Politechnika Śląska

Rozprawa Doktorska

Interpretowalne metody analizy niezawodności i przeżycia z zastosowaniem do predykcyjnego utrzymania ruchu

Marek Hermansa

Promotor: dr hab. Beata Sikora, prof. PŚ
Promotor pomocniczy: dr inż. Łukasz Wróbel

Gliwice 2025

Spis treści

1 Wprowadzenie

Analiza niezawodności (ang. *reliability analysis*) i analiza przeżycia (ang. *survival analysis*) to dziedziny inżynierii i statystyki, których przedmiotem jest modelowanie danych typu czas-do-zdarzenia (ang. *time-to-event*). W odróżnieniu od klasycznych problemów klasyfikacji i regresji, opierających się na obserwacjach kompletnych, metody analizy niezawodności i przeżycia umożliwiają pracę z danymi niepełnymi. W przypadku takich danych informacja o momencie wystąpienia zdarzenia będącego przedmiotem zainteresowania jest dla części obserwacji niekompletna. Dane tego typu określane są jako cenzurowane (ang. *censored*), ponieważ obserwacja zdarzenia zostaje przerwana przed jego wystąpieniem z przyczyn niezależnych od badanego procesu [1, 2]. Zastosowanie klasycznych metod klasyfikacji i regresji do danych cenzurowanych wymagałoby odrzucenia takich obserwacji, co prowadziłoby do utraty informacji o procesie degradacji oraz do obciążenia wyników analizy. Ta różnica metodologiczna sprawia, że analiza przeżycia wymaga zastosowania specjalistycznych technik statystycznych oraz algorytmów uczenia maszynowego (ang. *machine learning*, ML).

Analiza niezawodności, jako zbiór metod analitycznych, dotyczy badania trwałości i funkcjonalności systemów technicznych i koncentruje się na analizie czasów do awarii maszyn, urządzeń i komponentów, co umożliwia prognozowanie ich żywotności oraz planowanie strategii konserwacji [3]. Analiza przeżycia, wywodząca się ze statystyki medycznej, czyli dziedziny biostatystyki, zajmuje się modelowaniem czasów do wystąpienia określonych zdarzeń medycznych, takich jak nawrót choroby czy zgon pacjenta [4]. Pomimo różnego pochodzenia i kontekstów zastosowań, oba podejścia opierają się na podobnych fundamentach matematycznych i stosują analogiczne metody statystyczne do modelowania procesów czasowych, w których znaczenie ma obecność danych cenzurowanych.

Metody analizy niezawodności i przeżycia znajdują zastosowanie w różnych dziedzinach nauki i przemysłu. W medycynie wykorzystuje się je do oceny sku-

teczności terapii i analizy czynników prognostycznych [1], natomiast w inżynierii niezawodności wspomagają projektowanie systemów oraz planowanie konserwacji [5, 6]. W ekonomii i finansach umożliwiają modelowanie ryzyka kredytowego i analizę czasu do bankructwa [7], a w socjologii i demografii służą do badania momentów wystąpienia zdarzeń społecznych, takich jak znalezienie pracy, rozpad związku czy zakończenie edukacji [7, 8, 9]. Szczególnie ważnym obszarem zastosowań jest predykcyjne utrzymanie ruchu (ang. *Predictive Maintenance, PdM*), w którym metody te wykorzystywane są do przewidywania wystąpienia awarii maszyn i urządzeń przemysłowych [10, 3].

Charakterystyczną cechą danych w analizie niezawodności i przeżycia jest ich cenzurowany charakter, wynikający z ograniczeń czasowych badań, zróżnicowanych warunków obserwacji oraz heterogeniczności analizowanych obiektów. W predykcyjnym utrzymaniu ruchu cenzurowanie oznacza, że dla części urządzeń nie obserwuje się awarii w analizowanym okresie: mogą być one nadal eksploatowane, wymienione profilaktycznie lub wycofane z innych przyczyn. Podobnie w medycynie, pacjenci mogą opuścić badanie przed wystąpieniem interesującego zdarzenia, a w analizach ekonomicznych przedsiębiorstwa mogą kontynuować działalność poza horyzont czasowy obserwacji. Ta niepełna informacja o czasach zdarzeń stanowi wyzwanie metodologiczne, którego tradycyjne metody klasyfikacyjne i regresyjne nie są w stanie odpowiednio uwzględnić, co prowadzi do obciążonych estymacji i nieadekwatnego modelowania [2].

W obszarze analizy niezawodności i przeżycia wyróżnia się kilka głównych kategorii metod, odmiennych pod względem podejścia metodologicznego, poziomu złożoności oraz stopnia interpretowalności. Przy czym, przyjęto rozróżnienie dwóch powiązanych pojęć: interpretowalność (ang. *interpretability*) oraz objaśnialność (ang. *explainability*). Interpretowalność jest cechą modelu i oznacza, że mechanizm generowania predykcji jest dla człowieka zrozumiały bez potrzeby dodatkowych wyjaśnień (np. drzewa decyzyjne, reguły). Objaśnialność natomiast odnosi się do technik post-hoc, które dostarczają wyjaśnień dla modeli nieinterpretowalnych lub dla pojedynczych predykcji (np. LIME, SHAP). W dalszej części pracy konsekwentnie stosuje się oba pojęcia: w odniesieniu do modeli o przejrzystym mechanizmie działania używa się terminu *interpretowalne*, natomiast w odniesieniu do wyjaśnień modeli typu „czarna skrzynka” (ang. *black box*) lub wyjaśnień pojedynczych predykcji — terminu *objaśnialne* (por. Rozdział 4).

Klasyczne metody statystyczne, reprezentowane przez estymator Kaplana-Meiera oraz model proporcjonalnych hazardów Coxa, stanowią podstawy analizy przeżycia [1]. Estymator Kaplana-Meiera umożliwia nieparametryczną estymację funkcji przeżycia, czyli funkcji opisującej prawdopodobieństwo niewystąpienia zdarzenia (np. awarii urządzenia, zgonu pacjenta) przed określonym momentem czasowym. Z kolei model Coxa pozwala analizować wpływ zmiennych objaśniających na funkcję hazardu, opisującą chwilowe ryzyko wystąpienia zdarzenia w danym momencie, pod warunkiem że nie wystąpiło ono wcześniej, przy założeniu proporcjonalności hazardów. Metody te cechują się wysoką interpretowalnością i są szeroko akceptowane zarówno w środowisku akademickim, jak i w praktyce przemysłowej ze względu na przejrzystość procesu wnioskowania statystycznego.

Rozwój uczenia maszynowego umożliwił zastosowanie zaawansowanych algorytmów do problemów analizy przeżycia. Wśród interpretowalnych metod w tej dziedzinie dominują przeżyciowe drzewa decyzyjne [11] i przeżyciowe reguły logiczne [12], przystosowane do specyfiki danych cenzurowanych. Równolegle rozwijane są bardziej złożone, nieinterpretowalne metody, takie jak przeżyciowe głębokie sieci neuronowe, zespoły klasyfikatorów oraz metody kernelowe [13, 10], które oferują wysoką skuteczność predykcyjną kosztem transparentności modeli.

Klasyczne metody statystyczne charakteryzują się interpretowalnością, ale napotykają na ograniczenia w modelowaniu wielowymiarowych, nieliniowych zależności typowych dla współczesnych danych przemysłowych i medycznych. Model Coxa zakłada proporcjonalność hazardów oraz liniowość wpływu zmiennych objaśniających na logarytm funkcji hazardu, co może okazać się nieadekwatne w przypadku złożonych procesów degradacyjnych. Z kolei estymator Kaplana-Meiera, jako metoda nieparametryczna, nie uwzględnia wpływu zmiennych towarzyszących na funkcję przeżycia, co ogranicza jego możliwości predykcyjne w analizach wielowymiarowych.

Zaawansowane metody uczenia maszynowego, takie jak głębokie sieci neuronowe czy zespoły klasyfikatorów, choć oferują wysoką skuteczność w modelowaniu złożonych wzorców danych, klasyfikowane są jako systemy typu „czarna skrzynka” ze względu na brak transparentności procesów decyzyjnych [14]. W zastosowaniach krytycznych, gdzie decyzje modeli mają bezpośredni wpływ na bezpieczeństwo, efektywność ekonomiczną lub życie ludzkie, brak zrozumienia mechanizmów decyzyjnych stanowi barierę we wdrażaniu tych technologii.

W kontekście potrzeby łączenia wysokiej skuteczności predykcyjnej z interpretowalnością modeli, szczególne znaczenie w analizie przeżycia zyskują zaawansowane metody regułowe i drzewiaste. Metody te umożliwiają modelowanie nieliniowych, wielowymiarowych zależności przy zachowaniu transparentności procesów wnioskowania.

Drzewa decyzyjne (ang. *decision trees*) można interpretować jako hierarchicznie uporządkowany zbiór reguł, gdzie każda ścieżka od korzenia do liścia odpowiada pojedynczej regule. Z kolei reguły logiczne oferują większą swobodę w modelowaniu zależności między atrybutami, eliminując ograniczenia typowe dla drzew. W przeciwieństwie do drzew decyzyjnych, które wymagają przejścia przez wszystkie węzły na ścieżce od korzenia do liścia, reguły umożliwiają wnioskowanie bez stałej kolejności warunków oraz mogą opierać się na podzbiorach atrybutów. Mogą one opisywać przecinające się zbiory przykładów, podczas gdy drzewa charakteryzują się rozłącznym podziałem przestrzeni decyzyjnej, co może ograniczać zdolność modelowania złożonych wzorców w danych. Dodatkowo, reguły logiczne lepiej radzą sobie z brakami danych, bez konieczności stosowania podziałów zastępczych (ang. *surrogate splits*) charakterystycznych dla drzew, ułatwiając interpretację modelu. Główną wadą metod regułowych jest zazwyczaj wyższa złożoność obliczeniowa procesu indukcji w porównaniu z rekursywnym algorytmem budowy drzew, jednak oferują one większą różnorodność reprezentacji odkrytej wiedzy [15].

Uwzględniając powyższe rozważania, niniejsza praca koncentruje się na interpretowalnych modelach w postaci reguł logicznych jako alternatywie dla tradycyjnych metod statystycznych oraz zaawansowanych algorytmów uczenia maszynowego typu „czarna skrzynka". Reguły logiczne stanowią kompromis między interpretowalnością a zdolnością modelowania złożonych zależności, oferując jednocześnie kompatybilność z procedurami operacyjnymi stosowanymi w środowiskach przemysłowych i medycznych.

W analizie niezawodności i przeżycia reguły logiczne stosuje się głównie do budowy modeli predykcyjnych umożliwiających przewidywanie czasów do wystąpienia zdarzeń. Jedyną znaną implementacją dedykowaną do danych cenzurowanych jest biblioteka RuleKit [12], rozwijana przez naukowców z Politechniki Śląskiej. Zawiera ona podstawowe algorytmy indukcji reguł przeżyciowych. Natomiast w zadaniach klasyfikacyjnych reguły logiczne znajdują szersze zastosowanie, obejmujące odkrywanie podgrup (ang. *subgroup discovery*) [16, 17], reguły akcji (ang. *action*

rules) [18, 19], reguły wyjątków (ang. *exception rules*) [20], zbiory kontrastowe (ang. *contrast-sets*) [21] oraz reguły asocjacyjne (ang. *association rules*) [22, 23].

