

Recenzja rozprawy doktorskiej

Autor: Mgr Katarzyna Kobylińska

Tytuł: Metody wyjaśnialnego uczenia maszynowego dla danych tabelarycznych z przykładami zastosowań w medycynie

Promotorzy: prof. dr hab. inż. Przemysław Biecek
prof. dr hab. n. med. Mariusz Adamek

Ogólna charakterystyka rozprawy

Przedłożona do recenzji rozprawa liczy 135 stron tekstu, składa się z pięciu rozdziałów, spisu rysunków, spisu tabel, słownika terminów oraz bibliografii. Bibliografia jest obszerna, obejmuje wykaz 145 pozycji literatury. Praca zawiera opis metodologii wyjaśnialnej sztucznej inteligencji (XAI) w zastosowaniach do eksperymentów klinicznych oraz danych medycznych, a także formułuje oryginalne wyniki związane z badaniami nad zastosowaniem XAI do nowych danych medycznych i formułowaniem wynikających z tych badań wniosków.

Tematykę pracy należy uznać za bardzo istotną i interesującą. Obszar badań XAI jest z jednej strony intensywnie rozwijany z użyciem formalnych narzędzi modelowania matematycznego, technik obliczeniowych oraz symulacyjnych, a z drugiej strony ma bardzo często duże znaczenie dla wykorzystania wyników tego typu modelowania matematycznego w praktyce. To ostatnie sformułowanie dotyczy często zastosowań sztucznej inteligencji w medycynie i praktyce klinicznej. Możliwość jasnej i przejrzystej interpretacji wyników modelowania XAI jest bardzo dużą motywacją do badań i skupia często większe zainteresowanie użytkowników opracowanych narzędzi niż same metody konstrukcji modeli. Wyniki badań w obszarze XAI mają też bardzo często aspekty interdyscyplinarne. Wyniki przedstawione

w recenzowanej rozprawie także należy scharakteryzować jako interdyscyplinarne, łączące metody matematyczne i obliczeniowe z ich weryfikacją i interpretacją kliniczną i medyczną. Dwóch promotorów pracy zapewnia dwie perspektywy interpretacji uzyskiwanych wyników.

Rozdział 1 stanowi wprowadzenie do pracy. Zawiera wiele informacji i faktów wspierających znaczenie zastosowania metod sztucznej inteligencji i uczenia maszynowego. Omawia także i pokazuje motywacje do formułowania podejść XAI w badaniach, w szczególności w badaniach medycznych. W punkcie 1.2 przedstawione jest stosowane w pracy podejście polegające na poszukiwaniu argumentów za tym, że XAI prowadzi do usprawnienia procesu modelowania dzięki możliwości interpretacji oraz zrozumienia mechanizmów klasyfikacji, zależności funkcjonalnych itd. Przedstawia się także przegląd kilkunastu pozycji literaturowych charakteryzujących różne podejścia do zastosowania technik XAI we wspieraniu analiz danych medycznych.

W rozdziale 1 formułuje się dalej cel pracy oraz cztery hipotezy (tezy) pracy. Cel pracy jest określony jako zbadanie technik XAI w kontekście poprawy jakości wnioskowania. Tezy pracy 1-3 związane są z efektywnością różnych aspektów metodologii XAI i weryfikowane są na bazie analiz danych eksperymentalnych / obserwacyjnych. Natomiast teza 4 dotyczy oryginalnej metody XAI automatycznego wyboru podzbioru modeli, autorstwa Doktorantki, oraz wykazania skuteczności i użyteczności tej metody. Każda z tez pracy jest ponadto powiązana z oryginalną publikacją naukową, której współautorką jest Doktorantka. Obok artykułów naukowych stanowiących bezpośrednie odniesienie do tez pracy Doktorantka jest także współautorką szeregu innych, wartościowych publikacji naukowych powiązanych z tematyką doktoratu. Publikacje te mają interdyscyplinarny charakter i zostały uzyskane jako wynik współpracy informatyków z zespołami naukowymi klinicystów i medyków. Ich krótkie omówienie zamieszczone jest w podrozdziale 1.4. Na końcu rozdziału 1, w podrozdziale 1.7 Doktorantka przedstawia charakterystykę swojego wkładu w powstanie czterech publikacji związanych z tezami pracy.

