

Kraków, 30.08.2022 r.

dr hab. inż. Marek Bogacki, prof. AGH
Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie
Wydział Geodezji Górniczej i Inżynierii Środowiska
Katedra Kształtowania i Ochrony Środowiska
Al. Mickiewicza 30, 30-059 Kraków

Recenzja

pracy doktorskiej mgr inż. Mariusza Filaka pt.: *Predykcja stężeń zanieczyszczeń powietrza w perceptronowych modelach regresyjnych*

1. Przedmiot recenzji

Przedmiotem recenzji jest rozprawa doktorska mgr inż. Mariusza Filaka zatytułowana: „Predykcja stężeń zanieczyszczeń powietrza w perceptronowych modelach regresyjnych”, wykonana na Wydziale Infrastruktury i Środowiska Politechniki Częstochowskiej pod kierunkiem Pana dr hab. Szymona Hoffmana. Celem recenzji jest ocena, czy rozprawa spełnia warunki określone w art. 187 ustawy z dnia 20 lipca 2018 r. *Prawo o szkolnictwie wyższym i nauce* (Dz.U. 2022, poz. 574). Wymogiem jest by rozprawa doktorska stanowiła oryginalne rozwiązanie problemu naukowego, by potwierdzała ogólną wiedzę teoretyczną kandydata w danej dyscyplinie naukowej oraz umiejętność samodzielnego prowadzenia przez doktoranta pracy naukowej. Ocenę wykonano pod kątem stawianych wymogów oraz innych formalnych kryteriów, w tym jakości edytorskiej.

2. Podstawa wykonania recenzji

Recenzja rozprawy doktorskiej mgr inż. Mariusza Filaka została wykonana na podstawie zlecenia Kierownika Rady Dyscypliny Naukowej Inżynieria Środowiska, Górnictwo i Energetyka Politechniki Częstochowskiej Pani dr hab. inż. Iwony Zawieja, prof. Pcz, z dnia 12 lipca 2022 roku (znak pisma: R.WIIS.BOD.41.2022.3) oraz dostarczonego egzemplarza pracy.

3. Wybór tematyki rozprawy doktorskiej

Czyste powietrze stanowi fundamentalne dobro niezbędne dla zapewnienia zdrowego życia człowieka. Stąd obowiązkiem państwa jest takie zarządzanie jakością powietrza, aby na całym obszarze kraju jego stan spełniał ustalone normami prawa kryteria. Ocenę poziomów substancji w powietrzu w Polsce dokonuje się w

zdefiniowanych strefach zgodnie z obowiązującymi przepisami prawa metodami pomiarowymi, a tam gdzie nie jest to wymagane przy pomocy modelowania matematycznego transportu i przemian chemicznych substancji w powietrzu.

Wyniki pomiarów wymaganych prawem stężeń zanieczyszczeń w powietrzu oraz podstawowych parametrów meteorologicznych uzyskane w ramach Państwowego Monitoringu Środowiska są standardowo poddawane walidacji wykonywanej przez służby Głównego Inspektora Ochrony Środowiska m.in. w kontekście spełnienia przez nie warunku minimalnego procentu ważnych danych pomiarowych w przyjętym w ramach oceny okresie uśredniania. Szeregi czasowe danych pomiarowych są więc oceniane m.in. pod kątem ich kompletności i wiarygodności (zakres niepewności). Istnieją w tym zakresie zdefiniowane kryteria oceny poprawności danych pomiarowych wyrażone w formie minimalnego procentu ważnych danych pomiarowych odniesionych do danego kryterium oceny jakości powietrza. Najczęstszymi przyczynami niespełnienia kryterium związanego z minimalnym procentem ważnych danych pomiarowych są zbyt duże braki danych pomiarowych w analizowanym szeregu czasowym wynikające m.in. z przerw w zasilaniu urządzeń pomiarowych, awarii analizatorów, awarii systemów gromadzenia i transmisji danych pomiarowych, prowadzonych procedur serwisowych i kalibracyjnych. Niespełnienie kryterium minimalnego procentu ważnych danych pomiarowych jest przyczyną braku możliwości miarodajnej oceny stanu jakości powietrza na danym obszarze w odniesieniu do danego zanieczyszczenia powietrza.

