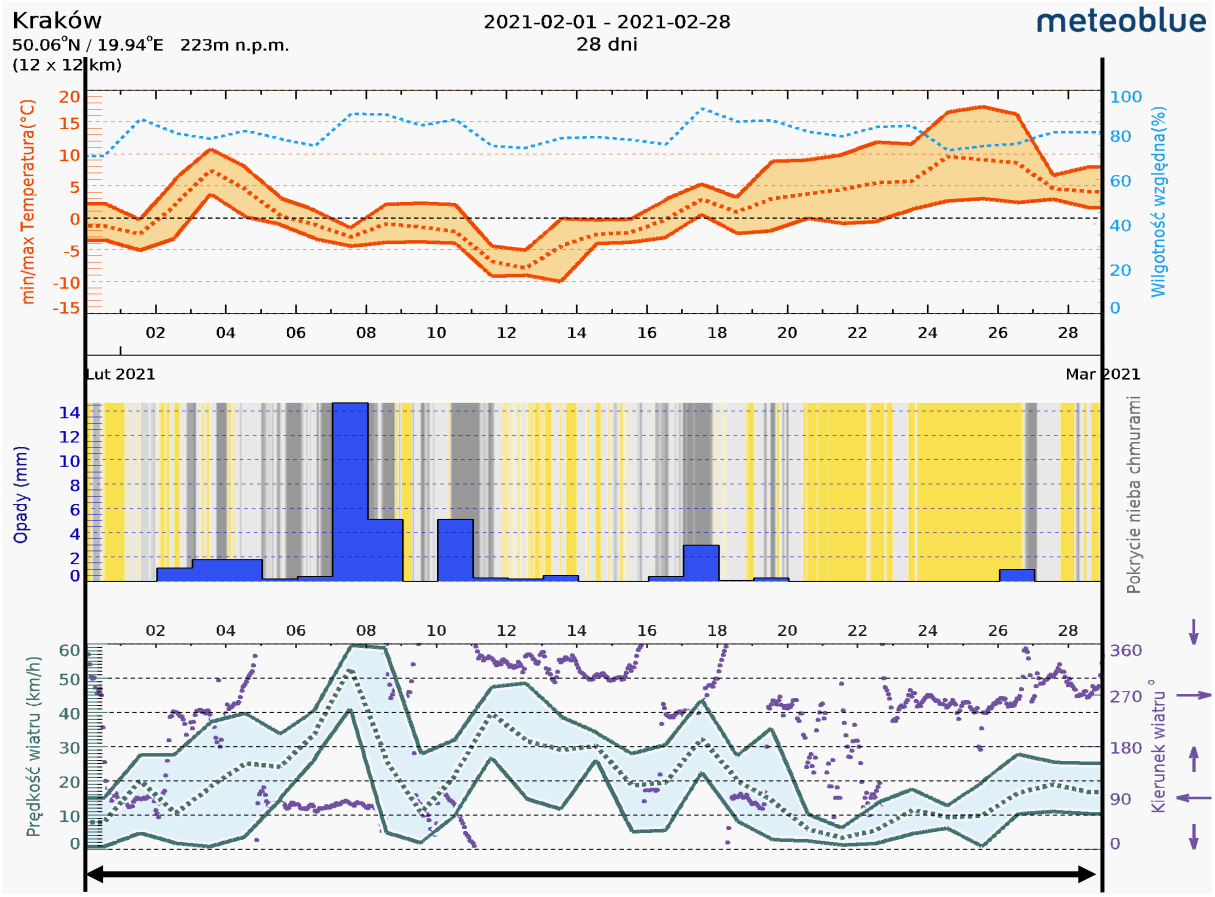
Rysunek 19. Pobornik nisko-objętościowy LV Leckel [materiał własny]



Rysunek 20. Wysięgnik z otwartym próbnikiem: (A) pobornik z filtrem z zaadsorbowanym pyłem, (B) pobornik bez filtra, (C) pobornik z czystym filtrem [materiał własny]


Kraków
50.06°N / 19.94°E 223m n.p.m.
(12 x 12 km)
2020-01-01 - 2020-12-31
366 dni
meteoblue
min/max Temperatura(°C)
Wilgotność względna(%)
Sty Lut Mar Kwi Maj Cze Lip Sie Wrz Paź Lis Gru
2020
2021
Opady (mm)
Pokrycie nieba chmurami
Prędkość wiatru (km/h)
Kierunek wiatru °


Kraków
50.06°N / 19.94°E 223m n.p.m.
(12 x 12 km)
2021-01-01 - 2021-01-31
31 dni
meteoblue
min/max Temperatura(°C)
Wilgotność względna(%)
02 04 06 08 10 12 14 16 18 20 22 24 26 28 30 01
Sty 2021
Lut 2021
Opady (mm)
Pokrycie nieba chmurami
Prędkość wiatru (km/h)
Kierunek wiatru °



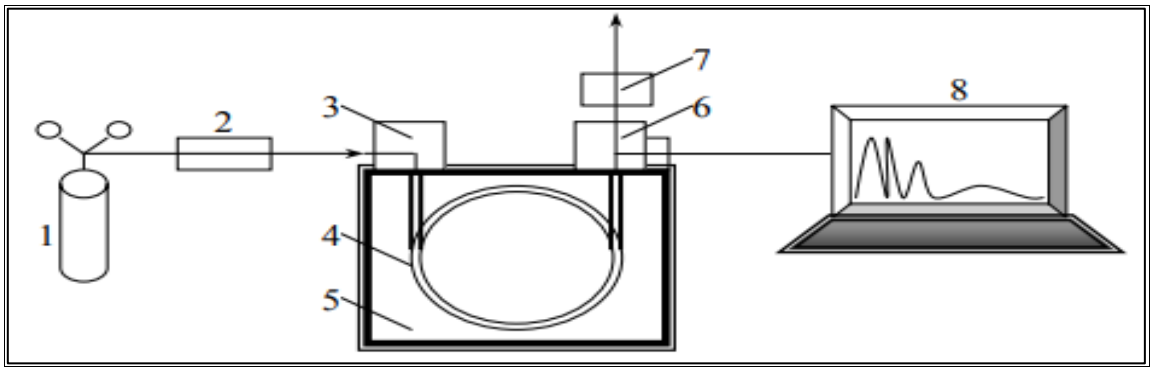
Rysunek 21. Dane meteorologiczne dla okresu kampanii pobierania próbek (z zaznaczonymi okresami pobierania) - Kraków [Meteoblue 2023]

Proces pobierania pyłu w Krakowie przeprowadzono zgodnie z Dyrektywą Parlamentu Europejskiego i Rady 2008/50/WE z dnia 21 maja 2008r w sprawie jakości powietrza i czystszego powietrza dla Europy [Dyrektywa Parlamentu Europejskiego i Rady 2008].

Przygotowanie i kondycjonowanie filtrów, pobór, i przechowywanie próbek odbywały się w tych samych warunkach, co w przypadku Wadowic (PN-EN 12341:2006a; PN-EN 14907:2006b).

2.3. Analiza zawartości WWA w pyle zawieszonym

Do badań zawartości WWA wykorzystano technikę chromatografii gazowej ze spektrometrią mas (GC-MS). Jest to metoda najczęściej stosowana w przypadku analiz ilościowych i jakościowych zarówno próbek gazowych jak i ciekłych. Podstawowe informacje uzyskuje się z widm masowych powstałych w spektrometrze mas dla analitów wychodzących z kolumny chromatograficznej. Identyfikacja analitów odbywa się na podstawie pomiaru masy cząsteczkowej analitu, a ściślej stosunku masy do ładunku (m/z). Uzupełnieniem informacji z widm masowych są wielkości retencyjne rozdzielanych związków. Wynik uzyskany w analizie chromatograficznej to chromatogram wykreślony jako zależność wysokości piku w czasie. Poniżej (rysunek 22) przedstawiono przykładowy schemat GC-MS [Namieśnik i Jamrógiewicz 1998].



Rysunek 22. Schemat chromatografu gazowego 1-zbiornik gazu nośnego, 2-regulator przepływu gazu, 3-dozownik, 4-kolumna, 5-termostat, 6-detektor, 7-przepływomierz, 8-komputer lub rejestrator [Bartulewicz i in. 1997]

Dozownik w układzie GC-MS odpowiada za wprowadzenie analitów w strumień fazy ruchomej oraz transport mieszaniny do kolumny. Proces rozdziału przebiega w kolumnie chromatograficznej pokrytej wewnątrz cienką warstwą polimerowej fazy stałej. Przepływ gazu nośnego w kolumnie chromatograficznej jest podstawowym czynnikiem napędowym procesu rozdzielania. Sile tej przeciwstawia się faza stacjonarna, która opóźnia przepływ analizowanych związków. Do najważniejszych cech wypełnienia fazy stacjonarnej zalicza się ich polarność, która odpowiada za oddziaływanie z cząsteczkami rozdzielanych związków . Oprócz tego, należy pamiętać o doborze właściwej długości i średnicy kolumny a także grubości filmu. Dodatkowo czynnikiem warunkującym prace kolumny jest jej zakres temperaturowy, w którym zachowana jest prawidłowa i wysoka wydajność rozdziału. Właściwości te decydują o szerokości uzyskanego piku i czasie retencji analitów. Detektor stanowi element odpowiedzialny za identyfikację rozdzielanych składników w kolumnie. Charakteryzuje się wysoką czułością, dobrą selektywnością i stabilnością wskazań sygnałów pochodzących od konkretnego analitu. Często stosowaną metodą w badaniach środowiskowych jest chromatografia gazowa sprzężona ze spektrometrią mas (GC-MS). W spektrometrze mas w pierwszym etapie zachodzi jonizacja związków zachodzi, następnie ich fragmentacja, w wyniku której powstają dodatnio naładowane jony o mniejszej masie. Jony te kierowane są do analizatora, gdzie grupowane są pod względem określonego stosunku masy do ładunku. Czas przejścia danego analitu od momentu nastrzyku do momentu pojawienia się sygnału analitycznego w detektorze określany jest czasem retencji i jest typowy dla pojedynczego związku, a rozbudowa układu o spektrometrie mas pozwala uzyskać widma masowe charakterystyczne dla danego związku [Bartulewicz i in. 1997, Namieśnik i Jamrógiewicz 1998].

Pierwszym etapem przygotowania próbek było grupowanie filtrów na podstawie wartości stężeń frakcji PM_{10}. W celu uzyskania wartości stężeń WWA powyżej wartości LOQ, łączono wycinki filtrów uzyskanych w dniach o niskich wartościach stężeń pyłu PM_{10} (≤ 20 µg/m^3). Z każdego filtra o średnicy 44 mm wycinano mniejsze fragmenty o średnicy 18 mm. Uzyskane wycinki, pojedyncze lub grupowane umieszczano w fiolkach szklanych o objętości 20 ml.

Stężenie pyłu zawieszonego C_{PM10} wyznaczono za pomocą wzoru:

Równanie 8. Stężenie pyłu zawieszonego PM_{10}

$$C_{PM_{10}} = \frac{m_{śr,k} - m_{śr,p}}{V_{pow}}$$

gdzie $m_{śr,p}$ i $m_{śr,k}$ odpowiadają odpowiednio za średnią masę filtra przed poborem i po pobraniu pyłu [µg], a V_{pow} stanowi objętość zassanego powietrza spisaną bezpośrednio z próbnika [m^3].

Z racji tego, że dla WWA uzyskane stężenia odnoszą się do wycinka filtra, przeliczono uzyskane wyniki według wzoru:

Równanie 9. Stężenie wielopierścieniowych węglowodorów aromatycznych w pyle zawieszonym PM_{10}

$$C_{WWA} = \frac{C * V_{ekstr.} * S_{cał.}}{S_{wyc.} * V_{pow}}$$

gdzie:
- C – stężenie WWA w próbkach, [ng/ml]
- $V_{ekstr.}$ – objętość ekstraktu przygotowana do badań, [ml]
- V_{pow} – objętość zassanego powietrza, [m^3]
- $S_{cał.}$ – powierzchnia całkowita, aktywna filtra, [mm]
- $S_{wyc.}$ – powierzchnia wycinka filtra wykorzystana do badań, [mm]

Kolejny etap to ekstrakcja rozpuszczalnikowa. Przy użyciu wytrząsarki poziomej (prędkość obrotów 50 obr./min) przez 40 minut próbki ekstrahowano dwukrotnie mieszaniną rozpuszczalników organicznych cykloheksan: dichlorometan, odpowiednio w stosunku objętościowym 2 : 3. Uzyskane roztwory łączono ze sobą i odparowano do objętości 500 µl w termobloku AccuBlock Digital Dry Bath w temperaturze 30 °C w strumieniu argonu. Tak uzyskane mieszaniny odwirowywano z prędkością 12 000 obr./min w celu oddzielenia stałych zanieczyszczeń, które powstały podczas ekstrakcji. Z tak uzyskanego roztworu pobrano 200 µl cieczy, przeniesiono do fiolek chromatograficznych i poddano analizie GC-MS.

Przed przystąpieniem do analizy GC-MS sporządzono roztwory wzorcowe będące mieszaniną WWA o stężeniach w zakresie od 2,5 ng/ml do 10 000 ng/ml. Roztwory te zostały wykorzystane do walidacji metody oraz wyznaczenia krzywych kalibracyjnych.

Analizy wykonano podczas stażu na Uniwersytecie Jana Kochanowskiego w Kielcach za pomocą chromatografu gazowego sprzężonego ze spektrometrem mas Clarus 600 Gas Chromatograph, Clarus 600T Mass Spectrometer z wieżą wtryskową (rysunek 23). W tabeli 4 przedstawiono parametry pracy GC-MS.

Rysunek 23. Chromatograf gazowy ze spektrometrią mas Clarus 600 Gas Chromatograph, Clarus 600T Mass Spectrometer, UJK Kielce [materiał własny]

Tabela 4. Parametry pracy GC-MS

Parametry pracy chromatografu gazowego ze spektrometrią mas GC-MS	
Tryb pracy	FULL SCAN – identyfikacja wszystkich jonów w ustalonym zakresie mas cząsteczkowych
Kolumna chromatograficzna	Kolumna kapilarna HP-5ms – 30 m x 0,25 mm – 0,25 µm
Program temperaturowy	I – temperatura 60 °C przez 1 min II – wzrost temperatury 15 °C/min do 300 °C III – temperatura 300 °C przez 13 min
Temperatura dozownika	250 °C
Temperatura źródła jonów	250 °C
Temperatura linii transferowej	250 °C
Energia jonizacji	70 eV
Objętość dozowanej próbki	1 µl
Gaz nośny	He, 1 ml/min
Całkowity czas analizy	30 min

Analiza roztworów wzorcowych dla mieszaniny EPA 525 PAH Mix A pozwoliła na wyznaczenie czasów retencji poszczególnych składników mieszaniny oraz porównanie ich z biblioteką widm masowych. W tabeli 5 zaprezentowano jony główne, charakterystyczne oraz czas retencji dla analizowanych związków:

Tabela 5. Czas retencji, jon główny oraz jony charakterystyczne badanych analitów

Związek	Jon główny	Jony charakterystyczne	Czas retencji [min]
Naftalen	128	102	6,11
Acenaftylen	152	149, 126, 125, 98	8,45
Acenaften	153	149, 125	8,77
Fluoren	166	162, 139, 115	9,72
Fenantren	178	175, 152, 151	11,68
Antracen	178	175, 152, 151	11,79
Fluoranten	202	199, 174, 150	14,32
Piren	202	199, 198, 174, 150	14,82
Benzo[*a*]antracen	228	223, 202	17,60
Chryzen	228	223, 202, 200	17,69
Benzo[*b*]fluoranten	252	248, 226	19,95
Benzo[*k*]fluoranten	252	248, 247, 226	20,01
Benzo(a)piren	252	248, 226, 224	20,60
Indeno[*1,2,3-cd*]piren	276	272, 248	22,98
Dibenzo[*ah*]antracen	278	252	23,06
Benzo[*ghi*]perylen	276	266	23,64

2.4. Elementy walidacji metody oznaczania wielopierścieniowych węglowodorów aromatycznych

Procesem walidacji określa się weryfikację metody, którą przeprowadza się w celu potwierdzenia lub ustalenia oczekiwanej wartości testu.

Podczas walidacji należy ustalić, które parametry charakteryzują naszą metodą, wyznaczyć te parametry i sprawdzić stawiane oczekiwania względem tych wartości. Dzięki tym czynnościom uzyskujemy pewność, że proces analizy przebiega w sposób rzetelny i precyzyjny, a przede wszystkim daje wiarygodne wyniki analiz [Clark i in. 2009, Jakubowska 2020, Stefaniuk i in. 2015].

W celu wyznaczania efektywności ekstrakcji metody analitycznej sporządzono dodatkowe próbki o znanym stężeniu analitu. Na suche wycinki filtra (przygotowane według wcześniej opisanej procedury) przeniesiono odpowiednią objętość roztworu wzorcowego o

stężeniu 2 000 ng/ml w celu uzyskania odpowiednio 1 000 ng, 500 ng, 100 ng oraz 50 ng każdego analitu na filtrze. Następnie filtry pozostawiono na 15 minut, w otwartych fiolkach pod wyciągiem, w celu odparowania rozpuszczalnika. Następnie postępowano zgodnie z procedurą przygotowania próbek, aż do otrzymania gotowych ekstraktów, które po przeniesieniu do fiolek chromatograficznych, skierowano do analizy chromatograficznej. Analizę chromatograficzną każdego roztworu wykonano czterokrotnie. W celach walidacji metody wykorzystano wartości pól powierzchni uzyskanych dla konkretnych związków.

Dla danych z Wadowic wykorzystanych w pracy informacje dotyczące walidacji metody analitycznej zawarte są w artykule [Furman i in. 2021]. Badania próbek frakcji PM_{10} pyłu zawieszonego pobranych w Wadowicach zostały przeprowadzone na chromatografie gazowym ze spektrometem mas Trace 1310 Gas Chromatograph, ITQ 900 Mass Spectrometer z automatycznym samplerem TriPlus RSH Autosampler, dostępnym na Wydziale Energetyki i Paliw AGH (rysunek 24).

Rysunek 24. Chromatograf gazowy ze spektrometrią mas Trace 1310 Gaz Chromatograph, ITQ 900 Mass Spectrometer, [materiał własny]

2.4.1. Dokładność metody analitycznej

Dokładność metody określa stopień zgodności wyników uzyskanych z zastosowaniem danej metody analitycznej z wynikami rzeczywistymi (oczekiwanymi). Metoda niedokładna zazwyczaj obarczona jest błędem systematycznym stałym i zmiennym, odpowiednio niezależnym i zależnym od stężenia badanego analitu. Dokładność wyznaczana jest na podstawie rozbieżności pomiędzy wartością prawdziwą, a wyznaczoną (określaną jako procent odzysku):

Równanie 10. Dokładność metody analitycznej

$$\frac{x_i}{y} * 100\%$$

gdzie x_i określa oznaczoną ilość analitu w badanej próbce, a y określa znaną ilość analitu w badanej próbce. Odzysk dla próbek ze składnikiem głównym (o znacznie przewyższającym stężeniu) powinien wynosić 95 % - 100 %, a dla składników składowych na poziomie śladowym 80 % - 120% [Clark i in. 2009, Jakubowska 2020].

Dokładność dla próbek PM_{10} zbadano wykorzystując 4-krotne powtórzenie analiz dla próbek o znanej ilości analitu (1 000 ng, 500 ng, 100 ng oraz 50 ng). Wartości powierzchni pików dla każdego związku uśredniono i dla tej wartości, wykorzystując powyższy wzór, obliczono dokładność (procent odzysku) – wielkość efektywności ekstrakcji. Wyniki przedstawiono w tabeli 6.

Tabela 6. Dokładność (procent odzysku) dla poszczególnych związków

Związek	Dokładność (procent odzysku) %				Średnia %
	50 ng	100 ng	500 ng	1 000 ng	
Naftalen	94	89	96	91	**92**
Acenaftylen	91	97	96	86	**93**
Acenaften	92	95	91	90	**92**
Fluoren	89	94	93	89	**91**
Fenantren	89	89	99	92	**92**
Antracen	91	99	97	89	**94**
Fluoranten	92	95	96	90	**93**
Piren	93	96	95	98	**96**
Benzo[*a*]antracen	95	92	96	94	**94**
Chrysen	92	96	97	90	**94**
Benzo[*b*]fluoranten	92	94	96	91	**93**
Benzo[*k*] fluoranten	94	101	98	90	**95**
Benzo[*a*]piren	92	98	96	94	**95**
Indeno[*1,2,3-cd*]piren	96	94	96	100	**97**
Dibenzo[*ah*]antracen	80	88	92	91	**88**
Benzo[*ghi*]perylen	84	93	95	94	**91**

Najniższa dokładność została określona dla Dibenzo[*ah*]antracenu. Wszystkie wartości mieszczą się w zakresie 80 % - 120 % – spełniają kryteria akceptacji. W całym procesie preparacji próbek do badań następuje utrata średnio około 8 % analitu. W dalszych obliczeniach stężeń WWA była uwzględniana strata o wartości równej danemu związkowi.

2.4.2. Precyzja, powtarzalność metody analitycznej

Precyzja metody analitycznej określana jest jako stopień zgodności pomiędzy pojedynczymi, niezależnymi wynikami analizy, gdzie konkretna procedura stosowana jest dla wielokrotnie powtarzanych i niezależnych oznaczeń jednorodnej próbki. Najczęściej spotykaną miarą precyzji jest odchylenie standardowe *S*, względne odchylenie standardowe *RDS* lub współczynnik zmienności *CV* dla danej serii pomiarów próbki na stałym poziomie stężeń. Precyzja określana dla metody analitycznej i wyników pochodzących z jednego laboratorium oraz badań wykonywanych przez jednego laboranta nosi nazwę powtarzalności. Kryterium akceptacji określa wartość *CV* dla głównego składnika nie przekraczającą 3%, a dla składników składowych na poziomie śladowym jest to maksymalnie 15 % [Konieczka i in. 2004].

Powtarzalność metody została określona dla czterokrotnych powtórzeń kilku wybranych próbek PM_{10} z każdej listy sekwencyjnej, wszystkich próbek o zadanym stężeniu początkowym analitu oraz wszystkich próbek z krzywej kalibracyjnej na każdym poziomie stężeń. W pierwszej kolejności wyznaczono odchylenie standardowe *S* uzyskanych wyników. Wartość ta traktowana jest jako rozrzut otrzymanych wyników od wartości średniej:

Równanie 11. Odchylenie standardowe S

$$S = \sqrt{\frac{\sum_{i=1}^{n}(x_i - x_{śr})^2}{n-1}}$$

gdzie x_i określa wartość pojedynczego wyniku, $x_{śr}$ to średnia arytmetyczna ze wszystkich wyników, a n to ilość wszystkich wyników. Następnie obliczono względne odchylenie standardowe *RDS*:

Równanie 12. Następnie obliczono względne odchylenie standardowe RDS

$$RDS = \frac{S}{x_{śr}}$$

Ostatecznie wyznaczono precyzję metody wykorzystując do tego celu parametr współczynnika zmienności *CV*:

Równanie 13. Precyzja metody - współczynnik zmienności CV

$$CV = RDS * 100\%$$

Wyniki współczynnika zmienności zostały przedstawione w tabeli 7.

Tabela 7. Współczynnik zmienności CV

Związek	Współczynnik zmienności *CV* %	
	Wartość średnia uzyskana dla wyników z krzywej i próbek o znanym stężeniu	Wartość średnia uzyskana dla wyników z próbek PM_{10}
Nafltalen	4,30	6,92
Acenaftylen	3,37	7,31
Acenaften	5,51	12,92
Fluoren	3,97	11,06
Fenantren	2,69	6,15
Antracen	4,02	5,69
Fluoranten	3,72	6,28
Piren	3,98	6,12
Benzo[*a*]antracen	3,65	4,91
Chrysen	4,16	5,49
Benzo[*b*]fluoranten	5,84	3,55
Benzo[*k*] fluoranten	3,78	7,85
Benzo[*a*]piren	2,16	4,82
Indeno[*1,2,3-cd*]piren	2,70	7,82
Dibenzo[*a,h*]antracen	3,77	7,23
Benzo[*ghi*]perylen	2,85	6,60

Niższe *CV*, czyli większą precyzję odnotowano dla próbek z krzywej i próbek o znanym stężeniu analitu. Niemniej, różnica ta nie jest znaczna. Najwyższą wartość *CV* wyznaczono dla Acenaftenu i Fluorenu, odpowiednio 9,22 % i 7,52 %, najniższą dla Benzo[*a*]antracenu i Benzo[*a*]pirenu, odpowiednio 4,28% i 3,49%. Aby metodę analityczną uznać za precyzyjną, 95% wyników *CV* powinno nie przekraczać kryterium akceptacji [Stefaniuk i in. 2015]. Wszystkie otrzymane wartości *CV* mieszczą się w granicach kryterium akceptacji, co pozwala uznać metodę za precyzyjną.

2.4.3. Liniowość

Liniowość określa się dla stężeń obejmujących deklarowany lub przewidywany zakres stężeń badanej substancji, najlepiej stanowiący od 80 % do 120 % spodziewanego stężenia. Powyższą zależność przedstawia równanie:

Równanie 14. Liniowość metody analitycznej

$$y = a * x + b$$

gdzie *y* odpowiada za sygnał analityczny wzorca, *x* to stężenie wzorca, *a* i *b* to współczynniki równania, odpowiednio nachylenia i przesunięcia prostej. Wielkością wyznaczającą liniowość określa się jako współczynnik korelacji *r*. Dla substancji głównej oznaczanej wartość *r* powinna być większa niż 0,995, natomiast dla substancji śladowych powinna być większa niż 0,980 [Konieczka i in. 2004, Stefaniuk i in. 2015].

W niniejszej pracy, do wyznaczenia współczynnika korelacji wykorzystano wartości średnie dla 4 powtórzeń analiz każdego stężenia wzorca (od 2,5 ng/ml – 5000 ng/ml). Na podstawie otrzymanych wyników sporządzono krzywe kalibracyjne, aproksymując je równaniem regresji liniowej. W tym przypadku *y* odpowiada za powierzchnie piku otrzymaną dla danego analitu z wzorca, a *x* odpowiada za wartość stężenia tego analitu. Analiza wszystkich związków wykazała, że nie jest zachowana liniowość w całym zakresie badanych stężeń wzorca. Dla niektórych analitów, w celu zachowania jak najlepszej liniowości (jak najwyższego współczynnika korelacji *r*) odrzucono część punktów z krzywej kalibracyjnej stale pamiętając o zakresie przewidywanych stężeń w próbkach. Z tego względu, dla kilku związków niezbędne było określenie liniowości w 2 zakresach stężeń. Zebrane wyniki, funkcje liniowe oraz wartości współczynnika korelacji przedstawia tabela 8:

Tabela 8. Regresja liniowa krzywej kalibracyjnej

Związek	**Równania regresji liniowej (z uwzględnieniem liniowości w 2 zakresach)**		**R^2**	
Nafltalen	y = 219,55x – 225,47		0,991	
Acenaftylen	y = 189,03x – 537,63		0,994	
Acenaften	y = 153,37x – 258,76		0,996	
Fluoren	y = 105,08x – 90,47		0,998	
Fenantren	y = 243,94x – 3151,80		0,996	
Antracen	y = 174,65x – 580,38		0,997	
Fluoranten	y = 162,87x – 504,78	y = 253,31x – 7286,70	0,992	1,000
Piren	y = 166,08x – 4,67	y = 261,17x – 10463,00	0,994	0,998
Benzo[*a*]antracen	y = 135,43x + 45,34	y = 227,44x – 5483,40	0,997	1,000
Chrysen	y = 154,71x + 12,85	y = 221,50x – 6134,30	0,999	0,999
Benzo[*b*]fluoranten	y = 122,06x – 74,82	y = 189,66x – 9568,40	0,991	1,000
Benzo[*k*] fluoranten	y = 109,31x + 231,27	y = 201,1x – 6447,90	0,995	1,000
Benzo[*a*]piren	y = 134,10x – 0,22	y = 220,9x – 10094,00	0,995	0,999
Indeno[*1,2,3-cd*]piren	y = 76,32x + 16,35	y = 178,14x – 6336,80	0,998	1,000
Dibenzo[*ah*]antracen	y = 95.026x - 416.41		0,979	
Benzo[*ghi*]perylen	y = 87,25x + 85,59	y = 182,55x – 5394,20	0,996	0,998

Wartość współczynnika korelacji jest we wszystkich przypadkach większa niż 0,980. Porównując otrzymaną wartość do kryterium akceptacji można stwierdzić, że opracowana metoda analityczna jest liniowa w całym zakresie. Należy jednak pamiętać o tym, że dla niektórych związków wyznaczono dwa zakresy liniowości, w których to wartość *R* wyliczono osobno z uwzględnieniem osobnych funkcji liniowych. Funkcje te zostały wykorzystane do wyliczenia stężeń badanych analitów.

2.4.4. Granica wykrywalności i oznaczalności

Granica wykrywalności LOD (limit of detection) to najmniejsze stężenie lub ilość analitu w badanej próbce, które można wykryć konkretną metodą analityczną. Korzystając z LOD można jakościowo ocenić, czy dany składnik znajduje się w analizowanej próbce. Granica oznaczalności LOQ (limit of quantification) to najmniejsze stężenie lub ilość analitu w badanej próbce, które można oznaczyć ilościowo konkretną metodą z odpowiednią precyzją i dokładnością. Zazwyczaj wartość LOD jest 2-3 razy mniejsza od LOQ [Konieczka i in. 2004, Stefaniuk i in. 2015].

Do wyznaczenia LOQ przyjęto wartość powierzchni piku danego analitu 3 razy mniejszą niż najmniejsze pole piku uzyskane z krzywej kalibracyjnej. Następnie wykorzystując wzory krzywych kalibracyjnych obliczono minimalne stężenia analitów w próbce, które można oznaczyć jakościowo. LOD zostało obliczone jako LOQ/3. Zebrane dane dla LOD i LOQ metody analitycznej przedstawiono w tabeli 9.

Tabela 9. Granica wykrywalności i oznaczalności metody analitycznej

Związek	LOQ [ng/ml]	LOD [ng/ml]
Nafltalen	1,74	0,58
Acenaftylen	3,27	1,09
Acenaften	2,16	0,72
Fluoren	1,89	0,63
Fenantren	13,77	4,59
Antracen	4,50	1,50
Fluoranten	4,02	1,34
Piren	0,84	0,28
Benzo[*a*]antracen	0,39	0,13
Chrysen	0,63	0,21
Benzo[*b*]fluoranten	1,53	0,51
Benzo[*k*] fluoranten	0,42	0,14
Benzo[*a*]piren	0,87	0,29
Indeno[*1,2,3-cd*]piren	1,71	0,57
Dibenzo[*ah*]antracen	5,61	1,87
Benzo[*ghi*]perylen	0,75	0,25

W dalszych obliczeniach uwzględniano wartość LOQ jako graniczne stężenie, dla którego oszacowano zawartość analitu w próbce. Wszystkie wyniki poniżej LOQ nie zostały uwzględnione w dalszych analizach i modelowaniu.

