Warszawa, dnia 20-01-2025

Recenzja rozprawy doktorskiej mgr Katarzyny Kobylińskiej

dr hab.inż. Marta Zalewska

Zakład Profilaktyki Zagrożeń Środowiskowych, Alergologii i Immunolgii,

Wydział Nauk o Zdrowiu

Warszawski Uniwersytet Medyczny,

Żwirki i Wigury 61, 02- 091 Warszawa

marta.zalewska@wum.edu.pl

Ocena rozprawy doktorskiej Pani mgr Katarzyny Kobylińskiej pt. **„Metody wyjaśnialnego uczenia maszynowego dla danych tabelarycznych z przykładami zastosowań w medycynie"**, napisanej pod kierunkiem dwóch promotorów.

Promotorzy rozprawy:

prof. dr hab. inż. Przemysław Biecek

Instytut Informatyki, Uniwersytet Warszawski

prof. dr. hab. n. med. Mariusz Adamek

Wydział Nauk Medycznych, Śląski Uniwersytet Medyczny

Moja ocena składa się z następujących tematów:

1. Omówienie zawartości rozprawy
2. Merytoryczna ocena treści rozprawy
3. Ocena redakcyjnej strony pracy
4. Ocena wiedzy doktorantki
5. Podsumowanie oceny i moja rekomendacja w sprawie nadania stopnia doktora

1. Omówienie zawartości rozprawy

Rozprawa jest napisana w języku polskim i opiera się na czterech opublikowanych pracach w języku angielskim (ostatnia praca została opublikowana w listopadzie 2024), w trzech z nich doktorantka jest pierwszą autorką. Lista publikacji zawarta jest w Tabeli 1.1.

W Tabeli 1.2 znajduje się lista sześciu wybranych publikacji w ramach współpracy naukowej doktorantki. Większość prac jest opublikowana w międzynarodowych wydawnictwach o dużej renomie.

Łącznie Doktorantka jest współautorką 10 artykułów naukowych

Rozprawa składa się z 5 rozdziałów.

W rozdziale pierwszym (zawartym na stronach 17-29), po krótkim wprowadzeniu, doktorantka przedstawiła motywację do prowadzenia badań w tematyce pracy, następnie sformułowała cel pracy i hipotezy badawcze oraz wymieniła publikacje stanowiące podstawę niniejszej pracy wraz z omówieniem wkładu własnego; wyjaśniła strukturę swojej pracy i zamieściła podziękowania.

W rozdziale 2 (na stronach 31-46) zatytułowanym „Metody i Algorytmy" Autorka przedstawiła metody i algorytmy uczenia maszynowego: drzewo, las losowy oraz wzmocnienie gradientowe. Następnie skoncentrowała się na interpretowalnym uczeniu maszynowym. W dalszej części rozdziału zajęła się globalnymi wyjaśnieniami modeli predykcyjnych; skoncentrowała się na metodzie cząstkowej zależności oraz permutacyjnej ważności zmiennych. Następnie zajęła się lokalnymi wyjaśnieniami modeli predykcyjnych w tym metody SHapley Additive eXplanations, w dalszej części zajęła się dekompozycją modelu break-down i omówiła lokalny profil. Zakończyła rozdział omawiając proces budowy modelu.

