dr hab. Agnieszka Nowak – Brzezińska, Prof. UŚ Katowice, 1 grudnia 2022 r.

Uniwersytet Śląski
Wydział Nauk Ścisłych i Technicznych
Instytut Informatyki
e-mail: agnieszka.nowak-brzezinska@us.edu.pl

RECENZJA ROZPRAWY DOKTORSKIEJ
DLA RADY DYSCYPLINY INFORMATYKA TECHNICZNA I TELEKOMUNIKACJA
POLITECHNIKI ŚLĄSKIEJ

Tytuł rozprawy: Zwiększenie niezawodności pracy urządzeń bloków energetycznych poprzez zastosowanie analityki predykcyjnej

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

Uwagi wstępne

Recenzja opracowana została na podstawie pisma Przewodniczącego Rady Dyscypliny Informatyka Techniczna i Telekomunikacja Politechniki Śląskiej, Profesora dra hab. inż. Andrzeja Polańskiego, z dnia 4 października 2022 r.

Łączna objętość przedłożonej do recenzji pracy wynosi 179 stron.

1. Problem naukowy (teza) rozprawy

Rozprawa poświęcona jest szukaniu rozwiązań pozwalających na zwiększenie niezawodności pracy urządzeń w przemyśle energetycznym. Coraz większą wagę przywiązuje się do predykcyjnego utrzymania ruchu, które eliminuje awarie i pozwala unikać kosztownych przestojów. Zapewne to przyczyniło się bezpośrednio do tego, że Doktorant podjął się badań związanych z analityką predykcyjną i możliwości jakie daje ona w zakresie zastosowań praktycznych. Na chwilę obecną predykcyjne systemy utrzymania ruchu są wykorzystywane jeszcze w niewielkim stopniu, najczęściej w przypadku kluczowych maszyn i systemów lub tych pracujących z dużą intensywnością.
Głównym celem badań było znalezienie obszarów, gdzie możliwe i efektywne staje się wykorzystanie technik związanych w szczególności z predykcyjnym utrzymaniem ruchu oraz opracowanie narzędzi i modeli analitycznych realizujących te zadania. Szczegółowym celem badawczym było opracowanie analitycznego modelu na podstawie danych zbieranych w systemach IT oraz systemach technologicznych, umożliwiającego przewidywanie nadchodzących usterek i awarii. Na podstawie historycznych danych pomiarowych oraz informacji o awariach zaproponowany został heurystyczny model predykcyjny w oparciu o wybrane techniki eksploracji danych.
Uwzględniając podejmowaną w literaturze tematykę technik predykcji usterek została zaproponowana technika oparta o metodę regresji pozwalająca na wykrywanie potencjalnych awarii w przyszłości na postawie zaobserwowanych anomalii pomiędzy sygnałem zmierzonym a jego cyfrową rekonstrukcją za pomocą utworzonego modelu. Dzięki wykorzystaniu opracowanego modelu predykcji możliwe stało się przewidywanie i zapobieganie awariom

urządzeń oraz, w konsekwencji, lepsze wsparcie procesów planowania przeglądów urządzeń, remontów i zakupów.
Ze względu na fakt, że rozprawa dotyka dwóch z pozoru niezależnych obszarów: nauki i przemysłu, tezę rozprawy Doktorant rozłożył na trzy mniejsze tezy badawcze. Pierwsza z nich mówi, że metoda regresja pozwoli na skuteczne przewidywanie awarii i usterek nawet w warunkach dużych zbiorów danych z systemów przemysłowych. W drugiej tezie Doktorant postuluje, że zastosowanie algorytmu adaptacyjnego okna przesuwnego wraz z techniką cyklicznej aktualizacji modelu pozwala osiągnąć wysoką jakość predykcji ograniczając przy tym nadzór zewnętrzny. Trzecia teza mówi, że wsparcie jakie daje nam współcześnie inżynieria danych pozwala automatyzować proces budowania modeli i przenosić raz opracowane rozwiązania na inne typy urządzeń, zachowując zbliżone zdolności predykcyjne.
Uważam, że tak postawione tezy rozprawy zostały jasno sformułowane przez Doktoranta.

