



dr hab. inż. Rafał Doroz  
Instytut Informatyki, Wydział Nauk Ścisłych i Technicznych  
Uniwersytet Śląski  
ul. Będzińska 39  
41-200 Sosnowiec  
[rafal.doroz@us.edu.pl](mailto:rafal.doroz@us.edu.pl)

Sosnowiec, 25.01.2025

Recenzja rozprawy doktorskiej mgr. inż. Łukasza Chmielowskiego

### **Application of multimodal Neural Networks in solving problem of labeling bug reports**

przygotowanej pod kierunkiem naukowym promotora

prof. dra hab. inż. Roberta Burduka

Promotor pomocniczy: dr inż. Michał Kucharzak

#### **1. Tematyka rozprawy**

Rozprawa doktorska poświęcona jest zagadnieniu automatycznego przypisywania zgłoszeń błędów oprogramowania w dużych organizacjach zajmujących się rozwojem i utrzymaniem systemów informatycznych. Zgłoszenia błędów odgrywają kluczową rolę w procesie zarządzania jakością oprogramowania. Główne wyzwanie, zdefiniowane przez Autora, polega na efektywnym przypisaniu zgłoszenia do właściwej jednostki organizacyjnej odpowiedzialnej za jego rozwiązanie, a także na odpowiedniej klasyfikacji problemu, np. jako dotyczącego bezpieczeństwa, wydajności lub zarządzania pamięcią.

Celem jaki postawił przed sobą Autor rozprawy jest opracowanie oraz ocena skuteczności nowej metody automatycznego przypisywania zgłoszeń błędów, mającej przewyższać efektywność aktualnie stosowanych technik. W ramach badań przeprowadzono analizę literatury, co pozwoliło na identyfikację luk badawczych oraz ograniczeń istniejących metod.

#### **2. Ocena treści rozprawy**

Rozprawa doktorska składa się z ośmiu rozdziałów oraz trzech dodatków, obejmujących wprowadzenie, szczegółowy opis przeprowadzonych badań, prezentację wyników, a także omówienie wdrożeń i perspektyw dalszych badań. Struktura pracy opiera się na artykułach naukowych, przy czym jeden z nich, będący podstawą Rozdziału 6, nie został dotychczas

opublikowany.

WPLYNEŁO  
18-02-2025

RDN-III/75/2025



**Rozdział 1** wprowadza czytelnika w tematykę badań, przedstawiając kluczowe zagadnienia teoretyczne i metodologiczne. Autor omówił w nim kontekst badawczy, określił problem oraz uzasadnił jego istotność. Wskazał także główne wyzwania stojące przed realizacją pracy. Autor identyfikuje obszary wymagające dalszych badań, formułuje pytania badawcze oraz określa cele rozprawy. Istotnym elementem rozdziału jest również przegląd aktualnego stanu wiedzy w danej dziedzinie, uwzględniający najnowsze publikacje i stosowane rozwiązania. Na zakończenie przedstawiono metody badawcze, które posłużyły do realizacji analiz i eksperymentów opisanych w dalszych częściach pracy.

**Rozdział 2** analizuje wpływ metod wstępnego przetwarzania i wektoryzacji zgłoszeń błędów oprogramowania na dokładność ich przypisywania. Autor bada znaczenie takich technik jak stemming, lematyzacja oraz różne konfiguracje  $n$ -gramów w modelu TF-IDF.

Atutem pracy są praktyczne rekomendacje dotyczące optymalizacji przetwarzania tekstu oraz analiza liczby cech mających wpływ na jakość proponowanego modelu. Należy jednak zauważyć, że zawężenie analizy do ograniczonego zestawu metryk oceny oraz brak testów istotności statystycznej sprawia, że interpretacja wyników może być nieco mniej precyzyjna. Podsumowując, praca wnosi wkład w obszar klasyfikacji zgłoszeń błędów, jednak jej rezultaty wymagają bardziej dogłębnej analizy statystycznej oraz uwzględnienia szerszego zakresu miar oceny.

