

dr hab. inż. Robert Burduk, prof. uczelni
Politechnika Wrocławska
Wydział Elektroniki
Ul. Wybrzeże Stanisława Wyspiańskiego 27
50-370 Wrocław

Wrocław, dnia 4.05.2020 r.

**RECENZJA ROZPRAWY DOKTORSKIEJ
DLA RADY WYDZIAŁU TELEKOMUNIKACJI, INFORMATYKI
I ELEKTROTECHNIKI
UNIWERSYTETU TECHNOLOGICZNO-PRZYRODNICZEGO
W BYDGOSZCZY**

Tytuł rozprawy: **Zastosowanie metod uczenia maszynowego do wykrywania ataków sieciowych**

Autor rozprawy: **mgr inż. Marek Pawlicki**

1. Jakie zagadnienie naukowe jest rozpatrzone w pracy (teza rozprawy) i czy zostało ono dostatecznie jasno sformułowane przez autora?

Zakres recenzowanej rozprawy dotyczy dynamicznie rozwijającej się i aktualnej tematyki cyberbezpieczeństwa. W dysertacji Doktorant koncentruje się na dwóch zagadnieniach związanych z wykrywaniem ataków sieciowych. Pierwszy z wątków dotyczy problematyki zaburzenia poufności, integralności lub dostępności systemu, drugi wątek wykrywania ataków na algorytmy uczenia maszynowego używane w systemie wykrywania ataków sieciowych. Autor wyraźnie podzielił pracę na wspomniane powyżej dwa nurty, które jednoznacznie wpisują się w zaproponowany tytuł rozprawy. W rozdziale *Wstęp* został zdefiniowany cel pracy, który związany jest z autorskimi propozycjami metod wykrywania ataków sieciowych w ramach dwóch nurtów pracy. W rozdziale tym została również postawiona teza pracy, która zakłada, że zaproponowane przez Autora metody pozwolą na skuteczną detekcję ataków sieciowych. Teza pracy postawiona jest prawidłowo, jednak drobnym mankamentem jest fakt, iż odnosi się ona do ataków typu przeciwnego uczenia maszynowego (*adversarial machine learning*). Zaproponowana przez Autora metoda opisana w Rozdziale 6 dotyczy ataków typu unikania (*evasion*), czyli tylko podtypu ataków wspomnianych w tezie pracy.

2. Jaka jest przydatność rozprawy z punktu widzenia nauk technicznych, czy założenia przyjęte przez autora są uzasadnione?

Tematyka rozprawy mieści się w dyscyplinie naukowej Informatyka Techniczna i Telekomunikacja. W szczególności, co wcześniej wspomniano, Autor koncentruje się na dwóch problemach wykrywania ataków sieciowych. Z punktu widzenia nauk technicznych przydatność rozprawy dotyczy przede wszystkim sformułowania detektora ataków typu *evasion* oraz struktury szeregowej złożonej z bloków czynności wykorzystywanych w uczeniu maszynowym, którego celem jest wykrywanie ataków sieciowych. Opracowane przez Autora metody zostały zweryfikowane na drodze eksperymentu komputerowego, w którym użyto referencyjnych baz danych o ugruntowanej pozycji w społeczności zajmującej się uczeniem maszynowym i atakami sieciowymi. Założenia przyjęte przez Autora są uzasadnione, dotyczy to w szczególności takich aspektów jak: niezbalansowanie danych dotyczących ataków sieciowych, optymalizacja hiperparametrów w algorytmach uczenia maszynowego czy też wykrywanie ataków na modele uczenia maszynowego wykorzystywane w detekcji ataków sieciowych. Wymienione założenia dotyczą z jednej strony aktualnych problemów badawczych, a z drugiej znajdują odzwierciedlenie w problematyce cyberbezpieczeństwa.

Tematyka rozprawy jest interesująca, w pełni uzasadniona i posiada potencjał wykorzystania w rzeczywistych problemach dotyczących cyberbezpieczeństwa. Należy podkreślić, że zakres rozprawy uwzględnia aktywny udział Autora w międzynarodowych projektach badawczych: H2020 Infracress oraz H2020 Sparta.

3. Czy autor rozwiązał postawione zagadnienie i czy użył właściwej do tego metody?

Autor rozwiązał zagadnienie zdefiniowane w tezie pracy. W tym celu sformułował własne, autorskie koncepcje dotyczące struktury szeregowej bloków służących do budowy systemu sztucznej inteligencji w wykrywaniu ataków sieciowych, badań eksperymentalnych dotyczących problemu niezbalansowania danych w wspomnianym typie ataków oraz propozycji detektora ataków na metody uczenia maszynowego. Zastosowana metodologia sprowadza się do następujących sekwencji: zdefiniowanie problemu badawczego, propozycja rozwiązania postawionego problemu, zdefiniowanie środowiska eksperymentowania z wykorzystaniem adekwatnych do postawionego problemu baz danych, analiza otrzymanych wyników z wykorzystaniem różnego rodzaju wskaźników jakości klasyfikacji. Tego typu postępowanie jest typowym podejściem stosowanym przez społeczność naukową w analizie

zagadnień związanych z uczeniem maszynowym, tak więc zaprezentowane w dysertacji podejście jest prawidłowe i właściwe w kontekście tematyki pracy.