Rozdział 3 (str. 47-88) zatytułowany „Wyjaśnialne uczenie maszynowe w obszarze medycyny" składa się z trzech podrozdziałów, w których autorka omawia zastosowanie metod XAI do rzeczywistych problemów medycznych. W podrozdziale 3.1 autorka koncentruje się na sformułowaniu hipotezy badawczej "Metody wyjaśnialnego uczenia maszynowego post-hoc poprawiają jakość modelowania rozumianą jako dokładność predykcji lub poprawność modelu. Zastosowanie tych metod pozwala na osiągnięcie lepszych wyników modeli uczenia maszynowego w medycynie". W tym celu wykorzystuje swoją publikację [68] dotyczącą przeżywalności pooperacyjnej u osób chorujących na raka płuca i prezentuje model uczenia maszynowego przewidujący prawdopodobieństwo przeżycia pacjenta okresu trzech miesięcy po operacji na podstawie cech pacjenta dostępnych przed operacją (Tabela 3.39, str. 54). Na rysunku 3.2 (str. 56) autorka przedstawia porównanie modelu regresji logistycznej i modelu lasu losowego względem ważności zmiennych, a na rysunku 3.3 (str. 57) przedstawia różnice między modelami dla najbardziej istotnych zmiennych wykorzystując wykresy cząstkowej zależności PDP. Szczegółowe analizy zmiennych autorka nazywa „globalnymi metodami wyjaśniania" a dalej na str. 59 przedstawia „ lokalne metody wyjaśniania" w celu dogłębnego zrozumienia predykcji dla konkretnego pacjenta. Na rysunku 3.4 zamieszcza wyniki dekompozycji break-down i prezentuje wkłady poszczególnych zmiennych w predykcję dla wybranego pacjenta. Podrozdział kończy się podsumowaniem (str. 63). W podrozdziale 3.2 autorka koncentruje się na uzasadnieniu hipotezy badawczej "Metody wyjaśnialnego uczenia maszynowego post-hoc identyfikują słabe strony modeli. Metody te pozwalają na identyfikację ograniczeń i niedoskonałości modeli". Z kolei w podrozdziale 3.3 autorka zajmuje się trzecią hipotezą badawczą „Metody wyjaśnialnego uczenia maszynowego post-hoc wspomagają zrozumienie i interpretację wyników modelowania. Techniki te umożliwiają bardziej przejrzyste wyjaśnienie działania modeli i pomagają użytkownikom zrozumienie procesu podejmowania decyzji przez modele". Każdy z trzech podrozdziałów ma podobny układ i zawiera analizę konkretnego problemu medycznego.

Rozdział 4 omawia opracowaną przez Autorkę metodę automatycznego wyboru podzbioru najbardziej różnych modeli spośród wielu, dobrych modeli oraz zamieszcza opis własnego algorytmu Rashomon_DETECT. W rozdziale tym zawarta jest również analiza syntetycznych zbiorów danych oraz zastosowanie opracowanej metody do rzeczywistego problemu medycznego (studium przypadku).

Rozdział 5 zawiera podsumowanie, a praca kończy się spisem literatury zawierającym 145 pozycji.

2. Merytoryczna ocena treści rozprawy

Rozprawa Katarzyny Kobylińskiej ma charakter wybitnie interdyscyplinarny. Problemy natury medycznej nie tylko stanowią motywację rozwijania metod XAI, ale są w centrum uwagi Autorki. Metody XAI są narzędziem, pozwalającym na identyfikację czynników ryzyka, prognozowanie stanu pacjenta i wspomaganie decyzji terapeutycznych.

Rozprawa stanowi wkład do dwóch (lub nawet trzech) dyscyplin nauki: informatyki i nauk medycznych oraz nauk o zdrowiu. Dwóch promotorów reprezentuje dwie różne dziedziny.

Interdyscyplinarność należy uznać za zaletę pracy doktorantki. Moją trudnością jest to, że praca ze względów formalnych jest „przypisana" do dyscypliny informatyka. Ocena rozprawy powinna zatem brać pod uwagę w pierwszym rzędzie jej wkład do informatyki. Nie jestem specjalistką w tej dziedzinie i w mojej recenzji skoncentruję się na ocenie rozprawy z punktu widzenia nauk o zdrowiu i statystyki. Uważam, że wkład Autorki w nauki o zdrowiu jest znaczący i postaram się tę tezę uzasadnić. Jeśli chodzi o wkład w informatykę, ograniczę się do kilku ogólnych uwag, pozostawiając szczegółową ocenę recenzentowi, który jest specjalistą w tej dziedzinie.

2.1 Wkład rozprawy w dyscyplinę informatyka oraz zastosowania statystyki

Najważniejszą częścią rozprawy są Rozdziały 3 i 4.

Rozdział 3 dotyczy zastosowania znanych metod i narzędzi informatycznych do rozwiązywania realnych problemów medycznych. Autorka wykorzystuje metody uczenia maszynowego, przede wszystkim drzewa losowe, lasy losowe i *boosting* oraz klasyczne narzędzia statystyczne takie jak regresja logistyczna.

Wkład Autorki polega na wykorzystaniu technik wyjaśnialnej sztucznej inteligencji (XAI) do krytycznej analizy rezultatów otrzymanych przez różne metody. Takie podejście jest określane jako *post hoc* XAI. Podstawowe techniki *post hoc* XAI stosowane przez Autorkę pozwalają ocenić ważność poszczególnych zmiennych dla predykcji. Używana jest metoda cząstkowej zależności (profile PDP- partial dependence plot), ocena permutacyjnej ważności zmiennych (VI – *variable importance*), metoda SHAP, dekompozycja „*break-down*", lokalny profil *ceteris paribus* (CP). W celu porównania różnych konkurujących modeli pod względem jakości predykcji Autorka używa m.in. krzywych ROC i wskaźnika AUC.