2.5. Wykorzystanie algorytmów uczenia maszynowego w celu identyfikacji stopnia narażenia

Otrzymane dane dotyczące stężenia pyłu zawieszonego PM_{10} oraz wielopierścieniowych węglowodorów aromatycznych zostały wykorzystane w celu identyfikacji stopnia narażenia mieszkańców na zanieczyszczenia powietrza. Obliczenia zostały przeprowadzone w środowisku Azure Machine Learning. Usługa Azure Machine Learning to środowisko przeglądarkowe ułatwiające naukowcom i deweloperom tworzenie i wdrażanie wysokiej jakości modeli oraz zarządzania nimi. Platforma ta jest stworzona z myślą o aplikacjach wykorzystujących sztuczną inteligencję. Pozwala na szybkie przygotowywanie i uczenie modeli z wykorzystaniem gotowych i zintegrowanych modułów realizujących obliczenia,

obsługi struktur i ogólnodostępnych bibliotek (open source). Stworzone modele można w łatwy sposób udostępnić w środowisku *Web Service* na potrzeby współpracy. Środowisko udostępnia narzędzia pozwalające na łatwą analizę dokładności i wydajności przygotowanych modeli.

Głównym założeniem tej części było wyznaczenie granicznych wartości wskaźnika, które pozwoliły przypisać stopień zagrożenia: niski/wysoki lub niski/informowania/alarmowy. Do tego celu określiłem średnie stosunki poszczególnych związków WWA względem średniego stężenia Benzo[*a*]pirenu w próbkach pobranych w Krakowie oraz Wadowicach. Posiadając informacje dotyczące stosunków miedzy stężeniami tych związków, wyznaczyłem stężenia WWA odpowiadające stężeniu Benzo[*a*]pirenu równemu 1 ng/m^3. Otrzymałem dane przedstawiające rozkład stężeń WWA w powietrzu. Następnie obliczyłem wartości wskaźników dla otrzymanych danych: *MEQ* – 2,33 ng/m^3, *CEQ* – 6,00 ng/m^3, *TEQ* – 11,46 pg/m^3. Testowano 2 przypadki z podziałem zakresów zmienności wskaźników na 2 i 3 przedziały wartości:

- 2 kategorie:
  - Niski stopień narażenia: *MEQ* < 2,33 ng/m^3, *CEQ* < 6,00 ng/m^3, *TEQ* < 11,46 pg/m^3,
  - Wysoki stopień narażenia: *MEQ* ≥ 2,33 ng/m^3, *CEQ* ≥ 6,00 ng/m^3, *TEQ* ≥ 11,46 pg/m^3.
- 3 kategorie:
  - Niski stopień narażenia: *MEQ* < 1,16 ng/m^3, *CEQ* < 3,00 ng/m^3, *TEQ* < 5,73 pg/m^3,
  - Poziom informowania: 1,16 ng/m^3 ≥ *MEQ* < 3,49 ng/m^3, 3,00 ng/m^3 ≥ *CEQ* < 9,00 ng/m^3, 5,73 ng/m^3 ≥ *TEQ* < 17,19 pg/m^3,
  - Poziom alarmowy: *MEQ* ≥ 3,49 ng/m^3, *CEQ* ≥ 9,00 ng/m^3, *TEQ* ≥ 17,19 pg/m^3.

Ze względu na dużą złożoność wieloczynnikowych ocen stopnia narażenia zdrowia i życia ludzkiego, dużą liczbę kombinacji wielu czynników meteorologicznych i stężeniowych, interakcje pomiędzy nimi, niewystarczającą ilość dostępnych danych (krótki okres pobierania próbek w porównaniu z dostępnymi bazami danych, np.: GIOŚ) zwłaszcza w przypadku braku uwzględniania wpływu pojedynczych związków WWA na zdrowie ludzkie, przyjęto pewne uproszczenia oraz podstawowe parametry/dane wejściowe do modelu:

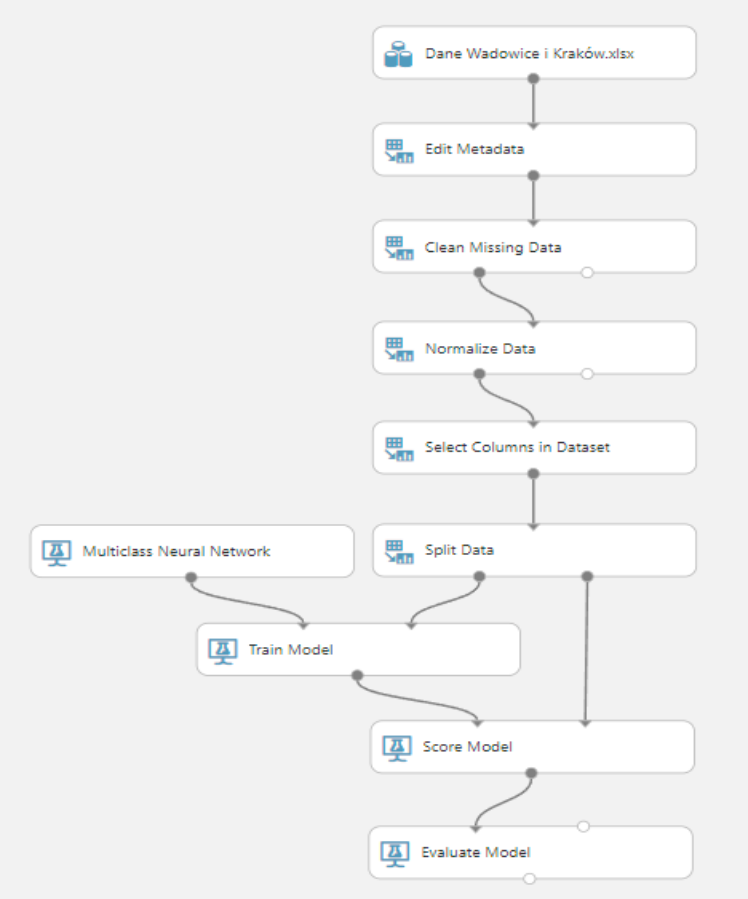
- stężenia średnie dobowe pyłu zawieszonego PM_{10} pobranego w Krakowie (2020-2021) oraz Wadowicach (2017) – dla Krakowa przyjęto wartości średnie dobowe jako sumę stężeń próbek 11h z każdego dnia lub próbki 23h. Dla Wadowic były to próbki 23h,
- wartości MEQ, CEQ i TEQ – dla każdego dnia pomiaru i średnich dobowych wartości stężeń WWA obliczono poszczególne wskaźniki narażenia – okresy uśredniania przyjęto tak jak w przypadku PM_{10},
- kategorie stopnia narażenia – zarówno dla 2 kategorii (niski, wysoki) jak i 3 kategorii (niski, informowania, alarmowy) przyjęto odpowiednie zakresy *MEQ*, *CEQ* i *TEQ*. Kategorie odnoszą się do wartości średnich uzyskanych dla stężenia Benzo[*a*]pirenu równego 1 ng/m^3 i proporcjonalnych wartości stężeń pozostałych WWA wyliczonych tylko dla 2 miejscowości (Kraków, Wadowice). Kategorie oszacowano względem jedynego związku WWA Benzo[*a*]pirenu, którego stężenie jest monitorowane i objęte normami.
- parametry meteorologiczne, wartości średnich dobowych z dnia pomiaru:
  - kierunek [N, E, S, W, NE, SE, SW, NW] i prędkość wiatru [m/s]
  - temperatura powietrza [°C]
  - ciśnienie atmosferyczne [hPa]
  - opad atmosferyczny [mm]

- w celu wykrycia dodatkowych zależności powiązanych z cyklem sezonowym, przypisano osobną kategorię – pora roku – dla zebranych próbek,
- w celu sprawdzenia wpływu miejsca pobierania na dokładność modelu, dodano kategorię – miejsce pobierania – przypisując do niej Kraków oraz Wadowice.

W pierwszej kolejności przeprowadzono testy w celu wybrania, który scenariusz podziału (2 lub 3 kategorie) będzie lepszy w przewidywaniu stopnia narażenia oraz jaki algorytm uczący pozwoli uzyskać większa dokładność. Do tego testu wykorzystano algorytmy:

- Sieci neuronowych (liczba ukrytych warstw – 1, liczba ukrytych węzłów – 100, szybkość uczenia się 0,1, ilość iteracji – 100, L_1 – 0,1)
- Regresji logistycznej (tolerancja optymalizacji – e^{-7}, L_1 – 1, L_2 – 2)
- Lasów losowych (liczba drzew decyzyjnych – 15, maksymalna głębokość drzew – 32, liczba losowych podziałów na węzły – 128, minimalna liczba próbek na węzeł – 1)
- Wektorów nośnych (dla 3 kategorii zastosowano dodatkowy moduł Azure Machine Learning – *One-vs-All Multiclass* – pozwalający zastosować algorytm dla więcej niż 2 kategorii) (liczba iteracji – 1, L_1 – 0,001),

Na rysunku 25 przedstawiono przykładowy schemat modelu stworzonego w Azure Machine Learning.



Rysunek 25. Przykładowy schemat modelu Azure Machine Learning

Wykorzystywany w badaniach model opierał się na schemacie blokowym, który zawierał następujące moduły:

a. Dane wejściowe – są to wszystkie zebrane dane (meteorologiczne, stężenie PM_{10}, stężenie B[*a*]P, wskaźniki *MEQ*, *CEQ*, *TEQ*), które model będzie wykorzystywał we wszystkich procesach. Dane wejściowe zostały przygotowane w formacie csv.,
b. Edit Metadata – edycja rodzaju danych – kategoryzacja danych. Istotnym jest czy informacja ma charakter ilościowy (np.: stężenie, opad atmosferyczny, temperatura), czy jakościowy (np.: kierunek wiatru – N, E, S, W, wskaźnik *MEQ*, *CEQ*, *TEQ* – wysoki, niski). Pliki csv. nie zawierają informacji o typie danych – zostaje on przypisany na tym etapie do konkretnych kolumn w zestawie danych,
c. Clean Missing Data – usunięcie lub zastąpienie brakujących danych w zestawie danych wejściowych,
d. Normalize Data – normalizacja danych – procedura wstępnej obróbki danych w celu dostosowania sygnałów do zakresu zmienności funkcji aktywacji oraz uniknięcia przesycenia sieci,
e. Select Columns in Dataset – wybór danych do dalszej analizy – określenie parametrów, które model będzie uwzględniał w analizach,
f. Split Data – podział danych wybranych na etapie Select Columns in Dataset w odpowiednich proporcjach na dane treningowe i testowe: 50/50, 60/40, 70/30, 80/20.
g. Algorytm uczenia maszynowego – odpowiedni algorytm względem, którego model będzie wykonywał proces uczenia/trenowania. Z racji stale rozpowszechniającego się zastosowania sztucznych sieci neuronowych, to właśnie ten algorytm został wykorzystany w modelach. Dodatkowo, z przeglądu literaturowego [Masood i Ahmad 2020, Suleiman i in. 2016, Suleiman i in. 2019] wynika, że wysoką dokładnością modelu charakteryzowały się również algorytmy: regresja logistyczna, lasy losowe, wektory nośne,
h. Train Model – uczenie się modelu na podzielonym wcześniej zbiorze danych przy użyciu wybranego algorytmu oraz konkretnej zmiennej, którą model będzie przewidywał,
i. Score Model – testowanie modelu – podłączenie danych, które zostały przeliczone do danych, które nie zostały jeszcze uwzględnione w modelowaniu w celu oceny zdolności do generalizacji przypadków. W przypadku modeli klasyfikacyjnych moduł generuje przewidywani. wartość dla klasy, jak również prawdopodobieństwo przewidywanej wartości. W przypadku modeli regresyjnych, moduł generuje tylko przewidywaną wartość liczbową.
j. Evaluate Model – rezultat – przegląd i analiza danych wyjściowych modelu.

Przed przystąpieniem do analiz, na etapie edycji danych, nadano parametrom takim jak Miejsce, Pora roku, Kierunek wiatru oraz Wskaźniki narażenia *MEQ*, *CEQ* i *TEQ* własności kategorii aby model nie traktował parametrów jako liczby, a jako osobne kategorie. Następnie, za pomocą modułu *Clean Missing Data* usunięto puste wiersze oraz dokonano normalizacji danych – *Normalize Data*. Wpływ i rodzaj normalizacji danych przedstawiono w dalszej części pracy – na tym etapie przyjęto normalizację ZScore – rodzaj normalizacji zmiennej losowej, w wyniku której zmienna otrzymuje średnią wartość oczekiwaną zero i odchylenie standardowe jeden. Z racji tego, że model mógł dokonywać w tym samym czasie obliczeń dla tylko jednego wskaźnika narażenia, w module *Select Columns in Dataset* wybrano najpierw wszystkie parametry dotyczące *MEQ*, następnie *CEQ*, a na końcu *TEQ*. Należy zaznaczyć, że stężenie Benzo[*a*]pirenu było uwzględniane jako jedyny WWA w zestawie parametrów, ponieważ tylko

względem tego związku planowane jest określanie stopnia narażenia. W kolejnym etapie (*Split Data*), podzielono wybrane dane na zbiory treningowe oraz testowe oraz przypisano do układu odpowiedni algorytm. Podział trening/test wynosił 50/50, 60/40 i 70/30. Końcowym etapem było oszacowanie jakości modelu za pomocą modułu *Score Model* oraz *Evaluate Model*. Zebrane wyniki przedstawiono w postaci macierzy błedów (tabela 16-23). Macierz błędów, inaczej określana jako tablica pomyłek (*confusion matrix*), stanowi fundament pod szereg metryk pozwalających ocenić jakość modelu. Rozmiar macierzy definiowany jest przez liczbę klas. Postać procentowa macierzy zawiera odsetek obserwacji, które klasyfikowane są do poszczególnych klas. Przypisanie tej samej kategorii przewidywanej względem kategorii rzeczywistej określa się mianem wyniku prawdziwego (np.: niski-niski, alarmowy-alarmowy). Natomiast przypisanie do kategorii rzeczywistej np.; poziom niski, kategorii przewidywanej np.: poziom wysoki określa się jako wynik fałszywie pozytywny (błąd pierwszego rodzaju, a przypisanie do kategorii rzeczywistej np.: poziom alarmowy, kategorii przewidywanej np.: poziom informowania/niski to wynik fałszywie negatywny (błąd drugiego rodzaju). W macierzy błędów przekątna (w tabelach poniżej) dla kategorii rzeczywistej i przewidywanej powinna przyjąć kolor zielony – wartości zbliżone do 100 %.

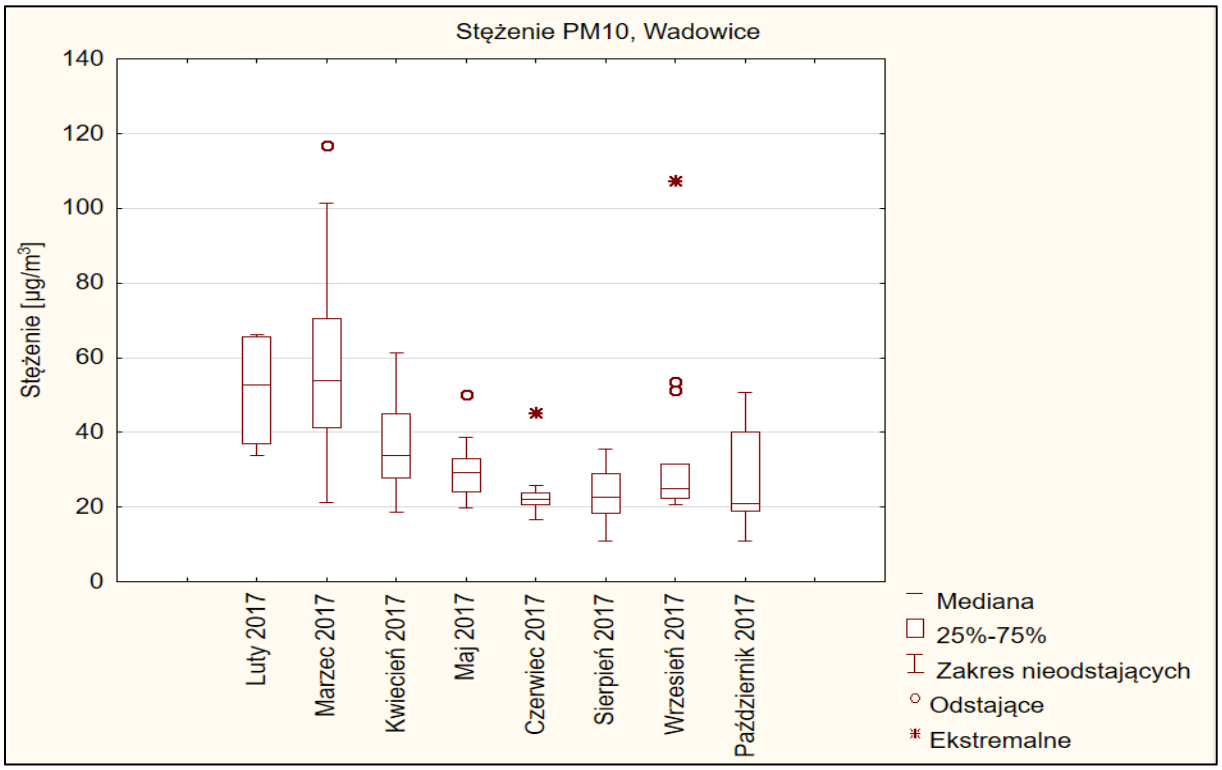
3. Wyniki badań

3.1. Pył zawieszony PM_{10}

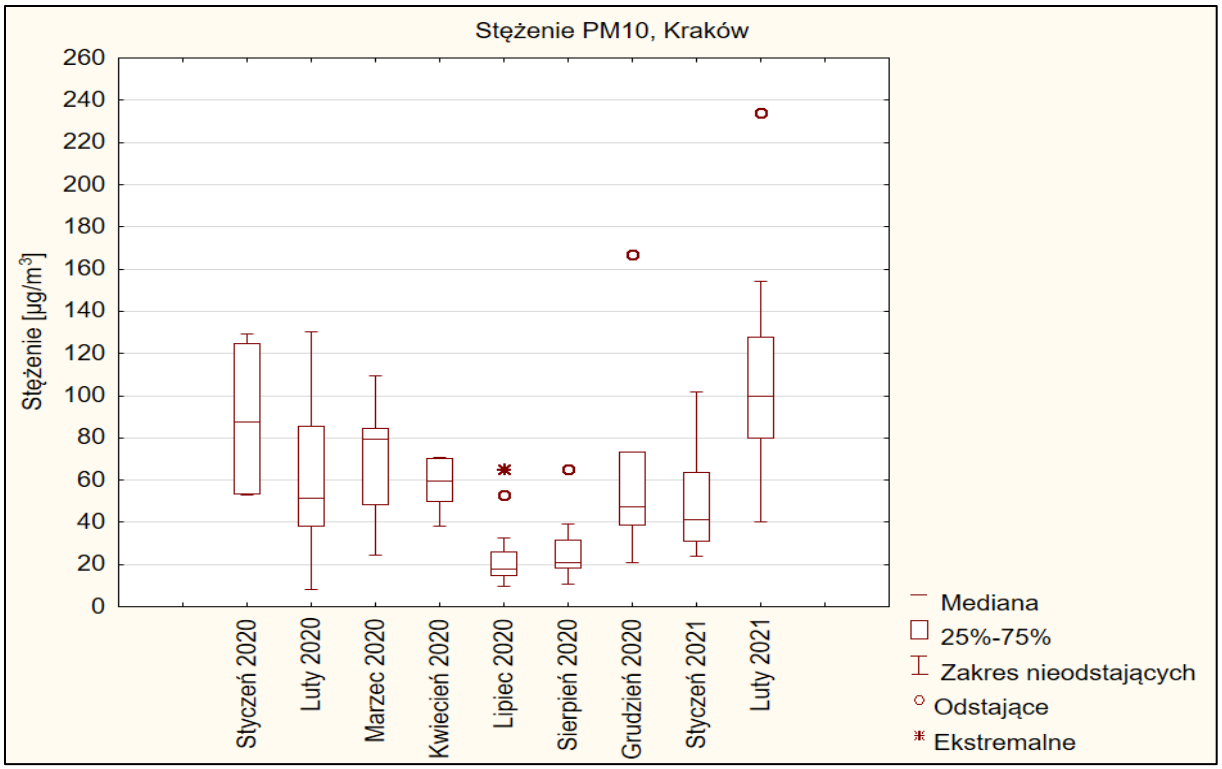
Rozkład średnich miesięcznych stężeń pyłu zawieszonego PM_{10} dla Wadowic oraz Krakowa przedstawiono na rysunkach 25 i 26. Aby lepiej oddać rozkład statystyczny wyników, sporządzono również wykresy zawierające zestaw danych dla okresów grzewczych i poza grzewczych w punktach pomiarowych, które zaprezentowane są na rysunkach 27 i 28.

Obliczone wartości średnie dla poszczególnych miesięcy mogą być niereprezentatywne ze względu na niewielką ilość dni pomiarowych w danym miesiącu. Wykresy dla Wadowic (rysunek 25, 27 i 29) zostały sporządzone z uwzględnieniem 5 dni pobierania PM_{10} w lutym, 27 dni w marcu, 25 dni w kwietniu i maju, 13 dni w czerwcu, 30 dni w sierpniu, 14 dni we wrześniu oraz 16 dni w październiku. Dla Krakowa sporządzono wykresy (rysunek 26, 28 i 30) z wykorzystaniem 5 dni pobierania próbek w styczniu 2020, 17 dni w lutym 2020, 6 dni w marcu, 7 dni w kwietniu, 16 dni w lipcu, 13 dni w sierpniu, 7 dni w grudniu i 13 dni w styczniu 2021 i lutym 2021.

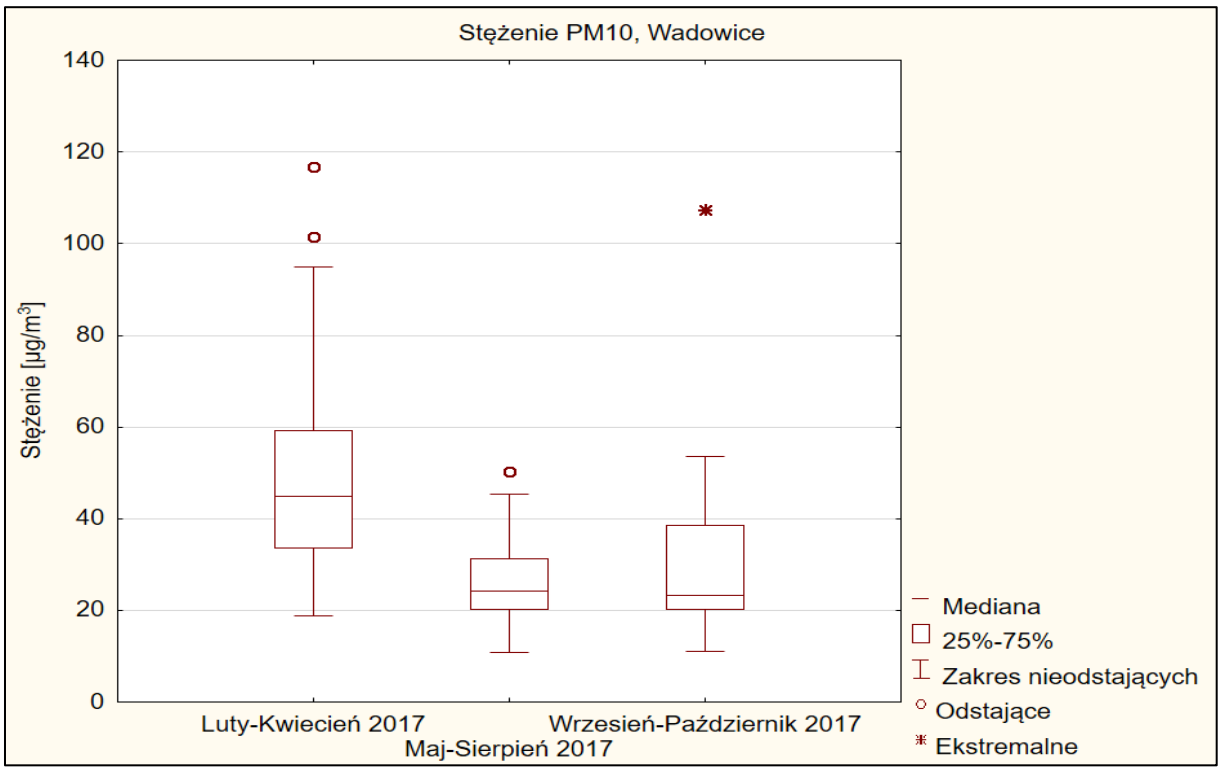
Dane odstające to wartości, które wykraczają poza rozstęp międzykwartylowy mediany *RQ* o wartość równą *Q1*-1,5*RQ* oraz *Q*3+1,5*RQ*. Natomiast dane ekstremalne to takie, które wykraczają o wartość *Q1*-3*RQ* oraz *Q*3+3*RQ*. Kwartyl pierwszy *Q1* dzieli dane w stosunku równym 25/75 – 25 % obserwacji jest niższa lub równa wartości *Q1*, 75 % obserwacji jest równa bądź większa od wartości *Q1*. Kwartyl drugi *Q2*, zwany medianą, dzieli obserwacje na dwie części w stosunku równym 50/50. Natomiast kwartyl trzeci *Q3* dzieli obserwacje w stosunku równym 75/25 – 75 % obserwacji jest niższa lub równa wartości *Q3*, 25 % obserwacji jest równa bądź większa od wartości *Q3*.



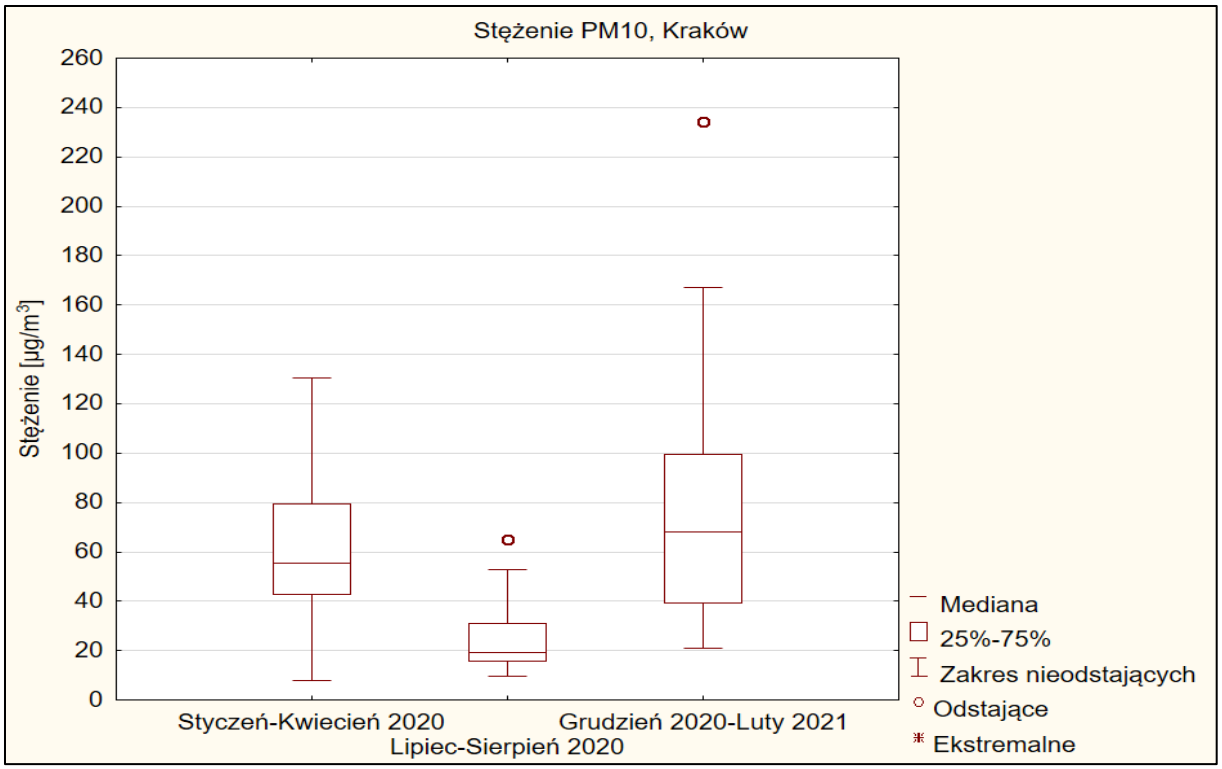
Rysunek 26. Rozkład stężeń PM_{10} dla Wadowic 2017. Wartości uzyskane dla poszczególnych miesięcy pomiarowych



Rysunek 27. Rozkład stężeń PM_{10} dla Krakowa 2020/2021. Wartości uzyskane dla poszczególnych miesięcy pomiarowych



Rysunek 28. Rozkład stężeń PM_{10} dla Wadowic (2017) z uwzględnieniem sezonu grzewczego i poza grzewczego



Rysunek 29.Rozkład stężeń PM_{10} dla Krakowa (2020/2021) z uwzględnieniem sezonu grzewczego i poza grzewczego

Dla Wadowic najwyższe stężenie PM_{10} wynosiło 116,77 µg/m^3 (13.03.2017), najniższe 10,80 µg/m^3 (19.08.2017). Odnotowane stężenia PM_{10} w pierwszym sezonie grzewczym (luty-kwiecień 2017) były w granicach 18,76 µg/m^3 (19.04.2017) – 116,77 µg/m^3 (13.03.2017), średnia dobowa wartość stężenia PM_{10} w tym okresie wyniosła 47,97 µg/m^3. W drugim sezonie grzewczym (wrzesień-październik) zakres ten wynosił 10,94 µg/m^3 (06.10.2017) – 107,29 µg/m^3 (20.09.2017), natomiast średnie dobowe stężenie PM_{10} było równe 30,56 µg/m^3. Sezon poza grzewczy (maj-sierpień), o średnim dobowym stężeniu PM_{10} równym 25,65 µg/m^3, charakteryzował się zakresem stężeń PM_{10} od 10,80 µg/m^3 (19.08.2017) do 50,26 µg/m^3 (10.05.2017). Największe średnie miesięczne stężenia PM_{10} odnotowano w marcu (58,22 µg/m^3), najniższe w sierpniu (23,31 µg/m^3).

