

# PROTOTYP MODELU SYSTEMU SAMOUCZĄCEGO DO PROGNOZOWANIA STĘŻENIA PYŁU PM10 W POWIETRZU ATMOSFERYCZNYM

Cezary ORŁOWSKI, Magdalena SARZYŃSKA, Arkadiusz SARZYŃSKI

**Streszczenie:** Artykuł przedstawia model prognozowania stężenia pyłu PM10 z wykorzystaniem algorytmów genetycznych. Model ten jest drugim z kolei do prognozowania pyłu (poprzedni wykorzystywał jednokierunkowe sieci neuronowe) i stanowi podstawę do budowy modelu samouczącego. Podczas budowy modelu uwzględniono oddziaływanie czynników meteorologicznych, a do implementacji wykorzystano algorytm genetyczny ze względu na specyfikę problemu prognozowania stężenia pyłu PM10. Następnie określono optymalną konfigurację algorytmu definiując strukturę osobnika oraz całej populacji, funkcję przystosowania, a także metody selekcji, krzyżowania oraz mutacji. W ostatnim kroku określono prawdopodobieństwa krzyżowania i mutacji oraz czas życia osobników (ilość generacji). Po zakończeniu procesu budowy, algorytm zastosowano do prognozowania stężenia pyłu w powietrzu atmosferycznym na obszarze aglomeracji Gdańskiej.

**Słowa kluczowe:** model samouczący, prognozowanie PM10, algorytmy genetyczne

## 1. Wprowadzenie

Projektowanie systemów dla potrzeb inteligentnych miast wymaga stosowania koncepcji modeli samouczących. Tylko takie modele i systemy będą w stanie przetwarzać nieograniczone zasoby danych, a jednocześnie stwarzać warunki do ich strojenia [1,2]. W pracy proponuje się podejście przyrostowe do projektowania takich modeli w ramach którego dziedziną jaką jest prognozowanie stężenia pyłu jest analizowana z punktu widzenia dwu niezależnych mechanizmów przetwarzania: sieci neuronowych i algorytmów genetycznych. W budowie tych modeli badane są mechanizmy uczenia i strojenia, aby znaleźć założenia do budowy modeli samouczących.

W pracy opisano proces budowy mechanizmu prognostycznego bazującego na algorytmach genetycznych. Przedmiotem zainteresowań jest pył zawieszony, który jest mieszaniną organicznych i nieorganicznych cząstek stałych, zawarty np. w spalinach samochodów. W skład pyłu wchodzi cząsteczki różnej wielkości. Wyróżnia się dwie główne grupy: cząstki o średnicy do 10 mikrometrów (PM10) oraz 2,5 mikrometra (PM2,5). Częściej stosowaną miarą poziomu zanieczyszczenia powietrza jest PM10 i dlatego stanie się przedmiotem badań. Nieustannie należy kontrolować stężenie pyłu zwieszonoego lub nawet starać się prognozować jego przyszłą wartość. Przewidywania takie mogą być podstawą do podejmowania działań operacyjnych, celem których jest obniżenie wartości zanieczyszczenia powietrza w przypadku jej prognozowanego przekroczenia. W artykule podjęta zostanie próba zbudowania modelu do przewidywania godzinowego poziomu zanieczyszczenia powietrza pyłem PM10 (24 godziny w przyszłość). Kolejnym krokiem będzie implementacja opracowanego modelu z wykorzystaniem algorytmów genetycznych. Przy użyciu wybranej technologii powinniśmy otrzymać przewidywane stężenie PM10 w

powietrzu z 24 godzinnym wyprzedzeniem, z możliwością do zaakceptowania granicą błędów.

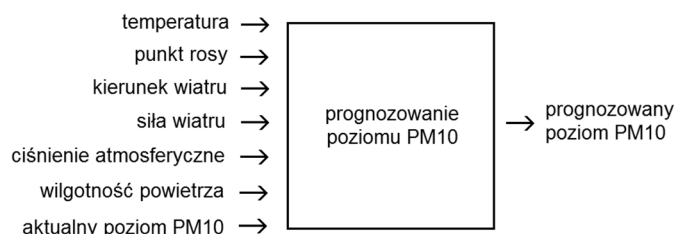
## 2. Budowa modelu do prognozowania pyłu

Pierwszym etapem prac nad prognozowaniem poziomu zanieczyszczenia powietrza przez pył zawieszony PM10 było określenie parametrów wpływających na jego stężenie (parametrów wejściowych modelu). Stwierdzono, że na przyszłą wartość pyłu wpływ mają wybrane warunki meteorologiczne oraz aktualny poziom PM10. Podstawą do takiego wniosku było przeprowadzenie analizy statystycznej współczynnika korelacji, pomiędzy kolejnymi parametrami wejściowymi oraz przewidywanym stężeniem PM10. Poniżej przedstawiono tabelę 1, zawierającą opis wyników ww. badania.

Tab. 1. Wpływ czynników meteorologicznych na prognozę stężenie PM10[3]

Czynnik	Wpływ na przyszłą wartość PM10
Temperatura	Wzrost temperatury powoduje spadek poziomu pyłu
Punkt rosy	Poziom pyłu maleje wraz ze wzrostem wartości punktu rosy
Kierunek wiatru	Istnieje wpływ kierunku wiatru na poziom pyłu. Wiatr wiejący z kierunku 315-45 stopni powoduje mniejsze wskazania poziomu pyłu, natomiast wiejący z kierunków 90-135 oraz 270-315 - większe.
Siła wiatru	Przy słabym wietrze poziom pyłu jest większy, przy silnym wietrze – mniejszy.
Ciśnienie atmosferyczne	Wzrost ciśnienia atmosferycznego powoduje wzrost poziomu PM10.
Wilgotność powietrza	Zanieczyszczenie powietrza pyłem rośnie wraz ze wzrostem wilgotności powietrza.

