

## RECENZJA ROZPRAWY DOKTORSKIEJ

Tytuł rozprawy:      Metodyka diagnozowania przyczyn powstawania wad wyrobów  
na podstawie zaawansowanego modelowania opartego na  
dużych zbiorach danych

Autor rozprawy:     mgr inż. Alicja Okuniewska

Promotor:            prof. dr hab. inż. Marcin Andrzej Perzyk

### Podstawa wykonania recenzji

Podstawą wykonania niniejszej recenzji jest pismo Przewodniczącego Rady Naukowej Dyscypliny Inżynieria Mechaniczna Politechniki Warszawskiej, prof. dra hab. inż. Roberta Sitnika, RDNIM.521.19.2023 z dnia 31 maja 2023 roku.

### Ocena formalna rozprawy

Przedstawiona do recenzji praca została wydana w formie książki. Rozprawa liczy 273 strony. Rozprawę poprzedza streszczenie w języku polskim i angielskim oraz słowa kluczowe.

Zasadnicza część rozprawy została podzielona na sześć rozdziałów.

*Rozdział 1* zawiera ciekawe wprowadzenie w tematykę podjętą w rozprawie i definicję obszaru badawczego. Autorka sformułowała także tezę badawczą i sprecyzowała cele badawcze rozprawy.

W *rozdziale 2* Autorka omawia funkcjonowanie sztucznych sieci neuronowych i główne założenia uczenia modeli opartych na dużych zbiorach danych. W rozdziale zawarty jest przegląd struktur, rodzajów i sposobów uczenia oraz konstruowania sztucznych sieci neuronowych.

*Rozdział 3* zawiera podstawowe informacje dotyczące drzew decyzyjnych. Autorka opisuje historię powstania metody, podstawy teoretyczne oraz przykłady zastosowań drzew regresyjnych w przemyśle odlewniczym.

*Rozdział 4* zawiera bardzo krótki opis metody maszyn wektorów wspierających obejmujący jej podstawy teoretyczne i przykłady zastosowań w procesach wytwarzania (odlewnia).

W *rozdziale 5* Autorka przedstawiła efekty prac badawczych związanych z wstępnym przetwarzaniem danych, zastosowaniem metod sztucznych sieci neuronowych, drzew decyzyjnych i maszyny wektorów wspierających, w wyniku których opracowana została metodologia diagnozowania przyczyn powstawania wad odlewów.

*Rozdział 6* zawiera podsumowanie dokonań rozprawy i wnioski. Autorka podsumowuje w nim wyniki przeprowadzonych badań wraz z parametryczną oceną uzyskanych wyników dla poszczególnych metod.

W pracy umieszczono także spisy rysunków i tabel.

Autorka w pracy powołała się na 153 źródła literaturowe oraz strony internetowe. Zacytowane pozycje literatury zostały dobrane we właściwy sposób. W przeważającej części źródła można uznać za aktualne.

Widoczny jest jednak nierównomierny podział pracy. Autorka zbyt ogólnie potraktowała analizowane zagadnienia dotyczące drzew decyzyjnych i metody maszyn wektorów wspierających. *Rozdział 3* dotyczący drzew decyzyjnych ma tylko 3 strony, *Rozdział 4* tylko 4 strony, brakuje w nich uporządkowania terminologii, szczegółowego opisu, definicji i klasyfikacji. Natomiast *Rozdział 2* dotyczący Sztucznych Sieci Neuronowych zajmuje 31 stron.

Układ przedstawionej do recenzji rozprawy jest prawidłowy, jednak nieco krytycznie oceniam dbałość Autorki o komunikatywność tekstu rozprawy i jakość edytorską rysunków. Autorka nie ustrzegła się także błędów edycyjnych. W pracy występują błędy gramatyczne, stylistyczne i interpunkcyjne. Autorka w rozprawie buduje długie wielokrotnie złożone zdania, czasami zajmujące 10 wierszy! Zdania czasami są sformułowane w taki sposób, że trudne jest ich zrozumienie.