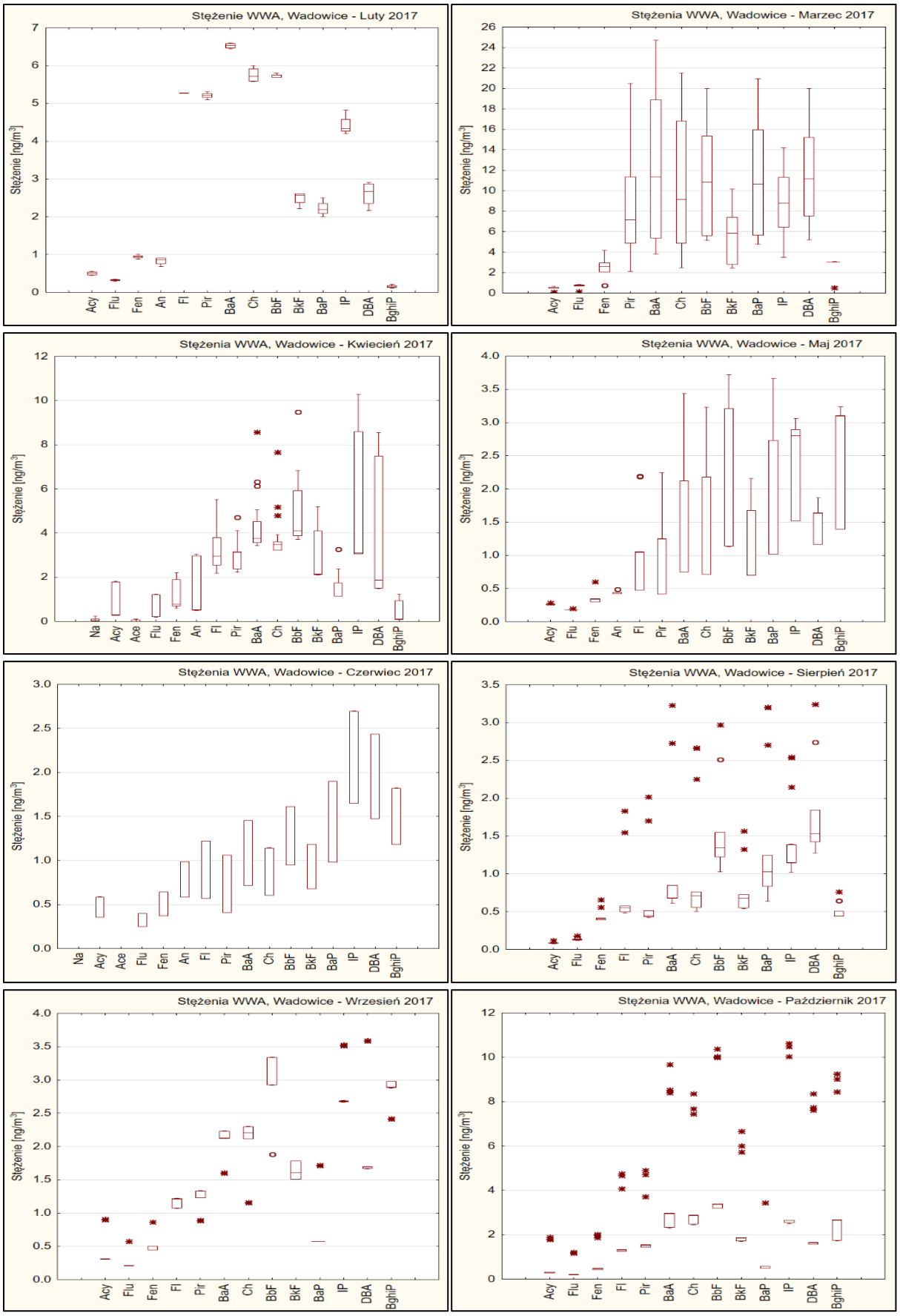
Dla Krakowa najwyższe stężenie PM_{10} odnotowano 24.02.2021, najniższe 11.02.2020, odpowiednio 234,14 µg/m^3 i 8,06 µg/m^3. Stężenie pyłu w sezonie grzewczym 2020 (styczeń-kwiecień) wynosiło od 8.06 µg/m^3 (11.02.2020) do 130,38 µg/m^3 (08.02.2020) ze średnią dobową wartością stężenia PM_{10} równą 65,01 µg/m^3. W sezonie grzewczym 2020/2021 (grudzień 2020, styczeń i luty 2021) wartości te znajdowały się w przedziale od 20,87 µg/m^3 (13.12.2020) do 234,14 µg/m^3 (24.02.2021), a średnie dobowe stężenie PM_{10} wyniosło 75.03 µg/m^3. Poza sezonem grzewczym (lipiec, sierpień 2020) zakres stężeń PM_{10} wynosił 9,64 µg/m^3 (23.02.2020) – 65,12 µg/m^3 (29.07.2020), natomiast średnie dobowe stężenie 23,29 µg/m^3. Najwyższe średnie miesięczne stężenie pyłu zawieszonego PM_{10} odnotowano w lutym 2021 – 106,18 µg/m^3, najniższe w lipcu 23,29 µg/m^3.

W badaniach [Samek i in. 2021] przeprowadzonych dla Krakowa w latach 2018/2019 dla dwóch stacji pomiarowych: Aleje Krasińskiego oraz Złoty Róg, wykazano podobne zróżnicowania sezonowe stężeń pyłu PM_{10}. Średnie dobowe stężenia PM_{10} były dwukrotnie wyższe w sezonie zimowym w porównaniu z sezonem letnim. W okresie letnim, średnie dobowe stężenia PM_{10} nie przekraczały 50 µg/m^3, natomiast w sezonie zimowym największe odnotowane wartości średnich dobowych stężeń pyłu wynosiły 157 µg/m^3 oraz 132 µg/m^3,

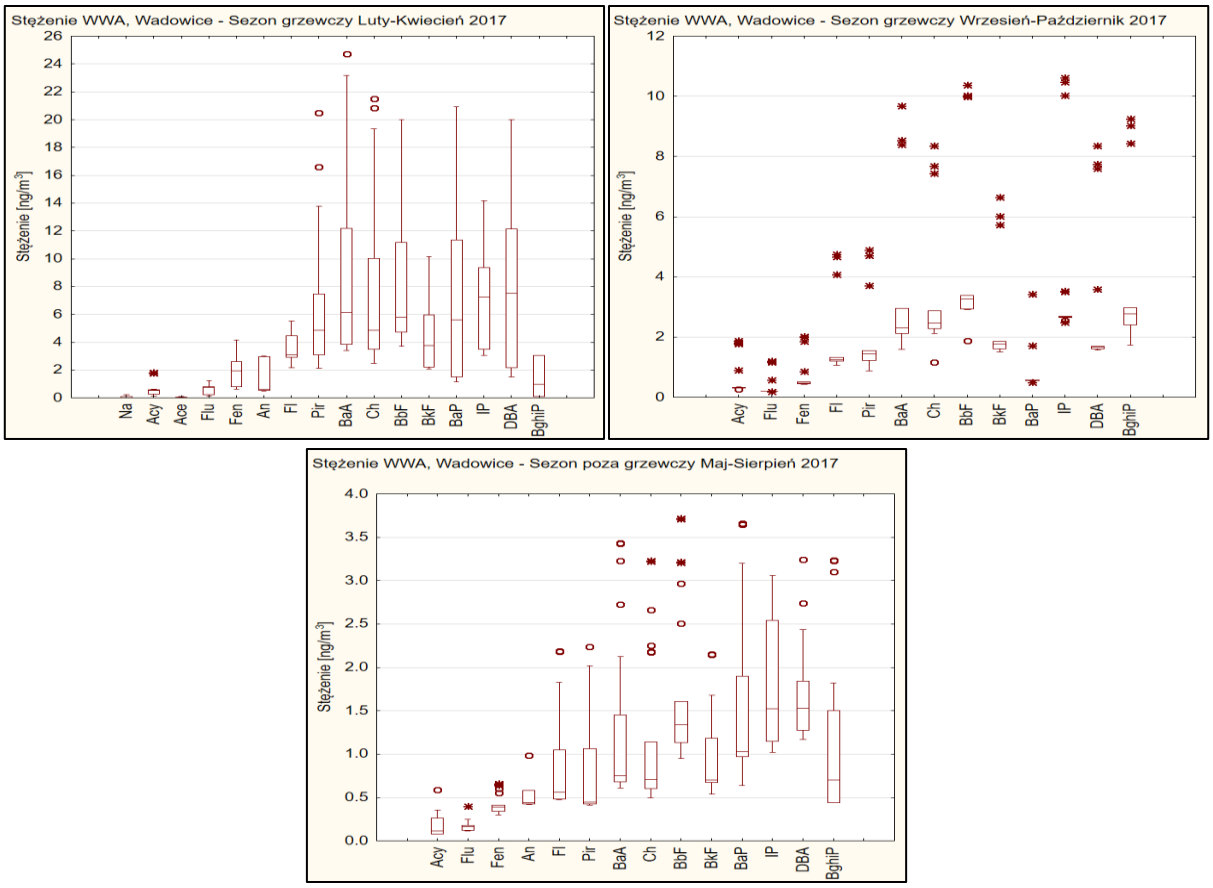
odpowiednio dla stacji pomiarowej Aleje Krasińskiego oraz Złoty Róg. Badania [Styszko i in. 2016] dotyczące stężeń PM_{10} w Krakowie przeprowadzone w 2014 roku wykazały najwyższe średnie dobowe stężenia w sezonie zimowym: 03.02.2014 – 136,6 $\mu g/m^3$, 04.02.2014 – 153,8 $\mu g/m^3$ oraz 06.02.2014 – 112,9 $\mu g/m^3$. Najniższe dobowe stężenie odnotowano w lecie – 23,5 $\mu g/m^3$. W artykule [Sekuła i in. 2021] przedstawiono procentowy udział dni, w których dopuszczalne średnie dobowe stężenie pyłu zawieszonego PM_{10} (50 $\mu g/m^3$) zostało przekroczone. Badania wykonano w okresach wrzesień 2017 – kwiecień 2018 oraz wrzesień 2018 – kwiecień 2019 dla 7 stacji pomiarowych znajdujących się na terenie Krakowa. Dla stacji pomiarowej Aleje Krasińskiego dopuszczalne normy średniego dobowego stężenia PM_{10} zostały przekroczone w ciągu 275 dni, co stanowi 75 % wszystkich dni pomiarowych dla tej stacji. Dla stacji Piastów było to 127 dni, Wadów – 115 dni, Złoty Róg – 176 dni, Kurdwanów – 163 dni, Dietla – 185 dni, Bulwarowa – 151 dni. Najwyższe średnie dobowe stężenie PM_{10} odnotowano na stacji Aleje Krasińskiego 05.03.2018, wynosiło 224 $\mu g/m^3$. W badaniach składu PM_{10} [Turek-Fijak i in. 2021] przeprowadzonych dla miasta Skała, znajdującego się około 20 km na północ od Krakowa, w latach 2017/2018 odnotowano średnie dobowe stężenie PM_{10} w zakresie 23 $\mu g/m^3$ - 301 $\mu g/m^3$, natomiast średnie stężenie w okresie pomiarowym wynosiło 84 $\mu g/m^3$. Przez cały okres pobierania próbek przez tylko 11 dni średnie dobowe stężenie PM_{10} nie przekraczało średnich dobowych norm stężenia PM_{10} (50 $\mu g/m^3$). Wysokie stężenia pyłów odnotowywane są również dla województwa Śląskiego [Kaleta i Kozielska 2023]. W latach 2018-2021 Dla Dąbrowy Górniczej zarejestrowano 113 dni z przekroczonymi dopuszczalnymi normami PM_{10} w 2018 roku, 69 dni w 2019 roku, 49 dni w 2020 roku i 65 dni w 2021 roku. Odpowiednio dla Częstochowy było to 72 dni, 39 dni, 20 dni i 40 dni. Dla Katowic liczby te wynosiły 93 dni (2018), 72 dni (2019), 48 dni (2020) i 52 dni (2021). Największa ilość dni ze stężeniem większym niż 50 $\mu g/m^3$ zaobserwowano w Pszczynie: 2018 – 139 dni, 2019 – 116 dni, 2020 – 96 dni oraz 2021 – 94 dni.

3.2. Wielopierścieniowe węglowodory aromatyczne WWA

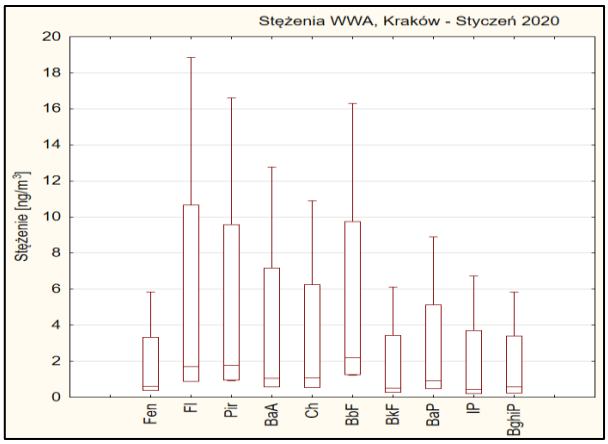
Rozkład stężeń wielopierścieniowych węglowodorów aromatycznych dla Wadowic i Krakowa został przedstawiony na rysunkach 30 i 32. Rysunki 31 i 33 prezentują rozkład stężeń WWA z uwzględnieniem sezonu grzewczego i poza grzewczego, odpowiednio dla Wadowic i Krakowa.

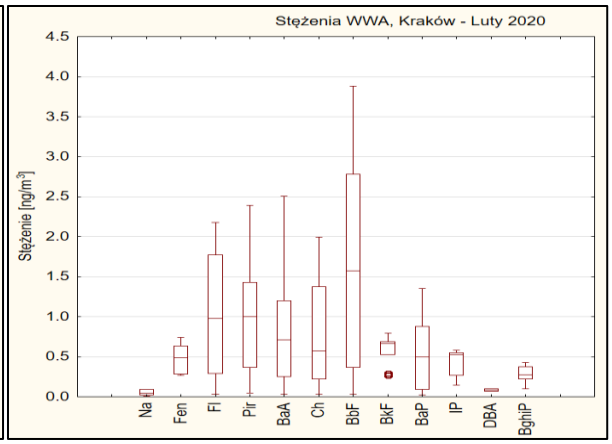


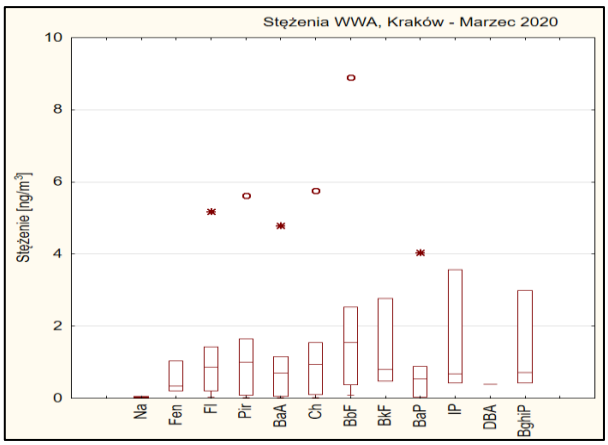
Rysunek 30. Rozkład stężeń WWA dla Wadowic (2017)
(– - mediana, □ - wartości z przedziału 25%-75%,
*T i ⊥ - wartości nieodstające, ○ - wartości odstające, * - wartości ekstremalne)*

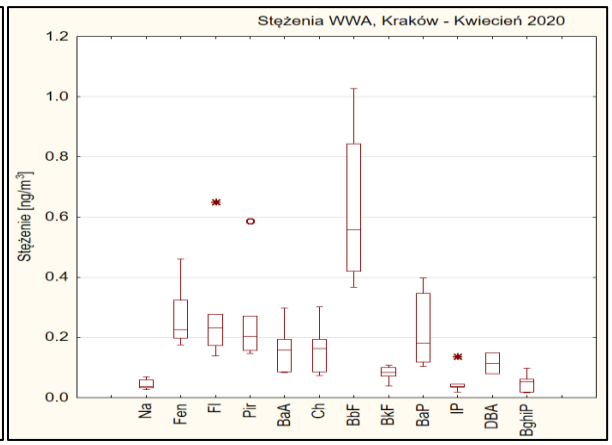


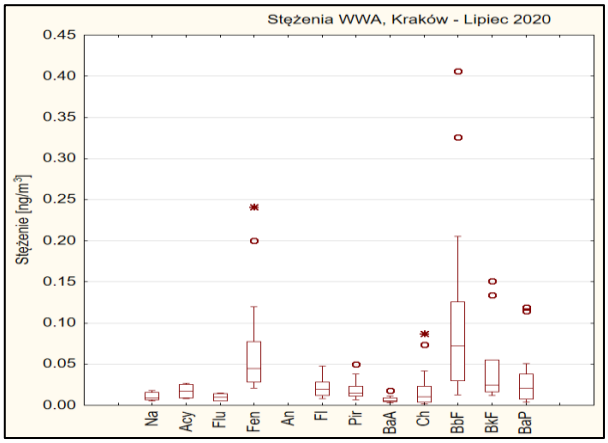
*Rysunek 31. Rozkład stężeń WWA dla Wadowic (2017) z uwzględnieniem sezonu grzewczego i poza grzewczego (– - mediana, □ - wartości z przedziału 25%-75%, T i ⊥ - wartości nieodstające, ○ - wartości odstające, * - wartości ekstremalne)*

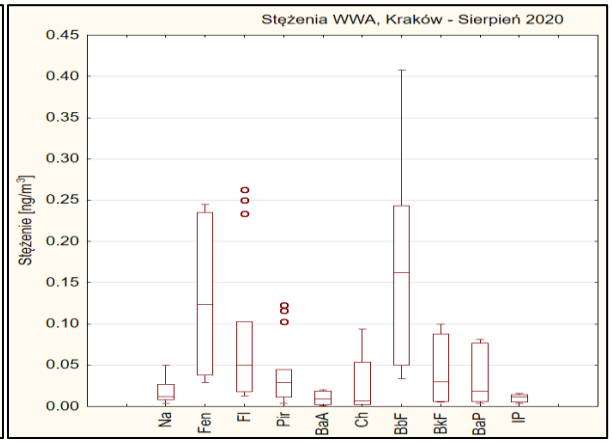

Stężenia WWA, Kraków - Styczeń 2020
Stężenie [ng/m3]
Fen
Fl
Pir
BaA
Ch
BbF
BkF
BaP
IP
BghiP

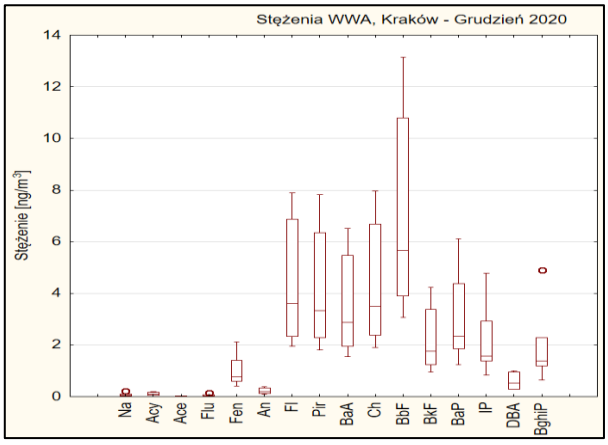

Stężenia WWA, Kraków - Luty 2020
Stężenie [ng/m3]
Na
Fen
Fl
Pir
BaA
Ch
BbF
BkF
BaP
IP
DBA
BghiP

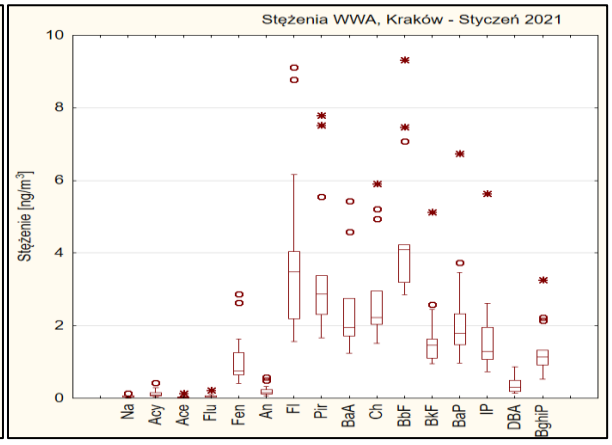

Stężenia WWA, Kraków - Marzec 2020
Stężenie [ng/m3]
Na
Fen
Fl
Pir
BaA
Ch
BbF
BkF
BaP
IP
DBA
BghiP

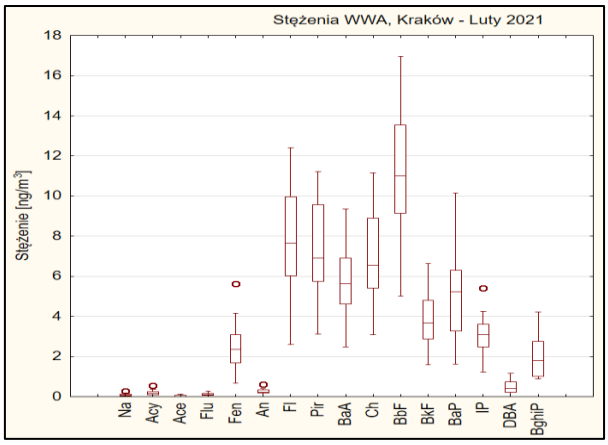

Stężenia WWA, Kraków - Kwiecień 2020
Stężenie [ng/m3]
Na
Fen
Fl
Pir
BaA
Ch
BbF
BkF
BaP
IP
DBA
BghiP


Stężenia WWA, Kraków - Lipiec 2020
Stężenie [ng/m3]
Na
Acy
Flu
Fen
An
Fl
Pir
BaA
Ch
BbF
BkF
BaP

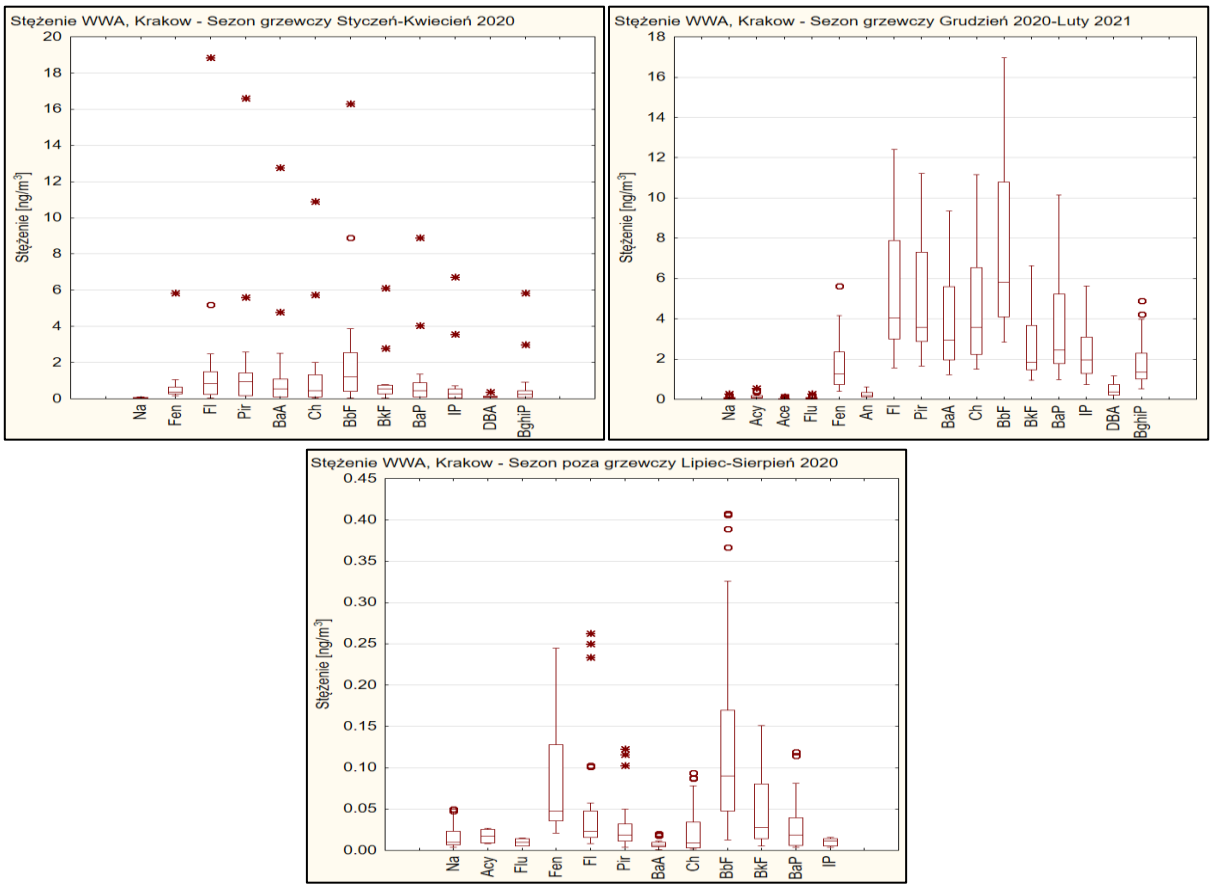

Stężenia WWA, Kraków - Sierpień 2020
Stężenie [ng/m3]
Na
Fen
Fl
Pir
BaA
Ch
BbF
BkF
BaP
IP


Stężenia WWA, Kraków - Grudzień 2020
Stężenie [ng/m3]
Na
Acy
Ace
Flu
Fen
An
Fl
Pir
BaA
Ch
BbF
BkF
BaP
IP
DBA
BghiP


Stężenia WWA, Kraków - Styczeń 2021
Stężenie [ng/m3]
Na
Acy
Ace
Flu
Fen
An
Fl
Pir
BaA
Ch
BbF
BkF
BaP
IP
DBA
BghiP



Rysunek 32. Rozkład stężeń WWA dla Krakowa (2020/2021)
(– - mediana, □ - wartości z przedziału 25%-75%,
*T i ⊥ - wartości nieodstające, ○ - wartości odstające, * - wartości ekstremalne)*



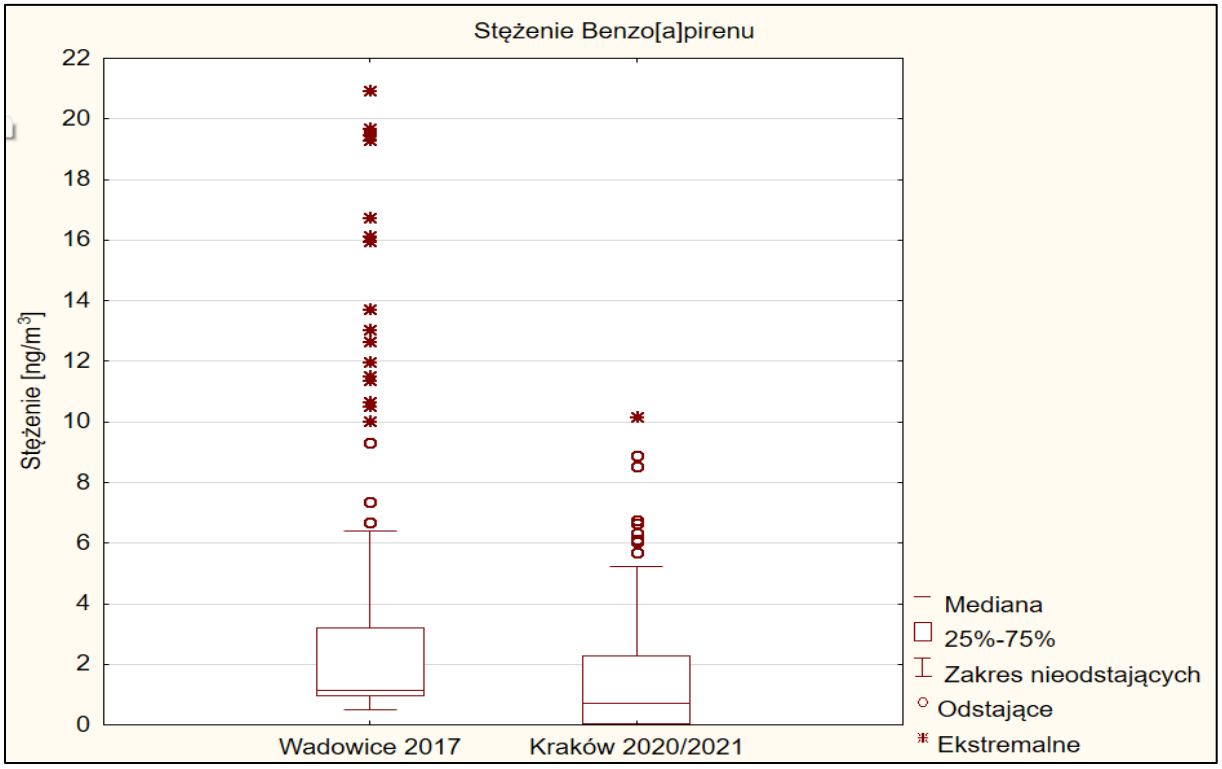
Rysunek 33.Rozkład stężeń WWA dla Krakowa (2020/2021) z uwzględnieniem sezonu grzewczego i poza grzewczego
(– - mediana, □ - wartości z przedziału 25%-75%,
*T i ⊥ - wartości nieodstające, ○ - wartości odstające, * - wartości ekstremalne)*

W oparciu o wyznaczone wartości stężeń frakcji PM_{10} pyłu zawieszonego, część filtrów dobowych została wytypowana do agregacji i poziom stężeń WWA został oznaczony w próbkach zagregowanych. Wartości dobowe WWA dla tych próbek zostały wyliczone na podstawie średniego stężenia WWA w próbce zagregowanej oraz stężenia pyłu w próbce dobowej.

W badaniach przeprowadzonych dla Wadowic zakres stężeń WWA wynosił od 0,08 ng/m^3 do 24,73 ng/m^3. Największe stężenia odnotowano dla Benzo[*a*]antracenu (24,73 ng/m^3 – 13.03.2017), Chryzenu (21,52 ng/m^3 – 13.03.2017) i Benzo[*a*]pirenu (20,93 ng/m^3 – 27.03.2017). Najwyższymi wartościami średniego dobowego stężenia w okresie grzewczym luty-kwiecień 2017 charakteryzował się Benzo[*a*]antracen (8,65 ng/m^3) oraz Benzo[*b*]fluoranten (8,32 ng/m^3), w okresie grzewczym wrzesień-październik 2017 był to Benzo[*b*]fluoranten (3,81 ng/m^3) oraz Indeno[*1,2,3-cd*]piren (3,47 ng/m^3). W przypadku sezonu poza grzewczego (maj-sierpień 2017) był to również Indeno[*1,2,3-cd*]piren (1,73 ng/m^3) oraz Benzo[*b*]fluoranten (1,70 ng/m^3).

Dla Krakowa, stężenia WWA w całym okresie pobierania próbek znajdowały się w przedziale od 0,01 ng/m^3 do 18,85 ng/m^3. Najwyższym dziennym stężeniem charakteryzował się Fenantren (18,85 ng/m^3 – 27.02.2020), Benzo[*b*]fluoranten (16,95 ng/m^3 – 24.02.20) oraz Piren (16,61 ng/m^3 – 27.02.2020). Największym średnim dobowym stężeniem w pierwszym (styczeń-kwiecień 2020), jak i drugim (grudzień 2020-luty 2021) sezonie grzewczym, dla Krakowa charakteryzował się Benzo[*b*]fluoranten (odpowiednio 2,06 ng/m^3 i 7,64 ng/m^3) oraz Fluoranten (1,55 ng/m^3 i 5,47 ng/m^3). W okresie poza grzewczym największe średnie dobowe stężenie przypadło również dla Benzo[*b*]fluorantenu (0,14 ng/m^3), a także dla Fenantrenu (0,09 ng/m^3). W okresie letnim wartości stężeń dla wielu związków były poniżej *LOQ* zarówna dla Wadowic, jak i Krakowa.

Rysunek 33 prezentuje rozkład stężenia Benzo[*a*]pirenu w Wadowicach oraz Krakowie.



