



**Politechnika
Śląska**

POLITECHNIKA ŚLĄSKA
WYDZIAŁ AUTOMATYKI, ELEKTRONIKI I INFORMATYKI
KIERUNEK: INFORMATYKA

STRESZCZENIE

Zwiększenie niezawodności pracy urządzeń bloków energetycznych
poprzez zastosowanie analityki predykcyjnej

Autor: mgr inż. Marek Molęda

Promotor: dr hab. inż. Dariusz Mrozek, prof. PŚ

Konsultant: dr inż. Alina Momot

Gliwice, wrzesień 2022

Spis treści

1	Wstęp	2
1.1	Spis publikacji autora	3
1.2	Motywacja	4
1.3	Tezy badawcze	4
2	Rola procesów odtworzeniowych i utrzymaniowych	5
2.1	Opis strategii utrzymania ruchu w przemyśle	6
2.2	Predykcyjne utrzymanie ruchu	7
3	Klasyfikacja metod prognozowania stanu technicznego urządzeń	8
4	Opis opracowanych metod i algorytmów	10
4.1	Algorytm wykrywania anomalii wykorzystujący metodę regresji	10
4.1.1	Ogólna koncepcja metody	10
4.1.2	Model wykorzystujący metodę regresji	11
4.1.3	Proces tworzenia modelu	12
4.1.4	Przetwarzanie i standaryzacja wyników	13
4.2	Parametryzacja modelu i automatyzacja procesu uczenia	14
4.3	Zjawisko dryfu koncepcji - detekcja i sposoby zapobiegania	14
5	Wyniki badań	16
5.1	Algorytm wykrywający usterki wykorzystujący metodę regresji na podstawie zaobserwowanych anomalii	16
5.1.1	Graficzna analiza wybranych wyników	16
5.1.2	Normalizacja i wizualizacja wyników	17
5.1.3	Porównanie wyników z innymi metodami	19
5.2	Strojenie modelu oraz przenaszalność rozwiązania	20
5.2.1	Selekcja cech i strojenie modelu	20
5.2.2	Jakość detekcji modelu	21
5.2.3	Porównanie wyników	23
5.3	Aktualizacja modelu i przeciwdziałanie zjawisku dryfu koncepcji	23
5.3.1	Porównanie wyników uzyskanych przez model stacjonarny oraz adaptacyjny	23
5.3.2	Zastosowanie filtrowania CUSUM	26
6	Podsumowanie wyników badań	29

1 Wstęp

Temat pracy związany jest z projektowaniem opartych na metodach zaawansowanej eksploracji danych rozwiązań informatycznych wspierających procesy produkcyjne w elektrowni. Zaawansowane technologie pozwalają wykorzystać potencjał istniejących zbiorów danych, dając szansę na usprawnienie tych obszarów, które do tej pory nie były wspierane przez technologie informatyczne ze względu na wysokie koszty lub brak odpowiednich narzędzi. Jednym z obszarów zastosowania tego rodzaju zaawansowanej analityki jest predykcyjne utrzymanie ruchu (ang. *Predictive maintenance*), które polega na wczesnym wykrywaniu usterek na podstawie istniejących pomiarów. Predykcyjne utrzymanie ruchu poprzez identyfikację pierwszych symptomów potencjalnego zagrożenia w urządzeniach produkcyjnych na długo przed wystąpieniem awarii pozwala na ograniczenie zarówno nieplanowanych, jak i planowanych przestoju, zwiększając tym samym współczynnik dyspozycyjności urządzenia lub całej instalacji.

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. Z uwagi na skomplikowane procesy technologiczne, szczególna uwaga została skupiona na metodach modelowania niewymagających w dużym stopniu znajomości wiedzy inżynierskiej, a pozwalających osiągnąć skalowalne rozwiązania spełniające zadane oczekiwania odnośnie wiarygodności modelu. 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 podstawie 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.

Zakres badań w obszarze predykcyjnego utrzymania ruchu w elektrowni obejmuje analizę pracy wybranych urządzeń energetycznych wraz z dokumentacją dotyczącą ich awarii oraz stanu technicznego. Jednym z rozpatrywanych zagadnień jest sposób hierarchizacji i prezentacji danych z systemów analitycznych w sposób zrozumiały dla odbiorcy. Często istotne informacje znajdują się poza percepcją adresata, ginąc w szumie informacyjnym. Zaproponowany został także sposób normalizacji wyników oraz sposób interakcji z użytkownikiem poprzez wizualizację wyników oraz wyzwalanie alarmów. Prawidłowa ocena wiarygodności i wartości informacji jest podstawą do zmiany sposobu podejmowania decyzji z intuicyjnego na oparty na analizie danych. Istotnymi zagadnieniami po-

dejmowanymi w pracy były zatem problemy związane z wdrożeniem i późniejszym utrzymaniem rozwiązania. Zaproponowano proces tworzenia i parametryzacji modeli predykcyjnych. Zaproponowane metody zostały zaaplikowane w eksperymencie mającym potwierdzić możliwość automatyzacji procesu tworzenia oraz przenoszenia rozwiązania na inne urządzenia.

W kwestii zagadnień związanych z utrzymaniem rozwiązania zaproponowany został mechanizm ograniczający negatywny wpływ zjawiska dryfu koncepcji na zdolności predykcyjne modelu.

Uzyskane w efekcie eksperymentów wyniki potwierdziły skuteczność proponowanej metody w zadaniu wykrywania poważnych usterek urządzeń energetycznych. W toku prac zaprojektowano system pozwalający na wykrywanie i identyfikację usterek oraz opracowano metody pozwalające na automatyzację procesu tworzenia i aktualizacji modeli predykcyjnych.

1.1 Spis publikacji autora

Wyniki badań niniejszej rozprawy doktorskiej prezentowane były w następujących publikacjach:

1. Moleda Marek, Dariusz Mrozek. "Big data in power generation." International Conference: Beyond Databases, Architectures and Structures. Springer, Cham, 2019. Punktacja MNiSW: 15 pkt.
2. Moleda Marek, Alina Momot, Dariusz Mrozek. "Predictive maintenance of boiler feed water pumps using SCADA data." *Sensors* 20.2 (2020): 571. Punktacja MNiSW: 100 pkt. Impact Factor: 3.275
3. Moleda Marek, Alina Momot, Dariusz Mrozek. "Regression Methods for Detecting Anomalies in Flue Gas Desulphurization Installations in Coal-Fired Power Plants Based on Sensor Data." International Conference on Computational Science (ICCS). Springer, Cham, 2020. Punktacja MNiSW: 140 pkt.
4. Moleda, Marek, Alina Momot, Dariusz Mrozek. "Concept Drift and Avoiding its Negative Effects in Predictive Modeling of Failures of Electricity Production Units in Power Plants." 2020 28th International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS). IEEE, 2020. Punktacja MNiSW: 140 pkt.
5. Moleda, Marek, Dariusz Mrozek, Weiping Ding. "Evolution of maintenance approaches towards smart power industry - a review". *Journal of Engineering Applications of Artificial Intelligence* (w trakcie recenzji). Punktacja MNiSW: 140 pkt. Impact Factor: 6.212

1.2 Motywacja

System energetyczny wymaga wysokiej dyspozycyjności bloków energetycznych w celu zapewnienia zarówno wymaganego poziomu produkcji energii elektrycznej, jak i zabezpieczenia odpowiedniej rezerwy mocy. Klasyczne podejścia i procesy związane z utrzymaniem ruchu w elektrowni oraz zarządzaniem aktywami, mogą zostać uzupełnione lub zastąpione rozwiązaniami opartymi na zaawansowanych metodach eksploracji danych. Rozwiązania takie pozwalają na podejmowanie decyzji w sposób w mniejszym stopniu bazującym na intuicji, a w większym na wiedzy i faktach. Czynnikiem sprzyjającym rozwojowi takiego podejścia jest ciągły rozwój technologiczny pozwalający na aplikację nowoczesnych rozwiązań oraz rozszerzenia wykorzystania istniejących systemów do celów innych niż pierwotnie zakładano.

W związku z postępującą komputeryzacją i cyfryzacją w zakładach przemysłowych przetwarzane są duże liczby danych różnego typu, zarówno w systemach technologicznych, jak i systemach informatycznych. Przetwarzane są dane pomiarowe z systemów produkcyjnych, dane z relacyjnych baz danych, obrazy z monitoringu, czy ręcznie tworzone rejestry i dokumenty. Zakres przetwarzanych danych, złożoność procesów technologicznych, fakt rozproszenia źródeł danych oraz niejasne korelacje między nimi utrudniają ich optymalne wykorzystanie, dające korzyści z uzyskanej wiedzy w wyniku analizy danych. Wiedza dotycząca możliwości wykorzystania tych danych, ich powiązań z procesami biznesowymi oraz opracowanie innowacyjnych narzędzi analitycznych będzie optymalizowała korzyści wynikające z wdrożenia nowoczesnych rozwiązań z obszaru predykcyjnego utrzymania ruchu, a także pokaże ich potencjalny wpływ na procesy biznesowe i technologiczne realizowane w TAURON Wytwarzanie S.A.

Wdrożenie predykcyjnego utrzymania ruchu wymaga wysokich nakładów w postaci inwestycji w dodatkowe systemy pomiarowe, a także systemu informatycznego oraz wyspecjalizowanej wiedzy eksperckiej. Dlatego główną motywacją w pracy jest wykorzystanie istniejących repozytoriów danych oraz w ograniczonym stopniu wiedzy eksperckiej w celu utworzenia systemu realizującego zadania z obszaru predykcyjnego utrzymania ruchu.

1.3 Tezy badawcze

Biorąc pod uwagę zarówno cele badawcze, jak i wdrożeniowe projektu oraz wyżej wymienione wyzwania, zostały sformułowane następujące hipotezy badawcze:

1. Zastosowanie opartych na regresji metod redundancji analitycznej operujących na dużych zbiorach danych z systemów przemysłowych pozwala na realizację zadania przewidywania awarii i usterek w sytuacji:
 - ograniczonej liczby analizowanych cech i zdarzeń,

- zmiennych warunków pracy wynikających z częstych remontów i wpływu czynników zewnętrznych.
2. Algorytm adaptacyjnego okna przesuwnego oparty na analizie średnich wartości cząstkowych elementów tego okna pozwala zachować parametry jakościowe modelu predykcyjnego wykrywającego usterki, jednocześnie ograniczając potrzebę okresowej nadzorowanej aktualizacji modeli.
 3. Proces inżynierii danych i tworzenia modeli predykcyjnych oparty na danych generowanych maszynowo z urządzeń przemysłowych może zostać zautomatyzowany dla zastosowania względem nowych urządzeń, zachowując zbliżone zdolności predykcyjne.

