

dr hab. inż. Damian Krenczyk, prof. PŚ
Katedra Automatykacji Procesów Technologicznych
i Zintegrowanych Systemów Wytwarzania
Wydział Mechaniczny Technologiczny
Politechnika Śląska
ul. Konarskiego 18A
44-100 Gliwice

Gliwice, 12.08.2022r.

RECENZJA

rozprawy doktorskiej mgr. inż. Michała Pajęckiego pt. „*Diagnozowanie wybranych zagrożeń bezpieczeństwa pracy w przedsiębiorstwach produkcyjnych z zastosowaniem technik eksploracji danych*”

Podstawa opracowania

Recenzję opracowano na podstawie pisma dr. hab. inż. Sławomira Błasiaka, prof. PŚk, Dyrektora Naukowego Dyscypliny Inżynieria Mechaniczna Politechniki Świętokrzyskiej nr MAA-5100/2022, z dnia 15.06.2022r.

Obszar problemowy i aktualność tematyki rozprawy

Podjęta przez Doktoranta w rozprawie tematyka dotyczy zagadnień związanych z problemami zagrożeń bezpieczeństwa pracy, w szczególności identyfikacji istotnych czynników dotyczących zdarzeń wypadkowych oraz wzorców wypadków przy pracy w przedsiębiorstwach produkcyjnych. Analiza incydentów w oparciu o charakterystykę działalności wytwórczej, a zatem w większości w branżach zaawansowanych technologicznie, jest bardzo ważna dla znajdowania ich przyczyn oraz zarządzania planowaniem zapobiegawczym. W centrum rozważań Autora znajduje się zagadnienie identyfikacji wzorców definiujących zagrożenia bezpieczeństwa pracy poprzez opracowanie modeli pozwalających na zdiagnozowanie cech mających istotne znaczenie w generowaniu zagrożeń. Jest to niewątpliwie jeden z obszarów prac naukowo-badawczych, w którym konieczne jest ciągłe wypracowywanie i wdrażanie właściwych w danych warunkach metod i rozwiązań. Potrzeba ciągłego rozwoju wynika przede wszystkim z szybkiego postępu i ciągłego doskonalenia systemów produkcyjnych, stymulowanych potrzebami zdobywania i utrzymywania przewagi konkurencyjnej przez podniesienie efektywności i elastyczności nowoczesnych przedsiębiorstw. Obserwowany postęp technologiczny, gwałtowny wzrost poziomu automatyzacji oraz produkcji zorientowanej na systemy cyber-fizyczne, integrujące w kompletnym łańcuchu wartości m.in. wytwarzanie, logistykę, magazynowanie i dystrybucję, powoduje zmianę charakteru zagrożeń i konieczność ponownego zdefiniowania czynników ryzyka. Konieczne staje się zatem poszukiwanie nowych metod tworzenia modeli przyczyn wypadków, stanowiących ważną podstawę teoretyczną analizy

wypadków, pozwalających na predykcję zagrożeń oraz kształtowanie bezpiecznego środowiska pracy, a w efekcie pozwalające zapewnić możliwość utrzymania oczekiwanej zdolności produkcyjnej systemu. W tym kontekście Doktorant trafnie wskazuje na potrzebę bardziej ogólnego podejścia do badania zagrożeń bezpieczeństwa pracy, podkreślając zaobserwowaną tendencję prowadzenia badań na podstawie szczegółowej analizy pojedynczych zdarzeń, czy bazujących na danych zagregowanych. Identyfikuje również wybrany do badań obszar działalności gospodarczej, przeprowadzając szczegółową analizę stanu wypadkowości w poszczególnych sekcjach PKD. Ponadto celnie zauważa tendencję do wykorzystania technik eksploracji danych w modelowaniu wybranych aspektów wypadkowości. Jest to zgodne z postulatami aktualnie formułowanymi w literaturze przedmiotu, mówiącymi o tym, że w erze sztucznej inteligencji i *big data* jednym z ważniejszych obszarów badań w temacie skutecznego zapewnienia bezpieczeństwa pracy jest właśnie wykorzystanie modeli przyczynowości wypadków bazujących na metodach eksploracji danych.