Metody uczenia maszynowego i sztucznej inteligencji wykorzystywane w Rozdziale 3 nie są nowe. Wkład Autorki polega na empirycznym udowodnieniu skuteczności tych metod, pokazaniu, że *post hoc* XAI poprawia jakość modelowania rozumianą jako dokładność predykcji (podrozdział 3.1), pomaga w walidacji modeli oraz wyborze modelu (podrozdział

3.2) i, poprzez dostarczenie zrozumiałych dla użytkownika wyjaśnień, rozszerza stosowalność AI (podrozdział 3.3). Podsumowując, należy uznać, że Rozdział 3 zawiera skromny wkład w informatykę jednak ma duże znaczenie dla zastosowań w medycynie i w zdrowiu publicznym. Materiał zamieszczony w podrozdziale 3.1 jest oparty na artykule [68], który był prezentowany na konferencji informatycznej AIME 2019 i został opublikowany w *Lecture Notes in Computer Science*.

Rozdział 4 zawiera propozycję nowej procedury informatycznej, która pozwala na porównywanie modeli ze zbioru Rashomon w celu wyboru podzbioru najbardziej różnych modeli spośród modeli o jakości predykcyjnej zbliżonej do najlepszej. Autorka wprowadza nową miarę rozbieżności pomiędzy profilami, opartą na porównaniu zgodności znaku pochodnej (wskaźnik PDI zdefiniowany wzorem (4.5)). Autorka przedstawia skonstruowany przez siebie algorytm, wykorzystujący miarę PDI do identyfikacji najbardziej różnych modeli w zbiorze Rashomon (Rashomon_DETECT na str. 101). Przeprowadzone są doświadczenia symulacyjne (podrozdział 4.4.1). Zastosowanie algorytmu w przykładowym problemie medycznym przedstawione jest w podrozdziale 4.4.2.

Idea miary PDI jest motywowana tym, że najłatwiej interpretowalną cechą profilu jest kierunek zależności (czy wzrost zmiennej objaśniającej wpływa na zwiększenie, czy na zmniejszenie przewidywanej wartości zmiennej odpowiedzi). Metoda, proponowana przez Autorkę w Rozdziale 4, jest więc nowym krokiem ułatwiającym stosowanie AI przez użytkowników (na przykład lekarzy, pracowników zdrowia publicznego), co jest zgodne z ideą XAI. Metoda ta pozwala na konfrontację modeli uczenia maszynowego z wiedzą dziedzinową (medyczną lub zdrowia publicznego).

Wkład Autorki został potwierdzony faktem opublikowania artykułu [70], na którym opiera się Rozdział 4, w renomowanym piśmie informatycznym *IEEE J. Biomed. Health Informatics* (IF=6.7) w roku 2024.

2.2 Wkład rozprawy w dyscyplinę nauk o zdrowiu

Zastosowania AI w Zdrowiu Publicznym są szalenie ważne i ciągle rozwijane. Sztuczna inteligencja daje możliwość przetwarzania i analizowania ogromnych ilości danych medycznych znacznie przekraczającą możliwości człowieka. Te możliwości AI odgrywają kluczową rolę w diagnozowaniu chorób, zalecaniu leczenia i przewidywaniu reakcji pacjenta. W rozdziale 3 autorka skupiła się na pokazaniu skuteczności i dużych możliwości metod XAI w odniesieniu do zastosowań w naukach medycznych.

Autorka pokazuje zastosowanie tych metod w trzech ważnych problemach medycznych. Jeden z tych problemów dotyczy przeżywalności pooperacyjnej u osób chorujących na nowotwór płuca (podrozdział 3.1.2). Drugi problem dotyczy badań przesiewowych mających na celu wykrycie raka płuca na wczesnym etapie rozwoju (podrozdział 3.2.2). Trzeci dotyczy przewidywania ryzyka zgonu u pacjentów Oddziału Intensywnej Terapii OIT (podrozdział 3.3.2).

Podrozdziały 3.2 i 3.3 są oparte na pracach [69] i [75] opublikowanych odpowiednio w Applied Science (IF=2.5) i Cells (IF=5.1). Są to wysokiej klasy pisma o profilu interdyscyplinarnym i biologicznym.

Choć metody uczenia maszynowego i sztucznej inteligencji zastosowane w Rozdziale 3 rozprawy nie są nowe to mają duże znaczenie w naukach o zdrowiu, a ich uzupełnienie technikami XAI jest bardzo wartościowe.