Na podstawie wyników badań, stwierdzono, iż w prognozowaniu należy uwzględnić podane w tabeli 1 czynniki meteorologiczne, uzupełniając je o aktualne stężenie PM10 zanieczyszczenia powietrza. Na wyjściu modelu, w oparciu o dane wejściowe, powinno się otrzymać przewidywane stężenie PM10 w powietrzu po upływie 24 godzin. Uproszczony model przedstawiono na rysunku 1.



Rys. 1. Model systemu do prognozowania pyłu PM10

Na rysunku 1 przedstawiono strukturę modelu do prognozowania stężenia pyłu PM10 z 24 godzinnym wyprzedzeniem. Zobrazowano 7 parametrów wejściowych oraz 1 parametr wyjściowy modelu.

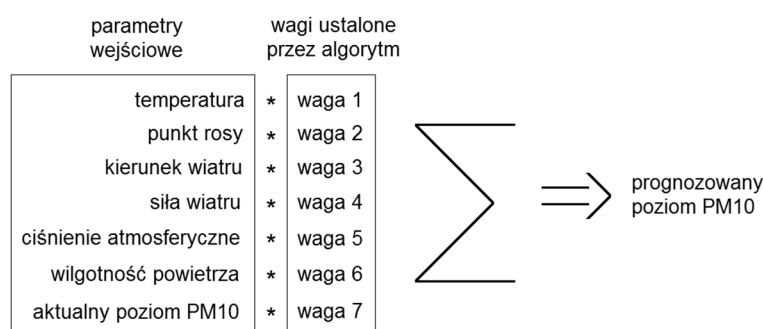
## 3. Implementacja modelu – algorytm genetyczny

Algorytmy genetyczne są to algorytmy, których twórcy brali inspirację ze środowiska naturalnego. Algorytmy te opierają swoje działanie na podstawowych zasadach ewolucji, takich jak selekcja, krzyżowanie, czy mutacja osobników. Zasady te zostaną szczegółowo

opisane w dalszej części artykułu, w odniesieniu do problemu prognozowania PM10.

Algorytmy genetyczne służą do znajdowania rozwiązań dla problemów, których nie można rozwiązać w wielomianowym czasie, a więc głównie takich, dla których nie ma precyzyjnego, dedykowanego algorytmu. Przykładem takich problemów mogą być wszelkiego rodzaju problemy optymalizacyjne, symulacje czy prognozy, bazujące na wielu niezależnych parametrach wejściowych. Korzystając z tej własności, podjęta została próba wykorzystania algorytmu genetycznego do przewidywania przyszłego stężenia PM10.

Idea proponowanego rozwiązania jest taka, iż algorytm genetyczny zostanie wykorzystany do znalezienia wag, przez które ostatecznie będą przemnożone parametry wejściowe. Wymnożone składniki zostaną zsumowane i wygenerują przewidywany poziom zawartości PM10 w powietrzu atmosferycznym (rysunek 2).



Rys. 2. Schemat prognozowania poziomu PM10

Na rysunku 2 przedstawiono zasadę działania algorytmu do prognozowania stężenia pyłu PM10. Parametry zgrupowane pierwszą ramą zostaną kolejno przemnożone przez kolejne wagi z drugiej ramy. W następnym kroku wyniki mnożenia zostaną zsumowane generując przewidywane stężenie pyłu PM10. Wagi z drugiej ramy będą wygenerowane przez algorytm genetyczny.

#### 4. Konfiguracja algorytmu genetycznego dla prognozowania poziomu zawartości pyłu PM10

##### 4.1. Populacja

Istotnym elementem, który należy określić dla algorytmu genetycznego jest populacja osobników, czyli opis, ilość oraz stan początkowy osobników w populacji. Dla opisywanego problemu prognozowania, każdy osobnik powinien posiadać 7 chromosomów, odpowiadających wagom, przez które zostaną przemnożone kolejne parametry wejściowe.

Wielkość populacji powinna być na tyle duża, żeby wszystkie procesy takie jak np. krzyżowanie przebiegały bez zakłóceń, czyli żeby różnorodność osobników w populacji, podlegających selekcji, była na wysokim poziomie. Należy również pamiętać o tym, aby nie była zbyt duża, ponieważ może to spowodować wolną pracę algorytmu. Metodą eksperymentalną określono początkową wielkość populacji na 200 oraz maksymalną ilość tworzonych w nowej generacji osobników na 50, które były wystarczające, aby zwracać odpowiednio zróżnicowane wyniki, ale nie obciążać zbyt mocno algorytmu.

Uruchamiając algorytm, pierwszym krokiem jest wygenerowanie populacji początkowej,

która składa się z osobników o losowo dobranych cechach. Ma to zapewnić różnorodność populacji przy tworzeniu kolejnych pokoleń. Wszystkie cechy osobników zostały dobrane losowo, jako liczby rzeczywiste z przedziału od -1 do 1.

