

Gdańsk, 04.12.2023

Dr hab. inż. Paweł Kudela, prof. instytutu
Zakład Mechaniki Struktur Inteligentnych
Instytut Maszyn Przepływowych
im. Roberta Szewalskiego
Polskiej Akademii Nauk
ul. Fiszera 14
80-231 Gdańsk
email: pk@imp.gda.pl



**Recenzja rozprawy doktorskiej
Pana mgr. inż. Jakuba Białka
pt. „Wyjaśnianie predykcji i szacowanie wpływu dryfu danych na jakość
modeli Uczenia Maszynowego prognozujących zapotrzebowanie na
energię na przykładzie modelu prognozującego zapotrzebowanie na
ciepło w Warszawskiej Sieci Ciepłowniczej”**

Podstawa opracowania

Podstawą opracowania niniejszej recenzji jest pismo nr RND.IŚGiE.134.2023, Przewodniczącego Rady Naukowej Dyscypliny Inżynieria Środowiska, Górnictwo i Energetyka, Politechniki Warszawskiej, prof. dr. hab. inż. Tomasza Wiśniewskiego z dnia 24.10.2023, wraz z dołączoną do niego rozprawą doktorską Pana mgr. inż. Jakuba Białka.

Zasadność podjęcia tematu

Rozprawa doktorska Pana mgr. inż. Jakuba Białka związana jest z predykcją zapotrzebowania na ciepło w sieciach ciepłowniczych. Jest to zagadnienie niezwykle złożone i istotne zarówno z ekonomicznego punktu widzenia jak i z punktu widzenia komfortu cieplnego końcowego użytkownika sieci ciepłowniczej. Zagadnienie to poruszane jest w kontekście Warszawskiej Sieci Ciepłowniczej (WSC) i analizą danych pomiarowych pochodzących z rzeczywistych warunków jej pracy. Zespół z Instytutu Techniki Ciepłej Politechniki Warszawskiej wraz z Veolia Energia Warszawa S.A. opracował system wspomaganie decyzji dotyczących ustawień WSC, w którym jeden z algorytmów predykcji zapotrzebowania na ciepło bazuje na sztucznych sieciach neuronowych. Jest to typowy model określany mianem czarnej skrzynki (z ang. *black-box*), co powoduje trudności w interpretowaniu jego decyzji.

Doktorant badał ten problem za pomocą narzędzi wyjaśniających modele predykcyjne XAI (z ang. *eXplainable Artificial Intelligence*).

Drugi nurt rozprawy doktorskiej związany jest z dryfem danych, który ma istotny wpływ na jakość predykcji opracowanego modelu prognozującego zapotrzebowanie na ciepło. Wykrycie dryfu cech czy dryfu koncepcji może być przesłanką do przeprowadzenia ponownego treningu modelu.

Ogólne cele swojej pracy Doktorant formułuje następująco (str. 14):

1. *Analiza użyteczności narzędzi wyjaśniających modele predykcyjne XAI dla modeli prognozujących zapotrzebowanie na energię na przykładzie modelu prognozującego zapotrzebowanie na ciepło w Warszawskiej Sieci Ciepłowniczej (WSC).*
2. *Przeprowadzenie weryfikacji modelu Sztucznej Sieci Neuronowej prognozującego zapotrzebowanie na ciepło¹ w WSC za pomocą wybranych narzędzi XAI i na tym przykładzie stworzenie spójnych wytycznych do przeprowadzenia tego typu analiz na modelach prognozujących zapotrzebowanie na energię.*
3. *Badanie dryfu danych na modelu prognostycznym WSC. Teoretyczna i praktyczna analiza sposobów wykrywania dryfu, określania jego rodzaju, szacowania wielkości dryfu oraz wpływ dryfu na dokładność predykcji modelu.*

Doktorant postawił dwie tezy (str. 14):

Teza 1. *Możliwe jest wykorzystanie metod XAI do stworzenia interpretowalnych wyjaśnień predykcji dla złożonych modeli typu black-box opartych o algorytmy Sztucznych Sieci Neuronowych prognozujących zapotrzebowanie na energię i stworzenie procesu weryfikacji wzorców wyuczonych przez model poprzez konfrontację wyjaśnień predykcji z oczekiwaniami z praw fizyki i wiedzy technicznej.*

Teza 2. *Możliwe jest stworzenie algorytmów, których zadaniem jest weryfikacja czy założenia postawione w procesie tworzenia modeli SSN prognozujących zapotrzebowanie na energię są spełniane w produkcyjnym okresie korzystania modelu, w szczególności czy występuje dryf danych i jaki jest jego wpływ na jakość predykcji modelu.*

Cele i tezy pracy zostały sformułowane poprawnie i precyzyjnie. Temat wpisuje się w obszar coraz popularniejszych metod uczenia maszynowego w zastosowaniu do praktycznych problemów techniki.

