

Białystok, 13 grudnia 2024 rok

Prof. dr hab. inż. Joanna Ejdyś  
Wydział Inżynierii Zarządzania  
Politechnika Białostocka

## RECENZJA

rozprawy doktorskiej mgr inż. Macieja Niemira  
pt. *Doskonalenie jakości elektronicznego katalogu z wykorzystaniem sztucznej inteligencji*

### Podstawa opracowania recenzji

Przedmiotem recenzji jest rozprawa doktorska mgr inż. Macieja Niemira zatytułowana *Doskonalenie jakości elektronicznego katalogu z wykorzystaniem sztucznej inteligencji*. Promotorem rozprawy jest dr hab. inż. Beata Mrugalska, prof. PP

Podstawę sporządzenia recenzji stanowiło pismo Dziekana Wydziału Inżynierii Zarządzania Politechniki Poznańskiej dr hab. inż. Marcina Butlewskiego, prof. PP z dnia 18 października 2024 roku (nr WIZ-08/310/2024) powołujące się na Uchwałę nr 7/2024-2028 Rady Dyscypliny Nauk o Zarządzaniu i Jakości Politechniki Poznańskiej z dnia 14 października 2024 roku.

Recenzja stanowi element postępowania o nadanie Panu mgr inż. Maciejowi Niemirowi stopnia naukowego doktora w dziedzinie nauk społecznych w dyscyplinie nauki o zarządzaniu i jakości.

### Charakterystyka ogólna rozprawy

Recenzowana rozprawa mieści się w szeroko rozumianej problematyce zarządzania jakością, a w szczególności jakością danych. Problematyka rozprawy została osadzona w teorii zarządzania wiedzą.

**Przedmiotem badań** Autora w recenzowanej rozprawie doktorskiej było opracowanie modelu automatycznej walidacji danych produktowych uwzględniającego sztuczną inteligencję, wspierającego doskonalenie jakości danych zawartych w katalogach elektronicznych.

W obliczu rosnącej ilości danych o produktach i braku możliwości manualnego zarządzania nimi, podjęty przez Autora rozprawy temat jest bardzo aktualny i ważny zarówno z teoretycznego, jak i praktycznego punktu widzenia. Uzasadniając wybór problematyki badawczej, mgr inż. Maciej Niemirowi wskazał następujące przesłanki:

- duże ryzyko popełniania błędów i odrzucania danych przez sztuczną inteligencję w procesie walidacji danych;
- dostrzegane możliwości wykorzystania sztucznej inteligencji jako tzw. co-pilota dla osób wprowadzających dane do systemów informatycznych;

- dostrzegany potencjał AI w zakresie wspierania doskonalenia jakości danych produktów umożliwiającą identyfikację potencjalnych nieprawidłowości i sugerowanie poprawek;
- potrzebę kompleksowego podejścia do automatycznej walidacji danych produktowych w elektronicznych katalogach, które łączyłyby metody regułowe z technikami bazującymi na sztucznej inteligencji

Rozprawa doktorska składa się z 240 stron tekstu zasadniczego. Struktura pracy jest bardzo rozbudowana i obejmuje osiem rozdziałów zatytułowanych kolejno: 1. Wprowadzenie; 2. Przegląd literatury i podstaw teoretycznych; 3. Eksploracja standardów i wytycznych w kontekście podstawowych danych; 4. Analiza problemów jakościowych w danych produktowych dostępnych online – badania własne; 5. Analiza problemu jakości danych produktowych; 6. Opracowanie modelu walidacji danych produktowych; 7. Weryfikacja koncepcji modelu w podziale na 3 warstwy; 8. Podsumowanie i kierunki dalszych badań. Integralnymi elementami pracy są również: bibliografia, spis rysunków, spis tabel oraz abstrakt.

W strukturze pracy wyróżnione zostały dwa rozdziały o charakterze teoriopoznawczym (rozdział 1 i 2), rozdziały prezentujące wyniki badań własnych Autora (rozdział 3, 4, 5), rozdział koncepcyjny (rozdział 6) oraz rozdział weryfikujący opracowany model teoretyczny (rozdział 7).

Bibliografia wykorzystana na potrzeby rozprawy doktorskiej liczy 233 pozycje, z których zdecydowana większość to pozycje zagraniczne. Literatura jest zgodna z tematyką pracy, aktualna i wyczerpująca.

