

Poznań, 23 grudnia 2025 r.

Prof. UAM dr hab. Jacek Marciniak

Zakład Sztucznej Inteligencji

Wydział Matematyki i Informatyki

Uniwersytet im. Adama Mickiewicza w Poznaniu

ul. Uniwersytetu Poznańskiego 4

61-614 Poznań

Recenzja rozprawy doktorskiej

pt. Nowe metody uwzględnienia niepewności i nieprecyzyjności danych w
klasyfikacji

autorstwa mgr inż. Dawida Kosiora

przygotowanej pod kierunkiem promotora dr hab. Barbary Pękali, prof. UR

Recenzję rozprawy doktorskiej mgr inż. Dawida Kosiora, ubiegającego się o stopień doktora w dziedzinie nauk inżynieryjno-technicznych, w dyscyplinie informatyka techniczna i telekomunikacja, sporządzono na podstawie powołania przez Radę Naukową Instytutu Badań Systemowych Polskiej Akademii Nauk (pismo z dnia 10.10.2025 r.). Recenzję przygotowano zgodnie z przepisami ustawy z dnia 20 lipca 2018 r. – Prawo o szkolnictwie wyższym i nauce.

1. Tematyka i cele rozprawy

1.1. Tematyka badawcza i jej aktualność

Tematyka rozprawy doktorskiej dotyczy zagadnień związanych z uwzględnianiem niepewności i nieprecyzyjności danych w procesie klasyfikacji, ze szczególnym wykorzystaniem aparatu teorii przedziałowych zbiorów rozmytych. Autor koncentruje się na opracowaniu oraz analizie miar zawierania, podobieństwa i entropii dla danych o charakterze przedziałowym, a następnie na ich zastosowaniu w zadaniach klasyfikacyjnych, istotnych z punktu widzenia praktycznych problemów diagnostyki medycznej.

Podjęta tematyka wpisuje się w aktualny nurt badań nad metodami analizy danych obarczonych niepewnością, które nie mogą być w sposób satysfakcjonujący modelowane za pomocą klasycznych reprezentacji punktowych. W szczególności dotyczy to danych pochodzących z obszaru medycyny, gdzie nieprecyzyjność pomiarów, brak pełnej informacji lub subiektywny charakter ocen eksperckich stanowią istotne wyzwanie dla standardowych metod uczenia maszynowego.

Na uwagę zasługuje fakt, że autor nie ogranicza się do zastosowania istniejących rozwiązań, lecz podejmuje próbę systematycznego rozwinięcia aparatu teoretycznego związanego z przedziałowymi zbiorami rozmytymi, w szczególności poprzez wprowadzenie miar interpretowanych w ujęciu klasycznym, optymistycznym (możliwości) oraz pesymistycznym (konieczności). Takie podejście pozwala na wieloaspektową analizę niepewności i jej wpływu na jakość procesów decyzyjnych.

Z punktu widzenia dyscypliny informatyka techniczna i telekomunikacja tematyka rozprawy została dobrana trafnie. Łączy ona elementy badań teoretycznych z ich praktyczną weryfikacją na rzeczywistych danych, odpowiadając jednocześnie na aktualne potrzeby związane z projektowaniem systemów klasyfikacyjnych działających w warunkach niepełnej lub nieprecyzyjnej informacji.

1.2. Cele i teza rozprawy

Autor rozprawy sformułował główny cel pracy jako opracowanie metod i algorytmów umożliwiających uwzględnienie niepewności i nieprecyzyjności danych w procesie klasyfikacji, ze szczególnym naciskiem na rozwiązania oparte na przedziałowych zbiorach rozmytych oraz ich zastosowanie w algorytmach klasyfikacyjnych. Cel ten ma charakter zarówno teoretyczny, jak i aplikacyjny, przy czym zasadniczy ciężar pracy położono na rozwój aparatu metodologicznego.