Reasumując, recenzowana praca koncentruje się na problemach kształtowania środowiska pracy oraz bezpieczeństwa pracy, dotyczy takich szczegółowych zagadnień jak: metod modelowania zagrożeń bezpieczeństwa pracy w przedsiębiorstwach produkcyjnych, wczesna identyfikacja oraz predykcja zagrożeń technicznych i zdrowotnych w środowisku pracy oraz zarządzanie ryzykiem. Przedstawione w rozprawie zagadnienia oraz wspomniane obszary problemowe mieszczą się w dyscyplinie inżynieria produkcji (wg klasyfikacji obowiązującej w momencie otwarcia przewodu doktorskiego) oraz szerzej w dyscyplinie inżynieria mechaniczna. Wpisują się także z powodzeniem w zidentyfikowaną lukę badawczą, związaną z niewielką liczbą badań prowadzonych na podstawie dużego zbioru obejmującego zdarzenia jednostkowe oraz potrzebą zastosowania technik *data mining* do modelowania aspektów wypadkowości w odniesieniu do przedsiębiorstw produkcyjnych. W tym kontekście Doktorant prawidłowo formułuje cel pracy jako opracowanie metody pozwalającej na *identyfikację wzorców wypadków przy pracy (tj. wzorców wypadkowych) w przedsiębiorstwach produkcyjnych na podstawie cech charakteryzujących uszkodzonych przy wykorzystaniu zasobów danych o wypadkach zaistniałych w przeszłości i technik data mining*.

Wybór tematyki badawczej poruszanej przez Doktoranta w przedstawionej do recenzji rozprawie uważam za uzasadniony, a omawiane w niej treści są aktualne i wpasowują się w zagadnienia związane z dyscypliną inżynieria mechaniczna.

Ocena rozprawy

Oceniana rozprawa liczy 158 stron, składa się z 9 rozdziałów, wprowadzenia, sześciu załączników, listy cytowanej bibliografii liczącej 120 pozycji, spisu rysunków i tablic oraz streszczeń w języku polskim i angielskim. W załącznikach, poza wartościami wybranych wskaźników zagrożeń wyznaczonych w ramach analizy stanu wypadkowości w sekcji przetwórstwa przemysłowego oraz wzoru katy wypadku, zamieszczono wyznaczone w procesie identyfikacji wzorców wypadkowych wartości wskaźników zdolności dyskryminacyjnej oraz wartości prawdopodobieństw dla obserwacji.

W pracy można wyróżnić cztery zasadnicze części. Część pierwsza obejmuje wprowadzenie, opis problematyki bezpieczeństwa pracy w Polsce, analizę literatury naukowej oraz sformułowanie celu i zakresu pracy (rozdziały 1-3). Rozdział 1 poświęcono zagadnieniom liczebności wypadków przy pracy w Polsce w latach 2008-2019. Zidentyfikowano obszary działalności gospodarczej, w których