### **Problem badawczy, pytania badawcze, cele, metody badawcze**

Przedmiotem badań było opracowanie modelu automatycznej walidacji danych produktowych uwzględniającego sztuczną inteligencję, wspierającego doskonalenie jakości danych zawartych w katalogach elektronicznych.

W oparciu o wyniki przeprowadzonych studiów literaturowych, Autor zidentyfikował lukę poznawczą w zakresie potrzeby kompleksowego podejścia do automatycznej walidacji danych produktowych w elektronicznych katalogach, które łączyłyby metody regułowe z technikami bazującymi na sztucznej inteligencji.

Główny problem badawczy został sformułowany w formie pytania: Jak można udoskonalić jakość danych produktowych bazując na modelu walidacji, który wykorzysta potencjał sztucznej inteligencji? Ponadto uszczegóławiając problem badawczy, Autor sformułował następujące pytania badawcze:

1. Jakie standardy i praktyki są stosowane w sektorze e-commerce w kontekście procesu wprowadzania danych o produktach?
2. Jakie problemy jakościowe można znaleźć w danych produktowych w elektronicznych katalogach?
3. Jaka struktura modelu walidacji danych produktowych zapewni skorzystanie z metod sztucznej inteligencji w elektronicznych katalogach?
4. Jakie algorytmy walidacyjne mogą pozytywnie wpłynąć na doskonalenie jakości danych produktów w elektronicznych katalogach?

**Celem rozprawy doktorskiej** było udoskonalenie jakości danych produktowych poprzez opracowanie i adaptację modelu walidacji, który uwzględnił możliwość korzystania z nowoczesnych rozwiązań z zakresu sztucznej inteligencji. Cel rozprawy charakteryzuje się oryginalnością.

Dość niefortunnie został sformułowany przedmiot badań, który co do treści pokrywa się z celem rozprawy doktorskiej i de facto stanowi cel badań: opracowanie modelu automatycznej walidacji danych produktowych uwzględniającego sztuczną inteligencję, wspierającego doskonalenie jakości danych zawartych w katalogach elektronicznych.

Metodyka badawcza obejmowała 5 zadań badawczych:

Zadanie 1. Analiza literatury przedmiotu.

Zadanie 2. Eksploracja standardów i analiza sektora e-commerce w kontekście przyjętych norm i praktyk w odniesieniu do procesu wprowadzania danych o produktach.

Zadanie 3. Identyfikacja wyzwań związanych z walidacją danych na podstawie analizy bazy danych, będącej przedmiotem badań.

Zadanie 4. Opracowanie struktury wielopoziomowego modelu walidacji.

Zadanie 5. Selekcja, adaptacja i ocena algorytmów walidacyjnych do zastosowania w zaproponowanym modelu walidacji.

Metodyka badań opisana w podrozdziale 1.7 została zaprezentowana w postaci schematu realizacji pracy na rysunku 3. Wydaje się, że bardziej prawidłowym podejściem mogłoby być opracowanie metodyki uwzględniającej wcześniej zdefiniowane zadania badawcze.

### **Ocena merytoryczna rozprawy**

Uwzględniając przyjętą przez Autora strukturę rozprawy rozdział 1 stanowi opis metodycznych elementów pracy, które zostały już omówione.

Rozdział 2 został poświęcony problematyce jakości. Autor w pierwszej kolejności dokonał przeglądu definicji terminu jakość (podrozdział 2.1), następnie odniósł pojęcia jakości do danych (podrozdział 2.2) i omówił wymiary jakości danych (podrozdział 2.3), ostatecznie zawężając rozważania do danych produktowych (podrozdział 2.4). Wartościowym elementem tej części rozprawy jest przegląd wymiarów jakości danych. Podrozdział 2.4 stanowi pogłębione uzasadnienie ważności podjętej przez Autora tematyki pracy w kontekście wpływu jakości danych (danych niskiej jakości) na funkcjonowanie organizacji i gospodarki. W kolejnych podrozdziałach odnoszących się do danych o produktach Autor omówił:

- metody pozyskiwania danych o produktach (podrozdział 2.5);
- podstawowe dane produktowe opisujące i identyfikujące produkt (podrozdział 2.6);
- główne źródła błędów w danych produktowych (podrozdział 2.7);
- systemy zarządzania danymi produktowymi i wyzwania w kontekście problematyki jakości (podrozdział 2.8);
- taksonomię sposobów oczyszczania danych produktowych (podrozdział 2.9);
- rolę walidacji danych produktowych (podrozdział 2.10);

- metody walidacji danych (podrozdział 2.11);
- koncepcje i rozwiązania walidacji danych (podrozdział 2.12).

