

Kraków, 21 grudnia 2022 r.

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

Recenzja pracy doktorskiej mgr. inż. Jarosława Hryszki

pt.

*Zastosowanie predykcji defektów w komercyjnych projektach
rozwoju oprogramowania — od rozważań teoretycznych
do codziennej praktyki*

Niniejsza recenzja została sporządzona na podstawie pisma Przewodniczącego Rady Dyscypliny Informatyka Techniczna i Telekomunikacja na Uniwersytecie Jagiellońskim prof. dr. hab. inż. Macieja Ogorzałka z 19 października 2022 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. Jarosława Hryszki dotyczy zastosowania algorytmów uczenia maszynowego do predykcji lokalizacji usterek w kodzie źródłowym projektu programistycznego o znacznej skali. Deklarowanym celem ogólnym rozprawy jest *opracowanie metodyki zapewnienia jakości* wykorzystującej tego rodzaju predykcję, *wdrożenie tej metodyki w komercyjnym środowisku wytwarzania oprogramowania i zbadanie jej wpływu na końcową jakość produktu*. Cel został określony jasno i precyzyjnie, niestety nie da się tego powiedzieć o tezie rozprawy. Niemniej jednak z treści dysertacji (zwłaszcza z podsumowania) w dość oczywisty sposób wynika, że teza taka mogłaby brzmieć następująco.

Predykcja lokalizacji defektów przy pomocy uczenia maszynowego prowadzi do wykrywania znacznej ich części we wczesnych etapach rozwoju projektu programistycznego. W konsekwencji pozwala ona skonstruować metodykę zapewniania jakości istotnie podnoszącą jakość produktu bez znaczącego wzrostu kosztów.

Problem zapewnienia jakości produktu z utrzymaniem kosztów produkcji na niskim poziomie ma naturalnie fundamentalne znaczenie w każdej dziedzinie związanej z wytwórczością, w szczególności także w inżynierii oprogramowania. Badania Doktoranta stanowią wartościowy wkład w rozwiązanie tego problemu. Uwzględniają one nie zarówno aspekt naukowy, jak i organizacyjno-wdrożeniowy zastosowania metod predykcji defektów w komercyjnych projektach programistycznych, a mianowicie:

- przygotowanie danych, wybór metody uczenia maszynowego i ewaluację jej wyników w docelowym środowisku,
- wybór metryk oprogramowania określających zadanie predykcji,
- analizę wymaganych elementów środowiska projektowego,
- analizę korzystnych i niekorzystnych cech konkretnych projektów, w tym identyfikację wąskich gardeł i propozycje zmian usuwających te wąskie gardła,
- symulację zmian w kosztach projektu spowodowanych przez implementację predykcji defektów przy użyciu wybranego narzędzia wspierającego,
- wdrożenie predykcji defektów w wybranym projekcie komercyjnym,
- określenie metodyki zapewniania jakości opartej na predykcji defektów o nazwie DPDQA i zaproponowanie sposobu ewaluacji tej metodyki.

Takie całościowe ujęcie tematu jest z pewnością godne podkreślenia. Zaproponowana metodyka jest kompletna i wszechstronnie zweryfikowana w środowisku komercyjnym. Doktorant umiejętnie wykorzystał własne doświadczenie wskazując wszystkie elementy procesu i środowiska projektu komercyjnego, które mają istotny wpływ na powodzenie zaproponowanej metodyki. Wśród mniej oczywistych zwrócić można uwagę na wyżej wspomnianą symulację przewidywanego wpływu DPDQA na koszty zapewnienia jakości, która pozwoliła uzyskać ciekawy efekt „socjologiczny” w postaci przekonania kadry kierowniczej firmy partnerskiej do zastosowania DPDQA w ważnym projekcie komercyjnym. W dalszej konsekwencji umożliwiło to Doktorantowi dokonanie oceny faktycznego wpływu metodyki *a posteriori*. Wyniki tej oceny jednoznacznie wykazują skuteczność zaproponowanego rozwiązania.

W zakresie uczenia maszynowego, Doktorant wybrał do porównania klasyczne metody klasyfikacji (naiwny klasyfikator bayesowski, drzewo decyzyjne, probabilistyczna sieć neuronowa i las losowy). Jest to uwarunkowane użytym narzędziem wspierającym (DePress), nie budzi też wielkich zastrzeżeń wobec całościowego podejścia do rozwiązania problemu. Natomiast treść pracy pozostawia pewne niewyjaśnione kwestie:

- 1) określenie *probabilistyczna sieć neuronowa* oznacza całą rodzinę modeli statystycznych — w pracy brak informacji o architekturze użytej sieci;
- 2) w kolejnych badaniach użyto różnych zestawów modeli — w rozdziale 4 jest to naiwny klasyfikator bayesowski, drzewo decyzyjne i probabilistyczna sieć neuronowa, w rozdziale 5 dochodzi las losowy, w rozdziale 6 z zestawu znika drzewo decyzyjne — jakie jest uzasadnienie dokonanych wyborów?

