

Warszawa, 28.02.2026 r.

prof. dr hab. inż. Robert Nowak
Instytut Informatyki
Wydział Elektroniki i Technik Informacyjnych
Politechnika Warszawska
ul. Nowowiejska 15/19
00-665 Warszawa
robert.nowak@pw.edu.pl

**Recenzja rozprawy doktorskiej mgr. inż. Cezarego Maszczyka
zatytułowanej „Zastosowanie głębokiej eksploracji danych, w tym
dyskretnego głębokiego uczenia, w indukcji reguł logicznych zawierających
złożone warunki elementarne”**

Recenzja powstała na prośbę Przewodniczącego Rady Dyscypliny Informatyka Techniczna i Telekomunikacja Politechniki Śląskiej, zgodnie z uchwałą nr 167/2025 Rady Dyscypliny Informatyka Techniczna i Telekomunikacja Politechniki Śląskiej, z dnia 28.10.2025, na podstawie:

- rozprawy doktorskiej liczącej 236 stron, z 2025 r,
- dorobku naukowego Doktoranta uwzględnionego w bazach Web of Science, Scopus i Google Scholar,
- rozwiązań udostępnionych w Internecie.

1 Główne wyniki rozprawy

Przedstawiona rozprawa doktorska przedstawia trzy nowe algorytmy indukcji reguł oparte na strategii sekwencyjnego pokrywania. Algorytmy rozpatrują złożone warunki w przesłankach, co zwiększa złożoność obliczeniową, natomiast może prowadzić do lepszych i krótszych reguł.

- Metoda *ComplexConditions* dodaje do typowych warunków (gdzie atrybut jest porównywany ze stałą) warunki relacji atrybutów (np. $x_1 > x_2$) lub wewnętrznych alternatyw (np. $\text{kolor} \in \{\text{czerwony, zielony, żółty}\}$).
- Metoda *MofNRules* rozwija pierwszą wprowadzając warunek M-z-N, (np. dokładnie 2-z- $\{v_1, v_2, v_3, v_4, v_5\}$).
- Metoda *DeepRules* skupia się na indukcji reguł z zagnieżdżonymi wyrażeniami (koniunkcje, alternatywy).

Przedstawione metody zostały zaimplementowane i przetestowane na 74 publicznie dostępnych zbiorach danych tabelarycznych, pochodzących z UC Irvine Machine Learning Repository, Kabbie oraz OpenML.

Rozpatrywane są problemy klasyfikacji i regresji, dodatkowo wydzielono problemy analizy przeżycia.

Przedstawione wyniki zostały porównane z modelem generowanym przez algorytm *RuleKit* (klasyczny algorytm generowania reguł) oraz z drzewem decyzyjnym (Python, scikit-learn). Została przeprowadzona dyskusja.

Analizowano pesymistyczną złożoność obliczeniową przedstawionych metod.

Problem badawczy jest postawiony właściwie, jest on interesujący i istotny. Ma on charakter naukowy i ma znaczenie praktyczne. Zaproponowane rozwiązania są poprawne. Została dostarczona implementacja przedstawionych metod na platformie *RuleMiner* rozwijanej w Instytucie Sztucznej Inteligencji i Cyberbezpieczeństwa.

Rozprawa jest napisana w języku polskim, ma 10 rozdziałów. W rozdziałach 1-5 przedstawiono problem badawczy i podjęto próbę uzasadnienia podejścia opartego o reguły w porównaniu z modelami subsymbolicznymi, takimi jak sztuczne sieci neuronowe. Kandydat poprawnie analizował literaturę światową pokazując podobne rozwiązania. Nowe metody utworzone przez doktoranta są opisane w Rozdziale 6, na 51 stronach. Rozdział 7 i 8, opisują wyniki eksperymentów numerycznych na publicznie dostępnych zbiorach danych. Rozdział 9 pokazuje elementy implementacyjne, zaś rozdział 10 to podsumowanie. Rozprawa cytuje 199 innych prac.