Główna teza rozprawy zakłada, że uwzględnienie niepewności i nieprecyzyjności danych może prowadzić do istotnej poprawy jakości działania klasyfikatorów opartych na algorytmie k-najbliższych sąsiadów, zarówno w warunkach danych niekompletnych, jak i w procesie optymalizacji zbioru cech. Teza ta ma charakter weryfikowalny i została sprawdzona w pracy poprzez analizę zaproponowanych miar oraz ich eksperymentalne zastosowanie w rzeczywistych zadaniach klasyfikacyjnych.

Cele szczegółowe rozprawy obejmują w szczególności: opracowanie nowych miar zawierania, podobieństwa i entropii dla przedziałowych zbiorów rozmytych, zaprojektowanie algorytmów klasyfikacji oraz metod wspomagających analizę danych (takich jak imputacja braków danych i selekcja cech), a także ocenę ich skuteczności na rzeczywistych zbiorach danych medycznych. Należy podkreślić, że cele te są ze sobą logicznie powiązane, a część aplikacyjna pracy pełni rolę weryfikacyjną wobec zaproponowanych rozwiązań teoretycznych.

Sformułowane cele oraz teza rozprawy są jasne, spójne i pozostają w zgodzie z zakresem oraz strukturą przeprowadzonych badań, a ich realizacja została w pracy konsekwentnie przeprowadzona.

2. Struktura i zawartość rozprawy

Rozprawa doktorska została podzielona na siedem rozdziałów, których układ jest logiczny i podporządkowany realizacji sformułowanych celów badawczych. Struktura pracy prowadzi czytelnika od wprowadzenia teoretycznego, poprzez rozwój aparatu metodologicznego, aż do jego weryfikacji w zadaniach aplikacyjnych oraz prezentacji wyników badań.

W rozdziałach 2 i 3 autor przedstawia podstawy teoretyczne pracy. Rozdział 2 wprowadza pojęcia związane z teorią zbiorów rozmytych, ze szczególnym uwzględnieniem przedziałowych zbiorów rozmytych, relacji porządku oraz operatorów agregacji, które stanowią formalną bazę dalszych rozważań. Rozdział 3 ma kluczowe znaczenie dla rozprawy, gdyż zawiera definicje oraz własności zaproponowanych miar zawierania, podobieństwa i entropii, analizowanych w ujęciu klasycznym, optymistycznym (możliwości) oraz pesymistycznym (konieczności).

Rozdział 4 ma charakter przeglądowy i poświęcony jest omówieniu wybranych metod klasyfikacji oraz miar oceny ich jakości. Pełni on rolę uzupełniającą, porządkującą wiedzę niezbędną do właściwego osadzenia zaproponowanych w dalszej części pracy rozwiązań na tle istniejących metod.

Część aplikacyjna rozprawy została przedstawiona w rozdziale 5, który zawiera opis autorskich metod wspierających klasyfikację danych nieprecyzyjnych, w szczególności uogólnionego klasyfikatora IV-kNN, metod imputacji brakujących danych oraz algorytmu selekcji cech. Zaprezentowane rozwiązania zostały zweryfikowane eksperymentalnie na rzeczywistych zbiorach danych medycznych.

W rozdziale 6 autor przedstawia dodatkowe kierunki badań i zastosowań zaproponowanego aparatu metodologicznego, w tym system wspomagający ocenę ryzyka chorób sercowo-naczyniowych oraz implementację biblioteki programistycznej wspierającej przedziałowe wnioskowanie rozmyte. Rozdział ten stanowi rozszerzenie głównych rozważań i pokazuje możliwość dalszego wykorzystania opracowanych metod.

Całość rozprawy zamyka rozdział 7, zawierający podsumowanie uzyskanych wyników oraz wskazanie potencjalnych kierunków dalszych badań. Przyjęta struktura pracy jest spójna, a proporcje pomiędzy częścią teoretyczną i aplikacyjną należy uznać za właściwe dla rozprawy doktorskiej w dyscyplinie informatyka techniczna i telekomunikacja.