2. Zawartość rozprawy

We wstępie, Doktorant bardzo dobrze osadził tematykę rozprawy, podkreślając jej praktyczny (tj. wdrożeniowy charakter i znaczenie dla przemysłu). Skupiając się na przemyśle energetycznym opisał w zwięzły sposób, jakie procesy w nim zachodzą i jakiego rodzaju dane poddawane są analizie. Doktorant zadbał o określenie celów pracy (ogólnego jak i szczegółowego), źródeł danych oraz zdefiniowanie problemu badawczego i tez badawczych. W rozdziale drugim przedstawił przegląd strategii utrzymania ruchu, akcentując przede wszystkim grupę metod predykcyjnego utrzymania ruchu, nie zapominając rzecz jasna o pozostałych tj. korekcyjnej czy prewencyjnej. Dla analityki predykcyjnej przedstawiono, bazując na literaturze światowej, najważniejsze techniki tj. analizę wibracji, termografię, analizę oleju, monitorowanie parametrów procesu, analizę akustyczną, analizę prądu silnika, wizualną inspekcję. Ważną część tego rozdziału stanowi klasyfikacja metod prognozowania stanu technicznego urządzeń, z podziałem na metody oparte na modelu, na analizie danych czy analizie sygnału. Największą wartość, potwierdzającą istotny wkład Autora w analizę literatury dziedzinowej, stanowi w mojej opinii rozdział 2.4., gdzie przedstawiono bardzo skompresowaną analizę prac poświęconych postawionemu problemowi badawczemu. Autor w formie tabeli ujął bardzo krótką charakterystykę tego, co dotąd odkryto w badanym temacie. W moim odczuciu, byłoby ogromną korzyścią dla odbiorców tej pracy gdyby ta synteza miała nieco szersze ramy. Rozdział trzeci zawiera opis metodologii badań oraz wykorzystanych metod i algorytmów przy opracowanych eksperymentach i obliczeniach. Zaczynając od studium przypadku, przez metody symulacji komputerowej, do właściwej eksploracji danych Doktorant poświęcił także uwagę narzędziom użytym w procesie implementacyjnym i przedwdrożeniowym (KNIME, PowerBI, Python), jednakże znów zrobił to w bardzo skondensowanej wersji, nie rozwijając tematu szerzej.
Jednak, niewątpliwie najbardziej znaczące treści znalazły się w sekcji 3.3., gdzie Doktorant szczegółowo opisał opracowane metody i algorytmy, zaczynając od ogólnego opisu metod uczenia maszynowego i przechodząc do algorytmu wykrywania anomalii wykorzystującego metodę regresji.
Opisany algorytm zakłada tworzenie cyfrowego modelu każdego z sygnałów (czujników opisujących stan danego urządzenia) oraz analizę różnic występujących między wartościami rzeczywistymi a estymowanymi. Zakładając, że nieprawidłowe stany urządzeń powinny być możliwe do przewidzenia jako te, w których obserwujemy największe różnice, algorytm wykrywania anomalii wydaje się prosty w implementacji. Doktorant zaproponował definicję modelu wykorzystującego regresję, opisał dość szczegółowo proces tworzenia tegoż modelu, od pozyskania danych i ich wstępnej obróbki, przez proces budowania samego modelu (wymagającego wyboru sygnałów dla których tworzone będą modele) oraz ocenę jakości modelu (z wykorzystaniem macierzy pomyłek).

