

Kraków, 13 marca 2026 r.

Dr hab. Maciej Smółka, prof. AGH  
Akademia Górniczo-Hutnicza im. Stanisława Staszica  
Wydział Informatyki  
Al. A. Mickiewicza 30, 30-059 Kraków

## **Recenzja pracy doktorskiej mgr. inż. Hammeda Mojeeda**

pt.

*An Interactive Machine-Learning-Based Multi-objective Optimization  
Framework for Software Overtime Planning*

Niniejsza recenzja została sporządzona na podstawie pisma Przewodniczącego Rady Dyscypliny Informatyka Techniczna i Telekomunikacja na Politechnice Gdańskiej dr. hab. inż. Jacka Raka, prof. PG z 26 listopada 2025 r. Recenzja jest opracowana zgodnie z wymaganiami określonymi w ustawie z dnia 20 lipca 2018 r. *Prawo o szkolnictwie wyższym i nauce* (Dz. U. 2018 poz. 1668 z późniejszymi zmianami).

### **Przedmiot rozprawy**

Rozprawa mgr. inż. Hammeda Mojeeda dotyczy zastosowania algorytmów uczenia maszynowego i optymalizacji wielokryterialnej do planowania nadgodzin w procesie produkcji oprogramowania. Deklarowanym celem ogólnym rozprawy jest *opracowanie interaktywnego algorytmu wielokryterialnego wykorzystującego uczenie maszynowe do optymalnego planowania nadgodzin*. Cel został sformułowany jasno i precyzyjnie, natomiast miejsce tezy zajmują następujące pytania badawcze.

- RQ1** Jak efektywnie jest w stanie predykcyjny model uczenia maszynowego oddać subiektywne preferencje menadżera projektu w planowaniu nadgodzin?
- RQ2** W jakim zakresie połączenie interaktywnego memetycznego algorytmu optymalizacji wielokryterialnej jest w stanie zastąpić ciągłą interakcję algorytmu z człowiekiem?
- RQ3** Jak wygląda porównanie rozwiązań znajdowanych przez wspomniane połączenie algorytmów z rozwiązaniami produkowanymi przez tradycyjne metody angażujące człowieka w trakcie działania pod względem opóźnienia projektu, kosztu i miar jakości kodu?
- RQ4** Jak proponowane rozwiązanie skaluje się względem rozmiarów i złożoności projektu?

Na podstawie powyższych pytań oraz odnoszących się do nich treści zawartych w pracy można wywnioskować następującą prawdopodobną tezę rozprawy.

*Subiektywne preferencje menadżera projektu w planowaniu nadgodzin można efektywnie estymować przy pomocy metod uczenia maszynowego. Pozwala to stworzyć skuteczny algorytm konstrukcji planów nadgodzin optymalizujących jednocześnie kilka kryteriów, wśród których fundamentalną rolę pełni estymowana subiektywna ocena kadry zarządzającej. Skonstruowana metoda będzie konkurencyjna w stosunku do metod angażujących człowieka w trakcie swojego działania pod względem nie tylko czasu i kosztu, ale również jakości otrzymanych rozwiązań. Ponadto skonstruowana metoda będzie się poprawnie skalować do projektów o różnym rozmiarze i złożoności.*

Problem optymalnego planowania nadgodzin jest jednym z istotnych zagadnień organizacji pracy, w szczególności zarządzania projektem informatycznym. Jest on naturalnie formułowany jako zadanie optymalizacji dyskretnej, w której kryteriami optymalności są koszty związane z opóźnieniami projektu, nadmiernym obciążeniem pracowników i złą jakością powstałego kodu. Kryteria te są wzajemnie od siebie zależne, ale na ogół pełne podejście do problemu wymaga uwzględnienia ich wszystkich, co skutkuje formułowaniem problemu optymalnego planowania nadgodzin jako zadania optymalizacji wielokryterialnej i, co za tym idzie, poszukiwaniem rozwiązań niezdominowanych w sensie Pareto. Badania Doktoranta stanowią istotny wkład w analizę tego problemu. Obejmują one:

- przeprowadzenie wszechstronnego i systematycznego przeglądu literatury na temat zastosowań algorytmów optymalizacji wielokryterialnej i uczenia maszynowego w planowaniu projektów programistycznych;
- zaprojektowanie mechanizmu zbierania i przechowywania subiektywnych preferencji kadry zarządzającej projektem;
- opracowanie modelu predykcyjnego uczenia maszynowego do estymacji subiektywnej oceny przydziału nadgodzin ze strony kadry zarządzającej;
- opracowanie interaktywnego algorytmu optymalizacji wielokryterialnej o nazwie ML-iMOSFLA wykorzystującego powyższy model predykcyjny w optymalizacji przydziału nadgodzin;
- przeprowadzenie weryfikacji eksperymentalnej skuteczności opracowanego algorytmu przy pomocy zestawu projektów informatycznych o zróżnicowanej wielkości i złożoności.