Reguły akcji formalizują rekomendacje dotyczące zmian wartości atrybutów w celu osiągnięcia pożądanego rezultatu. Transformacje te reprezentowane są przez reguły postaci „jeśli $(X, a \rightarrow b)$ to $(Y, c \rightarrow d)$", gdzie modyfikacja atrybutu X z wartości a do b powoduje zmianę atrybutu Y z wartości c do d. Reguły wyjątków identyfikują nietypowe wzorce w danych poprzez kontrast między regułą bazową (opisującą typowe zależności w danych) a regułą wyjątku (opisującą odstępstwa od tej normy). Odkrywanie podgrup koncentruje się na identyfikacji specyficznych grup obserwacji wyróżniających się nietypowymi wartościami względem zmiennej celu. Zbiory kontrastowe opisują różnice między grupami, identyfikując atrybuty o znacząco różnych rozkładach. Reguły asocjacyjne odkrywają współwystępowanie zdarzeń lub wartości atrybutów w danych transakcyjnych [24].

W predykcyjnym utrzymaniu ruchu, a także w innych obszarach, takich jak medycyna, szczególną rolę pełnią reguły akcji oraz reguły wyjątków ze względu na ich użyteczność w podejmowaniu decyzji operacyjnych. Reguły akcji umożliwiają generowanie konkretnych rekomendacji dotyczących interwencji prewencyjnych, np. „jeśli temperatura łożyska przekroczy 80 °C, a wibracje pozostaną poniżej 5 mm/s, to zwiększenie częstotliwości smarowania z miesięcznej na tygodniową może przedłużyć czas do awarii z 30 do 90 dni". W medycynie analogiczne reguły mogą sugerować modyfikacje terapii, np. „jeśli pacjent zmieni dawkę leku z 10mg na 15mg dziennie, to prawdopodobieństwo remisji wzrośnie z 60% do 85%".

Reguły wyjątków identyfikują sytuacje wymagające szczególnej uwagi, dotyczące np. urządzeń wykazujących nietypowe wzorce degradacji lub pacjentów o niestandardowej odpowiedzi na terapię. W środowiskach przemysłowych umożliwia to wykrywanie przypadków wymagających indywidualnego podejścia do konserwacji, a w medycynie wspiera identyfikację pacjentów wymagających spersonalizowanych protokołów leczenia [23].

W analizie przeżycia dodatkowym kierunkiem rozwoju metod regułowych są zespoły reguł (ang. *rule ensembles*), które łączą zalety uczenia zespołowego (ang. *ensemble learning*) z transparentnością modeli opartych na regułach. Zespoły reguł mogą przewyższać pojedyncze reguły pod względem dokładności predykcyjnej, zachowując jednocześnie możliwość analizy mechanizmów podejmowania decyzji.

W przeciwieństwie do klasycznych metod zespołowych, zespoły reguł umożliwiają wgląd w proces agregacji predykcji na poziomie poszczególnych reguł składowych.

Wyzwaniem w rozwoju interpretowalnych metod analizy niezawodności i przeżycia jest ich zastosowanie dla rzeczywistych danych, które cechuje wysoka heterogeniczność, wielowymiarowość oraz złożoność procesów degradacyjnych. Walidacja skuteczności takich metod w warunkach rzeczywistych zastosowań wymaga uwzględnienia nie tylko dokładności predykcji, ale również możliwości zrozumienia mechanizmów prowadzących do przewidywanych zdarzeń przez ekspertów odpowiedzialnych za podejmowanie decyzji operacyjnych.

Współczesny przemysł charakteryzuje się wysoką złożonością maszyn oraz rosnącym znaczeniem technologii cyfrowych. Wymaga to zastosowania zaawansowanych metod analitycznych do optymalizacji procesów utrzymania ruchu [3]. Era Przemysłu 4.0 integruje fizyczne procesy produkcyjne z cyfrowymi systemami sterowania typu SCADA (ang. *Supervisory Control and Data Acquisition*) oraz Internetem Rzeczy (ang. *Internet of Things*, IoT) otwierając nowe możliwości dla zastosowania interpretowalnych metod analizy niezawodności w praktyce przemysłowej [25]. Rosnąca ilość danych generowanych przez czujniki IoT oraz systemy monitorowania stwarza potrzebę rozwoju metod analitycznych łączących wysoką skuteczność predykcyjną z transparentnością procesów decyzyjnych, niezbędną w środowiskach przemysłowych [10].

Problem interpretowalności w systemach analizy niezawodności nabiera szczególnego znaczenia dla odpowiedzialności prawnej, regulacji branżowych oraz konieczności budowania zaufania personelu operacyjnego do systemów automatycznych. Decyzje podejmowane na podstawie analiz niezawodności często dotyczą kosztownych operacji remontowych, planowania przestojów produkcyjnych czy alokacji zasobów. W takich sytuacjach zrozumienie podstaw decyzji modelu umożliwia efektywne zarządzanie ryzykiem operacyjnym.

W kontekście transparentności modeli uczenia maszynowego (szczegółowo omówionej w Rozdziale 4), w niniejszej pracy skupiono się na modelach interpretowalnych, opracowując interpretowalne metody analizy niezawodności i przeżycia oparte na regułach logicznych.

1.1 Motywacja

Przegląd literatury naukowej ujawnia lukę badawczą w zakresie interpretowalnych metod uczenia maszynowego dedykowanych do analizy niezawodności i przeżycia. Większość istniejących badań koncentruje się na trzech głównych nurtach: tradycyjnych metodach statystycznych analizy przeżycia, podstawowych metodach interpretowalnych (takich jak proste drzewa i reguły przeżyciowe) oraz zaawansowanych algorytmach uczenia maszynowego o wysokiej złożoności. Istotny obszar zaawansowanych interpretowalnych metod przystosowanych do specyfiki danych cenzurowanych pozostaje jednak pomijany. W obszarach takich jak medycyna i inżynieria niezawodności częsta obecność niepełnych obserwacji wymaga zastosowania dedykowanych metod analitycznych. Dostępne interpretowalne metody nie wykorzystują w pełni potencjału zaawansowanych technik regułowych, takich jak reguły akcji czy reguły wyjątków, co ogranicza ich zastosowanie w środowiskach wymagających wysokiej interpretowalności.

W nurcie statystycznym dominują klasyczne podejścia oparte na estymatorze Kaplana-Meiera, modelu Coxa oraz parametrycznych rozkładach przeżycia, które zapewniają wysoką interpretowalność, lecz mają ograniczenia w modelowaniu złożonych zależności [1]. Proste metody interpretowalne, takie jak drzewa przeżyciowe i reguły przeżyciowe, lepiej dostosowują się do różnorodnych wzorców danych, ale wciąż charakteryzują się ograniczeniami w zakresie reguł akcji, reguł wyjątków czy interpretowalnych zespołów. Natomiast w nurcie zaawansowanych metod uczenia maszynowego przeważają podejścia oparte na głębokich sieci neuronowych, zespołach klasyfikatorów oraz metodach kernelowych, które oferują wysoką skuteczność predykcyjną kosztem transparentności procesów decyzyjnych [13].

Szczególnie zauważalna jest nieobecność systematycznych badań nad zaawansowanymi algorytmami indukcji reguł dedykowanych dla danych cenzurowanych. Choć reguły logiczne znajdują zastosowanie w klasycznych problemach klasyfikacji i regresji [22], ich adaptacja do analizy przeżycia pozostaje w dużej mierze nieopracowana. Istnieją co prawda implementacje podstawowych reguł przeżyciowych, takie jak biblioteka RuleKit [12], ale brakuje kompleksowego podejścia do zaawansowanych mechanizmów reguł akcji, algorytmów indukcji reguł wyjątków oraz interpretowalnych zespołów regułowych w analizie przeżycia [23, 24].

Motywacją do podjęcia badań opisanych w niniejszej rozprawie jest potrzeba rozszerzenia istniejących interpretowalnych metod analizy niezawodności i przeżycia o zaawansowane mechanizmy reguł akcji, algorytmy obsługi wyjątków oraz interpretowalne zespoły regułowe. Choć podstawowe metody wykorzystujące reguły przeżyciowe są dostępne, istnieje potrzeba opracowania zaawansowanych rozwiązań do generowania rekomendacji opartych na regułach akcji, identyfikacji wyjątków w danych cenzurowanych oraz łączenia interpretowalności z niezawodnością zespołów modeli. Rozwój dziedzin stosujących analizę przeżycia oraz rosnące wymagania dotyczące transparentności systemów automatycznych generują potrzebę nowej klasy narzędzi analitycznych o dobrej skuteczności i interpretowalności. Opracowanie algorytmów indukcji reguł dedykowanych do danych cenzurowanych oraz integracja metod analizy przeżycia z technikami uczenia maszynowego może przyczynić się do rozwoju interpretowalnych metod analizy niezawodności. Takie podejście oferuje narzędzia skuteczne i zrozumiałe dla użytkowników końcowych w medycynie, inżynierii niezawodności oraz w predykcyjnym utrzymaniu ruchu.

1.2 Cel i teza pracy

Głównym celem niniejszej pracy jest opracowanie interpretowalnych metod analizy niezawodności i przeżycia oraz przedstawienie ich potencjalnej użyteczności w zastosowaniach medycznych oraz w predykcyjnym utrzymaniu ruchu. Praca koncentruje się na rozwoju nowych algorytmów indukcji reguł, uwzględniających specyfikę danych cenzurowanych i łączących tradycyjne techniki analizy przeżycia z regułami logicznymi jako interpretowalnymi modelami uczenia maszynowego. Opracowane metody mają zapewnić zarówno dobrą skuteczność predykcyjną, jak i interpretowalność, umożliwiającą ekspertom dziedzinowym zrozumienie i weryfikację mechanizmów decyzyjnych w obszarach wykorzystujących analizę przeżycia.

W ramach realizacji celu głównego sformułowano następujące cele szczegółowe:

1. Rozszerzenie podstaw teoretycznych integracji analizy przeżycia z indukcją reguł poprzez ujednolicenie definicji i interpretacji reguł dla danych cenzurowanych oraz zdefiniowanie kryteriów oceny, selekcji i agregacji predykcji w modelach i zespołach.

2. Rozwój pokryciowego algorytmu indukcji przeżyciowych reguł akcji oraz algorytmu rekomendacji opartych na przeżyciowych regułach akcji.

3. Opracowanie algorytmu indukcji przeżyciowych reguł wyjątków umożliwiającego identyfikację nietypowych wzorców w danych cenzurowanych.

4. Stworzenie interpretowalnego zespołu reguł przeżyciowych, łączącego zalety uczenia zespołowego z transparentnością modeli regułowych.

5. Walidacja opracowanych metod na zróżnicowanych zbiorach danych przemysłowych i medycznych, odzwierciedlających różnorodność i złożoność problemów analizy przeżycia.

6. Demonstracja potencjalnej użyteczności zaproponowanych metod poprzez analizę przypadków zastosowań w medycynie oraz w środowiskach przemysłowych.

Na podstawie sformułowanych celów główna teza pracy brzmi następująco:

Zastosowanie interpretowalnych metod uczenia maszynowego, w szczególności algorytmów indukcji reguł dostosowanych do specyfiki danych cenzurowanych, umożliwia opracowanie skutecznych i transparentnych narzędzi analizy niezawodności i przeżycia o szerokim zakresie zastosowań, czego potwierdzeniem jest ich potencjalna użyteczność w medycynie oraz w predykcyjnym utrzymaniu ruchu.