2 Rola procesów odtworzeniowych i utrzymaniowych

Remonty i prace eksploatacyjne są nieodzowną częścią prawie każdego procesu produkcyjnego w przemyśle. Stan wiedzy i procedury wykorzystywane w tej dziedzinie ewoluują w szybkim tempie od czasów pierwszej rewolucji przemysłowej. Istotny wpływ na wsparcie bieżących praktyk w zakresie procesów konserwacji mają nowoczesne technologie. Według raportu Deloitte, niewłaściwe strategie konserwacji mogą zmniejszyć całkowitą zdolność produkcyjną o 5 do 20 procent [1]. Podobnie opracowanie McKinsey prognozuje, że cyfryzacja procesów w przemyśle może zwiększyć dostępność aktywów o 5 do 15 procent i zmniejszyć koszty konserwacji o 18 do 25 procent [2]. Istotne przypadki zastosowania nowoczesnych technologii w przemyśle dotyczą przede wszystkim wsparcia procesu utrzymania ruchu, optymalizacji procesu produkcyjnego oraz poprawy jakości inspekcji [3]. Głównymi czynnikami wpływającymi na korzyść stosowania tych podejść są [4]:

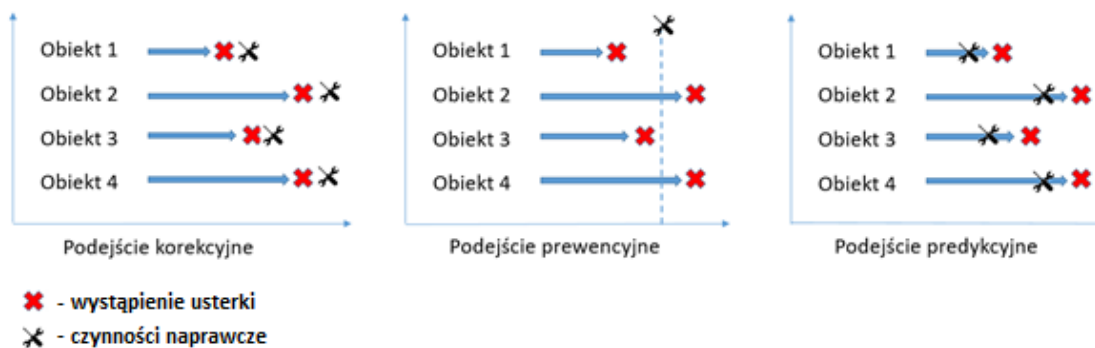
- poprawa dyspozycyjności,
- redukcja kosztów operacyjnych,
- poprawa bezpieczeństwa,
- ograniczenie negatywnego wpływu na środowisko,
- wzrost efektywności i jakości produkcji
- wydłużenie okresu eksploatacji aktywów.

Cyfryzacja i transformacja procesów pozwala na lepsze zrozumienie zachodzących zjawisk oraz podejmowanie trafniejszych i bardziej uzasadnionych decyzji, które w mniejszym stopniu opierają się na intuicji. Ponadto wiedza pochodząca ze zintegrowanych źródeł danych i zaawansowanej analityki umożliwia wdrażanie nowych strategii utrzymania ruchu, lepsze harmonogramowanie zadań, zwiększenie wydajności produkcji oraz podniesienie poziomu bezpieczeństwa. Branża energetyczna przechodzi w ostatnich latach istotne przeobrażenia. Przemiany te wynikają z konieczności znajdowania nowych źródeł energii, zmian klimatycznych i uwarunkowań ekologicznych. Sektor ten wy-

maga więc nowoczesnego podejścia do utrzymania prawidłowego działania urządzeń, aby zapobiec awariom, które mogą mieć poważne konsekwencje zarówno ekologiczne, jak i ekonomiczne.

2.1 Opis strategii utrzymania ruchu w przemyśle

Utrzymanie urządzeń w dobrym stanie technicznym stanowi istotną kwestię w zapewnieniu wysokiej wydajności procesu produkcyjnego. Prawidłowy serwis i konserwacja przyczyniają się do uzyskania wysokiego poziomu dyspozycyjności oraz ograniczają przerwy w produkcji. Z drugiej strony koszty utrzymania stanowią znaczące obciążenie finansowe dla przedsiębiorstwa. Dlatego istotne jest znalezienie równowagi w taki sposób, aby uzyskać akceptowalne wyniki produkcyjne przy jednoczesnej optymalizacji kosztów serwisowych. Zgodnie z definicją znajdującą się w Europejskim Standardzie EN13306 [5] techniki utrzymania należy rozumieć jako *ogół działań technicznych, administracyjnych oraz zarządczych mających na celu utrzymanie urządzenia w okresie jego eksploatacji w stanie, w którym może pełnić wymaganą funkcję lub przywrócenie go do takiego stanu*. Jako zakres czynności konserwacyjnych należy mieć na uwadze działania związane z inspekcjami, monitorowaniem stanu urządzenia, cyklicznymi przeglądami, wymianą części, naprawami, remontami, jak również planowaniem i nadzorowaniem wszystkich tych działań. Strategie utrzymania ruchu mogą zostać sklasyfikowane uwzględniając czas, kiedy wykonywana jest naprawa względem momentu wystąpienia awarii. Wyróżniamy trzy podstawowe podejścia, które przedstawiono na rys. 1 (rysunek na podstawie [6]). Są to podejścia korekcyjne, prewencyjne oraz predykcyjne. W niniejszej pracy szczególna uwaga została skupiona na technikach z obszaru predykcyjnego utrzymania.



Rysunek 1: Podstawowe strategie konserwacji

2.2 Predykcyjne utrzymanie ruchu

Predykcyjne utrzymanie ruchu (ang. *Predictive maintenance*) zakłada dokonywanie czynności serwisowych wtedy, kiedy jest to wymagane, przeważnie jest to czas na krótko przed wystąpieniem spodziewanej usterki. Istotą tego podejścia jest przewidywanie wskaźników zdrowia urządzenia w przyszłości, na podstawie bieżących wyników uzyskanych z okresowych badań diagnostycznych lub monitoringu parametrów urządzenia. Zastosowanie tej metody wymaga posiadania dużego zakresu dobrej jakości danych o stanie technicznym urządzenia, dla którego przewidujemy przyszłe działanie na podstawie bieżących i historycznych pomiarów. [7, 8, 9]. Podstawowym zadaniem tej metody jest wczesne wykrywanie potencjalnych usterek na podstawie pierwszych jej symptomów (niekoniecznie widocznych w czasie inspekcji) [10, 11, 12] jak przedstawiono na rys. 2, co w efekcie prowadzi do ograniczenia zarówno planowanych, jak i nieplanowanych przestoju tzn. :

- ograniczenia zbędnych przeglądów i napraw jeśli urządzenie jest w dobrej kondycji,
- minimalizacji ryzyka awarii poprzez bieżące monitorowanie stanu urządzenia.



Rysunek 2: Koszty naprawy w zależności od czasu wykrycia usterki

Tradycyjne techniki wykorzystywane przy predykcyjnym utrzymaniu ruchu obejmują [13, 14, 15, 16]: analizę wibracji, termografię, analizę oleju, monitorowanie parametrów procesu, analizę akustyczną, analizę prądu silnika i wizualną inspekcję.

Predykcyjne utrzymanie ruchu wymaga posiadania wyczerpujących informacji dotyczących aktualnego stanu urządzenia oraz dużego zaangażowania ekspertów w analizowanie danych. Dzięki

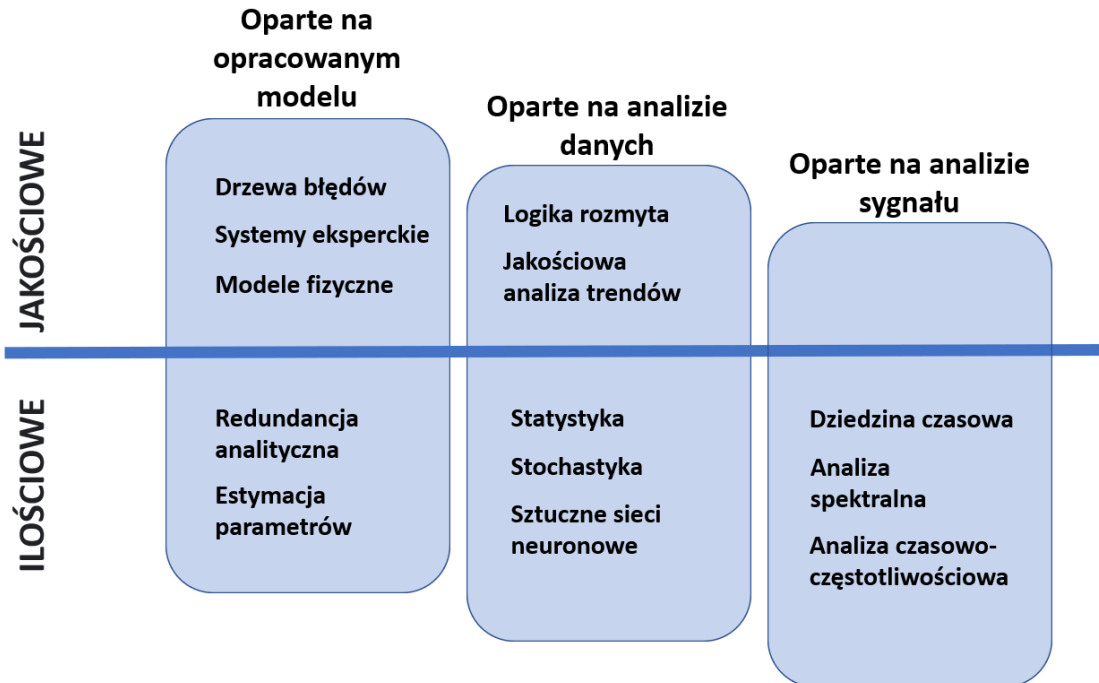
zaawansowanym technologiom z pogranicza IIoT (ang. *Industrial internet of things*) lub Big Data uzyskujemy możliwość łatwego dostępu i integracji danych z istniejących systemów. Rozwijające się techniki przetwarzania danych, w tym sztucznej inteligencji dają natomiast możliwość zautomatyzowania czasochłonnego procesu analizy stanu urządzenia. Systemy OT (ang. *Operational technology*), w tym repozytoria SCADA (ang. *Supervisory control and data acquisition*) mogą być wykorzystywane jako źródła danych procesowych dla systemów wspierających predykcyjne utrzymanie ruchu. W kwestii procesu wspierania decyzji możemy skodyfikować wiedzę ekspercką, np. w postaci reguł i klasyfikatorów, aby monitorować stan zdrowia urządzenia i wykrywać anomalie będące symptomem usterek, literatura przedstawia liczne zastosowania tego typu podejścia [13, 14, 15, 16]. W ten sposób można stworzyć system konserwacji predykcyjnej oparty na monitorowaniu stanu technicznego w czasie rzeczywistym. Podejście predykcyjne jest swojego rodzaju kompromisem pomiędzy podejściem korekcyjnym, a prewencyjnym. Jednak potencjalne koszty dodatkowego opomiarowania, wdrożenia lub architektury sprawiają, że ewentualne zastosowanie powinno być przeanalizowane na podstawie ewentualnych skutków awarii, jakich można by uniknąć. Techniki predykcyjne warto stosować dla urządzeń, których awarie mają wysoki wpływ na produkcję, np. urządzenia krytyczne jak turbina czy kocioł. Przy stosunkowo niskich kosztach wdrożenia, można rozważyć objęcie systemem szerszą grupę urządzeń.

3 Klasyfikacja metod prognozowania stanu technicznego urządzeń

Metody wykrywania i diagnozowania awarii można sklasyfikować ze względu na przyjęte podejście, wykorzystywane zasoby oraz stosowane techniki. Najczęściej metody są pogrupowane ze względu na przyjęte podejście - wyróżniamy tutaj podejścia oparte na analizie danych (ang. *Data-driven*) oraz podejścia oparte na opracowanym modelu (ang. *Model-driven*). Niektórzy autorzy wyróżniają również podgrupy, jak np. metody oparte na wiedzy, analizie sygnału lub probabilistyczne. Zgodnie z taksonomią przyjętą w literaturze i biorąc pod uwagę charakterystykę opisanych modeli, zaproponowano klasyfikację dwuwymiarową, jak przedstawiono na rys. 3.

