

dr hab. inż. Joanna Domańska
Instytut Informatyki
Teoretycznej i Stosowanej PAN
ul. Bałtycka 5, 44-100 Gliwice

Recenzja rozprawy doktorskiej

mgr Eryki Probierz

pt. "Problematic Internet Use - causes, picture of the phenomenon and its prevention. Interdisciplinary approach"

Cel, zakres i charakter rozprawy

Przedstawiona do recenzji rozprawa doktorska została wykonana pod kierunkiem naukowym prof. Ireny Pilch z Uniwersytetu Śląskiego oraz prof. Adama Gałuszki z Politechniki Śląskiej.

Rozprawa ma charakter interdyscyplinarny. Termin interdyscyplinarność odnosi się do działalności badawczej obejmującej problemy mieszczące się w zakresie zainteresowania więcej niż jednej dyscypliny naukowej. Interdyscyplinarność recenzowanej rozprawy implikuje wykorzystanie do rozwiązania problemów postawionych w rozprawie metod badawczych typowych dla różnych dyscyplin: w niniejszym przypadku jest to psychologia oraz informatyka.

Rozprawa dotyczy problematyki negatywnych skutków korzystania z Internetu, który stał się w obecnych czasach wszechobecnym narzędziem w praktycznie każdej dziedzinie życia. Autorka rozprawy motywuje swoje badania zarówno istotnością zagadnienia Problematicznego Użycia Internetu (PUI), jak również brakiem badań mających na celu utworzenie i przetestowanie skuteczności programu profilaktycznego w oparciu o dane uzyskane w formularzu internetowym.

Celem pracy było przedstawienie aktualnego stanu wiedzy na temat problematycznego korzystania z Internetu, potencjalnych czynników jego występowania, budowa modelu tego zjawiska oraz utworzenie programu profilaktycznego, który mógłby obniżyć ryzyko jego wystąpienia.

W ramach nadrzędnego celu pracy doktorantka wyróżniła trzy bardziej szczegółowe tezy, wykazanie prawdziwości których ułatwiło potwierdzenie osiągnięcia celu pracy.

Pierwsza z nich zakłada, że analiza współzależności zmiennych z problematycznym korzystaniem z Internetu dostarcza istotnych informacji na temat związku tego zjawiska z proponowanymi konstruktami.

Druga teza zakłada, że proponowany program profilaktyczny z możliwością personalizacji pozwala zmniejszyć ryzyko wystąpienia PUI u osób badanych.

Trzecia teza zakłada, że wybrane metody sztucznej inteligencji oraz rozwiązania w zakresie tworzenia stron internetowych pozwalają na częściową automatyzację programu profilaktycznego oraz jego realizację za pośrednictwem Internetu.

Ze względu na interdyscyplinarny charakter rozprawy integralną częścią przeprowadzonych w ramach pracy badań była implementacja rozwiązań z zakresu informatyki, które posłużyły do wykazania prawdziwości trzeciej tezy.

Zawartość rozprawy

Forma drukowana recenzowanej rozprawy obejmuje 213 stron. Praca została napisana w języku angielskim i składa się z wprowadzenia, sześciu rozdziałów, podsumowania pracy, wykazu literatury oraz ośmiu załączników zawierających materiały uzupełniające treść pracy. Tytuł odzwierciedla zawartość pracy, a przyjęty układ pracy jest właściwy.

Wprowadzenie - oprócz motywacji autorki do zajęcia się tematyką opisaną w rozprawie - zawiera również poprawnie sformułowany cel pracy oraz tezy pomocnicze.

Rozdział pierwszy zawiera definicję Problematycznego Użycia Internetu, przedstawia również skalę tego problemu.

Rozdział drugi opisuje stan wiedzy związanej ze wskaźnikami uważanymi za przyczyny występowania PUI. Zmienne te są używane w badaniach opisanych w pierwszej części rozdziału szóstego. Rozdział ten opisuje również problematyczne korzystanie z Internetu w kontekście pandemii Covid-19.

W rozdziale trzecim zawarto przegląd metod związanych z zapobieganiem problematycznemu korzystaniu z Internetu.

Rozdział czwarty opisuje aktualne badania nad PUI celem uwypuklenia nowości proponowanych w pracy rozwiązań.

Rozdział piąty przedstawia metody analizy danych wykorzystywane w badaniach nad PUI.

Rozdział szósty zawiera opis przeprowadzonych w ramach pracy badań. Jest on podzielony na dwie części. Część pierwsza dotyczy predykcji PUI, natomiast część druga opisuje skuteczność zaproponowanego programu prewencyjnego.

Bibliografia zawarta w rozprawie składa się z 262 pozycji. Cytowane w tekście rozprawy pozycje oraz analiza ich zawartości potwierdzają dostateczną znajomość stanu wiedzy doktorantki - związanego z tematyką rozprawy.