Zakres i treść rozprawy

Rozprawa doktorska napisana jest w języku polskim. Jej główna część liczy 97 stron i składa się z pięciu rozdziałów.

W rozdziale 1 Doktorant wprowadza czytelnika w badany problem, formułuje cel i zakres pracy oraz stawia tezy, krótko omawia narzędzia XAI i własny wkład w rozwój dziedziny.

Rozdział 2 jest krótki (liczy tylko 4 strony) i zawiera opis modelu prognozującego zapotrzebowanie na ciepło, architekturę zastosowanej sieci neuronowej oraz charakterystykę danych wejściowych.

¹ Uzupełnienie recenzenta

Rozdział 3 traktuje o wyjaśnianiu predykcji modelu prognostycznego za pomocą narzędzi XAI dostępnych w literaturze, tj. *Permutation Importance* (PI), *Local Interpretable Model-agnostic Explanations* (LIME), *Partial Dependence* (PD), *Shapley Additive exPlanations* (SHAP), przy czym Doktorant skupia uwagę na ostatniej z metod ze względu na generowanie wyjaśnień o lepszej interpretowalności w porównaniu do pozostałych metod.

Rozdział 4 dotyczy monitorowania modelu prognostycznego, błędu predykcji oraz dryfu danych określanego jako zjawisko odpowiedzialne za spadek jakości modelu z czasem. Może to być spowodowane zmianą danych i relacji między zmiennymi wejściowymi i wyjściowymi, co Doktorant analizuje z podziałem na różne rodzaje dryfu danych.

Rozdział 5 to zwięzłe podsumowanie, liczące 2 strony, dotyczące analiz narzędzi XAI oraz dryfu danych przeprowadzonych przez Doktoranta wraz ze wskazaniem dalszych kierunków badań.

Bibliografia liczy 46 pozycji. W tej liczbie Doktorant jest współautorem dwóch artykułów. Przy czym odszyfrowanie tego nie jest oczywiste, ponieważ spis bibliografii jest niespójny, tj. na niektórych pozycjach podawani są wszyscy współautorzy, a na innych używany jest skrót *et al.*

Układ rozprawy doktorskiej nie budzi większych zastrzeżeń, z tym, że Doktorant w dużej mierze polega na odwoływaniu się do poz. [1] bibliografii, której to publikacji jest współautorem. Należy dodać, że praca [1] nie ma charakteru otwartego. We wprowadzeniu dowiadujemy się, że zarysowane zagadnienia będą rozważane w kontekście istniejącego modelu prognozującego zapotrzebowanie na ciepło w WSC z podaniem odwołania do pracy [1], po czym w dalszej części okazuje się, że w pracy [1] analizowano więcej modeli. Analizowany zbiór danych oraz architektura sieci neuronowej mogłyby być opisane w rozprawie doktorskiej bardziej przejrzysto i szczegółowo. Jest to niezbędne, jeśli chcielibyśmy odtworzyć uzyskane wyniki.

Ocena strony redakcyjnej

Język rozprawy jest poprawny, występują jedynie nieliczne błędy interpunkcyjne i literowe, co sprawia, że rozprawę dobrze się czyta. Doktorant w sposób przejrzysty przedstawia omawiane zagadnienia. Skład rozprawy jest estetyczny, rysunki zasadniczo mają jednolity charakter.

Natomiast w niektórych miejscach Doktorant wprowadza skrót, np. *Uczenie Maszynowe* (UM), którego później nie stosuje (np. str. 62); Niektóre skróty pochodzą z języka polskiego, np. *Sztuczna Inteligencja* (SI), a niektóre z języka angielskiego, np. *eXplainable Artificial Intelligence* (XAI), co wprowadza pewną niespójność.