Ze względu na rodzaj wykorzystywanej informacji oraz sposób projektowania modelu prognostycznego metody prognozowania stanu technicznego urządzeń możemy podzielić na:

- Oparte na opracowanym modelu - grupa technik opartych na modelach wymaga, aby system był zaprojektowany w taki sposób, aby zakodować wiedzę ekspercką oraz zachodzące zjawiska w postaci cyfrowej. Model przeważnie jest projektowany w sposób deterministyczny, np. przy użyciu równań i modelowania matematycznego, tak aby odtworzyć zachowanie systemu. Podejścia te są czasem określane jako białe skrzynki (ang. *White box*), gdzie zależności między wejściami i wyjściami są starannie zaprojektowane i przewidywalne.



Rysunek 3: Klasyfikacja metod prognostycznych w obszarze predykcyjnego utrzymania ruchu

- Oparte na analizie danych - podejścia wykorzystujące duże zestawy danych historycznych oraz techniki, takie jak uczenie maszynowe do tworzenia reguł wnioskowania. Zwyczajowo modele utworzone tą techniką - szczególnie przy wykorzystaniu sztucznych sieci neuronowych - są nazywane czarnymi skrzynkami (ang. *Black box*) ze względu na ograniczony wgląd w strukturę i mechanizmy modelu. Proces projektowania opiera się na starannej selekcji danych treningowych i wyborze odpowiedniej architektury przy procesie uczenia. Wymagają one znacznie mniejszej wiedzy branżowej kosztem nieco większego wkładu analityka zajmującego się przetwarzaniem danych.
- Oparte na analizie sygnału - grupa metod bardzo zbliżonych do tradycyjnie stosowanych metod diagnostycznych. Opierają się one na założeniu, że w mierzonym sygnale (oraz jego cechach statystycznych) odzwierciedlają się stany awaryjne urządzenia. Stosowane techniki opierają się na badaniu pojedynczego sygnału lub pomiaru, wykorzystując techniki eksploracji danych (ekstrakcja cech) oraz dekompozycji/transformacji badanego sygnału.

Ze względu na charakter danych przetwarzanych przez model oraz wzajemne relacje pomiędzy danymi, metody prognozowania stanu technicznego możemy podzielić na:

- Metody ilościowe - podejścia ilościowe koncentrują się na określeniu relacji między wejściem a wyjściem systemu. Metody z tej grupy odpowiadają na pytanie "ile?", determinując tym samym rodzaj przetwarzanych danych do wartości liczbowych. Opierają się na metodach matematycznych, statystycznych lub stochastycznych uwzględniając potencjalną niepewność wyników.
- Metody jakościowe - w modelach jakościowych relacje w systemie są wyrażone przez jakościowe funkcje poszczególnych części systemu, przeważnie sformułowanych w postaci grafów, przepływów lub reguł typu "IF - THEN". Metody jakościowe odpowiadają na pytania: "co?", "jak?", "dlaczego?". Istotą metod z tej grupy jest lepsze poznanie charakteru badanego zjawiska, skupiając się przede wszystkim na zadaniach detekcji i klasyfikacji zdarzeń.

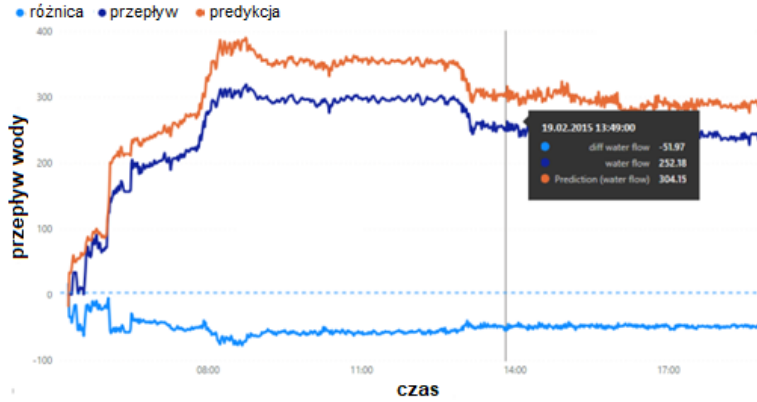
4 Opis opracowanych metod i algorytmów

4.1 Algorytm wykrywania anomalii wykorzystujący metodę regresji

W ramach niniejszej rozprawy opracowano algorytm pozwalający na wykrywanie z wyprzedzeniem usterek urządzeń na podstawie analizy generowanych przez nie danych procesowych. Opracowana metoda z uwagi na niewielką liczbę zarejestrowanych usterek krytycznych wykorzystuje mechanizm wykrywania anomalii, gdzie wykorzystano techniki regresji jako podstawę tworzenia modeli.

4.1.1 Ogólna koncepcja metody

Koncepcja algorytmu wykrywania anomalii zakładała utworzenie cyfrowego modelu każdego z sygnałów pochodzących z czujników podłączonych do urządzenia (czyli sygnałów opisujących stan tego urządzenia) oraz analizę różnic występujących pomiędzy wartościami rzeczywistymi (zmierzonymi) i wartościami oczekiwanymi sygnału (estymowanymi). Wartość oczekiwana jest estymowana na podstawie bieżących wskazań pozostałych czujników (z wyłączeniem czujnika, dla którego wartość estymujemy). Schemat działania odpowiada metodzie analitycznej redundancji i jest analogiczny do koncepcji pełnej rekonstrukcji sygnału (ang. *Full Signal ReConstruction*, FRSC) w metodzie modelowania tzw. "normalnego zachowania" przedstawionej w artykule [12]. Założono, że w czasie poprzedzającym zarejestrowaną usterkę różnica między wartościami estymowaną i rzeczywistą będzie rosła. Przykładowy wykres wartości szacowanych, rzeczywistych oraz ich różnicy dla pomiaru przepływu wody na wyjściu pompy pokazano na rys. 4. Na wykresie widać okres, w którym uszkodzony był zawór minimalnego przepływu, stąd widoczny był duży błąd predykcji.



Rysunek 4: Wartości rzeczywista i estymowana dla pomiaru przepływu wody na wyjściu pompy oraz residuum

Definicja 1 (Model sygnału). *Modelem sygnału w niniejszej pracy będziemy nazywać estymowaną wartość danego sygnału (pomiaru z czujnika) odwzorowaną na podstawie funkcji wartości pozostałych sygnałów z wyłączeniem sygnału, który jest estymowany oraz jego wartości historycznych.*

4.1.2 Model wykorzystujący metodę regresji

Do tworzenia cyfrowych modeli poszczególnych sygnałów wykorzystano metodę regresji wielomianowej. Niewątpliwą zaletą algorytmów opartych na regresji wielomianowej jest szybkość obliczeń oraz przejrzystość. Innymi czynnikami decydującymi o wyborze metod wykorzystujących regresję były zależności liniowe i wysoka korelacja między sygnałami z czujników zamontowanych na monitorowanych urządzeniach przemysłowych.

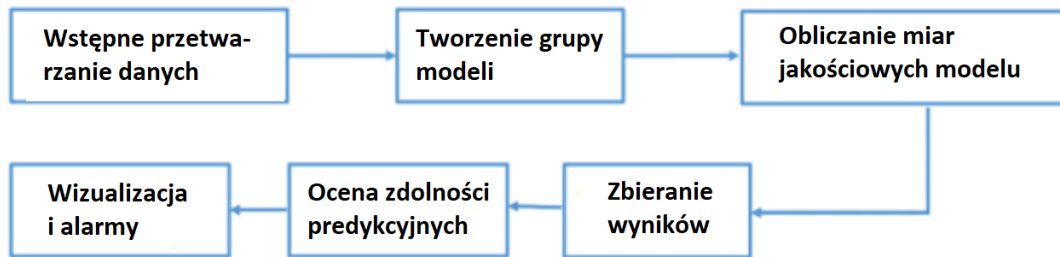
Dla zbioru wszystkich danych wejściowych, gdzie $x = [x_1, x_2, x_3, \dots, x_m]$ jest wektorem pojedynczych pomiarów z m czujników, estymujemy wektor k' wyników dla najbardziej skorelowanych sygnałów objętych monitoringiem, gdzie każda ze składowych jest rekonstrukcją danego sygnału będąc liniową funkcją wszystkich pozostałych zmiennych, tj,

$$\hat{x}_i = f_i(x) = \sum_{\substack{j=1 \\ j \neq i}}^m a_j^{(i)} x_j + a_0^{(i)} + \varepsilon^{(i)} \quad \forall i \in \{1, 2, \dots, k'\}, k' \leq m, \quad (1)$$

gdzie $\varepsilon^{(i)}$ jest wartością residuum dla i -tej estymacji, natomiast współczynniki $a_j^{(i)}$ przyjmują wartości obliczone metodą najmniejszych błędów kwadratowych [17].

4.1.3 Proces tworzenia modelu

Schemat przedstawiony na rys. 5 ilustruje ogólny algorytm wykonywany podczas budowania modelu analitycznego dla zespołu pomp zasilających. W proponowanym algorytmie utworzono grupę modeli opisujących monitorowane urządzenie, dla których wyniki przedstawiane są w postaci znormalizowanej (jako wielokrotność błędu RMSE obliczoną na etapie uczenia modelu). Uwzględniając błąd maksymalny i wyznaczając różne progi alarmowe, usterki mogą być wykrywane na podstawie obserwowanych anomalii (można zaobserwować znaczne odchylenia wartości sygnału od oczekiwanej poprzedzające zarejestrowane awarie).



Rysunek 5: Schemat blokowy etapów działania algorytmu

W procesie budowania modelu predykcyjnego urządzenia na podstawie zestawu danych procesowych o długości N (przedstawionym w Algorytmie 1) brany jest pod uwagę zestaw k sygnałów, dla których tworzone są modele sygnałów (x_1, \dots, x_k) , gdzie $k \leq m$ (m to liczba wszystkich czujników). Dla każdego z tych sygnałów utworzony zostaje model sygnału wykorzystujący metodę regresji, który rekonstruuje ten sygnał na podstawie danych z pozostałych czujników zgodnie z równaniem 1.

W przedstawionym algorytmie 1 opisującym tworzenie grupy modeli przekazywanymi parametrami są: bieżący znacznik czasu T_0 , długość okna czasowego zbioru treningowego ΔT , zbiór k' wybranych sygnałów, próg determinacji stanu pracy τ_1 , próg współczynnika determinacji τ_2 oraz maksymalny stopień wielomianu dla zbudowanego modelu regresji.

Algorithm 1: Tworzenie grupy modeli z eliminacją modeli o słabej jakości

Data: $X, k', T_0, \Delta T, deg, \tau_1, \tau_2$
Result: Modele regresji: $\{f_i\}_{i=1}^k$, współczynniki błędów: $\{Sc_i\}_{i=1}^k, k \leq k'$
 $f = \emptyset$; // pusty zbiór modeli
 $X = [x_1, x_2, x_3, \dots, x_m] = \{x_{ij} : i \in \{1, \dots, m\}, j \in \{1, \dots, N\}\}$;
 $X \leftarrow trim(X, [T_0 - \Delta T, T_0])$; // ustaw zakres zbioru uczącego X
 $X \leftarrow filter(X, missingvalues, \tau_1)$; // Filtruj tylko czas operacji
for $i = 1 \dots k'$ **do**
 $O_i = x_i$; // ustaw wektor wyjściowy wartości objaśnionych
 $I_i = X/x_i$; // wyklucz zmienną objaśnianą
 wyznacz model regresji $f_i(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_m)$ dla $S_i = \{I_i, O_i\}$
 wyznacz współczynniki błędów $Sc_i = (MAE_i, RMSE_i, R_i^2)$ dla $S_i = \{I_i, O_i\}$
 if $R_i^2 > \tau_2$ **then**
 zapisz model f_i w formacie PMML ;
 zapisz współczynniki błędów Sc_i ;
 end
end

4.1.4 Przetwarzanie i standaryzacja wyników

W celu znormalizowania wyników wprowadzono współczynnik znormalizowanego błędu względnego NRE (ang. *Normalized relative error*) do określenia stopnia odchylenia dla i -tej zmiennej w zbiorze danych ($i \in \{1, 2, \dots, k\}$).