Dobór i wykorzystanie literatury przedmiotu należy uznać za poprawne i adekwatne do podejmowanej problematyki. Autor odwołuje się zarówno do klasycznych prac z zakresu teorii zbiorów rozmytych i ich uogólnień, jak i do nowszych publikacji dotyczących analizy danych niepewnych oraz metod klasyfikacji, co pozwala właściwie osadzić uzyskane wyniki na tle istniejącego stanu wiedzy.

3. Ocena merytoryczna i oryginalne osiągnięcia rozprawy

Rozprawa doktorska mgr inż. Dawida Kosiora ma charakter metodologiczno-aplikacyjny, przy czym jej zasadniczy wkład naukowy koncentruje się na rozwoju aparatu teoretycznego związanego z przedziałowymi zbiorami rozmytymi oraz miarami definiowanymi na takich zbiorach. Część aplikacyjna pracy pełni w głównej mierze funkcję weryfikacyjną wobec zaproponowanych rozwiązań teoretycznych i stanowi ich praktyczną ilustrację.

Kluczowym osiągnięciem rozprawy jest opracowanie i analiza nowych miar zawierania, podobieństwa oraz entropii dla przedziałowych zbiorów rozmytych, rozpatrywanych w trzech ujęciach: klasycznym, optymistycznym (możliwości) oraz pesymistycznym (konieczności). Zaproponowane miary zostały formalnie zdefiniowane, a ich własności zostały wykazane w postaci twierdzeń wraz z dowodami.

W rozdziale 3.1 autor przedstawia miary zawierania się przedziałów i przedziałowych zbiorów rozmytych, w tym wskaźnik pierwszeństwa w ujęciu klasycznym, możliwości oraz konieczności. Część zaprezentowanych konstrukcji stanowi rozwinięcie i uogólnienie wcześniejszych prac autora, jednak w rozprawie zostały one istotnie rozszerzone o nowe warianty oraz pełną analizę teoretyczną, w tym uzupełnione dowody własności, co wykracza poza zakres wcześniejszych publikacji.

Rozdział 3.2 poświęcony jest miarom podobieństwa dla przedziałowych zbiorów rozmytych. Autor proponuje nowe metody konstrukcji miar podobieństwa, w szczególności oparte na relacjach możliwości i konieczności, a także przedstawia nową metodę konstrukcji miary podobieństwa w ujęciu klasycznym, wykorzystującą wskaźnik pierwszeństwa. Wprowadzone rozwiązania mają charakter oryginalny i stanowią istotny wkład w rozwój metod porównywania danych nieprecyzyjnych.

Istotnym elementem rozprawy są również wyniki zaprezentowane w rozdziale 3.3, dotyczące entropii przedziałowych zbiorów rozmytych. Autor analizuje istniejące podejścia do definiowania entropii, a następnie proponuje nowe miary entropii w ujęciu klasycznym, optymistycznym oraz pesymistycznym. W pracy przedstawiono dwie metody konstrukcji przedziałowych miar entropii w każdym z tych ujęć, poparte odpowiednimi twierdzeniami i dowodami. W szczególności należy podkreślić, że zaproponowane miary entropii możliwości i konieczności stanowią oryginalny rezultat pracy.

Opracowany aparat teoretyczny został następnie wykorzystany w części aplikacyjnej rozprawy. W rozdziale 5 autor proponuje uogólniony klasyfikator IV-kNN, stanowiący rozszerzenie klasycznego algorytmu k-najbliższych sąsiadów na dane o charakterze przedziałowym. Nowość zaproponowanego rozwiązania polega na zastosowaniu autorskich miar podobieństwa oraz relacji porządku do porównywania obiektów i podejmowania decyzji klasyfikacyjnych, z jednoczesnym zachowaniem informacji o poziomie niepewności na kolejnych etapach algorytmu.