**Rozdział 3** opisuje metodę przypisywania zgłoszeń błędów oprogramowania do odpowiednich grup programistów lub analityków, w której kluczową rolę odgrywa wykorzystanie struktury organizacyjnej przedsiębiorstwa. Analizowana jest skuteczność automatyzacji w porównaniu do decyzji podejmowanych przez ludzi, a także czynniki wpływające na efektywność tego procesu. Przedstawiono również minimalne wymagania wdrożeniowe oraz wpływ czynników ludzkich na podejmowanie decyzji. Zaproponowane rozwiązanie łączy predykcje modelu z decyzjami użytkowników, co umożliwi częściową automatyzację przy zachowaniu nadzoru człowieka. Autor wskazuje potencjalne korzyści, takie jak odciążenie zespołów i skrócenie czasu obsługi zgłoszeń, jednocześnie zwracając uwagę na wyzwania związane z przejrzystością działania modeli ML.

Choć zaprezentowane podejście jest interesujące, jego opis pozostaje dość ogólny. Brakuje szczegółowej analizy zastosowanych modeli i kryteriów oceny skuteczności, a omówienie wdrożenia i wpływu automatyzacji na procesy organizacyjne jest jedynie zarysowane.

Podsumowując, rozdział podejmuje istotne zagadnienia związane z automatyzacją przypisywania zgłoszeń błędów, jednak wymaga bardziej szczegółowego opracowania, zwłaszcza w zakresie metodologii i wyników badań.



**Rozdział 4** przedstawia innowacyjne podejście do oceny modeli uczenia maszynowego stosowanych w przypisywaniu zgłoszeń błędów oprogramowania, podkreślając kluczowe znaczenie zależności czasowych w procesie ewaluacji. Autor wskazuje na ograniczenia tradycyjnych metod, takich jak losowe podziały danych czy walidacja krzyżowa, które mogą prowadzić do nierealistycznych wyników i zawyżonej oceny skuteczności modeli. Zaproponowana metoda konstruuje zbiory treningowe i testowe w oparciu o chronologię zgłoszeń i lepiej odwzorowuje rzeczywiste warunki w jakich pracują systemy zarządzania błędami. Mocną stroną rozdziału jest dobrze uargumentowana konieczność uwzględnienia zależności czasowych oraz wykazanie, że nowa metoda lepiej odwzorowuje rzeczywiste działanie systemów ML. Słabością natomiast pozostaje ograniczona analiza wpływu różnych parametrów oraz brak szerszego zestawu metryk oceny.

**Rozdział 5** analizuje wykorzystanie metod XAI w automatyzacji klasyfikacji błędów, podkreślając ich potencjał w usprawnianiu procesu przypisywania zgłoszeń do odpowiednich zespołów. Automatyzacja zarządzania błędami jest kluczowa ze względu na wysokie koszty ręcznej analizy, a wprowadzenie metod XAI może dodatkowo zwiększyć przejrzystość podejmowanych decyzji. Autor ocenił skuteczność XAI w kontekście klasyfikacji błędów na podstawie raportów tworzonych przez testerów i użytkowników. Badania przeprowadzono na dwóch zestawach danych: pierwszym dotyczył klasyfikacji błędów bezpieczeństwa w firmie telekomunikacyjnej, natomiast drugi obejmował podział zgłoszeń na kategorie takie jak awarie, problemy z pamięcią, wydajnością i bezpieczeństwem. Wyniki eksperymentów wykazały, że stosowanie XAI nie powoduje degradacji dokładności predykcji w porównaniu do klasycznych metod typu „czarna skrzynka”, a jednocześnie pozwala generować reguły decyzyjne ułatwiające interpretację wyników.

Mocną stroną rozdziału jest podkreślenie unikalności podejścia – dotychczas nie stosowano XAI w kontekście przypisywania zgłoszeń błędów, mimo że istnieją badania dotyczące przewidywania defektów oprogramowania, które często są błędnie utożsamiane z tym zagadnieniem. W pracy zaprezentowano eksperymentalne porównanie standardowych metod klasyfikacji z metodami XAI, wskazując na ich praktyczną przydatność.

Podsumowując, rozdział wnosi wkład w rozwój inteligentnych systemów wspierających zarządzanie błędami, wskazując, że XAI może skutecznie wspierać klasyfikację zgłoszeń bez utraty jakości predykcji.