Uwagi techniczne i redakcyjne:

- Str.10: „...z dużym prawdopodobieństwem taka korelacja nie utrzyma się w kolejnych lata”
>> „...z dużym prawdopodobieństwem taka korelacja nie utrzyma się w kolejnych latach”

Rys. 1: brak źródła

Str. 11: „Dobłą zmienną kodującą te aktywności jest...” >> „Dobłą zmienną kodującą te aktywności jest...”

Str. 12: „(czyli takiego przekształcania zmiennych wejściowych i ich grup, które zwiększają ich zdolność do predykcji celu)” >> „(czyli takiego przekształcania zmiennych wejściowych i ich grup, które zwiększają ich zdolność do predykcji celu)”

Str. 12: „ang. *eXplainable Artificial Intelligence - XAI*” >> „ang. *eXplainable Artificial Intelligence - XAI*”

Str. 16: „Weryfikacji modelu SSN prognozującego...” >> „Weryfikacja modelu SSN prognozującego...”

Str. 20: „warstwa wejściowa o wymiarze równej liczbie zmiennych wejściowych tj. 154,” >> „warstwa wejściowa o wymiarze równym liczbie zmiennych wejściowych, tj. 154,”

Str. 27: „Lokalnie dokładny, interpretowalny model utworzonych za pomocą metody LIME wskazuje...” >> „Lokalnie dokładny, interpretowalny model utworzony za pomocą metody LIME wskazuje...”

Str. 29: „... należy wielokrotnie wywoływać analizowany model black-box aby otrzymać predykcje.” >> : „... należy wielokrotnie wywoływać analizowany model black-box, aby otrzymać predykcje.”

Str. 35: „... na temat sposobu działania modelu tj. rozumienie ...” >> : „... na temat sposobu działania modelu, tj. rozumienie ...”

Str. 35: „... byłaby to zadanie równie żmudne...” >> „... byłoby to zadanie równie żmudne...”

Str. 35: „Nie zdefiniowano dotąd konkretnego procesu tj. zestawu kroków, które...” >>

„Nie zdefiniowano dotąd konkretnego procesu, tj. zestawu kroków, które...”

Str. 36: „... dla każdej obserwacji x i każdej zmiennej wejściowej ...” >> „... dla każdej obserwacji x i każdej zmiennej wejściowej ...”

Str. 41: „3.2.2.2 Temperatury zewnętrzna” >> „3.2.2.2 Temperatura zewnętrzna”

Str. 42: „... w letnią czy jesienną noc nie wiąże się z automatycznym ...” >>

„... w letnią czy jesienną noc nie wiąże się z automatycznym ...” >>

Str. 52: Czcionki na Rys. 24 są zbyt małe.

Str. 56: „W końcu poprawne odwzorowanie dobowych profili zapotrzebowania (Rysunek 25).”
Brak czasownika w zdaniu.

Str. 57: „Zdaniem autora przedstawiona w tej części rozprawy analiza jest spójna, pozwala na weryfikację wzorców wyuczonych przez model dla najbardziej istotnych zmiennych oraz zbudowanie wiedzy co do sposobu działania.” Co do sposobu działania czego?

Str. 65: Brakuje jednostek w odniesieniu do temperatury.

Str. 67: „Jak wspomniano we Wprowadzeniu (0) ...” ?

Str. 67: „Porównując jednak funkcje gęstości prawdopodobieństwa, dryf pokazany na rysunku niżej jest wyraźnie większy”; Odwołanie do rysunku powinno zawierać jego numer.

Str. 69: „... w którym przecinają się funkcje gęstości prawdopodobieństwa” >> „... w którym przecinają się funkcje gęstości prawdopodobieństwa”

Str. 72: „Wykres niżej obrazuje ...”; Odwołanie do rysunku powinno zawierać jego numer.

Str. 73: W podpisie rysunku 39 występuje „częstotliwość”, natomiast w tekście „częstość”.

Str. 74: „Podobnie stanie się gdy np. zmieniają się straty przesyłu...” >> „Podobnie stanie się gdy, np. zmieniają się straty przesyłu...”