#### 4.2. Funkcja przystosowania

Ważnym elementem, bez którego algorytm genetyczny nie będzie działał poprawnie jest definicja funkcji przystosowania. Funkcja ta jest wykorzystywana do określenia, jakie szanse na przeżycie ma dany osobnik. Dla każdego problemu, funkcja przystosowania jest definiowana indywidualnie. Dla problemu prognozowania funkcja ta zwróci, jaki jest błąd prognozowania generowany przez przypisane do danego osobnika wagi. Wagi zostaną przemnożone przez wartości pobrane ze zbioru testującego.

Zbiór testujący składa się z danych pomiarowych obejmujących pełne 2 lata kalendarzowe. Zawiera wartości odpowiadające wszystkim parametrom wejściowym modelu oraz zaobserwowane stężenie pyłu PM10 po 24 godzinach.

Dane testujące składają się z 8 kolumn. Kolejne kolumny odpowiadają odpowiednio: aktualnemu stężeniu pyłu PM10, temperaturze, punkowi rosy, wilgotności powietrza, ciśnieniu atmosferycznemu, kierunkowi wiatru, sile wiatru oraz wartości pyłu PM10 po 24 godzinach. Tak przygotowane dane mogą w łatwy sposób posłużyć do sprawdzenia przystosowania osobników w populacji.

Źródłem informacji o zawartości pyłu PM10 była baza danych otrzymana przez fundację ARMAAG, natomiast dane meteorologiczne zostały pobrane z otwartego źródła w sieci internet. Stacja pomiarowa PM10 jest umiejscowiona przy ulicy Ostrzyckiej w Gdańsku (dzielnica Szadółki). Punkt pomiarów meteorologicznych znajduje się na lotnisku w gdańskim Rębichowie. Wybór wyżej wymienionych stacji nie był przypadkowy. Starano się dobrać stacje tak, aby odległość między nimi była jak najmniejsza. Stosując ten zabieg można było uzyskać podobne warunki na obu stacjach.

#### 4.3. Proces selekcji

Kolejnym istotnym elementem algorytmu genetycznego jest selekcja osobników. Dzięki niej możemy wybrać osobniki kwalifikujące się do budowy kolejnej generacji. Możemy wyróżnić dwie strategie oceny populacji: Pittsburgh oraz Michigan [4]. Strategia Pittsburgh polega na tym, iż ocenie podlega cała populacja, a osobniki współpracują ze sobą, aby jak najbardziej zbliżyć się do osiągnięcia celu. Druga strategia – Michigan – zakłada zupełnie inne podejście. Osobniki w populacji rywalizują ze sobą o przetrwanie i są oceniane indywidualnie. Podczas projektowania algorytmu do prognozowania pyłu wykorzystamy podejście Michigan, a więc będziemy starali się znaleźć takiego osobnika, którego wagi zwracają najmniejszy błąd predykcji.

Istotnym przy tworzeniu kolejnych generacji, jest nie dopuścić do wpadnięcia w minimum lokalne, czyli sytuacji, w której algorytm stwierdzi, że znalazł najlepsze rozwiązanie, jednak będzie istniało inne, lepsze. Dzieje się tak, gdy funkcja przystosowania dla osobników w populacji zaczyna zwracać zbieżne wyniki. Oznacza to, iż należy jakoś zabezpieczyć algorytm przed tym zjawiskiem. Istnieje kilka metod selekcji, wykorzystywanych w algorytmach genetycznych.

Pierwszą sprawdzoną metodą selekcji, była metoda ruletki. W metodzie tej, prawdopodobieństwo wylosowania osobników do reprodukcji w kolejnym pokoleniu, równe jest ich przystosowaniu. Osobniki z lepszym przystosowaniem mają większą szansę na

przekazanie genów kolejnym pokoleniom [5]. Prawdopodobieństwo wylosowania przedstawia wzór 1 [6].

$$p_i = \frac{f_i}{F} \quad (1)$$

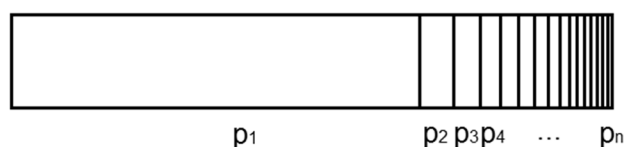
gdzie:  $p_i$  – prawdopodobieństwo wylosowania osobnika  $i$   
 $f_i$  – ocena przystosowania osobnika  $i$   
 $F$  – suma przystosowań wszystkich osobników

Należy pamiętać, iż metoda ruletki zadziała poprawnie, tylko, gdy miary oceny dostosowania osobników będą większe lub równe zero. Z każdą metodą, bazującą na prawdopodobieństwie, wiąże się pewien błąd stochastyczny. Dla metody ruletki wariancja tego błędu wynosi (wzór 2)[2]:

$$\sigma_i^2 = n * p_i * (1 - p_i) \quad (2)$$

gdzie:  $n$  – rozmiar populacji

Modyfikacją metody ruletki jest metoda rangowa. Jest ona bardzo podobna do metody ruletki - osobniki w populacji są posortowane wg. oceny przystosowania, ale miarą dostosowania osobników są rangi. Prawdopodobieństwo wyboru osobników do reprodukcji odbywa się na podstawie nadanych rang. Skutkiem operowania na rangach, a nie na prawdopodobieństwach jest wyeliminowanie przypadku, gdy jeden osobnik będzie miał bardzo dobre przystosowanie, natomiast pozostałe osobniki słabe. Nie wystąpi wtedy sytuacja, że „dobry” osobnik będzie wybierany prawie zawsze, a pozostałe prawie nigdy (rysunek 3), a więc przypadek wpadnięcia w minimum lokalne.