Autorka stwierdza, że dzięki metodom XAI uzyskała więcej informacji na różnych etapach tworzenia modelu, co z kolei przyczyniło się do zwiększenia wiarygodności modelowania predykcyjnego. Istotną zaletą rozwijanych przez Autorkę metod wyjaśnialnej sztucznej inteligencji XAI jest możliwość konfrontacji rezultatów modeli z diagnozą lekarską oraz weryfikacji zgodności z wiedzą medyczną. To co w metodach AI zawiera się w tzw. czarnej skrzynce i jest niedostępne dla użytkownika w metodach XAI zostaje ujawnione. Na podstawie opracowanych przez autorkę materiałów pomocniczych dla współpracujących z nią lekarzy mogę stwierdzić, że narzędzia w postaci opracowanych tabel i wykresów mają znaczący wkład w postawienie słusznej diagnozy. Dużym osiągnięciem autorki jest zaoferowanie chirurgom modelu wspierającego ich diagnozę w odniesieniu do określenia prawdopodobieństwa 3-miesięcznego przeżycia po operacji nowotworu płuca na podstawie cech pacjenta dostępnych przed operacją. Autorka wyraźnie udowodniła, że globalne metody XAI, przyczyniły się do uchwycenia prawidłowych relacji między zmiennymi objaśniającymi a odpowiedzią modelu (rysunek 3.2 i rysunek 3.3)

Na podkreślenie zasługuje fakt, że Autorka przywiązuje dużą wagę do analizy *lokalnych* metod wyjaśniania (profil *ceteris paribus* i dekompozycja modelu breakdown. Jest to szczególnie ważne z naukach o zdrowiu i w medycynie, gdyż pozwala podejmować zidywidualizowane decyzje (na przykład decyzje o zastosowaniu terapii dla konkretnego pacjenta). Poprzez analizę lokalnych profili możliwe jest wskazanie istotnych cech u pacjenta (rysunek 3.4 i rysunek 3.5). Dużym osiągnięciem Autorki jest przekonanie współpracujących lekarzy o potrzebie zastosowań metod XAI do poprawy jakości leczenia i poprawy wpływu na zdrowie publiczne. Zaproponowane przez Autorkę metody XAI przyczyniają się do poszerzenia wiedzy odnośnie predykcji na podstawie dostępnych cech monitorowanych u pacjenta i poprawiają skuteczność leczenia. W prosty sposób autorka przekazuję użytkownikom AI wiedzę o znaczeniu poszczególnych cech dla predykcji. Rycina dwukolorowa ułatwia lekarzom uświadomienie, które z cech badanych u pacjenta mają negatywny wkład i pogarszają rokowania w porównaniu do średniej predykcji modelu, a które poprawiają rokowania.

Autorka rozprawy podejmuje problem różnorodności i wielowymiarowości danych pojawiający się w zagadnieniach zdrowia publicznego w Rozdziale 4. Osiągnięciem Autorki jest umiejętność tłumaczenia trudnych zagadnień i wyników analiz dokonanych przez AI na język dostępny dla odbiorców. Również dużym osiągnięciem Autorki jest uświadomienie lekarzom potrzeby stosowania modeli matematycznych w połączeniu z metodami XAI. Te osiągnięcia docenione zostały w prestiżowych wydawnictwach. Podrozdziały 3.2 i 3.3 są oparte na pracach [69] i [75] opublikowanych odpowiednio w Applied Science (IF=2.5) i Cells (IF=5.1). Są to wysokiej klasy pisma o profilu interdyscyplinarnym i biologicznym.

W Rozdziale 4, w podrozdziale 4.4.2. Autorka zajmuje się rzadkim zespołem immunologicznym zagrażającym życiu pacjentom. Jako cechę diagnostyczną wybiera czas przeżycia pacjentów, który jest dobrą oceną skuteczności leczenia. Bada dziewięć różnych modeli, 5 modeli rf (lasy losowe), 4 modele gbm (wzmocnienia gradientowego, gradient boosting). Spośród tych modeli wybrała 3 modele najbardziej różniące się od siebie pod

względem zaproponowanej miary PDI. Stwierdziła, że jeden z modeli gbm wykazuje największe oscylacje niezgodne z wiedzą medyczną podczas gdy dwa modele rf są bardziej stabilne. Te spostrzeżenia mają istotną wartość dla lekarzy, gdyż ułatwiają bezpieczniejsze wnioskowanie na podstawie wielu modeli.