zannotowano najwięcej zdarzeń wypadkowych. Dla wybranej sekcji (C - Przetwórstwo przemysłowe), dla lat 2008-2017 wyznaczono wybrane wskaźniki wypadkowości oraz wskaźniki zagrożeń, co pozwoliło na wskazanie działów działalności charakteryzujących się dużą dotkliwością skutków wypadków przy pracy. Stanowiło to także kryterium wyboru obszaru działalności (dział 16) do dalszych badań. W rozdziale przedstawiono także uwarunkowania prawne i definicje dla wybranych zagadnień związanych z wypadkami przy pracy. Następnie przedstawiono krótki przegląd proponowanych w literaturze modeli przyczynowości wydarzeń wypadkowych, pozwalający na trafne zidentyfikowanie luki badawczej związanej ze stosunkowo dużą liczbą prowadzonych badań opartych na szczegółowej analizie pojedynczych zdarzeń wypadkowych w porównaniu do liczby badań opartych na dużych zbiorach danych jednostkowych. W tym miejscu Doktorant stwierdza, w odniesieniu do proponowanego przez siebie podejścia do analizy danych, że „w tym celu zastosowano wybrane techniki eksploracji danych”. W mojej ocenie zabrakło tutaj głębszego uzasadnienia takiego wyboru. Czynniki motywacyjne można znaleźć w kilku miejscach w dalszej części pracy, a słuszność wyboru wydają się potwierdzać uzyskane wyniki, jednak ich brak przed sformułowaniem celów badawczych może sprawiać wrażenie zbyt skrótowego potraktowania kwestii związanych z określeniem czynników motywacyjnych. Na podkreślenie zasługują jednak trafnie zasygnalizowane przez Autora przyczyny małej liczby badań na większych zbiorach danych (w odniesieniu do zdarzeń wypadkowych w Polsce), u których podstawy leżą trudności w pozyskaniu jednostkowych danych od instytucji gromadzących je na potrzeby badań naukowych. Biorąc pod uwagę opisane w kolejnych rozdziałach wyniki działań w tym zakresie, na uznanie zasługuje wytrwałość i nieustępliwość Doktoranta. Rozdział 2 poświęcono analizie literatury naukowej przedmiotu. Doktorant rozszerzył tutaj, już wcześniej zasygnalizowaną, tematykę prowadzonych badań w omawianym obszarze, tym razem odrębnie analizując wyniki przedstawiane w literaturze polskiej i światowej. W odniesieniu do drugiej grupy skoncentrowano się przede wszystkim na publikacjach z zakresu badań nad zagrożeniami bezpieczeństwa pracy z zastosowaniem technik eksploracji danych, prowadzonych na niezagregowanych zbiorach danych. Zwrócono uwagę na brak możliwości adaptacji proponowanych rozwiązań i wniosków do problematyki polskiej, czy brak uzasadnienia takich działań ze względu na specyficzne uwarunkowania prawne, społeczne i ekonomiczne. Zasygnalizowano także ważne kwestie związane z tym, że prowadzone badania dotyczyły zdarzeń w innych sekcjach gospodarki (niż wybrane przez Doktoranta). Pomimo że trafnie dostrzeżone zostały problemy i różnice związane ze specyfiką uwarunkowań, mój niedosyt budzi brak podsumowania tego rozdziału. W rozdziale 3 sformułowano cel, hipotezę oraz zakres pracy, które poprzedzone zostały opisem zidentyfikowanych luk badawczych w odniesieniu do istniejących analiz wypadków przy pracy w przedsiębiorstwach produkcyjnych. Trafnie i logicznie sformułowano ponadto cząstkowe zadania badawcze.

Drugą część pracy (rozdziały 4 – 8) Doktorant poświęcił metodyce tworzenia modeli klas ukrytych wybranych zagrożeń bezpieczeństwa pracy na podstawie jednostkowego zbioru danych o zdarzeniach wypadkowych. W rozdziale 4 opisano zagadnienia związane z odkrywaniem wiedzy z danych. Uporządkowano i zdefiniowano pojęcia analizy danych i eksploracji danych oraz przedstawiono najczęściej wykorzystywane metody. Zaprezentowano, opracowaną na podstawie analizy literaturowej, koncepcję *uproszczonego procesu odkrywania wiedzy w danych*, będącego