W rozdziale 2 pracy zebrane są metody, pojęcia i algorytmy stosowane w publikacjach związanych z pracą doktorską. Jak podkreślono na wstępie rozdziału 2, metodologia pracy skupia się na modelach predykcyjnych w zakresie uczenia nadzorowanego. Charakteryzuje się pojęcia często stosowane w uczeniu nadzorowanym, w zadaniach predykcji i klasyfikacji, takie jak drzewo czy las losowy. Wspomina się o metodzie konstrukcji drzewa decyzyjnego przez wzmocnienie gradientowe. Następnie, w podrozdziale 2.2, charakteryzuje się pojęcie wyjaśnialnego uczenia maszynowego oraz omawia się rodzaje interpretowalnych (wyjaśnialnych) modeli uczenia maszynowego. Kolejne dwa podrozdziały, 2.3 i 2.4, poświęcone są dwóm podejściom do interpretowalności, globalnemu (podrozdział 2.3) i lokalnemu (podrozdział 2.4). W podejściu globalnym wyszczególnione są dwie metody, cząstkowych zależności oraz oceny ważności zmiennych z wykorzystaniem permutacji. W podejściu lokalnym opisuje się metodę SHAP, metodę dekompozycji break-down oraz metodę konstrukcji lokalnego profilu (*ceteris paribus*). Wreszcie,

w ostatnim podrozdziale 2.5, przedstawia się zbiór ogólnych uwag i komentarzy dotyczących procesu budowy modelu predykcji, klasyfikacji i interpretacji.

W rozdziale 3 opisane są wyniki uzyskane w publikacjach związanych z tezami 1-3 pracy. Wyniki te związane są z zastosowaniem modeli XAI dla praktycznych danych eksperymentalnych w obszarze medycyny. Skuteczność metodologii XAI i prawdziwość tez wykazuje się przez analizę przykładów. Na wstępie podkreśla się, że podejście XAI może być efektywnie zastosowane na każdym etapie budowania modelu predykcji lub klasyfikacji, formułowania modelu, walidacji oraz implementacji.

Podrozdział 3.1 poświęcony jest tezie 1 i ściśle związany jest z publikacją Doktorantki: Kobylińska, K., Mikołajczyk, T., Adamek, M., Orłowski, T., & Biecek, P. (2019). *Explainable machine learning for modeling of early postoperative mortality in lung cancer*. In Artificial Intelligence in Medicine: Knowledge Representation and Transparent and Explainable Systems: AIME 2019 International Workshops, KR4HC/ProHealth and TEAAM, Poznan, Poland, June 26–29, 2019, Revised Selected Papers (pp. 161-174). Springer International Publishing. W pracy tej skoncentrowano się na metodologii budowy i walidacji modeli predykcyjnych wczesnej śmiertelności pooperacyjnej pacjentów z rakiem płuc. Przez wczesną śmiertelność pooperacyjną rozumiano spełnienie lub niespełnienie warunku przeżycia przez pacjenta 3 miesięcy po operacji. W budowie i walidacji modelu wykorzystywano paradygmaty XAI. Analizy bazowały na bardzo szerokiej bazie danych obserwacyjnych, obejmującej historię palenia tytoniu przez pacjentów, cechy radiologiczne przeprowadzonych badań obrazowania medycznego, cechy patologiczne guzków w płucach, wyniki badań, czynniki ryzyka raka, objawy i choroby współistniejące. Dane te zarejestrowano dla ponad 30 000 pacjentów. Do analiz zastosowano metodologię opisaną w rozdziale 2 pracy. Jako podstawę do analizy predykcyjnej przyjęto model regresji logistycznej oraz model lasu losowego. W podejściu globalnym narysowano wykresy ważności cech (zmiennych) a także wykresy zależności cząstkowych dla predykcji. Z kolei dla podejścia lokalnego wykreślono schematy dekompozycji breakdown (dla szeregu zmiennych) oraz profile ceteris-paribus dla zmiennych APTT, Age i Packyears. W konkluzji do podrozdziału 3.1 podkreśla się wagę porównania różnych modeli oraz użycia technik XAI dla dokonania wiarygodnych analiz praktycznych, dużych danych medycznych.