1.3 Elementy oryginalne i wkład naukowy

Niniejsza praca wnosi wkład w rozwój dziedziny poprzez opracowanie nowych, interpretowalnych metod analizy niezawodności i przeżycia, które wypełniają zidentyfikowaną lukę badawczą w obszarze łączącym analizę przeżycia z interpretowalnym uczeniem maszynowym. Główne elementy oryginalności i wkładu naukowego dotyczą przede wszystkim aspektów metodologicznych:

1. **Algorytmy przeżyciowych reguł akcji**: Opracowano pierwszy w literaturze pokryciowy algorytm indukcji reguł akcji, umożliwiający zastosowanie strategii sekwencyjnego pokrywania do danych cenzurowanych, oraz komplementarny algorytm rekomendacji, który wykorzystuje wygenerowane reguły

akcji do tworzenia interpretowalnych rekomendacji poprzez rozwiązywanie konfliktów między pokrywającymi się regułami.

2. **Algorytm indukcji przeżyciowych reguł wyjątków**: Opracowano oryginalny algorytm stanowiący rozszerzenie klasyfikacyjnych reguł wyjątków na potrzeby analizy przeżycia, który indukuje reguły złożone z trzech elementów (reguły bazowej, reguły referencyjnej oraz reguły wyjątku) i wykorzystuje sekwencyjne wyszukiwanie wyjątków w procesie indukcji reguł.

3. **Interpretowalny zespół reguł przeżyciowych**: Zaproponowano pierwszy algorytm łączący interpretowalność modeli regułowych z niezawodnością technik uczenia zespołowego, który rozszerza paradygmat lasów losowych poprzez zastąpienie drzew decyzyjnych modelami reguł przeżyciowych z wykorzystaniem próbkowania bootstrap i agregacji funkcji przeżycia.

Ponadto praca zawiera zarówno elementy teoretyczne — w tym rozszerzenie istniejących algorytmów indukcji reguł o nowe mechanizmy dedykowane do analizy przeżycia — jak i praktyczne, obejmujące demonstrację potencjalnej użyteczności opracowanych metod w rzeczywistych zastosowaniach medycznych oraz w predykcyjnym utrzymaniu ruchu.

Zgodnie z najlepszą wiedzą autora, przedstawione w pracy algorytmy stanowią pierwsze tak kompleksowe badania w zakresie pokryciowych algorytmów indukcji reguł akcji, reguł wyjątków oraz zespołów reguł stosowanych do analizy przeżycia. Opracowane metody otwierają nowy kierunek badań w obszarze łączącym interpretowalne uczenie maszynowe z analizą niezawodności i przeżycia, dostarczając podstaw do budowy narzędzi dla ekspertów dziedzinowych i wyznaczając punkt wyjścia dla dalszych prac badawczych w tej dziedzinie.

1.4 Układ pracy

Prezentowana rozprawa doktorska składa się z siedmiu rozdziałów, w których przedstawiono podstawy teoretyczne, metodologię, implementację oraz potencjalne zastosowania interpretowalnych metod analizy niezawodności i przeżycia w medycynie oraz w predykcyjnym utrzymaniu ruchu.

Rozdział 2 wprowadza kontekst zastosowań analizy niezawodności i przeżycia w środowisku przemysłowym, ze szczególnym uwzględnieniem predykcyjnego

utrzymaniu ruchu jako strategii eksploatacyjnej wykorzystującej zaawansowane metody analizy danych typu czas-do-zdarzenia. Omówiono w nim strategie utrzymania ruchu, stosowane technologie i metody, specyfikę danych przemysłowych o naturalnie cenzurowanym charakterze oraz przykłady zastosowań metod analizy przeżycia w obszarze przemysłowym i medycznym.

Rozdział 3 stanowi podstawę teoretyczną rozprawy. Przedstawiono w nim podstawowe pojęcia analizy niezawodności jako dziedziny inżynierii oraz analizy przeżycia jako zestawu metod statystycznych. Omówiono podobieństwa i różnice między tymi dziedzinami, funkcje statystyczne, rozkłady prawdopodobieństwa oraz metody estymacji stosowane w analizie niezawodności i przeżycia. Rozdział zawiera także szczegółową prezentację klasycznych metod analizy przeżycia, w tym estymatora Kaplana-Meiera, modelu proporcjonalnych hazardów Coxa oraz parametrycznych modeli przeżycia. Przedstawiono w nim ponadto zastosowania metod analizy przeżycia w inżynierii niezawodności, w tym przykład wykorzystania estymatora Kaplana-Meiera w analizie trwałości pomp przemysłowych. Stanowi to podstawę dla dalszych analiz i zastosowań omawianych w kolejnych częściach pracy.

Rozdział 4 koncentruje się na problematyce objaśnialności i interpretowalności w uczeniu maszynowym, ze szczególnym uwzględnieniem ich znaczenia w predykcyjnym utrzymaniu ruchu. Przedstawiono w nim przegląd metod objaśnialnego i interpretowalnego uczenia maszynowego, omawiając podstawowe pojęcia oraz ich praktyczne implikacje. Rozdział szczegółowo analizuje różnice między modelami interpretowalnymi a modelami typu „czarnej skrzynki", prezentuje współczesne techniki wyjaśniania decyzji modeli uczenia maszynowego i podkreśla znaczenie interpretowalności w praktyce przemysłowej, zwłaszcza dla odpowiedzialności prawnej i budowania zaufania do systemów automatycznych.

Rozdział 5 rozwija zagadnienia przedstawione we wcześniejszych rozdziałach, koncentrując się na regułach logicznych jako interpretowalnej metodzie analizy danych cenzurowanych i predykcji awarii. Omówiono w nim klasyczne metody indukcji reguł oraz zespołów reguł, przedstawiając szczegółowy przegląd literatury dotyczący zastosowań w analizie niezawodności i przeżycia. Rozdział zawiera także teoretyczne podstawy integracji metod analizy przeżycia z algorytmami indukcji reguł oraz szczegółowy opis czterech autorskich algorytmów opracowanych w ramach pracy doktorskiej, każdy przedstawiony z formalnym opisem matematycznym oraz pseudokodem.

Rozdział 6 prezentuje wyniki eksperymentów oraz analizę przypadków użycia, które stanowią empiryczną walidację teorii i metod omówionych we wcześniejszych częściach pracy. Bazując na teoretycznych podstawach analizy niezawodności i przeżycia, roli objaśnialności w modelach uczenia maszynowego oraz interpretowalnych metodach opisanych wcześniej, eksperymenty koncentrują się na ocenie skuteczności zaproponowanych podejść w medycynie i predykcyjnym utrzymaniu ruchu. Przeprowadzono testy algorytmów na zróżnicowanych zbiorach danych odzwierciedlających typowe wyzwania w analizie przeżycia i niezawodności, a wyniki eksperymentów dla każdego z czterech autorskich algorytmów zestawiono z metodami referencyjnymi.

Rozdział 7 stanowi podsumowanie pracy doktorskiej, przedstawiając syntezę najważniejszych wyników, wniosków oraz wkładu do dziedziny. Omówiono w nim główne osiągnięcia w zakresie rozwoju interpretowalnych metod analizy niezawodności i przeżycia oraz zakres potencjalnych zastosowań w medycynie i predykcyjnym utrzymaniu ruchu oraz kierunki dalszych badań. Podkreślono znaczenie integracji metod statystycznych z technikami uczenia maszynowego w kontekście interpretowalności, wskazano otwarte problemy badawcze wymagające dalszych prac, a także przedstawiono refleksje nad ograniczeniami opracowanych metod i propozycje ich przyszłego rozwoju.

2 Predykcyjne utrzymanie ruchu

Interpretowalne metody uczenia maszynowego znajdują zastosowanie w predykcyjnym utrzymaniu ruchu, które stanowi ważny obszar analizy niezawodności i przeżycia. Metody modelowania danych typu czas-do-zdarzenia dostarczają narzędzi statystycznych niezbędnych do opisu procesów degradacji urządzeń przemysłowych, przy jednoczesnym uwzględnieniu charakterystycznej dla środowisk przemysłowych obecności danych cenzurowanych [3]. We współczesnym przemyśle, który charakteryzuje się rosnącą złożonością maszyn i dynamicznym rozwojem technologii cyfrowych, predykcyjne utrzymanie ruchu stanowi strategię eksploatacyjną, umożliwiającą optymalizację harmonogramów konserwacji poprzez przewidywanie momentów wystąpienia awarii.

Era Przemysłu 4.0, wraz z powszechnym wdrażaniem rozwiązań Internetu Rzeczy i systemów SCADA, umożliwiła gromadzenie dużych ilości danych operacyjnych w czasie rzeczywistym, co stworzyło podstawę do implementacji zaawansowanych strategii predykcyjnych opartych na metodach analizy niezawodności [25, 10]. Przejście ku Przemysłowi 5.0 dodatkowo podkreśla znaczenie predykcyjnego utrzymania ruchu jako ważnego elementu zrównoważonego, zorientowanego na człowieka podejścia do zarządzania produkcją [26]. Integracja fizycznych procesów produkcyjnych z cyfrowymi systemami zarządzania umożliwia ciągłe monitorowanie parametrów technicznych maszyn i urządzeń, generując dane o naturalnie cenzurowanym charakterze, typowym dla analizy przeżycia. Niniejszy rozdział przedstawia strategie utrzymania ruchu, technologie i metody stosowane w predykcyjnym utrzymaniu ruchu, charakterystykę danych przemysłowych oraz zastosowania metod analizy przeżycia w środowisku przemysłowym.

2.1 Strategie utrzymania ruchu

Współczesne podejścia do zarządzania utrzymaniem ruchu można systematyzować zgodnie z normą PN-EN 13306:2018 [27], wyróżniając cztery podstawowe strategie eksploatacyjne:

- strategia do wystąpienia uszkodzenia,
- strategia prewencyjnego utrzymania ruchu,
- strategia utrzymania na podstawie stanu technicznego,
- strategia predykcyjnego utrzymania ruchu.

Każda z tych strategii charakteryzuje się specyficznym podejściem do planowania działań konserwacyjnych, różnym poziomem zaawansowania technologicznego oraz odmiennymi skutkami ekonomicznymi i operacyjnymi. Wybór odpowiedniej strategii zależy od charakterystyki eksploatowanych urządzeń, wymagań bezpieczeństwa, dostępności zasobów oraz celów biznesowych organizacji [28].

Strategia do wystąpienia uszkodzenia (ang. *Run to Failure, RTF*), nazywana również konserwacją reaktywną (ang. *Reactive Maintenance*), polega na eksploatacji urządzeń do momentu wystąpienia awarii, bez podejmowania działań prewencyjnych. Podejście to charakteryzuje się niskimi kosztami bieżącej eksploatacji, ale wiąże się z wysokim ryzykiem nieplanowanych przestojów oraz potencjalnie wysokimi kosztami napraw w przypadku awarii. RTF znajduje uzasadnienie ekonomiczne w przypadku urządzeń o niskiej krytyczności dla procesu produkcyjnego, łatwej wymienialności oraz niskich kosztach zakupu, gdy koszt monitorowania i konserwacji prewencyjnej przekraczałby korzyści z uniknięcia awarii.

Strategia prewencyjnego utrzymania ruchu (ang. *Preventive Maintenance*, PM) polega na planowaniu działań konserwacyjnych w regularnych odstępach czasowych lub po osiągnięciu określonych progów eksploatacyjnych, niezależnie od rzeczywistego stanu technicznego urządzenia. Harmonogramy konserwacji są ustalane na podstawie zaleceń producentów, danych historycznych lub norm branżowych [29]. PM umożliwia planowanie działań konserwacyjnych i alokację zasobów, jednak może prowadzić do nadmiernego serwisowania urządzeń znajdujących się w dobrym stanie technicznym lub do niewystarczającej konserwacji, gdy procesy degradacji przebiegają szybciej niż przewidziano w harmonogramie.

