

Wrocław 3.01.2024 r.

dr hab. inż. Katarzyna Pentoś, prof. uczelni
Uniwersytet Przyrodniczy we Wrocławiu
Wydział Przyrodniczo-Technologiczny
ul. Norwida 25, 50-375 Wrocław
email: katarzyna.pentos@upwr.edu.pl
tel. 608527533

RECENZJA

rozprawy doktorskiej mgr inż. Patryka Hary

**Prognozowanie plonu i zawartości białka w nasionach grochu siewnego (*Pisum sativum* L.)
z zastosowaniem sztucznych sieci neuronowych oraz regresji wielorakiej w oparciu o
wyniki doświadczeń odmianowych**

Promotor: prof. dr hab. inż. Gniewko Niedbała

Promotor pomocniczy: dr inż. Magdalena Piekutowska

1. Podstawa formalna recenzji

Podstawą formalną recenzji jest uchwała nr 990/ITP-PIB/2023 Rady Naukowej Instytutu Technologiczno-Przyrodniczego – PIB w Falentach z dnia 22 listopada 2023 roku oraz pismo nr DON-RN.61/2023 z dnia 18 grudnia 2023 roku.

2. Charakterystyka pracy doktorskiej

Praca obejmuje 112 stron, zawiera autoreferat oraz 3 załączniki - kopie artykułów naukowych wchodzących w skład rozprawy doktorskiej.

Praca doktorska „Prognozowanie plonu i zawartości białka w nasionach grochu siewnego (*Pisum sativum* L.) z zastosowaniem sztucznych sieci neuronowych oraz regresji wielorakiej w oparciu o wyniki doświadczeń odmianowych” jest pracą przygotowaną w oparciu o cykl trzech spójnych tematycznie artykułów naukowych opublikowanych w recenzowanych czasopismach o zasięgu międzynarodowym, których tematyka jest zgodna z tematem rozprawy doktorskiej. Publikacje charakteryzują się łącznym IF wynoszącym 11,1 i łączną punktacją Ministerstwa Edukacji i Nauki wynoszącą 350 pkt. Publikacje zostały przygotowane przez doktoranta wraz z promotorem i promotorem pomocniczym. We wszystkich publikacjach doktorant jest pierwszym autorem a jego wiodąca rola w realizacji badań jest potwierdzona przez zakres prac wykonanych przez poszczególnych współautorów zamieszczony w każdej publikacji. Doktorant był odpowiedzialny m.in. za opracowanie koncepcji i metodyki badań, ich wykonanie, opracowanie wyników i przygotowanie tekstu artykułów. Wszystkie artykuły zostały opublikowane w czasopismach wydawnictwa MDPI. Autoreferat obejmuje wprowadzenie,

przegląd literaturowy omawiający aktualny stan wiedzy związanej z tematyką pracy doktorskiej, cel pracy oraz postawione przez autora hipotezy badawcze i omówienie publikacji stanowiących podstawę rozprawy. Zakończenie rozprawy stanowi przedstawienie i dyskusja wyników oraz podsumowanie. Ta część pracy zakończona jest bibliografią, obejmującą 89 pozycji pochodzących przede wszystkim z ostatnich pięciu lat.

Przedłożona do recenzji rozprawa pt. „Prognozowanie plonu i zawartości białka w nasionach grochu siewnego (*Pisum sativum* L.) z zastosowaniem sztucznych sieci neuronowych oraz regresji wielorakiej w oparciu o wyniki doświadczeń odmianowych” obejmuje 3 oryginalne prace naukowe z IF, opublikowane w czasopismach z listy Ministerstwa Edukacji i Nauki:

1. Hara, P.; Piekutowska, M.; Niedbała, G. Selection of Independent Variables for Crop Yield Prediction Using Artificial Neural Network Models with Remote Sensing Data. *Land* 2021, *10*, 609, doi:10.3390/land10060609.
2. Hara, P.; Piekutowska, M.; Niedbała, G. Prediction of Protein Content in Pea (*Pisum sativum* L.) Seeds Using Artificial Neural Networks. *Agriculture* 2022, *13*, 29, doi:10.3390/agriculture13010029.
3. Hara, P.; Piekutowska, M.; Niedbała, G. Prediction of Pea (*Pisum sativum* L.) Seeds Yield Using Artificial Neural Networks. *Agriculture* 2023, *13*, 661, doi:10.3390/agriculture13030661.

