

Dr hab. inż. Maciej Kusy, prof. PRz

Rzeszów, 9 stycznia 2026 r.

Katedra Podstaw Elektroniki

Wydział Elektrotechniki i Informatyki

Politechnika Rzeszowska im. Ignacego Łukasiewicza

al. Powstańców Warszawy 12

35-029 Rzeszów

e-mail: mkusy@prz.edu.pl

Recenzja rozprawy doktorskiej

mgr. inż. Szymona Niewiadomskiego

pt. „ Fixing data quality issues in CMDB in IT and OT by using machine learning algorithms. Forecasting for IT procurement”

Prawną podstawą przygotowania recenzji rozprawy doktorskiej mgr. inż. Szymona Niewiadomskiego jest uchwała Rady Dyscypliny Naukowej Informatyka Techniczna i Telekomunikacja (RDNITiT) Politechniki Wrocławskiej (PWr) nr 325/15/RDND03/2024-2028 z dnia 22 października 2025 r., w której Rada Dyscypliny powołuje recenzentów. O treści uchwały zostałem poinformowany pismem Przewodniczącego RDNITiT – prof. dr. hab. Wojciecha Bożejko.

Niniejsza recenzja została przygotowana na podstawie rozprawy doktorskiej. Przedmiotowa praca została zrealizowana w ramach programu „Doktorat wdrożeniowy” na Politechnice Wrocławskiej we współpracy z firmą Gaz-System S.A., pod kierunkiem dr. hab. inż. Grzegorza Mzyka, prof. PWr.

Rozprawa, licząca łącznie 105 stron, obejmuje: wprowadzenie, problematykę związaną z czyszczeniem danych, opis zagadnienia strategii zakupowej, charakterystykę implementacji stworzonego w ramach pracy narzędzia oraz bibliografię. Zawiera także dodatki (A – D), listę symboli oraz łącznie 19 rysunków i 8 tabel. Literatura obejmuje 69 prac, wśród których Doktorant jest współautorem pięciokrotnie.

WPLYNĘŁO

12-01-2026

RDN-IT/8/2026

1. Tematyka pracy

Oceniana rozprawa koncentruje się na dwóch procesach zarządzania usługami IT, które należą do szczególnie wymagających pod względem wdrożeniowym: zarządzaniu konfiguracją oraz zarządzaniu pojemnością. Autor proponuje modułowe i wdrażane lokalnie ramy uczenia maszynowego, które w istotny sposób podnoszą jakość baz danych zarządzania konfiguracją (ang. Configuration Management Databases, CMDB). Rozwiązanie to uwzględnia jednocześnie rygorystyczne wymagania w zakresie bezpieczeństwa, audytowalności oraz ograniczenia kadrowe charakterystyczne dla operatorów infrastruktury krytycznej.

Drugim obszarem badań przedstawionych w rozprawie jest zagadnienie prognozowania na potrzeby zakupów IT w warunkach ograniczeń popytu, cen, dostępności oraz finansowania. Autor formułuje problem minimalizacji kosztów z wykorzystaniem estymatorów cen i popytu, a następnie proponuje algorytm genetyczny do wyznaczania dopuszczalnych harmonogramów zakupów, zaprojektowany w sposób uzasadniony z punktu widzenia charakteru rozważanego problemu decyzyjnego.

Doktorant wykazuje również, iż algorytmy uczenia maszynowego, zaprojektowane w sposób przejrzysty i analitycznie uzasadniony, mogą istotnie poprawić jakość danych przechowywanych w bazach CMDB zarówno w obszarze IT, jak i OT. Zaproponowane metody nie opierają się na nieprzejrzystych heurystykach ani modelach typu „czarna skrzynka”, lecz są osadzone w jednoznacznie zdefiniowanych ramach matematycznych.

Uważam, że podjęta w rozprawie doktorskiej mgr. inż. Szymona Niewiadomskiego tematyka badawcza jest z naukowego punktu widzenia ważna i aktualna.

2. Wkład Autora

Doktorant wnosi następujące elementy wkładu teoretycznego:

- sformułowanie nowych rekurencyjnych algorytmów uczenia, które w ramach nieparametrycznego modelu probabilistycznego uwzględniają mechanizmy zapominania oraz wycofywania;
- analityczne i eksperymentalne porównanie metod opartych na odległości oraz korelacji w kontekście detekcji anomalii w heterogenicznych danych konfiguracyjnych;
- opracowanie kryteriów walidacji odporności i wiarygodności modeli w środowiskach operacyjnych.