**Rozdział 6** koncentruje się na zastosowaniu multimodalnych sieci neuronowych w procesie klasyfikacji zgłoszeń błędów, prezentując podejście integrujące różne typy danych wejściowych.



Autorzy analizują problem automatycznego przypisywania zgłoszeń do odpowiednich zespołów deweloperskich, co jest kluczowe dla efektywnego zarządzania błędami w oprogramowaniu.

W części eksperymentalnej rozdziału Autor porównał skuteczność multimodalnych sieci neuronowych z tradycyjnymi metodami, takimi jak w pełni połączone sieci neuronowe. Wyniki wskazują, że integracja różnych modalności danych pozwala osiągnąć wyższą dokładność predykcji, zwłaszcza w przypadku modeli wykorzystujących dane systemowe i kategoryczne. Autor zauważa jednak, że efektywność tego podejścia zależy od jakości i kompletności dostępnych danych – w przypadku brakujących lub niespójnych informacji skuteczność modeli jest niższa. Dodatkowo przeprowadzona została analiza statystyczna, potwierdzająca istotne różnice między wynikami uzyskanymi przez multimodalne sieci neuronowe a klasycznymi metodami referencyjnymi. Interesującym elementem rozdziału jest również zastosowanie automatycznych enkoderów do redukcji nadmiernego dopasowania i poprawy stabilności modelu.

Podsumowując, rozdział stanowi istotny wkład w rozwój metod zarządzania błędami w oprogramowaniu, pokazując, że wykorzystanie multimodalnych sieci neuronowych może znacząco poprawić skuteczność klasyfikacji zgłoszeń.

**Rozdział 7** przedstawia architekturę, środowisko badawcze i proces orkiestracji systemu klasyfikacji raportów o błędach, koncentrując się na technicznych aspektach integracji w środowisku przedsiębiorstwa. Omówiono mechanizmy gromadzenia, selekcji i przechowywania informacji. Następnie opisano obsługę modeli predykcyjnych, w tym ich analizę, klasyfikację, wstępne przetwarzanie i filtrowanie danych, a także sposób przydzielania zasobów sprzętowych i wdrażania w środowisku produkcyjnym.

Dalsza część rozdziału dotyczy procesu ponownego trenowania modeli oraz działań wdrożeniowych, takich jak instalacja certyfikatu bezpieczeństwa, usuwanie zbędnych plików oraz konfiguracja narzędzi automatyzujących procesy. Całość podkreśla znaczenie skutecznego zarządzania danymi, automatyzacji i integracji systemu w rzeczywistych warunkach pracy.

**Rozdział 8** to podsumowanie wyników. Głównym celem pracy było opracowanie metody umożliwiającej automatyczne przypisywanie zgłoszeń błędów do odpowiednich grup deweloperskich. W ramach badań analizowano także przypisywanie raportów do kategorii związanych z bezpieczeństwem oraz możliwość zastosowania metod wyjaśnialnej sztucznej inteligencji. Kluczowym zagadnieniem było wykorzystanie multimodalnych sieci neuronowych do poprawy dokładności klasyfikacji. Wyniki eksperymentów, opartych na danych pochodzących z

tytułów, opisów oraz dzienników systemowych, wykazały przewagę nowego podejścia nad klasycznymi metodami.

Autor potwierdził postawioną tezę zakładającą, że istnieje metoda automatycznego przypisywania raportów o błędach, przewyższająca stosowane dotychczas rozwiązania. W badaniach porównano wyniki predykcji generowanych przez modele uczenia maszynowego z decyzjami ludzkich ekspertów, wskazując na znaczącą poprawę dokładności dzięki zastosowaniu nowoczesnych metod.

Autor podkreśla, że dodatkowym osiągnięciem pracy było odkrycie istotnych aspektów procesu zarządzania własnością intelektualną w firmie, co doprowadziło do przygotowania kilku wewnętrznych zgłoszeń patentowych. Jeden z projektów przeszedł analizę pod kątem nowości i został opublikowany w literaturze naukowej. Co więcej, wybrane elementy systemu zaprojektowanego w ramach badań zostały wdrożone w środowisku produkcyjnym dużego przedsiębiorstwa technologicznego. Stopniowa implementacja rozwiązania pozwoliła na pełną automatyzację przypisywania raportów o błędach do odpowiednich działów, co znacząco usprawniło proces ich obsługi.