Znalezienie metody uzupełniania brakujących danych pomiarowych w szeregach czasowych opisujących np. zmienność stężeń zanieczyszczeń powietrza w danym punkcie przestrzeni w założonym okresie czasu stanowi interesujące wyzwanie badawcze. Trudność w aproksymacji brakujących danych w analizowanym szeregu czasowym związana jest przede wszystkim z koniecznością uwzględnienia w tworzonym modelu aproksymacyjnym możliwie wielu niezależnych czynników determinujących poziom stężenia badanego zanieczyszczenia w danym punkcie czasowym. Rozwiązanie tak zdefiniowanego problemu nie jest zagadnieniem trywialnym i wymaga zastosowania wyspecjalizowanych narzędzi matematycznych. Jednym z takich narzędzi mogą być sieci neuronowe.

Podjęte w ramach rozprawy doktorskiej Pana mgr inż. Mariusza Filaka badania wpisują się w pełni w tę interesującą tematykę. Doktorant postanowił opracować perceptronowy model regresyjny funkcjonujący na historycznych danych pomiarowych, który umożliwi oszacowanie wartości stężenia badanego zanieczyszczenia w danym punkcie czasowym (zmienna objaśniana) na podstawie historycznych danych pomiarowych stanowiących zmienne objaśniające, np. stężenia innych zanieczyszczeń powietrza, czy dane meteorologiczne mierzone na danej stacji monitoringu jakości powietrza.

Z naukowego punktu widzenia podjęty przez Doktoranta temat jest bardzo ciekawy a zarazem wymagający znajomości zarówno zagadnień z zakresu szeroko rozumianej ochrony powietrza jak i technik obliczeniowych, w tym metod

statystycznych. Uzyskane wyniki badań posiadają znaczący potencjał poznawczy oraz aplikacyjny, szczególnie w obszarze poprawy jakości historycznych danych pomiarowych.

Tematyka rozprawy doktorskiej mieści się w dziedzinie nauk inżynieryjno-technicznych oraz szeroko rozumianej dyscyplinie: Inżynieria Środowiska, Górnictwo i Energetyka.

4. Charakterystyka oraz wartość merytoryczna rozprawy

4.1. Charakterystyka ogólna rozprawy

Rozprawa liczy sumarycznie 118 stron wraz ze spisem treści, streszczeniami, wykazem skrótów i oznaczeń, bibliografią oraz spisem tabel i rysunków. Zasadnicza treść rozprawy licząca ok. 91 stron jest zawarta w siedmiu rozdziałach. Na dwunastu stronach zamieszczono spis literatury wykorzystywanej w pracy, który obejmuje 134 pozycje, dotyczące zarówno najnowszych badań naukowych prowadzonych w ośrodkach zagranicznych, jak i najważniejszych pozycji literatury pochodzącej z polskich badań w tym obszarze wiedzy. W pracy zamieszczono 30 tabel i 62 rysunki.

4.2. Syntetyczny opis przeprowadzonych badań

Rozprawa doktorska mgr inż. Mariusza Filaka pod tytułem „Predykcja stężeń zanieczyszczeń powietrza w perceptronowych modelach regresyjnych” wg jej autora miała na celu stworzenie narzędzia matematycznego opartego na neuronowym modelu regresyjnym służącego do autouzupelniania braków danych w historycznych szeregach czasowych wyników pomiarów stężeń zanieczyszczeń powietrza zarejestrowanych na stacji monitoringu jakości powietrza w wybranej lokalizacji. Przy czym cechą osobliwą stworzonego przez autora modelu miała być większa dokładność odwzorowania brakujących/modelowanych danych w porównaniu do istniejących i stosowanych obecnie tradycyjnych metod statystycznych. Badania nad zwiększeniem dokładności aproksymacji brakujących danych pomiarowych w analizowanym historycznym szeregu czasowym autor realizował przy użyciu sztucznych sieci neuronowych o architekturze perceptronu wielowarstwowego wykorzystując w tym celu środowisko obliczeniowe programu Statistica. Badania prowadzono na historycznych danych pomiarowych pochodzących z dwóch automatycznych stacji monitoringu jakości powietrza zlokalizowanych w Zabrze i Złotym Potoku (woj. Śląskie). Dane pomiarowe o rozdzielczości czasowej 1-godziny pochodziły z lat 2011-2016 i dotyczyły takich parametrów (zmiennie objaśniające) jak: stężenia w powietrzu ozonu (O_3), tlenku azotu (NO), dwutlenku azotu (NO_2), dwutlenku siarki (SO_2), pyłu zawieszonego (PM_{10}), tlenku węgla (CO) – nie dotyczy stacji Złoty Potok, oraz prędkości wiatru (PW), temperatury powietrza (T), natężenia promieniowania