Str. 75: „Jednocześnie granica przesunęła się w kierunku niższych temperatur ...” >> „Jednocześnie granica przesunęła się w kierunku niższych temperatur ...”

Str. 78: „Do tej pory analizowano zagadnienie związane z dryfem cech i dryfem koncepcji w odosobnieniu tzn. występowało tylko jedno z dwóch zjawisk.” >> „Do tej pory analizowano zagadnienie związane z dryfem cech i dryfem koncepcji w odosobnieniu, tzn. występowało tylko jedno z dwóch zjawisk.”

Str. 82: „... ale już sama wiedza o takim dryfie koncepcji pozwala podjąć pewne działania zaradcze np. zmniejszenie zaufania operatorów ...” >> : „... ale już sama wiedza o takim dryfie koncepcji pozwala podjąć pewne działania zaradcze, np. zmniejszenie zaufania operatorów ...”

Str. 84: „Klasyczne modele regresji UM zwracają tylko wartość punktową predykcji – tj. tylko wartość oczekiwaną y.” >> „Klasyczne modele regresji UM zwracają tylko wartość punktową predykcji – tj. tylko wartość oczekiwaną y.”

Str. 86: „Wtedy wyuczona z danych wartość (...) modelu monitorującego jest dokładniejsza (i różna od) predykcji (...) modelu monitorowanego.” >> „Wtedy wyuczona z danych wartość (...) modelu monitorującego jest dokładniejsza (i różna) od predykcji (...) modelu monitorowanego.”

Str. 94: „(przykładowo – jeśli zostanie zarejestrowana niewidziana wcześniej prędkość wiatru) Predykcje w takich ...” >> „(przykładowo – jeśli zostanie zarejestrowana niewidziana wcześniej prędkość wiatru). Predykcje w takich ...”

Str. 94: „Zaproponowane metody te będą działały poprawnie ...” >> „Zaproponowane metody będą działały poprawnie ...”

Bibliografia: Opis bibliograficzny, jak już wspominałem wcześniej, jest sformatowany w niekonsekwentny sposób.

Merytoryczna ocena rozprawy

Rozprawa doktorska Pana mgr. inż. Jakuba Białka poświęcona jest aktualnemu i ważkiemu tematowi związanemu pośrednio z prognozowaniem zapotrzebowania ciepła. W szczególności rozprawa traktuje o interpretowalności modeli prognozujących zapotrzebowanie na ciepło oraz dryfie danych, które to zagadnienia są istotne dla operatorów sieci ciepłowniczych.

Obecnie mamy do czynienia z renesansem metod uczenia maszynowego; podążanie za nowymi rozwiązaniami w bardzo szybko rozwijającej się dziedzinie nie jest łatwe. Mimo to, Doktorant stosuje nowoczesne narzędzia i czerpie inspiracje z aktualnej literatury o czym świadczy bibliografia, w której większość pozycji pochodzi z ostatnich dziesięciu lat. Należy jednak zwrócić uwagę, że Doktorant korzysta z gotowych narzędzi XAI, natomiast nie tworzy własnych narzędzi.

Metodyka prowadzenia badań, którą posługuje się Doktorant jest poprawna, to jest (i) Doktorant rozpoczyna pracę od zarysowania szerokiej problematyki dotyczącej prognozowania ciepła w sieciach ciepłowniczych, wyzwani i dostępnych metod związanych z interpretowalnością danych oraz z dryfem danych, (ii) formułuje problem szczegółowy, (iii) opisuje dostępne techniki i wybiera najbardziej obiecujące, (iv) przeprowadza analizy użyteczności wybranych narzędzi, (v) proponuje rozwiązania praktyczne i wyciąga trafne wnioski. Świadczy to o umiejętności samodzielnego prowadzenia pracy naukowej.

W tym kontekście należy podkreślić użyteczność przeprowadzonych przez Doktoranta analiz wpływu interakcji pomiędzy zmiennymi wejściowymi na interpretację wyjaśnień wygenerowanych przez metodę SHAP dla tych zmiennych i w ogólności zaproponowanie spójnego i systematycznego procesu analizowania modeli prognozujących zapotrzebowanie na ciepło bazujących na sztucznych sieciach neuronowych.