Podsumowując, rozdział podkreśla wartość naukową i praktyczną przeprowadzonych badań. Opracowane rozwiązania przyczyniły się do poprawy efektywności klasyfikacji zgłoszeń błędów oraz stanowią istotny wkład w rozwój metod uczenia maszynowego w tej dziedzinie.

Dołączone do pracy aneksy to:

**Aneks A:** Szablon zgłoszenia błędu, wykorzystywany w badaniach.

**Aneks B:** Przewodnik użytkownika dotyczący dostępu do predykcji przypisywania zgłoszeń błędów oprogramowania opartych na uczeniu maszynowym.

**Aneks C:** Potwierdzenie wdrożenia wybranych elementów pracy doktorskiej w przedsiębiorstwie.

Praca stanowi ciekawe i kompleksowe opracowanie nowatorskich metod wspierających proces przypisywania zgłoszeń błędów w oparciu o techniki uczenia maszynowego, multimodalne sieci neuronowe oraz wyjaśnialną sztuczną inteligencję, jednocześnie dostarczając praktycznych rozwiązań wdrożonych w rzeczywistym środowisku biznesowym.



### 3. Ocena oryginalności rozprawy

Głównym celem rozprawy było opracowanie metody automatycznego przypisywania raportów o błędach w oprogramowaniu do odpowiednich grup deweloperskich, co miało na celu usprawnienie procesu klasyfikacji zgłoszeń i poprawę efektywności ich analizy na różnych poziomach organizacyjnych. Istotnym aspektem badań było także przypisywanie raportów do kategorii związanych z bezpieczeństwem, co pozwala na lepsze zarządzanie priorytetami w procesie rozwoju oprogramowania. Autor podjął się również analizy potencjalnych zastosowań metod wyjaśnialnej sztucznej inteligencji (XAI) w kontekście klasyfikacji zgłoszeń błędów, co pozwala na lepsze zrozumienie decyzji modelu przez użytkowników końcowych.

Aby osiągnąć zakładane cele, autor zbadał możliwość zastosowania multimodalnej sieci neuronowej jako narzędzia do automatycznego przypisywania raportów do właściwych zespołów odpowiedzialnych za ich rozwiązanie. W tym celu przeprowadzono szereg eksperymentów, w których analizowano skuteczność różnych metod klasyfikacji. Dane wejściowe obejmowały zarówno tytuły i opisy zgłoszeń błędów, jak i informacje pochodzące z dzienników systemowych. W celu reprezentacji tekstowych zastosowano metody przetwarzania języka naturalnego.

Przeprowadzone eksperymenty potwierdziły skuteczność zaproponowanego podejścia. W szczególności zastosowanie multimodalnej sieci neuronowej znacząco poprawiło dokładność klasyfikacji w porównaniu do metod referencyjnych. W badaniach porównano wyniki klasyfikacji uzyskiwane przez model uczenia maszynowego z wynikami klasyfikacji dokonywanej przez ekspertów, co pozwoliło ocenić, na ile rozwiązanie może zastąpić manualny proces analizy zgłoszeń błędów. Wyniki te potwierdziły, że opracowane podejście jest w stanie przewyższyć tradycyjne metody przypisywania raportów, co stanowi znaczący postęp w dziedzinie automatycznej analizy błędów w oprogramowaniu.

Dodatkowym aspektem analizy było uwzględnienie zależności czasowych w procesie klasyfikacji raportów o błędach. Zwrócono uwagę na wyzwania związane z duplikacją zgłoszeń oraz czasem potrzebnym na ich rozwiązanie w dynamicznie rozwijających się systemach. Badania pokazały, że ograniczenia wynikające z konieczności trenowania modeli na historycznych danych mogą mieć wpływ na ich skuteczność, co należy uwzględnić w przyszłych pracach nad optymalizacją algorytmów.

Jednym z kluczowych osiągnięć pracy było nie tylko opracowanie teoretycznych rozwiązań, ale także ich praktyczna implementacja. W wyniku badań opracowane metody zostały wdrożone w



systemach dużego przedsiębiorstwa technologicznego, co stanowi istotne potwierdzenie ich użyteczności w rzeczywistych zastosowaniach przemysłowych.