Strategia utrzymania na podstawie stanu technicznego (ang. *Condition Based Maintenance*, CBM) wykorzystuje ciągłe lub okresowe monitorowanie parametrów technicznych urządzeń w celu podejmowania decyzji dotyczących prac konserwacyjnych. Działania prewencyjne są inicjowane w momencie przekroczenia określonych progów alarmowych, które wskazują na pogorszenie stanu technicznego. CBM wymaga wdrożenia systemów monitorowania oraz ustalenia kryteriów decyzyjnych, ale pozwala na optymalizację częstotliwości konserwacji, dostosowując je do rzeczywistego stanu urządzeń.

Strategia predykcyjnego utrzymania ruchu stanowi najbardziej zaawansowane podejście, które wykorzystuje modelowanie predykcyjne do przewidywania przyszłego stanu technicznego urządzeń. Predykcyjne utrzymanie ruchu integruje dane z systemów monitorowania z zaawansowanymi metodami analitycznymi, takimi jak uczenie maszynowe czy analiza przeżycia, co umożliwia prognozowanie czasów do awarii oraz optymalizację harmonogramów konserwacji [30, 31]. Strategia ta pozwala na minimalizację zarówno kosztów utrzymania, jak i ryzyka nieplanowanych przestojów, dzięki planowaniu działań prewencyjnych w momencie optymalnym względem przewidywanej awarii.

2.2 Metody i technologie predykcyjnego utrzymania ruchu

Predykcyjne utrzymanie ruchu wykorzystuje szerokie spektrum metod analitycznych i technologii monitorowania, dopasowanych do rodzaju urządzeń oraz charakteru procesów degradacyjnych. Współczesne systemy predykcyjnego utrzymania ruchu łączą tradycyjne techniki inżynierskie z zaawansowanymi metodami analizy danych, tworząc kompleksowe rozwiązania umożliwiające skuteczne przewidywanie awarii oraz optymalizację strategii konserwacyjnych.

Tradycyjne metody analizy technicznej w predykcyjnym utrzymaniu ruchu obejmują szeroki zakres technik diagnostycznych dostosowanych do rodzaju urządzeń przemysłowych. Analiza wibracji stanowi podstawową metodę diagnozowania stanu łożysk, przekładni, sprzęgieł oraz elementów wirujących, umożliwiając wykrywanie niezrównoważenia, współosiowości czy luzów mechanicznych na podstawie charakterystycznych częstotliwości sygnałów wibracyjnych [32].

Termografia wykorzystuje promieniowanie podczerwone do identyfikacji problemów związanych z przegrzewaniem komponentów elektrycznych, połączeń, układów chłodzenia oraz elementów mechanicznych, pozwalając na bezkontaktową ocenę rozkładu temperatury na powierzchni urządzeń [33, 34, 35]. Analiza oleju umożliwia ocenę stanu układów smarowania poprzez badanie właściwości fizycznych i chemicznych płynów eksploatacyjnych oraz wykrywanie cząstek zużycia pochodzących z wewnętrznych elementów maszyn [36].

Uzupełnieniem tradycyjnych metod są analiza ciśnienia w układach hydraulicznych i pneumatycznych [37, 38], monitoring natężenia prądu i napięcia w systemach elektrycznych [39] oraz badania ultradźwiękowe służące do wykrywania nieszczelności, wyładowań elektrycznych czy problemów z łożyskami na wczesnym etapie degradacji [40, 41]. Metody te charakteryzują się wysoką specjalizacją i wymagają eksperckiej wiedzy do interpretacji wyników, lecz jednocześnie zapewniają wgląd w specyficzne mechanizmy degradacji, charakterystyczne dla różnych klas urządzeń przemysłowych [3].

Rozwój technologii IoT umożliwił implementację zaawansowanych systemów czujników, które zapewniają ciągłe monitorowanie parametrów operacyjnych maszyn i urządzeń w czasie rzeczywistym. Nowoczesne czujniki bezprzewodowe mierzą m.in. wibracje, temperaturę, ciśnienie, prąd, napięcie, przepływ oraz inne parametry fizyczne, a następnie przesyłają dane do centralnych systemów analitycznych za pośrednictwem protokołów komunikacyjnych, takich jak LoRaWAN, Zigbee czy 5G [42, 43]. Integracja z systemami typu SCADA pozwala łączyć dane z poziomu urządzeń przemysłowych z nadrzędnymi systemami zarządzania produkcją, tworząc kompleksową infrastrukturę monitorowania. Platformy IoT wspierają zarówno przetwarzanie danych w chmurze obliczeniowej, jak i wykorzystanie technik przetwarzania brzegowego (ang. *edge computing*) do analizy w czasie rzeczywistym, co stanowi podstawę zaawansowanych analiz predykcyjnych obejmujących całe zakłady przemysłowe [30]. Możliwość integracji danych z wielu źródeł — od prostych czujników temperatury po złożone systemy wizyjne — zapewnia algorytmom w predykcyjnym utrzymaniu ruchu dostęp do zróżnicowanych informacji. Coraz częściej do symulacji i predykcji zachowania urządzeń w różnych scenariuszach operacyjnych stosuje się także technologie cyfrowych bliźniaków (ang. *digital twins*) [44, 45].

Metody uczenia maszynowego znajdują coraz szersze zastosowanie w predykcyjnym utrzymaniu ruchu, umożliwiając automatyczne wykrywanie wzorców degradacji oraz przewidywanie awarii na podstawie danych historycznych i bieżących pomiarów [46, 10]. Algorytmy klasyfikacji służą do kategoryzowania stanów technicznych urządzeń, natomiast metody regresji umożliwiają przewidywanie wartości parametrów technicznych w czasie. Techniki głębokiego uczenia — w szczególności sieci LSTM (ang. *long short-term memory*) oraz konwolucyjne (ang. *convolutional neural network*, CNN) — pozwalają na modelowanie złożonych zależności w wielowymiarowych szeregach czasowych pochodzących z systemów monitorowania [47]. Z kolei metody zespołowe [48] oraz zaawansowane metody inżynierii cech, takie jak transformacje falkowe (ang. *continuous wavelet transform*, CWT) [49] czy analiza głównych składowych [50], zwiększają dokładność predykcji, uwzględniając jednocześnie niepewność modeli i zmienność warunków operacyjnych [51].

Analiza anomalii w predykcyjnym utrzymaniu koncentruje się na identyfikacji odstępstw od typowych wzorców operacyjnych. Obejmuje zarówno podejścia statystyczne, jak i techniki uczenia maszynowego dostosowane do analizy szeregów czasowych, co umożliwia wczesne wykrywanie symptomów degradacji przed wystąpieniem krytycznych awarii [52].

2.3 Charakterystyka danych w predykcyjnym utrzymaniu ruchu

Dane gromadzone w systemach predykcyjnego utrzymania ruchu charakteryzują się specyficznymi cechami, które wynikają z natury procesów przemysłowych oraz ograniczeń praktycznych związanych z monitorowaniem urządzeń w warunkach operacyjnych. Zrozumienie charakterystyki tych danych pomaga w wyborze odpowiednich metod analitycznych oraz skutecznego wdrażania strategii predykcyjnego utrzymania ruchu w środowiskach przemysłowych.

Charakterystyczną cechą danych w predykcyjnym utrzymaniu ruchu jest ich cenzurowany charakter, wynikający z faktu, że w analizowanym okresie znaczna części urządzeń nie ulega awarii. Urządzenia te mogą pozostawać w eksploatacji do końca okresu obserwacji, zostać wymienione profilaktycznie w ramach planowanych modernizacji lub wycofane z użycia z przyczyn niezwiązanych z awarią techniczną.

Taka niekompletna informacja o czasach awarii stanowi wyzwanie metodologiczne, ponieważ tradycyjne metody analizy danych zakładają dostępność kompletnych obserwacji dla wszystkich analizowanych przypadków [2, 53]. Problematyka danych cenzurowanych jest dobrze znana w niezawodności przemysłowej oraz w analizie danych gwarancyjnych [54].

Ignorowanie cenzurowanego charakteru danych przemysłowych w analizach predykcyjnych prowadzi do błędów w estymacji czasów do awarii. Pominięcie obserwacji cenzurowanych skutkuje utratą informacji o trwałości urządzeń, które pracowały bezawaryjnie przez cały okres obserwacji, co prowadzi do pesymistycznego obciążenia modeli w kierunku zaniżania przewidywanych czasów eksploatacji. Z kolei traktowanie obserwacji cenzurowanych jako kompletnych, poprzez przypisanie im czasu obserwacji jako momentu awarii, wprowadza błąd w przeciwnym kierunku — prowadzi do niedoszacowania ryzyka awarii [55].

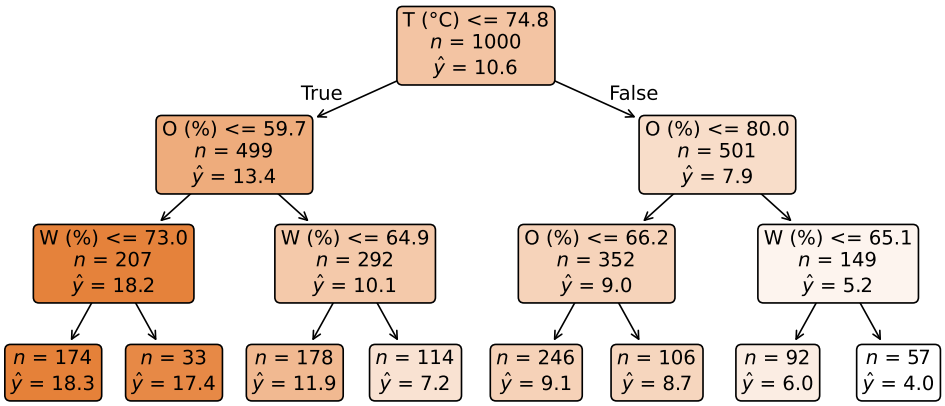
Specyfika danych cenzurowanych w predykcyjnym utrzymaniu ruchu uzasadnia stosowanie metod analizy przeżycia, opracowanych specjalnie do pracy z niekompletnymi obserwacjami czasów do zdarzeń [56, 1]. Metody te pozwalają wykorzystać pełną informację zawartą w danych, obejmując zarówno obserwowane awarie, jak i okresy bezawaryjnego funkcjonowania urządzeń. Integracja analizy przeżycia z danymi przemysłowymi umożliwia tworzenie modeli predykcyjnych, które szacują prawdopodobieństwo awarii oraz pozostały czas użytkowania urządzeń, uwzględniając niepewność wynikającą z cenzurowania danych [57].

Dane w systemach predykcyjnego utrzymaniu ruchu często charakteryzują się wysoką wielowymiarowością, obejmującą parametry operacyjne (temperatura, ciśnienie, prędkość obrotowa), warunki środowiskowe (temperatura otoczenia, wilgotność), dane eksploatacyjne (czas pracy, liczba cykli) oraz informacje kontekstowe (typ urządzenia, wiek, historia konserwacji). Taka struktura danych wymaga budowy modeli predykcyjnych, które uwzględniają wpływ wielu czynników na procesy degradacji oraz zastosowania odpowiednich metod analizy wielowymiarowej dostosowanych do specyfiki danych cenzurowanych [58].