Rysunek 34. Rozkład stężenia Benzo[a]pirenu w całym okresie pobierania próbek w Wadowicach (2017) i Krakowie (2020/2021)

Zakres stężeń dla Benzo[*a*]pirenu w Wadowicach wynosił 0,49 ng/m^3 (09.10.2027) – 20,93 ng/m^3 (27.03.2017), a wartość średniodobowa w okresie badań wyniosła 3,32 ng/m^3. Największe stężenie B[*a*]P oznaczono 27.03.2017 oraz 09.03.2017 i wynosiło, odpowiednio 20,93 ng/m^3 oraz 19,69 ng/m^3. Dla Krakowa średnie miesięczne stężenie Benzo[*a*]pirenu wyniosło 1,56 ng/m^3. Największe wartości oznaczono 24.02.2021 (10,16 ng/m^3) oraz

27.02.2020 (8.91 ng/m^3). Większość pomiarów mieści się w zakresie 25%-75% rozkładu. Występują jednak dni, w których stężenia znacznie wykraczają poza przedział. Podobne zmienności sezonowe zauważono dla wielu dużych miast w Polsce i na świecie. Wojewódzki Inspektorat Ochrony Środowiska wykazał w latach 2014-2018 stężenia Benzo[a]pirenu na wysokim poziomie dla wielu miast w Polsce. Najwyższe średnioroczne stężenie B[a]P odnotowano dla stacji w Brzeszczach (22,70 ng/m^3, 2017 rok), najniższe dla stacji w Muszynie-Złockiem (2,00 ng/m^3, 2018 rok). Wykazano przekroczone dopuszczalne normy B[a]P na terenach uzdrowisk. Na obszarach Aglomeracji Krakowskiej zakres średniorocznych stężeń Benzo[a]pirenu wynosił od 3,60 ng/m^3 do 8,30 ng/m^3, odpowiednio dla stacji Wadów (2017 rok) i stacji na ulicy Bulwarowej (2015 rok). Dla Małopolski maksymalne stężenia w poszczególnych latach wykazano na stacjach: Nowy Targ (15.20 ng/m^3, 2014 rok), Nowy Sącz (12,00 ng/m^3, 2015 rok), Nowy Sącz (9,70 ng/m^3, 2016 rok), Brzeszcze (22,70 ng/m^3, 2017 rok) oraz Nowy Targ (18,30 ng/m^3, 2018 rok) [Główny Inspektorat Ochrony Środowiska 2019]. W badaniach na terenie Górnego Śląska [Kozielska i in. 2013], na obszarach charakteryzujących się wysokim natężeniem ruchu zakres stężeń B[a]P wynosił 7,90 ng/m^3–11,10 ng/m^3. Analizy wykonane dla Warszawy i Gliwic [Rogula-Kozłowska i in. 2017] wykazały wysokie stężenia Benzo[a]pirenu również w pomieszczeniach zamkniętych: Warszawa – 1,11 ng/m^3, Gliwice –3,27 ng/m^3. W wielu badanych przypadkach Benzo[a]piren stanowi 1-20% wszystkich WWA [Jamhari i in. 2014, Kozielska i in. 2016, Wojewódzki Inspektorat Ochrony Środowiska 2014].

3.3. Profile źródeł zanieczyszczeń powietrza

W wielu analizowanych opracowaniach oraz w niniejszej pracy największy udział w pyle zawieszonym wykazują WWA zaliczane do tzw. cięższych węglowodorów (4 pierścienie i więcej) [Kozielska i in. 2013, Kozielska i in. 2016]. Obecność cięższych WWA w powietrzu świadczy o emisji zanieczyszczeń ze źródeł, takich jak transport (spalanie paliw w silnikach), ogrzewania domów paliwem o niskiej wartości opałowej [Jamhari i in. 2014, Simoneit 2015, Yunker i in. 2002]. Udział poszczególnych WWA oraz och wzajemne stosunki można wykorzystać do oszacowania pochodzenia pyłu zawieszonego. Stosunki stężeń WWA określane są jako wskaźniki diagnostyczne. Dla przykładu Fenantren, Fluoren i Piren uznawane są najczęściej za znaczniki spalania węgla. Benzo[a]piren i Fluoren są emitowane podczas spalania drewna. Fluoren, Piren, Benzo[b]fluoranten i Benzo[k]fluoranten są charakterystyczne dla spalania paliw w silnikach diesla [Yunker i in. 2002, Kulshrestha i in. 2019]. Analizując stężenia poszczególnych WWA, a także uwzględniając dane dotyczące wskaźników i ich konkretnych wartości lub zakresów z tabeli 2, w tabelach 10 i 11 przedstawiono procentowy udział źródeł emisji w oparciu o średnie dobowe stężenia WWA w próbkach pobranych, odpowiednio w Wadowicach 2017 i Krakowie 2020/2021. Wartości wykraczających poza zakresy wskaźników z tabeli 2, przyporządkowano do osobnej kategorii źródła (*inne*):

Tabela 10. Źródła zanieczyszczeń dla próbek PM_{10} względem wskaźników diagnostycznych dla Wadowic 2017

nr	Wskaźnik	Źródło	Udział źródła %
(1)	$\frac{Fluoren}{Fluoren + Piren}$	Silniki benzynowe Silniki diesla Inne	100 0 0
(2)	$\frac{Fluoranten}{Fluoranten + Piren}$	Spalanie paliwa (benzyna/diesel) Spalanie węgla/drewna Inne	79 21 0
(3)	$\frac{Benzo[b]fluoraten}{Benzo[k]fluoraten}$	Spalanie drewna Silniki benzynowe Przemysł (huty) Węgiel/koks Inne	0 12 0 0 88
(4)	$\frac{Piren}{Benzo[a]piren}$	Silniki benzynowe Silniki diesla Spalanie drewna Inne	5 4 10 81
(5)	$\frac{Benzo[a]piren}{Benzo[a]piren + Chryzen}$	Spalanie drewna Ogrzewanie domów (drewno/węgiel) Transport Inne	0 61 39 0
(6)	$\frac{Indeno[1,2,3-cd]piren}{Indeno[1,2,3-cd]piren + Benzo[ghi]perylen}$	Transport Silniki diesla Silniki benzynowe Gaz ziemny Spalanie oleju Spalanie węgla Spalanie drewna Inne	0 0 0 0 0 11 8 81
(7)	$\frac{Benzo[a]antracen}{Benzo[a]antracen + Chryzen}$	Transport Silniki diesla i benzynowe Inne	88 0 12

Tabela 11.Źródła zanieczyszczeń dla próbek PM_{10} względem wskaźników diagnostycznych dla Krakowa 2020/2021

nr	Wskaźnik	Źródło	Udział źródła %
(1)	$\frac{Fluoren}{Fluoren + Piren}$	Silniki benzynowe Silniki diesla Inne	92 4 4
(2)	$\frac{Fluoranten}{Fluoranten + Piren}$	Spalanie paliwa (benzyna/diesel) Spalanie węgla/drewna Inne	100 0 0
(3)	$\frac{Benzo[b]fluoraten}{Benzo[k]fluoraten}$	Spalanie drewna Silniki benzynowe Przemysł (huty) Węgiel/koks Inne	0 4 0 0 96
(4)	$\frac{Piren}{Benzo[a]piren}$	Silniki benzynowe Silniki diesla Spalanie drewna Inne	1 2 8 89
(5)	$\frac{Benzo[a]piren}{Benzo[a]piren + Chryzen}$	Spalanie drewna Ogrzewanie domów (drewno/węgiel) Transport Inne	0 65 29 6
(6)	$\frac{Indeno[1,2,3-cd]piren}{Indeno[1,2,3-cd]piren + Benzo[ghi]perylen}$	Transport Silniki diesla Silniki benzynowe Gaz ziemny Spalanie oleju Spalanie węgla Spalanie drewna Inne	0 0 0 0 0 41 11 48
(7)	$\frac{Benzo[a]antracen}{Benzo[a]antracen + Chryzen}$	Transport Silniki diesla i benzynowe Inne	61 11 28

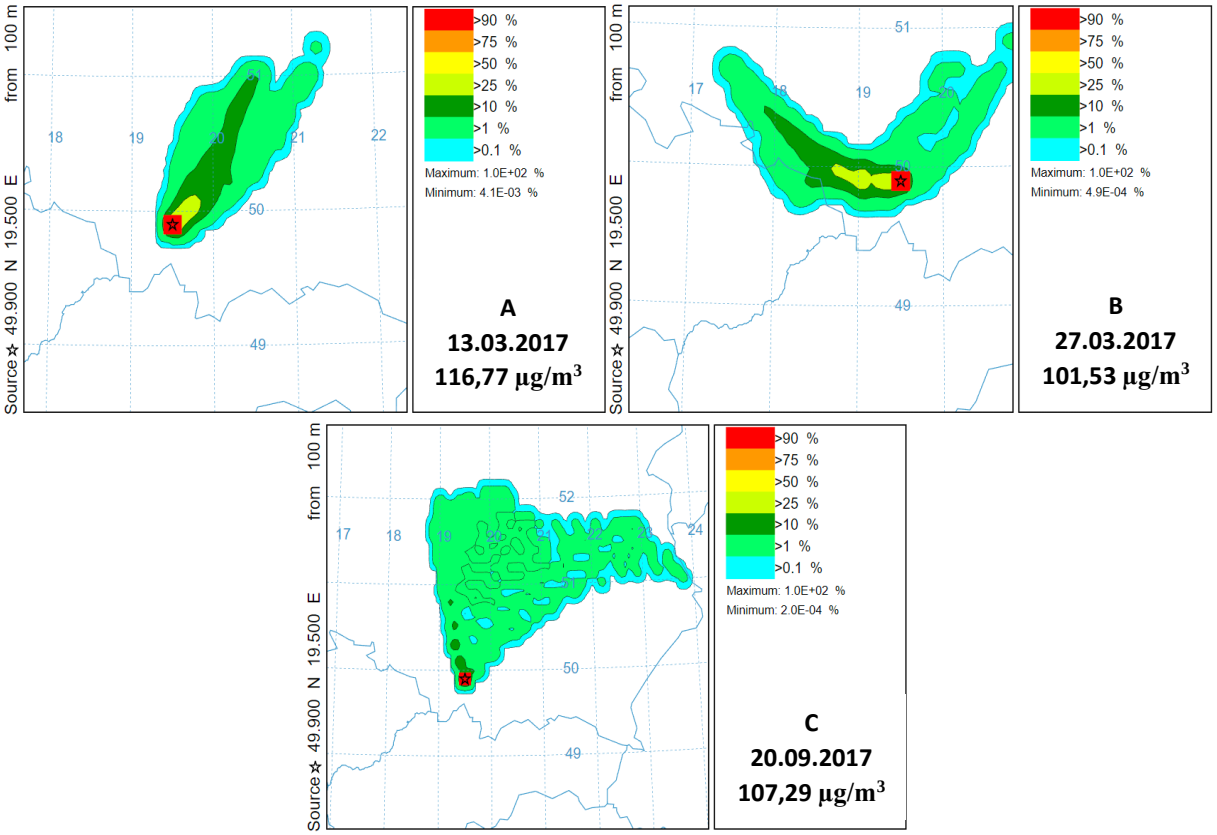
Dla Wadowic i Krakowa wykazano zbliżone wartości wskaźników diagnostycznych. Wskaźniki (1), (2) i (7) wskazują jako główne źródło zanieczyszczeń transport, w szczególności spalanie paliwa w silnikach spalinowych. Oba miejsca pobierania charakteryzowały się dużym natężeniem ruchu. Podobne obserwacje odnotowano dla miejsc ściśle związanych z transportem ulicznym [Simoneit 2015]. Należy jednak pamiętać, że stężenia Fluorenu w okresie grzewczym wyszły poniżej LOQ, z tego względu należy traktować wskaźnik z ostrożnością. Dla wskaźnika (3) nie uzyskano informacji dotyczących konkretnego źródła. 88 % wyników dla Wadowic i 96 % wyników dla Krakowa zostało zaklasyfikowanych jako inne, nieznane źródło. Podobne założenia otrzymano dla wskaźnika (4). Wskaźnik (5) ułatwia najczęściej identyfikacje źródeł takich jak spalanie drewna i węgla w celach grzewczych. W niniejszym opracowaniu wskazuje na dominacje źródeł ogrzewania gospodarstw domowych w okresie zimowym. Z racji tego, że znaczna część wyników dotyczy miesięcy chłodniejszych – 61 % wyników dla Wadowic oraz 65 % wyników dla Krakowa wskazuje na ogrzewanie domów jako główne źródło tych zanieczyszczeń. W pozostałym okresie (lato) dominuje transport. Zbliżone dane uzyskano dla (6), gdzie dla Krakowa około 41 % dni odpowiada spalaniu węgla, 11% dni spalaniu drewna. W przypadku Wadowic, spalanie węgla jako główne źródło badanych WWA odnotowano dla 11% próbek pyłu. Dla wskaźnika (6), nie można zidentyfikować z dużym prawdopodobieństwem pochodzenia zanieczyszczeń ze względu na wysoki procent udziału

nieznanego źródła dla obu miast. Dodatkową informacją o źródłach WWA dla Wadowic i Krakowa jest większe stężenie Benzo[*ghi*]perylenu od Dibenzo[*ah*]antracenu, co świadczy o wysokim udziale transportu w analizowanych próbkach, co również potwierdzają dane literaturowe dotyczące badań jakości powietrza w miejscach z dużym udziałem transportu ulicznego [Kozielska i in. 2016, Siudek i Frankowski 2018].

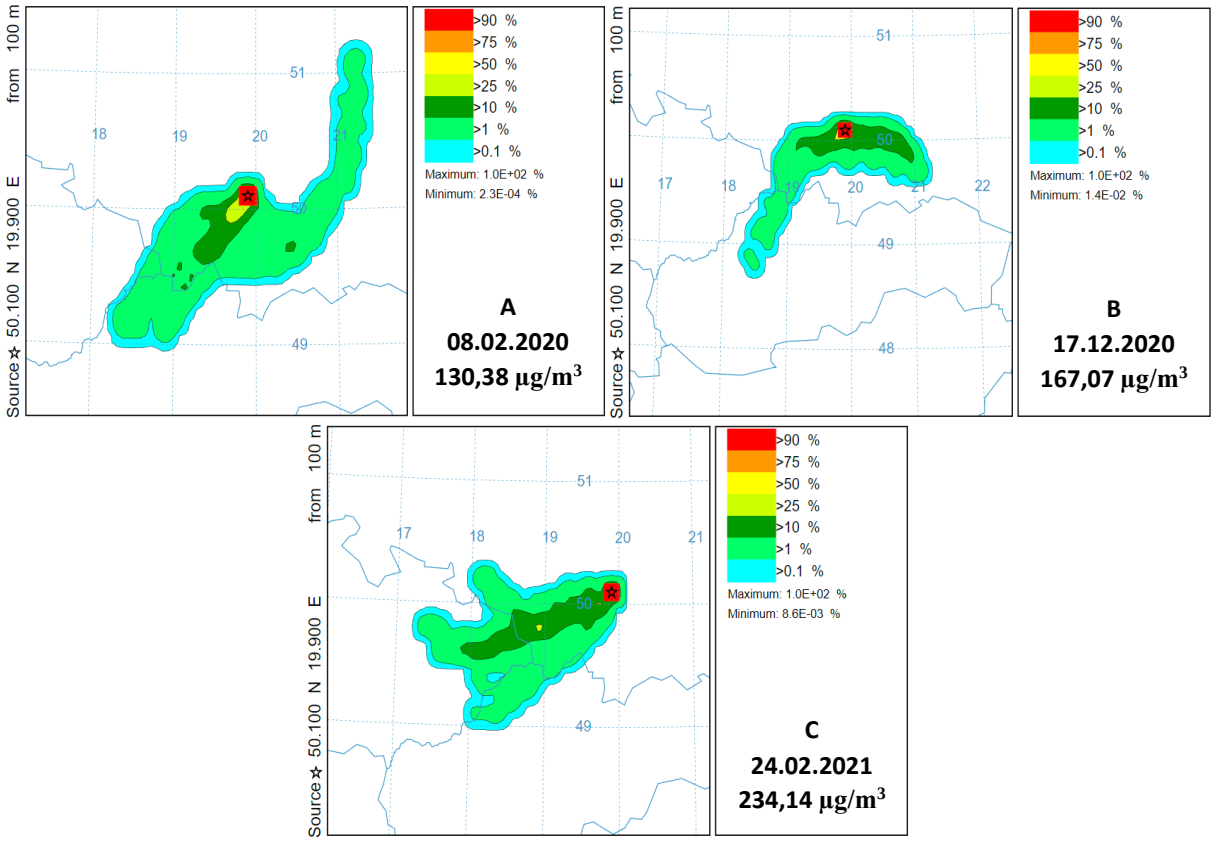
3.4. Ocena kierunku napływu zanieczyszczeń

Aby uzyskać informacje na temat możliwości transportu zanieczyszczeń z wybranych obszarów z okolic badanych miejsc, wykonano analizy częstotliwości występujących kierunków napływu mas powietrza. Analizy wykonano za pomocą modelu HYSPLIT NOAA Air Resources Laboratory (*Hybrid Single-Particle Lagrangian Integrated Trajectory model*) opracowanego przez NOAA Air Resources Laboratory (*National Oceanic and Atmospheric Administration*) [Draxler i Hess 1997, Draxler i Hess 1998, Draxler 1999, Rolph i in. 2017, Stein i in. 2015]. Ze względu na analizy wykonywane w skali regionalnej do zasilenia modelu wykorzystano dane przygotowywane przez operacyjny system prognoz globalnych GFS (*Global Forecasting System*) o rozdzielczości 0,25°/0,25° i 127 warstwach. Są to dane o najwyższej dostępnej rozdzielczości pokrywającej analizowane obszary. GFS jest to system prognozowania pogody krótko i średnioterminowej opracowany przez Narodowe Centrum Prognozowania Środowiska *(NCEP - National Centers for Environmental Prediction).* Dane z NCEP zawierają informacje na temat prędkości i kierunku wiatru, temperatury, wilgotności, ciśnienia, wysokości nad poziomem morza, opadów atmosferycznych, zachmurzenia itd. Dane GFS posiadają rozdzielczość czasową 3h.

W niniejszej pracy wykorzystano 2 wersje dostępnych danych pokrywające różne przedziały czasowe: wersję GFS.v1 (0,25 stopnia, system globalny, dane gromadzone od 13.05.2016 do 12.06.2019) oraz GFS.v2 (0,25 stopnia, system globalny, dane gromadzone od 13.06.2019). Wersja pierwsza została wykorzystana do symulacji trajektorii dla Wadowic 2017, wersję drugą użyto do symulacji trajektorii dla Krakowa 2020/2021. Symulacja polegała na wygenerowaniu 24 trajektorii wstecznych dla każdego przypadku o długości 12 h. Za punkt początkowy przyjęto współrzędne dla Wadowic: 49°90’N, 19°50’E oraz dla Krakowa: 50°10’N, 19°90’E. Do analizy trajektorii wstecznej wybrano dni z najwyższym dziennym stężeniem pyłu PM_{10} – rysunek 35 dla Wadowic i rysunek 36 dla Krakowa:



Rysunek 35. Mapy częstotliwości występowania trajektorii wstecznych dla wybranych dni dla Wadowic. Skala barwna odpowiada procentowi trajektorii przecinający dany punkt w okresie 24 h



Rysunek 36. Mapy częstotliwości występowania trajektorii wstecznych dla wybranych dni dla Krakowa: 08.02.2020 (rysunek 35A), 17.12.2020 (rysunek 35B), 24.02.2021 (rysunek 35C). Skala barwna odpowiada procentowi trajektorii przecinający dany punkt w okresie 24 h

W tabeli 12 przedstawiono średnie dobowe stężenia PM_{10}, WWA, B[a]P oraz informacje o dobowych warunkach meteorologicznych dla analizowanych przypadków:

Tabela 12.Średnie dobowe wartości stężeń oraz dane meteorologiczne dla Krakowa i Wadowic opowiadające analizowanym przypadkom

Rys.	Data	Miejsce	Kierunek wiatru	Prędkość wiatru [m/s]	Temperatura powietrza [°C]	Stężenie PM_{10} [µg/m^3]	Stężenie WWA [ng/m^3]	Stężenie B[a]P [ng/m^3]
34A	13.03.17	Wadowice	NE	2,7	3,0	116,77	156,86	19,48
34B	27.03.17	Wadowice	W	2,2	5,5	101,53	150,06	20,93
34C	20.09.17	Wadowice	N	4,0	10,5	107,29	21,06	0,57
35A	08.02.20	Kraków	SW	0,7	0,4	130,38	4,78	0,10
35B	17.12.20	Kraków	E	1,2	2,3	167,07	67,24	6,11
35C	24.02.21	Kraków	SW	0,7	8,2	234,14	83,67	10,16

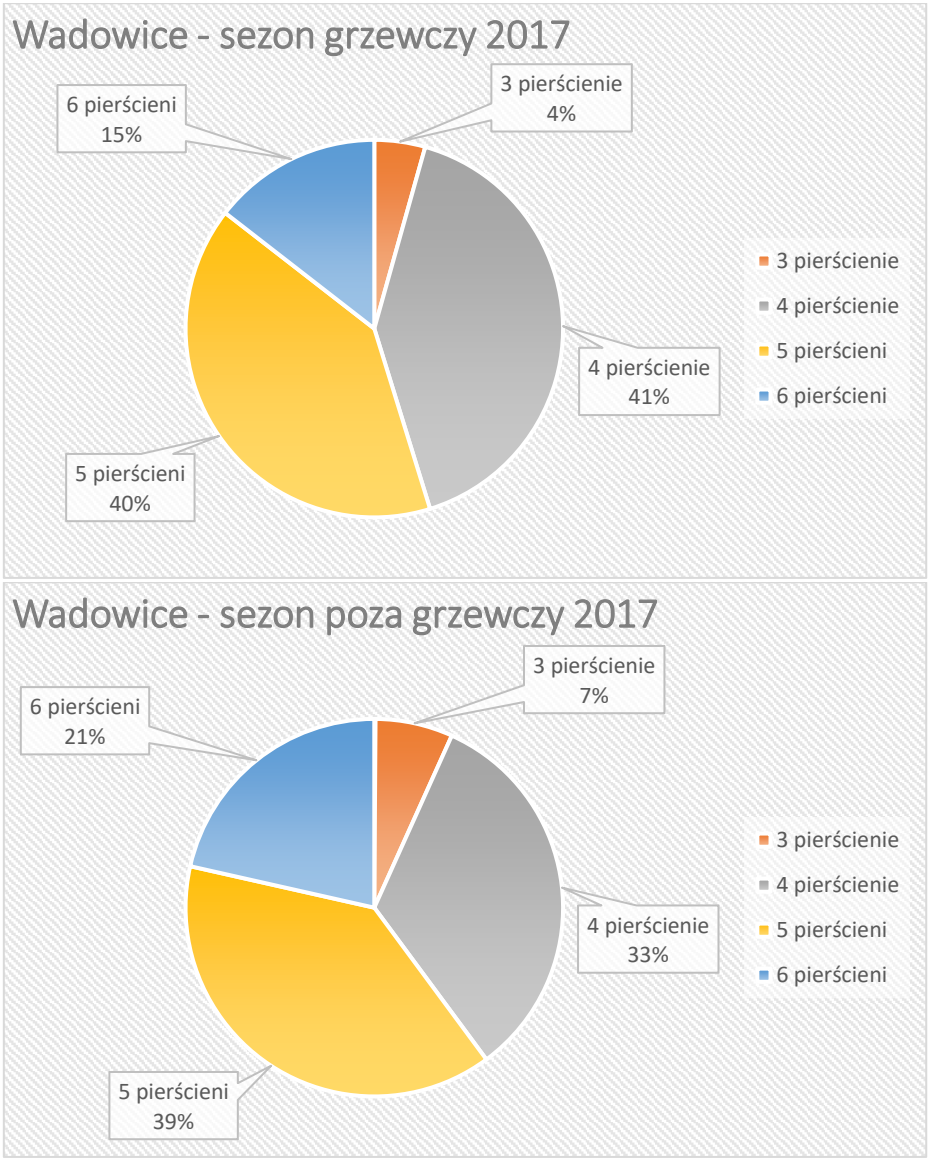
Trajektorie dla Wadowic wykazały napływ mas powietrza z obszarów znajdujących się na północ, północny wschód oraz zachód od Wadowic. Trajektorie północne i północno-wschodnie przechodzą nad obszarami Krakowa (gęstość zaludnienia 2456 os./km^2 – stan na 30.06.2022), Katowic (gęstość zaludnienia 1717 os./km^2 – stan na 31.12.2021), Olkusza (gęstość zaludnienia 1371 os./km^2 – stan na 31.12.2019), Chrzanowa (gęstość zaludnienia 955 os./km^2 – stan na 31.12.2019) [Główny Urząd Statystyczny 2023]. Wymienione miejscowości należą do miast o najwyższych wartościach dobowych stężeń pyłów zawieszonych oraz B[a]P w Polsce [Główny Inspektorat Ochrony Środowiska (GIOŚ)]. 13.03.2017 oraz 27.03.2017 odnotowano wysokie dobowe stężenie PM_{10} (odpowiednio 116,77 µg/m^3 i 101,53 µg/m^3) oraz B[a]P (odpowiednio 19,48 ng/m^3 i 20,93 ng/m^3). 20.09.2017 pomimo wysokiego dziennego stężenia PM_{10} (107,39 µg/m^3) zarejestrowano niskie dobowe wartości stężeń WWA i B[a]P (odpowiednio 21.06 ng/m^3 i 0,57 ng/m^3). Pomimo napływu mas powietrza do Wadowic 20.09.2017 z terenów przemysłowych zlokalizowanych na Śląsku (Katowice, Dąbrowa Górnicza) występowały niskie dobowe stężenie wszystkich WWA oraz B[a]P. Tego dnia odnotowano silny wiatr (4 m/s), co przyczyniło się do rozcieńczenia zanieczyszczeń. Wpływ na takie wartości stężeń ma również ukształtowanie terenu oraz obecność rozległych terenów leśnych na obszarach obejmujących przepływ analizowanych mas powietrza. Na północ od Wadowic znajdują się tereny Terczyńskiego Parku Krajobrazowego o powierzchni 151,54 km^2 i Rudniańskiego Parku Krajobrazowego o powierzchni 58,14 m^2 oraz Parku Krajobrazowego Dolinki Krakowskie o powierzchni 206,86 km^2 [Zespół Parków Krajobrazowych Województwa Małopolskiego 2023]. Obszary północne charakteryzują się niską gęstością zaludnienia. Na wschód, zachód i północny zachód od Wadowic rozciągają się kotliny o wyższej gęstości zaludnienia z dominującymi paleniskami domowymi.

Analiza trajektorii mas powietrza dla Krakowa, dla dni o największym dobowym stężeniu PM_{10} wykazała napływ mas powietrza z kierunku południowo zachodniego oraz wschodniego. Obszary te charakteryzują się wysoką gęstością zamieszkania: kierunek wschodni – Tarnów 1463 os./km^2 (stan na 31.12.2021), Brzesko 1421 os./km^2 (stan na 31.12.2019) [Główny Urząd Statystyczny 2023], kierunek południowo zachodni – Bielsko Biała 1352 os./km^2 (stan na 31.12.2021), Wadowice 1729,6 os./km^2 (stan na 30.06.2021) [Główny Urząd Statystyczny

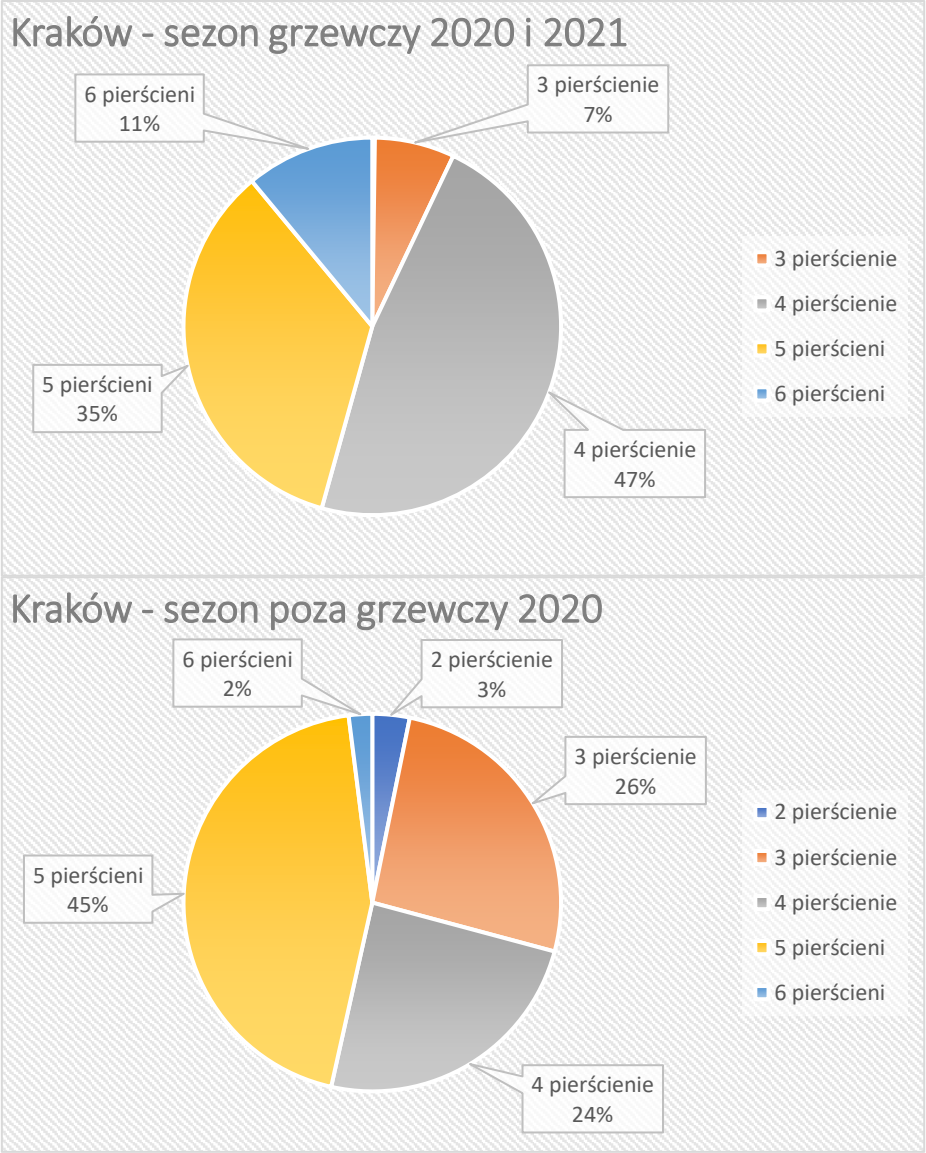
2023]. Raporty GIOŚ klasyfikują wymienione miejscowości wysoko na liście miast o największym dobowym stężeniu pyłów zawieszonych oraz Benzo[a]pirenu w Polsce [Główny Inspektorat Ochrony Środowiska (GIOŚ)]. Najwyższe dobowe stężenie PM_{10} odnotowano z terenów południowo zachodnich 24.02.2021 oraz z terenów wschodnich 17.12.2020, odpowiednio 234,14 $\mu g/m^3$ i 167,07 $\mu g/m^3$. 08.02.20 również wykazano kierunek południowo zachodni trajektorii wstecznych. Pomimo wysokiego dobowego stężenia PM_{10} (130,38 $\mu g/m^3$), dobowe stężenie WWA oraz B[a]P były na niskim poziomie, odpowiednio 4,78 ng/m^3 i 0,10 ng/m^3. Czynnikiem wpływającym na tego typu obserwacje jest ukształtowanie terenu. Tereny na południowy zachód ku południu (SW/S) charakteryzują się wyżynnym i górzystym ukształtowaniem terenu oraz mniejszą gęstością zaludnienia: Beskid Mały, Beskid Makowski, Beskid Wyspowy. Na tych obszarach dochodzi do kumulacji zanieczyszczeń w dolinach w wyniku utrudnionego naturalnego procesu przewietrzania i utrudnionego procesu przemieszczania się mas powietrza. Tereny na południowy wschód ku zachodowi (SE/W) to obszary kotliny rozciągającej się od Bielsko-Białej, Oświęcimia do Jastrzębia-Zdrój. Podobne ukształtowanie dominuje w kierunku wschodnim od Krakowa – Niepołomice, Bochnia, Tarnów. Obszary te charakteryzują się wysoką gęstością zaludnienia, gdzie dominują paleniska domowe mające główny wpływ na jakość powietrza. Monitoringi powietrza prowadzone w tych miastach dowodzą wysokich dobowych stężeń PM_{10} i B[a]P [European Environment Agency 2017, Główny Inspektorat Ochrony Środowiska 2017].