Zaproponowane metody zostały zweryfikowane na rzeczywistych zbiorach danych medycznych, w szczególności w zadaniu diagnostyki raka piersi. Przeprowadzone eksperymenty wskazują, że uwzględnienie niepewności w postaci danych przedziałowych oraz zastosowanie miar w ujęciu możliwości prowadzi do poprawy jakości klasyfikacji w porównaniu z rozmytymi wariantami klasycznego algorytmu k-NN. Uzyskane wyniki, choć ilościowo umiarkowane, mają istotne znaczenie w kontekście zastosowań medycznych, gdzie nawet niewielka poprawa jakości klasyfikacji może mieć praktyczne konsekwencje decyzyjne.

Kolejnymi istotnymi osiągnięciami pracy są: opracowanie metody imputacji brakujących danych opartej na przedziałowej mierze entropii oraz zaproponowanie algorytmu selekcji cech IVRelief, będącego rozszerzeniem klasycznego algorytmu Relief na dane przedziałowe. W obu przypadkach autor wykazał, że wykorzystanie zaproponowanych miar pozwala na poprawę jakości klasyfikacji, co potwierdzają przeprowadzone eksperymenty.

Na uwagę zasługuje również przedstawiony w pracy system prewencji depresji, oparty na autorskim teście diagnostycznym oraz zaproponowanych metodach klasyfikacji. Rozwiązanie to stanowi interesujący przykład zastosowania opracowanego aparatu metodologicznego do złożonego problemu praktycznego i pokazuje potencjał dalszych zastosowań zaproponowanych metod.

Podsumowując, rozprawa zawiera spójny i oryginalny wkład naukowy w zakresie metod analizy danych nieprecyzyjnych z wykorzystaniem przedziałowych zbiorów rozmytych. Zaproponowane rozwiązania teoretyczne są poprawne formalnie, a ich praktyczna przydatność została potwierdzona w badaniach eksperymentalnych. Zakres i charakter uzyskanych wyników odpowiadają wymaganiom stawianym rozprawom doktorskim w dyscyplinie informatyka techniczna i telekomunikacja.

4. Uwagi krytyczne i pytania

Rozprawa doktorska mgr inż. Dawida Kosiora została przygotowana starannie i charakteryzuje się wysokim poziomem merytorycznym. Poniższe uwagi krytyczne oraz pytania nie podważają ogólnej wartości pracy, lecz mają na celu doprecyzowanie wybranych aspektów metodologicznych oraz wskazanie potencjalnych kierunków dalszych badań.

W pierwszej kolejności należy zwrócić uwagę na drobne uchybienia redakcyjne, takie jak pojedyncze literówki występujące w tekście rozprawy. Mają one jednak charakter incydentalny i nie wpływają na odbiór ani zrozumiałość pracy.

Pewnej ostrożności terminologicznej wymaga fragment dotyczący charakterystyki krzywej ROC, gdzie pojawia się niejednoznaczność w użyciu pojęć związanych z miarami jakości klasyfikacji. W szczególności relacja pomiędzy swoistością (specificity) a wskaźnikiem fałszywych pozytywów (false positive rate) mogłaby zostać w tym miejscu doprecyzowana, aby uniknąć możliwych nieporozumień interpretacyjnych.

W części aplikacyjnej rozprawy autor koncentruje się na algorytmie k-najbliższych sąsiadów, uzasadniając jego wybór zarówno właściwościami metody, jak i wynikami wcześniejszych badań. W kontekście aktualnego stanu badań warto jednak zapytać, czy autor rozważał odniesienie zaproponowanych rozwiązań do innych metod klasyfikacji, takich jak regresja logistyczna, maszyny wektorów nośnych lub współczesne metody oparte na sieciach neuronowych, które stanowią obecnie standardowe punkty odniesienia w zadaniach klasyfikacji danych, lub o uzasadnienie ich pominięcia.