2.4 Przykład zastosowania drzewa przeżyciowego w analizie danych przemysłowych

W celu ilustracji zastosowania metod analizy przeżycia w predykcyjnym utrzymaniu ruchu przedstawiono przykład analizy danych pochodzących z monitorowania 15 przemysłowych pomp wirowych. Analiza wykorzystuje drzewo przeżyciowe jako interpretowalną metodę uczenia maszynowego, która uwzględnia wpływ warunków operacyjnych na niezawodność urządzeń [11, 59]. Zbiór danych obejmuje zarówno obserwacje awarii, jak i dane cenzurowane. Drzewa przeżyciowe stanowią rozszerzenie klasycznych drzew decyzyjnych do analizy danych typu czas-do-zdarzenia [60].

Zbiór danych zawiera informacje o czasach obserwacji (w miesiącach), statusie urządzenia (awaria lub brak informacji o awarii, tj. cenzurowanie) oraz trzech zmiennych objaśniających: temperaturze pracy (°C), wilgotności względnej środowiska (%) oraz obciążeniu roboczym (% mocy nominalnej). Dane te odzwierciedlają typowe parametry monitorowane w systemach predykcyjnego utrzymania ruchu oraz ilustrują wpływ warunków operacyjnych na niezawodność urządzeń przemysłowych [61]. Podobne analizy niezawodności, uwzględniające wpływ warunków środowiskowych, stosuje się również w przypadku turbin gazowych [62].



Rysunek 2.1: Drzewo przeżyciowe dla zbioru danych pomp wirowych. Oznaczenia: T — temperatura pracy (°C), O — obciążenie robocze (%), W — wilgotność względna (%), n — liczba obserwacji w węźle, $\hat{y}$ — przewidywana mediana czasu przeżycia (miesiące).

Drzewo przeżyciowe wygenerowane dla tego zbioru danych, przedstawione na Rysunku 2.1, charakteryzuje się hierarchiczną strukturą podziałów opartych na progowych wartościach zmiennych objaśniających. Główny podział w korzeniu drzewa następuje na podstawie temperatury pracy, gdzie próg 74.8 °C oddziela pompy pracujące w wysokiej temperaturze (>74.8 °C) od urządzeń eksploatowanych w standardowych warunkach termicznych (⩽74.8 °C). Podział ten odzwierciedla krytyczny wpływ temperatury na trwałość komponentów mechanicznych oraz efektywność smarowania.

W węzłach potomnych drzewo wprowadza dalsze podziały na podstawie obciążenia roboczego oraz wilgotności względnej. Dla pomp pracujących w niższej temperaturze (⩽74.8 °C) próg obciążenia wynosi 59.7% mocy nominalnej, z kolejnymi podziałami przy wilgotności 73.0% i 64.9%. Dla pomp w wyższej temperaturze (>74.8 °C) główny podział następuje przy obciążeniu 80.0%, z dodatkowymi podziałami przy obciążeniu 66.2% oraz wilgotności 65.1%, które stanowią czynniki ryzyka wpływające na procesy degradacyjne.

Liście drzewa reprezentują grupy pomp o podobnych profilach niezawodności, z oszacowanymi funkcjami przeżycia specyficznymi dla każdej kombinacji warunków operacyjnych. Grupa o najwyższym ryzyku, reprezentowana przez skrajną prawą ścieżkę w drzewie, charakteryzuje się medianą czasu do awarii wynoszącą 4.0 miesiące. Najbardziej niezawodną grupę stanowią pompy eksploatowane w najkorzystniejszych warunkach operacyjnych (skrajna lewa ścieżka), dla których mediana czasu do awarii wynosi 18.3 miesiąca. Pozostałe liście drzewa odpowiadają grupom o pośrednich poziomach niezawodności, z medianami w zakresie od 6.0 do 17.4 miesięcy.

Drzewo przeżyciowe jest modelem interpretowalnym, którego struktura podziałów umożliwia priorytetyzację działań konserwacyjnych na podstawie rzeczywistych warunków operacyjnych urządzeń. Struktura drzewa pozwala identyfikować krytyczne kombinacje czynników wpływających na niezawodność, co umożliwia optymalizację harmonogramów konserwacji oraz modyfikację warunków eksploatacji w celu przedłużenia żywotności urządzeń. Metoda ta łączy zalety interpretowalności charakterystycznej dla tradycyjnych podejść inżynierskich z możliwościami modelowania złożonych zależności oferowanymi przez uczenie maszynowe. Dalsze rozszerzenia tej metodologii obejmują zespoły drzew przeżyciowych oraz zaawansowane techniki typu boosting dostosowane do danych cenzurowanych [63].

3 Analiza niezawodności i przeżycia

Analiza niezawodności i przeżycia — wprowadzona w Rozdziale 1 i ilustrowana przykładami PDM w Rozdziale 2 — dostarcza podstaw teoretycznych do modelowania danych typu czas-do-zdarzenia w kontekstach technicznych i medycznych [5, 64]. To podejście analityczne rozwijało się początkowo w medycynie, gdzie służyło do analizy długoterminowych wyników leczenia pacjentów. Następnie znalazło również zastosowanie w inżynierii niezawodności, gdzie umożliwia precyzyjne modelowanie procesów degradacji urządzeń przemysłowych. Wspólne fundamenty teoretyczne obu zastosowań sprzyjają transferowi wiedzy i metod pomiędzy domeną medyczną a techniczną, gdyż podstawowy problem badawczy — modelowanie czasu do wystąpienia krytycznego zdarzenia w obecności niepełnych obserwacji — pozostaje taki sam niezależnie od obszaru zastosowań. Uniwersalność tego podejścia sprawia, że metody analizy przeżycia znajdują zastosowanie w zarządzaniu ryzykiem operacyjnym, planowaniu strategii konserwacyjnych oraz optymalizacji cykli życia systemów technicznych.

Ze względu na cenzurowany charakter danych (opisany w Rozdziale 1), tradycyjne metody klasyfikacyjne i regresyjne nie są odpowiednie do analizy danych przeżyciowych, co uzasadnia potrzebę zastosowania dedykowanych metod. Cenzurowanie stanowi wyzwanie metodologiczne, ponieważ dla części obserwacji informacja o czasie wystąpienia zdarzenia jest niepełna — wiemy jedynie, że zdarzenie nie wystąpiło do momentu zakończenia obserwacji lub że nastąpiło w określonym przedziale czasowym. Ignorowanie obserwacji cenzurowanych prowadzi do obciążenia estymatorów w kierunku zaniżania czasu pozostałego do wystąpienia zdarzeń, co w praktyce przemysłowej może skutkować nieadekwatnym planowaniem konserwacji i zwiększonym ryzykiem nieplanowanych przestojów. Z kolei prostsze podejścia alternatywne, takie jak traktowanie obserwacji cenzurowanych jako kompletnych na podstawie ostatniego znanego czasu obserwacji, prowadzą do przeciwnego rodzaju obciążenia — przeszacowania rzeczywistych czasów wystąpienia

zdarzeń. Specjalistyczne metody analizy przeżycia rozwiązują te problemy poprzez modelowanie procesu cenzurowania i wykorzystanie pełnej dostępnej informacji, w tym także informacji zawartej w obserwacjach niekompletnych.

Różnicę między tradycyjnymi danymi regresyjnymi a danymi cenzurowanymi można zilustrować prostym przykładem. W badaniu trwałości 6 urządzeń, w pierwszym scenariuszu obserwowane są wszystkie awarie — urządzenia uległy awarii odpowiednio po 12, 18, 24, 30, 36 oraz 42 miesiącach eksploatacji. Dla takich kompletnych danych można zastosować standardową regresję liniową, uzyskując średni czas do awarii wynoszący 27 miesięcy. W drugim scenariuszu, częściej spotykanym w praktyce, u części obiektów występuje cenzurowanie prawostronne (obserwację zakończono przed wystąpieniem zdarzenia) — badanie kończy się po 30 miesiącach, w wyniku czego obserwowane są awarie tylko trzech pierwszych urządzeń (12, 18, 24 miesiące), podczas gdy pozostałe trzy nadal funkcjonują (dane cenzurowane). Zastosowanie standardowej regresji tylko do obserwowanych awarii dałoby średni czas 18 miesięcy, co stanowi znaczące niedoszacowanie w porównaniu z rzeczywistą wartością 27 miesięcy. Pominięcie obserwacji cenzurowanych skutkuje więc systematycznym błędem w kierunku zaniżania czasów do zdarzeń.

Zaprezentowany przykład ilustruje potrzebę stosowania specjalistycznych metod analizy przeżycia, które wykorzystują informację zawartą w obserwacjach cenzurowanych. W rzeczywistych zastosowaniach odsetek takich obserwacji może sięgać kilkudziesięciu procent, co sprawia, że nie można ich pominąć — w szczególności w przypadku kosztownych badań medycznych czy długoterminowych analiz przemysłowych. Niniejszy rozdział przedstawia teoretyczne fundamenty analizy niezawodności i przeżycia, omawiając podstawowe pojęcia, funkcje statystyczne, rozkłady prawdopodobieństwa oraz metody estymacji stosowane w modelowaniu danych cenzurowanych. Stanowią one punkt wyjścia dla opracowania interpretowalnych algorytmów uczenia maszynowego, które zostaną zaprezentowane w dalszych częściach pracy.

3.1 Podstawowe pojęcia

Analiza niezawodności i analiza przeżycia, mimo odmiennych obszarów zastosowania, opierają się na wspólnych fundamentach teoretycznych i wykorzystują analogiczne metody statystyczne do modelowania danych typu czas-do-zdarzenia.

W tej sekcji zostaną wprowadzone podstawowe pojęcia dla obu podejść, które zostaną zilustrowane przykładem badania niezawodności.

pompa	t (miesiące)	d (status)
A	3	1
B	7	1
C	9	1
D	12	0
E	12	0
F	12	0
G	12	0
H	12	0

Tabela 3.1: Dane w przykładzie pomp: czas obserwacji t (w miesiącach) i status d, gdzie $d = 1$ oznacza awarię, a $d = 0$ obserwację cenzurowaną (brak awarii do końca obserwacji $t = 12$).

W ilustracyjnym przykładzie hipotetycznego badania niezawodności ośmiu identycznych pomp przemysłowych obserwowanych przez 12 miesięcy w celu określenia ich czasów do awarii przyjęto następujące dane syntetyczne: pompa A uległa awarii po 3 miesiącach, pompa B po 7 miesiącach, pompa C po 9 miesiącach, a pompy D, E, F, G oraz H działały nadal na koniec okresu obserwacji. Dla pomp nadal działających wiadomo jedynie, że przetrwały co najmniej 12 miesięcy, natomiast dokładny czas ich ewentualnej awarii pozostaje nieznany. Zestawienie danych przedstawiono w Tabeli 3.1. Przykład ten posłuży do zilustrowania podstawowych pojęć analizy przeżycia przedstawionych poniżej.

Niezawodność (ang. *reliability*) definiuje się jako zdolność systemu, komponentu lub struktury do ciągłego wykonywania zamierzonej funkcji w określonych warunkach operacyjnych przez określony okres eksploatacji, zgodnie z normą ISO 8402 (1986) [65]. W przykładzie z pompami niezawodność opisuje prawdopodobieństwo, że pompa będzie działać bez awarii przez określony czas, np. 6 miesięcy. Miara ta jest stosowana w planowaniu strategii konserwacyjnych i ocenie ryzyka operacyjnego.