Definicja 2 (Znormalizowany błąd względny). *Znormalizowany błąd względny jest ilorazem modułu błędu estymacji pomniejszonego o średni błąd bezwzględny określony dla zbioru treningowego względem średniego odchylenia standardowego również określonego dla zbioru treningowego:*

$$NRE_i = \frac{|x_i(t) - \hat{x}_i(t)| - MAE_i}{RMSE_i}, \quad (2)$$

gdzie $x_i(t)$ i $\hat{x}_i(t)$ to wartość bieżąca i jej estymacja uzyskana za pomocą funkcji regresji ($t \in (T_0, T_0 + \Delta T)$).

Wybierając zmienną o maksymalnej wartości NRE_{max} (zwanej *maksymalnym znormalizowanym błędem względnym*), można zidentyfikować sygnał, który prawdopodobnie wskazuje na przyczynę nadchodzącej awarii.

Definicja 3 (Maksymalny znormalizowany błąd względny). *Maksymalny znormalizowany błąd względny jest maksymalną wartością spośród wszystkich współczynników NRE obliczonych dla wszyst-*

kich modeli f_i dla danego znacznika czasowego w obrębie danej grupy modeli wskazujący jednocześnie na indeks modelu sygnału będący prawdopodobnym źródłem potencjalnej usterki.

$$NRE_{max} = \max(NRE_1, NRE_2, \dots, NRE_k). \quad (3)$$

4.2 Parametryzacja modelu i automatyzacja procesu uczenia

Jednym z celów pracy jest opracowanie algorytmów i modeli z ograniczonym wykorzystaniem wiedzy eksperckiej, zarówno w kwestii wiedzy branżowej, jak i wiedzy analitycznej. Dla rozwiązania tak zdefiniowanego zadania opracowano szereg metod pozwalających na automatyzację procesu tworzenia modeli, co pozwala na wdrożenie opracowanego algorytmu dla różnych urządzeń pracujących w różnych środowiskach. Szczególnie istotnym krokiem w procesie tworzenia modelu jest jego parametryzacja pod względem ustalenia współczynników modelu oraz zbioru cech wejściowych, co ma istotne znaczenie dla jakości zwracanych wyników.

W ramach rozprawy doktorskiej opracowano rozwiązania w zakresie:

- wstępnego przetwarzania i gromadzenia danych w oparciu o kodowanie KKS;
- automatyzacji procesu etykietowania i selekcji danych;
- selekcji cech z wykorzystaniem m.in. algorytmu genetycznego;
- dopasowania wielomianowego modelu w celu poprawy jakości modelu.

4.3 Zjawisko dryfu koncepcji - detekcja i sposoby zapobiegania

Zjawisko dryfu koncepcji (ang. *Concept drift*) w analityce predykcyjnej związane jest ze zmianą w perspektywie czasu właściwości statystycznych procesu lub obiektu, który jest przedmiotem predykcji, co ma wpływ na jakość uzyskiwanych wyników przez utworzony model [18]. Wpływ czynników zewnętrznych, remonty, wymiana podzespołów, zmieniające się warunki pracy mają istotny wpływ na charakterystykę systemu co sprawia, że zmiany w funkcjonowaniu modelu predykcyjnego są nieuchronne.

W ramach prowadzonych prac opracowano następujące rozwiązania mające na celu zniwelowanie zjawiska dryfu koncepcji:

- **Algorytm sum skumulowanych (CUSUM)** – jest używany do wykrywania zmian w szeregach czasowych. Algorytm opiera się na następujących równaniach 4:

$$\begin{aligned} g_0 &= 0, \\ g_t &= \max(0, g_{t-1} + (r_t - v)), \end{aligned} \quad (4)$$

gdzie r_t jest aktualnie obserwowaną wartością, v jest dopuszczalną wielkością zmiany, a t jest aktualnym czasem. Gdy wartość g_t przekroczy ustalony próg λ , wyzwalany jest alarm, a wartość g_t ustawiana jest na zero. W kontekście przeprowadzonych badań algorytm ten służy do odfiltrowania stałych lub wolno rosnących błędów wynikających z pogorszenia właściwości predykcyjnych modelu (czyli domyślnie dryfu koncepcji) od wykrywania gwałtownych zmian (domyślnie awarii).

- **Aktualizacja "na ślepo"** – jest to metoda należąca do grupy metod adaptacyjnych polegająca na okresowej aktualizacji modelu w ustalonych odstępach czasowych. W ramach badań analizowany były wyniki otrzymane przy zastosowaniu różnych konfiguracji okresów odświeżania i długości zbiorów treningowych.
- **Zastosowanie algorytm MSC-ADWIN** – w ramach prac opracowano algorytm na podstawie algorytmu adaptacyjnego okna (ang. *Adaptive Window - ADWIN*, którego celem był dobór zakresu danych testowych przy aktualizacji modelu, w taki sposób aby eliminować zakresy, dla których wykryto zmianę stanu urządzenia (remont, awarię).

Algorithm 2: Zmodyfikowany algorytm MSC-ADWIN

Input : W : zbiór danych wejściowych;
 $step$: rozmiar minimalnego podzbioru cząstkowego;
Set $i_{max} := 0$; $res_{max} := 0$; $p := \lfloor |W|/step \rfloor$;
for $i \leftarrow 1$ **to** p **do**
 –Obliczam średnie wartości cząstkowe–
 $\hat{\mu}_{W_0} :=$ average of the range[1 : i];
 $\hat{\mu}_{W_1} :=$ average of the range[$i + 1$: p];
 oblicz m_h wg. wzoru ??
 if $|\hat{\mu}_{W_0} - \hat{\mu}_{W_1}| * m_h > res_{max}$ **then**
 $i_{max} := i$;
 $res_{max} := |\hat{\mu}_{W_0} - \hat{\mu}_{W_1}| * m_h$;
 end
end
Output: $i_{max} * step$

5 Wyniki badań

Wszystkie prezentowane wykonano z wykorzystaniem rzeczywistych historycznych danych pomiarowych dotyczących badanych urządzeń przemysłowych w Elektrowni Jaworzno III oraz Elektrowni Łaziska. Kolejność przedstawianych wyników odpowiada sekwencji badań, dając sukcesywnie odpowiedzi na kolejne zagadnienia badawcze oraz zgłębiając naturę badanego przedmiotu. Całość badań podzielona jest na trzy etapy, w których:

1. oceniana jest jakość predykcyjna zaproponowanego algorytmu wykrywającego usterki,
2. implementowane są elementy pozwalające na automatyzację procesu tworzenia modeli oraz dobór parametrów modelu,
3. zaproponowany jest mechanizm cyklicznej aktualizacji modelu w kontekście zjawiska dryfu koncepcji.

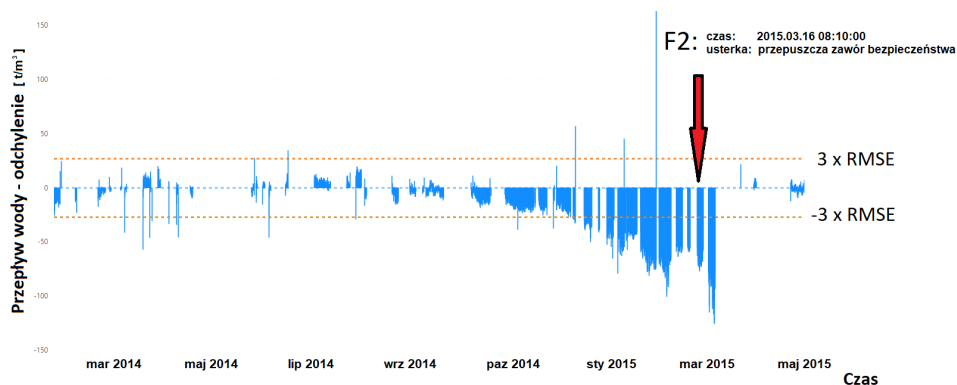
5.1 Algorytm wykrywający usterki wykorzystujący metodę regresji na podstawie zaobserwowanych anomalii

Problemem badawczym rozpatrywanym w niniejszej części było znalezienie modelu pozwalającego na wykrywanie rzadkich usterek o istotnym znaczeniu mogących powodować postój urządzenia. Studium przypadku obejmuje analizę pracy zespoły pomp zasilających w okresie 2013-2017, dla których odnotowano dwa przypadki wystąpienia usterek krytycznych, powodujących wyłączenie urządzenia i w efekcie postój całego bloku energetycznego.

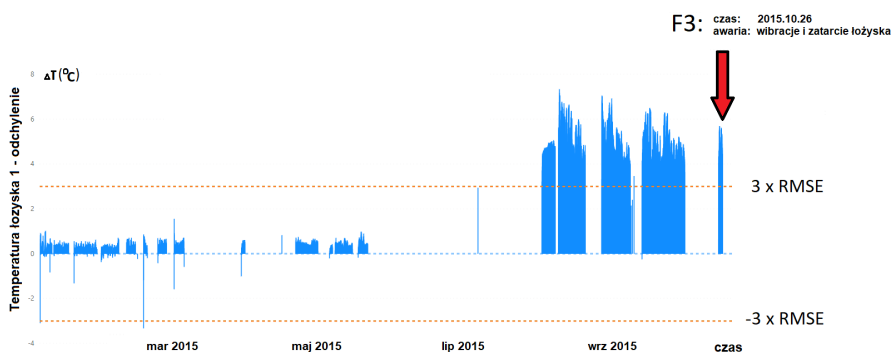
5.1.1 Graficzna analiza wybranych wyników

Analizując różnice między wartością rzeczywistą, a wartością szacunkową poszczególnych sygnałów można zauważyć znaczące odchylenia w okresach poprzedzających czas kiedy zarejestrowano awarie urządzenia. W pierwszym przykładzie (awaria F2) pokazanym na rys. 6 zaobserwowano stopniowy spadek wydajności pompy spowodowany nieszczelnością zaworu minimalnego przepływu. Usterka zaworu powodowała nieszczelność i pracę pompy z mniejszą wydajnością. Stosując zaproponowany próg alarmowy jako trzy odchylenia standardowe ($3 \times RMSE$), można byłoby wykryć to zdarzenie trzy miesiące wcześniej.

Inny przykład (awaria F3) dotyczy nagłego wzrostu odchylenia temperatury łożyska. Anomalia ta wystąpiła po okresie przestoju urządzenia. Przyczyną awarii był brak współosiowości wałów przekładni i silnika oraz stopienie się stopu łożyska. Widocznymi skutkami awarii były drgania i dym wydobywający się z łożyska. Różnice w temperaturach łożyska w czasie przed wystąpieniem tej awarii pokazano na rys. 7.



Rysunek 6: Różnica między rzeczywistą, a szacunkową wartością pomiaru dla przepływu wody.