Wkład autora w przedstawione rozwiązania jest wiodący, pomimo tego, że zostały one opublikowane wyłącznie jako wieloautorskie publikacje w czasopismach naukowych lub w materiałach konferencyjnych. Kandydat prezentuje odpowiednią wiedzę w dyscyplinie Informatyka Techniczna i Telekomunikacja.

2 Elementy krytyczne

Poniżej zamieszczam elementy rozprawy, które należy dodatkowo wyjaśnić. Przedstawione niedociągnięcia nie zmieniają pozytywnej konkluzji końcowej analizowanej rozprawy. Stanowią podstawę do dyskusji na publicznej obronie.

2.1 Eksperymenty

W rozprawie w sposób właściwy wskazano rozwiązania podobne i opisano fakt, że poszukiwanie lepszych przesłanek reguł w metodach symbolicznych jest problemem nad którym pracuje wiele zespołów od dłuższego czasu. Wskazano istniejące prace, które zbliżone metody, zarówno dla warunków relacji, warunków M-z-N i indukcji reguł z zagnieźdzonymi wyrażeniami. Z drugiej strony w badaniach numerycznych do porównania używany jest algorytm indukcji reguł pozbawiony omawianych rozszerzeń oraz drzewo decyzyjne. Nie wiadomo więc jaki jest wynik porównania nowych algorytmów i dostarczonej implementacji z podobnymi metodami opisanymi w literaturze.

W opisie własnych metod Kandydata są krótko omówione różnice w koncepcji dostarczonego rozwiązania, natomiast nie jest pokazana zmiana (lub brak zmiany) wyników analiz, w porównaniu z wynikami metod podobnych. Ten element wymaga dyskusji na obronie.

2.2 Wdrożenie

Dodanie modułu z działającym algorytmem do istniejącej platformy, opisane w rozdziale 9, jest istotnym elementem wdrożenia, ale nie wdrożeniem. Prosiłbym o podanie problemów, które były rozwiązane za pomocą nowych metod, albo problemów, których rozwiązanie jest planowane.

2.3 Złożoność obliczeniowa

Przedstawione analizy pesymistycznej złożoności obliczeniowej są one poprawne, bazują na szacowaniu wielkości przestrzeni rozpatrywanych warunków dla przesłanek. Man-kamentem pracy jest, moim zdaniem, zbyt mało uwagi przy analizie średniej złożoności obliczeniowej zaproponowanych metod. Średnia złożoność obliczeniowa ma większe znaczenie praktyczne, niż złożoność pesymistyczna. W omawianej rozprawie doktorskiej poświęcono średniej złożoności obliczeniowej jedynie jeden akapit (na stronie 115), twierdząc, że założenia przyjęte do obliczeń złożoności obliczeniowej pesymistycznej są mało prawdopodobne do spełnienia dla rzeczywistych zbiorów danych. Analizy powinny być szersze.

Nieco więcej informacji na temat złożoności średniej mogłoby dać dodanie pomiarów złożoności dostarczonych metod dla zbiorów danych o różnej wielkości. Takie dane są dostępne, zrealizowano i opisano wiele eksperymentów numerycznych.

2.4 Duża złożoność obliczeniowa, heurystyki

Przeszukiwanie pełne nowych warunków, np. dla warunków M-z-N, jest mało wydajne, jak wykazano w pracy. Typowym rozwiązaniem w takich przypadkach jest dostarczenie heurystyk. Być może doktorant mógłby jakieś zaproponować i przetestować ich skuteczność.

2.5 Błędy

W tekście jest trochę błędów i literówek.

Przykładowo, na stronie 97 w sekcji 3.1 „Motywacja” jest zdanie „Warunki wewnętrznej alternatywy ... np. $wiek > 20 \vee wiek < 35$ ”, Przedstawiony warunek nie jest użyteczny, jest tautologią. Nie można tego potraktować jako prostą omyłkę (wstawiono $\vee$ zamiast $\wedge$), bo w opisie jest słowo „alternatywa”.