Jako przykłady niedociągnięć mogą posłużyć następujące elementy rozprawy:

1. Na str.5 i 6 w streszczeniu użyto skrótu SNN dla Sztucznych Sieci Neuronowych, powinno być ANN w streszczeniu w j. angielskim i ANN lub SSN w streszczeniu w j. polskim.
2. Opisy na rys. 2.1.2, 2.1.3, 2.3.3, 2.5.4 i 4.1.1 są w j. angielskim.
3. Na str. 22 błędnie wpisano nazwę nagrody Turninga – powinno być Turinga.
4. Na stronie 29 w opisie do wzoru 2.2 błędnie wpisano „*j* –funkcja aktywacji” powinno być  $\varphi$ . Na stronach 28 i 29 Autorka stosuje różne oznaczenia tej samej wielkości – we wzorze 2.1 funkcja aktywacji to *j*, we wzorze 2.2 oznaczona jako  $\varphi$ .
5. Na str.35 przy pierwszym użyciu nie wyjaśniono skrótów GRNN, PNN i MLP.

6. Czasami niektóre stwierdzenia są niezrozumiałe dla czytelnika. Na przykład str. 44 „W badanym procesie występuje inny problem, którym jest identyfikacja zapisanej obserwacji i konkretnym fizycznym odlewem.”, str. 45. „Przechodząc do procesów odlewniczych, których modelowanie z zastosowaniem zaawansowanych metod uczących się jest uzasadnione przez złożony charakter tych procesów.”.
7. Kolejne przykłady trudnych do zrozumienia zdań:
  - str. 137 „Niektóre działania mogą być określone sztuką, jako, że realizowana jest wizja twórcy w oparciu o jego eksperckie doświadczenie niż podążanie za formalnymi procedurami będące niejako rutynowym.”,
  - str. 57 „Istotną nowością pracy jest między innymi stworzenie metodyki zawierającej metody eksploracji danych w sektorze produkcji i wytwarzania, w myśl koncepcji tzw. „Przemysłu 4.0.”, istotny wpływ na stworzenie sprawnego systemu kontroli procesu celem diagnostyki wadliwych produktów realizując cele koncepcji inteligentnej fabryki.”.
8. Na str. 53 są błędne odesłania do wzorów 4.3 - 4.6, w pracy wpisano 5.3 - 5.6.
9. W tekście rozprawy odesłanie do Rys. 5.2.1.57 ze str. 86 znajduje się dopiero na str. 114 i 115.
10. W tekście rozprawy brak jest odesłania do Tab. 5.59.
11. Na str. 231 są odesłania do rys. od 5.7.7.4 do 5.7.7.8 – w pracy nie ma rysunków o takich numerach, powinno być od rys.5.2.7.4 do 5.2.7.8.

### **Ocena merytoryczna rozprawy**

Tematyka przedstawiona w pracy jest aktualna i dotyczy coraz powszechniej stosowanych metod analizy danych oraz wykorzystania metod sztucznej inteligencji w praktyce przemysłowej, na przykładzie procesu odlewania ciśnieniowego.

Autorka w *Rozdziale 1* przedstawia tezę, cele i zakres swojej rozprawy. Głównym celem rozprawy jest dowiedzenie, że diagnozowanie przyczyn powstawania wad wyrobów wymaga zastosowania określonego zbioru metod zaawansowanego modelowania opartego na dużych zbiorach danych. Autorka przeprowadza analizę możliwych rozwiązań w celu wyboru określonych metod analizy dużych zbiorów danych. Jest świadoma istniejących ograniczeń w stosowaniu tych metod z uwagi na jakość i ilość analizowanych danych oraz konieczność wyboru odpowiednich parametrów modeli i poziomów dokładności.

Do analizy zastosowano dane rzeczywiste, pochodzące z procesu odlewania ciśnieniowego odlewów dla przemysłu motoryzacyjnego zarejestrowane w jednej z odlewni. Podczas badań analizowano jakość odlewów wyrażoną wynikiem testu szczelności. Autorka przeprowadza prace badawcze związane z wstępnym przetwarzaniem danych, a następnie z zastosowaniem metod sztucznych sieci neuronowych, drzew decyzyjnych i maszyn

wektorów wspierających oraz wielowymiarowej optymalizacji parametrów procesu w celu opracowania metodyki diagnozowania przyczyn powstawania wad odlewów.