3.5. Profile narażenia na WWA

W celu dokładniejszej analizy jakości powietrza stworzono profile uwzględniające procentową zawartość WWA w zależności od liczby pierścieni, a co za tym idzie, ich masy (Wadowice – rysunek 37, Kraków – rysunek 38):



Rysunek 37.Procentowy udział WWA pod względem liczby pierścieni dla Wadowic 2017 z uwzględnieniem sezonu grzewczego i poza grzewczego



Rysunek 38. Procentowy udział WWA pod względem liczby pierścieni dla Krakowa 2020/2021 z uwzględnieniem sezonu grzewczego i poza grzewczego

Dla Krakowa odnotowano ponad 90 % udział ciężkich WWA w stosunku do wszystkich WWA wyznaczonych w próbkach pyłu. Spowodowane jest to adsorpcją i kondensacją ciężkich WWA na powierzchni pyłów. W sezonie grzewczym dla Krakowa około 93 % stanowią ciężkie WWA, w sezonie poza grzewczym 71 %. Dla Wadowic w sezonie grzewczym i poza grzewczym był to, odpowiednio 96 % i 93 % udział. W podobnych badaniach środowiskowych [Kozielska i in. 2016, Siudek i Frankowski 2018] odnotowano analogiczne udziały WWA o budowie 4-6 pierścieniowej – oszacowano udział ciężkich WWA na poziomie 90 % w całym okresie pobierania próbek. W większości przypadków wartości dla lekkich WWA nie przekraczają 10 %. Znacznie mniejszy udział lekkich WWA jest wynikiem ich występowania głównie w fazie gazowej. W sezonie cieplejszym, ciężkie związki stanowią mniejszy udział w stosunku do związków lekkich, co oznacza że związki lekkie, które występują głównie w fazie gazowej, w sezonie cieplejszym stanowiły większy udział w całkowitej masie WWA.W tabelach 13 i 14 przedstawiono profil uwzględniający procentowy udział węglowodorów kancerogennych w stosunku do sumarycznego stężenia wszystkich WWA w pyle zawieszonym PM_{10}, odpowiednio dla Wadowic i Krakowa:

Tabela 13. Udział węglowodorów kancerogennych – Wadowice 2017

Sezon grzewczy 2017	
Sumaryczne średnie miesięczne stężenie węglowodorów kancerogennych	33,64 ng/m^3
Sumaryczne średnie miesięczne stężenie wszystkich węglowodorów	42,12 ng/m^3
Procentowy udział węglowodorów kancerogennych	80 %
Sezon poza grzewczy 2017	
Sumaryczne średnie miesięczne stężenie węglowodorów kancerogennych	8,37 ng/m^3
Sumaryczne średnie miesięczne stężenie wszystkich węglowodorów	10,43 ng/m^3
Procentowy udział węglowodorów kancerogennych	80 %

Tabela 14. Udział węglowodorów kancerogennych – Kraków 2020/2021

Sezony grzewcze 2020 i 2020/2021	
Sumaryczne średnie miesięczne stężenie węglowodorów kancerogennych	15,98 ng/m^3
Sumaryczne średnie miesięczne stężenie wszystkich węglowodorów	25,92 ng/m^3
Procentowy udział węglowodorów kancerogennych	63 %
Sezon poza grzewczy 2020	
Sumaryczne średniomiesięczne stężenie węglowodorów kancerogennych	0,27 ng/m^3
Sumaryczne średniomiesięczne stężenie wszystkich węglowodorów	0,51 ng/m^3
Procentowy udział węglowodorów kancerogennych	53%

W Wadowicach węglowodory silnie kancerogenne stanowiły 80 % udział względem wszystkich WWA w obu sezonach pobierania próbek. Dla Krakowa był to 63 % udział w sezonie grzewczym i 53 % udział w sezonie poza grzewczym. Pomimo niższych sumarycznych średnich miesięcznych stężeń WWA w okresach letnich w obu miastach, węglowodory silnie kancerogenne stanowiły więcej niż 50%. W Katowicach, dla pomiarów frakcji pyłu zawieszonego z zakresu PM_1 – PM_{10} było to od 55 % dla okresu zimowego i 63 % dla okresu letniego [Kozielska i in. 2016]. Jedną z istotnych przyczyn obserwacji tego typu zależności w większości miast w Polsce jest dominująca rola źródła emisja zanieczyszczeń – transport, gdzie nastepuje znaczna emisja WWA o największej masie [Kozielska i in. 2013].

W tabeli 15 zaprezentowano wartości *MEQ*, *CEQ* i *TEQ* uzyskane dla pobranych próbek z Wadowic, Krakowa oraz innych miejsc w Polsce i na świecie:

Tabela 15. Wskaźniki narażenia MEQ, CEQ, TEQ [Kozielska i in. 2016, Li i in. 2009, Sánchez-Jiménez i in. 2012]

Lokalizacja	**Okres pomiaru**	**Frakcja PM**	***MEQ* [ng/m^3]**	***CEQ* [ng/m^3]**	***TEQ* [pg/m^3]**
Kraków	sezon poza grzewczy 2020	PM_{10}	0,08	0,06	0,62
	sezon grzewczy 2020 i 2020/2021		4,65	5,33	24,10
Wadowice	sezon poza grzewczy 2017	PM_{10}	3,05	11,27	14,00
	sezon grzewczy 2017		22,08	72,23	94,00
Katowice	lato 2012	PM_{10}	2,06	2,43	128
	wiosna 2012		3,54	4,16	413,9
Delhi, Indie	lato 2007/2008	PM_{10}	7,87	23,09	42,82
	zima 2007		20,07	59,75	106,42
Zagrzeb, Chorwacja	lato 2007	PM_{10}	0,14	0,10	0,56
	zima 2008		4,91	3,64	16,12
Madryt. Hiszpania	lato 2009	PM_1	0,09	0,12	0,43
	zima 2009		0,51	0,51	2,55
Atlanta, USA	styczeń-marzec 2004	$PM_{2.5}$	0,53	0,48	2,19
	październik-grudzień 2004		0,97	0,90	3,52

Dla wszystkich badanych miejsc dużo niższe wskaźniki uzyskano podczas okresów letnich lub sezonów poza grzewczych. Analiza wyników niniejszej pracy oraz danych zebranych z innych miast Polski i świata [Kozielska i in. 2016, Li i in. 2009, Sánchez-Jiménez i in. 2012] wykazuje, że niezależnie od sezonu pomiarowego (letni/zimowy, grzewczy/pozagrzewczy) oraz od frakcji pyłu zawieszonego, istnieje duże zagrożenie dla człowieka generowane obecnością wielopierścieniowych weglowodorów aromatycznych w powietrzu. Największe niebezpieczeństwo wystepuje na obszarach gęsto zaludnionych (np.: Indie) oraz o wysokim udziale transportu i spalaniu paliw stałych jako głównych źródłach zanieczyszczeń.

3.6. Kategoryzacja wskaźników narażenia przy użyciu algorytmów uczenia maszynowego

Próby stworzenia modeli bazujących na algorytmach uczenia maszynowego stanowią często spotykany dział badań środowiskowych, których celem jest przewidywanie stężeń pyłów zawieszonych, tlenków siarki, azotu oraz próby identyfikacji klasy jakości powietrza na podstawie stężeń wybranych zanieczyszczeń powietrza oraz parametrów meteorologicznych w 4-8 stopniowej skali [Ali i in. 2022, Chen i Chiu 2021, Pimpunchat i in. 2014, Sani i in. 2021]. Przykłady zastosowania algorytmów uczenia maszynowego w dziedzinie nauk środowiskowych znane są od wielu lat [Łozowicka Stupnicka i Talarczyk 2005, Ośródka i in. 1995]. Badania te opierały się na predykcjach poziomu stężenia konkretnego zanieczyszczenia powietrza (PM_{10}, SO_2, NO_2, CO) oraz stanu jakości powietrza z podziałem na 5 klas. Przed rozpowszechnieniem się działu uczenia maszynowego, do celów prognozowania jakości powietrza wykorzystywano statystyczne modele prognoz. Przykładem tego jest system ARMAX (ang. *AutoRegressive Moving Average with eXogenous input*), który w latach 1993-2002 był najczęściej używanym modelem prognoz w Krakowie służącym do predykcji stężeń np.: pyłów zawieszonych, dwutlenku siarki, dwutlenku azotu [Łozowicka Stupnicka i Talarczyk 2005]. Zastosowanie algorytmów uczenia maszynowego pozwala uwzględnić w przewidywaniach złożoną i nieliniową relację pomiędzy poziomem zanieczyszczeń a czynnikami wpływającymi na wartości ich stężeń (np.: warunki meteorologiczne) oraz uzyskać model o wysokiej zgodności wartości modelowanych z wartościami rzeczywistymi.

W niniejszej pracy wybiegam poza powszechnie spotykane w literaturze rozwiązania, proponując nowe podejście, w którym będę starał się opracować model pozwalający zidentyfikować stopień narażenia społeczności za pomocą wskaźników *MEQ*, *CEQ* i *TEQ*. Wartości tych współczynników uzależnione są od obecności 16 podstawowych WWA (rozdział *1.3),* określonych przez US EPA za najbardziej szkodliwe [U.S. EPA 2003]. W analizach opierałem się na danych dotyczących średniego dobowego stężenia pyłu zawieszonego PM_{10}, Benzo[*a*]pirenu, wskaźnikach narażenia *MEQ*, *CEQ*, *TEQ*, warunkach meteorologicznych, oraz modelach wykorzystujących algorytmy uczenia maszynowego. Wyznaczyłem zakresy zmienności wskaźników *MEQ*, *CEQ* oraz *TEQ*, na których pracował model. Opracowałem dwa osobne zakresy dla 2 (niski/wysoki) oraz 3 (niski/informowania/alarmowy) kategorii narażenia społeczności względem stężenia Benzo[*a*]pirenu równego 1 ng/m^3. Do modelowania wykorzystałem algorytmy wektorów nośnych, regresji logistycznych, lasów losowych oraz sztucznych sieci neuronowych. W pierwszej kolejności przetestowałem dla obu przypadków (2 i 3 kategorii) dokładność modelu dla każdego algorytmu względem podziału danych na zbiór treningowy i testowy w stosunkach równych 50/50, 60/40 oraz 70/30. Obliczenia zrealizowane zostały w środowisku Azure Machine Learning. Dokładny opis znajduje się rozdziale *2.5*.

3.6.1. Wektory nośne

Pierwszym testowanym algorytmem był algorytm wektorów nośnych. Efektywność modeli uzyskanych z wykorzystaniem tej metody zaprezentowane są w formie macierzy błędu w tabelach 16-17:

Tabela 16. Macierz błędu dla algorytmu wektorów nośnych dla 2 kategorii (Niski, Wysoki)

Wektory nośne – 2 kategorie

Podział:		**50/50**		**Podział:**		**60/40**		**Podział:**		**70/30**	
MEQ %		**Przewidywana kategoria**		**MEQ %**		**Przewidywana kategoria**		**MEQ %**		**Przewidywana kategoria**	
		Niski	Wysoki			Niski	Wysoki			Niski	Wysoki
Aktualna kategoria	Niski	72,5	27,5	**Aktualna kategoria**	Niski	66,2	33,8	**Aktualna kategoria**	Niski	80,0	20,0
	Wysoki	26,7	73,3		Wysoki	22,1	77,9		Wysoki	21,2	78,8
Dokładność		72,9		**Dokładność**		71,8		**Dokładność**		79,4	
CEQ %		**Przewidywana kategoria**		**CEQ %**		**Przewidywana kategoria**		**CEQ %**		**Przewidywana kategoria**	
		Niski	Wysoki			Niski	Wysoki			Niski	Wysoki
Aktualna kategoria	Niski	72,0	28,0	**Aktualna kategoria**	Niski	77,2	22,8	**Aktualna kategoria**	Niski	74,1	25,9
	Wysoki	46,8	53,2		Wysoki	33,3	66,7		Wysoki	30,6	69,4
Dokładność		63,8		**Dokładność**		72,5		**Dokładność**		72,0	
TEQ %		**Przewidywana kategoria**		**TEQ %**		**Przewidywana kategoria**		**TEQ %**		**Przewidywana kategoria**	
		Niski	Wysoki			Niski	Wysoki			Niski	Wysoki
Aktualna kategoria	Niski	73,3	26,3	**Aktualna kategoria**	Niski	66,2	33,8	**Aktualna kategoria**	Niski	78,9	21,1
	Wysoki	30,5	69,5		Wysoki	33,8	66,2		Wysoki	28,0	72,0
Dokładność		71,8		**Dokładność**		66,2		**Dokładność**		75,7	

Tabela 17. Macierz błędu dla algorytmu wektorów nośnych dla 3 kategorii (Niski, Informowania, Alarmowy)

Wektory nośne – 3 kategorie

Podział:		**50/50**			**Podział:**		**60/40**			**Podział:**		**70/30**		
MEQ %		**Przewidywana kategoria**			**MEQ %**		**Przewidywana kategoria**			**MEQ %**		**Przewidywana kategoria**		
		Niski	Info.	Alarm.			Niski	Info	Alarm.			Niski	Info.	Alarm.
Aktualna kategoria	Niski	69,8	19,0	11,2	**Aktualna kategoria**	Niski	73,3	9,3	17,4	**Aktualna kategoria**	Niski	71,9	18,0	10,0
	Info.	30,6	30,6	38,8		Info.	33,3	31,6	35,1		Info.	28,4	35,8	35,8
	Alarm.	6,00	16,4	77,6		Alarm.	10,8	9,7	79,5		Alarm.	20,0	6,3	73,7
Dokładność		62,0			**Dokładność**		64,7			**Dokładność**		62,9		
CEQ %		**Przewidywana kategoria**			**CEQ %**		**Przewidywana kategoria**			**CEQ %**		**Przewidywana kategoria**		
		Niski	Info.	Alarm.			Niski	Info.	Alarm.			Niski	Info.	Alarm.
Aktualna kategoria	Niski	54,1	1,6	44,3	**Aktualna kategoria**	Niski	74,5	2,0	23,5	**Aktualna kategoria**	Niski	59,5	0,0	40,5
	Info.	47,2	0,0	52,8		Info.	28,6	3,5	67,9		Info.	30,4	4,3	65,3
	Alarm.	19,5	0,0	80,5		Alarm.	23,4	0,0	76,6		Alarm.	19,1	0,0	80,9
Dokładność		55,3			**Dokładność**		61,5			**Dokładność**		57,0		
TEQ %		**Przewidywana kategoria**			**TEQ %**		**Przewidywana kategoria**			**TEQ %**		**Przewidywana kategoria**		
		Niski	Info.	Alarm.			Niski	Info.	Alarm.			Niski	Info.	Alarm.
Aktualna kategoria	Niski	70,3	17,2	12,5	**Aktualna kategoria**	Niski	71,2	11,5	17,3	**Aktualna kategoria**	Niski	65,8	21,0	13,2
	Info.	22,4	30,7	46,9		Info.	27,5	27,5	45,0		Info.	36,7	13,3	50,0
	Alarm.	9,1	30,3	60,6		Alarm.	15,7	33,3	51,0		Alarm.	10,3	20,5	69,2
Dokładność		55,9			**Dokładność**		51,7			**Dokładność**		52,3		

Model dla dwóch kategorii narażenia (niski, wysoki) bazujący na algorytmie wektorów nośnych wykazał najwyższą dokładność (79,4 %) dla wskaźnika *MEQ* dla podziału danych na zbiór treningowy i testowy w stosunku 70/30, natomiast najniższą dokładność (63,8 %) dla wskaźnika *CEQ* dla podziału danych na zbiór treningowy i testowy w stosunku 50/50. Model dla 3 kategorii najwyższą dokładność uzyskał dla wskaźnika *MEQ* (64,7 %), najniższą dla *TEQ* (51,7 %) – w obu przypadkach podział danych na zbiór treningowy i testowy wynosił 60/40. W środowisku Azure Machine Learning w automatyczny sposób za pomocą porównania dokładności obu zbiorów (treningowego i testowego) można stwierdzić, czy model jest nadmiernie dopasowany. Dokładność zbioru testowego powinna być porównywalna lub nieznacznie mniejsza od dokładności dla zbioru treningowego. Sprawdzając dokładności zbioru testowego i treningowego za pomocą modułu *Evaluate Model* dla wszystkich przypadków podziału danych na zbiór treningowy i testowy oraz dla każdego współczynnika narażenia, nie wykazano nadmiernego dopasowania modelu. Dokładność zbioru treningowego dla modelu wektorów nośnych we wszystkich przypadkach była większa od dokładności zbioru testowego o 3-5 %, co wskazuje na poprawny poziom zdolności do generalizacji modelu.

Modele wektorów nośnych w większości przypadków stosowane są w prognozach stężeń pyłów zawieszonych. Przykładem tego są badania [Hou i in. 2014] dotyczące stężeń $PM_{2.5}$ i PM_{10} przeprowadzonych w latach 2010-2012 w Pekinie, gdzie dokładność predykcji wynosiła 87 % dla przewidywanych średnich dobowych stężeń pyłu $PM_{2.5}$ oraz 70 % dla pyłu PM_{10}. W badaniach [Hou i in. 2014] podzielono dane na zestaw treningowy i testowy w stosunku równym 80/20. Korzystano z danych dotyczących średnich dobowych stężeń pyłów zawieszonych z kampanii pomiarowej trwającej 300 dni. Dla porównania, w niniejszej pracy doktorskiej, wykorzystano dane z 252 dni oraz podzielono dane na zbiory treningowe i testowe w stosunkach: 50/50, 60/40 i 70/30. Najwyższą dokładność uzyskano dla podziału danych na zbiór treningowy i testowy w stosunku 70/30 dla 3 kategorii narażenia (79,4 %). W badaniach [Hou i in. 2014] stwierdzono, że niezbędnym jest zwiększenie liczby danych wejściowych, co może mieć dodatni wpływ na jakość modelu. Z powodu podobnej ilości próbek wykorzystanych w niniejszej rozprawie doktorskiej, w celu zwiększenia dokładności modelu wektorów nośnych, w przyszłych badaniach korzystniejszym rozwiązaniem będzie znaczne zwiększenie okresu pobierania próbek. Podobne wnioski uzyskano w badaniach [Li i Tao 2017] przeprowadzonych w Chengguan (Chiny) w 2015 roku, gdzie do badań wykorzystano dane z 365 dni pomiarowych. Do badań użyto średnie dobowe stężenia PM_{10} oraz dane meteorologiczne. Dokładność modelu przewidywanych stężeń PM_{10} wynosiła 67 % dla podziału na zbiór treningowy i testowy równego 80/20. Model wektorów nośnych posiadał wysoką dokładność głównie w przypadku wysokich stężeń PM_{10}. W przeprowadzonych przeze mnie obliczeniach, dla mniejszej liczby próbek (252) uzyskano przeważnie wyższą dokładność dla 2 kategorii narażenia oraz każdego wskaźnika i każdego podziału danych na zbiór treningowy i testowy. W przypadku 3 kategorii narażenia dokładność modelu we wszystkich przypadkach była niższa niż w badaniach [Li i Tao 2017]. Nie zaobserwowano podobnych zależności wysokiej trafności predykcji wskaźników *MEQ, CEQ* i *TEQ* od stężenia PM_{10}. W badaniach [Suleiman i in. 2019] wykazano, że wektory nośne mają tendencję do nadmiernego dopasowania i należy z ostrożnością podchodzić do stosowania tego typu algorytmów w badaniach środowiskowych. W niniejszej pracy doktorskiej nie wykazano nadmiernego dopasowania modelu, jednak poziom uzyskanej dokładności (< 85 %) składania do przetestowania innych algorytmów uczenia maszynowego. W badaniach [Chen i Chiu 2021] dla przewidywania klas jakości powietrza w 4-stopniowej skali, wykorzystano algorytm wektorów nośnych oraz regresji logistycznej. Wykazano taką samą dokładność obu modeli (88

%) unikając zjawiska nadmiernego dopasowania przy zastosowaniu algorytmów regresji logistycznej. Dlatego też jednym z kolejnych testowanych algorytmów była regresja logistyczna (tabela 18-19).

3.6.2. Regresja logistyczna

Drugim testowanym algorytmem był algorytm regresji logistycznej. Dokładność modeli bazujących na tym algorytmie została przedstawiona w postaci macierzy błędu w tabelach 18-19:

Tabela 18. Macierz błędu dla algorytmu regresji logistycznej dla 2 kategorii (Niski, Wysoki)

Regresja logistyczna – 2 kategorie

Podział:		**50/50**		**Podział:**		**60/40**		**Podział:**		**70/30**	
MEQ %		**Przewidywana kategoria**		**MEQ %**		**Przewidywana kategoria**		**MEQ %**		**Przewidywana kategoria**	
		Niski	Wysoki			Niski	Wysoki			Niski	Wysoki
Aktualna kategoria	Niski	71,4	28,6	**Aktualna kategoria**	Niski	71,6	28,4	**Aktualna kategoria**	Niski	78,2	21,8
	Wysoki	19,8	80,2		Wysoki	20,6	79,4		Wysoki	19,2	80,8
Dokładność		75,7		**Dokładność**		79,4		**Dokładność**		79,4	
CEQ %		**Przewidywana kategoria**		**CEQ %**		**Przewidywana kategoria**		**CEQ %**		**Przewidywana kategoria**	
		Niski	Wysoki			Niski	Wysoki			Niski	Wysoki
Aktualna kategoria	Niski	85,0	15,0	**Aktualna kategoria**	Niski	81,0	19,0	**Aktualna kategoria**	Niski	84,5	15,5
	Wysoki	37,7	62,3		Wysoki	31,7	68,3		Wysoki	32,7	67,3
Dokładność		75,1		**Dokładność**		75,4		**Dokładność**		76,6	
TEQ %		**Przewidywana kategoria**		**TEQ %**		**Przewidywana kategoria**		**TEQ %**		**Przewidywana kategoria**	
		Niski	Wysoki			Niski	Wysoki			Niski	Wysoki
Aktualna kategoria	Niski	73,7	26,3	**Aktualna kategoria**	Niski	76,6	23,4	**Aktualna kategoria**	Niski	84,2	15,8
	Wysoki	24,4	75,6		Wysoki	26,2	73,8		Wysoki	18,0	82,0
Dokładność		74,6		**Dokładność**		75,4		**Dokładność**		83,2	

Tabela 19. Macierz błędu dla algorytmu regresji logistycznej dla 3 kategorii (Niski, Informowania, Alarmowy)

Regresja logistyczna – 3 kategorie

Podział: 50/50

MEQ %		Przewidywana kategoria: Niski	Info.	Alarm.
Aktualna kategoria	Niski	74,6	7,9	17,5
	Info.	30,6	22,4	47,0
	Alarm.	6,0	4,4	89,6
Dokładność		65,9		

CEQ %		Przewidywana kategoria: Niski	Info.	Alarm.
Aktualna kategoria	Niski	60,7	0,0	39,3
	Info.	30,6	2,8	66,7
	Alarm.	11,0	0,0	89,0
Dokładność		62,0		

TEQ %		Przewidywana kategoria: Niski	Info.	Alarm.
Aktualna kategoria	Niski	65,6	20,3	14,1
	Info.	32,7	22,4	44,9
	Alarm.	10,6	22,7	66,7
Dokładność		54,2		

Podział: 60/40

MEQ %		Przewidywana kategoria: Niski	Info.	Alarm.
Aktualna kategoria	Niski	77,4	9,3	13,3
	Info.	29,8	26,3	43,9
	Alarm.	4,8	3,6	91,6
Dokładność		69,3		

CEQ %		Przewidywana kategoria: Niski	Info.	Alarm.
Aktualna kategoria	Niski	68,6	0,0	31,4
	Info.	21,4	3,6	75,0
	Alarm.	9,4	0,0	90,6
Dokładność		65,7		

TEQ %		Przewidywana kategoria: Niski	Info.	Alarm.
Aktualna kategoria	Niski	67,3	17,3	15,4
	Info.	30,0	15,0	55,0
	Alarm.	13,7	13,7	72,5
Dokładność		54,5		

Podział: 70/30

MEQ %		Przewidywana kategoria: Niski	Info.	Alarm.
Aktualna kategoria	Niski	78,7	9,0	12,3
	Info.	25,3	26,9	47,8
	Alarm.	6,3	3,2	90,5
Dokładność		69,3		

CEQ %		Przewidywana kategoria: Niski	Info.	Alarm.
Aktualna kategoria	Niski	70,3	0,0	29,7
	Info.	30,4	0,0	69,6
	Alarm.	10,6	0,0	89,4
Dokładność		63,6		

TEQ %		Przewidywana kategoria: Niski	Info.	Alarm.
Aktualna kategoria	Niski	63,2	18,4	18,4
	Info.	30,0	13,3	56,7
	Alarm.	7,7	17,9	74,4
Dokładność		53,3		

Model dla 2 kategorii opierający się na algorytmach regresji logistycznej wykazał zarazem najniższą, jak i najwyższą dokładność dla wskaźnika *TEQ*. Najwyższą dokładność (83,2 %) uzyskano dla podziału danych na zbiór treningowy i testowy w stosunku 70/30, natomiast najniższą dokładność (74,6 %) otrzymano w przypadku podziału danych na zbiór treningowy i testowy w stosunku 50/50. Dla modeli klasyfikujących poziom narażenia na 3 kategorie wykazano najwyższą dokładność dla wskaźnika *MEQ* (69,3 %) dla podziału danych na zbiór treningowy i testowy 60/40 oraz 70/30. Modele 3 kategorii najniższą dokładność uzyskały dla wskaźnika *TEQ* (53,3 %) i podziału danych na zbiór trening/test w stosunku 70/30. Nie wykazano przetrenowania modelu. Dokładność zbioru treningowego dla modeli regresji logistycznej we wszystkich przypadkach była większa od dokładności zbioru testowego o 3-9 % – oprawny poziom zdolności do generalizacji modelu.

W badaniach [Chen i Chiu 2021] porównywano dokładność modelu regresji logistycznej względem ilości klas jakości powietrza. Przeprowadzono analizy dla 2-stopniowej i 4-stopniowej skali jakości powietrza. W modelach wykorzystano dane dotyczące średniego dobowego stężenia $PM_{2.5}$, PM_{10} oraz dane meteorologiczne z 1227 dni uzyskane ze stacji pomiarowych z 16 różnych miast w Tajwanie w latach 2016-2020. Dokładność modelu dla 2-stopniowej skali dla wszystkich stacji pomiarowych była w granicach od 78 % do 98 % (średnia dokładność wynosiła 89 %), dla 4-stopniowej wahała się od 74 % do 97 % (średnia dokładność wynosiła 88 %). Dla porównania wykonano testy dokładności wektorów nośnych. Średnia dokładność takiego modelu dla 4-stopniowej skali z uwzględnieniem wszystkich 16 stacji wynosiła 88 %. W tym przypadku na jakość modelu może mieć znaczny wpływ miejsce pobierania próbek pyłów zawieszonych. W badaniach [Pimpunchat i in. 2014] przeprowadzonych w latach 2005-2010 w Tajlandii wykorzystano algorytmy regresji logistycznych w przewidywaniach 3-stopniowej skali jakości powietrza. Wykazano 93 % dokładność modelu bazującego na średnich dobowych stężeniach PM_{10}, tlenków azotu, tlenków węgla, tlenków siarki oraz warunkach meteorologicznych. W badaniach [Balas i in 2021] opierających się na predykcjach 5-stopniowej skali jakości powietrza wykazano dokładność algorytmów regresji logistycznej równą 80,6 %. Podobne wnioski wynikają z artykułu [Jung i in. 2021]. Badania prowadzono w latach 2009-2018 w Cheonan w Korei Południowej. Model bazował na średnich dobowych stężeniach $PM_{2.5}$, PM_{10}, ozonu, tlenków węgla, tlenków azotu oraz tlenków siarki, a także danych meteorologicznych. Podział danych na zbiór treningowy i testowy wynosił 75/25. Wykazano niezadowalającą dokładność modeli regresji logistycznej (< 85 %). Podobnych obserwacji dokonano w niniejsze pracy doktorskiej, gdzie najwyższa dokładność modelu nie przekraczała 80 % dla 2 kategorii i 70 % dla 3 kategorii dla wszystkich wskaźników narażenia i każdego przypadku podziału danych na zbiór treningowy i testowy. Wymienione badania [Balas i in. 2021, Jung i in. 2021, Pimpunchat i in. 2014] opierały się na kilkuletnich kampaniach pomiarowych liczących kilka tysięcy danych. Niniejsza praca bazuje na zestawie składającym się z 252 dni pomiarowych. We wszystkich przypadkach dokładność modeli nie przekroczyła progu 85 %. Jednak niewielka ilość danych w przypadku regresji logistycznych nie zawsze stanowi problem w odniesieniu do jakości modelu. Dla przykładu, w artykule [Ali i in. 2022] wykazano dokładność modeli regresji logistycznej równą 84 %. Badania dotyczyły przewidywania klasy jakości powietrza w 8-stopniowej skali. W badaniach wykorzystano dane z 145 dni pomiarowych: dane meteorologiczne, stężenia średnie dobowe $PM_{2.5}$ i PM_{10}. Wyniki porównano dla algorytmu wektorów nośnych, którego model uzyskał dokładność równą 89 %. Korzystniejszym rozwiązaniem problemu niskiej dokładności modelu regresji logistycznej w niniejszej pracy jest zastosowanie innych algorytmów uczenia maszynowego, czego dowodzą powyższe

artykuły [Balas i in 2021, Y-J i in. 2021], gdzie modele lasów losowych charakteryzowały się wyższymi dokładnościami. W tabelach 20-21 zaprezentowano macierze błędów dla modeli lasów losowych.