Prace zostały opublikowane w latach 2021-2023 w czasopismach o zasięgu międzynarodowym, przypisanych do dyscypliny rolnictwo i ogrodnictwo.

Tematyka pracy i przeprowadzone badania wpisują się w trendy ogólnoświatowe dotyczące optymalizacji produkcji rolniczej. Jest ona niezbędna w świetle zmian klimatycznych i kurczących się zasobów, które można wykorzystywać w produkcji żywności. W ostatnich latach w modelowaniu i optymalizacji procesów związanych z rolnictwem coraz częściej stosuje się metody sztucznej inteligencji, w tym uczenia maszynowego, które zastępują stosowane dotąd metody statystyczne. Dużą zaletą badań doktoranta jest zastosowanie tego interdyscyplinarnego podejścia i wykorzystanie sztucznych sieci neuronowych do predykcji plonowania grochu siewnego w kontekście ilościowym i jakościowym. Dzięki zastosowaniu nowoczesnych narzędzi informatycznych możliwe było osiągnięcie znacząco dokładniejszej predykcji niż w przypadku regresji liniowej.

3. Cel pracy

W autoreferacie doktorant sformułował cel pracy i hipotezy badawcze. Jako cel główny pracy autor wskazał budowę modeli prognozujących plon nasion grochu siewnego oraz zawartego w nich białka na 20 dni przed zbiorem. W tym celu autor zastosował wieloraką regresję liniową oraz sieci neuronowe typu perceptron wielowarstwowy (MLP). Celem dodatkowym było wskazanie zmiennych niezależnych, które w największym stopniu wpływają na ilość i jakość plonu nasion grochu.

Uwzględniając analizę źródeł literaturowych i planowany zakres badań sformułowano następujące hipotezy badawcze:

H1: Modele bazujące na sztucznych sieciach neuronowych cechują się większą dokładnością prognozy w odniesieniu do modeli zbudowanych przy użyciu wielorakiej regresji liniowej.

H2: Pięcioletnie badania polowe w warunkach doświadczeń odmianowych pozwalają na budowę modeli prognozujących plon nasion grochu i zawartego w nich białka z błędem MAPE nieprzekraczającym 10%.

H3: Analiza wrażliwości sieci neuronowej jest właściwym narzędziem do wartościowania czynników plonotwórczych (ilościowych i jakościowych).

Zarówno cele pracy jak i hipotezy zostały sformułowane prawidłowo.

4. Ocena merytoryczna pracy

Oceniając rozprawę doktorską w zakresie kryteriów merytorycznych stwierdzam, że recenzowana praca dotyczy interesującego i aktualnego zagadnienia. Prognozowanie plonów, zarówno ilościowe jak i jakościowe jest istotne zarówno dla rolnika jak i dla firm przetwarzających płody rolne oraz jednostek rządowych. Pozwala na przeprowadzenie działań optymalizujących produkcję roślinną oraz planowanie powierzchni do przechowywania i zapewnienie środków niezbędnych w produkcji roślinnej. Dlatego też opracowywanie dokładnych modeli pozwalających na skuteczną predykcję ma kluczowe znaczenie w tym zakresie. W rozdziale 1. autoreferatu doktorant przeanalizował znaczenie modeli matematycznych pozwalających na przewidywanie plonu i podkreślił korzyści wynikające z zastosowania uczenia maszynowego w tym zakresie. W końcowej części tego rozdziału autor przedstawił opinię, że modele pozwalają uzyskać rzetelne informacje o wielkości zbiorów, co jest niezbyt fortunnym sformułowaniem. Modele te pozwalają tylko z pewną dokładnością przewidzieć wybrane parametry plonów.