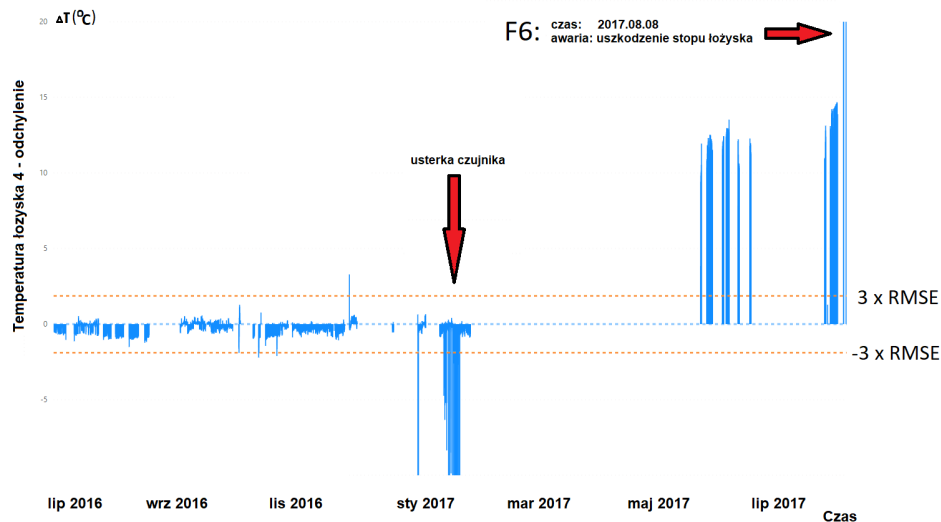


Rysunek 7: Różnica między rzeczywistą a szacunkową wartością łożyska nr 1.

Trzeci przykład (F6) i druga analizowana awaria krytyczna widoczna na rys. 8 przedstawia uszkodzenie stopu łożyska, które spowodowało wzrost temperatury łożyska. Podobnie jak w poprzednim przykładzie, znaczącą zmianę charakterystyki pracy układu zaobserwowano po dłuższym przestoju. W tym przypadku tendencję wzrostową krzywej odchylenia można zaobserwować na kilka dni przed awarią.

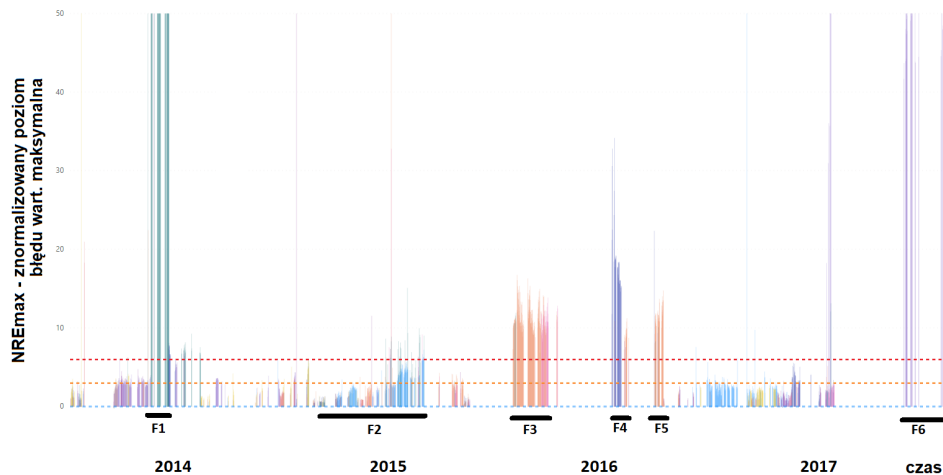
5.1.2 Normalizacja i wizualizacja wyników

Proponowana metoda normalizacji wyników poprzez wprowadzenie współczynnika *NRE* pozwala na wstępną interpretację i priorytetyzację wyników dla odbiorcy końcowego. Wyniki działania algorytmu przy zastosowaniu 30-dniowego okna treningowego dla całego pozostałego okresu przedstawiono na rys. 9. Symbolami F1-F6 oznaczono wykryte (przewidywane) awarie urządzeń, które



Rysunek 8: Różnica między rzeczywistą a szacunkową wartością łożyska nr 4.

rzeczywiście wystąpiły w okresie monitorowania. Dwie poziome linie w kolorze pomarańczowym i czerwonym reprezentują odpowiednio pierwszy i drugi poziom alarmowy. Wszystkie wykryte awarie są opisane bardziej szczegółowo w tabeli 1. Kolor linii na wykresie wskazuje sygnał, dla którego występuje największe odchylenie względne (NRE_{max}).



Rysunek 9: Znormalizowany błąd względny (NRE) z identyfikacją źródła awarii

Tabela 1: Awarie i czas ich zarejestrowania.

Id	Opis	Czas rejestr.	Pocz. usterki	Źródło
F1	Nieprawidłowy pomiar temperatury oleju smarowego przed chłodnicą	25.12.2013	27.11.2013	temp. oleju smar.
F2	Nieszczelność zaworu minimalnego przepływu	27.02.2015	06.01.2015	przepływ wody
F3	Wibracje i dymienie z wewnętrznego łożyska silnika pompy	14.01.2016	26.10.2015	temp. łoż. 1
F4	Wyciek oleju z wału z zewnętrznego łożyska silnika	03.02.2016	18.01.2016	temp. oleju smar.
F5	Słaby przepływ wody chłodzącej przez chłodnicę	08.04.2016	30.03.2016	temp. łoż. 1
F6	Zwiększona temperatura łożyska nr 4 przekładni VOITH	07.08.2017	22.05.2017	temp. łoż. 4

5.1.3 Porównanie wyników z innymi metodami

W celu porównania uzyskanych wyników z innymi algorytmami opisanymi w literaturze pokrewnej, utworzono modele oparte na drzewach decyzyjnych i klasyfikatorach opartych na sztucznych sieciach neuronowych typu MLP (ang. multi-layer perceptron). W obu przypadkach eksperymenty przeprowadzono, dostarczając surowe dane, dodając etap ekstrakcji cech oraz opcjonalnie krok redukcji wymiarowości za pomocą algorytmu PCA. Ekstrakcja cech polegała na obliczeniu dodatkowych zmiennych dla każdego sygnału wejściowego przy użyciu takich funkcji, jak kurtoza, skośność, odchylenie standardowe, minimum, maksimum, wariancja i średnia. Do zbioru danych przypisano etykiety informujące o wystąpieniu awarii dla każdej próbki. Eksperymenty przeprowadzono z wykorzystaniem techniki 10-krotnej walidacji krzyżowej. Zbiór danych został podzielony na 10 równych części, przy czym obliczenia były wykonywane w pętli, gdzie jedna część stanowiła zbiór testowy, a pozostałe 9 części - zbiór treningowy.

Wyniki eksperymentu zawierające takie metryki ewaluacyjne, jak dokładność, pole powierzchni pod krzywą (AUC), czułość i specyficzność, przedstawiono w tabeli 2.

Tabela 2: Jakość klasyfikacji algorytmów

	Dokładność	AUC	Czułość	Swoistość
Proponowana metoda	0,86	0,89	0,67	0,95
Drzewa decyzyjne	0,76	0,72	0,58	0,85
PCA + Drzewa decyz.	0,59	0,56	0,43	0,67
MLP	0,67	0,64	0,2	0,90
PCA + MLP	0,65	0,60	0,37	0,79

5.2 Strojenie modelu oraz przenaszalność rozwiązania

Celem badawczym realizowanym na tym etapie było sprawdzenie uniwersalności wypracowanego algorytmu pod kątem przeniesienia i uruchomienia go na innym urządzeniu, przy czym celami drugorzędnymi było wprowadzenie dodatkowych kroków parametryzujących model podczas procesu uczenia oraz automatyzacja procesu tworzenia modelu. Dodatkowymi elementami, o których mowa są wykorzystanie systemu kodowania KKS w procesie zbierania danych, zastosowanie zautomatyzowanego przepływu danych oraz metod optymalizacji parametrycznej.

Aby udowodnić możliwość przenoszenia prezentowanego rozwiązania na inne urządzenia, wykorzystujemy algorytm zastosowany wcześniej dla predykcji usterek pomp zasilających do analizy danych pochodzących ze sprężarek natleniających. Eksperymenty przeprowadzono na rzeczywistych danych z czujników monitorujących wybrane sprężarki natleniające, które są podzespołami instalacji odsiarczania spalin w elektrowni.

5.2.1 Selekcja cech i strojenie modelu

Dopasowanie wielomianowe. Wyniki uzyskane podczas parametryzacji stopnia wielomianu na 3-miesięcznym zbiorze testowym, przedstawiono w tabeli 3. W przypadku większości zmiennych najlepsze wyniki predykcji uzyskano przy niskim stopniu wielomianu. Im wyższy jest stopień wielomianu, tym model jest bardziej wrażliwy na zmiany w zbiorze danych wejściowych. W przypadku wyższych stopni wielomianu model predykcyjny ujawnia więcej nietypowych sytuacji, jednak działa bardzo niestabilnie i jest podatny na zakłócenia.

Tabela 3: Miary błędów (MSE) w zależności od parametru stopnia wielomianu. Najlepsze wyniki oznaczono kolorem żółtym.

kod sygnału	stopień wielomianu				
	1	2	3	4	5
1HTG01CP025	0,0001	0,0002	0,0002	0,0002	0,0011
1HTG01CS025	2,7962	1,2322	1,8621	29,408	264,76
1HTG01CS026	2,6363	2,1789	2,9565	16,536	34,689
1HTG01CG025	0,00001	0,00001	0,00002	0,0001	0,0002
1HTG01CT025	0,0087	0,0111	0,0196	0,0286	0,1813
1HTG01CT026	0,0067	0,008	0,0173	0,0106	0,0281
1HTG01CT001	12,433	12,643	14,953	19,346	30,349
1HTG01CT020	0,0818	0,0745	0,0791	0,1373	1,1435
1HTG01CE001	0,1006	0,1084	0,1286	0,1659	3,1217
1HTG01CE002	8,0852	8,4233	8,0673	30,904	154,71

Wybór cech. W kolejnym kroku strojenia wybieramy optymalny zbiór cech wejściowych, aby zminimalizować błąd predykcji, a tym samym zmaksymalizować efektywność procesu predykcyjnego. Dzięki odpowiedniemu doborowi parametrów można zwiększyć dokładność modelu o około

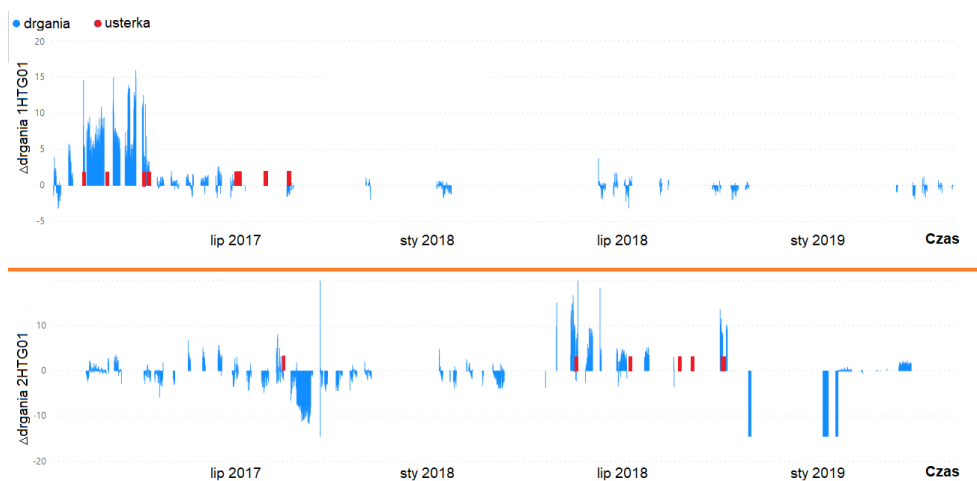
20% (tzn. zmniejszyć błąd średnio-kwadratowy MSE średnio o 20%), co pokazano w tabeli 4. Do optymalizacji parametrycznej wykorzystano dwa algorytmy - eliminację wsteczną (ang. *Backward feature elimination*) i algorytm genetyczny. Zastosowanie obydwu metod optymalizacji dla przedstawianego urządzenia dało takie same wyniki jeśli chodzi o określenie optymalnego zbioru cech oraz poprawę dokładności predykcji. Czas obliczeń w przypadku zastosowania metody eliminacji wstecznej (29 785 kombinacji) w porównaniu z algorytmem genetycznym (27 535 kombinacji) był również podobny.