3.6.3. Lasy losowe

Kolejnym algorytmem wykorzystanym w niniejszej pracy doktorskiej były lasy losowe. Wyniki przeprowadzonych obliczeń dla tego algorytmu zawierają tabele 20-21:

Tabela 20. Macierz błędu dla algorytmu lasów losowych dla 2 kategorii (Niski, Wysoki)

Lasy losowe – 2 kategorie

Podział:		**50/50**		**Podział:**		**60/40**		**Podział:**		**70/30**	
MEQ %		**Przewidywana kategoria**		**MEQ %**		**Przewidywana kategoria**		**MEQ %**		**Przewidywana kategoria**	
		Niski	Wysoki			Niski	Wysoki			Niski	Wysoki
Aktualna kategoria	Niski	94,5	5,5	**Aktualna kategoria**	Niski	81,1	18,9	**Aktualna kategoria**	Niski	90,9	9,1
	Wysoki	0,00	100,0		Wysoki	0,0	100,0		Wysoki	0,0	100,0
Dokładność		97,2		**Dokładność**		90,1		**Dokładność**		95,3	
CEQ %		**Przewidywana kategoria**		**CEQ %**		**Przewidywana kategoria**		**CEQ %**		**Przewidywana kategoria**	
		Niski	Wysoki			Niski	Wysoki			Niski	Wysoki
Aktualna kategoria	Niski	89,0	11,0	**Aktualna kategoria**	Niski	88,6	11,4	**Aktualna kategoria**	Niski	87,9	12,1
	Wysoki	5,2	94,8		Wysoki	6,3	93,7		Wysoki	4,1	95,9
Dokładność		91,50		**Dokładność**		90,8		**Dokładność**		91,6	
TEQ %		**Przewidywana kategoria**		**TEQ %**		**Przewidywana kategoria**		**TEQ %**		**Przewidywana kategoria**	
		Niski	Wysoki			Niski	Wysoki			Niski	Wysoki
Aktualna kategoria	Niski	91,6	8,4	**Aktualna kategoria**	Niski	92,2	7,8	**Aktualna kategoria**	Niski	91,2	8,8
	Wysoki	2,4	97,6		Wysoki	3,1	96,9		Wysoki	4,0	96,0
Dokładność		94,4		**Dokładność**		94,4		**Dokładność**		93,5	

Tabela 21. Macierz błędu dla algorytmu lasów losowych dla 3 kategorii (Niski, Informowania, Alarmowy)

Lasy losowe – 3 kategorie

Podział:		**50/50**			**Podział:**		**60/40**			**Podział:**		**70/30**		
MEQ %		**Przewidywana kategoria**			**MEQ %**		**Przewidywana kategoria**			**MEQ %**		**Przewidywana kategoria**		
		Niski	Info.	Alarm.			Niski	Info.	Alarm.			Niski	Info..	Alarm.
Aktualna kategoria	Niski	98,5	1,5	0,0	**Aktualna kategoria**	Niski	100,0	0,0	0,0	**Aktualna kategoria**	Niski	100,0	0,0	0,0
	Info.	17,0	83,0	0,0		Info.	2,6	97,4	0,0		Info.	17,2	82,8	0,0
	Alarm.	3,0	0,0	97,0		Alarm.	2,0	3,9	94,1		Alarm.	2,6	0,0	97,4
Dokładność		95,2			**Dokładność**		97,2			**Dokładność**		94,4		
CEQ %		**Przewidywana kategoria**			**CEQ %**		**Przewidywana kategoria**			**CEQ %**		**Przewidywana kategoria**		
		Niski	Info.	Alarm.			Niski	Info.	Alarm.			Niski	Info.	Alarm.
Aktualna kategoria	Niski	95,1	4,9	0,0	**Aktualna kategoria**	Niski	98,0	2,0	0,0	**Aktualna kategoria**	Niski	97,3	2,7	0,0
	Info.	25,0	72,2	2,8		Info.	25,0	75,0	0,0		Info.	21,7	78,3	0,0
	Alarm.	7,3	1,2	91,5		Alarm.	9,4	1,5	89,1		Alarm.	8,5	0,0	91,5
Dokładność		88,8			**Dokładność**		89,5			**Dokładność**		90,7		
TEQ %		**Przewidywana kategoria**			**TEQ %**		**Przewidywana kategoria**			**TEQ %**		**Przewidywana kategoria**		
		Niski	Info.	Alarm.			Niski	Info.	Alarm.			Niski	Info.	Alarm.
Aktualna kategoria	Niski	93,8	6,2	0,0	**Aktualna kategoria**	Niski	88,5	11,5	0,0	**Aktualna kategoria**	Niski	97,4	2,6	0,0
	Info.	8,2	89,8	2,0		Info.	5,0	95,0	0,0		Info.	16,7	83,3	0,0
	Alarm.	6,0	7,6	86,4		Alarm.	3,9	5,9	90,2		Alarm.	5,1	0,0	94,9
Dokładność		89,9			**Dokładność**		90,9			**Dokładność**		92,5		

Modele oparte na algorytmach lasów losowych w większości przypadków uzyskały dokładność większą niż 90 %. Dla 2 kategorii najniższą (90,1 %) oraz najwyższą (97,2 %) dokładność uzyskano dla wskaźnika *MEQ* oraz podziału danych na zbiór treningowy i testowy w stosunku, odpowiednio 60/40 oraz 50/50. W przypadku 3 kategorii, najwyższą dokładnością charakteryzuje się współczynnik *MEQ* (97,2 %, podział danych na zbiór treningowy i testowy – 60/40), a najniższą współczynnik *CEQ* (88,8 %, podział danych na zbiór treningowy i testowy – 50/50). Dla modeli lasów losowych stwierdzono nadmierne dopasowanie – dokładności zbiorów testowych posiadały wyższe wartości od dokładności zbiorów treningowych o 14-16 %.

W badaniach [Johney i in. 2020] podjęto próby klasyfikacji stopnia jakości powietrza w 6-stopniowej skali. Parametrami wejściowymi były dane meteorologiczne oraz stężenia średnie dobowe $PM_{2.5}$ i PM_{10}. 70 % danych stanowiło zbiór treningowy, natomiast zbiór testowy to 30 % danych. Wykazano dokładność modelu lasów losowych rzędu 97 %. Wysokie dokładności modelu lasów losowych uzyskano również w badaniach [Sani i in. 2021] przeprowadzonych w Bangladeszu w latach 2016-2019 odnotowano dokładność modelu równą 99,9 %. Bazowano na danych meteorologicznych, średnich dobowych stężeniach $PM_{2.5}$, PM_{10}, tlenków siarki, tlenków azotu i tlenków węgla, przewidywano jakość powietrza w 5-stopniowej skali. Stosunek danych treningowych do testowych wynosił 75/25. Dla porównania, w niniejszej pracy doktorskiej, dla podziału danych na zbiór treningowy i testowy równego 70/30 i 2 kategorii uzyskano dokładności modelu dla *MEQ*, *CEQ* i *TEQ* równe, odpowiednio 95,3 %, 91,6 % oraz 93,5 %, natomiast dla 3 kategorii uzyskano dokładności modelu 94,4 %, 90,7 % oraz 92,5 %. Autorzy publikacji [Sani i in. 2021] zwracają uwagę, że z tak wysoką dokładnością modelu wiąże się ryzyko przetrenowania. Zjawisko przetrenowania pojawia się, gdy model jest nadmiernie dopasowany do zbioru uczącego – model podejmuje decyzje poprawnie tylko dla danych znajdujących się w zbiorze treningowym. Wystarczy drobna zmiana, nawet jednego parametru lub jednej danej, a wynik predykcji zupełnie odbiega od poprawnego. Pomimo charakterystycznej dla tych algorytmów wysokiej dokładności modeli, istnieją przypadki, gdzie uzyskane dokładności modeli lasów losowych są dużo niższe niż innych algorytmów. Przykładem tego są badania [Shah i in 2020] przewidywania 5-stopniowej skali jakości powietrza dla Seulu. Porównywano ze sobą modele bazujące na algorytmach wektorów nośnych i lasów losowych. W tym celu zebrano dane dotyczące średnich dobowych stężeń $PM_{2.5}$ PM_{10}, CO, NO_2, O_3, SO_2 w latach 2014-2020. Algorytm wektorów nośnych uzyskał dokładność predykcji jakości powietrza równą 95 %, natomiast algorytm lasów losowych 76 %. W badaniach zaprezentowano również predykcje 3-stopniowej skali poziomu hałasu jaki może pojawić się w budynkach w Seulu. Do tego celu wykorzystano dane z innych artykułów naukowych. Model bazujący na algorytmach wektorów nośnych uzyskał 98 %, a model bazujący na lasach losowych 85 %. Pomimo wysokiej dokładności modeli lasów losowych odnotowano nadmierne dopasowanie. W niniejszej rozprawie doktorskiej, przeprowadzono testy dokładności dla zbioru treningowego oraz testowego, które wykazały niższą dokładność zbioru treningowe niż testowego o 14-16 %, co wskazuje na brak zdolności do generalizacji modelu. Z tego powodu podjęto decyzję wykorzystania w pracy kolejnego algorytmu uczenia maszynowego – sztucznych sieci neuronowych. W dalszej części pracy (tabele 22-23) zaprezentowano wyniki dla modeli sieci neuronowych.

3.6.4. Sieci neuronowe

Ostatnim algorytmem zastosowanym w modelach były sztuczne sieci neuronowe. Dokładność modeli sieci neuronowych prezentują macierze błędów przedstawione tabelach 22-23:

Tabela 22. Macierz błędu dla algorytmu sieci neuronowych dla 2 kategorii (Niski, Wysoki)

Sieci neuronowe – 2 kategorie

Podział:		**50/50**		**Podział:**		**60/40**		**Podział:**		**70/30**	
MEQ %		**Przewidywana kategoria**		**MEQ %**		**Przewidywana kategoria**		**MEQ %**		**Przewidywana kategoria**	
		Niski	Wysoki			Niski	Wysoki			Niski	Wysoki
Aktualna kategoria	Niski	89,0	11,0	**Aktualna kategoria**	Niski	93,2	6,8	**Aktualna kategoria**	Niski	78,2	21,8
	Wysoki	11,6	88,4		Wysoki	17,6	82,4		Wysoki	1,9	98,1
Dokładność		88,7		**Dokładność**		88,0		**Dokładność**		87,9	
CEQ %		**Przewidywana kategoria**		**CEQ %**		**Przewidywana kategoria**		**CEQ %**		**Przewidywana kategoria**	
		Niski	Wysoki			Niski	Wysoki			Niski	Wysoki
Aktualna kategoria	Niski	91,0	9,0	**Aktualna kategoria**	Niski	78,5	21,5	**Aktualna kategoria**	Niski	91,4	8,6
	Wysoki	20,8	79,2		Wysoki	9,5	90,5		Wysoki	22,4	77,6
Dokładność		85,9		**Dokładność**		83,8		**Dokładność**		85,0	
TEQ %		**Przewidywana kategoria**		**TEQ %**		**Przewidywana kategoria**		**TEQ %**		**Przewidywana kategoria**	
		Niski	Wysoki			Niski	Wysoki			Niski	Wysoki
Aktualna kategoria	Niski	85,3	14,7	**Aktualna kategoria**	Niski	92,2	7,8	**Aktualna kategoria**	Niski	77,2	22,8
	Wysoki	6,1	93,9		Wysoki	35,4	64,6		Wysoki	2,0	98,0
Dokładność		89,3		**Dokładność**		79,6		**Dokładność**		86,9	

Tabela 23. Macierz błędu dla algorytmu sieci neuronowych dla 3 kategorii (Niski, Informowania, Alarmowy)

Sieci neuronowe – 3 kategorie

Podział:		50/50			Podział:		60/40			Podział:		70/30		
MEQ %		**Przewidywana kategoria**			**MEQ %**		**Przewidywana kategoria**			**MEQ %**		**Przewidywana kategoria**		
		Niski	Info..	Alarm.			Niski	Info.	Alarm.			Niski	Info.	Alarm.
Aktualna kategoria	Niski	83,1	16,9	0,0	**Aktualna kategoria**	Niski	83,0	17,0	0,0	**Aktualna kategoria**	Niski	82,1	17,9	0,0
	Info.	21,3	78,7	0,0		Info.	30,8	66,7	2,6		Info.	13,8	86,2	0,0
	Alarm.	1,5	9,0	89,5		Alarm.	2,0	3,9	94,1		Alarm.	2,6	5,1	92,3
Dokładność		84,3			**Dokładność**		82,5			**Dokładność**		86,9		
CEQ %		**Przewidywana kategoria**			**CEQ %**		**Przewidywana kategoria**			**CEQ %**		**Przewidywana kategoria**		
		Niski	Info.	Alarm.			Niski	Info.	Alarm.			Niski	Info.	Alarm.
Aktualna kategoria	Niski	70,5	16,4	13,1	**Aktualna kategoria**	Niski	80,4	9,8	9,8	**Aktualna kategoria**	Niski	83,8	10,8	5,4
	Info.	47,2	36,1	16,7		Info.	42,9	50,0	7,1		Info.	43,5	47,8	8,7
	Alarm.	2,4	3,7	93,9		Alarm.	6,3	1,5	92,2		Alarm.	8,5	0,0	91,5
Dokładność		74,3			**Dokładność**		79,7			**Dokładność**		79,4		
TEQ %		**Przewidywana kategoria**			**TEQ %**		**Przewidywana kategoria**			**TEQ %**		**Przewidywana kategoria**		
		Niski	Info.	Alarm.			Niski	Info.	Alarm.			Niski	Info.	Alarm.
Aktualna kategoria	Niski	82,8	10,9	6,3	**Aktualna kategoria**	Niski	82,7	11,5	5,8	**Aktualna kategoria**	Niski	89,5	7,9	2,6
	Info.	18,4	73,5	8,1		Info.	27,5	67,5	5,0		Info.	20,0	80,0	0,0
	Alarm.	1,5	13,6	84,9		Alarm.	2,0	3,9	94,1		Alarm.	2,6	5,1	92,3
Dokładność		81,0			**Dokładność**		82,5			**Dokładność**		87,9		

Dla 2 kategorii modele sieci neuronowych uzyskały najwyższą (89,3 %) oraz najniższą (79,6 %) dokładność dla wskaźnika *TEQ*, odpowiednio dla podziału danych na zbiór treningowy i testowy w stosunku 50/50 oraz 60/40. W przypadku 3 kategorii najwyższą dokładnością charakteryzuje się współczynnik *TEQ* (87,9 %) dla podziału danych na zbiór treningowy i testowy równego 70/30. Najniższą dokładność (74,3 %) posiada współczynnik *CEQ* dla podziału danych na zbiór treningowy i testowy równego 50/50. Nie odnotowano nadmiernego dopasowania modelu – dokładność zbioru treningowego była wyższa od dokładności zbioru testowego o 2-4 %.

W badaniach [Farhadi i in. 2020] do predykcji klas jakości powietrza wykorzystano warunki meteorologiczne i wartości średnich dobowych stężeń PM_{10} oraz tlenku węgla uzyskane w latach 2013-2015 w Teheranie. W modelach użyto algorytmy sztucznych sieci neuronowych. Model predykcji stężeń PM_{10} uzyskał dokładność 83 % dla ciepłych okresów badawczych, a model predykcji stężeń CO 74 %. W przypadku przewidywania klas jakości powietrza uzyskano dokładność modelu nie przekraczającą 57 %. Pomimo dużej ilości danych wejściowych (2 lata), model predykcji uzyskał niezadowalające dokładności. W niniejszej pracy, pomimo znacznie mniejszego zbioru danych (252 dni) wykazano wyższe dokładności modelu. W większości przypadków dla 2 kategorii, model sieci neuronowych uzyskał dokładność większa niż 85 %. Dla 3 kategorii, dokładność ta wahała się od 74,3 % dla wskaźnika *CEQ*, do 87,9% dla wskaźnika *TEQ*. W artykule [Skrzypski i Jach-Szakiel 2008] skupiających się na wykorzystaniu sztucznych sieci neuronowych do prognozowania klas stanu jakości powietrza w obszarach miejsko-przemysłowych w Łodzi w latach 2004-2007 wykazano wskaźnik błędnych predykcji 1,9% dla prognozowania średnich stężeń dobowych PM_{10} oraz 9,2% dla prognozowanych dobowych stężeń maksymalnych. Udowodniono, że modele neuronowe mogą zostać z powodzeniem wykorzystywane w systemach wczesnego ostrzegania związanych z jakością powietrza. Badania [Balas i in 2021] dowodzą wysokiej skuteczności sztucznych sieci neuronowych w aspekcie kategoryzacji jakości powietrza w 5 stopniowej skali. Badania bazowały na danych zebranych w latach 2016-2019 w Bangladeszu (warunki meteorologiczne, stężenia średnie dobowe $PM_{2.5}$, PM_{10}, tlenków siarki, tlenków azotu i tlenków węgla). Wyniki porównano dla tych samych badań z wykorzystaniem algorytmów regresji logistycznej oraz lasów losowych. Dla wszystkich algorytmów stosunek danych treningowych i testowych wynosił 75/25. Wykazano dokładność modelu bazującego na sztucznych sieciach neuronowych na poziomie 94 %. Najniższe dokładności uzyskał algorytm regresji logistycznych (80,6 %). W niniejszej pracy, porównując te same algorytmy (sieci neuronowych, lasów losowych, regresji logistycznych) oraz podział danych na zbiór treningowy i testowy równy 70/30 modele dla 2 kategorii i regresji logistycznej posiadały dokładności równe 79,4 % (*MEQ*), 76,6 % (*CEQ*), 83,2 % (*TEQ*), lasów losowych 95,3 % (*MEQ*), 91,6 % (*CEQ*), 93,5 % (*TEQ*), sieci neuronowych 87,9 % (*MEQ*), 85,0 % (*CEQ*), 86,9 % (*TEQ*). Dla 3 kategorii, dokładność modeli regresji logistycznej wynosiła 69,3 % (*MEQ*), 63,6 % (*CEQ*), 53,3 % (*TEQ*), lasów losowych 94,4 % (*MEQ*), 90,7 % (*CEQ*), 92,5 % (*TEQ*), sieci neuronowych 86,9 % (*MEQ*), 79,4 % (*CEQ*), 87,9 % (*TEQ*). W przypadku lasów losowych odnotowano nadmierne dopasowanie. Wysoką dokładność modeli z zastosowaniem sieci neuronowych zauważono również w badaniach przeprowadzonych w Delhi w latach 2016-2018 dla $PM_{2.5}$ [Masooda i Ahmad 2020]. Algorytm ten obejmujący dane dotyczące zanieczyszczeń powietrza i dane meteorologiczne wykazał dużo lepsze wyniki prognoz w porównaniu z modelami wektorów nośnych (odpowiednio 86 % i 73 %). W niniejszej pracy, dla znacznie mniejszej ilości danych (252 dni), dokładność modeli wektorów nośnych dla 3 kategorii oraz podziału danych na zbiór treningowy i testowy 70/30 wynosiła 62,9 % (*MEQ*),

57,0 % (*CEQ*), 52,3 % (*TEQ*). Dla prognozy stężeń PM_{10} i $PM_{2.5}$ przeprowadzonych w Londynie w latach 2007 i 2012 wykorzystano sztuczne sieci neuronowe, lasy losowe oraz wektory nośne [Suleiman i in. 2019]. Wszystkie modele uzyskały dokładność rzędu 95%, jednak w przypadku lasów losowych oraz wektorów nośnych zaobserwowano zjawisko przetrenowania modelu. Algorytmy sztucznych sieci neuronowych sprawdziły się najlepiej dla przewidywania stężeń PM_{10} i $PM_{2.5}$ w obszarach przydrożnych. W tabeli 24 zaprezentowano zebrane dane dla wszystkich zastosowanych algorytmów oraz 2-stopniowej i 3-stopniowej klasyfikacji stopnia narażenia.

Tabela 24. Podsumowanie otrzymanych dokładności dla każdego algorytmu dla wskaźników MEQ, CEQ i TEQ

Kategorie	Parametry		Sieci neuronowe			Lasy losowe			Wektory nośne			Regresja logistyczna		
			MEQ	***CEQ***	***TEQ***	***MEQ***	***CEQ***	***TEQ***	***MEQ***	***CEQ***	***TEQ***	***MEQ***	***CEQ***	***TEQ***
2 Niski Wysoki	Podział Trening/Test→	50/50	88,7	85,9	89,3	97,2	91,5	94,4	72,9	63,8	71,8	75,7	75,1	74,6
	Brak walidacji krzyżowej	60/40	88,0	83,8	79,6	90,1	90,8	94,4	71,8	72,5	66,2	79,4	75,4	75,4
	Z uwzględnieniem B[a]P	70/30	87,9	85,0	86,9	95,3	91,6	93,5	79,4	72,0	75,7	79,4	76,6	83,2
	Podział Trening/Test 70/30 Brak walidacji krzyżowej Bez uwzględniania B[a]P		77,6	69,2	79,4	81,3	73,8	78,5	72,9	68,2	77,6	69,2	68,2	73,8
3 Niski Informowania Alarmowy	Podział Trening/Test→	50/50	84,3	74,3	81,0	95,2	88,8	89,9	62,0	55,3	55,9	65,9	62,0	54,2
	Brak walidacji krzyżowej	60/40	82,5	79,7	82,5	97,2	89,5	90,9	64,7	61,5	51,7	69,3	65,7	54,5
	Z uwzględnieniem B[a]P	70/30	86,9	79,4	87,9	94,4	90,7	92,5	62,9	57,0	52,3	69,3	63,6	53,3
	Podział Trening/Test 70/30 Brak walidacji krzyżowej Bez uwzględniania B[a]P		57,9	59,8	55,1	59,8	66,4	60,7	49,0	51,7	49,0	51,4	55,1	53,3

Dodatkowym aspektem wpływającym na jakość modelu jest wyznaczenie korelacji pomiędzy danymi oraz ewentualną redukcją ilości zmiennych wejściowych. Środowisko Azure Machine Learning pozwala zrobić to w sposób automatyczny za pomocą modułów *Filter Based Feature Selection* oraz *Permutation Feature Importance*. Zasada działania modułu *Filter Based Feature Selection* bazuje na wyliczeniu korelacji pomiędzy każdym z predyktorów, a zmienną zależną. Moduł pracuje tylko na danych numerycznych pomijając wszystkie pozostałe dane (np.: kategorie). W niniejszej pracy w module *Filter Based Feature Selection* wyznaczono współczynniki korelacji Pearsona. Wskaźnik ten może przyjmować wartości od -1 do 1. Wartość 0 oznacza brak korelacji, -1 lub 1 oznacza ścisłą korelację między zmiennymi, odpowiednio odwrotnie proporcjonalną i proporcjonalną. Moduł *Permutation Feature Importance* opiera się na wprowadzeniu do danych „szumu", czyli modyfikacji w sposób losowy kolejności danych dla każdej zmiennej. Po każdej takiej operacji dla pojedynczej zmiennej, sprawdzana jest dokładność modelu – im większy spadek dokładności tym większa istotność danej zmiennej. Moduł *Permutation Feature Importance* pracuje na każdym rodzaju danych (numeryczne, kategorie) oraz przypisuje wagi dla każdej zmiennej. Waga 0 oznacza brak zależności, waga -1 lub 1 oznacza ścisłą zależność, odwrotnie proporcjonalną lub proporcjonalną.

W tabeli 25 przedstawiono ocenę istotności parametrów dla 3 kategorii (niski/informowania/alarmowy), podziału na zbiór danych treningowych i testowych równego 70/20 dla algorytmu sztucznych sieci neuronowych z wykorzystaniem modułów *Filter Based Feature Selection* oraz *Permutation Feature Importance:*

Tabela 25. Ocena istotności parametrów na dokładność modelu

	Parametr ***(tylko wartości liczbowe)***	**Filter Based Feature Selection**	**Parametr** ***(wartości liczbowe i kategorie)***	**Permutation Feature Importance**
		Pearson		**Waga**
MEQ	Benzo[*a*]piren	0,56	Miejsce	0,26
	Temperatura	0,40	Benzo[*a*]piren	0,14
	PM_{10}	0,34	Temperatura	0,09
	Prędkość wiatru	0,29	PM_{10}	0,09
	Opad atm.	0,01	Prędkośc wiatru	0,05
			Pora roku	0,02
			Kierunek wiatru	0,01
			Opad atm.	0
CEQ	Benzo[*a*]piren	0,51	Miejsce	0,28
	Temperatura	0,45	Benzo[*a*]piren	0,09
	PM_{10}	0,19	Temperatura	0,03
	Prędkość wiatru	0,19	PM_{10}	0,02
	Opad atm.	0,04	Pora roku	0,01
			Prędkośc wiatru	0,01
			Opad	0
			Kierunek wiatru	0
TEQ	Benzo[*a*]piren	0,55	Miejsce	0,31
	Temperatura	0,41	Temperatura	0,15
	PM_{10}	0,32	Benzo[*a*]piren	0,15
	Prędkość wiatru	0,32	PM_{10}	0,12
	Opad atm.	0,02	Pora roku	0,06
			Prędkość wiatru	0,06
			Kierunek wiatru	0,02
			Opad atm.	0

Moduł *Filter Based Feature Selection* dla wszystkich wskaźników wykazał największą korelację wskaźnika *MEQ* ze stężeniem Benzo[a]pirenu, temperaturą oraz stężeniem PM_{10}. Moduł *Permitation Feature Importance* przypisał największą wagę dla parametrów miejsce, Benzo[a]piren, temperatura oraz stężenie PM_{10}. Test ten dowiódł, że do stworzenia modelu określającego stopień narażenia społeczności na wybrane ksenobiotyki, niezbędne jest określenie miejsca pobierania próbek.

Podobne wnioski płyną z artykułu [Pimpunchat i in. 2014], gdzie sprawdzono jaki wpływ mają dane na predykcję stężenia PM_{10}. Badania wykonano w Tajlandii w latach 2005-2010. Jako dane wejściowe wykorzystano średnie dobowe stężenia tlenków węgla, tlenków azotu, tlenków siarki, PM_{10} oraz dane meteorologiczne. Testy opierały się na algorytmach regresji logistycznej. Za pomocą współczynnika korelacji Pearsona stwierdzono, że największy wpływ na przewidywane wartości stężeń pyłu zawieszonego posiadają wszystkie zanieczyszczenia powietrza. Najmniejszy współczynnik Pearsona otrzymano dla temperatury oraz kierunku i prędkości wiatru. W artykule [Li i in. 2020] skupiającym się na badaniach metali ciężkich w pyle zawieszonym PM_1 w Chinach w latach 2019-2020 stwierdzono, że algorytmy sieci neuronowych nie we wszystkich przypadkach mogą zostać zastosowane do prognoz stężeń metali. Najwyższą dokładność uzyskano dla Pb, Tl i Zn, najniższą dla Ti i V. Wykazano, że model dużo lepiej sprawdza się, gdy parametrem wejściowym było stężenie pyłu $PM_{2.5}$ w porównaniu do stężeń pyłu PM_1. Przewidywane stężenia metali wykazały tendencję zgodną z wdrożonymi nadzwyczajnymi środkami zaradczymi związanymi z pandemią COVID19. Wyniki te potwierdziły wiarygodność symulacji opierających się na algorytmach sieci neuronowych dla PM_1.

Z racji pojawiającego się zjawiska przetrenowania modelu wykorzystującego algorytm lasów losowych oraz niskich dokładności algorytmów regresji logistycznej i wektorów nośnych, odrzucono te algorytmy w dalszych obliczeniach. W kolejnej części pracy zdecydowano się na zastosowanie 3 kategorii (niski/informowania/alarmowy) oraz wyłącznie algorytmu sztucznych sieci neuronowych. Dodatkowo, wykonano test dla nowego podziału danych o stosunku danych treningowych i testowych równym 80/20. Zastosowano również dodatkowy moduł walidacji krzyżowej dostępny w Azure Machine Learning. Poniżej (tabela 26) przedstawiono wyniki dla tego etapu:

Tabela 26. Macierz błędu dla algorytmu sieci neuronowych dla 3 kategorii (Niski, Informowania, Alarmowy) dla nowego podziału i walidacji krzyżowej

Sieci neuronowe – 3 kategorie

Podział:		**70/30**			**Podział:**		**80/20**			**Walidacja krzyżowa**				
MEQ %		**Przewidywana kategoria**			**MEQ %**		**Przewidywana kategoria**			**MEQ %**		**Przewidywana kategoria**		
		Niski	Info.	Alarm.			Niski	Info.	Alarm.			Niski	Info.	Alarm.
Aktualna kategoria	Niski	82,1	17,9	0,0	**Aktualna kategoria**	Niski	88,3	10,9	0,8	**Aktualna kategoria**	Niski	89,7	10,3	0,0
	Info.	13,8	86,2	0,0		Info.	11,5	82,3	6,2		Info.	5,9	94,1	0,0
	Alarm.	2,6	5,1	92,3		Alarm.	0,70	3,00	96,3		Alarm.	0,0	0,0	100,0
Dokładność		86,9			**Dokładność**		89,7			**Dokładność**		94,4		
CEQ %		**Przewidywana kategoria**			**CEQ %**		**Przewidywana kategoria**			**CEQ %**		**Przewidywana kategoria**		
		Niski	Info.	Alarm.			Niski	Info.	Alarm.			Niski	Info.	Alarm.
Aktualna kategoria	Niski	83,8	10,8	5,4	**Aktualna kategoria**	Niski	84,0	8,0	8,0	**Aktualna kategoria**	Niski	86,0	9,0	5,0
	Info.	43,5	47,8	8,7		Info.	35,7	64,3	0,0		Info.	25,4	71,6	3,0
	Alarm.	8,5	0,0	91,5		Alarm.	0,0	6,1	93,9		Alarm.	1,8	2,9	95,3
Dokładność		79,4			**Dokładność**		84,7			**Dokładność**		87,7		
TEQ %		**Przewidywana kategoria**			**TEQ %**		**Przewidywana kategoria**			**TEQ %**		**Przewidywana kategoria**		
		Niski	Info.	Alarm.			Niski	Info.	Alarm.			Niski	Info.	Alarm.
Aktualna kategoria	Niski	89,5	7,9	2,6	**Aktualna kategoria**	Niski	88,1	7,1	4,8	**Aktualna kategoria**	Niski	96,3	3,7	0,0
	Info.	20,0	80,0	0,0		Info.	8,8	84,3	6,9		Info.	5,3	94,7	0,0
	Alarm.	2,6	5,1	92,3		Alarm.	0,8	4,6	94,6		Alarm.	0,0	3,8	96,2
Dokładność		87,9			**Dokładność**		89,4			**Dokładność**		95,8		

Dotychczas model opierał się na podziale danych pomiędzy zbiór treningowy i testowy w odpowiednim stosunku. Koncepcja metody opierającej się na takim podziale jest prosta i nie wymaga dodatkowych zasobów obliczeniowych. Wadą takiego rozwiązania jest utrata sporej części danych, która trafia do zbioru treningowego, co w przypadku niewielkich zbiorów danych może znacznie wpłynąć na dokładność modelu. Dodatkowo, istnieje szansa, że do zbioru treningowego nie trafią istotne (z punktu widzenia jakości modelu) ważne informacje. Alternatywą rozwiązania problemu utraty danych jest zastosowanie w modelu walidacji krzyżowej. Walidacja krzyżowa w najprostszym przypadku opiera się na metodzie *leave-one-out* – czyli odrzuceniu pojedynczych obserwacji ze wszystkich dostępnych danych i wygenerowaniu pojedynczego modelu. Następnie odrzucana jest kolejna obserwacja ze zbioru danych wejściowych i generowany jest drugi model. Operacja *leave-one-out* powtarza się tyle razy, ile dostępnych jest obserwacji. Wadą takiego rozwiązania jest duże obciążenie obliczeniowe, ze względu na konieczność wygenerowania osobnych modeli. W praktyce częściej spotykane jest zastosowanie walidacji krzyżowej opierającej się na metodzie *K-Fold,* która polega na podzieleniu dostępnego zbioru danych wejściowych na określoną liczbę *K* podzbiorów, charakteryzujących się równą lub bardzo zbliżoną liczebnością. Następnie powtarzany jest proces odrzucenia pierwszego podzbioru i wygenerowania osobnego modelu, odrzucenia drugie podzbioru i wygenerowania drugiego modelu itd. W niniejszej pracy zastosowano walidację krzyżową opierającą się na metodzie *K-Fold*. Moduł walidacji krzyżowej dostępnej w środowisku Azure Machine Learning ma przypisaną z góry wartość *K* wynoszącą 10 podzbiorów. Wykonana analiza dla takiego zestawienia zwiększyła dokładność modelu dla każdego wskaźnika narażenia (tabela 26). Dokładność sieci neuronowych dla wskaźników *MEQ*, *CEQ* i *TEQ*, z zastosowaniem walidacji krzyżowej wynosiła, odpowiednio 94,4 %, 87,7 % oraz 95,8 % i była większa niż uzyskana dokładność dla modeli sieci neuronowych z zastosowaniem nowego podziału danych na zbiór treningowy i testowy równego 80/20 (*MEQ* – 89,7 %, *CEQ* – 84,7 %, *TEQ* – 89, 4 %). Dodatkowo walidacja krzyżowa pozwala uniknąć zjawiska nadmiernego dopasowania.