słonecznego (I) i wilgotności względnej (W_{wzg}) – nie dotyczy stacji Zabrze.

W celu wyboru najbardziej korzystnego do modelowania typu sieci przetestowano trzy rodzaje sieci, a mianowicie sieć liniową, perceptronu wielowarstwowego (MLP – Multi-Layer Perceptron) składającego się z 10 neuronów w warstwie ukrytej oraz sieć radialnych funkcji bazowych (RBF – Radial Basis Functions) składającą się z 50 neuronów w warstwie ukrytej. Jako kryterium wyboru rodzaju sieci do dalszych badań posłużyły wartości dwóch błędów odnoszących wartości modelowanych stężeń do wartości stężeń rzeczywistych a mianowicie: średniego błędu bezwzględnego (MAE – Mean Absolute Error) oraz pierwiastka z błędu średniokwadratowego (RMSE – Root Mean Squared Error). Badania przeprowadzono na arbitralnie wybranym przez Doktoranta epizodzie 24-godzinnym stężeń O_3 zanotowanym na stacji Zabrze w dniu 6.04.2015 r. Wyniki z zamodelowania przedmiotowego epizodu w formie wartości uzyskanych błędów MAE i RMSE przekonały Doktoranta do wyboru do dalszych badań sieci perceptronu wielowarstwowego MLP jako obciążonego najmniejszymi błędami.

W kolejnym kroku badań Doktorant starał się odpowiedzieć na pytanie jaka ilość neuronów w warstwie ukrytej sieci MLP gwarantuje najlepsze dopasowanie wartości stężenia modelowanego do wartości stężenia rzeczywistego? Obliczenia porównawcze przeprowadzono dla modeli posiadających od 3 do 10 neuronów w warstwie ukrytej. Jako miarę dokładności odwzorowania wartości rzeczywistej stężenia badanego zanieczyszczenia powietrza Doktorant wykorzystał wartość błędu MAE oraz współczynnik korelacji Pearsona (r). Na podstawie przeprowadzonych w tym zakresie badań do dalszych analiz Doktorant zarekomendował jako najbardziej odpowiedni model oparty na 10 neuronach.

W ten sposób wybrany model posłużył do przeprowadzenia dalszych badań, których celem była odpowiedź na pytanie: jak zmienia się dokładność modelowania zmiennej objaśnianej w zależności od faktu wykorzystania podczas modelowania pełnozakresowego zbioru danych w porównaniu do modelowania wykonanego na zbiorach danych stanowiących równoliczne podzbiory danego zbioru pełnozakresowego?

Badania modelowe przeprowadzono więc na danych pomiarowych dotyczących stężeń 1-godzinnych 6 analizowanych zanieczyszczeń powietrza zarejestrowanych w latach 2011-2016 na stacjach monitoringu jakości powietrza zlokalizowanych w Zabrzu i Żłotym Potoku, tworząc dla każdego zanieczyszczenia osobny pełnozakresowy zbiór przypadków zawierających kompletne dane wszystkich zmiennych objaśniających (wejść modelu), a w dalszych wariantach obliczeń dzieląc zbiór pełnozakresowy na 2, 4 i 8 podzbiorów równolicznych.