W rozdziale 2. doktorant przedstawił przegląd literaturowy związany z tematyką pracy. Pierwszy podrozdział dotyczy znaczenia uprawy grochu siewnego i potwierdza celowość wyboru tej rośliny i sensowność budowania modeli predykcyjnych jej plonowania. W podrozdziale 2.2. doktorant przedstawił metody uczenia maszynowego jako skuteczne narzędzie w prognozowaniu plonu roślin uprawnych. Doktorant skupił się na sztucznych sieciach neuronowych, ponieważ to narzędzie wykorzystał w swojej pracy. Zaletą tego opisu jest wskazanie wielu ciekawych przykładów zastosowania sieci neuronowych do prognozowania plonu roślin uprawnych. Jednak w tym podrozdziale autor umieścił kilka dyskusyjnych sformułowań. Wydaje się, że wynikają one głównie z tego, że doktorant utożsamia sieci neuronowe z perceptronem wielowarstwowym. Np. stwierdzenie, że sieci neuronowe składają się z minimum trzech warstw jest pewnym uogólnieniem, ponieważ istnieją sieci składające się z dwóch warstw, np. sieci Kohonena. Opis redukcji zmiennych niezależnych poprzez sprowadzanie wag do 0 też nie jest poprawny, ponieważ redukcja wag

nie dotyczy warstwy wejściowej, która nie składa się z neuronów tylko z punktów dostępowych, które nie mają wag synaptycznych. Taka redukcja wag dotyczy kolejnych warstw w sieci. Trudno zgodzić się ze stwierdzeniem, że „Za pomocą modeli sztucznych sieci neuronowych istnieje możliwość pokazania ukrytych powiązań pomiędzy zmiennymi wejściowymi”. Myślę, że doktorantowi chodziło o powiązania między zmiennymi wejściowymi i wyjściowymi.

W rozdziałach 5 – 7 doktorant zwięźle opisał treść publikacji przedstawionych jako praca doktorska.

Wybór zmiennych niezależnych do prognozowania wielkości plonu przy pomocy sztucznych sieci neuronowych

Pierwszy artykuł przedstawiony przez doktoranta jako element pracy doktorskiej został opublikowany w czasopiśmie Land po tytule „Selection of Independent Variables for Crop Yield Prediction Using Artificial Neural Network Models with Remote Sensing Data”. Jest to artykuł przeglądowy, w którym przeanalizowany jest sposób doboru zmiennych niezależnych dla modeli do predykcji plonu. W pracy tej przedstawiony został schemat blokowy pokazujący etapy budowania modeli predykcyjnych (rys. 1). Wynika z niego, że dobór zmiennych niezależnych modelu jest jedną z kluczowych decyzji podejmowanych przez projektanta. Przygotowując ten artykuł, doktorant dokonał bardzo szerokiego przeglądu literaturowego (175 pozycji bibliograficznych). Dzięki temu zidentyfikował potencjalne parametry wpływające na plon oraz dość dokładnie omówił naturę związku między tymi parametrami a plonem. Są to czynniki związane z warunkami pogodowymi i klimatycznymi, glebą i uprawą. Omówiono także możliwość wykorzystania wskaźników związanych z produktywnością roślin. Doktorant przeanalizował także ograniczenia związane z wykorzystaniem omówionych parametrów jako zmiennych niezależnych modeli predykcyjnych. Praca ta jest niezwykle ważnym etapem w realizacji doktoratu, ponieważ dzięki niej doktorant mógł w sposób przemyślany i prawidłowy zaplanować eksperymenty, które były podstawą dalszych etapów badań.