Zdarzenie w analizie przeżycia oznacza specyficzny, interesujący z punktu widzenia analizy, wynik procesu obserwacyjnego. Stanowi ono uniwersalne pojęcie stosowane w różnych kontekstach aplikacyjnych. W analizie niezawodności zda-

rzeniem jest awaria urządzenia (jak w przypadku pomp A, B, C z przykładu), natomiast w analizie przeżycia może to być np. zgon pacjenta, nawrót choroby czy inne krytyczne zdarzenie medyczne. Dzięki tej uniwersalności możliwe jest stosowanie tych samych metod analitycznych w różnych dziedzinach, przy czym pojęcie awarii jest konkretną realizacją zdarzenia w kontekście inżynierskim [66].

Cenzurowanie występuje, gdy dokładny moment wystąpienia zdarzenia nie jest znany, co prowadzi do niekompletnych danych czasowych. W przykładzie, obserwacje pomp D–H są cenzurowane prawostronnie — wiadomo, że przetrwały co najmniej 12 miesięcy, lecz dokładny czas ich awarii nie jest znany. Cenzurowanie prawostronne jest najczęstsze w praktyce i pojawia, gdy obserwacja kończy się przed wystąpieniem zdarzenia. Cenzurowanie lewostronne zachodzi, gdy zdarzenie nastąpiło przed rozpoczęciem obserwacji, ale jego dokładny moment nie jest znany. Z kolei cenzurowanie obustronne (interwałowe) ma miejsce, gdy obserwacja podlega jednocześnie ograniczeniom lewostronnym i prawostronnym — wiadomo jedynie, że zdarzenie nastąpiło w pewnym przedziale czasowym, np. między wizytami kontrolnymi, lecz jego dokładny moment pozostaje nieznany. Obecność danych cenzurowanych uniemożliwia stosowanie standardowych metod regresyjnych i wymaga specjalistycznych technik analizy przeżycia [64].

Czas-do-zdarzenia to nieujemna zmienna losowa reprezentująca okres od początku obserwacji do wystąpienia interesującego zdarzenia. W analizie niezawodności określany jest jako czas-do-awarii (ang. *time-to-failure*, TTF) — w przykładzie wynosi odpowiednio 3, 7 i 9 miesięcy dla pomp A, B, C. W analizie przeżycia używa się pojęcia czasu przeżycia (ang. *survival time*). W przypadku obserwacji cenzurowanych znamy jedynie dolną granicę tego czasu [53].

Awaria (ang. *failure*) definiowana jest jako utrata zdolności systemu do realizacji zamierzonej funkcji, zgodnie z normą IEC 50(191) (1990) [67]. W inżynierii przyczyną awarii mogą być m.in. zużycie materiału, korozja czy przeciążenie. W przykładzie z pompami awarie urządzeń A, B i C mogły wynikać ze zużycia łożysk, uszkodzenia wirnika lub innych procesów degradacyjnych, podczas gdy pompy D–H nadal spełniają swoją funkcję na koniec okresu obserwacji.

3.2 Funkcje i rozkłady w analizie niezawodności i przeżycia

Analiza niezawodności i przeżycia wykorzystuje funkcje matematyczne oraz rozkłady prawdopodobieństwa dostosowane do modelowania danych cenzurowanych typu czas-do-zdarzenia [64, 29]. W tej sekcji omówione zostaną metody analityczne, obejmujące funkcję przeżycia, funkcję hazardu oraz wybrane rozkłady prawdopodobieństwa, a także zilustrowane zostanie ich zastosowanie w analizie danych przemysłowych na przykładzie estymatora Kaplana-Meiera.

Funkcja przeżycia S, która w analizie niezawodności nazywana jest również funkcją niezawodności R, definiowana jest jako prawdopodobieństwo, że czas do zdarzenia T przekroczy określony moment t. W literaturze obie notacje używane są zamiennie, w zależności od dziedziny zastosowania. Funkcja ta wyraża się wzorem:

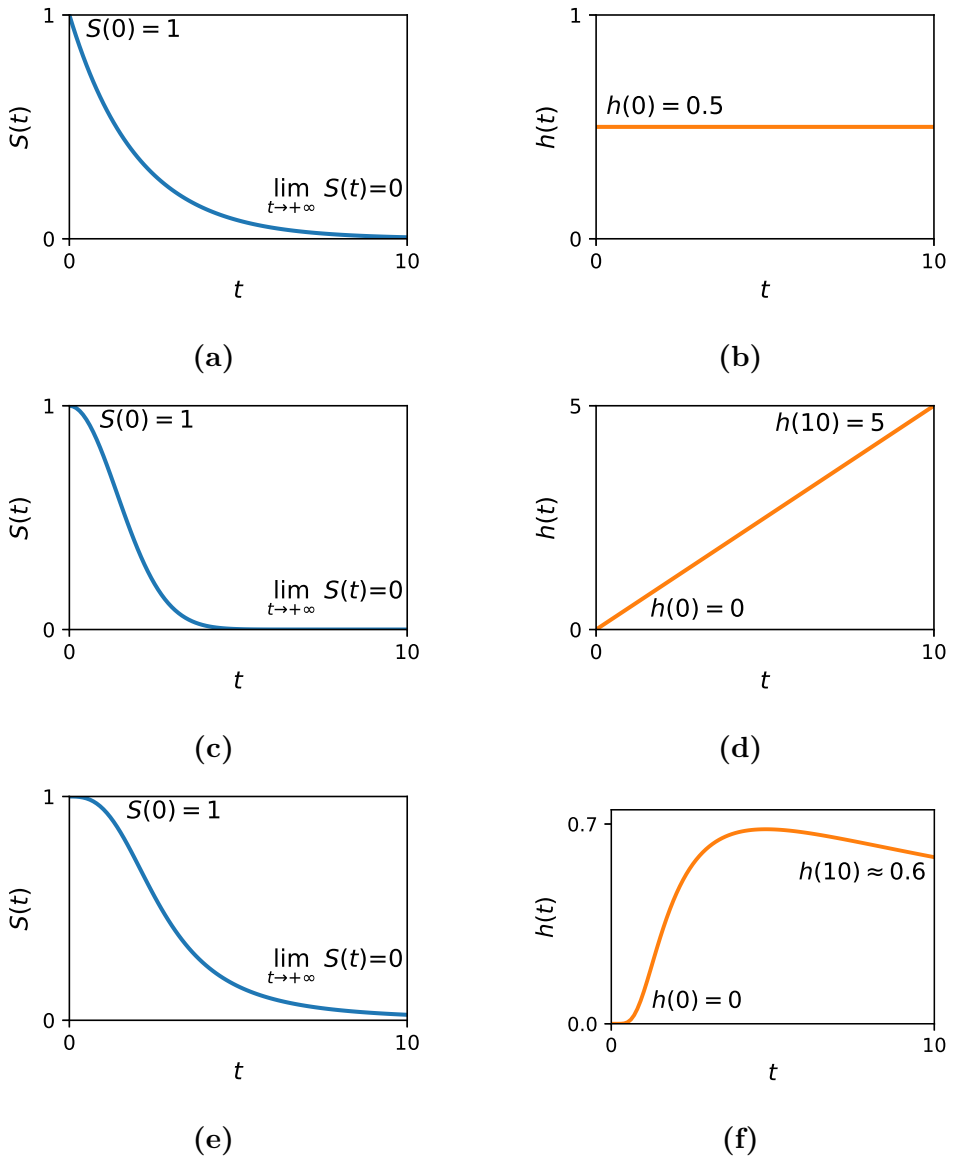
$$S(t) = P(T > t). \tag{3.1}$$

Funkcja przeżycia odzwierciedla odsetek jednostek — pacjentów w analizie medycznej lub urządzeń w analizie niezawodności — które przetrwają lub pozostaną sprawne po upływie czasu t. Jest to funkcja nierosnąca, przyjmująca wartości od 1 (wszystkie jednostki sprawne w $t = 0$) do 0 (gdy wszystkie jednostki ulegną zdarzeniu) [64, 54]. Graficzną reprezentację funkcji przeżycia nazywa się krzywą przeżycia lub, w analizie niezawodności, krzywą niezawodności. Przykładową funkcję przeżycia dla rozkładu wykładniczego przedstawiono na Rysunku 3.1a.

Funkcja hazardu h opisuje chwilowe ryzyko wystąpienia zdarzenia w czasie t, pod warunkiem że jednostka przetrwała do tego momentu. Matematycznie definiuje się ją następującym wzorem:

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leqslant T < t + \Delta t \mid T \geqslant t)}{\Delta t}. \tag{3.2}$$

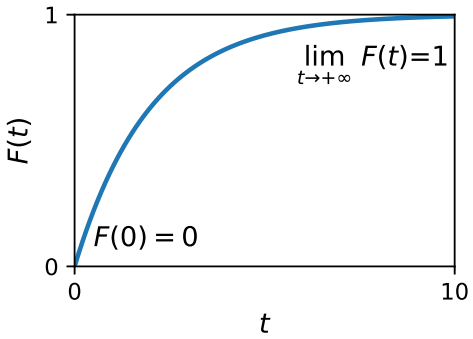
Funkcja hazardu umożliwia analizę dynamiki zdarzeń w czasie — może wskazywać zwiększone ryzyko zgonu w późniejszych stadiach choroby (analiza przeżycia) lub wzrost prawdopodobieństwa awarii wynikający ze zużycia materiału (analiza niezawodności). Jej przebieg odzwierciedla mechanizmy degradacji, co prowadzi do stosowania odpowiednich rozkładów prawdopodobieństwa [64]. Przykładową funkcję hazardu dla rozkładu wykładniczego przedstawiono na Rysunku 3.1b.



Rysunek 3.1: Wykresy funkcji przeżycia S i odpowiadających im funkcji hazardu h dla trzech rozkładów prawdopodobieństwa stosowanych w analizie niezawodności i przeżycia (ozn. $S(\infty) = \lim_{t\to+\infty} S(t)$): (a) rozkład wykładniczy ($\lambda = 0.5$) ze stałym hazardem; (b) funkcja hazardu dla rozkładu wykładniczego; (c) rozkład Weibulla ($\lambda = 2$, $k = 2$) z liniowo rosnącym hazardem; (d) funkcja hazardu dla rozkładu Weibulla; (e) rozkład log-normalny ($\mu = 1$, $\sigma = 0.5$) z rosnącym, a następnie malejącym hazardem; (f) funkcja hazardu dla rozkładu log-normalnego. Osie oznaczono jako t (czas) oraz odpowiednio $S(t)$ lub $h(t)$ (wartość funkcji przeżycia lub hazardu).

Dystrybuanta skumulowana (ang. *cumulative distribution function*, CDF) to prawdopodobieństwo wystąpienia zdarzenia nie później niż w chwili t. Funkcja ta zdefiniowana jest wzorem:

$$F(t) = 1 - S(t). \tag{3.3}$$



Rysunek 3.2: Dystrybuanta skumulowana (CDF) dla rozkładu wykładniczego.

Funkcja CDF określa skumulowane ryzyko wystąpienia zdarzenia do chwili t i stanowi uzupełnienie funkcji przeżycia. Wzajemne relacje między funkcją przeżycia S, funkcją hazardu h i dystrybuantą skumulowaną F tworzą zestaw metod umożliwiający pełną charakterystykę procesów degradacji w modelowaniu danych typu czas-do-zdarzenia [29]. Wykres przykładowej funkcji CDF dla rozkładu wykładniczego przedstawiono na Rysunku 3.2.