Tabela 4: Wyniki w postaci błędu średnio-kwadratowego (MSE) uzyskane w efekcie wprowadzenia kroku optymalizacji selekcji cech.

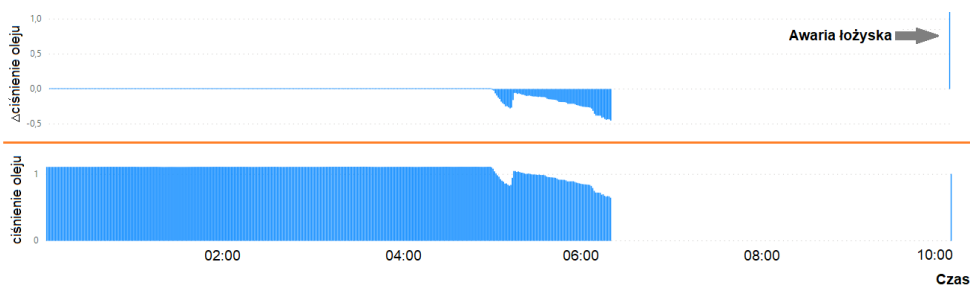
kod sygnału	wszystkie cechy		Eliminacja wsteczna		Algorytm genetyczny		Poprawa %
	liczba cech	MSE	liczba cech	MSE	liczba cech	MSE	
1HTG01CP025	9	0,0001	2	0,0001	2	0,0001	9,76%
1HTG01CS025	9	1,2322	7	1,0971	7	1,0971	12,3%
1HTG01CS026	9	2,1789	5	1,7033	5	1,7033	27,92%
1HTG01CG025	9	0,00001	3	0,00001	3	0,00001	57,77%
1HTG01CT025	9	0,0087	7	0,0073	7	0,0073	19,15%
1HTG01CT026	9	0,0067	6	0,0053	6	0,0053	26,18%
1HTG01CT001	9	12,433	4	10,352	4	10,352	20,09%
1HTG01CT020	9	0,0745	3	0,0499	3	0,0499	49%
1HTG01CE001	9	0,1006	7	0,099	7	0,099	1,58%
1HTG01CE002	9	8,067	3	4,8747	3	4,8747	65,49%

5.2.2 Jakość detekcji modelu

Prognoza długoterminowa. W badanym okresie w rejestrach awarii odnotowano kilka istotnych usterek. W przypadku sprężarki 1HTG01 zgłaszanymi zdarzeniami były w lutym 2017 r. - uszkodzenie łożyska, a następnie stwierdzenie głośnej pracy urządzenia i drgań łożyska w wyniku czego w kwietniu 2017 r. przeprowadzono remont urządzenia. Po remoncie model prognostyczny nie wykazywał żadnych odchyżeń, mimo zgłoszeń zwiększonych drgań w okresie lipiec-sierpień. Analiza rzeczywistych pomiarów drgań nie wykazała żadnych nieprawidłowości w pracy urządzenia. Sprężarka 2HTG01 charakteryzowała się bezawaryjną pracą przez dłuższy okres do maja 2018 r., kiedy to odnotowano podwyższoną temperaturę oleju, a w kolejnych tygodniach pojawiło się kilka usterek związanych z podwyższonym poziomem drgań. Sprężarka została wyremontowana w październiku 2018 r. Wpływ remontu jest widoczny na wykresie na rys. 10 jako znaczne zmniejszenie wartości odchylenia (błędu). W przypadku obu sprężarek punkty awarii widoczne na rys. 10 jako zarejestrowane zdarzenia pokrywają się z większymi odchyleniami wskazanymi przez model predykcyjny dla badanej zmiennej, co świadczy o prawidłowym działaniu modelu.



Rysunek 10: Anomalie dla pomiarów drgań ze wskazaniem na osi czasu usterek i działań serwisowych.



Rysunek 11: Odchylenie i rzeczywiste wskazanie dla ciśnienia oleju za filtrem przed awarią łożyska.

Przewidywanie krótkoterminowe. Jedną z awarii, która została zarejestrowana 23 lutego w sprężarce 1HTG01 (3 dni po wystąpieniu awarii), było uszkodzenie łożyska. Uszkodzenie nastąpiło w czasie, gdy zarejestrowano podwyższone poziomy drgań, co spowodowało konieczność przeprowadzenia kompleksowego remontu. Jednak już przed samą awarią można zauważyć znaczny wzrost odchylenia od wartości szacunkowych. Na około godzinę przed włączeniem urządzenia 20 lutego widoczny jest znaczny wzrost odchylenia. Fizyczny efekt zaobserwowanej awarii miał wpływ na rzeczywisty spadek ciśnienia za filtrem oleju, co widać na rys. 11. Awaria wymagająca postoju wystąpiła trzy dni później podczas kolejnego rozruchu sprężarki.

5.2.3 Porównanie wyników

W tabeli 5 przedstawiono efektywność modeli predykcyjnych zbudowanych zarówno dla sprężarek 1HTG01 i 2HTG02 oraz dla pompy zasilającej. Uzyskane wyniki dla sprężarek są nieco gorsze niż dla pompy. Warto jednak zauważyć, iż przy niewielkiej liczbie rzeczywistych usterek uzyskano dobry poziom swoistości, co przekłada się na niewielką liczbę wyników fałszywie dodatnich (błędów typu I). Czulość modelu prawdopodobnie można poprawić poprzez ulepszenie algorytmów generujących alarmy lub dalsze doskonalenie etapu strojenia.

Tabela 5: Wydajność modeli predykcyjnych zbudowanych dla różnych urządzeń w elektrowni.

	Dokładność	AUC	Czulość	Swoistość
Pompa zasilająca PZ3	0,86	0,89	0,67	0,95
Sprężarka 1HTG01	0,79	0,764	0,58	0,91
Sprężarka 2HTG01	0,93	0,762	0,47	0,98

5.3 Aktualizacja modelu i przeciwdziałanie zjawisku dryfu koncepcji

W niniejszej sekcji analizowane jest zjawisko dryfu koncepcji występujące w kontekście wyników predykcji zwracanych przez opracowany system diagnostyczny dla rzeczywistych danych pomiarowych pompy zasilającej. Celem badań na tym etapie było zbadanie opracowanych technik pozwalających na utrzymanie zdolności predykcyjnych modelu w perspektywie czasu oraz odróżnienie anomalii wywołanych usterką bądź awarią od sytuacji, w której model stracił własności predykcyjne (rozregulował się) i powinien być odświeżony. W części badawczej eksperymentalnie zweryfikowano algorytmy i techniki operujące na szeregach czasowych umożliwiające poprawę zdolności detekcyjnych modelu oraz pozwalające ograniczyć negatywne skutki zjawiska dryfu koncepcji. Podstawowym problemem badawczym jest odróżnienie zjawiska dryfu koncepcji spowodowanego nieaktualnością modelu od sytuacji, w której odchylenie jest spowodowane zbliżającą się usterką. W założeniu opracowywana metoda ma na celu zapewnienie wymaganej dokładności modeli predykcyjnych poprzez automatyczną aktualizację oraz aplikację metod operujących na szeregach czasowych w celu poprawy jakości zwracanych wyników. Zaproponowane rozwiązania mają na celu usprawnienie prac analitycznych i administracyjnych w sytuacji, gdy istnieje wiele modeli obejmujących setki urządzeń i tysiące pomiarów działających w zmiennym środowisku.

5.3.1 Porównanie wyników uzyskanych przez model stacjonarny oraz adaptacyjny

Zaprojektowany eksperyment weryfikuje skuteczność różnych metod aktualizacji modelu oraz porównuje zdolności wykrywania usterek dla poszczególnych wariantów.

W przypadku modeli wykorzystujących regresję, jakość modelu predykcyjnego określa się na podstawie zdolności do predykcji wartości, czego miarą są najczęściej współczynniki, takie jak średni błąd bezwzględny (MAE), błąd średnio-kwadratowy ($RMSE$) lub współczynnik determinacji (R^2). Dla systemów z obszaru predykcyjnego utrzymania ruchu bardziej istotną cechą projektowanego modelu niż odwzorowywanie sygnału jest zachowanie wymaganej zdolności do wykrywania anomalii. W przeprowadzonym eksperymencie porównano wyniki procesu predykcji awarii uzyskane dla różnych wariantów aktualizacji modelu oraz metody stacjonarnej (model nieaktualizowany). Badania eksperymentalne prowadzono dla różnych wariantów aktualizacji modeli predykcyjnych:

- Model stacjonarny – dane treningowe pochodziły z całego roku 2013, natomiast zbiór testowy zawierał dane z okresu 2014-2017.
- Modele adaptacyjne (aktualizacja "na ślepo") – operowały na zbiorze danych z lat 2014-2017, który został podzielony na podzbiory po 160 tys. próbek każdy (około 40 000 próbek odpowiada miesięcowi ciągłej pracy). W pierwszym wariantcie zbiór treningowy obejmował pierwsze 120 000 próbek, a następne 40 000 próbek włączono do zbioru testowego (przesunięcie okna przy jednym cyklu aktualizacji modelu wynosiło 40 000). W drugim wariantcie użyto mniejszego zbioru treningowego (pierwsze 40 000 próbek), natomiast zbiór testowy był równy kolejnym 120 000 próbkom (przesunięcie okna przy jednym cyklu aktualizacji modelu wynosiło 120 000).
- Model adaptacyjny z dostosowanym rozmiarem okna – algorytm jest niemal identyczny, jak w przypadku modelu aktualizowanego "na ślepo", jednak w tym przypadku dodatkową funkcją jest dopasowanie długości okna treningowego do takiego rozmiaru, przy którym wykryto najbardziej znaczącą zmianę wyników predykcji aktualizowanego modelu. Zastosowanie dodatkowego kroku pozwoli na uzyskanie spójnego zbioru treningowego zawierającego najaktualniejsze dane. Długość okna jest wyznaczana z przedziału od 5 do 80 tysięcy próbek.

W tabeli 6 przedstawiono wyniki uzyskane dla modelu stacjonarnego, natomiast w tabeli 7 - wyniki dla modeli adaptacyjnych. Eksperymenty przeprowadzono dla modeli predykcyjnych opracowanych dla sygnałów:

- T_{engine} – temperatury silnika,
- T_{air} – temperatury powietrza przed silnikiem,
- $T_{bearing}$ – temperatury łożyska,
- P_{output} – ciśnienia wody na wyjściu.