Dodatkowo, autor pracy przyczynił się do rozwoju zarządzania własnością intelektualną w badanej firmie. W trakcie badań zgłoszono trzy wewnętrzne wnioski patentowe dotyczące opracowanych metod klasyfikacji raportów błędów. Jeden z nich został poddany szczegółowej analizie pod kątem nowości i innowacyjności, co doprowadziło do jego publikacji w renomowanym czasopiśmie naukowym. Pozostałe wyniki badań również znalazły swoje miejsce w prestiżowych publikacjach, w tym w monografii wydanej przez Springer oraz w materiałach międzynarodowej konferencji naukowej.

Przedstawiona rozprawa stanowi cenny wkład w rozwój metod automatycznej klasyfikacji raportów o błędach w oprogramowaniu. Autorowi udało się nie tylko opracować nowatorskie podejście, ale także potwierdzić jego skuteczność w warunkach rzeczywistych oraz wdrożyć je w systemach przemysłowych. Wyniki badań wskazują na wyraźną poprawę efektywności przypisywania raportów w porównaniu do metod tradycyjnych, co ma duże znaczenie w kontekście usprawnienia zarządzania zgłoszeniami błędów w organizacjach.

#### **4. Uwagi dyskusyjne i polemiczne**

Nie kwestionując wartości wyników zawartych w rozprawie, poniżej przedstawiam kilka krytycznych uwag, spostrzeżeń które mogą przyczynić się do dalszej poprawy jakości pracy.

##### Rozdział 2

- Ze względu na tajemnicę handlową nie podano liczby próbek użytych do testów. Utrudnia to ocenę skuteczności metody. Brakuje informacji, czy zbiór danych był zbalansowany pod względem klas? Jeśli nie, to jakie kroki podjęto, aby zniwelować potencjalne problemy związane z niezbalansowanym zestawem danych?
- Dlaczego jako główną metodę wektoryzacji wybrano TF-IDF? Czy rozważano inne podejścia?
- Po wektoryzacji zastosowano Logistic Regression – dlaczego wybrano akurat ten model? Czy testowano inne klasyfikatory?
- Czy przeprowadzono testy statystyczne, aby potwierdzić, czy różnice między metodami są istotne, czy wynikają z losowych fluktuacji w danych?



### Rozdział 3

- W jaki sposób oceniano poprawność predykcji ludzkich? Czy były one analizowane pod kątem spójności między różnymi osobami dokonującymi przypisań?
- Jakie czynniki wpływają na różnice między predykcjami modelu a decyzjami ludzkimi? Czy analizowano wpływ np. doświadczenia pracowników na jakość przypisań?
- Jakie były najczęstsze przypadki błędnych predykcji modelu i jak można je poprawić?
- Czy analizowano wpływ zmieniającej się struktury organizacyjnej na jakość predykcji?

### Rozdział 4

- Czy przeprowadzono testy statystyczne dla potwierdzenia istotności różnic między metodami?
- Czy przeprowadzono testy z bardziej zaawansowanymi modelami, np. opartymi na sieciach neuronowych, które mogą uwzględniać czas jako zmienną wejściową?
- Brak informacji o balansie klas – czy dane były zrównoważone? Jakie metody zastosowano w przypadku klasyfikacji nierównomiernie rozłożonych klas?
- Niejasność w metodzie przesuwania okna czasowego – skoro testy obejmowały 10 miesięcy, jak możliwe było przesuwanie okna o jeden miesiąc w każdej serii? Czy nie zabrakło danych w ostatnich seriach?

### Rozdział 5

- Analiza XAI ogranicza się do drzew decyzyjnych. Czy brano pod uwagę inne techniki interpretowalności?
- Brak analizy interakcji użytkownika z XAI – czy testerzy ufają wyjaśnieniom generowanym przez model? Jakie są konsekwencje decyzji opartych na tych wyjaśnieniach?

### Rozdział 6

- Czy rozważano inne potencjalne sposoby radzenia sobie z brakującymi danymi, zamiast ich całkowitego wykluczenia?
- Brak testów statystycznych porównujących modele – czy różnice w wynikach są istotne statystycznie?
- Brak analizy generalizacji modelu – czy wyniki można przenieść na inne branże i firmy?