Należy uznać, że zaproponowane podejście uwzględnia wszystkie istotne składowe metodologii badań naukowych — od analizy kontekstu do zaproponowania i weryfikacji rozwiązania. Najważniejszą częścią oryginalnego dorobku Doktoranta jest

konstrukcja złożonego algorytmu ML-iMOSFLA będącego połączeniem metaheurystyki optymalizacji wielokryterialnej z modelem regresyjnym uczenia maszynowego wytrenowanym na nowo utworzonym zbiorze danych opisującym preferencje menadżerów projektów w zakresie przydziału nadgodzin. Warto tu zwrócić uwagę na interakcję ze środowiskiem komercyjnym w postaci zaangażowania grupy menadżerów projektów w przygotowanie tego zbioru danych. Przeprowadzona weryfikacja eksperymentalna wykazuje przewagę rozwiązania zaproponowanego przez Doktoranta nad metodami referencyjnymi.

### Uwagi merytoryczne

Bibliografia zebrana przez Doktoranta jest bogata i wszechstronna, ale nieco brakuje w niej pozycji najnowszych. Zestawienie nie zawiera artykułów naukowych z 2025 roku, a z roku 2024 znajdziemy tylko 2 artykuły, przy czym współautorem jednego z nich jest Doktorant.

Część formuł matematycznych jest przytoczona bez należytej dbałości o poprawność i precyzję. Poniżej przytaczam błędy i nieścisłości zawarte we wzorach i otaczającym je tekście.

- Definicja optymalności w sensie Pareto — wzór (2.3) — jest błędna, poprzedzający ją opis jest nieścisły i niewiele wyjaśnia.
- Definicja dominacji w sensie Pareto, czyli wzór (2.4), jest zasadniczo poprawna, ale w zapisie indeksy dolne w pewnym momencie przestają być indeksami.
- Wzór (3.3) nie jest zgodny z poprzedzającym opisem.
- Nie jest nigdzie wyjaśnione co oznaczają symbole  $P[i + 1]$ ,  $P[i - 1]$  i  $fk$  we wzorze (3.10).
- Nie wiadomo dlaczego w tym samym wzorze (3.10) domniemane cechy (jakie?) sąsiednich rozwiązań mają numery  $i + 1$  i  $i - 1$ . Jakie wobec tego rozwiązanie ma numer  $i$ ?
- Wielkość *crowding distance* we wzorze (3.10) ma oznaczenie  $Cd$ , a we wzorze (3.11)  $C_d$ .
- Oznaczenie  $P(i)$  u dołu strony 64 nie jest wyjaśnione.
- Nie wyjaśniono co oznacza  $P(i)^k$  we wzorze (3.12), zresztą cały ten wzór jest niezrozumiały.
- We wzorze (3.25)  $s$  nie może się zmieniać od 1 do  $S$ , bo ani  $s$  ani  $S$  nie są liczbami —  $s$  jest punktem (wielowymiarowym) ze zbioru  $S$ .
- pomijając indeks sumy wzór (3.25) jest poprawny tylko w przypadku zadań maksymalizacji — brak na ten temat komentarza w tekście.

- Oznaczenie  $set\{x\}$  wyjaśnione poniżej wzoru (3.25) w samym wzorze nie występuje. Prawdopodobnie miało to być wyjaśnienie symbolu  $Volume()$ .
- Wyrażenie w liczniku wewnętrznego ułamka we wzorze (3.26) jest całkowicie niezrozumiałe.
- Nie wiadomo co oznacza  $1 : ss$  w linii 2 Algorytmu 3.

Skrótowce i inne oznaczenia często nie są wyjaśniane przy pierwszym użyciu. Część z nich nigdzie w pracy nie jest wyjaśniona, np. MOWOA, FPA, LSD,  $Pred(25)$ , BBN, NRP, ASF, APP.