Automatyzacja procesu tworzenia modelu z ograniczonym wykorzystaniem wiedzy eksperckiej (która jest dobrem trudnodostępnym) możliwa była dzięki parametryzacji proponowanego algorytmu. Po pierwsze, trzeba wybrać grupę sygnałów spośród dziesiątek tysięcy punktów pomiarowych oraz wyznaczyć we właściwy sposób zbiór treningowy do wyuczenia modelu. Trzeba jednak przy tym wszystkim brać pod uwagę także czynniki zewnętrzne warunkujące pracę urządzeń (w literaturze przedmiotu nazywamy to zjawiskiem dryfu koncepcji). Należało więc uwzględnić metody zapobiegania tego typu zjawiskom. Doktorant nie tylko opisał dość szczegółowo algorytmy znane w literaturze, ale również zaproponował modyfikacje istniejących algorytmów (algorytmu sum skumulowanych oraz algorytmu okna adaptacyjnego).
Rozdział czwarty obejmuje prezentację oraz interpretację wyników badań. Doktorant demonstruje w ten sposób zdolności predykcyjne opracowanego modelu, pozwalającego na wczesne wykrywanie usterek. Historycznymi danymi, które wykorzystano do obliczeń były dane pomiarowe urządzeń przemysłowych Elektrowni Jaworzno III oraz Elektrowni Łaziska.
Uzyskane wyniki potwierdziły możliwość zastosowania zaproponowanej metody dla innych typów urządzeń znajdujących się w elektrowniach (studium przypadku obejmowało analizę pracy pomp zasilających blok energetyczny). Niezwykle ważną częścią tegoż rozdziału jest sekcja pt. „Porównanie wyników z innymi metodami" (strony 110 i 111). Ilekroć w przeglądanych pracach widzimy zadanie eksploracji z użyciem sztucznej inteligencji czy uczenia maszynowego, zakładamy od razu, że będą to klasyfikatory (modele klasyfikacji danych). Doktorant krótko przedstawia zastosowane metody tj. drzewa decyzyjne i sieci neuronowe, jednak nie podaje informacji szczegółowych dotyczących zarówno użytych algorytmów (parametrów) jak i środowiska wykonania badań. Można się domyślać, że Autor użył gotowych rozwiązań (np. w ramach środowiska Python), ale tego nie wiemy. Widzimy za to Tabelę 4.6. z wynikami przedstawiającymi średnie wartości metryk ewaluacyjnych, ale niewiele więcej. To rodzi od razu pytania o to, czy jakość predykcji była jedynym kryterium porównania proponowanej metody tj. analizy regresji z innymi metodami uczenia maszynowego, wymienionymi wyżej. Ważną treścią rozdziału czwartego jest także sekcja 4.2., w której Doktorant podjął próbę odpowiedzi na pytanie czy proponowane przez niego rozwiązanie jest na tyle uniwersalne by nadawało się do predykcji awarii czy usterek także innych urządzeń (np. sprężarek natleniających będących podzespołami instalacji odsiarczania spalin w elektrowniach). Wyniki wskazują na możliwość zastosowania proponowanej techniki także w innych urządzeniach aczkolwiek należy mieć świadomość, że to parametry typu zbiór cech wejściowych czy stopień wielomianu regresji będą ostatecznie decydować o skuteczności rozwiązania. Podsumowaniem przeprowadzonych badań, jest rozdział piąty, w którym Doktorant podkreśla to, co było przedmiotem badań, w sensie teoretycznym jak i praktycznym, a także potwierdza prawdziwość postawionych tez badawczych oraz, jak przystało na prawdziwego badacza, określa możliwe kierunki rozwoju badań.

3. Aktualność i ważność tematyki rozprawy

Predykcyjne utrzymanie ruchu, to jedna ze strategii utrzymania ruchu, polegająca na optymalnym użytkowaniu maszyn. Oparta jest na przewidywaniu stanów obiektów w przyszłości i podejmowaniu odpowiednich działań naprawczych, konserwujących i zapobiegawczych w odpowiednim czasie. Jest to możliwe dzięki ciągłej kontroli ich stanu technicznego oraz stałemu monitorowaniu pożądanych parametrów.
Utrzymanie predykcyjne opiera się na analizie dużej ilości danych dostarczanych bezpośrednio z maszyn a następnie analizie tych danych i wykrywaniu anomalii, które mogą świadczyć o ryzyku wystąpienia awarii. Mierzonymi parametrami mogą być np. temperatura, poziom wibracji, zużycie oleju, ciśnienie oraz wiele innych. Strumienie danych pochodzących

z różnego rodzaju czujników gromadzone są najczęściej w chmurze, a ich analiza z użyciem algorytmów sztucznej inteligencji i uczenia maszynowego umożliwia przewidywanie z dużą dokładnością ryzyka wystąpienia awarii. W efekcie, możliwe jest optymalne planowanie prac konserwacyjnych, remontowych, wymian i innych, które minimalizują liczbę wystąpień awarii i pozwalają osiągnąć dłuższe okresy między awariami.
W praktyce, aby skutecznie zarządzać działem utrzymania ruchu zgodnie z podejściem predykcyjnym niezbędny jest odpowiedni system informatyczny umożliwiający integrację danych gromadzonych w różnych źródłach i w różny sposób. To wszystko wyjaśnia potrzebę badań w tym zakresie, bowiem stabilność procesu produkcyjnego i efektywność wykorzystania parku maszynowego to wartości nie do przecenienia. Każdy planowany lub nieplanowany przegląd czy naprawa maszyny to zakłócenie procesu produkcji, które generuje dodatkowe koszty. Predykcyjne utrzymanie ruchu może stać się skuteczną odpowiedzią na nowe wyzwania i wiele wskazuje na to, że w nieodległej przyszłości stanie się standardem w podejściu do utrzymania ruchu.