W drugiej części rozprawy Doktorant zaproponował teoretyczne określenie absolutnej miary różnych rodzajów dryfu danych. W szczególności, Doktorant opracował teoretyczną metodę szacowania wpływu dryfu cech na jakość predykcji modelu oraz zaproponował i zweryfikował praktyczne metody szacowania wpływu dryfu cech na jakość predykcji analizowanego modelu, co należy uznać za oryginalny i wartościowy wkład. Dodatkowo, przeprowadzone analizy prezentują ogólną wiedzę teoretyczną Doktoranta w dyscyplinie nauk inżynierjno-technicznych, a w szczególności biegłość w metodach uczenia maszynowego, rachunku prawdopodobieństwa i statystyki.

Z drugiej strony, moim zdaniem, w rozprawie doktorskiej brakuje holistycznego podejścia do problemu prognozowania zapotrzebowania na ciepło, Doktorant skupia się jedynie na wycinku tego zagadnienia. W rezultacie istotne kwestie są jedynie zasygnalizowane lub kompletnie przemilczane. Przykładowo wpływ bezwładności cieplnej na funkcjonowanie systemu predykcji jest poruszony tylko w kilku słowach na stronach 42-43. Zagadnienie magazynów ciepła zostało pominięte.

Analiza atrybucji cech przeprowadzona w rozdziale 3 w sposób naturalny dostarcza wniosków na temat inżynierii cech w badanym modelu sieci neuronowej i aż prosi się o zaproponowanie własnego modelu predykcji zapotrzebowania ciepła z ulepszonymi danymi wejściowymi. Doktorant jednak jedynie wyciąga pojedyncze wnioski, takie jak na str. 45, że model potrzebuje więcej informacji na temat długoterminowych temperatur historycznych. Budzi to niedosyt, zwłaszcza, że Doktorant ma dostęp do zbioru danych do trenowania modelu z zakresu aż 6 lat. Opracowanie nowego modelu byłoby znakomitą okazją do rozszerzenia rozprawy doktorskiej, w której zakres przeprowadzonych analiz jest dosyć skromny.

Pozostając przy zbiorze danych, w tym miejscu pozwolę sobie na dygresję. Opracowane metody uczenia maszynowego czy architektury sieci neuronowej mogą być ze sobą porównywane w sposób jednoznaczny, jeśli są uzupełnione o zbiór danych, na których zostały wytrenowane. Zatem zbiory danych powinny być udostępnione w otwartej bazie danych. Natomiast, zbiory danych związanych z pracą sieci ciepłowniczych raczej nie są udostępniane w otwartych bazach, co w pewnym sensie hamuje rozwój nowych modeli. Problem porównywalności modeli częściowo poruszany jest również w cytowanej pracy [1].

Pytania i uwagi dyskusyjne

Zasadniczo, rozprawa nie zawiera większych błędów, a przedstawione uwagi i pytania mają charakter dyskusyjny.