Przedstawienie rozprawy doktorskiej, w której Autor nie formułuje i nie wykazuje poprawności (prawdziwości) tezy nie jest podejściem powszechnie stosowanym, ale nie ma wymogów formalnych, aby tezę do pracy wprowadzać. Uważam podejście Autora za jak najbardziej poprawne.

3. Struktura pracy

Praca napisana jest w języku angielskim. Zaprezentowane w dysertacji zagadnienia zostały opisane w 4 rozdziałach.

Rozdział 1 zawiera jedenaście sekcji. Doktorant przedstawia w nim kontekst biznesowy i organizacyjny funkcjonowania systemów CMDB w środowisku przedsiębiorstwa infrastrukturalnego, ze szczególnym uwzględnieniem specyfiki IT i OT oraz uwarunkowań wdrożeniowych. Omawia zarządzanie usługami IT w obszarach konfiguracji, pojemności i popytu, wskazując ich zależność od jakości i spójności danych konfiguracyjnych. Doktorant koncentruje się również na wspólnym modelu danych CMDB, jego strukturze, aspektach bezpieczeństwa, interoperacyjności oraz porównaniu rozwiązań oferowanych przez wiodących dostawców. Wprowadza grafową reprezentację CMDB, uzasadniając jej przydatność do modelowania relacji między elementami konfiguracji oraz realizacji celów biznesowych, takich jak poprawa jakości danych i wsparcie strategii zakupowych. W końcowej części rozdziału mgr inż. Szymon Niewiadomski prezentuje analizę procesów zarządzania jakością danych, wspieranych metodami sztucznej inteligencji, dokonuje przeglądu literatury i identyfikuje luki badawcze, które stanowią punkt wyjścia do rozważań przedstawionych w dalszych rozdziałach rozprawy.

Rozdział 2 składa się z czterech części poświęconych metodom automatycznego czyszczenia i poprawy jakości danych w bazach CMDB z wykorzystaniem technik uczenia maszynowego. Przedstawiono w nim algorytmy oparte na estymacji jądrowej gęstości prawdopodobieństwa, selekcji cech (w tym z użyciem miary podobieństwa Jaccarda) oraz klasyfikatorach neuronowych wraz z ich walidacją eksperymentalną. Omówiono także podejścia dystansowe i strukturalne do wykrywania rekordów podejrzanych oraz błędów relacyjnych w danych, w tym transformację struktur CMDB do postaci umożliwiającej analizę ML. Rozdział zamyka syntetyczna analiza zalet proponowanych metod, obejmująca m.in.: bezpieczeństwo, przetwarzanie on-premise, modułowość, wyjaśnialność oraz możliwość adaptacji i transferu wiedzy do innych domen.

Rozdział 3, w ramach ośmiu sekcji, poświęcony jest zagadnieniu strategii zakupowych IT i prezentuje kompletne rozwiązanie obejmujące sformułowanie problemu decyzyjnego, ideę algorytmu oraz jego implementację. Przedstawiono w nim modele wejściowe opisujące uwarunkowania rynkowe i finansowe, w tym estymatory funkcji cen, popytu oraz wpływ stóp procentowych na proces zakupowy. Centralnym elementem rozdziału jest algorytm przeznaczony do optymalizacji zakupów zasobów pamięci masowej w IT. Rozwiązanie opiera się na estymacji funkcji ceny i popytu, a następnie na heurystyce wykorzystującej algorytm genetyczny do wyznaczenia optymalnego planu zakupów. Skuteczność algorytmu oceniono eksperymentalnie w dwóch scenariuszach: z rabatami przy stałej cenie produktu oraz bez rabatów przy cenie malejącej w czasie. Uzupełnieniem części algorytmicznej jest opis praktycznych aspektów wdrożeniowych, takich jak: monitorowanie trendów cenowych, prognozowanie czasów dostaw oraz wpływ modeli licencyjnych i adopcji chmury. Rozdział zamyka syntetyczne podsumowanie oraz wskazanie otwartych problemów badawczych.