Na dokładność modeli wskaźnika narażenia ma znaczny wpływ rodzaj użytej normalizacji danych. W analizach wykorzystano 4 dostępne możliwości:

i. Brak normalizacji danych,
ii. Normalizację Min-Max – liniowe przeskalowanie każdej cechy do przedziału [0,1],
iii. Normalizację Gaussa – przeskalowanie wartości każdej cechy tak, aby miały średnią 0 i wariancję 1,
iv. Normalizację Binning (przedziałów) – tworzenie przedziałów o równym rozmiarze, a następnie normalizacja każdej wartości w każdym przedziale, dzieląc wartość przez całkowitą liczbę przedziałów.

Zestawienie danych dotyczących normalizacji z wykorzystaniem algorytmów sieci neuronowych oraz modułu walidacji krzyżowej przedstawia tabela 27. Nie uwzględniono wyników dla przypadku, gdzie nie użyto normalizacji – dokładność takiego modelu znacząco zmalała: *MEQ* – 32,4 %, *CEQ* – 42,7 %, *TEQ* – 34,1 %.

Tabela 27. Wpływ rodzaju normalizacji danych na macierz błędu dla algorytmu sieci neuronowych dla 3 kategorii (Niski, Informowania, Alarmowy) z uwzględnieniem walidacji krzyżowej

Sieci neuronowe – 3 kategorie – normalizacja danych

Normalizacja		**Min-Max**		
MEQ %		**Przewidywana kategoria**		
		Niski	Info.	Alarm.
Aktualna kategoria	Niski	87,5	9,4	3,1
	Info.	19,8	55,2	25,0
	Alarm.	0,7	6,7	92,6
Dokładność		80,7		
CEQ %		**Przewidywana kategoria**		
		Niski	Info.	Alarm.
Aktualna kategoria	Niski	86,0	5,0	9,0
	Info.	29,9	38,8	31,3
	Alarm.	3,5	0,6	95,9
Dokładność		81,8		
TEQ %		**Przewidywana kategoria**		
		Niski	Info.	Alarm.
Aktualna kategoria	Niski	85,7	10,3	4,0
	Info.	8,9	62,7	28,4
	Alarm.	2,3	8,5	89,2
Dokładność		80,4		

Normalizacja		**Gaussa**		
MEQ %		**Przewidywana kategoria**		
		Niski	Info.	Alarm.
Aktualna kategoria	Niski	84,4	14,8	0,8
	Info.	16,7	79,1	4,2
	Alarm.	1,5	3,0	95,5
Dokładność		87,2		
CEQ %		**Przewidywana kategoria**		
		Niski	Info.	Alarm.
Aktualna kategoria	Niski	80,2	11,5	8,3
	Info.	25,4	67,1	7,5
	Alarm.	2,9	2,9	94,2
Dokładność		84,4		
TEQ %		**Przewidywana kategoria**		
		Niski	Info.	Alarm.
Aktualna kategoria	Niski	87,3	9,5	3,2
	Info.	11,8	82,3	5,9
	Alarm.	1,5	6,9	91,6
Dokładność		87,4		

Normalizacja		**Binning**		
MEQ %		**Przewidywana kategoria**		
		Niski	Info.	Alarm.
Aktualna kategoria	Niski	91,4	7,8	0,8
	Info.	16,7	78,1	5,2
	Alarm.	0,7	1,5	97,8
Dokładność		90,2		
CEQ %		**Przewidywana kategoria**		
		Niski	Info.	Alarm.
Aktualna kategoria	Niski	88,4	8,3	3,3
	Info.	26,9	61,2	11,9
	Alarm.	2,9	4,2	92,9
Dokładność		85,5		
TEQ %		**Przewidywana kategoria**		
		Niski	Info.	Alarm.
Aktualna kategoria	Niski	90,5	8,7	0,8
	Info.	9,8	84,3	5,9
	Alarm.	1,5	3,1	95,4
Dokładność		90,5		

Dane wejściowe wykorzystywane do określania stopnia narażenia *MEQ, CEQ* i *TEQ* stanowią odrębne parametry o różnych zakresach wartości. Dla przykładu, ciśnienie atmosferyczne wahało się od 960 hPa do 1007,1 hPa, temperatura: -7,9 – 31,5 °C, stężenie PM_{10}: 8,06 – 234,14 $\mu g/m^3$. Proces normalizacji polega na zmianie wartości kolumn liczbowych w zestawie danych wejściowych w celu używania wspólnej skali bez zniekształcenia różnic w zakresach wartości lub utraty informacji. Zbyt duża różnica pomiędzy skalami poszczególnych parametrów może powodować problemy podczas próby połączenia tych wartości jako funkcji podczas modelowania. Na poprawę dokładności stworzonego modelu najlepiej wpływa zastosowanie normalizacji danych typu Gausa lub Binning. Dla modeli bazujących na algorytmach sieci neuronowych, walidacji krzyżowej oraz normalizacji typu Gaussa uzyskano dokładność modelu 87,2 % dla wskaźnika *MEQ*, 84,4 % dla wskaźnika *CEQ* oraz 87,4 % dla wskaźnika *TEQ*. Dla tego samego modelu z zastosowaniem normalizacji typu Binning uzyskano dokładność 90,2 % dla wskaźnika *MEQ*, 85,5 % dla wskaźnika *CEQ* oraz 90,5 % dla wskaźnika *TEQ*.

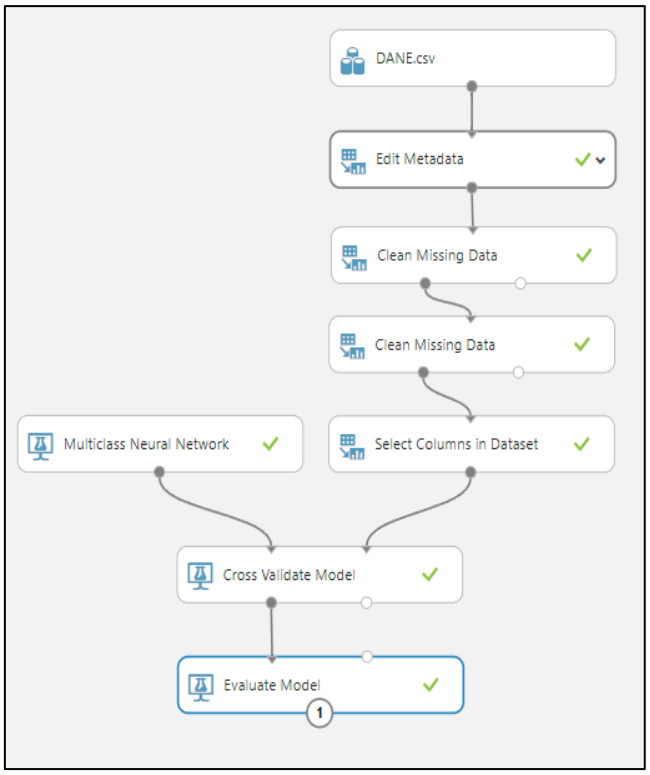
3.6.5. Implementacja modelu w środowisku *Deploy Web Serive*

Przeprowadzone badania pozwoliły stworzyć ostateczną wersję modelu identyfikacji stopnia narażenia względem wskaźników *MEQ, CEQ* i *TEQ*. Finalny model opierał się na algorytmach sztucznych sieci neuronowych (liczba ukrytych warstw – 1, liczba ukrytych węzłów – 100, szybkość uczenia się 0,1, ilość iteracji – 100, L1 – 0,1), walidacji krzyżowej (metoda K-Fold, 10 podzbiorów), normalizacji typu Binning. Kolejnym etapem pracy była implementacja powstałego modelu w ogólnodostępnym środowisku *Deploy Web Service*.

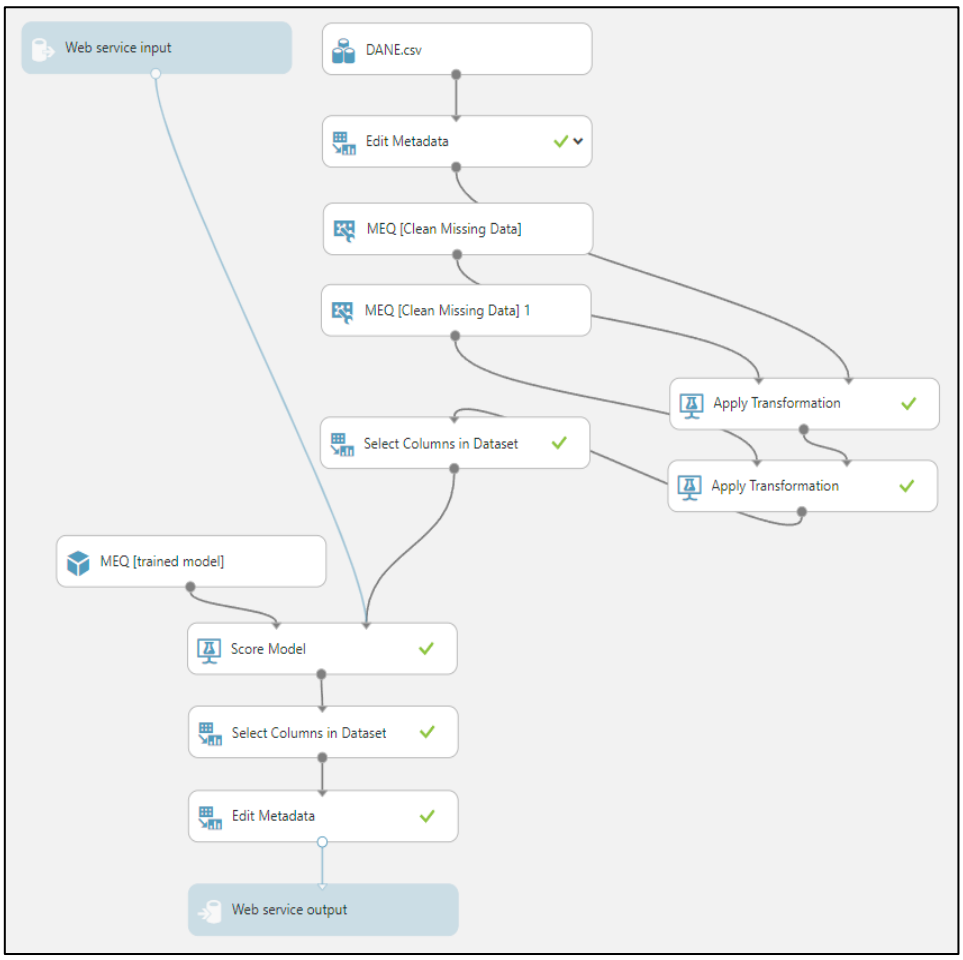
Środowisko Azure Machine Learning pozwala na publikację modelu, czyli zamienienie na publicznie dostępne API w formacie REST (*REpresentational State Transfer*), czyli protokołu powszechnie wykorzystywanego w przypadku tworzenia aplikacji internetowych. REST API są najpopularniejszymi sposobami interakcji programów i urządzeń w nowoczesnych technologiach obliczeniowych. Takie API może działać w dwóch trybach:

- pojedynczego zapytania – model nie będzie musiał przetwarzać dużej ilości danych (krótki czas oczekiwania na odpowiedź)
- zbiorowych zapytań – informacje do przetworzenia trafiają do odbiorcy w dużych porcjach.

W niniejszej pracy dokonano implementacji modelu, a następnie podjęto próbę wykorzystania go we własnej aplikacji za pomocą *Predictive Web Service* dostępnego w AML. Poniżej (rysunek 39 i rysunek 40) przedstawiono początkowy schemat modelu przed i po implementacji:



Rysunek 39. Początkowy schemat modelu przed wdrożeniem

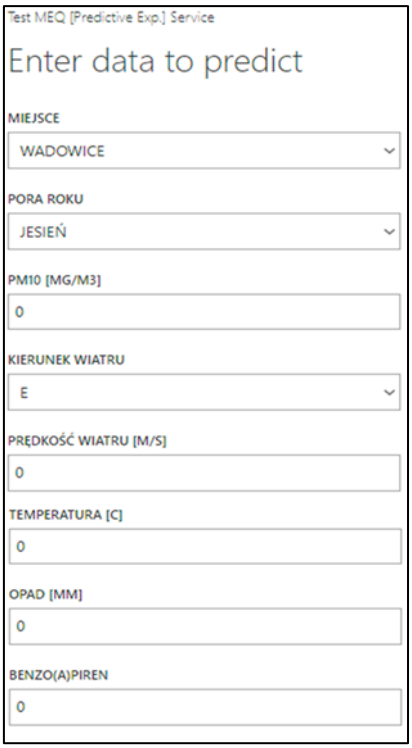


Rysunek 40. Końcowy schemat modelu po wdrożeniu

W rezultacie wdrażania, otrzymano zmieniony model początkowy (rysunek 40). Podstawowym elementem takiego układu są nowe bloki *Web Service input* oraz *Web Service output* odpowiednio wejście i wyjście uzyskanego API. Cały schemat został również rozbudowany o moduł *MEQ [trained model]* stanowiący cały podstawowy model (rysunek 39). W pierwszej kolejności po uzyskaniu zaimplementowanego modelu, w module *Select Columns in Dataset* należy określić parametry, które będą traktowane jako wejściowe. W tym przypadku został usunięty parametr *Wskaźnik narażenia MEQ* – jest to kategoria, którą chcemy przewidywać na podstawie pozostałych danych. Całość schematu została ręcznie uzupełniona o dodatkowy moduł *Select Columns in Dataset* po *Score Model*, gdzie została wybrana kategoria, którą chcemy przewidywać (*Scored Labels*) – upraszcza to uzyskane dane wyjściowe do pojedynczej informacji. Moduł *Edit Metadata* znajdujący się na końcu modelu odpowiada za zmianę nazwy *Scored Labels* na *Wskaźnik narażenia MEQ*. W rezultacie uzyskane API jest już gotowe do użytku (rysunek 41):

Rysunek 41. API zaimplementowanego modelu

Uzyskano w ten sposób 2 tryby modelu. Pierwszy z nich *REQUEST/RESPONSE* pozwala na wywołanie jednorazowe. Nadaje się ono świetnie do testów lub w sytuacji, w której użytkownicy będą generować dużą ilość pojedynczych zapytań. W przypadku, kiedy do modelu będą dostarczane dane w dużych ilościach, lepszym rozwiązaniem będzie skorzystanie z drugiego trybu *BATCH EXECUTION*. W pracy przeprowadzono test dla *REQUEST/RESPONSE* dla przykładowych warunków, które mogą pojawić się w danym miejscu. Po wybraniu ikony *Test* użytkownik proszony jest o wypełnienie danych wybranych na etapie *Select Columns in Dataset* (rysunek 42):



Rysunek 42. Platforma do uzupełnienia danych w BATCH EXECUTION

W danych *Miejsce* należy wybrać, gdzie wykonywana jest analiza. W niniejszej pracy są to tylko 2 miejsca: Wadowice lub Kraków. Należy zdefiniować *Porę roku* (Wiosna, Lato, Jesień, Zima) i *Kierunek wiatru* (N, NE, E, SE, S, SW, W, NW). Dla *PM_{10}, Prędkość wiatru, Temperatura, Opad* oraz *Benzo[a]piren* należy podać wartości odpowiednio średniego dobowego stężenia pyłu PM_{10} [$\mu g/m^3$ – API nie uwzględnia symbolu µ], prędkości wiatru [m/s], temperaturę [°C], opad atmosferyczny [mm] oraz stężenie Benzo[*a*]pirenu [ng/m^3]. Po uzupełnieniu wszystkich danych otrzymamy informacje, jaki w danym momencie, dla konkretnych parametrów wystąpi stopień narażenia *MEQ*. Poniżej (tabela 28) zaprezentowano przykłady uzyskanych stopni narażenia *MEQ*:

Tabela 28. Przykładowe stopnie narażenia MEQ uzyskane po implementacji modelu

nr	Miejsce	Pora roku	PM_{10} [$\mu g/m^3$]	Kierunek wiatru	Prędkość wiatru [m/s]	Temperatura [°C]	Opad [mm]	B[*a*]P [ng/m^3]	Stopień narażenia *MEQ*
(1)	Kraków	Zima	70	N	1,0	5	10	3	*Alarmowy*
	Wadowice	Zima	70	N	1,0	5	10	3	*Alarmowy*
(2)	Kraków	Wiosna	45	SW	1,4	12	0	1	*Niski*
	Wadowice	Wiosna	45	SW	1,4	12	0	1	*Alarmowy*
(3)	Kraków	Lato	10	S	2,0	22	30	0,1	*Niski*
	Wadowice	Lato	10	S	2,0	22	30	0,1	*Informowania*
(4)	Kraków	Jesień	50	SE	1,0	17	0	2	*Informowania*
	Wadowice	Jesień	50	SE	1,0	17	0	2	*Alarmowy*

Uzyskane analizy potwierdzają, że na dokładność modelu zasadniczy wpływ ma miejsce. Dla przewidywań nr (2), (3) i (4) z Krakowa wskaźnik narażenia uzyskiwany jest na niższym poziomie niż dla Wadowic dla tych samych parametrów wejściowych. Związane jest to ściśle z charakterystyką lokalnych źródeł emisji oraz występującymi różnymi stosunkami pomiędzy stężeniami Benzo[a]pirenu, a pozostałymi WWA, które zostały wyznaczone osobno dla Krakowa i Wadowic. Stosunki te posłużyły do wyznaczenia stężeń wszystkich WWA względem przyjętego średniego dobowego stężenia Benzo[a]pirenu 1 ng/m^3, a następnie do określenia wartości wskaźnika *MEQ*. W celu stworzenia nowego systemu identyfikującego stopień narażenia społeczności, należy rozszerzyć zestaw danych o nowe miejsca pobierania próbek.

PODSUMOWANIE

Przedstawiona dysertacja doktorska opierała się na interdyscyplinarnym podejściu łączącym analizy chemiczne, modele dyspersyjne oraz dział uczenia maszynowego. W pracy przedstawiono nowe podejście, w którym starałem się opracować model identyfikujący stopień narażenia społeczności na wybrane zanieczyszczenia powietrza względem.

W niniejszej rozprawie doktorskiej przeprowadzono badania próbek pyłu zawieszonego PM_{10} pobranego na terenie Wadowic w 2017 roku oraz na terenie Krakowa w latach 2020/2021. Uzyskany materiał został poddany analizie fizyko-chemicznej składu pod względem zawartości wielopierścieniowych węglowodorów aromatycznych za pomocą chromatografii gazowej sprzężonej ze spektrometrią mas GC-MS.

W pierwszej kolejności opracowano metodykę analizy wielopierścieniowych węglowodorów aromatycznych w pyle zawieszonym PM_{10}. Uzyskane wyniki metody GC-MS jednoznacznie stwierdzają jej wysoką dokładność, precyzję i czułość, a także liniowość. Odzysk badanych związków uzyskano na wysokim, zadowalającym poziomie.

W dalszej części wykonano analizy 252 próbek uzyskanych podczas kampanii poboru. Analizy obejmowały 16 podstawowych WWA, określonych przez US EPA (*United States Environmental Protection Agency*) za najbardziej szkodliwe: Acenaften, Acenaftylen, Antracen, Benzo[*a*]antracen, Benzo[a]piren, Benzo[*b*]fluoranten, Benzo[*ghi*]perylen, Benzo[*k*]fluoranten, Chryzen), Dibenzo[*a,h*]antracen, Fenantren, Fluoranten, Fluoren, Indeno[*1,2,3-cd*]piren, , Piren i Naftalen.

Otrzymane informacje dotyczące poziomów stężeń WWA zostały wykorzystane do określenia profili źródeł zanieczyszczeń, profili narażenia oraz wartości wskaźników ekwiwalentu toksyczności wielopierścieniowych węglowodorów aromatycznych zalecanych przez EPA: ekwiwalentu działania mutagennego względem B[*a*]P (ang. *mutagenic equivalent, MEQ*), ekwiwalentu działania toksycznego względem B[*a*]P (ang. *toxic equivalent, TEQ*) i ekwiwalentu działania kancerogennego względem 2,3,7,8-tetrachlorodienzo-p-dioksyny (ang. *carcinogenic equivalent, CEQ*).

Wartości wskaźników *MEQ, CEQ* i *TEQ* zostały wykorzystane do stworzenia modelu bazującego na algorytmach uczenia maszynowego w celu identyfikacji stopnia narażenia społeczności na wybrane ksenobiotyki. Obliczenia wykonano w ogólnodostępnym środowisku Azure Machine Learning.

W wyniku przeprowadzonych badań i analiz, wnioski z pracy można sformułować odnosząc się bezpośrednio do postawionych tez badawczych:

1. Stężenia pyłu zawieszanego PM_{10} oraz wielopierścieniowych węglowodorów aromatycznych wykazują wzajemne zależności, które są charakterystyczne dla poszczególnych miast oraz panujących warunków synoptycznych

Odnotowano sezonową zmienność stężenia, zarówno pyłu zawieszonego PM_{10} i wszystkich WWA. Najwyższe średnie dobowe stężenie PM_{10} w Wadowicach wynosiło 116,77 $\mu g/m^3$ (13.03.2017), natomiast w Krakowie było równe 234,14 $\mu g/m^3$ (11.02.2020). W obu przypadkach najniższe wartości średnich dobowych stężeń pyłu PM_{10} odnotowano w sezonach poza grzewczych. Zaobserwowano znaczne różnice w wartościach stężeń PM_{10} w zależności od miejsca pobierania próbek. Podobnych obserwacji dokonano dla stężeń WWA, gdzie zakres stężeń wszystkich WWA dla Wadowic i Krakowa wynosił, odpowiednio od 0,08 ng/m^3 do 24,73 ng/m^3 oraz 0,01 ng/m^3 do 18,85 ng/m^3. Najwyższe dobowe stężenie w Wadowicach

odnotowano dla Benzo[a]antracenu (24,73 ng/m^3 – 13.03.2017) oraz Chryzenu (21,52 ng/m^3 – 13.03.2017). W Krakowie dla Fenantrenu (18,85 ng/m^3 – 27.02.2020) i Benzo[b]fluorantenu (16,95 ng/m^3 – 24.02.2020). Największe poziomy stężeń Benzo[a]pirenu w Wadowicach oznaczono 27.03.2017 oraz 09.03.2017 i wynosiły, odpowiednio 20,93 ng/m^3 oraz 19,69 ng/m^3, a wartość średnia dobowa w okresie badań wyniosła 3,32 ng/m^3. Dla Krakowa średnie miesięczne stężenie Benzo[a]pirenu wyniosło 1,56 ng/m^3. Największe wartości oznaczono 24.02.2021 (10,16 ng/m^3) oraz 27.02.2020 (8.91 ng/m^3). Raporty Wojewódzkiego Inspektoratu Ochrony Środowiska wykazały, że wiele polskich miast charakteryzuje złą jakością powietrza, stężenia B[a]P znacznie przewyższają dopuszczalne normy (wartość średnioroczna 1 ng/m^3), a sam związek może stanowić do 20 % wszystkich WWA.

W niniejszej pracy wykazano zróżnicowane wartości wskaźników diagnostycznych. Wskaźniki te, to stosunki konkretnych stężeń WWA, za pomocą których można oszacować pochodzenie pyłu zawieszonego. W przypadku Wadowic i Krakowa, największym procentowym udziałem charakteryzował się transport, jako źródło zanieczyszczeń. Głównie wskazano na udział spalania paliw w silnikach spalinowych. Odnotowano typowe zróżnicowanie sezonowe źródeł emisji dla obu badanych miejscowości – w sezonie grzewczym (oprócz transportu) dominowało ogrzewanie domów (spalanie węgla, drewna), w sezonie poza grzewczym był to przede wszystkim transport.

Dla Wadowic wyznaczono 96 % udział ciężkich WWA w stosunku do wszystkich WWA zidentyfikowanych w próbkach w sezonie grzewczym oraz 93 % udział w sezonie poza grzewczym. W sezonie grzewczym i poza grzewczym dla Krakowa udział ten wynosił, odpowiednio 93 % oraz 71 %. W pyle zawieszonym zebranym w Wadowicach węglowodory silnie kancerogenne stanowiły 80 % udział względem wszystkich WWA w sezonie grzewczym i poza grzewczym. W Krakowie udziały te charakteryzowały się niższymi wartościami, wynosiły 63 % oraz 53 %, odpowiednio w sezonie grzewczym oraz poza grzewczym. Znaczne różnice zostały również odnotowane w wartościach wskaźników ekwiwalentów toksyczności WWA. Wartość MEQ, CEQ i TEQ dla Wadowic w okresie grzewczym wynosiły 22,08 ng/m3, 72,23 ng/m3 i 94,00 pg/m3, gdzie dla Krakowa były to wartości, odpowiednio 4,65 ng/m3, 5,33 ng/m3 i 24,10 pg/m3.

W badaniach przeprowadzonych dla Wadowic wykazano niższe stężenia PM_{10} niż w Krakowie. Nie jest to jednoznaczne z zawartością WWA. Kraków pomimo większych stężeń pyłu, charakteryzował się niższymi średnimi dobowymi stężeniami WWA, w tym B[a]P. Analizy dowiodły niższych udziałów kancerogennych WWA w próbkach pyłu pobranego w Krakowie. Wartości wskaźników ekwiwalentu toksycznego działania WWA dla Wadowic były znacznie wyższe niż wskaźniki dla Krakowa. Potwierdza to tezę, iż wzajemne zależności pomiędzy stężeniami PM_{10} i WWA są charakterystyczne dla poszczególnych miast.

2. Modele deterministyczne transportu zanieczyszczeń pomagają w określeniu lokalizacji ich źródeł (lokalne, napływowe)

W celu uzyskania informacji dotyczących transportu zanieczyszczeń z wybranych obszarów z okolic Wadowic oraz Krakowa, przeprowadzono analizy częstotliwości występujących kierunków napływu mas powietrza. Badania wykonano za pomocą modelu HYSPLIT. Do analizy trajektorii wstecznej wybrano dni z najwyższym dziennym stężeniem pyłu PM_{10}. Analizy bazowały na wygenerowaniu 24 trajektorii wstecznych o długości 12 h dla każdego z badanych miejsc.

Dla Wadowic, trajektorie wykazały napływ mas powietrza z terenów znajdujących się od miasta na północ, północny wschód oraz zachód. W przypadku Krakowa był to kierunek

południowo zachodni oraz wschodni. Pomimo napływu mas powietrza nad Wadowice z kierunków północnych, charakteryzujących się wysoko rozwiniętym przemysłem zlokalizowanym na Śląsku (Katowice, Dąbrowa Górnicza), odnotowano niskie średnie dobowe stężenia wszystkich WWA, łącznie z B[a]P. Na północ od Wadowic znajdują się również rozległe Parki Krajobrazowe (Tenczyński, Rudniański, Dolinki Krakowskie). Obszary te charakteryzuje się niską gęstością zaludnienia. Czynniki te mogą być głównym powodem niskiej zawartości szkodliwych wielopierścieniowych węglowodorów aromatycznych w pyle zawieszonym PM_{10}. Kierunek północno wschodni oraz zachodni od Wadowic to kotliny o znacznie wyższej gęstości zaludnienia z dominującymi paleniskami domowymi, co przekłada się na wyższe wartości szkodliwych WWA (w tym B[a]P), pomimo podobnych wartości średnich dobowych stężeń PM_{10}, co w przypadku kierunków północnych.

Analiza trajektorii wstecznych dla Krakowa, wykazała występowanie kierunków znad obszarów wyżynnych i górzystych charakteryzujących się niską gęstością zaludnienia (kierunek południowo zachodni). Są to tereny Beskidu Małego, Makowskiego i Wyspowego. Na tych obszarach dochodzi do kumulacji zanieczyszczeń w dolinach i ograniczenia ich dalszego transportu w wyniku utrudnionego naturalnego procesu przewietrzania. Pomimo znacznie wyższych stężeń pyłu zawieszonego PM_{10} w Krakowie, odnotowano dużo niższe wartości stężeń WWA oraz B[a]P niż w Wadowicach. Wiąże się to z wprowadzonym zakazem spalania paliw stałych – na mocy uchwały Nr XVII/243/16 Sejmiku Województwa Małopolskiego z dnia 15 stycznia 2016r. Kraków został objęty bezwzględnym zakazem spalania paliw stałych. 24 kwietnia 2017 roku Sejmik Województwa Małopolskiego przegłosował uchwałę o dopuszczeniu do spalania paliw o odpowiedniej jakości. Z racji tego, że otaczające Kraków miejscowości nie obowiązuje zakaz spalania paliw stałych, powstało określenie „obwarzanka Krakowskiego", obszaru, z którego obserwuje się napływ pyłu zawieszonego na wskutek spływów chłodnego powietrza i w konsekwencji kumulację zanieczyszczeń nad obszarem miasta.

Modele deterministyczne pozwoliły oszacować możliwość napływu zanieczyszczeń z sąsiednich obszarów badanych miejsc. W analizowanych przypadkach można stwierdzić, że obserwowane wysokie stężenia zanieczyszczeń pochodzą zarówno z emisji lokalnej, jak i z terenów znajdujących się w niedalekiej odległości od Wadowic i Krakowa.

3. Uczenie maszynowe pozwala określić stopień narażenia społeczności na wybrane zanieczyszczenia powietrza (wielopierścieniowe węglowodory aromatyczne) z wykorzystaniem algorytmów sztucznych sieci neuronowych oraz wskaźników mutagenności (*MEQ*), kancerogenności (*CEQ*) i ekwiwalentu toksycznego działania (*TEQ*) pozwalając na przewidywanie poziomu wskaźników narażenia wymagających informacji o stężeniach całej gamy związków WWA jedynie na podstawie danych meteorologicznych, stężenia frakcji PM_{10} pyłu oraz stężenia B[a]P.