Podrozdział 3.2, w którym analizuje się zastosowanie XAI na etapie walidacji modeli, odnosi się do tezy 2 oraz do publikacji Doktorantki: Kobylińska, K., Orłowski, T., Adamek, M., & Biecek, P. (2022). *Explainable machine learning for lung cancer screening models*. Applied Sciences, 12(4), 1926. Poruszana problematyka związana jest z modelowaniem ryzyka w badaniach typu przesiewowego. Porównuje się trzy modele przesiewowe ryzyka (LCRAT, BACH, PLCO) i wykorzystuje się bazę danych pacjentów chorujących na raka płuc, wspomnianą wcześniej w podrozdziale 3.1. Porównuje się predykcje modeli LCRAT, BACH, PLCO z wykorzystaniem podejścia XAI. Podobnie jak w poprzednich analizach wykorzystuje się podejście globalne oraz lokalne do interpretacji modeli. Porównuje się, dla trzech modeli, istotności cech, profile

zależności cząstkowych, a także wyznacza się dla nich diagramy break-down oraz lokalne profile typu ceteris paribus. W podsumowaniu do tego podrozdziału stwierdza się, że przeprowadzona analiza stanowi empiryczny dowód na efektywność wykorzystania metodologii XAI w zadaniach porównywania i walidacji modeli ryzyka w medycynie i praktyce klinicznej.

Podrozdział 3.3 pracy jest związany z tezą 3 oraz z publikacją, której współautorką jest Doktorantka: Lemańska-Perek, A., Krzyżanowska-Gołąb, D., Kobylińska, K., Biecek, P., Skalec, T., Tyszko, M., ... & Adamik, B. (2022). *Explainable artificial intelligence helps in understanding the effect of fibronectin on survival of sepsis*. *Cells*, 11(15), 2433. Badano model ryzyka zgonu na bazie zbioru danych obserwacyjnych 122 pacjentów z sepsą na oddziale intensywnej terapii w Szpitalu Uniwersyteckim (we Wrocławiu). Opracowano trzy rodzaje modelu predykcyjnego, model regresji logistycznej, model lasu losowego oraz model lasu losowego z algorytmem wzmocnienia gradientowego. Dokonano analizy XAI wyników modelowania. Wykreślono diagramy ważności cech, profile zależności cząstkowych, diagramy dekompozycja break - down, wykresy SHAP oraz wykresy ceteris paribus, dla istotnych cech charakteryzujących model ryzyka. W konkluzji tego podrozdziału stwierdzono przydatność opracowanego modelu dla praktycznego zastosowania w praktyce klinicznej.

Rozdział 4 poświęcony jest uzasadnieniu ostatniej tezy doktoratu. Jest on związany z publikacją Doktorantki: Kobylińska, K., Krzyziński, M., Machowicz, R., Adamek, M., & Biecek, P. (2024). *Exploration of the Rashomon Set Assists Trustworthy Explanations for Medical Data*. *IEEE Journal of Biomedical and Health Informatics*, 28(11), 6454-6465. W rozdziale tym opisana jest oryginalna metodologia analizy (charakterystyki) zbiorów Rashomon nawiązująca do technik XAI. Efekt Rashomon w analizach danych był przedmiotem badań szeregu autorów w literaturze. Doktorantka (wraz ze współautorami) podejmuje się dość złożonego zadania wprowadzenia odpowiednich wskaźników i metod dla charakteryzacji zbioru Rashomon. Trafnie ocenia, że dla zrealizowania projektu charakteryzacji zbioru Rashomon należy zdefiniować miary odległości pomiędzy modelami oraz poszukiwać par modeli o jak największych wzajemnych odległościach. Cenne jest także to, że dla zweryfikowania swoich wyników stosuje zarówno dane symulacyjne jak też rzeczywisty zbiór danych medycznych pacjentów z zespołem HLH. Wynik zastosowania podejścia zaproponowanego przez Doktorantkę można rozumieć jako technikę redukcji zbioru Rashomon, użyteczną w predykcji, klasyfikacji oraz w interpretacji wyników analiz.

Ocena rozprawy

Doktorantka podejmuje bardzo interesujący kierunek badań naukowych związanych z wyjaśnialną (interpretowalną) sztuczną inteligencją – XAI. Wykazuje się wiedzą dotyczącą współczesnych problemów uczenia maszynowego, sztucznej inteligencji, problemów interpretacji, praktycznych zadań analizy



danych obserwacyjnych, predykcji, klasyfikacji. Należy docenić szeroki, dość wyczerpujący przegląd publikacji naukowych z tego obszaru zamieszczony w pracy.