Predykcja zawartości białka w nasionach grochu siewnego (*Pisum Sativum* L.) przy użyciu modelu ANN

Drugi artykuł przedstawiony przez doktoranta do oceny w ramach pracy doktorskiej został opublikowany w czasopiśmie Agriculture po tytule „Prediction of Protein Content in Pea (*Pisum sativum* L.) Seeds Using Artificial Neural Networks”. W tej pracy autor zbudował dwa modele regresyjne do predykcji zawartości białka w nasionach grochu siewnego na podstawie 19 zmiennych niezależnych. Do budowy modeli zostały wykorzystane dwie techniki – sieć neuronowa MLP oraz wieloraka regresja liniowa. Dane do budowy modeli zostały zebrane prawidłowo i pochodziły z eksperymentu obejmującego 5 lat. Dobór zmiennych niezależnych nie budzi zastrzeżeń i został wykonany w oparciu o wcześniejsze analizy. Liczba zebranych danych była wystarczająca do budowy modeli predykcyjnych (1.155 poletek). Dane

zostały podzielone na dwa podzbiory, z których pierwszy został wykorzystany do budowy modeli a drugi do ich walidacji. Doktorant przeanalizował jakość uzyskanych modeli na podstawie kilku metryk błędu. Model oparty o regresję liniową charakteryzował się niską dokładnością, co jest typową sytuacją przy predykcji plonu. Natomiast model neuronowy miał wysoką dokładność i może zostać wykorzystany w praktyce. Dodatkowo, została wykonana analiza wrażliwości modelu neuronowego, która pozwoliła na ustalenie stopnia wpływu poszczególnych zmiennych niezależnych na zawartość białka w nasionach grochu. Tego typu analizy wnoszą dodatkową, wartościową informację o naturze relacji między zmiennymi niezależnymi i zależnymi. Wyniki zostały przedstawione w artykule w tabeli 6, która została nieprawidłowo podpisana jako parametry jakościowe modelu N1. W pracy została przedstawiona wartościowa dyskusja wyników a wnioski mają poparcie w uzyskanych wynikach.

Analiza porównawcza skuteczności sztucznych sieci neuronowych i wielorakiej regresji liniowej w przewidywaniu plonu nasion grochu siewnego (*Pisum Sativum* L.)

Trzeci artykuł przedstawiony jako część pracy doktorskiej został opublikowany w czasopiśmie Agriculture po tytule „Prediction of Pea (*Pisum sativum* L.) Seeds Yield Using Artificial Neural Networks”. W tym przypadku metodyka badań była zbieżna z wykorzystaną do predykcji zawartości białka w nasionach grochu. Inny był tylko przewidywany parametr - plonu nasion. Artykuł ten jest bliźniaczy z poprzednim artykułem również co do wyników, dyskusji i wniosków.

W rozdziale 8. autoreferatu doktorant przedstawił omówienie i dyskusję wyników zaprezentowanych w publikacjach. Doktorant podkreślił znaczenie predykcji plonu w kontekście jakościowym i ilościowym dla nowoczesnego rolnictwa. Odniósł się do zalet stosowania metod uczenia maszynowego w porównaniu do metod regresyjnych. Przeanalizował wybrane przez siebie grupy zmiennych niezależnych wykorzystanych do budowy modeli (fitofenologiczne, pogodowe, glebowe i agronomiczne). Następnie doktorant dość szczegółowo omówił modele regresyjne i neuronowe do predykcji plonu i zawartości białka w ziarnach grochu. Przeanalizował wartości metryk błędów i dokonał analizy porównawczej z wynikami innych autorów. Odnosząc się do publikacji 77 i 81 w spisie literatury, doktorant użył dość niefortunnego stwierdzenia, że autorzy zbudowali model składający się z jednej warstwy ukrytej. Nie wydaje się to możliwe i z pewnością model ten zawierał także warstwę wejściową i wyjściową. Analizując sposób, w jaki doktorant porównuje wykorzystane w badaniach metody modelowania, mogę stwierdzić, że posiada szeroką wiedzę na ich temat i dość swobodnie posługuje się tym warsztatem badawczym. Jednakże w tekście pojawiło się kilka stwierdzeń dyskusyjnych. Np. doktorant, odnosząc się do metod regresyjnych stwierdza, że „w regresji liniowej zakłada się brak korelacji między zmiennymi objaśniającymi”. Tymczasem, stosując dowolne metody budowy modeli regresyjnych, należy zwrócić uwagę na ewentualne korelacje między zmiennymi niezależnymi i dokonać