Ostatni rozdział rozprawy doktorskiej podsumowuje zagadnienia związane z wdrożeniem zaproponowanego rozwiązania. Przedstawiony w czterech sekcjach opis obejmuje takie zagadnienia jak: zarządzanie transformacją jako warunek wstępny wdrożenia, procedury implementacyjne ustalone na poziomie organizacyjnym, architekturę systemu, która została zaprojektowana jako rozwiązanie zorientowane usługowo, oraz długoterminową strategię utrzymania i wsparcia zaproponowanego w ramach pracy rozwiązania.

5. Ocena rozprawy doktorskiej

5.1. Uwagi językowe i redakcyjne

W tekście rozprawy można zidentyfikować uchybienia językowe i redakcyjne, w tym dotyczące stylu naukowego. Poniżej przedstawiam wybrane mankamenty.

- Stosowanie w pracy sformułowań, takich jak: „AI support for”, „AI inspection”, „AI suggestions” czy „AI can be operationally integrated”, które mają wydźwięk popularnonaukowy lub potoczny i są częściej spotykane w języku publicystycznym. W opracowaniach o charakterze naukowym zasadne jest stosowanie terminologii precyzyjnej, odnoszącej się bezpośrednio do modeli, metod lub systemów, np. „AI model”, „AI technique” czy „AI-based system”.
- Wielokrotne wyjaśnianie już wprowadzonych skrótów. Na przykład, pełnemu terminowi *bazy danych zarządzania konfiguracją* (ang. Configuration Management Databases) przypisano skrót CMDB. Rozwinięcie tego skrótu występuje na stronach 10, 12, 15 oraz

18 – dwukrotnie, a także na stronach 11, 14, 20, 21, 25, 27 – jednokrotnie (pomijam wystąpienie w nagłówku strony). Podobny problem występuje w przypadku skrótów AI, CI, CSDM czy ML. W pracach naukowych wystarczy jednokrotne rozwinięcie skrótu.

- Błędy interpunkcyjne i edytorskie, błędy składu typograficznego (nieprawidłowe formatowanie nagłówków, w wyniku którego standardowy odnośnik zawierający numer i nazwę rozdziału nakłada się na numer i nazwę sekcji, np. na stronach 82, 83, 86, 100), niesformatowane wzory i cytowania; niewiadome odniesienia (pod ilustracją Figure 3.4).

Nie jest konieczne odnoszenie się przez Doktoranta do zaprezentowanych powyżej uwag.

5.2. Uwagi krytyczne

Uwagi ogólne:

1. W dosłownym tłumaczeniu tytuł rozprawy brzmi: „*Poprawa jakości danych w bazach CMDB w obszarach IT i OT z wykorzystaniem algorytmów uczenia maszynowego. Prognozowanie na potrzeby zakupów IT*”. Zastosowanie w tytule dwóch odrębnych zdań budzi wątpliwości interpretacyjne, zwłaszcza że dla obu części nie wskazano jednoznacznego wspólnego mianownika. Dodatkowo niespójna interpunkcja (kropka po pierwszym członie i jej brak po drugim) potęguje wrażenie formalnej niejednorodności. Należy również zauważyć, że w tytule występują skróty (IT, OT, CMDB), które są wytłumaczone później, a pojęcie IT pojawia się dwukrotnie, co wpływa na jego czytelność i zwięźłość. Z perspektywy redakcyjnej zasadne wydaje się ujęcie tytułu w jednej, spójnej formie oraz jego skrócenie. Przykładowo tytuł mógłby zostać sformułowany jako: „*Zastosowanie algorytmów uczenia maszynowego do poprawy jakości danych oraz prognozowania procesów zakupowych w branży IT*”. Ponadto, w mojej opinii, *Prognozowanie na potrzeby zakupów IT* jest akceptowalnym, lecz słabo wyspecyfikowanym kontekstem wdrożeniowym.
2. W rozprawie nie sformułowano w sposób wystarczająco jednoznaczny problemu badawczego lub implementacyjnego ani celu pracy (co utrudnia ocenę zakresu), motywacji oraz elementów nowości zaproponowanego rozwiązania.
3. W ujęciu strukturalnym rozprawa zawiera wstęp (rozdział 1: *Introduction*). Nie posiada ona jednak klasycznego zakończenia w formie odrębnej części (np. *Conclusions/Summary*). Nie świadczy to o błędzie merytorycznym, lecz o nieukończeniu pracy w wymiarze „konstrukcyjnym”; brak syntezy uniemożliwia bowiem jednoznaczną ocenę wkładu naukowego Autora.