Obliczenia modelowe na tak zdefiniowanych zbiorach danych prowadzone były wg dwóch schematów sortowania zbiorów przypadków uwarunkowanych dostępnością do danych o stężeniach rzeczywistych modelowanego zanieczyszczenia, a więc:

1) przy dostępności do wartości stężeń rzeczywistych/pomiarowych danego

zanieczyszczenia powietrza (zmiennej objaśnianej) przygotowane zbiory przypadków sortowane były według rosnących stężeń rzeczywistych modelowanego zanieczyszczenia (RVS – Real Values Sorting),

- 2) przy braku dostępności do wartości stężeń rzeczywistych/pomiarowych danego zanieczyszczenia powietrza (zmiennej objaśnianej) zbiór wartości stężeń zmiennej objaśnianej musiał być wyznaczony na drodze obliczeń modelowych i posortowany według rosnących stężeń „predykcyjnych” modelowanego zanieczyszczenia (PVS – Predicted Values Sorting).

Sumarycznie zdefiniowano i przebadano 60 modeli (RVS – 30 modeli, PVS – 30 modeli) dla których Doktorant przeprowadził ocenę ich dokładności stosując jako miarę dokładności modelu uzyskane wartości błędów MAE i RMSE. Najmniejsze wartości błędów MAE i RMSE uzyskał dla modeli podzakresowych typu RVS przy modelach 8-podzakresowych, natomiast dla modeli typu PVS przy modelach 2-, 4-podzakresowych.

4.3. Wartość naukowa rozprawy

Podjęte przez Doktoranta wyzwanie badawcze nie należy do zagadnień trywialnych. Zaprojektowanie badań w tym zakresie tematycznym wymagało od Doktoranta dobrej znajomości zagadnień z zakresu szeroko rozumianej ochrony powietrza, statystyki oraz znajomości narzędzi badawczych/obliczeniowych z obszaru eksploracji danych. Przed wyborem metody badawczej a następnie narzędzia badawczego należało bowiem rozeznaczyć możliwości zastosowania różnych technik obliczeniowych do osiągnięcia założonego celu badawczego. Doktorant wybrał jako narzędzie do realizacji zdefiniowanego zadania sieci neuronowe. Wybór narzędzia badawczego należy uznać za słuszny, co potwierdziły wyniki końcowe przeprowadzonych badań. Ciekawym podejściem do badań było przeprowadzenie oceny możliwości zwiększenia „skuteczności predykcyjnej” zbudowanego modelu perceptronowego poprzez podział pełnozakresowego zbioru danych na kilka równolicznych podzbiorów, które stanowiły odrębne modele. Taki zabieg, jak się okazało w czasie prowadzonych badań pozwolił nawet o kilkadziesiąt procent poprawić dokładność modelowania zmiennej objaśnianej. Nie ulega więc wątpliwości, że przeprowadzone badania wnoszą wartość dodaną do aktualnego stanu wiedzy w tym zakresie. Pokazują duży potencjał zbadanej metody obliczeniowej do zastosowań związanych z uzupełnianiem brakujących danych w historycznych szeregach czasowych, szczególnie w zakresie danych pochodzących z systemów automatycznego monitoringu jakości powietrza.