## 5. Ocena strony formalnej rozprawy

Rozprawa została napisana w języku angielskim i stanowi opis cyklu publikacji naukowych, na których bazuje cała praca. Wszystkie artykuły, na podstawie których powstała rozprawa, również zostały opublikowane w języku angielskim. Układ pracy jest czytelny, co ułatwia śledzenie kolejnych etapów badań oraz wniosków wynikających z przeprowadzonych eksperymentów. Struktura pracy

została zachowana zgodnie z przyjętymi standardami akademickimi, a przedstawione zagadnienia są konsekwentnie rozwijane w kolejnych rozdziałach.

Niestety, strona językowa rozprawy budzi istotne zastrzeżenia. W tekście pojawia się bardzo wiele błędów gramatycznych, składniowych oraz leksykalnych, które utrudniają jej odbiór i negatywnie wpływają na czytelność pracy. W wielu miejscach autor stosuje nieprecyzyjne sformułowania oraz konstrukcje gramatyczne, które odbiegają od standardów języka akademickiego.

W całej rozprawie doktorskiej można zauważyć wiele usterek, które mogą wpływać na czytelność i interpretację wyników. Przykładowo, w niektórych miejscach wykresy i tabele mogą prowadzić do błędnej interpretacji danych – Fig. 4.1 zawiera odwrócone kolory zbiorów, co może być mylące, a Tabele 4.5 i 4.6 mają identyczny podpis, mimo że prezentują różne dane. Warto zadbać o konsekwentne i precyzyjne opisy, które jednoznacznie wskazują, jakie informacje są przedstawiane.

## **6. Konkluzja końcowa**

Rozprawa potwierdza szeroką wiedzę teoretyczną Doktoranta w dyscyplinie *Informatyka Techniczna i Telekomunikacja*, zwłaszcza w obszarze uczenia maszynowego i analizy danych tekstowych. Autor wykazał się znajomością nowoczesnych algorytmów klasyfikacji, w tym metod multimodalnych, a także umiejętnością prowadzenia eksperymentów oraz stosowania odpowiednich narzędzi do analizy wyników. Wykonane badania stanowią próbę rozwiązania istotnego problemu automatycznej klasyfikacji raportów o błędach w oprogramowaniu i ich przypisywania do odpowiednich zespołów deweloperskich.

Praca wnosi oryginalny wkład w rozwój metod automatycznej analizy zgłoszeń błędów, pomimo, że pewne jej aspekty wymagają dopracowania. W szczególności można zauważyć brak bardziej szczegółowej analizy porównawczej z innymi współczesnymi metodami stosowanymi w przemyśle oraz ograniczony zakres ewaluacji modeli w rzeczywistych warunkach produkcyjnych. Zaproponowane podejście ma istotne znaczenie dla automatyzacji zarządzania błędami w oprogramowaniu i może znaleźć zastosowanie w praktyce przemysłowej.

Strona formalna rozprawy, mimo dobrze przemyślanej struktury, budzi zastrzeżenia. Liczne błędy językowe, zarówno gramatyczne, jak i składniowe, mogą wpływać na odbiór pracy przez czytelników anglojęzycznych i obniżać jej wartość naukową. Prezentacja wyników jest poprawna,



choć momentami brakuje bardziej przejrzystych uzasadnień dla wybranych podejść oraz bardziej kompleksowej interpretacji uzyskanych rezultatów w kontekście aktualnego stanu wiedzy.

Uwagi krytyczne nie umniejszają wartości rozprawy, lecz mają na celu podkreślenie aspektów, które mogą zostać ulepszone w przyszłości. Pomimo wskazanych niedoskonałości, praca stanowi istotny wkład w rozwój metod klasyfikacji raportów o błędach i ich automatycznego przypisywania. Wobec powyższego stwierdzam, że rozprawa doktorska **spełnia wszystkie wymagania stawiane przez Prawo o szkolnictwie wyższym i nauce wobec rozpraw doktorskich. Wnoszę o dopuszczenie dysertacji do dalszych etapów przewodu doktorskiego, w tym do jej publicznej obrony.**

Rafał Doroz