Na podstawie uzyskanych wyników przeprowadzone zostają szczegółowe analizy i porównania wyników badań w celu wyboru odpowiednich metod uczenia i klasyfikacji do konkretnych zastosowań. Każda z użytych przez Autorkę metod pozwala na dość skuteczne prognozowanie wartości zmiennej zależnej, jednak metodą sztucznych sieci neuronowych uzyskano najlepsze dopasowanie i najmniejsze wartości błędu przewidywania RMSE dla dużych i małych zbiorów danych. Najlepsze wyniki otrzymywano dla zbiorów danych zawierających więcej obserwacji opisujących proces. Metoda drzew decyzyjnych okazała się być mniej precyzyjna, szczególnie dla małych zbiorów danych. Wyniki modelowania uzyskane za pomocą maszyn wektorów wspierających zasadniczo odzwierciedlają oczekiwane tendencje, jednak wartości wynikowe są mniej dokładne niż te uzyskane metodą sztucznych sieci neuronowych i drzew decyzyjnych. Ostatnia część planu badań dotyczy zaawansowanej analizy wyników i badania modelu w celu wielowymiarowej optymalizacji parametrów procesu. W wyniku przeprowadzonej analizy określono jakie wartości parametrów danej zmiennej wpływają na formowanie się wady w odlewie. Wskazano, że podwyższone wartości parametrów: *opóźnienie multiplikacji*, *temperatura wody miejskiej* i *czas dozowania stopu 2* oraz obniżone wartości parametrów: *czas przedmuchu* i *przepływ w obwodzie chłodzenia 14*, wpływają na osiągnięcie wyższych wartości przecieku (nieszczelności odlewu).

Rozprawa doktorska prezentuje dobrą ogólną wiedzę Autorki w dyscyplinie Inżynieria Mechaniczna. Autorka wykazała umiejętność samodzielnego prowadzenia pracy naukowej, a mianowicie: potrafiła dostrzec i sformułować problem w zakresie inżynierii mechanicznej, potrafiła przeprowadzić analizę literatury, potrafiła sformułować cele rozprawy, zastosowała odpowiednie narzędzia badawcze i poprawnie wyciągnęła wnioski.

Autorka wykazuje się bardzo dobrą znajomością problematyki sztucznej inteligencji i uczenia maszynowego potwierdzoną walidacją dla aplikacji przemysłowych.

W mojej opinii Autorka jest doświadczonym analitykiem danych dobrze posługującym się nowoczesnymi narzędziami programistycznymi, natomiast widoczny jest brak doświadczenia z zakresu analizowanego procesu odlewania. Niestety Autorka nie analizuje co jest przyczyną wykrytej nieszczelności odlewu, we wstępie wspomina tylko, że na ogół jest to wynik zbyt porowatej struktury związanej najczęściej z procesem krzepnięcia. Analiza wady danego odlewu może odpowiedzieć na pytanie czy jej pochodzenie jest wynikiem procesu wypełniania wnęki formy, procesu krzepnięcia czy np. stopnia zagazowania stopu.

Praca ma charakter interdyscyplinarny, w znacznej części obejmuje zagadnienia właściwe dla nauk ścisłych i informatyki, których metody są stosowane w obszarach inżynierii mechanicznej.