Rys. 3. Przypadek wpadnięcia w minimum lokalne

Rysunek 3 obrazuje problem wpadnięcia w minimum lokalne, który może wystąpić w metodzie ruletki. Całkowite prawdopodobieństwo zostało podzielone na prawdopodobieństwa wylosowania kolejnych osobników ( $p_1, p_2, \dots, p_n$ ). Do opisywanej sytuacji może dojść, gdy któryś z osobników, po losowym dobraniu wag, uzyska wysoki wynik przystosowania, a tym samym wysokie prawdopodobieństwa wylosowania do generowania nowych osobników ( $p_1$ ). Suma prawdopodobieństw wylosowania pozostałych osobników ( $p_2, p_3, p_4, \dots, p_n$ ) będzie mniejsza od prawdopodobieństwa osobnika najlepiej przystosowanego ( $p_1$ ) i będzie on prawie zawsze wybierany do reprodukcji, a więc algorytm zacznie generować zbieżne wyniki. Rozwiązaniem tego problemu jest opisana wcześniej metoda rangowa.

Kolejną sprawdzoną metodą selekcji jest metoda turniejowa. W metodzie tej, sortujemy populację w sposób losowy, aby zapewnić jednakowe szanse na przeżycie każdemu

osobnikowi. Następnie dzielimy osobniki na małe grupy i w każdej z nich wybieramy po dwa najlepiej przystosowane [7]. Wyselekcjonowane osobniki będą uczestniczyły w procesie reprodukcji. Metoda ta posiada kilka zalet. Po pierwsze, w kolejnych iteracjach przystosowanie populacji nie pogarsza się, a w najgorszym przypadku może pozostać na tym samym poziomie. Drugą zaletą jest różnorodność generowanej populacji i eliminacja problemu wczesnej zbieżności. Każdy osobnik w danej iteracji, może wygenerować tylko jednego osobnika generacji następnej. Wadą tej metody jest to, iż dążenie do wygenerowania rozwiązania optymalnego następuje dość powoli.

Porównanie metod rangowej oraz turniejowej dla prognozowania poziomu pyłu PM10 przedstawiono w poniższej tabeli 2.

Tab. 2. Porównanie metod rangowej oraz turniejowej do prognozowania poziomu PM10

	Metoda rangowa	Metoda turniejowa
Błąd minimalny	5,7455	5,6912
Błąd średni	6,1740	6,6086
Błąd < 5µg/m <sup>3</sup> (% osobników)	50,1194	48,6815
Błąd < 10µg/m <sup>3</sup> (% osobników)	82,8155	79,3769
Błąd < 15µg/m <sup>3</sup> (% osobników)	93,8919	91,8673

Obie metody (rangowa oraz turniejowa) zostały przebadane przy jednakowej konfiguracji algorytmu, tak, aby zachować obiektywne środowisko porównawcze. Na podstawie wyników badania możemy stwierdzić, iż metoda rangowa lepiej nadaje się jako metoda selekcji algorytmu genetycznego dla problemu prognozowania, niż metoda turniejowa. Pomimo tego, iż przy zastosowaniu metody turniejowej minimalny błąd jest mniejszy niż przy drugiej metodzie, to średni błąd uzyskany dla metody rangowej jest już znacznie mniejszy. Z tego też powodu metodą selekcji, która zostanie zastosowana w algorytmie genetycznym do prognozowania zawartości pyłu PM10 w powietrzu, będzie metoda rangowa.

#### 4.4. Proces krzyżowania

Kolejnym krokiem, po wybraniu najlepiej rokujących osobników, który należy przeprowadzić jest proces krzyżowania, w wyniku którego powstanie nowy osobnik. Istnieje wiele uniwersalnych metod krzyżowania, jednak są to tylko pewnego rodzaju ramy. Można z nich korzystać, jednak krzyżowanie należy zaprojektować indywidualnie dla danego problemu. Związane jest to z oryginalną budową osobników występujących w populacji. Przyjęliśmy wcześniej, iż każdy osobnik, będzie opisany 7 liczbami rzeczywistymi. Założono, iż można podjąć próbę skrzyżowania osobników na dwa sposoby.

Pierwszy z nich polega na losowym wyborze miejsca przecięcia zbioru chromosomów opisujących osobnika rozdzielając go na dwa podzbiory, pobrania wag z pierwszego podzbioru jednego osobnika, a następnie dopełnienia wagami z drugiego podzbioru, drugiego osobnika. W ten sposób powstanie potomek, składający się z części cech pochodzących od obu osobników.

Drugi sposób krzyżowania polega na budowie potomka, którego wagi będą średnimi wag rodziców. Porównanie obu metod krzyżowania przedstawiono w tabeli 3.

Tab. 3. Porównanie metod krzyżowania: podzbiorów oraz średnich.

	Metoda podzbiorów	Metoda średnich
Błąd minimalny	5,7552	5,6727
Błąd średni	7,4107	6,0467
Błąd < 5µg/m <sup>3</sup> (% osobników)	53,5537	53,9818
Błąd < 10µg/m <sup>3</sup> (% osobników)	85,8245	87,1311
Błąd < 15µg/m <sup>3</sup> (% osobników)	94,9538	94,8018

Badania metod krzyżowania przeprowadzone zostały przy takich samych parametrach konfiguracyjnych algorytmu. Na podstawie tabeli powyżej, możemy wywnioskować, iż dla problemu prognozowania stężenia pyłu PM10, metoda średnich generuje lepsze wyniki. Średni błąd prognozowania przy wykorzystaniu metody średnich wyniósł 6,0467 µg/m<sup>3</sup>, natomiast dla metody podzbiorów 7,4107µg/m<sup>3</sup>. Błąd minimalny jest również mniejszy dla metody średnich, a odsetek osobników z błędem poniżej 10µg/m<sup>3</sup> większy. W związku z powyższym metoda średnich została wybrana, jako metoda krzyżowania w projektowanym algorytmie.