4.4. Główne uwagi krytyczne do rozprawy

Pierwsza uwaga dotyczy zapisu pierwszej hipotezy badawczej, która brzmi następująco: „*archiwalne dane pomiarowe mogą być wykorzystane do modelowania aktualnych stężeń zanieczyszczeń*”. Takie sformułowanie hipotezy sugeruje czytającemu, że przedmiotowy model bazując na danych historycznych reprezentujących zmienne objaśniające jest w stanie zaprognozować zmienną objaśnianą co najmniej w teraźniejszości – a tak nie jest. Warunkiem predykcji zmiennej objaśnianej w danym punkcie czasowym, w tym w teraźniejszości lub przyszłości jest istnienie w tych punktach czasowych predyktorów. Stąd jeśli predyktorami są historyczne wartości pomiarów np. stężeń zanieczyszczeń powietrza, wówczas zmienna objaśniana w każdym przypadku będzie miała charakter historyczny. Pewną niedoskonałością pierwszej hipotezy badawczej jest również zastosowany w niej skrót myślowy, który nie doprecyzowuje o jakie stężenia zanieczyszczeń chodzi (w powietrzu atmosferycznym, wodzie, glebie, ...). Domyślnie wiadomo z treści rozprawy, że dotyczy ona modelowania stężeń zanieczyszczeń powietrza, ale powinno to zostać w hipotezie doprecyzowane. Autor hipotezy pominął również ważną kwestię z nią związaną a mianowicie, że archiwalne dane pomiarowe mogą służyć do określania zmiennej objaśnianej, ale tylko w lokalizacji z której pochodzą, gdyż w innym, uogólnionym przypadku hipoteza ta jest z gruntu nieprawdziwa. Definiując hipotezę badawczą warto ją więc precyzyjnie zapisać, aby była jednoznaczna, gdyż w toku badań należy ją zweryfikować. Autor rozprawy w pierwszym zdaniu rozdziału 6 pn.: *Podsumowanie i dyskusja* stwierdza: „*Archiwalne dane pomiarowe stanowią dobrą bazę do modelowania stężeń zanieczyszczeń powietrza*” a w podrozdziale 7.1. pn.: *Wnioski główne* zapisał wniosek nr 1: „*Archiwalne dane pomiarowe monitoringu powietrza stanowią dobrą bazę do aproksymowania stężeń zanieczyszczeń powietrza*”. Oba stwierdzenia mają być uzasadnieniem pierwszej hipotezy badawczej. Czy jednak taki zapis hipotezy badawczej pozwala ją w jakikolwiek sposób zweryfikować?

Podobnie ogólnie sformułowana jest druga hipoteza badawcza, która zakłada, że: „*do modelowania predykcyjnego można wykorzystać sztuczne sieci neuronowe, które stanowią wygodne i praktyczne narzędzie do modelowania w porównaniu do tradycyjnych metod*”. W hipotezie tej również nie doprecyzowano do modelowania czego można wykorzystać sieci neuronowe? Poza oczywistym skrótem myślowym, hipoteza ta w kontekście postawionego i zrealizowanego w toku całych badań celu nie została w żaden ani jakościowy ani ilościowy sposób udowodniona. W jaki zobiektywizowany sposób autor rozprawy chciał dowieźć prawdziwość postawionej hipotezy? W którym miejscu rozprawy, poza drugim wnioskiem głównym (rozdz. 7.1.) mówiącym, że: „*Sieci neuronowe stanowią praktyczne narzędzie do modelowania predykcyjnego poziomów stężeń zanieczyszczeń powietrza*” autor udowadnia postawioną hipotezę?

Czytając wstęp do rozprawy, jej część teoretyczną oraz fragment rozdziału *Podsumowanie i dyskusja* można odnieść wrażenie, że głównym celem badań było

stworzenie modelu, który może służyć np. do (cytuję): „uzupełniania luk występujących w danych pomiarowych, ... stąd potrzeba stosowania modelowania, które może być użyteczne do zapewnienia kompletności wymaganej do oceny jakości powietrza”. Należy w tym miejscu zadać pytanie, czy obowiązujące w Polsce prawo dopuszcza stosowanie metod matematycznych do poprawy kompletności danych pomiarowych wyrażanych w formie minimalnego procentu ważnych danych pomiarowych? A jeśli nie, to do jakich innych zastosowań może być wykorzystany opracowany model?

Ważnym elementem każdej rozprawy doktorskiej jest jej część teoretyczna, w której Doktorant powinien przedstawić aktualny stan wiedzy światowej w obszarze związanym z realizowaną tematyką badań. Analiza źródeł literaturowych jest podstawą nie tylko do rozeznania istniejącego stanu wiedzy w badanym obszarze, ale przede wszystkim do poprawnego zdefiniowania wyzwań badawczych czekających na rozwiązanie. Dobra znajomość stanu prowadzonych na świecie badań pozwala również badaczom w sposób bardziej zobiektywizowany weryfikować wyniki swoich badań i na ich tle wskazywać np. innowacyjność zaproponowanych przez siebie rozwiązań. Doktorant wprawdzie wykazał w rozprawie korzystanie z ponad 130 źródeł, jednak w dużej mierze są to źródła literaturowe o charakterze bardzo ogólnym związanym m.in. z prawodawstwem ochrony środowiska, wpływem zanieczyszczeń powietrza na zdrowie człowieka (co nie było tematem tej rozprawy), ogólnie znaną wiedzą z zakresu sztucznych sieci neuronowych, czy powszechnymi informacjami na temat modelowania matematycznego w zakresie oceny jakości powietrza. W przedstawionej rozprawie część teoretyczna stanowiła 10 stron, natomiast stan światowych badań w zakresie przedmiotu rozprawy został opisany jedynie na niecałej jednej stronie.