Podobna uwaga dotyczy metod imputacji brakujących danych oraz selekcji cech. W rozprawie przedstawiono autorskie rozwiązania oparte na przedziałowych miarach entropii oraz podobieństwa, jednak interesujące byłoby odniesienie uzyskanych wyników do prostszych metod bazowych, takich jak imputacja średnią lub medianą, bądź do alternatywnych algorytmów selekcji cech stosowanych w analizie danych niepewnych. Tego rodzaju porównania mogłyby dodatkowo wzmocnić interpretację uzyskanych rezultatów.

W rozdziale poświęconym systemowi prewencji depresji zaprezentowano autorski test diagnostyczny SenDD oraz jego wykorzystanie w procesie klasyfikacji. Rozwiązanie to stanowi interesujący przykład zastosowania opracowanego aparatu metodologicznego. Warto jednak zapytać, w jakim zakresie test ten był konstruowany z uwzględnieniem standardowych procedur stosowanych w psychometrii, takich jak ocena trafności i rzetelności narzędzia, oraz czy autor planuje dalsze badania w tym kierunku przed ewentualnym wykorzystaniem testu w praktyce diagnostycznej.

Odnosząc się do przeprowadzonych eksperymentów klasyfikacyjnych, można również zapytać o stabilność zaproponowanych metod w przypadku zwiększenia liczby danych uczących lub zastosowania ich do zbiorów danych pochodzących z innych źródeł, charakteryzujących się odmiennym poziomem niepewności. Kwestia ta może mieć znaczenie dla oceny możliwości uogólnienia zaproponowanych rozwiązań.

Przedstawione uwagi i pytania mają charakter uzupełniający i nie wpływają na ogólną, pozytywną ocenę rozprawy. Stanowią one raczej punkt wyjścia do dalszej dyskusji naukowej podczas publicznej obrony pracy.

5. Konkluzja recenzji

Przedłożona do oceny rozprawa doktorska mgr inż. Dawida Kosiora stanowi spójne i wartościowe opracowanie naukowe, podejmujące istotne zagadnienia związane z analizą danych nieprecyzyjnych z wykorzystaniem aparatu przedziałowych zbiorów rozmytych. Praca ma wyraźnie określony cel badawczy, poprawnie sformułowaną tezę oraz logiczną strukturę, a przedstawione wyniki świadczą o dojrzałości naukowej autora. Zaprezentowane w rozprawie rozwiązania teoretyczne, w szczególności w zakresie miar zawierania, podobieństwa i entropii dla przedziałowych zbiorów rozmytych, stanowią oryginalny wkład autora w rozwój metod analizy danych obarczonych niepewnością. Część aplikacyjna pracy potwierdza praktyczną

przydatność zaproponowanego aparatu metodologicznego, zwłaszcza w kontekście zadań klasyfikacyjnych w diagnostyce medycznej.

Przedstawione w rozprawie wyniki potwierdzają, że autor rozwiązał postawione problemy badawcze, wykazał się znajomością zagadnień teoretycznych właściwych dla dyscypliny informatyka techniczna i telekomunikacja oraz zdolnością samodzielnego prowadzenia pracy naukowej. Uwagi krytyczne sformułowane w niniejszej recenzji mają charakter uzupełniający i nie wpływają na ogólną, pozytywną ocenę rozprawy. Dotyczą one głównie doprecyzowania wybranych aspektów terminologicznych oraz rozszerzenia dyskusji porównawczej, co może stanowić punkt wyjścia do dalszych badań i dyskusji naukowej podczas publicznej obrony pracy.

Na podstawie przeprowadzonej analizy stwierdzam, że rozprawa doktorska mgr inż. Dawida Kosiora spełnia wymagania sformułowane w Ustawie z dnia 20 lipca 2018 r. – Prawo o szkolnictwie wyższym i nauce (Dz. U. 2018, poz. 1668, z późniejszymi zmianami). W związku z powyższym przedstawioną rozprawę oceniam pozytywnie i wnoszę o dopuszczenie mgr inż. Dawida Kosiora do dalszych etapów postępowania o nadanie stopnia naukowego doktora.



prof. UAM dr hab. Jacek Marciniak
Zakład Sztucznej Inteligencji, WMI, UAM