4. W pracy nie wprowadzono wyraźnego rozróżnienia pomiędzy częścią teoretyczną a eksperymentalną, co w niektórych fragmentach utrudnia jednoznaczną ocenę zakresu walidacji zaproponowanych rozwiązań. Wyraźniejsze rozgraniczenie tych aspektów poprawiłoby czytelność i spójność rozprawy.

Uwagi do poszczególnych rozdziałów:

1. Uwagi do Rozdziału *Introduction*

- Zasadniczą słabością ocenianego rozdziału jest jego nadmierna objętość. Składa się on aż z 11 sekcji i liczy 19 stron, co znacząco obniża jego przejrzystość. Informacje dotyczące: kontekstu biznesowego i organizacyjnego funkcjonowania systemów CMDB w przedsiębiorstwie infrastrukturalnym (z uwzględnieniem specyfiki IT i OT oraz uwarunkowań wdrożeniowych), zarządzania usługami, modelu danych CMDB i jego grafowej reprezentacji, a także analizy procesów zarządzania jakością danych wspieranych metodami sztucznej inteligencji należałoby przenieść do odrębnego rozdziału.
- Jak już wskazano w uwagach ogólnych, we wstępie nie sformułowano wprost problemu badawczego, celu pracy, motywacji ani zaproponowanych elementów nowości. Jednocześnie, na podstawie treści kolejnych rozdziałów, wkład Autora jest na tyle istotny, że sformułowanie tych elementów nie powinno sprawiać trudności. Ułatwiłoby to niewątpliwie czytanie rozprawy.
- Sekcja 1.1 zatytułowana jest „Dissertation Thesis”. W tej części pracy Autor nie formułuje jednak tezy rozprawy, lecz przedstawia wkład własny, co nie jest pojęciowo tożsame.
- Przegląd literatury przedstawiony w sekcji 1.10 ma charakter pobieżny i nie wykorzystuje w pełni potencjału tego fragmentu pracy.

2. Uwagi dotyczące rozdziału *Data Cleaning*

- W sekcji 2.1.1 Autor, w ujęciu niematematycznym, powinien jasno uzasadnić postawiony problem. Wynika to z braku wstępu do rozdziału 2, dzięki któremu przejście do sekcji opisującej algorytm do identyfikacji/korekcji błędnych danych byłoby płynne i bardziej zrozumiałe.
- Autor wprowadza i formalnie dowodzi dwa lematy: pierwszy dotyczący odporności algorytmu na błędne etykietowanie rekordów oraz drugi odnoszący się do przyrostowej aktualizacji estymatora z lokalną szerokością pasma. Choć oba lematy zostały poprawnie

sformułowane i udowodnione, w treści rozprawy nie wskazano w sposób jednoznaczny, w jaki sposób zostały one zweryfikowane empirycznie lub wykorzystane w dalszej części pracy, w szczególności w rozdziałach poświęconych implementacji. W konsekwencji ich rola praktyczna i związek z częścią wdrożeniową nie są w pełni czytelne.