Doktorant dzieli algorytmy optymalizacji wielokryterialnej na trzy — zapewne rozłączne — klasy: *Multi-Objective Evolutionary Algorithms* (MOEA), *Multi-Objective Swarm Intelligence Algorithms* (MOSIA) i *Multi-Objective Bio-Inspired Optimization Algorithms* (MBIOA). Nie jest jasne według jakich kryteriów algorytmy zaliczane są do którejś z tych klas. Np. jeśli MOWOA oznacza *multi-objective whale optimization algorithm*, to dlaczego został zaliczony do klasy MOSIA a nie MBIOA?

Na str. 40 w tabeli 2.8 jako odrębne modele występują *Regression Tree* i *Decision Tree*. Przy tym drugie z tych pojęć jest uogólnieniem pierwszego, więc dość prawdopodobne, że chodzi o ten sam model. W każdym razie brak na ten temat komentarza w tekście.

Nie jest jasne jaki model jest nazwany *the Generalized Linear Model* na stronie 43. GLM to zróżnicowana wewnętrznie klasa modeli statystycznych i żaden z członków tej klasy nie jest traktowany jako domyślny reprezentant niezależny od dziedziny zastosowania.

Na stronie 44 sieć neuronowa (ANN) została zaliczona do modeli nieparametrycznych, co jest oczywiście błędne.

Używanie w opisie metaheurystyk terminologii związanej z dziedziną stanowiącą ich inspirację utrudnia zrozumienie istoty algorytmu i jego porównanie z innymi metodami. Z tego względu z opisu strategii ML-iMOSFLA powinny być wyeliminowane takie terminy jak *frog* czy *memeplex*.

We wzorze (3.7) występuje stała  $\alpha$ . Praca nie zawiera żadnego komentarza na temat jej znaczenia. Nie wiadomo też jakiej wartości w praktyce używał Doktorant.

Algorytm 1 na stronie 67 opisuje bagging, ale nie las losowy — brakuje losowania predyktorów, które wyróżnia tę drugą metodę.

W celu wybrania modelu uczenia maszynowego do estymacji preferencji menadżera projektu Doktorant dokonał porównania szeregu szeroko używanych modeli. Porównanie jest dość wszechstronne i w miarę poprawne metodologicznie, ale pozostawia następujące niewyjaśnione kwestie.

- Jak dobierano parametry modeli takie jak liczba drzew w lesie losowym i boostingu gradientowym, intensywność uczenia w boostingu gradientowym, jądro w SVR, liczba sąsiadów w kNN, czy wielkość i architektura MLP?
- Którego z algorytmów boostingu gradientowego użyto?

Ponadto interpretowalność nie może być argumentem decydującym w porównaniu lasu losowego z boostingiem gradientowym (jak wynikałoby z komentarza na str. 83), bo metody pod tym względem się nie różnią.

Wybrany model lasu losowego poddany został dostrajaniu parametrów, co jest uzasadnione. Niestety następujące kwestie nie zostały w rozprawie wyjaśnione.

- Doktorant porównał ze sobą kilka strategii dostrajania. Czy testy porównawcze były testami z ustalonym budżetem?
- Jeśli tak, to jaki budżet maksymalny przyjęto?
- Jeśli nie, to dlaczego algorytm GS osiągnął gorsze wyniki niż HGS?
- Jakie były optymalne znalezione zestawy parametrów lasu losowego?

Co więcej, wyniki na str. 92 i dalszych nie pokazują istotnej różnicy między HGS a gHGS, więc stwierdzenie o przewadze predykcyjnej tego drugiego algorytmu jest nieuzasadnione. Natomiast oczywiście na jego korzyść przemawiają wyniki dotyczące czasu wykonania (tabela 4.21 i rysunek 4.13).

Stwierdzenia o zdolności modelu RFR-gHGS do uogólniania wiedzy (str. 94) nie są w pełni uzasadnione ze względu na to, że predykcja modelu nie została sprawdzona na danych pochodzących z projektu, na którym model nie był trenowany.

Doktorant dokonał porównania ML-iMOSFLA z innymi algorytmami, które wypadają na korzyść ML-iMOSFLA, przy czym szczególnie ciekawe jest porównanie z przygotowanym na potrzeby rozprawy wariantem HIL-iMOSFLA. Natomiast słabą stroną porównania jest to, że uwzględnia ono wyłącznie warianty algorytmu MOSFLA i nie obejmuje współcześnie szeroko używanych ewolucyjnych algorytmów optymalizacji wielokryterialnej takich jak różne warianty NSGA czy MOEA/D. Doktorant uzasadnia to wynikami pracy [30], w której rzekomo wykazano wyższość MOSFLA nad NSGA. Niestety porównanie opisane w tej pracy jest wadliwe i nie może być uznane za miarodajne. Brakuje też eksperymentu pokazującego skuteczność zaprojektowanej strategii na zbiorze danych, który wcześniej nie był użyty w procesie trenowania składowego modelu uczenia maszynowego.