Rozdział 2 stanowi ważny element pracy, porządkujący zawarte w nim treści. Stanowi wprowadzenie do dalszych rozważań. Autor wykorzystał bogatą literaturę krajową i zagraniczną. Tekst czyta się bardzo dobrze i zrozumiale.

W rozdziale 3 rozprawy Autor dokonał przeglądu standardów i wytycznych dotyczących danych produktowych. Na podstawie przeprowadzonych studiów literaturowych, katalog atrybutów produktów został zawężony do sześciu atrybutów: globalny unikalny identyfikator produktu (Global unique product ID), unikalna nazwa opisująca produkt (Unique name that describes the product). Analizowano, jak nazwy produktów są używane do ich identyfikacji i jak różnią się one w zależności od platformy, biorąc pod uwagę ich unikalność i zrozumiałość; nazwa marki (Brand name); zdjęcie produktu (Product image); zawartość netto (Net content) oraz klasyfikacja produktu (Product Classification). Opis przyjętej metodyki badań i sposobu prezentacji wyników jest poprawny i wyczerpujący. Treści uporządkowane w logiczny sposób pozwoliły Autorowi na wyciągnięcie wniosków odnoszące się do zróżnicowanego podejścia do kluczowych atrybutów produktowych w różnych systemach e-commerce. Wnioski z przeprowadzonych badań uzasadniły celowość opracowania mechanizmu walidacji.

W rozdziale 4 rozprawy Autor mgr inż. Maciej Niemir zaprezentował wyniki szczegółowej analizy danych produktowych, korzystając ze źródeł pochodzących z różnych baz danych, sieci, społeczności oraz platform cyfrowych dostępnych on-line. W szczególności obiektem zainteresowania Autora były następujące aspekty danych produktowych: poprawność numerów GTIN, nazwa produktu w wybranych sklepach elektronicznych, nazwa produktu/oferty na platformie zakupowej, nazwy produktów w różnych katalogach produktowych, zgodność nazwy ze zdjęciem produktu. Przeprowadzone badania pozwoliły Autorowi na sformułowanie wniosku odnoszącego się do złożoności problemów związanych z jakością danych produktowych, wynikających z braku standaryzacji i ogólnodostępnych baz referencyjnych.

Rozdział 5 został poświęcony analizie problemu jakości danych produktowych. Przeprowadzona analiza błędów danych została pogłębiona o identyfikację przyczyn powstawania tych błędów, co ma istotne znaczenie w kontekście ich eliminacji. Wartościowym elementem tej części pracy jest zaproponowana klasyfikacja problemów odnoszących się do danych produktowych. W odniesieniu do 45 głównych kategorii problemów, Autor dokonał oceny możliwości utworzenia algorytmów w celu eliminacji wykrytych problemów (tabela 26). Zidentyfikowane problemy zostały przyporządkowane do jednej z grup: problemy o niskiej złożoności algorytmicznej oraz problemy o wysokiej złożoności algorytmicznej.

Treści zawarte w rozdziałach 3-5 stanowiły dla Autora z jednej strony uzasadnienie potrzeby opracowania modelu walidacji, z drugiej były istotnym źródłem danych i informacji wejściowych do zaproponowanego w rozdziale 6 modelu walidacji.

Rozdział 6 został poświęcony opracowaniu modelu walidacji danych produktowych, którego celem było zaproponowanie rozwiązań umożliwiających skuteczną kontrolę danych produktowych wprowadzanych do bazy przy minimalnym udziale osób weryfikujących. Działania normalizacyjno-walidacyjne w modelu uwzględniały następujące elementy: normalizację danych, walidację danych oraz

sugestię poprawy jakości. Zaproponowany model przedstawiony na rysunku 16 jest logiczny i precyzyjnie opisany. Zaproponowany model oferuje w porównaniu do innych modeli opisach w literaturze szczegółową segmentację działań walidacyjnych, wprowadza techniki sztucznej inteligencji do sugestii poprawy jakości danych oraz angażuje użytkowników wprowadzających dane w proces poprawy, co jest zgodne z koncepcją ciągłego doskonalenia. Do oceny skuteczności zaproponowanego modelu Autor zastosował metodę przeliczenia potencjału poprawy danych. Istotnym elementem oceny skuteczności była suma wykrytych problemów jakościowych wszystkich zaproponowanych walidatorów dla każdego poziomu walidacji.