4. Jaki charakter ma rozprawa (teoretyczny, doświadczalny, konstrukcyjny), jaka jest jej pozycja w stosunku do stanu wiedzy czy poziomu techniki reprezentowanych przez literaturę światową?

Recenzowana praca ma charakter koncepcyjno-eksperymentalny. Autor zaproponował kilka rozwiązań problemu badawczego dotyczącego zwiększenia skuteczności metod uczenia maszynowego wykrywania ataków sieciowych oraz wrogich ataków na wspomniane metody. Uzyskane przez Autora rezultaty potwierdzają postawioną na wstępie pracy tezę badawczą, która została udowodniona w sposób eksperymentalny. Dysertacja zawiera niezbędne elementy eksperymentu komputerowego, które pozwalają na ocenę jakości metod uczenia maszynowego, a poruszana tematyka jest aktualna w stosunku do stanu wiedzy dotyczącego wykorzystania metod sztucznej inteligencji w zagadnieniach cyberbezpieczeństwa.

5. Na czym polega oryginalność rozprawy, co stanowi samodzielny i oryginalny dorobek autora?

Wkład Autora w rozwój metod związanych z obroną przed cyberatakami polega na:

- opracowaniu struktury szeregowej metody wykrywania ataków sieciowych wykorzystującej jedną z architektur rekurencyjnych sieci neuronowych (*Gated Recurrent Unit*) oraz modułu balansowania danych,
- zbadaniu wpływu metod balansowania danych za skuteczność metod uczenia maszynowego wykorzystywanych w analizie ruchu sieciowego,
- zbadaniu wpływu optymalizacji hiperparametrów sztucznych sieci neuronowych na skuteczność tych sieci w analizie ruchu sieciowego,
- opracowaniu metody przeciwnego uczenia maszynowego (*adversarial machine learning*), a w szczególności wykrywania ataków typu unikania (*evasion*) wykorzystującej jako klasyfikator sztuczną sieć neuronową,
- eksperymentalnej weryfikacji opracowanych metod z wykorzystaniem referencyjnych zbiorów danych (CICIDS2017, NSL-KDD).

6. Jakie są słabe strony rozprawy i jej główne wady?

W opisie eksperymentów dotyczących niezbalansowania zbioru danych Autor używa wskaźnika *support*. W pracy nie ma jednoznacznej definicji tego wskaźnika, poza krótką wzmianką na stronie 57. Wprowadzenie wskaźnika niezbalansowania danych (np. stosunku

liczby obiektów z klasy mniejszościowej do liczby obiektów z klasy większościowej) pozwoliło by na analizę niezależną od liczby obiektów, którą wyraża używany wskaźnik *support*. Dodatkowo nazwa *support* występuje w problematyce dotyczącej reguł asocjacyjnych, co bez precyzyjnej definicji tego wskaźnika może budzić pewne niejasności.

Autor dysertacji podczas analizy wyników oraz w ich przedstawieniu w formie tabelarycznej wykorzystuje trzy miary jakości klasyfikacji: *precision*, *recall* oraz *F1-score*. Miara *F1-score* jest średnią harmoniczną dwóch pozostałych, które z punktu widzenia macierzy pomyłek nie uwzględniają wyniku klasyfikacji prawdziwie negatywnej (*true negative*). Nie uwzględnienie wyniku klasyfikacji typu *true negative* powinno być wyjaśnione przez Autora w tekście pracy. Kontynuując ten wątek, w rozprawie nie użyto innych miar jakości klasyfikacji, w szczególności dedykowanych do danych niezbalansowanych takich jak: *balanced accuracy* lub *G-mean*.

Badania eksperymentalne dotyczące detektora ataków typu unikania (*evasion*) odnoszą się tylko do jednej architektury sztucznej sieci neuronowej wykorzystanej do uczenia modelu oraz jednej architektury głębokiej sieci neuronowej wykorzystanej do identyfikacji modelu uzyskanego w trakcie uczenia.

Analiza wyników zawarta w punkcie 6.2 wykonana jest w wykorzystaniu miary jakości klasyfikacji jaką jest dokładność (*accuracy*), w dalszych częściach rozdziału 6 wykorzystywane są miary *precision*, *recall* oraz *F1-score*, co jest pewną niekonsekwencją.