Przeprowadzone dotychczas badania i wdrożenia potwierdzają oszczędności rzędu od 10% do 40% w kosztach utrzymania ruchu oraz 5% w inwestycjach w sprzęt i maszyny. Zgodnie z praktyką, koszty usunięcia skutków awarii są zwykle co najmniej kilkukrotnie wyższe od nakładów poniesionych na zapobieganie jej wystąpieniu.

4. Oryginalny dorobek Doktoranta, jego znaczenie poznawcze oraz przydatność praktyczna dla nauki i techniki

Oryginalny dorobek rozprawy zawarty jest w rozdziałach 2-4 rozprawy. Do dorobku tego należy zaliczyć:

- Zaprojektowano i przeprowadzono eksperymenty potwierdzające tezy pracy. Tezy pracy zostały udowodnione. Doktorant tak zaprojektował, a następnie przeprowadził badania, aby możliwe stało się udowodnienie głównych tez badawczych pracy. Po pierwsze potwierdzono, że metody redundancji analitycznej oparte na regresji, zastosowane dla dużych zbiorów danych rzeczywistych, pozwalają na przewidywanie awarii czy mniejszych usterek. Po drugie, zaproponowany algorytm adaptacyjny okna przesuwnego, oparty na analizie średnich wartości cząstkowych elementów tego okna, pozwala zachować parametry jakościowe modelu predykcyjnego, a więc pozwala działać sprawnie nawet w przypadku dryfu koncepcji (zmian z powodu czynników zewnętrznych). Po trzecie, Doktorant potwierdził, że proces przygotowania i eksploracji danych, a także tworzenia modeli predykcyjnych, może zostać zautomatyzowany i przeniesiony na inne urządzenia przemysłowe z zachowaniem wysokiej jakości predykcji. Niewątpliwie, opracowana metoda ma duży potencjał wdrożeniowy, dając nadzieję na szerokie zastosowanie w praktyce m.in. w zadaniach monitoringu urządzeń np. w trybie online czy planowania remontów.
- W zakresie podjętej tematyki badawczej, którą jest optymalizacja niezawodności pracy urządzeń w przemyśle energetycznym, można odnotować spory ruch. Sam Doktorant w rozdziale 2.4. w ciekawy i dość przejrzysty sposób przedstawił dokonania innych badaczy w zakresie predykcyjnego utrzymaniu ruchu w przemyśle. Potwierdza to, że prowadzone przez niego badania są bardzo aktualne i wpisują się w aktualnie obierany nurt. W rozdziale drugim Doktorant zauważa popularność metod opartych na analizie

danych tj. systemów rozmytych, sztucznych sieci neuronowych w tym coraz bardziej popularnych algorytmów uczenia głębokiego. Wstępne badania (w których Doktorant zastosował algorytmy sieci neuronowych czy drzew decyzyjnych) nie pozwoliły uzyskać lepszych wyników niż zaproponowane podejście regresyjne. Jednakże należy się spodziewać, że w najbliższym czasie będziemy obserwować wzrost użycia sztucznej inteligencji i uczenia maszynowego w tego typu zadaniach.