Miarami wykorzystanymi po porównania jakości predykcji modeli były: średni błąd bezwzględny (MAE) i błąd średnio-kwadratowy ($RMSE$), gdzie symbole MAE i $RMSE$ opisują parametry

obliczone dla całego zbioru danych, MAE_0 i $RMSE_0$ - uzyskane wyniki dla części zbioru danych, dla której stwierdzono stan awaryjny, MAE_1 i $RMSE_1$ - uzyskane dla części zbioru danych dla stanu normalnej pracy. W celu określenia zdolności do wykrywania usterek/anomalii oznaczono zbiór danych znacznikami czasu, w których zarejestrowano zdarzenia serwisowe dla urządzenia. Następnie obliczono współczynniki RES_{MAE} i RES_{RMSE} , zgodnie z wzorami 5 i 6, które odnoszą się do stosunku wartości MAE i $RMSE$ uzyskanych dla okresu awaryjnego do okresu bezawaryjnego:

$$RES_{MAE} = \frac{MAE_1}{MAE_0}, \quad (5)$$

$$RES_{RMSE} = \frac{RMSE_1}{RMSE_0}. \quad (6)$$

Współczynniki RES_{MAE} i RES_{RMSE} określają w sposób ilościowy jakość modelu predykcyjnego, w założeniu lepszy model będzie posiadał wyższy współczynnik z uwagi na odnotowany wyższy błąd w okresie awarii względem błędu w okresie bezawaryjnym.

Analizując otrzymane wyniki, można zauważyć, że stosując model okresowego odświeżania z wykorzystaniem okna o stałym rozmiarze (wyniki przedstawiono w tabeli 7), uzyskuje się ogólnie niższy średni błąd bezwzględny (MAE) w porównaniu do modelu statycznego (wyniki w tabeli 6), przy zachowaniu nieco mniejszych, ale nadal zadowalających współczynników RES_{MAE} i RES_{RMSE} . Porównując wyniki dla różnych wariantów okna przesuwne o stałym rozmiarze, można zauważyć, że stosując krótsze przesunięcie okna (porównując okno 160/120 z oknem 160/40), uzyskano lepsze parametry MAE i $RMSE$, ale znacznie gorsze parametry RES_{MAE} i RES_{RMSE} . Zastosowanie modelu adaptacyjnego z dostosowanym rozmiarem okna (poprzez zastosowanie algorytmu MSC-ADWIN) poprawia parametry RES_{MAE} i RES_{RMSE} , podczas gdy parametry MAE i $RMSE$ pozostają na porównywalnym poziomie, jak w modelu adaptacyjnym aktualizowanym "na ślepo" w pierwszym wariancie (tj. rozmiar okna 160/120). Zastosowanie metody MSC-ADWIN dało najlepsze wyniki pod względem dopasowania modelu do celów wykrywania anomalii.

Tabela 6: Poziomy błędów przy zastosowaniu modeli statycznych

	T_{engine}	T_{air}	$T_{bearing}$	P_{output}
MAE_0	0,31	0,31	2,08	0,42
$RMSE_0$	0,43	0,45	2,35	0,54
MAE_1	2,36	1,81	7,40	0,47
$RMSE_1$	2,98	3,42	8,37	0,84
MAE	1,23	0,98	4,48	0,44
$RMSE$	2,02	2,32	5,88	0,69
RES_{MAE}	7,63	5,83	3,55	1,09
RES_{RMSE}	6,91	7,63	3,56	1,56

Tabela 7: Poziomy błędów przy zastosowaniu modeli adaptacyjnych

Stały rozmiar okna 160/120				
	T_{engine}	T_{air}	$T_{bearing}$	P_{output}
MAE_0	0,28	0,24	0,85	0,25
$RMSE_0$	0,40	0,45	1,16	0,43
MAE_1	0,87	0,74	0,92	0,81
$RMSE_1$	2,49	2,95	1,53	0,59
MAE	0,91	0,86	0,95	0,85
$RMSE$	1,70	2,01	1,34	0,51
RES_{MAE}	5,04	4,40	1,33	1,55
RES_{RMSE}	6,31	7,33	1,32	1,39
Stały rozmiar okna 160/40				
	T_{engine}	T_{air}	$T_{bearing}$	P_{output}
MAE_0	0,33	0,45	1,26	0,29
$RMSE_0$	0,44	0,58	1,61	0,45
MAE_1	2,41	1,35	2,99	1,28
$RMSE_1$	3,58	2,40	4,41	1,66
MAE	1,27	0,86	2,04	0,74
$RMSE$	2,42	1,67	3,19	1,17
RES_{MAE}	7,21	3,02	2,36	4,32
RES_{RMSE}	6,91	7,63	3,56	1,56
Adaptacyjny rozmiar okna 160/(5-80)				
	T_{engine}	T_{air}	$T_{bearing}$	P_{output}
MAE_0	0,26	0,23	0,96	0,27
$RMSE_0$	0,35	0,38	1,25	0,41
MAE_1	1,26	1,19	1,66	0,46
$RMSE_1$	2,33	2,57	2,52	0,78
MAE	0,71	0,66	1,27	0,35
$RMSE$	1,59	1,75	1,93	0,60
RES_{MAE}	4,87	5,20	1,72	1,72
RES_{RMSE}	6,57	6,73	2,02	1,90

5.3.2 Zastosowanie filtrowania CUSUM

Algorytm CUSUM służy do wykrywania zmian w szeregach czasowych. W kontekście zjawiska dryfu koncepcji odchylenie wartości błędu regresji ϵ_i może być spowodowane zarówno nadchodzącą usterką, jak i innymi czynnikami (powodującymi powstanie dryfu koncepcji). Aby odróżnić zjawisko dryfu koncepcji od prawidłowego działania algorytmu predykcyjnego założono, iż stałe błędy pojawiające się w szerokim horyzoncie czasowym oznaczają dryf koncepcji, natomiast nagłe zmiany następujące w krótkim okresie (np. w ciągu ostatnich dwóch tygodni) oznaczają zbliżającą się usterkę.

Zastosowanie algorytmu CUSUM w założeniach powinno wyzwać alarmy na podstawie gene-

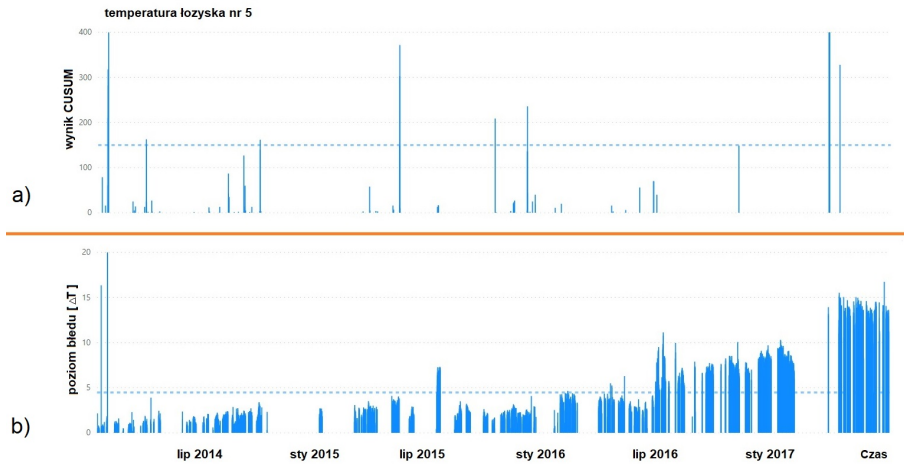
rowanych wyników poprzez filtrowanie stałej wartości błędu wynikającej z istnienia zjawiska dryfu koncepcji. Istotną kwestią jest odpowiedni dobór parametrów określających wrażliwość algorytmu na zmiany. Aby dostosować algorytm do dynamicznie zmieniających się warunków parametr v określający próg zmian, które mają być wykryte nie był ustalany statycznie, a był obliczany na podstawie bieżącego kontekstu (najnowszych danych, przy czym długość okna na potrzeby realizacji zadania ustalono na 2 tygodnie). Parametr v , który był wykorzystany do tego zadania, składał się z ruchomej średniej i odchylenia standardowego z dwóch tygodni w oknie przesuwnym:

$$v = AVG_{2\ weeks} + 3 \cdot ST_DEV_{2\ weeks}. \quad (7)$$

Algorytm CUSUM został wykorzystany do analizy danych wyjściowych (błąd predykcji ϵ_i) dla modelu stacjonarnego (nieaktualizowanego).

Na rys. 12 przedstawiono wizualny wynik działania algorytmu na przykładzie modelu predykcji temperatury łożyska; w górnej części (oznaczonej literą a) znajdują się piki oznaczające wyzwolone alarmy, natomiast w dolnej części (oznaczonej literą b) poziom błędu predykcji. Analizując wykres, można stwierdzić, że zastosowanie filtra CUSUM rozwiązuje częściowo problem przyrostowego dryfu, poprzez odseparowanie stałej wolno zmiennej części błędu. W celu potwierdzenia tej teorii porównano punkty wystąpienia alarmu z prawdopodobnymi przyczynami, tj. zarejestrowanymi zdarzeniami serwisowymi w tym okresie. Wyniki badania przedstawiono w tabeli 8.

Zastosowanie algorytmu CUSUM w roli detektora nagłych zmian w wynikach modelu predykcyjnego dało dobre rezultaty. Miejsca kiedy wyzwolane są alarmy pokrywają się lub poprzedzają okresy odnotowanych zdarzeń serwisowych. Dzięki algorytmowi CUSUM możliwe jest poprawne działanie algorytmu predykcji bez konieczności częstego odświeżania modelu. Oznacza to, iż w przypadku zastosowania wariantów aktualizacji modelu "na ślepo" (zarówno ze stałą długością zbioru uczącego, jak i długością wynikającą z zastosowania algorytmu MSC-ADWIN) możliwe jest wydłużenie cykli odświeżania modelu. Inną zaletą wynikającą z właściwości algorytmu CUSUM jest generowanie alarmów w postaci pojedynczych pików, co ma istotny wpływ na czytelność wyników dla odbiorców końcowych. Zarówno w przypadku prawidłowego wykrycia usterki, jak i fałszywego alarmu algorytm ogranicza się do jednorazowego wyzwolenia alarmu (nie sygnalizując stanu alarmowego cały czas).



Rysunek 12: Przykład zastosowania filtra CUSUM dla danych wyjściowych modelu sygnału: a) alarmy wygenerowane poprzez zastosowanie algorytmu CUSUM na szeregu czasowym błęd ϵ_i będącego wynikiem predykcji z użyciem modelu stacjonarnego, b) poziom błęd ϵ_i wygenerowanego przez model predykcyjny dla estymacji temperatury łożyska nr 5.

Tabela 8: Alarmy wygenerowane po zastosowaniu algorytmu CUSUM oraz powiązane zdarzenia

Data	P_{output}	T_{engine}	T_{air}	$T_{bearing}$	Prawdopodobna przyczyna
2014-01-23	✓	✓	✓	✓	Uszkodzony pomiar temp. łożyska oporowego
2014-03-24				✓	Wyciek oleju między silnikiem, a sprzęgłem
2014-06-26	✓				Nieznane
2014-07-19		✓	✓		Zanieczyszczona chłodnica
2014-08-11	✓	✓	✓		Wyciek z zaworu
2014-09-23				✓	Zanieczyszczona chłodnica
2015-04-25	✓	✓			Wyciek z zaworu bezpieczeństwa
2015-05-05		✓		✓	Zanieczyszczona chłodnica
2015-08-12	✓				Nieznane
2015-08-28			✓		Wyciek oleju
2015-10-05		✓	✓	✓	Remont silników elektrycznych
2015-11-10		✓			Podwyższone wibracje łożyska
2015-11-26				✓	Podwyższone wibracje łożyska
2015-12-02		✓			Podwyższone wibracje łożyska

6 Podsumowanie wyników badań

Główny problem badawczy podejmował wyzwanie utworzenia systemu wnioskowania pozwalającego na predykcyjne wykrywanie usterek w urządzeniach energetycznych elektrowni węglowych. Specyfika pracy urządzeń energetycznych sprawia, iż potencjalne usterki krytyczne mają charakter unikalny, a warunki i charakterystyka pracy systemu są bardzo zmienne.