zmniejszenia liczby zmiennych tak aby wyeliminować silne korelacje. Nie mogę też zgodzić się ze stwierdzeniem, że „W niniejszej rozprawie zastosowano wieloraką regresję liniową ponieważ w literaturze światowej praktycznie nie używa się innej metody porównawczej do ANN”. W literaturze światowej jest wiele przykładów porównywania modeli neuronowych z innymi metodami budowania modeli regresyjnych (regresje nieliniowe, maszyny wektorów nośnych, drzewa decyzyjne, programowanie genetyczne i wiele innych).

Autoreferat kończy podsumowanie i wnioski sformułowane przez doktoranta na podstawie wyników badań. Wnioski generalnie nie budzą zastrzeżeń i wynikają bezpośrednio z wyników przedstawionych w publikacjach. Mam jedynie uwagę do wniosku nr 7, który został sformułowany niepoprawnie językowo. Myślę, że w treści wniosku brak jest fragmentu tekstu.

5. Uwagi krytyczne i dyskusyjne o charakterze merytorycznym

- Doktorant słusznie stwierdza w podrozdziale 2.2., że jednym z kluczowych etapów budowania modeli w oparciu o uczenie maszynowe jest prawidłowy dobór zmiennych niezależnych. Istnieje wiele algorytmów, które pozwalają przeprowadzić taką selekcję, jednak doktorant nie skorzystał z takich metod. Proszę o wyjaśnienie, dlaczego doktorant przeprowadził selekcję wyłącznie w oparciu o znajomość modelowanych zależności i czy rozważał wykorzystanie algorytmów do selekcji parametrów wejściowych modeli.

- Doktorant dość szeroko opisał w autoreferacie sztuczne sieci neuronowe. Dlaczego nie zostały opisane metody regresyjne, które również były wykorzystywane w budowaniu modeli?

- W artykule „Prediction of Protein Content in Pea (*Pisum sativum* L.) Seeds Using Artificial Neural Networks” doktorant przedstawił parametry jakościowe modelu N1 w dwóch tabelach: 4 i 7. Proszę o sprecyzowanie metryk przedstawionych w tabeli 4 oraz wyjaśnienie ich związku z metrykami przedstawionymi w tabeli 7. Ta sama sytuacja dotyczy artykułu „Prediction of Pea (*Pisum sativum* L.) Seeds Yield Using Artificial Neural Networks”.

- W czasie budowy modelu regresyjnego do predykcji zawartości białka w ziarnach grochu, 7 z 19 zmiennych niezależnych zostało odrzuconych. Czy doktorant rozważał budowę modelu neuronowego w oparciu o pomniejszony zestaw parametrów?

- W rozdziale 8. doktorant zawarł stwierdzenie „Wybór odpowiedniego typu modelu ANN jest kolejnym istotnym elementem determinującym wysoką jakość predykcji. Testowanie metodą prób i błędów wykazało, że modelem najlepiej prognozującym plon i zawartość białka w grochu jest sieć typu MLP”. Jakie inne typy sieci doktorant testował i jakie były uzyskane wyniki?

- Proszę o wyjaśnienie jak technicznie, w środowisku Statistica, zostało wykonane dwuetapowe uczenie sieci neuronowych („W pierwszym etapie sztuczne sieci neuronowe uczone były metodą wstecznej propagacji błędów. Drugi krok obejmował uczenie metodą gradientów sprzężonych”).