- Znaczenie uzyskanych wyników dla danej dyscypliny naukowej można ocenić poprzez porównanie wyników jakości predykcji proponowanego rozwiązania z jakością klasyfikacji dostarczoną przez inne typy klasyfikatorów, m.in. drzewa decyzyjne czy sieci neuronowe. W tym kontekście, wyniki przemawiają na korzyść Doktoranta i proponowanego przez niego rozwiązania. Nie można jednak w tym momencie nie uwzględnić także innych kryteriów jak chociażby złożoność czasowa czy pamięciowa możliwych rozwiązań, a także ograniczeń wynikających z charakteru danych, które w ten sposób możemy analizować. Pamiętajmy więc, że rozwiązanie proponowane w rozprawie sprawdzać się będzie w zastosowaniu dla danych na skali ilościowej (numerycznej nie zaś jakościowej/kategorycznej). W przypadku danych pomiarowych, zdecydowanie jednak częściej spotykamy się z danymi na skali numerycznej, co pozwala potwierdzić duże znaczenie uzyskanych wyników dla danej dyscypliny naukowej.
- Pewnym globalnym miernikiem uznania istotności prac Autora mogą być indeksy bibliometryczne, tj. liczba punktów przyznanych publikacji, wskaźnik Impact Factor czy licznik cytowań. W dorobku Doktoranta odnotowujemy 4 publikacje o zasięgu międzynarodowym, na łączną sumę punktów ministerialnych o równą 400. Wśród tych publikacji znajdziemy m.in. artykuł w czasopiśmie Sensors (Core A, 140 pkt, IF 3,275) oraz dwie publikacje w ramach renomowanych konferencji wpisanych w ramy Core A, tj. MASCOTS (140 pkt) oraz ICCS (140 pkt). Mimo tego, że są to stosunkowo świeże publikacje, doczekały się cytowań (Scopus: 25 cytowań, WoS: 18 cytowań, ResearchGate: 29 cytowań). Dostrzegalność prac Autora jest zatem potwierdzona.

Oceniając całościowo dorobek rozprawy stwierdzam, że Autor w sposób systematyczny zrealizował zaplanowane badania oraz ocenił ich wyniki, dobrze je dokumentując. Doktorant wykazał się dużą starannością oraz pracowitością przygotowując rozprawę.

5. Wiedza Autora oraz znajomość współczesnej literatury z dyscypliny naukowej, której dotyczy rozprawa

Autor rozprawy wykazał się bardzo dobrą znajomością dorobku literaturowego dotyczącego zagadnień, którym poświęcona jest rozprawa. Przegląd tych zagadnień został wykonany w rozdziale pierwszym oraz drugim rozprawy. Podstawą oceny literatury światowej jest spis liczący 209 pozycji bibliografii. Doktorant uwzględnił tu zarówno prace odnoszące do najnowszych badań w temacie rozprawy jak i prace stanowiące genezę analizy regresji czy metodologii analizy predykcyjnej. Na uwagę zasługuje również sporo odwołań do prac (raportów) związanych z dokumentacją techniczną dotyczącą charakterystyki przemysłu energetycznego. Tak ujęty dobór literatury oceniam w pełni pozytywnie. Świadczy on o dostatecznej wiedzy doktoranta w danej dyscyplinie naukowej. Spis ten jest bardzo obszerny i aktualny (są to prace wydane w ostatnich latach).

6. Wady i słabe strony rozprawy, uwagi dyskusyjne

Rozprawa jako całość nie ma istotnych wad. Treść rozprawy, użyty język, sposób prezentacji wyników, a także sformułowane wnioski, które cechuje ceniona zwięzłość i przejrzystość, świadczą o wysokich umiejętnościach Autora do poprawnego i przekonującego przedstawienia uzyskanych wyników.
Wśród słabszych stron bądź uchybień można wymienić następujące:

a) Kilkukrotnie Doktorant powołuje się w pracy na wymagające definicji pojęcia, jednak dopiero znacząco dalej w rozprawie definiuje te pojęcia. Tak było np. z pojęciem „dryfu koncepcji" (pierwsze odwołanie mamy na stronach 2, 13, 63, 69 a właściwą definicję i wyjaśnienie pojęcia widzimy dopiero na stronie 83). Podobnie jest z pojęciem macierzy pomyłek.
b) Zdarza się w pracy niekonsekwencja ze strony Doktoranta w używaniu zamiennie pewnych pojęć bez wcześniejszej definicji. Przykładowo, w odniesienie do zbioru danych poddawanemu eksploracji, na stronie 62 Autor wyróżnia zbiór treningowy, testowy i ewaluacyjny, a na stronie 66 (będących oczywiście odpowiednikami tychże) zbioru uczącego, testowego i walidacyjnego.
c) Autor powinien wyjaśnić pojęcia: „residuum", „wiarygodność modelu", „skalowalne rozwiązania" czy „niespójność danych". Warto byłoby także zdefiniować wszystkie składowe (kolumny) prezentowane w tabelach, np. w tabeli 4.7. Można się domyślać, że w tabeli tej kolumna „Śr." Oznacza wartość średnią ale powinno być to określone przez Doktoranta.
d) Zdarza się (np. w Tabeli 4.7.), że część ułamkowa jest rozdzielona od całkowitej znakiem „ ." zamiast „ ,".
e) Korekta rozprawy jest dość dokładna i w całej pracy nie znalazłam błędów ortograficznych, czy znaczących interpunkcyjnych (kilka braków przecinków), co w opracowaniu liczącym 179 stron jest praktycznie nie do uniknięcia. Znalazłam jednak sporo błędów językowych. Wymienię kilka z nich:
   - str. 9 - *Liczby przetwarzanych danych* zamiast *ilość przetwarzanych danych*
   - str. 31 - *przewidywania usterek przewidywania usterek* zamiast *przewidywania usterek*
   - str 51 - *opisane w szczegółowo* zamiast *opisane szczegółowo*
   - str 45 - *principal component analisys* zamiast *principal component analysis*
   - str 46 – *suport* zamiast *support*
   - str 92 - *treningowego treningowego* zamiast *treningowego*

   i szereg innych, podobnych drobnych usterek, które nie obniżają w sposób znaczący wartości pracy.

f) Znalazłam następujące drobne błędy redakcyjne:
   - Wiele z załączonych rysunków jest nieczytelnych (np. Rysunek 3.12 lub 4.1., 4.13. i dalej 4.14.). Nie powinno się umieszczać tabel jako rysunków w pracach naukowych (np. Rysunek 2.5). Same tabele powinny być opatrzone opisem w postaci nagłówka tabeli nad nią bezpośrednio tak, jak to uczyniono dla Tabel 4.1. czy 4.2.

- Przed wybranymi wzorami tj. 2.2. oraz 3.14 można zauważyć wiele niepotrzebnych pustych linii.
- W pracy można znaleźć niedociągnięcia typu wdowy i bękarty.

g) Jako, że praca dotyczy dyscypliny informatyka spodziewałam się nieco większego wkładu Autora w zagadnienia związane zarówno z analizą proponowanych algorytmów w zakresie np. złożoności obliczeniowej (jako kryterium pozwalającego ocenić wagę danego algorytmu), jak i implementacji wskazanych algorytmów (np. szczegółów dotyczących konkretnego środowiska obliczeniowego). Proponowana metoda jest oceniona w kategoriach jakości klasyfikacji, ale jej ocena nie musi ograniczać się tylko i wyłącznie do tego aspektu. Innymi, potencjalnie interesującymi kategoriami oceny tego typy metod mogą być np. interpretowalność czy złożoność czasowa/pamięciowa. W rozprawie brakuje jednak nieco szerszej (w tym sensie, a więc wielokryterialnej) oceny tej metody. W pracy brakuje także, w mojej opinii, opisów użytej metodologii (przedstawione treści są bardzo ogólne) a także informacji dotyczących analizowanych zbiorów danych (np. ich wielkości). Taka wiedza pozwalałaby rzetelniej ocenić na ile przeprowadzone badania wpisują się w zakres współczesnych wyzwań informatyki (związanych z analizą danych złożonych, i charakteryzujących się dużym wolumenem tj. Big Data), ile wykonano eksperymentów, jak był czas obliczeń, jakie były parametry sprzętowe służące do przeprowadzenia badań eksperymentalnych czy wreszcie informacje o tym, czy przygotowane narzędzia stanowią spójny zestaw, który z powodzeniem może być wykorzystany jako samodzielne oprogramowanie. Autor użył zarówno środowiska Python, KNIME czy Power BI zatem zachodzi pytanie czy w ramach realizowanych prac stworzono jedno spójne narzędzie?

h) Wykrywanie anomalii jak i tematyka systemów eksperckich, bardzo często stosowanych w tworzeniu systemów diagnostycznych, mają bardzo bogaty zasób dotychczasowych rozwiązań. Zastanawia zatem fakt, dlaczego Doktorant nie podjął się w rozprawie szerszego użycia obu tych zagadnień. Jeśli zaś usiłował, to jakie były tego efekty?