4.5. Uwagi szczegółowe oraz pytania do Doktoranta

1. Przystępując do budowy jakiegokolwiek modelu statystycznego należy w pierwszej kolejności poznać naturę danych, na których będzie budowany model. Doktorant w rozdziałach 4.1.2. i 4.1.3. przeprowadził analizę opisową danych pomiarowych z obu analizowanych stacji monitoringu jakości powietrza. Miała jednak ona charakter bardzo uproszczony.

Pytanie:

- 1) dlaczego nie przeprowadzono głębszej analizy danych pomiarowych np. rozkładów poszczególnych danych, korelacji między nimi, itp.?
 - 2) jaki sens informacyjny ma procentowa średnia kompletność danych zarejestrowanych w okresie 2011-2016 w poszczególnych miesiącach na stacjach Zabrze oraz Złoty Potok (Tab. 4, Rys. 2 i Rys. 3)?
2. Wybór odpowiedniego typu sieci oraz odpowiedniej liczby warstw ukrytych i liczby neuronów wchodzących w skład poszczególnych warstw tej sieci jest zadaniem trudnym i najlepszym jego rozwiązaniem jest przetestowanie wielu różnych architektur sieci.

Pytanie:

- 1) jakie były przesłanki do przyjęcia w sposób arbitralny do porównania trzech konkretnych typów sieci tj.: liniowej - 10 neuronów, MLP – 10 neuronów, RBF – 50 neuronów?
 - 2) dlaczego uznano, że stężenie O_3 zarejestrowane na stacji Zabrze będzie najlepszym zbiorem do wyboru typu sieci?
 - 3) mając na uwadze bardzo małe różnice wartości pomiędzy błędami MAE i RMSE dla sieci MLP i RBF (Tab. 6), czy były podjęte jakieś próby przeliczenia na innym zbiorze danych zachowania się obu typów sieci?
 - 4) dlaczego porównanie wartości stężeń O_3 rzeczywistych i zamodelowanych przy pomocy sieci liniowej, MLP i RBF przedstawiono na mało czytelnym wykresie słupkowym (Rys. 8.) a nie na wykresie punktowym (Scatter plot) jak się to zwykle w tego typu przypadkach robi?
 - 5) czy badano zachowanie się sieci przy większej niż 10 liczbie neuronów?
3. Podział pełnozakresowych zbiorów danych na równoliczne podzbiory danych z dużym prawdopodobieństwem da inne wyniki błędów MAE i RMSE niż przy podziale danych na podzbiory wg innych kryteriów np. uwzględniających rozkład gęstości mierzonych stężeń zanieczyszczeń w powietrzu (rozkład prawdopodobieństwa), który wcześniej należało zidentyfikować. Czy były prowadzone analizy w tym kierunku, aby ocenić wpływ sposobu podziału zbiorów na dokładność wyników modelowania?

5. Poprawność redakcyjna rozprawy

Kolejność poszczególnych rozdziałów można uznać za prawidłową, za wyjątkiem niezrozumiałego połączenia dyskusji nad wynikami badań z podsumowaniem w rozdziale 6. Podsumowanie badań powinno być zamieszczone na końcu rozprawy wraz z wnioskami.