161

podstawą opracowywanego autorskiego rozwiązania. Dużą uwagę poświęcono opisowi wybranej metody analizy skupień, grupowania – metody analizy klas ukrytych LCA (Latent Class Analysis). Trafnie wskazano na podstawową różnicę pomiędzy innymi technikami (analizy czynnikowej, analizy cech ukrytych, analizy profili ukrytych). Niedosyt budzi jednak zbyt zwięzłe uzasadnienie wyboru metody LCA, chociaż poparte odwołaniem do prac z wynikami wcześniejszych badań pilotażowych oraz analiz Doktoranta. Wskazane byłoby tutaj zamieszczenie chociażby krótkiego podsumowania wskazanych prac. Bardzo szczegółowo natomiast przedstawiono opis modelu klas ukrytych oraz proces estymacji parametrów modelu LCA wraz z zagadnieniami identyfikowalności i oceny modelu. W rozdziale 5 zawarto szczegółową charakterystykę pokaźnego zbioru danych, pozyskanych przez Autora z GUS, obejmującego informacje o poszkodowanych w wypadkach przy pracy w latach 2008-2017. W kolejnych podrozdziałach opisano poszczególne etapy, składające się na proponowany sposób postępowania, a mianowicie proces selekcji i transformacji danych, którego celem było przygotowanie danych do analizy LCA. Pozytywnie oceniam szczegółowy, dobrze umotywowany opis poszczególnych etapów oraz sformułowanych kryteriów ich stosowania, m.in. wykluczenia z dalszych analiz zmiennych nieniosących żadnej wartości informacyjnej, selekcji danych związanych bezpośrednio z procesem produkcyjnym, agregacji wybranych kategorii i wartości czy wykluczania obserwacji. Rozdział uzupełniają szczegółowo opisane testy statystyczne i ich wyniki wskazujące na potrzebę agregacji wartości dla zmiennej obserwowalnej określającej wielkość przedsiębiorstw, a w efekcie, trafne w mojej ocenie, prowadzenie niezależnych analiz dla dwóch podzbiorów danych obejmujących mikro i małe oraz średnie i duże przedsiębiorstwa. W rozdziale 6, będącym kontynuacją opisu kolejnych etapów proponowanego rozwiązania, zamieszczono definicję wskaźnika zdolności dyskryminacyjnej (AR) oraz proponowany algorytm selekcji zmiennych obserwowalnych do budowy modelu LCA w oparciu o ten wskaźnik, które uważam za jedno z najważniejszych osiągnięć Doktoranta. O ile wykorzystywane w algorytmie parametry związane z ograniczeniem liczby klas (C_{min} , C_{max}) są proste do wyznaczenia, zabrakło tutaj głębszej dyskusji w odniesieniu do parametrów s (liczby początkowych modeli) oraz AR^* (próg wykluczenia zmiennej). Rozdział kończy dyskusja o aspektach, koniecznych do wzięcia pod uwagę podczas podejmowania decyzji o wyborze liczby klas ukrytych w modelu. Rozdział 7 w całości poświęcono opisowi procesu budowy modeli LCA z zastosowaniem omówionych w poprzednich rozdziałach sposobów postępowania oraz identyfikacji na jego podstawie wzorców wypadkowych. Na uwagę zasługuje bogato ilustrowane przedstawienie uzyskanych wyników w poszczególnych krokach, niezależnie dla obu podzbiorów danych. Dyskusyjne w mojej ocenie jest jednak poczynione założenie o przyjętej wartości progu AR^* na poziomie 0,2. Warty podkreślenia są sformułowane przez Doktoranta reguły (kryteria uporządkowania cech), według których przeprowadzono etykietowanie poszczególnych klas ukrytych. Dla obu wymienionych wyżej grup przedsiębiorstw pokazano przebieg i uzyskane wyniki w procesie selekcji zmiennych obserwowalnych oraz wartości wybranych statystyk oceny modeli LCA według liczby klas. Szczegółowo omówiono przebieg wyboru modelu, a powstałe wzorce zagrożeń bezpieczeństwa pracy zobrazowano w postaci mapy cieplnej. Szczegółowej charakterystyce opisowej poddano ponadto profile osób poszkodowanych w utworzonym modelu LCA. Dużo miejsca poświęcono także porównaniu otrzymanych wzorców (dla obu grup) ze względu na stopień uszkodzenia. Część tę kończy rozdział 8, w którym pokazano sposób przypisania ofiary zdarzenia wypadkowego

14

do określonej klasy ukrytej, na podstawie cech ją charakteryzujących, z zastosowaniem utworzonych modeli. W jego wyniku możliwe staje się utworzenie i zaprezentowanie sylwetek poszkodowanych poprzez wartości charakteryzujących ich zmiennych obserwowalnych.

Trzecia część rozprawy obejmuje podsumowanie i sformułowanie wniosków końcowych: W rozdziale 9 wskazano na przydatność technik eksploracji danych do identyfikacji wzorców wypadków w przedsiębiorstwach produkcyjnych na podstawie cech charakteryzujących poszkodowanych. Poddano dyskusji wyniki badań uzyskane z zastosowaniem proponowanego podejścia bazującego na metodzie LCA i sformułowano wnioski. Na podstawie wyznaczonych wzorców wskazano na cechy charakteryzujące poszkodowanych pracowników oraz otoczenia pracy w odniesieniu do dotkliwości wypadków i rodzaju obrażeń oraz zalecenia w zakresie bezpieczeństwa pracy. Wyodrębnione wzorce, sylwetki, jak również sformułowane zalecenia stanowią cenny wkład w wiedzę w obszarze kształtowania środowiska i bezpieczeństwa pracy w odniesieniu do przedsiębiorstw produkcyjnych. Autor rozprawy przedstawia także ograniczenia, lecz adresuje je jedynie w kierunku struktury pozyskanego zbioru danych, pomijając ograniczenia samej metody. Rozprawę kończy sformułowanie propozycji przyszłych prac badawczych w omawianym obszarze.