Ocena rozprawy

Ze względu na interdyscyplinarny charakter rozprawy poniższa ocena rozprawy dotyczy przede wszystkim wkładu w dyscyplinę Informatyka. Nie dotyczy to zauważonych niedociągnięć edycyjnych, które mogą prowadzić do pogorszenia czytelności rozprawy. Należą do nich m.in.:

- brak tekstu wprowadzającego do niektórych podsekcji, np.: 6.1.4.1, 6.1.4.2, 6.2.4.1, 6.2.4.2,
- zbyt duże zagnieżdżanie podsekcji,
- zbyt małe różnice rozmiaru tytułów poszczególnych stopni zagłębienia podsekcji,
- brak odnośników do niektórych tabel i rysunków,
- brak konsekwencji w nazewnictwie (Chapter One, Chapter 6),

- brak Introduction w spisie treści,
- brak numeracji w wykazie literatury,
- drobne błędy, np.: błędny numer rusunku (Figure 1510), błędne oznaczenie parametru w tabelach (A zamiast α), w tabeli 26 boosting powinno być opisane jako gradient boosting.

Badania związane z dyscypliną Informatyka dotyczyły trzeciej tezy pracy. Teza ta zakładała, że wybrane metody sztucznej inteligencji oraz rozwiązania z zakresu tworzenia stron internetowych pozwolą na częściową automatyzację programu prewencyjnego. Teza ta jest oryginalna, a uzyskane i opisane w pracy wyniki potwierdzają jej prawdziwość.

Zaimplementowany algorytm oraz strona internetowa pozwoliły na realizację programu prewencyjnego on-line. Zgadzam się z doktorantką, że wykorzystanie rozwiązań z zakresu Informatyki miało wpływ na dostępność programu oraz możliwość jego realizacji podczas pandemii Covid-19. Wg mnie przydałaby się jednak poszerzona dyskusja nad wpływem wykorzystania Internetu do badań nad jego problematycznym użyciem. Zgadzam się również z autorką rozprawy w kwestii tego, że wykorzystanie metod uczenia maszynowego do danych związanych z naukami społecznymi pozwala wykazać duży potencjał takiego podejścia. Podczas analizy części rozprawy dotyczącej wykorzystania metod uczenia maszynowego, nasunęło mi się kilka uwag/spostrzeżeń - które przedstawiam poniżej:

- doktorantka prawidłowo wybrała miary oceny modeli klasyfikacji - oprócz najczęściej używanej Accuracy zaprezentowała również wartości takich metryk jak Recall czy Precision; proszę o wyjaśnienie dlaczego autorka nie zaprezentowała macierzy pomyłek, które umożliwiają szerszą ocenę skuteczności algorytmów oraz pozwalają na zaprezentowanie sposobu w jaki modele mylą się pomiędzy poszczególnymi klasami;
- oczekuję szerszego wyjaśnienia czym różnią się techniki uczenia zespołowego bagging oraz boosting (a także gradient boosting), jak również jak wykorzystywany jest w tych algorytmach prosty algorytm ML - drzewa decyzyjne - w pracy jest to opisane zbyt lakonicznie. Proszę wyjaśnić również co oznacza losowość w lasach losowych;
- brakuje mi również wyjaśnienia jakie funkcje straty są wykorzystywane w gradient boosting oraz co jest ważne w takiej funkcji;
- proszę wyjaśnić dlaczego w wynikach regresji ML autorka pracy wykorzystuje loss function Gaussian dla Gradient Boosting Regression, skoro w opisie algorytmu podaje (jako jedyną wymienioną z nazwy, a więc domyślnie najbardziej popularną - jak też jest) MSE. Dlaczego w innych algorytmach są wykorzystywane takie a nie inne funkcje (w zależności od algorytmu: distance, loss, penalty ...);
- dlaczego regresory i klasyfikatory oparte na uczeniu zespołowym (Boosting Regression i Random Forest Regression) posiadają inną liczbę drzew? Jak porównywać je skoro mają zupełnie inny rozmiar?
- dlaczego autorka pracy przy algorytmie KNN wybrała N równe 2;

- dlaczego w tabeli 28 autorka pracy wykorzystuje różne liczby próbek w zbiorze treningowym;
- jak autorka pracy tłumaczy najgorsze wyniki uzyskane dla klasy 4 - general (KNN dla klasy 4 ma recall 0.2, czyli tylko taką część próbek jest w stanie dobrze zaklasyfikować jako klasa 4 spośród wszystkich instancji tej klasy).

Oczekuję od doktorantki ustosunkowania się do powyższych uwag podczas publicznej obrony pracy doktorskiej. Przedstawione powyżej uwagi nie wpływają jednakże na pozytywną ocenę merytorycznej części pracy oraz nie podważają dobrej oceny umiejętności autorki do poprawnego i przekonującego przedstawienia uzyskanych przez siebie wyników.

Wnioski końcowe

Rozprawa doktorska dotyczy ważnego problemu oraz stanowi wartościowe osiągnięcie naukowe Autorki. Dysertacja doktorska powinna również potwierdzać ogólną teoretyczną wiedzę Doktorantki w danej dyscyplinie oraz umiejętność samodzielnego prowadzenia pracy naukowej. Wg mojej opinii przedstawiona do recenzji praca spełnia wymienione wymagania.

Biorąc pod uwagę powyższe stwierdzam, że niniejsza rozprawa spełnia warunki określone w art.13.1 ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz.U. 2017 poz. 1789) i wnioskuję o dopuszczenie Autorki rozprawy do publicznej obrony.

Wojciech