Głównym założeniem części uczenia maszynowego było uwzględnienie wielopierścieniowych węglowodorów aromatycznych oraz wyznaczonych wartości wskaźników: mutagenności (*MEQ*), toksyczności (*CEQ*) i ekwiwalentu toksycznego działania (*TEQ*) w celu wykorzystania uczenia maszynowego do określenia stopnia narażenia społeczności na podstawie analizy zanieczyszczenia powietrza. Sporządzono modele uwzględniające 4 najczęściej wykorzystywane algorytmy uczenia maszynowego: wektory nośne, regresję logistyczną, lasy losowe, sztuczne sieci neuronowe. Największą dokładność uzyskano dla sieci neuronowych oraz lasów losowych zarówno dla 2 kategorii (niski/wysoki) i 3 kategorii (niski/informowania/alarmowy). W przypadku lasów losowych odnotowano

zjawisko przeuczenia się modelu. Podjęto decyzję o wykorzystaniu w dalszych analizach wyłącznie algorytmu sztucznych sieci neuronowych. Wybrano układ z 3 kategoriami w celu dokładniejszego oszacowania poziomów narażenia. Aby uniknąć utraty znacznej ilości danych na etapie podziału danych na zbiór treningowy i testowy, zastosowano walidację krzyżową. Zauważono, że na wyniki badań ma znaczny wpływ miejsce, z którego pobrano próbki. Dokładność uzyskanych modeli sieci neuronowych była na zadowalającym poziomie (> 85 %). Zaimplementowany model REST API wykazał zasadnicze różnice w stopniu narażenia *MEQ* dla tych samych parametrów wejściowych (pora roku, temperatura, stężenie PM_{10} i B[a]P, kierunek i prędkość wiatru, opad atmosferyczny) podczas zmiany parametru Miejsce (tabela 28). Ma to związek z profilem źródeł emisji zanieczyszczeń dla Wadowic oraz Krakowa, a co za tym idzie, różnymi stosunkami pomiędzy stężeniem B[a]P, a pozostałymi WWA, które zostały wykorzystane do wyznaczanie wartości granicznych zakresów wskaźników *MEQ, CEQ* i *TEQ*.

Zaproponowany model wskaźników oceny jakości powietrza stwarza możliwości uwzględniające stopień narażenia na kancerogenne węglowodory aromatyczne znajdujące się w powietrzu. Zastosowanie sztucznych sieci neuronowych do wyznaczania stopnia narażenia pozwala określić stan narażenia w oparciu o obserwacje ograniczone do parametrów meteorologicznych, stężenia PM_{10} i B[a]P w stanie rzeczywistym. Otrzymane wyniki analiz dla próbek pobranych w Wadowicach oraz Krakowie są zadowalające i skłaniają do dalszych badań.

Identyfikacja stopnia narażenia społeczności w odniesieniu do zawartości kancerogennych WWA jest niezbędna. Obecne systemy alarmowe przewidują wyłącznie stężenia pyłów zawieszonych oraz Benzo[a]pirenu, pomijając pozostałe wielopierścieniowe węglowodory aromatyczne, które posiadają silne właściwości kancerogenne i mutagenne. Użycie modeli bazujących na algorytmach sieci neuronowych w kompleksowej ocenie stanu narażenia i predykcji stopni narażenia może skutecznie pomóc w podejmowaniu decyzji w celu zapobiegania negatywnym i groźnym skutkom zanieczyszczenia powietrza mimo braku pomiarów pozostałych WWA. Wskaźniki *MEQ*, *CEQ* i *TEQ* mogą stanowić podstawę codziennych informacji o jakości powietrza oraz wczesnego ostrzegania i alarmowania o sytuacjach, które mogą powodować znaczne zagrożenie dla zdrowia człowieka, a także środowiska. Modele sieci neuronowych mogą być niezbędną częścią instrumentarium zarządzania bezpieczeństwem ekologicznym na całym świecie.

Bibliografia

Ali AN, Nassreddine G, Younis J. 2022. Air Quality prediction using Multinomial Logistic Regression. Journal of Computer Science and Technology Studies. 4(2), 71–78. doi: 10.32996/jcsts.2022.4.2.9.

Balas VE, Hassanien AE, Chakrabarti S, Mandal L. 2021. Proceedings of International Conference on Computational Intelligence, Data Science and Cloud Computing. Lecture Notes on Data Engineering and Communications Technologies. Springer. 62. doi: 10.1007/978-981-33-4968-1.

Bartulewicz J, Gawłowski J, Bartulewicz E. 1997. Zastosowanie chromatografii gazowej i cieczowej do analizy zanieczyszczeń środowiska. Państwowa Inspekcja Ochrony Środowiska.

Borchers N, Horsley J, Palmer A, Morgan G, Tham R, Johnston F. 2019. Association between fire smoke fine particulate matter and asthma-related outcomes: Systematic review and meta-analysis. Environmental Research. 179:108777. doi:10.1016/j.envres.2019.108777.

ChessBasse. 2013. A play based on Kasparov vs Deep Blue [online]. Chessbase.com. [Dostęp 21.02.2023]. Dostępny w https://en.chessbase.com/post/a-play-based-on-kasparov-vs-deep-blue-180813

Clark RB, Lewinski MA, Loeffelholz MJ, Tibbetts RJ. 2009. Cumitech 31A. Verification and validation of procedures in the clinical microbiology laboratory. Coordinating ed, Sharp SE. ASM Press. Washington. DC.

Chen CWS, Chiu LM. 2021. Ordinal Time Series Forecasting of the Air Quality Index. Entropy. 23(9):1167. doi: 10.3390/e23091167.

Draxler RR, Hess GD. 1997. Description of the HYSPLIT_4 modeling system. NOAA Technical Memorandum ERL ARL-224. NOAA Air Resources Laboratory. Silver Spring. MD.

Draxler RR, Hess GD. 1998. An overview of the HYSPLIT_4 modeling system of trajectories, dispersion, and deposition. Australian Meteorological Magazine. 47:295–308.

Draxler RR. 1999. HYSPLIT4 user's guide. NOAA Technical Memorandum ERL ARL-230. NOAA Air Resources Laboratory. Silver Spring. MD

Durant J, Busby W, Lafleur A, Penman B, Crespi C. 1996. Human cell mutagenicity of oxygenated, nitrated and unsubstituted polycyclic aromatic hydrocarbons associated with urban aerosols. Mutat Res. 371(3–4):123–57

Dyrektywa Parlamentu Europejskiego i Rady 2008/50/WE z dnia 21 maja 2008r w sprawie jakości powietrza i czystszego powietrza dla Europy. https://eur-lex.europa.eu/legal-content/pl/LSU/?uri=CELEX:32008L0050)

European Environment Agency. 2017. Air quality in Europe – 2017 report. 13/2017. ISBN 978-92-9213-921-6. doi:10.2800/850018.

Farhadi R, Hadavifar M, Moeinaddini M, Amintoosi M. 2020. Prediction of Air Pollutants Concentration Based on Meteorological Factors in Warm and Cold Season by Artificial Neural Network and Linear Regression, Case Study: Tehran. Journal of Natural Environment. 73(1):115–127. doi:10.22059/jne.2020.278331.1681.

Frydrychowicz W, Szymańska K. 2008. Zagadnienie sztucznych sieci neuronowych w dynamicznych procesach niestandardowej ekonomii. Scientific Bulletin of Chelm Section of Mathematics and Computer Science.

Furman P, Styszko K, Skiba A, Zięba D, Zimnoch M, Kistler M, Kasper-Giebl A, Gilardoni S. 2021. Seasonal variability of PM_{10} chemical composition including 1,3,5-triphenylbenzene, marker of plastic combustion and toxicity in Wadowice, South Poland. Aerosol and Air Quality Research. 21(3):1–12. doi:10.4209/aaqr.2020.05.0223.

Główny Inspektorat Ochrony Środowiska (GIOŚ). https://powietrze.gios.gov.pl/pjp/archives

Główny Inspektorat Ochrony Środowiska. 2016. Pyły drobne w atmosferze. Kompendium wiedzy o zanieczyszczeniu powietrza pyłem zawieszonym w Polsce. Biblioteka Monitoringu Środowiska. Warszawa. ISBN:978-83-61227-73-1.

Główny Inspektorat Ochrony Środowiska. 2017. Zanieczyszczenie powietrza wielopierścieniowymi węglowodorami aromatycznymi na stacjach tła miejskiego w 2016 roku. Narodowy Fundusz Ochrony Środowiska i Gospodarki Wodnej. Warszawa.

Główny Inspektorat Ochrony Środowiska. 2019. Pięcioletnia Ocena Jakości Powietrza w Województwie Małopolskim. Raport Wojewódzki za lata 2014-2018. Regionalny Wydział Monitoringu Środowiska w Krakowie. Kraków.

Główny Urząd Statystyczny. 2023. Dane o Wadowicach [online]. Dostęp 22.02.2023. Dostępny w https://stat.gov.pl/obszary-tematyczne/ludnosc/ludnosc/ludnosc-stan-i-struktura-ludnosci-oraz-ruch-naturalny-w-przekroju-terytorialnym-stan-w-dniu-30-06-2021,6,30.html

Google Maps 2022. Zdjęcia satelitarne dla Krakowa i Wadowic [online]. Dostęp 07.10.2022. Dostępny w:
Kraków –
https://www.google.com/maps/@50.0654071,19.9159554,4858m/data=!3m1!1e3
Wadowice –[
https://www.google.com/maps/search/wadowice+rynek/@49.8830198,19.4840738,4877m/data=!3m2!1e3!4b1.

Henderson H. 2007. The Automated Mathematician. Artificial Intelligence: Mirrors for the Mind, Milestones in Discovery and Invention. Infobase Publishing. 93–94. ISBN 9781604130591.

Holtzer M, Grabowska B. 2010. Podstawy ochrony środowiska z elementami zarządzania środowiskowego. Wydawnictwo AGH. Kraków.

IARC. 1987. International Agency for Research on Cancer. Overall evaluation of carcinogenicity: An updating of IARC Monographs. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. 7(1-42).

Jakubowska M. 2020. Walidacja metod analitycznych. Raport z walidacji. Katedra Chemii Analitycznej WIMiC AGH.

Jamhari AA, Sahani M, Latif MT, Chan KM, Tan HS, Khan MF, Tahir NM. 2014. Concentration and source identification of polycyclic aromatic hydrocarbons (PAHs) in PM_{10}

of urban, industrial and semi-urban areas in Malaysia. Atmosphetic Environment. 86:16–27. doi:10.1016/j.atmosenv.2013.12.019

Janka RM. 2022. Zanieczyszczenia pyłowe i gazowe. Podstawy obliczania i sterowania poziomem emisji. Wydawnictwo Naukowe PWN. ISBN: 978-83-01-17455-2.

Jo E-J, Lee W-S, Jo H-Y, Kim C-H, Eom J-S, Mok J-H, Kim M-H, Lee K, Kim K-U, Lee M-K, Park H-K. 2017. Effects of particulate matter on respiratory disease and the impact of meteorological factors in Busan, Korea. Respiratory Medicine. 124:79–87. doi: 10.1016/j.rmed.2017.02.010.

Johney A, Samitha SJ, Namboothiri LV. 2020. Foresight of Health Risk Based on Air Pollutants' Air Quality Index Values. International Journal of Innovative Technology and Exploring Engineering (IJITEE). 9(6). doi: 10.35940/ijitee.F3320.049620.

Jung Y-J, Cho K-W, Lee J-S, Oh C-H. 2021. Comparative analysis of multiple classification models to improve PM_{10} prediction performance. International Journal of Electrical and Computer Engineering (IJECE). 11(3):2500–2507. doi: 10.11591/ijece.v11i3.pp2500-2507.

Kaleta D, Kozielska B. 2023. Spatial and Temporal Volatility of $PM_{2.5}$, PM_{10} and PM_{10}-Bound B[a]P Concentrations and Assessment of the Exposure of the Population of Silesia in 2018–2021. International Journal of environmental Research and Public Health. 20(1):138. doi: 10.3390/ijerph20010138.

Kaplan A, Haenlein M. 2019. Siri, Siri in my Hand, who's the Fairest in the Land? On the Interpretations, Illustrations and Implications of Artificial Intelligence. Business Horizons. 62(1):15–2. doi:10.1016/j.bushor.2018.08.004

Kubiak M. 2013. Wielopierścieniowe węglowodory aromatyczne (WWA) – ich występowanie w środowisku i w żywności. Probl Hig Epidemiol. 94(1):31–36.

Kulshrestha M, Singh R, Ojha V. 2019. Trends and source attribution of PAHs in fine particulate matter at an urban and a rural site in Indo-Gangetic plain. Urban Climate. 29:1–14. doi:10.1016/j.uclim.2019.100485.

Kobylec M. 2003. Sztuczne sieci neuronowe. Rozpoznawanie obrazu [online]. Poltynk.pl Dostęp 21.02.2023. Dostępny w http://www.poltynk.pl/marcin/obraz.html

Konieczka P, Namieśnik J, Zygmunt B, Bulska E, Świtaj-Zawada A, Naganowska A, Kremer E, Rompa M. 2004. Ocena i kontrola jakości wyników analitycznych. Centrum Doskonałości Analityki i Monitoringu Środowiska. Gdańsk.

Kozielska B. 2018. Health hazards from polycyclic aromatic hydrocarbons bound to submicrometer particles in Gliwice (Poland). MATEC Web of Conferences 247:1–8. doi:10.1051/matecconf/201824700034.

Kozielska B, Rogula-Kozłowska W, Pastuszka JS. 2013. Traffic emission effects on ambient air pollution by PM2,5-related PAH in Upper Silesia, Poland. International Journal of Environment and Pollution, 53(3–4):245–264. doi:10.1504/IJEP.2013.059920.

Kozielska B, Rogula-Kozłowska W, Rogula-Kopiec P, Jureczko I. 2016. Wielopierścieniowe węglowodory aromatyczne w różnych frakcjach pyłu zawieszonego w powietrzu obszarów

zdominowanych emisją komunikacyjną. Inżynieria Ekologiczna. 49:25–32. doi:10.12912/23920629/64512.

Li H, Dai Q, Yang M, Li F, Liu X, Zhou M, Qian X. 2020. Heavy metals in submicronic particulate matter (PM_1) from a Chinese metropolitan city predicted by machine learning models. Chemosphere. 261:127571. doi:10.1016/j.chemosphere.2020.127571.

Li M, Feng Y, Wang K, Yong WF, Yu L, Chung TS. 2017. Novel Hollow Fiber Air Filters for the Removal of Ultrafine Particles in $PM_{2.5}$ with Repetitive Usage Capability. Environmental Science and Technology. 51(17):10041–10049. doi:10.1021/acs.est.7b01494.

Li Y, Tao Y. 2017. PM_{10} Concentration Forecast Based on Wavelet Support Vector Machine. 2017 International Conference on Sensing, Diagnostics, Prognostics, and Control. doi:10.1109/sdpc.2017.79.

Li Z, Porter EN, Sjödin A, Needham LL, Lee S, Russell AG, Mulholland JA. 2009. Characterization of PM2,5-bound polycyclic aromatic hydrocarbons in Atlanta-Seasonal variations at urban, suburban, and rural ambient air monitoring sites. Atmospheric Environment, 43(27):4187–4193. doi:10.1016/j.atmosenv.2009.05.031.

Łozowicka-Stupnicka T. 2000. Ocena ryzyka i zagrożeń w złożonych systemach człowiek – obiekt techniczny – środowisko. Monografia 270. Politechnika Krakowska.

Łozowicka-Stupnicka T, Talarczyk M. 2005. Zastosowanie modeli sieci neuronowych w ocenie i prognozowaniu jakości powietrza. Inżynieria Środowiska.

Łupkowski Ł. 2010. Test Turinga. Perspektywa sędziego. Wydawnictwo Naukowe UAM. Poznań 2010. ISBN 978-83-232-2208-8.

Masooda A, Ahmad K. 2020. A model for particulate matter ($PM_{2.5}$) prediction for Delhi based on machine learning approaches. Procedia Computer Science. 167:2101–2110. doi:10.1016/j.procs.2020.03.258.

Maternowska M. 2019. Nowe technologie i ich wpływ na łańcuchy dostaw. Sztuczna inteligencja. Studia Ekonomiczne. Katowice 288:59–73. ISSN 2083-8611.

Matuszko D, Piotrowicz K. 2015. Cechy klimatu miasta a klimat Krakowa. Kraków: Instytut Geografii i Gospodarki Przestrzennej Uniwersytetu Jagiellońskiego. 221–241.

Meteoblue. 2023. Archiwalne dane pogodowe dla Krakowa i Wadowic za rok 2022 rok [online]. Dostęp 22.02.2023.
Dostępny w:
Kraków
https://www.meteoblue.com/pl/pogoda/historyclimate/weatherarchive/krak%c3%b3w_polska_3094802?fcstlength=1y&year=2022&month=2
Wadowice
https://www.meteoblue.com/pl/pogoda/historyclimate/weatherarchive/wadowice_polska_3082722

Namieśnik J, Jamrógiewicza Z. 1998. Fizykochemiczne metody kontroli zanieczyszczeń środowiska. Wydawnictwo Naukowo-Techniczne. Warszawa.

Nisbet C, LaGoy P. 1992. Toxic equivalency factors (TEFs) for polycyclic aromatic hydrocarbons (PAHs). Regul Toxicol Pharmacol. 16(3):290–300. doi:10.1016/0273-2300(92)90009-x.

Niżewski P, Boniecki P, Dach J. 2007. Ocena zastosowania prognostycznej sieci neuronowej w modelowaniu emisji gazowych. Journal of Research and Applications in Agricultural Engineering. 52(2):71–74.

Osowski S. 2000. Sieci neuronowe do przetwarzania informacji. Oficyna Wydawnictwa Politechniki Warszawskiej. Warszawa.

Ośródka K, Ośródka L, Wojtylak M. 1995. Zastosowanie metody sztucznych sieci neuronowych do prognozy zanieczyszczeń powietrza w aglomeracji miejsko-przemysłowej. Wiadomości Instytutu Meteorologii i Gospodarki Wodnej. 18(3-4):5–18.

Pimpunchat B, Sirimangkhala K, Junyapoon S. 2014. Modeling Haze Problems in the North of Thailand using Logistic Regression. Journal of Mathematical and Fundamental Sciences. 46(2):183–193. doi: 10.5614/j.math.fund.sci.2014.46.2.7.

PN-EN 12341:2006a. PN-EN 14907:2006b. 2006. Jakość powietrza – Oznaczanie frakcji PM 10 pyłu zawieszonego – Metoda odniesienia i procedura badania terenowego do wykazania równoważności stosowanej metody pomiarowej z metodą odniesienia. KT 280. Jakości Powietrza.

Pohl A. 2019. Wielopierścieniowe węglowodory aromatyczne – charakterystyka, występowanie i identyfikacja źródeł ich pochodzenia w środowisku. Laboratorium-Przegląd Ogólnopolski. 27(4):10–14.

Poloczek Ł, Wilkosz M, Czech P, Saternus M, Kania H. 2021. Wykorzystanie sztucznych sieci neuronowych typu mlp do predykcji zanieczyszczenia powietrza na podstawie danych pogodowych ze stacji pomiarowej. 14:222–240. doi:10.53052/9788366249837.20.

Rogula W, Żeliński J. 2004. Zastosowanie sztucznych sieci neuronowych do identyfikacji mechanizmów dominujących w rozprzestrzenianiu zanieczyszczeń powietrza. Ochrona Powietrza i Problemy Odpadów. 38(4)129–139. ISSN 1230-7408.

Rogula-Kozłowska W, Kozielska B, Klejnowski K. 2013. Concentration, Origin and Health Hazard from Fine particle-Bound PAH at Three Characteristic Sites in Southern Poland. Springer. 91(3):349-55. doi:10.1007/s00128-013-1060-1.

Rogula-Kozłowska W, Kozielska B, Majewski G, Rogula-Kopiec P, Mucha W, Kociszewska K. 2017. Submicreon particle-bound polycyclic aromatic hydrocarbons in the Polish teaching rooms: Concentrations, origin and health hazard. Jurnal of Environmental Sciences. 64:235-244. doi:10.1016/j.jes.2017.06.022.

Rolph G, Stein A, Stunder B. 2017. Real-time Environmental Applications and Display sYstem: READY. Environmental Modelling & Software. 95:210–228. doi:10.1016/j.envsoft.2017.06.025

Rosenblatt F. 1992. Principle of neurodynamics. Spartan, New York.

Rutkowski I. 2020. Inteligentne technologie w marketingu i sprzedaży - zastosowania, obszary i kierunki badań. Journal of Marketing and Market Studies. 6:3–12. Doi:10.33226/1231-7853.2020.6.1.

Sani SH, Shopon Md, Rakib SH. 2021. ir Quality Index Prediction Using Azure IoT & Machine Learning for Smart Cities. Lecture Notes on Data Engineering and Communications Technologies. 62. doi: 10.1007/978-981-33-4968-1_56.

Samek L, Styszko K, Stęgowski Z, Zimnoch M, Skiba A, Turek-Fijak A, Gorczyca Z, Furman P, Kasper-Giebl A, Różański K. 2021. Comparison of PM_{10} Sources at Traffic and Urban Background Sites Based on Elemental, Chemical and Isotopic Composition: Case Study from Krakow, Southern Poland. Atmosphere 2021, 12(10):1364. doi: 10.3390/atmos12101364.

Sánchez-Jiménez A, Heal MR, Beverland IJ. 2012. Correlations of particle number concentrations and metals with nitrogen oxides and other traffic-related air pollutants in Glasgow and Lon-don. Atmospheric Environment. 54:667–678. doi: 10.1016/j.atmosenv.2012.01.047.

Sekuła P, Bokwa A, Ustrnul Z, Zimnoch M, Bochenek B. 2021. The impact of a foehn wind on PM_{10} concentrations and the urban boundary layer in complex terrain: a case study from Kraków, Poland. Tellus B: Chemical and Physical Meteorology. 73(1):1933780. doi: 10.1080/16000889.2021.1933780.

Sejmik Województwa Małopolskiego. 2016. Uchwała Nr XVIII/243/16 w sprawie wprowadzenia na obszarze Gminy Miejskiej Kraków ograniczeń w zakresie eksploatacji instalacji, w których następuje spalanie paliw. http://edziennik.malopolska.uw.gov.pl/legalact/2016/812/.

Sejmik Województwa Małopolskiego. 2017a. Uchwała Nr XXXV/527/17 Sejmiku Województwa Małopolskiego z dnia 24 kwietnia 2017r w sprawie wprowadzenia na obszarze Gminy Miejskiej Kraków, w okresie od dnia 1 lipca 2017 roku do dnia 31 sierpnia 2019 roku, zakazów w zakresie eksploatacji instalacji.

Sejmik Województwa Małopolskiego. 2020. Załącznik nr 2 do Uchwały Nr XXV Sejmiku Województwa Małopolskiego z dnia 28 września 2020r. Program ochrony powietrza dla Województwa Małopolskiego. Małopolska w zdrowej atmosferze.

Shah SK. Tariq Z, Lee J, Lee Y. 2020. Real-Time Machine Learning for Air Quality and Environmental Noise Detection. 2020 IEEE International Conference on Big Data (Big Data). doi: 10.1109/bigdata50022.2020.9377939.

Skrzypski J, Jach-Szakiel E. 2008. Prognozowanie klas stanu jakości powietrza jako instrument zarządzania bezpieczeństwem ekologicznym w aglomeracjach miejsko-przemysłowych. Politechnika Łódzka.

Simoneit B. 2015. Triphenylbenzene in Urban Atmospheres, a New PAH Source Tracer. Polycyclic Aromatic Compounds. 35: 3–15. doi:10.1080/10406638.2014.883417.

Skiba A, Furman P, Durak J, Dobrowolska N, Guzik N, Zięba D, Kistler M, Kasper-Giebl A, Styszko K. 2018. Chemical composition of atmospheric aerosols collected in Krakow Agglomeration. Knowledge Technology and Society AGH ISC 2018. 43–51.

Skiba A, Styszko K, Furman P, Dobrowolska N, Kistler M, Kasper-Giebl A, Zięba D. 2019. Polycyclic aromatic hydrocarbons (PAHs) associated with PM_{10} collected in Wadowice, South Poland. E3S Web of Conferences, 108, p. 02007. doi:10.1051/e3sconf/201910802007.

Siudek P, Frankowski M. 2018. The role of sources and atmospheric conditions in the seasonal variability of particulate phase PAHs at the urban site in Central Poland. Aerosol Air Qual Res. 18:1405–1418. doi:10.4209/aaqr.2018.01.0037.

Sobczyk P, Tabor J, Pagacz P. 2019. Metody wykrywania obserwacji odstających i ich zastosowanie do detekcji nadużyć. Uniwersytet Jagielloński.

Stefaniuk E, Bosacka K, Hryniewicz W. 2015. Walidacja i weryfikacja metod i testów diagnostycznych w laboratorium mikrobiologicznym. Postępy Mikrobiologii. 54(4):415–424.

Stein AF, Draxler RR, Rolph GD, Stunder BJB, Cohen MD, Ngan F. 2015. NOAA's HYSPLIT atmospheric transport and dispersion modeling system, Bull. American Meteorological Society. 96:2059-2077. doi:10.1175/BAMS-D-14-00110.1.

Suleiman A, Tight MR, Quinn AD. 2016. Hybrid Neural Networks and Boosted Regression Tree Models for Predicting Roadside Particulate Matter. Environmental Modeling & Assessment. 21(6):731–750. doi:10.1007/s10666-016-9507-5.

Suleiman A, Tight MR, Quinn AD. 2019. Applying machine learning methods in managing urban concentrations of traffic-related particulate matter (PM_{10} and $PM_{2.5}$). Atmospheric Pollution Research. 10(1):134–144. doi:10.1016/j.apr.2018.07.001.

Styszko K, Szramowiat K, Kistler M, Kasper-Giebl A, Socha S, Rosenber EE, Gołaś J. 2016. Polycyclic aromatic hydrocarbons and their nitrated derivatives associated with PM_{10} from Kraków city during heating season. E3S Web of Conferences, 10:00091. doi:10.1051/e3sconf/20161000091.

Szlachtowska E, Kosiorowski D, Mielczarek D. 2016. Ocena jakości aplikacyjnej odpornego algorytmu analizy skupień TCLUST na przykładzie zbioru danych dotyczących jakości powietrza w Krakowie. Przegląd statystyczny. Kraków. 63(1):67–80.

Szramowiat K, Styszko K, Kistler M, Kasper-Giebl A, Gołas J. 2016. Carbonaceous species in atmospheric aerosols from the Krakow area (Malopolska District): Carbonaceous species dry deposition analysis. E3S Web of Conferences. 10:4–11. doi:10.1051/e3sconf/20161000092.

Szramowiat-Sala, Styszko K, Gołaś J. 2014. Determination of chemical composition of atmospheric aerosols conducted on the basis of particulate matter samples from Malopolska, South Poland. AGH Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie. Katedra Chemii Węgla i Nauk o Środowisku. Wydział Energetyki i Paliw. 100 lecie odnowienia tradycji Politechniki Warszawskiej. 82–89.

Tadeusiewicz R, Szaleniec M. 2015. Leksykon sieci neuronowych. Wydawnictwo Fundacji „Projekt Nauka”. Wrocław.

Tesauro G. 1994. TD-Gammon, a Self-Teaching Backgammon Program, Achieves Master-Level Play. Neural Computation; 6(2):215–219. doi: https://doi.org/10.1162/neco.1994.6.2.215

Tomski A. 2022. Statystyka w badaniach klinicznych cz.2 [online]. Statystyka.az.pl. Dostęp 15.01.2023. Dostępny w https://www.statystyka.az.pl/regresja-logistyczna.php

Tsai T, Lin Y, Hwang B, Nakayama S, Tsai C, Sun X, Ma C, Jung C. 2019. Fine particulate matter is a potential determinant of Alzheimer's disease: A systemic review and meta-analysis. Environmental Research 177:108638. doi:10.1016/j.envres.2019.108638.

U.S. EPA. 2003. List of PAHs recommended for analytical measurement to quantify. Total PAHs. United States Environmental Protection Agency.

Urząd Statystyczny w Krakowie. 2023. Dane o Krakowie [online]. Dostęp 22.02.2023. Dostępny w https://krakow.stat.gov.pl/

WeatherSpark. 2023. Całoroczny klimat i średnie warunki pogodowe dla Krakowa i Wadowic [online]. Dostęp 22.02.2023.
Dostępny w:
Kraków
https://pl.weatherspark.com/y/85104/%C5%9Arednie-warunki-pogodowe-w:-Krak%C3%B3w-Polska-w-ci%C4%85gu-roku
Wadowice
https://pl.weatherspark.com/y/84835/%C5%9Arednie-warunki-pogodowe-w:-Wadowice-Polska-w-ci%C4%85gu-roku

Weinmayr G, Pedersen M, Stafoggia M, Andersen Z, Galassi C, Munkenast J, Jaensch A, Oftedal B, Krog N, Aamodt G, Pyko A, Pershagen G, Korek M, Faire U, Pedersen N, Östenson C, Rizzuto D, Sørensen M, Tjønneland A, Bueno-de-Mesquita B, Vermeulen R, Eeftens M, Concin H, Lang A, Wang M, Tsai M, Ricceri F, Sacerdote C, Ranzi A, Cesaroni G, Forastiere F, Hoogh K, Beelen R, Vineis P, Kooter I, Sokhi R, Brunekreef B, Hoek G, Raaschou-Nielsen O, Nagel G. 2018. Particulate matter air pollution components and incidence of cancers of the stomach and the upper aerodigestive tract in the European Study of Cohorts of Air Pollution Effects (ESCAPE). Environment International. 120:163–171. doi:10.1016/j.envint.2018.07.030.

Hou W, Li Z, Zhang Y, Xu H, Zhang Y, Li K, Li D, Wei P, Ma Y. 2014. Using support vector regression to predict PM_{10} and $PM_{2.5}$. IOP Conf. Series: Earth and Environmental Science. 17: 012268. doi:10.1088/1755-1315/17/1/012268.

Widrow B, Hoff M. 1960. Adaptive switching circuits. IRE WESCON Convention Record. 4:96–194.

Wilkosz M, Poloczek Ł, Czech P, Saternus M, Kania H. 2021. Prognozowaniu zanieczyszczenia powietrza atmosferycznego przy użyciu szeregów czasowych i różnych typów sztucznych sieci neuronowych. 11:295–308. doi:10.53052/9788366249837.28.

Willett K, Gardinali P, Sericano J, Wade T, Safe S. 1997. Characterization of the H4IIE rat hepatoma cell bioassay for evaluation of environmental samples containing polynuclear aromatic hydrocarbons (PAHs). Arch Environ Contam Toxicol. 32(4):442–8. doi: 10.1007/s002449900211.

Wojewódzki Inspektorat Ochrony Środowiska w Krakowie. 2014. Pięcioletnia ocena jakości powietrza pod kątem jego zanieczyszczenia: SO2 , NO2 , NOx , CO, benzenem, O3, pyłem PM_{10}, pyłem $PM_{2,5}$ oraz As, Cd, Ni, Pb i B(a)P w Województwie Małopolskim uwzględniająca wymogi dyrektyw: 2008/50/WE i 2004/107/WE oraz decyzji 2011/850/UE.

World Health Organization. 2016. World health statistics - monitoring health for the SDGs. p. 1.121. doi:10.1017/CBO9781107415324.004.

World Health Organization. 2021. WHO global air quality guidelines. Particulate matter ($PM_{2.5}$ and PM_{10}), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. Geneva.

Yunker B, Macdonald R, Vingarzan R, Mitchell H, Goyette D, Sylvestre S. 2002. PAHs in the Fraser River basin: a critical appraisal of PAH ratios as indicators of PAH source and composition. Organic Geochemistry. 33:489–515. doi:10.1016/S0146-6380(02)00002-5.

Zaciera M. 2007. Badania nad oznaczaniem nitrowych pochodnych wielopierścieniowych węglowodorów aromatycznych w powietrzu. Praca Doktorska. Katowice. Uniwersytet Śląski.

Zespół Parków Krajobrazowych Województwa Małopolskiego. 2023. Instytucja Województwa Małopolskiego.

Zylinski M. 2018. Uczenie maszynowe od podstaw z Azure ML Studio (classic). Platforma Udemy.

Spis rysunków

Spis równań

Spis tabel