Dalsze kierunki podjętej przez Autorkę tematyki, dotyczącej rozwoju wyjaśnialnej sztucznej inteligencji w odniesieniu do zdrowia publicznego i zastosowań medycznych, powinny zmierzać w kierunku wykorzystania technik przetwarzanie języka naturalnego w celu zrozumienia złożonych danych medycznych. Recenzowana praca nie obejmuje tego obiecującego kierunku badań, ale może być przedmiotem dalszych prac Autorki.

3. Ocena redakcyjnej strony pracy

Praca doktorska Kobylińskiej jest obszerna, zawiera bogatą bibliografię i rzetelny przegląd literatury. Analizowane przez Autorkę problemy medyczne i epidemiologiczne są opisane szczegółowo i profesjonalnie.

Omawiane i stosowane w rozprawie metody uczenia maszynowego są przedstawione raczej skrótowo w Rozdziale 2. W założeniu, prezentacja znanych w literaturze narzędzi informatycznych powinna być zrozumiała na poziomie intuicyjnym dla użytkowników tych metod. Użytkownicy oczekują od XAI wyjaśnienia wyników obliczeń bez zagłębiania się w szczegółowy opis działania algorytmów.

Poświęcony opisowi algorytmów Rozdział 2 budzi pewne zastrzeżenia natury redakcyjnej. Przytoczone wzory i definicje są prawdopodobnie zbędne dla czytelnika obeznanego z uczeniem maszynowym, zaś dla niespecjalisty w tej dziedzinie są zbyt abstrakcyjne, niedostatecznie objaśnione i w rezultacie niezrozumiałe. Podam kilka przykładów.

Wzór (2.1) i towarzyszący mu tekst na pewno nie wyjaśni, co to jest drzewo decyzyjne komuś, kto tego wcześniej nie wie.

To samo dotyczy wzoru (2.2). Nie jest objaśnione co to jest x_j, co znaczy $1(x_j)$, w dodatku po prawej stronie występuje mnożenie wektora gamma przez liczbę (?) $1(x_j)$,, więc lewa strona T(teta,x) jest wektorem? Dopiero ze wzorów (2.3)-(2.7) można domyślić się, że T(teta,x) powinno oznaczać predykcję y na podstawie x przez drzewo o parametrach teta.

Niestety, podobne problemy pojawiają się podrozdziale 2.3, w której Autorka objaśnia pojęcia odgrywające kluczową rolę w całości rozprawy, jak profil PDP i ważność zmiennej, mierzona wskaźnikiem VI (variable impotrance). Znaczenie symboli we wzorze (2.12) nie jest objaśnione i Autorka nie zadbała o uzgodnienie oznaczeń w podrozdziale (2.3.2) z tymi w podrozdziale (2.3.1). Można się tylko domyślić, że PD(f,j,z) oznacza to samo, co lewa strona wzoru (2.10) (ale dlaczego napisana inaczej?). Ale co znaczy z_i? Czy to jest wartość atrybutu j dla jednostki i ? Zdanie „[kreska u góry]PD(f, j) jest średnią wartością PDP policzoną po wszystkich zmiennych." jest całkowicie niezrozumiałe. Jeśli „średnia ... po wszystkich zmiennych" ma znaczyć „po wszystkich zmiennych j", to nie powinno zależeć od j.

Powyższe krytyczne uwagi dotyczą tylko redakcji Rozdziału 2 i nie podważają pozytywnej oceny najważniejszych wyników rozprawy, zamieszczonych w Rozdziałach 3 i 4, ale obowiązkiem recenzenta jest również ocena redakcji rozprawy.

4. Ocena wiedzy doktorantki

Doktorantka wykazała się dużą wiedzą z zakresu statystyki, uczenia maszynowego i modelowania. Na pochwałę zasługuje dobra orientacja Autorki rozprawy w zagadnieniach medycznych. Dokonany przez Autorkę rozprawy przegląd literatury jest kompetentny i wyczerpujący.

5. Podsumowanie oceny i moja rekomendacja w sprawie nadania stopnia doktora

Oceniam, że rozprawa doktorska Katarzyny Kobylińskiej zawiera znaczący wkład w nauki o zdrowiu. Z punktu widzenia potencjalnego użytkownika metod zaproponowanych przez Autorkę uważam, że informatyczna i statystyczna zawartość pracy jest wartościowa i zasługuje na pozytywną ocenę.

Wnoszę zatem o dopuszczenie Pani mgr Katarzyny Kobylińskiej do dalszych etapów postępowania doktorskiego.

Marta Zalewska

dr hab. inż. Marta Zalewska