Innym zauważalnym mankamentem merytorycznym rozprawy są dane zaprezentowane w tabeli 6.5 i oparte na nich rozumowanie opisane na stronach 117 i dalszych. Mianowicie wartości współczynnika korelacji Matthews (MCC) w tej tabeli są policzone błędnie. W wierszach z $TP = FP = 0$ prawidłowa wartość MCC jest nieokreślona (mianownik w formule definiującej jest równy 0), a nie równa -1 . Słusznie zdziwienie Doktoranta budzi wartość MCC osiągnięta przez las losowy połączony ze SMOTE i eliminacją cech, bo przy bardzo przyzwoitych wynikach klasyfikacji ($TP = 3435$, $FP = 147$, $TN = 3343$, $FN = 56$) w tabeli mamy $MCC = 0.0775$, gdzie prawidłowa wartość to około 0.94. Jest to oczywiście pomyłka obliczeniowa, natomiast trudno byłoby sensownie uzasadnić użycie w praktyce klasyfikatora o tak słabej wartości MCC. Na szczęście w tym przypadku zastosowany klasyfikator jest całkiem skuteczny, co potwierdzają kolejne wyniki (z wyjątkiem MCC, który zapewne również w przypadkach prezentowanych w tabeli 6.6 jest policzony niewłaściwie). Natomiast wnioski Doktoranta dotyczące zastosowania metryk opartych na „zapachach kodu” są oparte na niewielkim wzroście MCC i pozostają słuszne po właściwym obliczeniu wartości tego współczynnika.

Podsumowując, pomimo wystąpienia powyższych mankamentów z pewnością można uznać, że opracowanie i wdrożenie w firmie partnerskiej metodyki zapewnienia jakości DPDQA stanowi *oryginalne rozwiązanie w zakresie zastosowania wyników własnych badań naukowych w sferze gospodarczej*, co oznacza, że kryterium ustawowe dotyczące przedmiotu rozprawy (art. 187 ust. 2) zostało spełnione.

Ocena umiejętności samodzielnego prowadzenia pracy naukowej

Osiągnięciem mgr. inż. Jarosława Hryszki 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 w komercyjnym środowisku projektowym. 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. Jarosława Hryszki ukazuje jego szeroką wiedzę teoretyczną w obszarze inżynierii oprogramowania, zwłaszcza w obszarze wykrywania usterek w kodzie i identyfikacji kodu podatnego na błędy. Doktorant wykazał też znajomość klasycznych metod uczenia maszynowego używanych w zagadnieniach klasyfikacji binarnej, w tym technik oceny skuteczności klasyfikatorów. Szczególnie dogłębnie został przez Doktoranta rozpoznany obszar zastosowania uczenia maszynowego w analizie kodu pod kątem wykrywania defektów. Natomiast przedstawiając estymację kosztów implementacji opracowanej metodyki Doktorant wykazał znajomość podstawowych zagadnień z pogranicza informatyki technicznej i nauk ekonomicznych. Podsumowując, uznać można, że kryterium ustawowe dotyczące ogólnej wiedzy teoretycznej Doktoranta (art. 187 ust. 1), również zostało spełnione.

Uwagi edytorskie

Praca pisana jest zasadniczo poprawnym językiem formalnym, nie całkiem jednak wolnym od usterek, np.:

- Doktorant stara się unikać wtrętów obcojęzycznych i żargonowych, ale nie jest w tym zupełnie konsekwentny — zwraca uwagę np. odmieniany przez wszystkie przypadki „commit”, naleciałością żargonową jest też nazywanie naiwnego klasyfikatora bayesowskiego *naiwnym Bayesem*;
- pisownia nazwisk obcojęzycznych w przypadkach bywa niewłaściwa (*McCa-be'a* jest poprawnie, ale zamiast *Halstead'a* powinno być *Halsteada*);
- zamiast *przenaszalność* (str. 113) lepiej napisać *przenośność*.

Ponadto w rozprawie występują inne usterki o charakterze edytorskim, np.:

- rysunek 3.7 jest kompletnie nieczytelny;

- przy rysunku 4.5 brak opisu wyjaśniającego znaczenie obu wykresów (stosowny opis jest w tekście dopiero na następnej stronie);
- referencja [122] jest powtórzeniem referencji [121].
- referencja [157] jest powtórzeniem referencji [156].

Wniosek końcowy

Uwagi krytyczne przedstawione powyżej nie zmieniają ogólnie pozytywnej oceny pracy. Rozprawę doktorską mgr. inż. Jarosława Hryszki 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 Smolke