#### 4.5. Proces mutacji

Zastosowanie w algorytmie tylko procesu krzyżowania, może powodować znajdowanie nieoptymalnego rozwiązania, poprzez wpadanie w minima lokalne. Należy wprowadzić pewien mechanizm losowości podczas generowania nowych osobników, który spowoduje wyskoczenie z takiego minimum. Do tego służy mutacja osobników. Podobnie jak w środowisku naturalnym, mutacja może przynieść korzyści, tworząc osobnika z cechami lepszymi niż rodzic, lub stratę, generując osobnika słabiej przystosowanego. Proces mutacji, polega na zamianie określonej liczby chromosomów nowymi, losowymi chromosomami[8].

Metodę tę wprowadzono w algorytmie, oraz zbadano liczbę wag(chromosomów), które należy zamienić, aby otrzymać najlepsze rozwiązanie. Poniżej przedstawiono tabelę 4, przedstawiającą wyniki przeprowadzonego badania.

Tab. 4. Wyniki badania ilości nowych chromosomów podczas procesu mutacji

	Ilość nowych chromosomów						
	1	2	3	4	5	6	7
Błąd minimalny	5,6882	5,6374	5,5254	5,5724	5,5497	5,5305	5,6151
Błąd średni	6,5758	6,2532	5,8664	5,9886	5,9455	5,8708	5,9759
Błąd < 5µg/m <sup>3</sup> (% osobników)	51,2109	55,5474	57,4879	53,8691	55,2095	57,1976	55,8009
Błąd < 10µg/m <sup>3</sup> (% osobników)	86,8946	86,4384	86,5510	87,4521	87,1086	86,4947	87,1874
Błąd < 15µg/m <sup>3</sup> (% osobników)	96,2661	94,4301	95,1284	95,6522	95,2861	94,5427	94,3625

Badanie zostało przeprowadzone dla jednakowej konfiguracji algorytmu genetycznego. Na podstawie powyższej tabeli, możemy wywnioskować, iż algorytm prognozuje najlepiej (z najmniejszym błędem), podczas mutacji z podmianą 3 chromosomów. Błąd minimalny oraz średni jest najmniejszy, a odsetek osobników w przedziałach do 5 i 10µg/m<sup>3</sup> jest największy. W związku z powyższym, mutacja z wprowadzeniem 3 losowych chromosomów zostanie zastosowana w tworzonym algorytmie.

#### 4.6. Podsumowanie konfiguracji

Na tym etapie tworzenia algorytmu określono wszystkie podstawowe elementy, wchodzące w jego skład. Ustalono budowę osobnika, wielkość populacji, funkcję przystosowania, metody selekcji, krzyżowania oraz mutacji. Szczegóły konfiguracji przedstawiono poniżej w tabeli 5.

Tab. 5. Szczegółowa konfiguracja algorytmu genetycznego

Element konfiguracji	Konfiguracja
Budowa osobnika	Składa się z 7 wag, będących mnożnikami parametrów wejściowych
Wielkość populacji początkowej	200 osobników
Maksymalna ilość osobników generowanych	50 osobników
Funkcja przystosowania	Korzystając ze zbioru testującego, określa na podstawie wag danego osobnika, jaki jest błąd przez niego generowany
Metoda selekcji	Metoda rangowa
Metoda krzyżowania	Wagi potomka jako średnie z wag rodziców
Metoda mutacji	Modyfikacja 3 losowych chromosomów rodzica

Schemat krokowy działania algorytmu do prognozowania poziomu zawartości pyłu PM10 z wykorzystaniem algorytmu genetycznego:

- Krok 1. Generowanie populacji początkowej
- Krok 2. Wyzerowanie licznika generowanych osobników
- Krok 3. Obliczenie przystosowania osobników
- Krok 4. Generowanie z określonym prawdopodobieństwem nowego osobnika poprzez krzyżowanie 2 innych (wyselekcjonowanych)
- Krok 5. Generowanie z określonym prawdopodobieństwem nowego osobnika poprzez mutację innego (wyselekcjonowanego)  
*[Jeżeli wygenerowano mniej niż 50 osobników - idź do 'Krok 3']*
- Krok 6. Usunięcie osobników nie nadających się do przeżycia  
*[Jeżeli stworzono mniej niż 500 generacji - idź do 'Krok 2']*
- Krok 7. Zwrócenie wyniku działania algorytmu

#### 4.7. Proces ewolucji

Podstawą działania algorytmów genetycznych jest proces ewolucji. Wraz z kolejnymi pokoleniami powstają osobniki lepiej przystosowane do przeżycia, czyli w naszym przypadku lepiej odzwierciedlające poziom zawartości pyłu PM10. Podstawowa wersja algorytmu genetycznego, podczas tworzenia nowej generacji osobników, zastępuje starą generację. Algorytm do prognozowania poziomu pyłu PM10 został rozszerzony o możliwość życia osobników przez kilka generacji. W wyniku czego, zaistniała możliwość korzystania z cech dobrze przystosowanych osobników przez kilka pokoleń. Należało przeprowadzić badania jaki jest efektywny (zwracający najmniejszy błąd prognozowania) czas życia osobników (ilość generacji). Poniżej zamieszczono tabelę 6 przedstawiającą efekty badań nad efektywnym czasem życia osobników.