6. Uwagi edytorskie

W autoreferacie doktorant popełnił drobne błędy językowe, które nie wpływają na jakość merytoryczną pracy. Poniżej przykłady takich błędów:

- strona 8: jest „Dzięki tym informacja” zamiast „Dzięki tym informacjom”
- strona 9: jest „w odniesieniu do modele” zamiast „w odniesieniu do modeli”
- strona 11: jest „uprawa grochu wpisuję się” zamiast „uprawa grochu wpisuję się”
- strona 12: jest „Istotną elementem” zamiast „Istotnym elementem”
- strona 16: niefortunne jest stwierdzenie „opłacalność produkcji konkretnego gatunku jest na granicy opłacalności”
- strona 21: jest „w nasion grochu” zamiast „w nasionach grochu”
- strona 21: jest „opublikowany została” zamiast „opublikowany został”
- strona 21: jest „Baza danych stworzono” zamiast „Bazę danych stworzono”
- strona 25: jest „Powierzchnia odpowiedzi...przedstawiono” zamiast „Powierzchnię odpowiedzi...przedstawiono”
- strona 25: jest „Poszczególne rodzaje błędów oraz ich wartości został zilustrowane” zamiast „Poszczególne rodzaje błędów oraz ich wartości zostały zilustrowane”
- strona 26: jest „ukazała ciekawe i mało znany wpływ” zamiast „ukazała ciekawy i mało znany wpływ”
- strona 35: jest „Tempo ... zależny jest” zamiast „Tempo ... zależne jest”
- strona 35: jest „ukierunkowane będą na redukcji i/lub zastosowaniu dodatkowych zmiennych” zamiast „ukierunkowane będą na redukcję i/lub zastosowanie dodatkowych zmiennych”
- strona 37: jest „Zadania te powinno być” zamiast „Zadania te powinny być”

Sformułowane w recenzji uwagi krytyczne i dyskusyjne a także uwagi edytorskie nie obniżają wartości merytorycznej pracy. Podsumowując stwierdzam, że praca pod względem merytorycznym nie budzi żadnych zastrzeżeń i napisana jest na dobrym poziomie naukowym.

7. Wniosek końcowy

Uważam, że rozprawa doktorska mgr inż. Patryka Hary jest bardzo cennym opracowaniem, w którym potwierdzono główne tezy pracy i rozwiązano problem naukowy. Doktorant wykazał się znajomością warsztatu badawczego, umiejętnością analizy doniesień literaturowych, dobrym planowaniem eksperymentu. Doktorant potrafił prawidłowo dobrać metody analizy danych i przedstawić wyniki badań w formie publikacji. Na podkreślenie zasługuje fakt, że praca ma charakter interdyscyplinarny i doktorant musiał wykazać się wiedzą z zakresu rolnictwa oraz

uczenia maszynowego. Tematyka i zakres rozprawy doktorskiej są związane z ważnym problemem, który dotyczy prognozowania plonu roślin uprawnych. Doktorant jako obiekt badań wybrał groch siewny, czyli roślinę o dość dużym znaczeniu w uprawach w Polsce. Wyniki badań mogą mieć zastosowanie praktyczne w budowaniu narzędzi doradczych dla nowoczesnego rolnictwa. Biorąc po uwagę informacje zaprezentowane powyżej stwierdzam, że rozprawa doktorska mgr inż. Patryka Hary pt. *„Prognozowanie plonu i zawartości białka w nasionach grochu siewnego (*Pisum sativum* L.) z zastosowaniem sztucznych sieci neuronowych oraz regresji wielorakiej w oparciu o wyniki doświadczeń odmianowych”* spełnia wszystkie wymagania określone w art. 187 Ustawy z dnia 20 lipca 2018 roku Prawo o szkolnictwie wyższym i nauce (tekst jednolity DZ. U. z 2021 roku poz. 478) stawiane rozprawom doktorskim. Dlatego wnioskuję o przyjęcie rozprawy doktorskiej mgr inż. Patryka Hary i dopuszczenie jej do publicznej obrony przed Radą Naukową Instytutu Technologiczno-Przyrodniczego – PIB w Falentach. Mając na uwadze dużą wartość naukową pracy oraz jej interdyscyplinarny charakter wnioskuję o wyróżnienie recenzowanej rozprawy doktorskiej.

dr hab. inż. Katarzyna Pentoś, prof. uczelni