Wobec powyższego problemu przeprowadzono szereg badań poświęconych zaprojektowaniu modelu analitycznego realizującego to zadanie oraz ewaluacji otrzymanych wyników. Specyfika systemów przemysłowych wiąże się z generowaniem w sposób maszynowy dużych rozmiarów danych, np. z czujników, systemów zabezpieczeń, czy elementów wykonawczych. Mnogość generowanych danych do analizy stanowi niewątpliwą zaletę względem potencjalnego stanowiska badawczego w laboratorium. Opierając się o tę zaletę zaproponowany został algorytm wykorzystujący efektywną obliczeniowo metodę regresji, który schematem działania (tj. wykrywania usterek) zbliżony jest do metody redundancji analitycznej. Występujące istotne korelacje pomiędzy sygnałami procesowymi sprawiły, iż zaproponowana metoda daje rezultaty o bardzo dobrej jakości. Natomiast oparcie głównego trzonu algorytmu o mechanizm wykrywający anomalie pozwolił na ograniczenie wpływu na wyniki niepewnych lub niekompletnych danych o awariach. Przeprowadzone badania na zespole pomp zasilających wykazały zdolność zaproponowanego modelu do wykrywania usterek krytycznych oraz innych zdarzeń mających wpływ na proces produkcyjny. Analiza otrzymanych wyników wykazała również, iż zaproponowane podejście bardzo dobrze nadaje się też do wykrywania awarii czujników oraz do procesu ich kalibracji. Innym zastosowaniem jest wykrywanie usterek mających wpływ na wydajność urządzenia. Takim przypadkiem było wykrycie rozszczelnienia zaworu zasilającego. Algorytm nie generował alarmów w przypadku pomniejszych usterek, jak wycieki oleju czy niewielkie wycieki wody, oraz przy okresowych czynnościach eksploatacyjnych, jak wymiana filtrów oleju. Porównanie wyników z innymi metodami opierającymi się na klasyfikacji dało satysfakcjonujące wyniki na korzyść opracowanego podejścia, przy czym dodatkowym atutem jest klarowny sposób działania i możliwość łatwej interpretacji wyników.

Oceniając wyniki działania zaproponowanego algorytmu dla zespołu pomp zasilających, należy stwierdzić, że główna teza badawcza mówiąca iż:

Teza 1: *Zastosowanie metod redundancji analitycznej opartych na regresji operujących na dużych zbiorach danych z systemów przemysłowych pozwala na realizację zadania przewidywania awarii i usterek w sytuacji:*

- *ograniczonej liczby analizowanych cech i zdarzeń,*
- *zmiennych warunków pracy wynikających z częstych remontów i wpływu czynników zewnętrznych*

została udowodniona.

Charakter wdrożeniowy projektu oraz osadzenie rozwiązania w realiach produkcyjnych wymagają opracowania mechanizmów pozwalających na łatwe wdrożenie i późniejsze utrzymanie rozwiązania. Podążając ku tym celom realizowany był wątek badawczy poświęcony możliwościom przenoszenia rozwiązania oraz aktualizacji opracowanego modelu. Szereg przeprowadzonych badań nad różnymi sposobami aktualizacji modelu pozwalają na świadome dobranie strategii odświeżania do wymaganych parametrów jakościowych. Na szczególną uwagę zasługują wyniki uzyskane przy zastosowaniu autorskiej metody MSC-ADWIN opierającej się na zmodyfikowanym algorytmie okna adaptacyjnego. Wykorzystując właściwości algorytmu, na podstawie analizy średnich cząstkowych całego zbioru treningowego dopasowywany był rozmiar okna treningowego w taki sposób, aby docelowy zakres był jak najbardziej spójny pod względem stabilności przewidywanych wartości. Wyniki badań zrealizowanych dla sprężarek natleniających wykazały, iż zastosowanie zaproponowanej metody znacząco poprawia parametry jakościowe modelu regresji zachowując wysokie zdolności predykcyjne. Zastosowanie metody adaptacyjnego okna wraz z techniką cyklicznej aktualizacji modelu sprawia, iż utrzymanie modelu nie wymaga znaczącego wkładu pracy przez administratora lub analityka nadzorującego pracę systemu. Zaprezentowane wyniki poświadczają zasadność drugiej z tez badawczych niniejszej pracy, tj.:

Teza 2: *Algorytm adaptacyjnego okna przesuwnego oparty na analizie średnich wartości cząstkowych elementów tego okna pozwala zachować parametry jakościowe modelu predykcyjnego wykrywającego usterki, jednocześnie ograniczając potrzebę okresowej nadzorowanej aktualizacji modeli.*

Ważnym wątkiem w podjętych badaniach z projektowego punktu widzenia jest zapewnienie mechanizmów pozwalających na łatwe wdrożenie i utrzymanie modelu. Proces tworzenia modelu we wstępnej fazie eksperymentalnej obejmował ręcznie wykonywane czynności związane z parametryzacją i doбором cech. Parametry modelu były wyznaczone na podstawie dodatkowych eksperymentów lub wnioskowane na podstawie analizy deskryptywnej. Z uwagi na czasochłonność wspomnianego procesu inżynierii danych, a także konieczność posiadania specjalistycznej wiedzy eksperckiej, podjęto próby zautomatyzowania tego procesu.

Znając w praktyce wszystkie etapy tworzenia modelu dla zespołu pomp zasilających, te same kroki zostały odtworzone stosując zautomatyzowane przepływy danych dla dwóch agregatów sprężarek natleniających. Wprowadzone usprawnienia pozwalały na selekcję oraz filtrowanie cech dla zbioru treningowego oraz dostrojenie parametrów modelu. Zastosowane heurystyki pozwoliły na zbudowanie modelu wykazującego zbliżone właściwości predykcyjne w porównaniu z wcześniejszym eksperymentem (dotyczącym pomp zasilających). Porównanie wpływu kroku parametryzacji w ramach eksperymentów na sprężarkach zasilających, zgodnie z oczekiwaniami, dało znacząca poprawę jakości predykcji. Automatyzacja procesu inżynierii danych wymaga od użytkowników wprowadzania dobrej jakości danych wykorzystywanych później do analizy (np. do etykietowania czasów usterek, postojów). Dane generowane maszynowo mają natomiast tę zaletę, iż mają stałą strukturę

oraz są generowane w znaczących ilościach. Dzięki temu ewentualne braki lub pomyłki w danych z systemów informatycznych mogą być kompensowane dużą liczbą danych z systemów przemysłowych. Zastosowanie powyższych rozwiązań w zadaniu automatyzacji procesu budowania modeli w kontekście uzyskanych wyników dla agregatów sprężarek natleniających pozwala na potwierdzenie postawionej tezy:

Teza 3: *Proces inżynierii danych i tworzenia modeli predykcyjnych oparty na analizie danych generowanych maszynowo z urządzeń przemysłowych może zostać zautomatyzowany dla zastosowania względem nowych urządzeń, zachowując zbliżone zdolności predykcyjne*

Literatura

- [1] E. D. Chris Coleman, Satish Damodaran, “Predictive maintenance and the smart factory.” <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/process-and-operations/us-cons-predictive-maintenance.pdf>, 2017. Last accessed 15 Apr 2021.
- [2] S. Bradbury, B. Carpizo, M. Gentzel, D. Horah, and J. Thibert, “Digitally enabled reliability: Beyond predictive maintenance.” <https://www.mckinsey.com/business-functions/operations/our-insights/digitally-enabled-reliability-beyond-predictive-maintenance#>, 2018. Last accessed 15 Apr 2021.
- [3] IoT-analytics, “Industrial AI Market Report 2020-2025.” <https://iot-analytics.com/the-top-10-industrial-ai-use-cases/>, 2019. Last accessed 15 Apr 2021.
- [4] M. M. Mark Haarman, “Predictive Maintenance 4.0 Beyond the hype: PdM 4.0 delivers results.” <https://www.pwc.be/en/documents/20180926-pdm40-beyond-the-hype-report.pdf>, 2018. Last accessed 15 Apr 2021.
- [5] “Maintenance - maintenance terminology,” standard EN13306, European Committee for Standardization, Brussels, 2010.
- [6] “Predictive Maintenance in Manufacturing Overview.” <https://docs.microsoft.com/en-us/previous-versions/azure/industry-marketing/manufacturing/predictive-maintenance-overview>. Last accessed 14 Apr 2021.
- [7] I. H. Afefy, “Reliability-centered maintenance methodology and application: a case study,” *Engineering*, vol. 2, no. 11, pp. 863–873, 2010.
- [8] M. Bevilacqua and M. Braglia, “The analytic hierarchy process applied to maintenance strategy selection,” *Reliability Engineering & System Safety*, vol. 70, no. 1, pp. 71–83, 2000.
- [9] L. Swanson, “Linking maintenance strategies to performance,” *International journal of production economics*, vol. 70, no. 3, pp. 237–244, 2001.
- [10] M. Nouri, B. K. Fussell, B. L. Ziniti, and E. Linder, “Real-time tool wear monitoring in milling using a cutting condition independent method,” *International Journal of Machine Tools and Manufacture*, vol. 89, pp. 1–13, 2015.
- [11] C. Yang, J. Liu, Y. Zeng, and G. Xie, “Real-time condition monitoring and fault detection of components based on machine-learning reconstruction model,” *Renewable energy*, vol. 133, pp. 433–441, 2019.

-
- [12] J. Tautz-Weinert and S. J. Watson, “Using SCADA data for wind turbine condition monitoring—a review,” *IET Renewable Power Generation*, vol. 11, no. 4, pp. 382–394, 2016.
- [13] Y. Feng, Y. Qiu, C. J. Crabtree, H. Long, and P. J. Tavner, “Monitoring wind turbine gearboxes,” *Wind Energy*, vol. 16, no. 5, pp. 728–740, 2013.
- [14] Y. Qiu, Y. Feng, P. Tavner, P. Richardson, G. Erdos, and B. Chen, “Wind turbine SCADA alarm analysis for improving reliability,” *Wind Energy*, vol. 15, no. 8, pp. 951–966, 2012.
- [15] Y. Qiu, L. Chen, Y. Feng, and Y. Xu, “An approach of quantifying gear fatigue life for wind turbine gearboxes using supervisory control and data acquisition data,” *Energies*, vol. 10, no. 8, p. 1084, 2017.
- [16] Y. Qiu, Y. Feng, J. Sun, W. Zhang, and D. Infield, “Applying thermophysics for wind turbine drivetrain fault diagnosis using SCADA data,” *IET Renewable Power Generation*, vol. 10, no. 5, pp. 661–668, 2016.
- [17] D. A. Freedman, *Statistical Models: Theory and Practice. Revised Edition*. New York: Cambridge University Press, 2009.
- [18] J. Gama, *Knowledge discovery from data streams*. Boca Raton, US: CRC Press, 2010.