Tab. 6. Wyniki badania czasu życia osobników

Czas życia	Błąd minimalny	Błąd średni	Błąd < 5µg/m <sup>3</sup> (% osobników)	Błąd < 10µg/m <sup>3</sup> (% osobników)	Błąd < 15µg/m <sup>3</sup> (% osobników)
0	>1000	>1000	0,0056	0,0056	0,0056
1	5,5157	6,3344	42,4814	73,3724	87,745
2	5,487	6,0855	50,3267	81,8597	92,3688
3	5,5109	5,8592	47,6008	83,0536	93,6585
4	5,4971	5,8358	50,2703	83,7914	92,8137
5	5,478	5,7983	50,1408	84,2363	93,6528
6	5,4885	5,7459	50,6251	84,9121	93,6641
7	5,5013	5,7254	51,7177	85,2895	93,4726
8	5,4963	5,7667	51,1771	84,9403	93,7993

Na podstawie wyników z tabeli 6 możemy stwierdzić, iż dodanie parametru czasu życia osobników, istotnie wpłynęło na działanie algorytmu. W początkowym etapie badania sprawdzono czas życia w zakresie od 0 do 8 generacji, ponieważ większy wydaje się być nienaturalny (gdyby stwierdzono, iż górna granica jest zbyt niska, zakres zostałby rozszerzony). Można zauważyć, iż wyniki dla różnych czasów życia nieznacznie się różnią, jednak najlepsze wyniki uzyskano dla czasu życia równego 7 generacji. Błąd średni wyniósł 5,7254 µg/m<sup>3</sup>, a odsetek osobników z błędami poniżej 5 i 10µg/m<sup>3</sup> jest największy wśród innych badanych przypadków. W związku z powyższym, czas życia równy 7 generacji zostanie zastosowany podczas realizacji algorytmu.

#### 4.8. Prawdopodobieństwo krzyżowania i mutacji

Ostatnią czynnością, którą należało przeprowadzić, było ustalenie, z jakim prawdopodobieństwem w każdej iteracji, przeprowadzać procesy krzyżowania oraz mutacji. W tym celu przeprowadzono badanie dla wcześniej ustalonej konfiguracji algorytmu, w którym sprawdzono permutacje dwóch zbiorów ( $M = \{10\%, 20\%, 30\%, 40\%, 50\%, 60\%, 70\%, 80\%, 90\%, 100\%\}$ ,  $K = \{10\%, 20\%, 30\%, 40\%, 50\%, 60\%, 70\%, 80\%, 90\%, 100\%\}$ ) odpowiadających prawdopodobieństwom mutacji oraz krzyżowania osobników. Poniżej zamieszczono tabelę 7, przedstawiającą błędy prognozowana pyłu PM10 algorytmu genetycznego, dla wszystkich możliwości prawdopodobieństwa mutacji oraz krzyżowania (Błąd – procentowy udział w całości).

Tab. 7. Badanie prawdopodobieństwa mutacji i krzyżowania osobników

Mut/ Krzyż	Min [µg/ m <sup>3</sup> ]	Błąd <5 µg/m <sup>3</sup>	Błąd <10 µg/m <sup>3</sup>	Błąd <15 µg/m <sup>3</sup>	Śred [µg/ m <sup>3</sup> ]	Mut/ Krzyż	Min [µg/ m <sup>3</sup> ]	Błąd <5 µg/m <sup>3</sup>	Błąd <10 µg/m <sup>3</sup>	Błąd <15 µg/m <sup>3</sup>	Śred [µg/ m <sup>3</sup> ]
10%/10%	5,577	57,288	85,819	94,188	6,515	60%/10%	5,524	57,474	86,450	94,577	6,429
10%/20%	5,675	54,584	87,041	94,351	6,363	60%/20%	5,631	54,613	86,005	95,410	6,012
10%/30%	5,719	55,311	85,864	93,895	6,582	60%/30%	5,657	58,020	85,290	93,473	5,946
10%/40%	5,690	51,718	87,334	95,686	6,792	60%/40%	5,622	55,035	87,199	94,599	5,873
10%/50%	5,617	55,018	87,154	94,976	6,483	60%/50%	5,632	53,086	87,199	95,579	5,914
10%/60%	5,548	56,629	87,221	94,424	6,823	60%/60%	5,606	57,057	85,701	94,149	5,873
10%/70%	5,602	56,629	86,343	94,334	6,444	60%/70%	5,516	55,615	87,430	95,309	5,813
10%/80%	5,734	53,283	87,255	94,379	6,682	60%/80%	5,536	55,074	87,908	95,140	5,819
10%/90%	5,740	53,323	87,086	94,554	6,639	60%/90%	5,532	57,417	86,467	94,402	5,828
10%/100%	5,629	54,241	87,300	94,740	6,435	60%/100%	5,501	56,955	87,447	94,796	5,859
20%/10%	5,655	54,128	87,621	94,526	6,229	70%/10%	5,529	54,314	87,728	95,709	6,345