Publikacje, których wyniki i tezy zawarte są w pracy, powstałe z udziałem Doktorantki, są oryginalne. Przedstawiają nowe koncepcje i podejścia do problematyki wykorzystania XAI w analizie danych medycznych. Bazują na rzeczywistych zbiorach danych. Ukazały się w recenzowanych czasopismach naukowych lub/i w materiałach międzynarodowych konferencji naukowych.

Publikacje są wieloautorskie, jednak Doktorantka przedstawiła charakterystykę swojego wkładu w powstanie każdej z tych publikacji. Bardzo pozytywnym aspektem udziału Doktorantki w opracowaniu publikacji jest różnorodność realizowanych przez nią prac (projektów) naukowych. We wspomnianych artykułach występuje w różnych rolach, opracowuje implementacje algorytmów, dokonuje analiz danych, organizuje współpracę z zespołami medycznymi, formułuje problemy badawcze, zajmuje się edycją tekstu publikacji. Różnorodność realizowanych zadań i prac naukowych w wieloautorskich publikacjach bardzo dobrze świadczy o kompetencjach Doktorantki i o dojrzałości jej warsztatu naukowego, a także o umiejętności podejmowania pracy w zespołach naukowych.

Doktorantka jest także współautorką szeregu innych interesujących publikacji naukowych z zakresu sztucznej inteligencji.

Podsumowując, Doktorantka wykazuje się umiejętnością modelowania matematycznego w obszarze sztucznej inteligencji i uczenia maszynowego, a także kompetencjami w zakresie technik informatycznych. Potrafi postawić oryginalne tezy naukowe oraz sformułować i rozwiązać oryginalne problemy naukowe z zakresu informatyki technicznej.

Za najciekawszy z przedstawionych wyników, w mojej opinii, należy uznać rozdział 4. Zaproponowana metodologia powinna być interesująca dla innych badaczy. Istnienie zbioru modeli Rashomon jest częstym przypadkiem w analizie danych zwłaszcza medycznych.

Uwagi i pytania do doktoratu

- Rozdział drugi stanowi interesujące zebranie stosowanych metod i pojęć. Jednak brakuje w nim trochę precyzji i jasności stosowanych pojęć matematycznych. Na przykład dla wzorów (2.2), (2.3) brakuje wytłumaczenia znaczenia stosowanych tam symboli. Sformułowanie na stronie 33 „... można wyrazić drzewo następująco” jest niejasne. Sformułowanie na stronie 35 „Model zespołowy drzewiasty można zapisać jako sumę drzew” jest żargonowe.

- W rozdziale 4 stosuje się niekonsekwentnie termin: zbiór Rashomon, lub zbiór Rashomona. Powinno się jednolicie stosować pierwszy z tych terminów.
- Czy dla predyktorów i klasyfikatorów regresji logistycznej i lasów losowych stosowano gotowe oprogramowanie? Z jakiego środowiska? Czy opracowywano całościową implementację dla eksperymentów obliczeniowych? Czy Doktorantka może przedstawić jakieś doświadczenia płynące z tych prac?
- Wykres ważności zmiennych przedstawiony na rysunku 3.2 różni się od analogicznego wykresu (Fig. 2) w publikacji: Kobylińska, K., Mikołajczyk, T., Adamek, M., Orłowski, T., & Biecek, P. (2019). *Explainable machine learning for modeling of early postoperative mortality in lung cancer*. In Artificial Intelligence in Medicine: Knowledge Representation and Transparent and Explainable Systems: AIME 2019 International Workshops, KR4HC/ProHealth and TEAAM, Poznan, Poland, June 26–29, 2019, Revised Selected Papers (pp. 161-174). Czy Doktorantka mogłaby skomentować te różnice?
- Czy Doktorantka zna lub interesuje się dalszymi (innymi niż opisane w pracy) zastosowaniami XAI w sztucznej inteligencji? Na przykład zastosowaniami XAI w głębokich sieciach (głębokim uczeniu) lub w transferze wiedzy?

Konkluzja

Uwagi krytyczne i dyskusyjne nie podważają oryginalności i wartości pracy. Praca zawiera oryginalne wyniki badań naukowych, opublikowane w wysokiej rangi czasopismach naukowych i przedstawione na międzynarodowych konferencjach naukowych.

Stwierdzam, że rozprawa spełnia wynikające z ustawy warunki stawiane pracom doktorskim i wnioskuję o jej dopuszczenie do publicznej obrony.