## Uwagi krytyczne i pytania do Autorki

1. Autorka nie przedstawiła informacji o akwizycji danych. W jaki sposób dane były rejestrowane, w jakiej formie, w jakim czasie, jakich procesów dotyczą?
2. W pracy brak jest informacji o zmiennych niezależnych. Podane są tylko ich wartości. Co do większości domyślam się czego dotyczą. Nie rozumiem czego dotyczą zmienne niezależne: *Koncentrat* rys.5.2.1.16, *Temperatura termoregulatora* 2.1, 2.2, 2.3, rys.5.2.1.39-41. Jaki sens ma zmienna *Dzienny numer wtrysku* rys.5.2.1.13?
3. Na str. 57 w opisie błędnie podano zakres normalizacji danych od 0,1 do 1. Powinno być od 0 do 1 co 0,1.
4. W tabelach od 5.1 do 5.56 przedstawiających wyniki normalizacji:
  - nieprawidłowo oznaczono przedziały, powinno być  $<0.0;0.1)$ ,  $<0.1;0.2)$  itd. lub  $<0.0;0.1>$ ,  $(0.1;0.2>$  itd.,
  - dlaczego wyodrębniono z przedziałów 1, a nie zrobiono tego samego dla 0?
  - w nazwach kolumn powinna występować liczba produktów i liczba obserwacji, a nie ilość,
  - dlaczego po przeprowadzeniu normalizacji min-max w tabelach 5.9, 5.13, 5.14, 5.21, 5.24, 5.27 i 5.54 dla „przedziału 1.0” są same zera?
  - nie wiadomo co oznaczają kolorowe obszary w ostatniej kolumnie.
5. Na rysunkach od rys.5.2.1.1 do 5.2.1.57 i 5.2.2.1 błędnie nazwano oś odciętych. To co przedstawiono na wykresach to nie jest *Liczba obserwacji [szt.]*. Można się tylko domyślać, że wszystkie obserwacje posortowano od najmniejszej do największej wartości osobno dla każdej zmiennej niezależnej i nadano im kolejne numery. W związku z tym oś odciętych to numer obserwacji.
6. Na str. 117 popełniono błąd w opisie, napisano: „Czwarty zbiór danych do badań zawierał w sobie wartości zbioru drugiego, czyli obserwacje, dla których zmienna zależna przyjmowała **wartości  $< 7,5$**  oraz taką samą ilość obserwacji pochodzących z górnego zakresu zbioru trzeciego”. Według założeń na str. 115: „Drugii zbiór danych do badań natomiast zawierał tylko obserwacje, dla których zmienna zależna przyjmowała **wartości  $\geq 7,5$** ”. To samo dotyczy opisu na str. 118.
7. Na str. 231 napisano: „Na wykresie (rys.5.7.7.4.) należy zwrócić uwagę, że wyższe wartości zmiennej niezależnej - opóźnienie multiplikacji [ms] sprzyjają powstawaniu wyższych wartości przecieku.”. Na rysunku 5.2.7.4 do którego powinno być odesłanie nie ma informacji o przecieku. Na osi odciętych jest numer przeprowadzonej optymalizacji, a na osi rzędnych opóźnienie multiplikacji. Czytelnik musi się domyśleć, że w przywołanej w podpisie rysunku tablicy 5.105 znajdzie te informacje. Jednak w opisach tej tablicy nie ma żadnych informacji o przecieku. Są natomiast kolorowe

wypełnienia dla wyników optymalizacji każdej zmiennej niezależnej. Można się domyślać, że kolory odpowiadają wartości przecieku, jednak nie jest to wyjaśnione w opisie.

### **Wniosek końcowy**

Wymienione uwagi nie zmieniają w niczym faktu, że Autorka zrealizowała postawiony cel rozprawy oraz rzetelnie wykonała postawione zadania.

Podjęta tematyka diagnozowania przyczyn powstawania wad wyrobów na podstawie modelowania opartego na dużych zbiorach danych jest niezwykle istotna i aktualna. Zaproponowana w rozprawie metodyka diagnozowania przyczyn wad odlewów może zostać wykorzystana w praktyce przemysłowej, co wskazuje, że praca ma charakter aplikacyjny.

Moim zdaniem Autorka pracy wykazała się umiejętnością samodzielnego prowadzenia prac badawczych i wiedzą w zakresie odpowiadającym stosownym przepisom odnośnie do wymogów stawianych rozprawom doktorskim w dyscyplinie Inżynieria Mechaniczna.

W podsumowaniu stwierdzam, że przedstawiona do oceny rozprawa pt. „Metodyka diagnozowania przyczyn powstawania wad wyrobów na podstawie zaawansowanego modelowania opartego na dużych zbiorach danych” spełnia wymagania określone w art. 187, ustawy z dnia 20 lipca 2018 r. – Prawo o szkolnictwie wyższym i nauce (Dz. U. 2018 poz. 1668 ze zm.) i wnioskuję o dopuszczenie Pani mgr inż. Alicji Okuniewskiej do publicznej obrony.