Reasumując, uważam, że kompozycja rozprawy, układ rozdziałów i podrozdziałów są poprawne oraz podporządkowane postawionemu i konsekwentnie zrealizowanemu przez Doktoranta celowi. Lista cytowanych źródeł literaturowych obejmuje pozycje istotne i aktualne. Strona edycyjna pracy również nie budzi zastrzeżeń.

Oryginalne osiągnięcia rozprawy

W swojej pracy Doktorant poszukuje odpowiedzi na postawiony problem badawczy, dotyczący kształtowania środowiska i bezpieczeństwa pracy, w szczególności identyfikacji istotnych czynników zdarzeń wypadkowych w przedsiębiorstwach produkcyjnych. Identyfikuje stan badań i luki badawcze w odniesieniu do istniejących analiz wypadków przy pracy. Odpowiedzią na postawiony problem badawczy, są:

- opracowanie metodyki konstruowania modeli klas ukrytych LCA zagrożeń bezpieczeństwa pracy na podstawie jednostkowego zbioru danych wejściowych,
- zaproponowanie sposobu wyznaczania wskaźnika zdolności dyskryminacyjnej AR dla zmiennych obserwowalnych, umożliwiającą ocenę znaczenia tych zmiennych w procesie budowy modelu klas ukrytych,
- opracowanie algorytmu selekcji zmiennych obserwowalnych do budowy modelu LCA,
- analiza stanu wypadkowości w sekcji przetwórstwa przemysłowego wraz ze wskazaniem działań szczególnie niebezpiecznych pod kątem zagrożeń bezpieczeństwa pracy,
- określenie specyfiki zagrożeń w wybranym dziale sekcji przetwórstwa przemysłowego oraz wyodrębnienie wzorców zagrożeń bezpieczeństwa pracy wraz ze sformułowanymi zaleceniami dla pracodawców i służb BHP.

Należy podkreślić, że recenzowana rozprawa porządkuje także wyniki zebrane w pracach Doktoranta, cytowanych w bibliografii. Wyniki pracy odpowiadają także aktualnym potrzebom przedsiębiorstw i poza wartością naukową mają dużą wartość użyteczną.