- Jak jest powód wykorzystania współczynnika podobieństwa Jaccarda do selekcji cech?
 - W sekcji 2.1.8 przedstawiono zastosowanie sieci neuronowej typu feed-forward do problemu klasyfikacji binarnej. Jakie były przesłanki wyboru architektury sieci, w szczególności zastosowania dwóch warstw ukrytych po cztery neurony oraz sigmoidalnej funkcji aktywacji? W jaki sposób opis hiperparametrów dla pozycji *Neural network architecture* odnosi się do przedstawionych ustawień modelu (strona 46)?
 - Podsekcje 2.1.9, 2.1.10 oraz 2.2.3 przedstawiają wyniki eksperymentów. Czy eksperymenty te mają bezpośredni związek z teoretycznymi zagadnieniami przedstawionymi w sekcjach 2.1 oraz 2.2?
 - Algorytm 1 zaproponowany w sekcji 2.3.2 wymaga bardziej szczegółowego objaśnienia. Zasadne byłoby ponumerowanie poszczególnych kroków procedury oraz wprowadzenie w tekście odwołań do tych kroków wraz z ich opisem. Przedstawiony algorytm charakteryzuje się znaczną złożonością pod względem zastosowanych symboli, funkcji, operacji i operatorów, co - bez odpowiedniego komentarza i odniesienia do zagadnień teoretycznych - utrudnia jego interpretację. Uzupełnienie opisu o prosty przykład ilustrujący działanie algorytmu niewątpliwie zwiększyłoby jego czytelność i wartość poznawczą. Jak jest zastosowanie tego algorytmu? Czy sekcja 2.3.4 przedstawia jego przykładowe zastosowanie? Jeśli tak, to co jest rezultatem Algorytmu 1 zgodnie z listingiem na stronie 61?
3. Uwagi dotyczące rozdziału *Purchasing Strategy*
- W sekcji 3.6 Doktorant nie podał interpretacji wykresów na rysunkach 3.7, 3.8 i 3.9.
 - Trudno zweryfikować wyniki przedstawione w krótkim akapicie u góry strony 81 (nad rysunkiem 3.9).
4. Uwagi do rozdziału *Implementation*
- Wobec braku jednoznacznie sformułowanego celu pracy, który precyzyjnie wyznaczałby jej zakres i zastosowane środki oraz umożliwiał spójną weryfikację uzyskanych rezultatów, trudno jest jednoznacznie określić, jakie rozwiązanie jest przedmiotem rozważań w tym rozdziale. W konsekwencji nie jest jasne, co Autor ma na myśli, pisząc:

- „The effective deployment of the proposed solution...” Czy mowa tu o algorytmie opisanym w rozdziale 3?
- „The availability of this unified dataset was a necessary condition for applying the algorithms developed in this dissertation.” O jakie algorytmy dokładnie chodzi?
- „... the solution exposes a RESTful API, which acts as an intermediary layer between the underlying machine learning components and external enterprise applications.” O jakich komponentach jest tu mowa?

5.3. Uwagi drobne

1. Opisując wybrane metody sztucznej inteligencji, warto byłoby odwołać się do odpowiednich źródeł literaturowych, zwłaszcza że przedstawione podejścia reprezentują aktualny stan wiedzy.
2. Przy tak zaawansowanym aparacie matematycznym wykorzystanym przy wyjaśnianiu zagadnień teoretycznych w rozdziale 2, dodatek C jest - w mojej opinii - zbędny.
3. W opisie zagadnienia wektoryzacji przedstawionym w dodatku D zasadne byłoby uzupełnienie rozważań o przykłady ilustrujące działanie klas **CountVectorizer** oraz **TfidfVectorizer** (biblioteka **scikit-learn**), służących do transformacji danych tekstowych lub kategoriowych do postaci wektorów numerycznych odpowiednich dla modeli uczenia maszynowego.
4. Etykiety wykorzystane do opisu rysunków 3.7, 3.8 i 3.9 są zbyt małe/niewyraźne.

Nie jest konieczne odnoszenie się przez Doktoranta do zaprezentowanych powyżej uwag drobnych.

5.4. Ocena ogólna

Formułując ocenę ogólną, należy podkreślić, iż mgr inż. Szymon Niewiadomski posiada bardzo dobrą znajomość procesów zarządzania usługami IT. Rozumie również zagadnienia prognozowania na potrzeby zakupów IT w warunkach ograniczeń popytu, cen, dostępności zasobów oraz finansowania. Potrafi w tym kontekście (i) sformułować problem minimalizacji kosztów z wykorzystaniem estymatorów cen i popytu, a następnie (ii) zaproponować skuteczny algorytm do wyznaczania dopuszczalnych harmonogramów zakupów, zaprojektowany w sposób uzasadniony z punktu widzenia charakteru rozważanego problemu decyzyjnego.

Doktorant wykazuje, że algorytmy ML, zaprojektowane w sposób przejrzysty i analitycznie uzasadniony, mogą istotnie poprawić jakość danych przechowywanych w bazach

CMDB zarówno w obszarze IT, jak i OT. Co istotne, proponowane metody nie opierają się na nieprzejrzystych heurystykach ani modelach typu „czarna skrzynka”, lecz są osadzone w jednoznacznie zdefiniowanych ramach matematycznych.