20%/20%	5,652	55,379	86,236	94,475	6,203	70%/20%	5,725	53,790	85,982	94,954	6,171
20%/30%	5,562	54,815	87,582	95,314	6,048	70%/30%	5,659	54,584	87,013	94,678	5,960
20%/40%	5,545	54,905	87,756	95,044	6,066	70%/40%	5,594	53,075	87,323	95,934	5,929
20%/50%	5,589	55,350	87,024	94,903	6,051	70%/50%	5,543	56,161	87,340	94,588	5,871
20%/60%	5,655	53,357	87,801	94,993	6,421	70%/60%	5,648	53,424	87,362	95,264	5,887
20%/70%	5,559	56,043	87,424	94,582	6,080	70%/70%	5,573	56,297	86,720	94,548	5,816
20%/80%	5,606	55,576	86,979	94,334	6,355	70%/80%	5,518	55,671	87,334	95,303	5,801
20%/90%	5,582	54,562	87,672	95,095	6,328	70%/90%	5,567	55,570	87,103	94,655	5,748
20%/100%	5,617	55,356	87,165	94,458	6,057	70%/100%	5,552	53,903	87,846	95,714	5,847
30%/10%	5,695	55,790	86,151	93,918	6,323	80%/10%	5,727	54,365	86,444	94,323	6,292
30%/20%	5,640	53,588	87,233	95,264	6,221	80%/20%	5,606	54,185	87,176	95,292	6,137
30%/30%	5,581	56,848	86,681	94,239	6,086	80%/30%	5,625	52,427	87,525	95,883	5,994
30%/40%	5,610	53,441	87,486	95,331	6,114	80%/40%	5,600	55,091	86,523	95,117	5,853
30%/50%	5,570	53,824	87,458	95,725	5,971	80%/50%	5,608	55,283	87,373	94,638	5,816
30%/60%	5,546	56,634	87,233	94,543	6,050	80%/60%	5,580	54,004	87,492	95,489	5,724
30%/70%	5,527	54,928	87,751	95,247	5,863	80%/70%	5,573	56,133	86,681	94,548	5,792
30%/80%	5,533	56,190	87,447	94,627	5,981	80%/80%	5,585	54,573	87,030	95,309	5,790
30%/90%	5,614	53,666	87,672	95,326	6,018	80%/90%	5,535	56,184	87,075	95,123	5,760
30%/100%	5,568	54,691	87,886	94,903	5,975	80%/100%	5,517	56,432	87,441	94,740	5,732
40%/10%	5,567	56,342	86,866	94,531	6,267	90%/10%	5,765	52,619	86,568	95,010	6,404
40%/20%	5,660	55,046	86,844	94,329	6,190	90%/20%	5,517	56,049	87,272	95,061	6,087
40%/30%	5,655	56,122	85,554	94,498	5,899	90%/30%	5,557	55,317	87,278	95,044	5,976
40%/40%	5,584	54,235	86,883	95,556	6,008	90%/40%	5,552	55,159	87,768	95,151	5,971
40%/50%	5,630	55,469	86,985	94,256	5,863	90%/50%	5,590	54,528	87,616	95,134	5,938
40%/60%	5,507	56,178	87,283	95,072	5,992	90%/60%	5,551	54,477	87,768	95,174	5,905
40%/70%	5,555	55,108	87,745	95,050	5,865	90%/70%	5,520	56,854	86,912	94,779	5,855
40%/80%	5,589	54,292	88,179	94,965	5,903	90%/80%	5,554	55,429	87,413	94,920	5,722
40%/90%	5,565	53,407	87,689	95,855	6,024	90%/90%	5,532	55,739	87,723	95,005	5,739
40%/100%	5,566	54,156	87,559	95,748	5,886	90%/100%	5,537	55,390	87,255	95,444	5,760
50%/10%	5,579	54,489	87,097	95,489	6,409	100%/10%	5,549	54,810	87,649	95,219	6,258
50%/20%	5,603	55,654	87,064	94,661	6,041	100%/20%	5,652	54,461	85,813	95,281	6,040
50%/30%	5,527	55,823	87,644	95,033	6,005	100%/30%	5,624	53,464	87,052	95,782	6,067
50%/40%	5,559	55,959	86,979	94,689	5,951	100%/40%	5,581	57,592	85,971	94,109	5,925
50%/50%	5,571	55,435	87,565	94,841	5,969	100%/50%	5,564	54,117	87,925	95,399	5,836
50%/60%	5,631	53,987	87,537	95,005	5,858	100%/60%	5,514	55,649	87,723	95,078	5,815
50%/70%	5,606	54,950	86,495	95,050	5,832	100%/70%	5,557	57,693	85,813	94,278	5,808
50%/80%	5,517	54,798	87,914	95,489	5,864	100%/80%	5,587	54,061	87,294	95,202	5,865
50%/90%	5,558	55,255	87,830	94,971	5,840	100%/90%	5,580	55,379	87,435	94,734	5,786
50%/100%	5,553	56,516	87,199	94,610	5,808	100%/100%	5,559	57,485	86,360	94,216	5,780

Na podstawie zamieszczonej tabeli 7, można stwierdzić, iż zastosowanie różnych wartości prawdopodobieństwa mutacji oraz krzyżowania, inaczej wpływa na wysokość błędu prognozowania. Algorytm znajduje lepsze wartości dla większych prawdopodobieństw mutacji oraz krzyżowania. Po przeprowadzeniu badań, najlepszy wynik uzyskano dla prawdopodobieństw: mutacji 90% oraz krzyżowania 80%. Średni błąd prognozy wyniósł 5,722 $\mu\text{g}/\text{m}^3$ .