W analizie niezawodności i przeżycia szczególną rolę odgrywają wyspecjalizowane rozkłady prawdopodobieństwa dostosowane do modelowania czasów do zdarzeń [66]. Rozkład wykładniczy, charakteryzujący się stałą funkcją hazardu $h(t) = \lambda, \lambda > 0$, opisuje procesy bezpamięciowe (ang. *memoryless*) typowe dla awarii komponentów elektronicznych. Rozkład Weibulla, o funkcji hazardu $h(t) = \alpha\lambda t^{\alpha-1}$, umożliwia modelowanie zróżnicowanych wzorców intensywności zdarzeń — monotonicznych (rosnących, malejących, stałych) oraz niemonotonicznych (np. najpierw malejących, następnie rosnących) — w zależności od wartości parametru kształtu α. Stała $\alpha > 0$ jest parametrem kształtu kontrolującym monotoniczność hazardu, a czynnik $t^{\alpha-1}, t > 0$, jest stały dla $\alpha = 1$, rośnie dla $\alpha > 1$, a maleje dla $0 < \alpha < 1$, stąd funkcja h ma odpowiednio stały, rosnący lub malejący przebieg. Rozkład log-normalny znajduje zastosowanie w modelowaniu procesów multiplikatywnych, charakterystycznych dla degradacji biologicznej lub zmęczenia materiału. Graficzne

przedstawienie funkcji przeżycia i hazardu dla tych rozkładów zaprezentowano na Rysunku 3.1.

3.3 Metody estymacji i porównywania funkcji przeżycia

Analiza niezawodności i przeżycia wykorzystuje specjalistyczne metody estymacji funkcji przeżycia i hazardu oraz techniki porównawcze dostosowane do specyfiki danych cenzurowanych. W niniejszej sekcji omówiono podejścia statystyczne obejmujące zarówno metody parametryczne, jak i nieparametryczne do szacowania funkcji przeżycia i hazardu, a także testy porównawcze umożliwiające ocenę różnic w niezawodności między grupami lub systemami.

Metody nieparametryczne charakteryzują się brakiem założeń dotyczących konkretnego rozkładu prawdopodobieństwa czasów do zdarzeń. Estymator Kaplana-Meiera stanowi najczęściej stosowaną technikę nieparametryczną do szacowania funkcji przeżycia S na podstawie danych cenzurowanych i jest definiowany wzorem:

$$\hat{S}(t) = \prod_{t_i \leqslant t} \left(1 - \frac{d_i}{n_i}\right) \tag{3.4}$$

gdzie d_i oznacza liczbę zdarzeń w chwili t_i, a n_i to liczba jednostek w zbiorze ryzyka w chwili t_i [56]. W analizie niezawodności estymator Kaplana-Meiera oznacza się jako $\hat{R}$. Alternatywną metodą jest estymator Nelsona-Aalena, który szacuje skumulowaną funkcję hazardu:

$$\hat{H}(t) = \sum_{t_i \leqslant t} \frac{d_i}{n_i} \tag{3.5}$$

Oba estymatory uwzględniają dane cenzurowane i prowadzą do podobnych rezultatów, przy czym estymator Kaplana-Meiera jest częściej stosowany ze względu na prostą interpretację funkcji przeżycia w kategoriach prawdopodobieństwa [68].

Metody parametryczne zakładają określony rozkład czasów do zdarzeń, co umożliwia bardziej szczegółowe modelowanie procesów degradacji. Estymacja parametrów rozkładu, takich jak λ w rozkładzie wykładniczym czy α i λ w rozkładzie Weibulla, przeprowadzana jest zazwyczaj metodą estymacji największej

wiarygodności (MLE). Podejście parametryczne pozwala na uzyskanie precyzyjnych prognoz w sytuacjach, gdy dane dobrze odpowiadają wybranemu rozkładowi teoretycznemu, jednak wymaga weryfikacji zgodności założeń z rzeczywistymi obserwacjami [69].

Porównywanie funkcji przeżycia między grupami wymaga zastosowania testów statystycznych dostosowanych do danych cenzurowanych. Najczęściej stosowaną metodą oceny istotności różnic między funkcjami przeżycia dwóch lub więcej grup jest test log-rank. Test ten opiera się na porównaniu liczby zdarzeń obserwowanych i oczekiwanych w każdym punkcie czasowym, a następnie na obliczeniu statystyki χ^2 w celu sprawdzenia hipotezy zerowej, zgodnie z którą krzywe przeżycia są jednakowe we wszystkich porównywanych grupach. Test log-rank charakteryzuje się efektywnością w przypadku danych cenzurowanych oraz brakiem wymagań dotyczących konkretnego rozkładu czasów do zdarzeń [64].

Przykład zastosowania estymatora Kaplana-Meiera

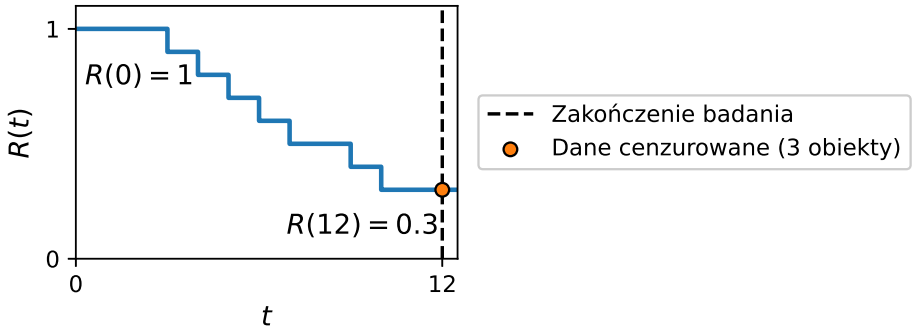
W celu ilustracji zastosowania metod analizy przeżycia w modelowaniu danych cenzurowanych przedstawiono przykład wykorzystania estymatora Kaplana-Meiera do analizy danych z badania niezawodności 10 przemysłowych pomp. Przykład pokazuje sposób szacowania funkcji niezawodności w obecności obserwacji cenzurowanych oraz interpretację wyników w kontekście praktycznego zarządzania ryzykiem.

Zbiór danych przedstawiono w Tabeli 3.2. Dla każdej pompy podano czas obserwacji t (w miesiącach) oraz status d, gdzie $d = 1$ oznacza wystąpienie zdarzenia, a $d = 0$ wskazuje na obserwację cenzurowaną (brak zdarzenia w czasie obserwacji).

Estymator Kaplana-Meiera (wzór 3.4) umożliwia oszacowanie funkcji przeżycia (niezawodności) z uwzględnieniem zarówno obserwowanych zdarzeń, jak i danych cenzurowanych. Procedura obliczeniowa polega na uporządkowaniu czasów zdarzeń (3, 4, 5, 6, 7, 9, 10 miesięcy) oraz na mnożeniu czynników z definicji, gdzie d_i oznacza liczbę zdarzeń w chwili t_i, a n_i to liczba pomp w zbiorze ryzyka tuż przed t_i, rozumianym jako zbiór obiektów, dla których nie odnotowano zdarzenia i które pozostają pod obserwacją bezpośrednio przed t_i. Przykładowe obliczenia: $\hat{R}(3) = 1 \cdot (1 - \frac{1}{10}) = 0.9$ oraz $\hat{R}(4) = 0.9 \cdot (1 - \frac{1}{9}) \approx 0.8$.

Pompa	t (miesiące)	d (status)
1	3	1
2	5	1
3	12	0
4	7	1
5	9	1
6	12	0
7	10	1
8	4	1
9	12	0
10	6	1

Tabela 3.2: Status pomp po określonym czasie. Wartość $d = 0$ oznacza dane cenzurowane, $d = 1$ odpowiada wystąpieniu zdarzenia.



Rysunek 3.3: Krzywa niezawodności R uzyskana za pomocą estymatora Kaplana-Meiera. Krzywa schodkowa ilustruje spadek prawdopodobieństwa niezawodności w czasie, z punktem oznaczającym dane cenzurowane dla $t = 12$ miesięcy.

Rezultaty analizy, przedstawione na Rysunku 3.3, dostarczają informacji dla planowania konserwacji. Krzywa niezawodności wskazuje, że prawdopodobieństwo bezawaryjnej pracy pompy przez 8 miesięcy wynosi około 50%. Może to stanowić podstawę do określenia progów ryzyka dla działań prewencyjnych. Metoda Kaplana-Meiera pozwala wykorzystać pełną informację zawartą w danych, w tym obserwacje cenzurowane, które odnoszą się do pomp nadal działające po 12 miesiącach obserwacji.

Zaprezentowane funkcje matematyczne oraz przykład ich zastosowania stanowią podstawę analizy niezawodności i przeżycia, umożliwiając modelowanie i predykcję w obecności danych cenzurowanych. Połączenie teoretycznych podstaw z narzędziami estymacji tworzy zestaw metod analitycznych przydatnych w dalszych zastosowaniach omawianych w kolejnych sekcjach rozdziału.

3.4 Modelowanie zależności od zmiennych objaśniających

W analizie niezawodności i przeżycia bada się także wpływ zmiennych objaśniających — takich jak warunki środowiskowe, charakterystyki jednostek czy zastosowane interwencje — na czas do wystąpienia zdarzenia. W niniejszej sekcji omówiono metody modelowania tych zależności, obejmujące zarówno klasyczne podejścia statystycznych, jak i współczesne techniki uczenia maszynowego dostosowane do specyfiki danych cenzurowanych.

Model proporcjonalnych hazardów Coxa jest najczęściej stosowaną metodą w analizie wielowymiarowej danych przeżycia, pozwalającą na analizę wpływu zmiennych objaśniających bez konieczności specyfikacji bazowego rozkładu hazardu [1]. Model definiuje funkcję hazardu jako:

$$h(t \mid X) = h_0(t)e^{\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p} \tag{3.6}$$

gdzie h_0 reprezentuje bazową funkcję hazardu, a $X = (X_1, X_2, \ldots, X_p)$ to wektor zmiennych predykcyjnych z odpowiadającymi współczynnikami regresji β_i, dla $i = 1, \ldots, p$. Interpretacja modelu opiera się na ilorazach hazardów (ang. *hazard ratio*) — dla wzrostu zmiennej X_i o jedną jednostkę jej skali pomiaru iloraz hazardów wynosi e^{β_i}, co wynika z log-liniowej postaci modelu. Zaletą tego podejścia

jest semi-parametryczny (ang. *semi-parametric*) charakter — model nie zakłada parametrycznej formy funkcji hazardu h_0, a jedynie proporcjonalność hazardów oraz log-liniowy efekt zmiennych objaśniających [70].

Alternatywnym podejściem jest parametryczny model przyspieszonego czasu awarii (ang. *accelerated failure time model*, AFT), który zakłada multiplikatywny wpływ zmiennych objaśniających na czas do zdarzenia. Model wyraża się wzorem:

$$\log(T) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \sigma\epsilon \tag{3.7}$$

gdzie T to czas do zdarzenia, β_0 to wyraz wolny, który ustala poziom bazowy i odpowiada logarytmowi czasu do zdarzenia dla obserwacji referencyjnej przy $X = (0, \ldots, 0)$; β_i to współczynniki regresji dla zmiennych X_i $(i = 1, \ldots, p)$, σ to parametr skali, a ϵ to błąd o określonym rozkładzie. W ujęciu parametrycznym AFT wymaga specyfikacji rozkładu składnika losowego (np. Weibulla, log-normalnego), lecz istnieją również semi-parametryczne wersje niewymagające pełnej parametryzacji rozkładu [71, 72]. Interpretacja na skali czasu jest następująca: wzrost zmiennej X_i o 1 jednostkę mnoży typowy czas do zdarzenia (np. medianę) przez e^{β_i}; gdy $e^{\beta_i} > 1$ czas się wydłuża, a gdy $e^{\beta_i} < 1$ skraca. Dla porównania, model Coxa opisuje efekt względny — wzrost X_i o 1 jednostkę zmienia hazard przez czynnik e^{β_i}, czyli wpływa na zdarzenia w każdym momencie w sposób proporcjonalny (iloraz hazardów) [73, 74]. Przez efekt względny rozumie się multiplikatywną zmianę hazardu (iloraz hazardów), natomiast przez efekt absolutny — multiplikatywną zmianę typowego czasu do zdarzenia (iloraz czasów).