### Uwagi edytorskie

Praca pisana jest językiem formalnym na akceptowalnym poziomie poprawności. Tekst zawiera jednak pewną liczbę usterek edytorskich. Poniżej podano przykłady.

- Zdanie na str. 19 sugeruje podział zadań uczenia maszynowego na zadania estymacji i klasyfikacji. Zapewne omyłkowo zastąpiono *regression* przez *estimation*.
- Na str. 22 w linii 2 zamiast jednego *equality* powinno być *inequality*, a słowo *functions* pojawia się zapewne omyłkowo.

- We wzorze (2.1) powinno być  $h(x) = 0$ .
- Na str. 31 napisano, że lista artykułów jest w dodatku A, a faktycznie znajduje się w tabeli 2.4.
- Skrótowiec MOEA czasami (str. 36) zamienia się w MOAE.
- Na str. 58 fragment [85, 189]. *Memetic algorithms [...] within it [3]* jest powtórzony dwukrotnie.
- Gdzieś w pobliżu wzoru (3.5) przydałaby się uwaga o tym ile jest równe *maxovertime*.
- Rysunki 4.19 i 4.20 mają takie same podpisy.
- Na str. 103 zamiast *confidence level* powinno być *significance level*.

Bibliografia jest obszerna i zasadniczo dobrze przygotowana. Zawiera jednak następujące usterki:

- niejasne jest według jakiej reguły uporządkowano pozycje,
- referencja [36] jest duplikatem [22],
- referencja [106] jest duplikatem [105],
- referencje [24] i [77] mają niekompletne dane bibliograficzne,
- w referencji [129] część opisu jest pisana alfabetem prawdopodobnie japońskim.

Powyżej opisane mankamenty nie dominują jednak nad pozytywnymi stronami rozprawy. Można zatem uznać, że opracowanie i weryfikacja strategii optymalnego przydziału nadgodzin ML-iMOSFLA stanowi *oryginalne rozwiązanie problemu naukowego*, co oznacza, że kryterium ustawowe dotyczące przedmiotu rozprawy (art. 187 ust. 2) zostało spełnione.

## **Ocena umiejętności samodzielnego prowadzenia pracy naukowej**

Osiągnięciem mgr. inż. Hameda Mojeeda jest sformułowanie ważnego i aktualnego problemu z dziedziny informatyki technicznej, rozległa analiza istniejącego stanu badań w dziedzinie problemu, opracowanie oryginalnego rozwiązania tego problemu przy użyciu metod naukowych właściwych dla dziedziny i weryfikacja tego rozwiązania. Przygotowana przez Doktoranta rozprawa prezentuje pełny kontekst problemu na akceptowalnym poziomie przejrzystości, logiczności i ścisłości. Pozwala to uznać ustawowe kryterium umiejętności samodzielnego prowadzenia pracy naukowej (art. 187 ust. 1) za spełnione.

## **Ocena prezentacji ogólnej wiedzy teoretycznej Doktoranta**

Rozprawa doktorska mgr. inż. Hammeda Mojeeda ukazuje jego wiedzę teoretyczną w obszarze inżynierii oprogramowania, zwłaszcza w obszarze algorytmów wspierających planowanie projektów. Doktorant wykazał też znajomość współczesnych metod uczenia maszynowego oraz algorytmów stochastycznych optymalizacji wielokryterialnej. Można więc uznać, że kryterium ustawowe dotyczące ogólnej wiedzy teoretycznej Doktoranta (art. 187 ust. 1) również zostało spełnione.

## **Wniosek końcowy**

Uwagi krytyczne przedstawione powyżej nie zmieniają ogólnie pozytywnej oceny pracy. Rozprawę doktorską mgr. inż. Hammeda Mojeeda uważam za wartościową i stwierdzam, że spełnia ona wymagania obowiązującej ustawy *Prawo o szkolnictwie wyższym i nauce*. Wnoszę zatem o dopuszczenie Doktoranta do dalszych etapów postępowania.

Moniej Snothe