Styl rozprawy jest poprawny, chociaż zdarzają się w tekście również zapisy, których w pracach naukowych należy unikać. Przykładem może być zapis nieprecyzyjny w stylu: „... stężenia predykcyjne stosunkowo dobrze dopasowane do wartości rzeczywistych” (rozd. 5.2.6.). Stosowana w rozprawie terminologia jest prawidłowa. Można byłoby jedynie we wstępie do rozprawy doprecyzować pojęcie słowa „predykcja”, aby w sposób jednoznaczny zakotwiczyć to pojęcie w przyjętej metodyce obliczeń (model regresyjny) i aby nie sugerować czytelnikowi innego znaczenia słowa „predykcja” rozumianego powszechnie jako proces wnioskowania o przyszłych wielkościach zmiennych losowych w określonym przyszłym momencie (okresie), gdy nie jest znana wielkość wyjściowa.

Doktorant nie ustrzegł się w pracy pewnych potknięć literowych lub pomyłek. Przykładowo opisując nazwy zanieczyszczeń na str. 76 (rozd. 5.2.4.) powinno być:

„modelowania stężeń SO_2 ” zamiast NO_2 lub w rozdz. 5.2.3. Doktorant cytując wartości błędów MAE i RMSE z tabel 23 i 24 źle je wpisał w tekst.

Dobór i wykorzystanie w rozprawie źródeł literaturowych pozostawia duży niedosyt. Nie dotyczy to liczby pozycji, lecz ich adekwatności do tematyki rozprawy. Jest on konsekwencją przyjętego przez Doktoranta sposobu przygotowania części teoretycznej rozprawy.

Sposób zredagowania rozprawy doktorskiej pozostawia również dużo do życzenia, szczególnie w zakresie optymalnego i estetycznego zagospodarowania każdej strony. Wiele stron jest pustych w 20-40%. Doktorant nie zoptymalizował wielkości rysunków i tabel w taki sposób, aby lepiej wpasowywały się w układ strony. Na dużej części stron znajdują się same rysunki lub tabele bez tekstu. Rysunki znacznie lepiej prezentowałyby się w pracy bez czarnej ramki. Rysunki 2 i 3 wyglądałyby lepiej bez tła.

Wszystkie te uwagi nie mają wprawdzie znaczenia merytorycznego, ale wpływają na sposób odbioru pracy przez czytającego i powinny być w przypadku ewentualnej publikacji niniejszej pracy wzięte pod uwagę.

6. Wnioski końcowe

Podsumowując, przedstawioną do recenzji rozprawę doktorską Pana mgr Mariusza Filaka oceniam jako wartościową. Doktorant wykazał się dobrą znajomością narzędzi badawczych oraz innowacyjnością w podejściu do założonego celu badawczego. Przedstawiona do recenzji rozprawa nie jest pozbawiona pewnych drobnych błędów, o których wspomniano wcześniej. Pewne rozwiązania warsztatowe zastosowane przez Doktoranta w ramach badań można było z pewnością inaczej zaaranżować, ale wszystkie te nieznaczne niedomagania zauważone w trakcie recenzji mieszczą się w granicach akceptowalnych dla warsztatu badacza i nie mają znamion błędów kardynalnych. Doktorant przedstawioną do recenzji rozprawą dowiódł umiejętności samodzielnego prowadzenia pracy naukowej poprzez zaplanowanie badań, stworzenie wiarygodnej bazy danych, przeprowadzenie analizy danych i właściwe opracowanie wyników.

Biorąc pod uwagę wszystkie powyższe argumenty, pomimo wylistowanych w recenzji drobnych zastrzeżeń mających w pewnej mierze charakter subiektywnej oceny recenzenta stwierdzam, że przedłożona do recenzji praca doktorska mgr inż. Mariusza Filaka zatytułowana: „*Predykcja stężeń zanieczyszczeń powietrza w perceptronowych modelach regresyjnych*”, stanowi oryginalne rozwiązanie istotnego problemu naukowego oraz spełnia wymogi określone w art. 187 ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz.U. z 2022 r., poz. 574 ze zm.). W związku z tym może ona stanowić podstawę do dalszych działań w postępowaniu o nadanie stopnia doktora w dziedzinie nauk inżynieryjno-technicznych w dyscyplinie Inżynieria Środowiska Górnictwo i Energetyka.

Bożena Hanel