2.6 Inne

Podział problemów rozwiązywanych przez nadzorowane metody uczenia maszynowego na klasyfikację, regresję i analizę przeżycia jest nietypowy. Przyjęło się wydzielać dwa zadania dla uczenia nadzorowanego: klasyfikację i regresję, różniąc je typem przewidywanej (objaśnianej) zmiennej. Typowo przewidywanie wartości zmiennej, która nie jest czasem i przewidywanie czasu wystąpienia zdarzenia to podziały wewnątrz regresji. Ten element nie musi być dyskutowany, to kwestia przejętej konwencji.

Treści przedstawiane w rozdziałach wstępnych (rozdziały 1-5), opracowane na podstawie literatury i stanowiące wprowadzenie do własnych metod są poprawne, ale lektura jest nieco utrudniona, ponieważ tytuł rozdziału sugeruje omawianie rozłącznych treści, wyliczanie (np. rozdział 2: „wybrane typy reguł decyzyjnych”, albo rozdział 3: „typy warunków ...”), tak jednak nie jest, treści poszczególnych sekcji często wiążą się ze sobą, rozszerzają lub zawężają poprzedni, np. warunki skośne (rozdział 3.5) i warunki relacji atrybutów (rozdział 3.6) są powiązane, jeżeli relacją będzie operator mniejszości. Ten element także jest kwestią przyjętej konwencji i nie musi być dyskutowany.

3 Ocena dorobku

Dorobek publikacyjny Kandydata jest niewielki, ale wystarczający na tym etapie rozwoju naukowego. W bazach publikacji są ujęte 4 pozycje, w trzech Kandydat jest pierwszym autorem.

- Gudyś, A. Maszczyk, C., et al., „RuleKit2: Faster and simpler rule learning”, SoftwareX, 2025, bez cytowań (Scopus).
- Maszczyk, C., Sikora, M. Wróbel, Ł., „Classification, Regression, and Survival Rule Induction with Complex and M-of-N Elementary Conditions”, Machine Learning and Knowledge Extraction, 2024, 3 cytowania (Scopus).
- Maszczyk, C., Macha, D., Sikora, M., „New Approach to Constructive Induction—Towards Deep Discrete Learning”, Lecture Notes in Networks and Systems, 2023, bez cytowań.
- Maszczyk, C., Kozielski, M., Sikora, M., „Rule-based approximation of black-box classifiers for tabular data to generate global and local explanations”, Proceedings of the 17th Conference on Computer Science and Intelligence Systems, 2022, 2 cytowania (Scopus).

Liczba cytowań jest niewielka (kilka), publikacje ukazały się w ostatnich latach. Wszystkie publikacje są wieloautorskie.

W mojej ocenie dorobek naukowy wystarczający, aby dopuścić Kandydata do dalszych etapów przewodu doktorskiego.

4 Podsumowanie

Temat badawczy uważam za bardzo istotny, teza jest poprawna i oryginalna, wykazana w stopniu wyczerpującym. Opracowane rozwiązanie dostarcza wyniki na poziomie prezentowanym w literaturze światowej, a ponadto ma wysoki poziom gotowości technologicznej oraz ma aspekty nowości.

Pod względem trzech podstawowych kryteriów stwierdzam że,

- rozprawa zawiera oryginalne rozwiązanie problemu badawczego,
- zgadzam się, że kandydat posiada ogólną wiedzę teoretyczną w dyscyplinie ITT,
- potwierdzam umiejętność samodzielnego prowadzenia pracy naukowej przez kandydata.

Stwierdzam, że recenzowana rozprawa doktorska mgr. inż. Cezarego Maszczyka spełnia warunki określone w aktualnych przepisach i wnioskuję do Rady Naukowej Dyscypliny Informatyka Techniczna i Telekomunikacja Politechniki Śląskiej o dopuszczenie tej rozprawy doktorskiej do publicznej obrony.

Zastępca Dyrektora
Instytutu Informatyki ds. Nauki

Prof. dr hab. inż. Robert Nowak