Rozdział 7 rozprawy obejmuje wyniki weryfikacji koncepcji modelu w podziale na 3 warstwy: Poziom 1: Normalizacja danych; Poziom 2: Walidacja techniczno – logiczna; Poziom 3: Sugestie poprawy jakości. Ważnym elementem tej części pracy jest ocena wpływu walidatorów na dane w zaproponowanym modelu.

Do najważniejszych osiągnięć naukowych przeprowadzonego postępowania badawczego należy zaliczyć:

- rzetelnie przeprowadzone studia literaturowe dotyczące różnych aspektów jakości danych produktowych, w tym metod walidacji oraz problemów występujących w zarządzaniu danymi w katalogach elektronicznych;
- przegląd obowiązujących standardów oraz praktyk stosowanych w sektorze e-commerce w zakresie wprowadzania danych produktowych;
- zidentyfikowanie i sklasyfikowanie 45 specyficznych problemów jakościowych, związanych z różnymi atrybutami, takimi jak nazwa, marka, klasyfikacja produktu, zdjęcia produktów czy zawartość netto;
- opracowanie struktury wielopoziomowego modelu walidacji danych produktowych wraz z opisem elementów modelu;
- graficzne przedstawienie i opis procesu ewaluacji walidatorów;
- zrealizowanie procesu weryfikacji koncepcji modelu wielopoziomowej walidacji danych produktowych na przykładzie wybranych walidatorów;
- opracowanie zestawu walidatorów dla trzech poziomów: normalizacja danych, walidacja danych oraz sugestie poprawy jakości;
- zastosowanie wielu algorytmów sztucznej inteligencji oraz narzędzi wspomagających analizę tekstu i obrazu.

Podczas publicznej obrony proszę o udzielenie odpowiedzi na poniższe pytania:

1. Do jakich innych obszarów zastosowań mógłby być użyteczny opracowany model walidacji?
2. Jaki jest potencjał komercjalizacyjny opracowanego modelu walidacji?

## Ocena rozprawy pod względem techniczno-redakcyjnym

Pod względem techniczno-redakcyjnym praca została napisana poprawnie i bardzo starannie. Rysunki oraz tabele zostały wykonane estetycznie i przejrzyste. Praca wyróżnia się pod względem językowo-redakcyjnym. Autor posiada umiejętność bardzo logicznego i ustrukturyzowanego podejścia do prezentowania treści. Jednocześnie słownictwo jakim operuje pozwala na zrozumienie treści przez osoby nie będące ekspertami w dziedzinie danych produktowych.

## Wniosek końcowy

Za istotny, a tym samym oryginalny walor naukowy recenzowanej rozprawy doktorskiej należy uznać podjętą przez Autora udaną próbę zbudowania wielopoziomowego modelu walidacji danych produktowych wraz z opisem elementów modelu oraz jego weryfikacją.

Autor rozprawy podjął się rozwiązania oryginalnego problemu naukowego, dotychczas nierozpoznanego w literaturze przedmiotu, związanego z modelowym ujęciem walidacji danych produktowych. Wartością dodaną w stosunku innych modeli walidacji jest zastosowanie algorytmów sztucznej inteligencji na potrzeby analizy tekstu i obrazu. Uzyskane wyniki mają istotny walor aplikacyjny.

Recenzowana rozprawa Pana mgr inż. Macieja Niemira spełnia wymagania stawiane rozprawom doktorskim zgodnie z obowiązującymi w tym zakresie przepisami prawnymi, w zakresie wykazania się przez Kandydata do stopnia doktora ogólną wiedzą teoretyczną w dyscyplinie naukowej nauki o zarządzaniu i jakości oraz umiejętnością samodzielnego prowadzenia pracy naukowej. Autor wykazał się dociekliwością i rzetelnością w procesie planowania, realizowania i analizowania wyników badań naukowych. Recenzowana praca stanowi oryginalne rozwiązanie problemu naukowego.

Uwzględniając fakt, że Pan mgr inż. Maciej Niemir podjął się aktualnej i ważnej problematyki naukowej oraz zrealizował przyjęte cele rozprawy, wnoszę o dopuszczenie do kolejnych etapów postępowania o nadanie stopnia naukowego doktora nauk społecznych w dyscyplinie nauki o zarządzaniu i jakości.

Jednocześnie, uwzględniając rzetelność zaprojektowanego i zrealizowanego postępowania badawczego oraz oryginalność problemu naukowego, wnioskuję o wyróżnienie rozprawy doktorskiej Pana mgr inż. Macieja Niemira.



prof. dr hab. inż. Joanna Ejdyś