- 1) Współcześnie, w wielu przypadkach, dąży się do podejścia kompleksowego, z ang. *end-to-end*, w którym nowoczesne architektury sieci neuronowych przetwarzają dane wejściowe w postaci niemal surowej, np. w formie sygnałów, obrazów czy animacji, gdzie ekstrakcja cech jest niejako wbudowana w model. Czy w takich przypadkach narzędzia XAI są nadal użyteczne? W jaki sposób można by je wykorzystać?
- 2) Wydaje się, że zastosowanie *Nonlinear Autoregressive Neural Networks* lub *Recurrent Neural Networks* z *Gated Rcurrent Units* lub komórkami *Long Short Term Memory* ma duży potencjał w predykcji zapotrzebowania na ciepło. Jaka jest opinia Doktoranta na ten temat? Czy planowane są prace w przyszłości w tym kierunku?
- 3) Str. 27: Miesiąc w analizowanym modelu kodowany jest gorącojedynkowo, a na rys. 5 mamy $month_7 = 3$. Skąd to wynika? Jak należy rozumieć tabelę, w której zestawiono w kolumnach cechy i wartości, gdzie np. $month_8 = 0$, ale wartość jest „True”?
- 4) Str. 46: W analizie zmiennych kalendarzowych Doktorant dyskutuje problem zbyt wielu danych wejściowych dotyczących reprezentacji dnia i podaje liczbę wejść 364. Skąd wzięła się ta liczba skoro 365 dni 5 godzin i 48 minut to czas, w którym w przybliżeniu Ziemia okrąży Słońce.
- 5) Str. 47: Doktorant omawia przekształcenie dnia roku funkcją cosinus i stwierdza, że dodanie przekształcenia funkcją sinus nie poprawiało istotnie jakości modelu, ale nie podaje żadnych danych potwierdzających to stwierdzenie.
- 6) Str. 61: Doktorant wymienia rodzaje dryfu i w przypadku *manifestation shift* przytacza pozycję [45], ale w oryginalnym artykule autorzy nie posługują się takim terminem.
- 7) Str. 67: „Porównując jednak funkcje gęstości prawdopodobieństwa, dryf pokazany na rysunku niżej jest wyraźnie większy”. Według jakiej miary dryf jest wyraźnie większy?
- 8) Str. 91: „Do trenowania modelu regresji kwantylowej wykorzystano zbiór treningowy i referencyjny. Jako wejść użyto tych samych cech, co cechy wykorzystane do trenowania modelu [1]. Wykorzystano ten sam algorytm co algorytm [1] z odpowiednio zmodyfikowaną funkcją straty”.
Ponieważ w pracy [1] analizowano różne modele należałoby wyspecyfikować, o który model dokładnie chodzi. Na czym polegała modyfikacja funkcji straty?
- 9) Str. 92: Równanie (29) definiujące błąd średniokwadratowy jest niepoprawne, co zapewne wynika ze skopiowania zawartości równania (28).
- 10) Czy zastosowany model predykcji zapotrzebowania na ciepło można zastosować do innej sieci ciepłowniczej niż Warszawska Sieć Ciepłownicza?
- 11) Z jakich narzędzi programistycznych i bibliotek korzystał Doktorant podczas trenowania modelu sieci neuronowej?
- 12) Jeśli dobrze zrozumiałem, to rozważana sieć neuronowa została ponownie wytrenowana na innym zakresie danych pomiarowych. Jakie były przesłanki żeby ponownie wytrenować sieć neuronową? Jakie uzyskano wyniki trenowania w stosunku do poprzedniego zbioru danych?
- 13) W rozprawie doktorskiej pojawia się nieścisłość, ponieważ Doktorant określa Warszawską Sieć Ciepłowniczą jako największą w Europie, natomiast w pracy [1] oraz na stronie internetowej energiadlwarszawy.pl Warszawska Sieć Ciepłownicza

określana jest jako największa w Unii Europejskiej. Czy faktycznie i to pierwsze stwierdzenie jest prawdziwe?

- 14) Inna nieścisłość pojawia się w odniesieniu do danych wejściowych do sieci neuronowej, gdzie na str. 20 podana jest liczba 154 zmiennych wejściowych, podczas gdy na str. 22 mowa jest o 152 cechach wejściowych. Proszę o wyjaśnienie tej kwestii.
- 15) Załóżmy hipotetyczną sytuację modernizacji zaworów termostatycznych w grzejnikach, gdzie nowe zawory termostatyczne reagują na temperaturę zewnętrzną zamiast regulacji manualnej przez użytkowników. Jakich implikacji należałoby się spodziewać w kontekście dryfu danych?

Podsumowanie

Rozprawa doktorska Pana mgr. inż. Jakuba Białka dotyczy aktualnego i oryginalnego tematu z obszaru prognozowania zapotrzebowania ciepła. Z uwagi na stosowane narzędzia, temat ten dobrze wpisuje się w intensywnie rozwijane metody uczenia maszynowego. Rozwiązując postawiony problem badawczy Doktorant posłużył się klasyczną metodyką prowadzenia badań naukowych i wykazał się umiejętnością samodzielnego i systematycznego ich prowadzenia. Uzyskane wyniki mają użyteczny charakter.

Z formalnego punktu widzenia stwierdzam, że przedstawiona do recenzji praca spełnia w stopniu minimalnym wszystkie wymogi stawiane pracom doktorskim wynikające z Ustawy z dnia 20 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule naukowym w zakresie sztuki z późniejszymi zmianami. W związku z powyższym wnioskuję o dopuszczenie Doktoranta do publicznej obrony Jego rozprawy doktorskiej.



Paweł Kudela