## 5. Podsumowanie

W artykule przedstawiono zastosowanie algorytmów genetycznych do prognozowania pyłu PM10. Proces ten był drugim, który (po wykorzystaniu sieci neuronowych) był zastosowany do prognozowania pyłu. Sekwencja stosowania obu inteligentnych mechanizmów przetwarzania była konieczna, aby pozyskać dane które stanowią będą podstawę do projektowania systemu samouczącego do prognozowania poziomu zawartości pyłu PM10.

Z wykorzystaniem algorytmu genetycznego zaprojektowano rozwiązanie zwracające z 24 godzinnym wyprzedzeniem stężenie pyłu PM10 w powietrzu atmosferycznym. Prace rozpoczęto od budowy modelu prognozowania pyłu PM10. Następnie, w oparciu o model, zaimplementowano algorytm genetyczny, konfigurując kolejno poszczególne elementy: budowę osobnika, wielkość populacji, funkcję przystosowania, metody selekcji, krzyżowania oraz mutacji. Ostatnim etapem budowy algorytmu było znalezienie optymalnego czasu życia osobnika, który określono na 7 generacji, oraz prawdopodobieństw mutacji i krzyżowania, które wyniosły kolejno 90% i 80%.

Ostatecznie, z wykorzystaniem algorytmu genetycznego, udało się zaprojektować i zaimplementować rozwiązanie, zapewniające możliwość prognozy stężenia pyłu PM10 na 24 godziny w przyszłość, z błędem wynoszącym  $5,722\mu\text{g}/\text{m}^3$ . Rozwiązanie zostało zastosowane w systemie do zarządzania jakością powietrza w Gdańsku.

Poprzedni zbudowany mechanizm przetwarzania wykorzystujący sztuczne sieci neuronowe, prognozował poziom pyłu PM10, z błędem wynoszącym średnio  $7,82265\mu\text{g}/\text{m}^3$ . Algorytm ten, podobnie jak opisany algorytm genetyczny posiada doskonałe zastosowanie w rozwiązywaniu problemów takich jak problemy prognozowania wartości.

Obie metody(a być może większa ilość w przyszłości), wykorzystane zostaną do przewidywania stężenia pyłu PM10. Uzyskane w tym procesie wartości przekazane zostaną do modułu strojenia, który na podstawie danych wejściowych(danych mających wpływ na przyszłe stężenie pyłu PM10) oraz rzeczywistego stężenia pyłu PM10 określi, który z algorytmów przy panujących warunkach lepiej nadaje się do użycia(possiada mniejszy błąd predykcji). Strojenie algorytmu odbywa się dynamicznie, a kolejne iteracje(wyniki prognoz przekazane do modułu strojenia) zwiększają dokładność doboru odpowiedniej metody do przewidywania stężenia pyłu PM10, dla określonych danych wejściowych.

## Literatura

1. Kowalczyk Z., Orłowski C.: Design of Knowledge-based Systems in Environmental Engineering, Cybernetics and Systems, vol. 35 nr 5-6, 2004, s. 487-498.
2. Orłowski C., Kowalczyk Z., Knowledge management based on dynamic and self-adjusting fuzzy models, in: Knowledge-Based Intelligent Information and Engineering Systems, Springer, Berlin–Heidelberg 2006.
3. Orłowski C., Sarzyński A.: A model for forecasting pm10 levels with the use of artificial neural networks, in: Information Systems Architecture and Technology - The Use of IT Technologies to Support Organizational Management in Risky Environment, Wrocław 2014.
4. Ishibuchi H., Nakashima T., Murata T., Comparison of the Michigan and Pittsburgh approaches to the design of fuzzy classification systems, Electronics and Communications in Japan, 1997.

5. Miller B. L., Goldberg D. E.: Genetic Algorithms, Selection Schemes, and the Varying Effects of Noise, Evolutionary Computation, Illinois 1996.
6. Cant-Paz, E.: Efficient and Accurate Parallel Genetic Algorithms, Kluwer Academic Publishers, Dordrecht 2000.
7. Miller B. L., Goldberg D. E.: Genetic Algorithms, Tournament Selection and the Effects of Noise, Complex Systems, Illinois 1995.
8. Hinterding R., Gielewski H., Peachey T. C.: The Nature of Mutation in Genetic Algorithms, in: Eshelman L.J.: Proceedings of the 6th International Conference on Genetic Algorithms, 1995.

Prof. dr hab. inż. Cezary ORŁOWSKI  
Katedra Zastosowań Informatyki w Zarządzaniu  
Wyższa Szkoła Bankowa w Gdańsku  
80-266 Gdańsk, Aleja Grunwaldzka 238A  
tel.: 508 87 67 84  
e-mail: corlowski@wsb.gda.pl

Mgr inż. Magdalena SARZYŃSKA  
Wydział Matematyki, Fizyki i Informatyki  
Zakład Sztucznej Inteligencji  
Instytut Informatyki  
Uniwersytet Gdański  
80-309 Gdańsk, ul. Jana Bażyńskiego 8  
tel.: 669 21 61 91  
e-mail: msarzynska@sigma.ug.edu.pl

Mgr inż. Arkadiusz SARZYŃSKI  
Wydział Zarządzania i Ekonomii  
Politechnika Gdańska  
Katedra Zastosowań Informatyki w Zarządzaniu  
Politechnika Gdańska  
80-233 Gdańsk, ul. G. Narutowicza 11/12  
tel.: 511 15 50 45  
e-mail: sarzynski.arkadiusz@gmail.com