i) Na stronie 11 (rozdział 1.4.) Autor pisze, że „*brak dobrze opisanych zdarzeń, niespójność danych archiwalnych, czy braki spowodowane czułością systemu są wyzwaniami, które podejmuje niniejsza praca*". W jaki sposób niniejsza praca podejmuje temat niespójności danych? W rozprawie nie zdefiniowano pojęcia niespójności i metod inżynierii danych, które mogą wtedy w znaczący sposób poprawić skuteczność eksploracji.

j) Na stronie 65 Autor opisuje metody przygotowania danych do analizy, wśród których szczególnie istotne, przy niekompletności danych, są użyte wówczas metody uzupełnienia braków w danych. Autor powołuje się tutaj na uśrednienie bądź przypisanie wartości zerowej. Jednak to drugie rozwiązanie może być bardzo niebezpiecznym krokiem. Czy Autor nie rozważał także innych sposobów imputacji danych jak mediana czy chociażby użycie klasyfikatora kNN do uzupełnienia braków w danych w oparciu o sąsiedztwo?

k) Na stronie 69 Autor wspomina, że za pomocą skryptów Python realizowane były operacje polegające na konsolidowaniu plików z danymi procesowymi z systemów automatyki przemysłowej, łączenie danych procesowych z danymi o usterkach, a także

ekstrakcja cech. Czy stanowi to odrębne narzędzie, udostępnione podmiotom z sektora energetyki?

l) W proponowanych algorytmach (Algorytm 1, strona 75) użyto parametru *k*' oraz *k* i *m*. Jakie są kryteria wyboru *k'* wybranych sygnałów? Jak małe jest *k* w stosunku do *m*?

m) Na stronie 110 umieszczono porównanie wyników z innymi metodami z zakresu predykcji czy klasyfikacji, jak drzewa decyzyjne czy sieci neuronowe. Autor w Tabeli 4.6. prezentuje jakość klasyfikacji algorytmów. Czy wartość w tabeli określa wartość jakości klasyfikacji uzyskaną z uśrednienia wszystkich wykonanych eksperymentów? Jeśli tak, to w mojej ocenie, dla przejrzystości wyników, zdecydowanie bardziej miarodajnym wynikiem byłoby przedstawienie nie tylko wartości średniej, ale również minimalnej czy maksymalnej, by móc właściwie ocenić efektywność tych rozwiązań.

7. Wniosek końcowy

W podsumowaniu stwierdzam, że mimo drobnych uchybień przedstawionych powyżej, przedłożona do recenzji praca doktorska wykonana przez Pana mgra inż. Marka Molędę spełnia w mojej opinii wymagania stawiane kandydatom do stopnia naukowego doktora, określone w art. 14 ust. 2 pkt 2 ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz stopniach i tytule w zakresie sztuki (Dz. U. z 2017 r. poz. 1789), § 6 ust. 1 rozporządzenia Ministra Nauki i Szkolnictwa Wyższego z dnia 19 stycznia 2018 r. w sprawie szczegółowego trybu i warunków przeprowadzania czynności w przewodzie doktorskim, w postępowaniu habilitacyjnym oraz w postępowaniu o nadanie tytułu profesora (Dz. U. z 2018 r. poz. 261) w związku z art. 179 ust. 1 ustawy z dnia 3 lipca 2018 r. – Przepisy wprowadzające ustawę-Prawo o szkolnictwie wyższym i nauce (Dz. U. z 2018 r. poz. 1669, z późn. zm.). W mojej ocenie przedłożona rozprawa zawiera oryginalne rozwiązanie problemu. Doktorant osiągnął stawiany cel, wykazując się niezbędną wiedzą i umiejętnościami do samodzielnego rozwiązywania problemów naukowo-technicznych z wykorzystaniem metod informatycznych.

Wnoszę zatem o dopuszczenie Pana mgra inż. Marka Molędy do następnej fazy przewodu doktorskiego. Ponadto z uwagi na fakt, że niektóre wyniki rozprawy zostały opublikowane w renomowanych czasopismach o zasięgu światowym (Czasopismo Sensors, IF = 3,275), wnoszę o wyróżnienie rozprawy.