Recenzowana rozprawa napisana jest poprawnie pod względem językowym, stylistycznym oraz redakcyjnym. Niemniej jednak w pracy można znaleźć błędy językowe oraz redakcyjne, takie jak:

- str. 40 oraz 97 – błędne odwołanie do rysunków,
- str. 41 – „...problemu niezbalansowane danych liczba ...”,
- miary jakości klasyfikacji wymienione są na stronie 46, powtórnie wymienione i zdefiniowane na stronie 56,
- str. 67 – „w trakcie ...”, powinno być dużą literą (podobnie na str. 71),
- użycie kropki (.) jako separatora dziesiętnego zamiast przecinka (,),
- oraz inne drobne błędy np. brak spacji przy odnośnikach (str. 45, 71, 97, 104, 107).

Należy podkreślić, że przytoczone powyżej błędy językowe oraz redakcyjne nie pomniejszają wartości naukowej oraz oryginalności rozprawy.

7. Czy rozprawa świadczy o dostatecznej wiedzy autora i znajomości współczesnej literatury z zakresu dyscypliny naukowej, jakiej rozprawa dotyczy?

Rozdziały nr 2 oraz 5 opisują kolejno stan wiedzy dotyczący wykrywania ataków w ruchu sieciowym oraz problematykę osłabiania i dezinformowania algorytmów sztucznej inteligencji wykorzystywanych w systemach wykrywania ataków sieciowych. Przedstawiona w wymienionych rozdziałach treść wskazuje, że Autor rozprawy posiada wiedzę teoretyczną, która dotyczy omawianej w pracy problematyki i mieści się w aktualnym nurcie badań związanych z systemami wykrywania ataków sieciowych, w szczególności z zagadnieniami wykorzystania uczenia maszynowego przed cyberzagrożeniami. Treść tych rozdziałów odnosi się do aktualnych problemów cyberbezpieczeństwa i odpowiednio wprowadza czytelnika do zagadnień, które w dysertacji stanowią oryginalny wkład Autora.

Spis literatury liczy 137 pozycji, w tym dwie współautorskie. Dodatkowych pięć prac współautorskich Autora dysertacji wymienionych jest rozdziale *Wstęp*. Cytowane prace dobrane są prawidłowo, odnoszą się do omawianych w pracy problemów oraz świadczą o umiejętności korzystania z istniejącej literatury. Na podkreślenie zasługuje fakt, że ponad połowa pozycji pochodzi z lat 2015-2019.

8. Czy autor wskazał umiejętność poprawnego i przekonującego przedstawienia uzyskanych przez siebie wyników (zwięzłość, jasność, poprawność redakcyjna rozprawy)?

Autor wykazał umiejętności poprawnego i przekonującego przedstawienia autorskich metod oraz ich eksperymentalnej weryfikacji z wykorzystaniem referencyjnych zbiorów danych. Styl prezentacji jest zrozumiały (wyniki przedstawione są w formie tabel). Wyniki badań eksperymentalnych zostały również odpowiednio skomentowane w treści pracy. Znalezione błędy redakcyjne zostały wyszczególnione w punkcie 6 niniejszej recenzji.


9. Czy i jaka jest przydatność rozprawy dla gospodarki narodowej?

Przedstawione przez Autora dysertacji metody oraz wykonane badania eksperymentalne mają istotny wpływ na rozwój wiedzy dotyczącej wykorzystania algorytmów sztucznej inteligencji w zagadnieniach obrony przed cyberzagrożeniami. Przeciwdziałanie cyberataków jest obecnie jednym z głównych nurtów narodowej gospodarki cyfrowej, tak więc recenzowana dysertacja posiada duży potencjał wykorzystania oraz przydatności w rzeczywistych systemach zabezpieczenia przed atakami sieciowymi.

10. Czy rozprawa spełnia wymagania stawiane rozprawom doktorskim przez obowiązujące przepisy?

Reasumując stwierdzam, iż mgr inż. Marek Pawlicki posiada ogólną wiedzę teoretyczną z dyscypliny Informatyka Techniczna i Telekomunikacja. W szczególności wiedza ta dotyczy zagadnień związanych z cyberbezpieczeństwem oraz metodami uczenia maszynowego. Recenzowana praca zawiera sformułowaną tezę pracy, która została udowodniona na drodze badań eksperymentalnych. Lektura dysertacji pozwala stwierdzić, że Autor zaprezentował na jej łamach umiejętność samodzielnego prowadzenia pracy naukowej.

Wobec powyższego, recenzowana praca spełnia wymagania zdefiniowane przez Ustawę z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. U. z 2017 r. poz. 1789 z późniejszymi zmianami). Konkludując, wnoszę o przyjęcie rozprawy oraz dopuszczenie mgr inż. Marka Pawlickiego do publicznej obrony.


.....
podpis