W oparciu o solidny warsztat matematyczny Autor proponuje nowe rekurencyjne algorytmy uczenia, uwzględniające mechanizmy zapominania oraz wycofywania. Przeprowadza analityczne i eksperymentalne porównanie metod opartych na miarach odległości oraz korelacji w kontekście detekcji anomalii w heterogenicznych danych konfiguracyjnych, a także opracowuje kryteria walidacji odporności i wiarygodności modeli stosowanych w środowiskach operacyjnych.

Rozprawa doktorska mgr. inż. Szymona Niewiadomskiego została zrealizowana w ramach programu „Doktorat wdrożeniowy”. Mimo że prezentuje wysoki poziom naukowy, w szczególności w warstwie teoretycznej, zawiera jednak obszary wymagające dopracowania, które zostały wskazane we wcześniejszych komentarzach. Obejmują one zarówno kwestie o charakterze kluczowym, jak i drobniejsze detale redakcyjne i strukturalne. W ujęciu ogólnym należy zauważyć, że w pracy nie sformułowano jednoznacznie celu rozprawy, co utrudnia ocenę jej zakresu oraz stopnia realizacji założeń. Ponadto nie wyodrębniono rozdziału podsumowującego, który umożliwiłby syntetyczny powrót do problemu i celu pracy, podkreślenie najważniejszych osiągnięć, rozdzielenie wkładu teoretycznego i wdrożeniowego, ocenę skuteczności zaproponowanego rozwiązania oraz wskazanie jego ograniczeń i możliwych kierunków dalszych prac wdrożeniowych. Zauważalny jest również brak jednoznacznego rozgraniczenia części naukowej i eksperymentalnej rozprawy, co w niektórych fragmentach obniża czytelność prezentowanych wyników.

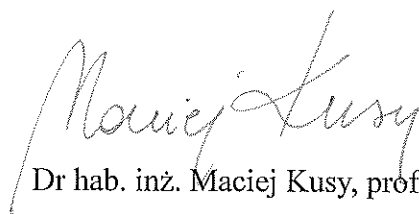
Niezależnie od opisanych przeze mnie niedociągnięć, praca posiada ogromny potencjał, który można jeszcze bardziej wyeksponować poprzez uwzględnienie przedstawionych wyżej uwag. Odpowiedzi na zawarte w recenzji komentarze i pytania niewątpliwie utwierdzą mnie w przekonaniu, iż opracowany materiał jest kompletny, a przede wszystkim unikalny.

Elementami wyróżniającymi całościowy dorobek naukowy Autora są liczne artykuły związane z tematyką rozprawy. Wśród pięciu prac wspólnych z Promotorem, Doktorant jest autorem pierwszym aż czterokrotnie. Należy również dodać, iż dwa artykuły zostały opublikowane w czasopiśmie znajdujących się w wykazie czasopism naukowych i konferencji MNiSW.

6. Podsumowanie

Rozprawa doktorska mgr. inż. Szymona Niewiadomskiego, przygotowana pod opieką naukową dr. hab. inż. Grzegorza Mzyka, prof. PWr, stanowi oryginalne rozwiązanie zagadnienia „ML w zarządzaniu danymi IT” w ramach programu „doktorat wdrożeniowy” oraz potwierdza wysoki poziom wiedzy teoretycznej i praktycznej Doktoranta w danej dyscyplinie naukowej. Na podstawie przedmiotowego przeglądu literatury, przedstawienia zagadnień teoretycznych i przygotowania opisu implementacyjnego rozwiązania uważam, iż wiedzę Autora w dyscyplinie *informatyka techniczna i telekomunikacja* należy ocenić bardzo wysoko.

Stwierdzam zatem, że recenzowana przeze mnie rozprawa doktorska mgr. inż. Szymona Niewiadomskiego pt. „Fixing data quality issues in CMDB in IT and OT by using machine learning algorithms. Forecasting for IT procurement” spełnia kryteria określone w ustawie z dnia 20 lipca 2018 r. „Prawo o szkolnictwie wyższym i nauce” (Dz. U. 2018 poz. 1668 z późn. zm.). Na tej podstawie wnioskuję do Rady Dyscypliny Naukowej Informatyka Techniczna i Telekomunikacja Politechniki Wrocławskiej o dopuszczenie Doktoranta do dalszych etapów przewodu doktorskiego.



Dr hab. inż. Maciej Kusy, prof. PRz