Uwagi ogólne i dyskusyjne

- W rozdziale 2 dokonano analizy literatury polskiej i światowej. Jak wspomniałem wcześniej, mój niedosyt budzi brak podsumowania tego rozdziału i próby głębszej analizy wyników oraz rozwinięcia uzasadnienia „braku możliwości adaptacji proponowanych rozwiązań” (bazujących na *data mining*) w kontekście wcześniej dokonanego wyboru zastosowania metod eksploracji danych w swoich badaniach przez Doktoranta. Uwaga dotycząca braku krótkiego podsumowania odnosi się także do pozostałych rozdziałów – jego zamieszczenie dałoby możliwość krótkiego spuentowania rozdziału i połączenia z tematyką kolejnych rozdziałów pracy.
- W rozdziale 4 (str. 54) opisując metody poszukiwania wartości parametrów modelu LCA, stwierdzono, w odniesieniu do funkcji (6), że „ponieważ optymalizowana funkcja może mieć wiele minimów lokalnych w procesie iteracyjnym, estymacja parametrów powinna być wykonywana wielokrotnie (powtórzenia) dla wielu różnych wartości początkowych”. Nie podano jednak informacji o sposobie szacowania liczby powtórzeń lub ich ograniczeniach.
- W odniesieniu do agregacji wybranych kategorii oraz zmiennych (podrozdział 5.3) Doktorant trafnie wskazał na taką potrzebę w celu uniknięcia mnogości potencjalnych wzorców odpowiedzi, co miało pozytywnie wpłynąć na możliwość profilowania wzorców. Dyskusyjna w mojej ocenie jest jednak dokonana agregacja wartości zmiennej P26 w odniesieniu do nowej wartości „Inny czynnik”, w ramach której zagregowano duży zbiór wejściowych wartości (na przykład, umieszczonych razem *Pojazdów* czy *Ludzi i innych organizmów żywych*).
- W podrozdziale 6.2 opisano *Algorytm selekcji zmiennych obserwowalnych do budowy modelu LCA*. Co w przypadku, gdy wynikiem funkcji *arg min* będzie zbiór argumentów o liczbie większej od 1? (patrz Załącznik 4 tab. 24 dla P07 i P16).
- Dyskusyjna jest kwestia przyjęcia wartości parametrów (Rozdział 7) „Założono, że wartość proggu AR^* wynosi 0,2 (...) $s=10$ estymacji...”. Czy wartości te wynikały z wcześniejszych badań/prób? Czy prowadzono testy dla innych wartości (*strojenia* algorytmu)?
- Opisane w rozdziałach 5 – 8 poszczególne etapy (i algorytmy) składają się na proponowany autorski sposób postępowania, co zostało wyeksponowane dopiero w rozdziale 9.2 jako „metodyka pozwalającej na skonstruowanie modeli klas ukrytych LCA wybranych zagrożeń bezpieczeństwa pracy na podstawie jednostkowego zbioru danych wejściowych”. Uważam, że wkład własny Doktoranta powinien zostać lepiej wyeksponowany już w tej części rozprawy.
- Zastanawiająca jest mała ("wykluczająca") zdolność dyskryminacyjna AR zmiennej P27 "Główna przyczyna wypadku" (podrozdział 7.2). Uważam, że w pracy powinien znaleźć się komentarz w tym zakresie.

Uwagi szczegółowe

Poza uwagami natury ogólnej i dyskusyjnej można sformułować następujące uwagi szczegółowe:

- Rozdział 1, str. 22, 23 - Wartości wskaźników dotkliwości wypadków (W_d) różnią się od wartości wyznaczonych dla danych z Tablicy 1.

- Rozdział 5.4 str. 78 – Pomimo wskazania w treści na zilustrowanie na rysunku 16 opisywanych różnic w rozkładach zmiennych obserwowalnych pomiędzy grupami przedsiębiorstw dla P01, nie zostały one tam uwzględnione.
- Rozdział 7.1 i 7.2, str. 92, 100 – Stwierdzenia „...wszystkie te zmienne mają wartość wskaźnika AR wynoszącą, w zaokrągleniu, co najmniej 0,5.” wymaga podania przyjętej dokładności zaokrąglenia. W przeciwnym przypadku wątpliwości budzi (rys. 23) np. wartość P06 dla 11 klas.
- Podrozdział „7.1. Grupa wk2 – przedsiębiorstwa średnie i duże” powinien posiadać numerację 7.2., kolejne podrozdziały również wymagają przenieumerowania.

Podsumowanie i wniosek końcowy

Przedstawione w niniejszej recenzji uwagi mają głównie charakter dyskusyjny i nie podważają mojej ogólnej pozytywnej oceny rozprawy. Na podstawie powyższej oceny stwierdzam, że przedstawiona do recenzji rozprawa stanowi oryginalne rozwiązanie aktualnego problemu naukowego, natomiast podejście Autora do zagadnień prezentowanych w pracy dowodzi, że posiada umiejętność samodzielnego prowadzenia badań naukowych oraz wykazuje ogólną wiedzę teoretyczną w dyscyplinie naukowej inżynieria mechaniczna.

Uważam, że opiniowana rozprawa mgr. inż. Michała Pajęckiego pt. **„Diagnozowanie wybranych zagrożeń bezpieczeństwa pracy w przedsiębiorstwach produkcyjnych z zastosowaniem technik eksploracji danych”** spełnia warunki stawiane przez ustawę o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki w odniesieniu do rozpraw doktorskich (Dz.U. 2017, poz. 1789 t.j., z późn. zm.). Wnioskuje o dopuszczenie jej do publicznej obrony przed Radą Naukową Dyscypliny Inżynieria Mechaniczna Politechniki Świętokrzyskiej.

Damian Kuczyński