Oba podejścia oferują komplementarne perspektywy analizy danych przeżycia. Model Coxa koncentruje się na względnych efektach zmiennych na hazard, podczas gdy model AFT dostarcza informacji o absolutnych efektach na czas do zdarzenia. Wybór między metodami zależy od charakteru danych, celów analizy oraz wymagań dotyczących interpretowalności wyników [75].

Rozwój uczenia maszynowego umożliwił zarówno adaptację klasycznych algorytmów do problemów analizy przeżycia, jak i opracowanie specjalistycznych technik dedykowanych danym cenzurowanym. Adaptacja klasycznych metod obejmuje dyskretyzację czasu i zastosowanie klasyfikacji binarnej, regresję z traktowaniem obserwacji cenzurowanych jako brakujących danych oraz techniki ważenia uwzględniające specyfikę cenzurowania [76]. Metody zespołowe, takie jak gradient boosting [77] czy bagging, znajdują zastosowanie poprzez agregację predykcji wielu esty-

matorów bazowych, czego przykładem są implementacje w algorytmach XGBoost [78] oraz w przeżyciowych lasach losowych [59].

Modele zespołowe łączą wiele estymatorów bazowych w celu zmniejszenia błędu generalizacji (ang. *generalization error*) i wariancji predykcji. W *baggingu* (*bootstrap aggregating*) trenuje się niezależne estymatory na próbkach *bootstrap* zbioru uczącego, a ich predykcje uśrednia, co zmniejsza wariancję. Dodatkowe losowe podpróbkowanie cech, stosowane w modelach lasów losowych [79], obniża korelację między estymatorami. *Boosting* buduje model łącznie (addytywnie) jako sumę wkładów kolejnych słabych estymatorów bazowych. W kolejnym kroku dopasowuje się nowy estymator do kierunku największej poprawy wartości funkcji straty (np. ujemnego gradientu w *gradient boosting* [77]) i dodaje go z małą wagą (współczynnik uczenia, tzw. *shrinkage*). Taka konstrukcja zmniejsza obciążenie (*bias*) modelu, lecz wymaga regularyzacji (płytkie drzewa, podpróbkowanie, wczesne zatrzymanie; ang. *shallow trees*, *subsampling*, *early stopping*), aby nie zwiększyć wariancji i nie doprowadzić do przeuczenia. *Stacking* polega na nauczeniu modelu łączącego (meta-klasyfikatora/regresora), który przyjmuje jako wejście wektor predykcji estymatorów bazowych uzyskanych w schemacie *out-of-fold* walidacji krzyżowej. Model łączący uczy się wag (np. w regresji liniowej/logistycznej) przypisanych do poszczególnych estymatorów bazowych, co określa ich wkład w finalną predykcję. Różnorodność zespołu osiąga się poprzez *bootstrap* i podpróbkowanie cech, co zmniejsza korelację estymatorów i obniża wariancję predykcji.

Dedykowane techniki uczenia maszynowego dla danych cenzurowanych obejmują drzewa przeżycia, które rozszerzają klasyczne drzewa decyzyjne o możliwość pracy z danymi cenzurowanymi, oraz lasy losowe przeżycia, łączące wiele drzew przeżycia przy zachowaniu interpretowalności [59]. Maszyny wektorów nośnych (ang. *support vector machines*, SVMs) zostały zaadaptowane poprzez modyfikację funkcji straty uwzględniającej ranking czasów zdarzeń [80]. Sieci neuronowe (ang. *neural networks*) znajdują zastosowanie w architekturach takich jak DeepSurv czy DeepHit, umożliwiających modelowanie złożonych, nieliniowych zależności kosztem interpretowalności [13, 81].

4 Wyjaśnialne metody analizy niezawodności i przeżycia

Tradycyjne metody analizy niezawodności i przeżycia, takie jak model proporcjonalnych hazardów Coxa czy estymator Kaplana-Meiera, omówione w Rozdziale 3, zapewniają wysoką interpretowalność, lecz ich zdolność do modelowania złożonych, nieliniowych zależności w danych jest ograniczona. Współczesne środowiska przemysłowe oraz analiza medyczna generują wielowymiarowe, złożone dane, które wymagają zaawansowanych modeli uczenia maszynowego zdolnych do uchwycenia subtelnych wzorców i interakcji między zmiennymi. Objaśnialność i interpretowalność (Sekcja 4.1.1) tych modeli umożliwiają budowanie zaufania, zapewnienie zgodności z regulacjami oraz podejmowanie świadomych decyzji w analizie czasu do wystąpienia krytycznych zdarzeń [10].

Rozwój metod obliczeniowych i dostępność dużych zbiorów danych umożliwiły zastosowanie uczenia maszynowego — dziedziny sztucznej inteligencji zajmującej się automatycznym rozpoznawaniem wzorców z danych — w analizie niezawodności i przeżycia. W tym kontekście interpretowalność i objaśnialność nabierają dodatkowego znaczenia. W przemyśle decyzje oparte na modelach dotyczą kosztownych operacji remontowych, planowania przestojów produkcyjnych czy alokacji zasobów, a w medycynie przewidywania czasów do wystąpienia zdarzeń bezpośrednio wpływają na decyzje kliniczne i życie pacjentów. Ograniczenia tradycyjnych metod analizy przeżycia, takich jak model Coxa czy estymator Kaplana-Meiera, w modelowaniu złożonych zależności nieliniowych wskazują na potrzebę rozwoju zaawansowanych, a jednocześnie interpretowalnych metod uczenia maszynowego.

Niniejszy rozdział przedstawia przegląd metod objaśnialnego i interpretowalnego uczenia maszynowego, omawiając różnice między interpretowalnością wbudowaną w modele a metodami post-hoc służącymi do wyjaśniania decyzji modeli typu „czarnej skrzynki". Szczególną uwagę poświęcono znaczeniu interpretowalności

w analizie niezawodności i przeżycia oraz jej praktycznym implikacjom w predykcyjnym utrzymaniu ruchu i medycynie, tworząc teoretyczne podstawy dla opracowania interpretowalnych algorytmów reguł logicznych przedstawionych w kolejnych rozdziałach.

4.1 Wprowadzenie do objaśnialności i interpretowalności w uczeniu maszynowym

W kontekście analizy niezawodności i przeżycia omówionych w poprzednich rozdziałach, metody uczenia maszynowego oferują zaawansowane podejście do modelowania danych typu czas-do-zdarzenia, rozszerzając możliwości tradycyjnych metod statystycznych zarówno w zastosowaniach przemysłowych, jak i medycznych. W przemyśle pozwalają prognozować awarie, optymalizować harmonogramy konserwacji oraz redukować przestoje, natomiast w medycynie wspierają przewidywanie czasów do wystąpienia zdarzeń klinicznych, planowanie terapii oraz ocenę skuteczności leczenia. Jednak wysoka skuteczność predykcyjna modeli uczenia maszynowego często wynika z ich złożoności, co prowadzi do sytuacji, w której proces decyzyjny modelu pozostaje niezrozumiały dla człowieka [82]. W predykcyjnym utrzymaniu ruchu i medycynie, w których decyzje modeli bezpośrednio wpływają na bezpieczeństwo operacyjne lub życie pacjentów, zrozumienie podstaw tych decyzji staje się nieodzowne [30, 83].

Celem niniejszej sekcji jest przedstawienie najważniejszych pojęć objaśnialności i interpretowalności w uczeniu maszynowym, wraz z ich rolą w analizie niezawodności i przeżycia w predykcyjnym utrzymaniu ruchu oraz w zastosowaniach biomedycznych. Przedstawiono w niej definicje tych terminów, analizę ich znaczenia w praktyce przemysłowej i medycznej oraz omówienie wyzwań związanych z ich implementacją.

4.1.1 Definicje interpretowalności i objaśnialności

W literaturze poświęconej analizie modeli uczenia maszynowego wyróżnia się m.in. interpretowalność i objaśnialność [84, 85]. Interpretowalność określa stopień, w jakim człowiek może zrozumieć przyczyny decyzji podejmowanych przez model uczenia maszynowego [84]. Model interpretowalny charakteryzuje się przejrzystym

mechanizmem decyzyjnym, który nie wymaga dodatkowych wyjaśnień. Przykładem jest drzewo decyzyjne, w którym każda decyzja jest jednoznacznie powiązana z określonymi zmiennymi wejściowymi [86]. Objaśnialność natomiast odnosi się do zdolności modelu do dostarczania wyjaśnień dotyczących jego funkcjonowania lub konkretnych predykcji, także w przypadku modeli nieinterpretowalnych [85].

Interpretowalność stanowi zatem cechę wewnętrzną modelu, podczas gdy objaśnialność może być uzyskana dzięki zewnętrznym technikom, takim jak analiza ważności cech (ang. *feature importance*) czy metod wizualizacji wpływu cech na predykcje [86]. Modele nieinterpretowalne, takie jak głębokie sieci neuronowe, mogą być poddawane technikom objaśnialności w celu lepszego zrozumienia ich decyzji [87].

4.1.2 Znaczenie interpretowalności i objaśnialności

Objaśnialność i interpretowalność odgrywają szczególną rolę w dziedzinach o podwyższonym ryzyku, takich jak predykcyjne utrzymanie ruchu i medycyna, gdzie decyzje podjęte na podstawie modeli uczenia maszynowego mogą mieć bezpośredni wpływ na bezpieczeństwo operacyjne, stabilność procesów produkcyjnych, koszty związane z przestojami, a także na życie i zdrowie pacjentów. W takich zastosowaniach równie ważna jak sama dokładność precyzji jest interpretacja mechanizmów stojących za estymacją czasu do awarii czy prognozy przeżycia — na przykład identyfikacja zmiennych determinujących wysokie ryzyko wystąpienia zdarzenia w krótkim horyzoncie czasowym. Transparentność modelu pozwala nie tylko zwiększyć skuteczność działań podejmowanych na jego podstawie, ale także budować zaufanie wśród użytkowników w środowiskach przemysłowych i medycznych [30].

Istotność tych zagadnień oddają następujące przykłady:

- Inżynierowie utrzymania ruchu muszą mieć możliwość weryfikacji i zaufania rekomendacjom generowanym przez model uczenia maszynowego, aby skutecznie planować i wdrażać działania prewencyjne. Podobnie, lekarze wymagają zrozumienia podstaw przewidywań modeli dotyczących prognozy pacjenta, aby podejmować właściwe decyzje kliniczne. Bez wiedzy o tym, na podstawie jakich zmiennych model lub wzorców model estymuje wysokie prawdopodobieństwo wystąpienia zdarzenia w określonym czasie, predyk-

cje te mogą zostać zignorowane lub błędnie zinterpretowane, co obniża ich użyteczność [83].

- W przypadku potwierdzenia wystąpienia zdarzenia retrospektywna analiza predykcji modelu umożliwia identyfikację czynników, które przyczyniły się do właściwej lub błędnej estymacji, a także opracowanie strategii poprawy jakości przewidywań. Objaśnialność pozwala ustalić, czy predykcja wynikała z konkretnych odczytów czujników przemysłowych bądź parametrów biomedycznych, co może wskazywać na potrzebę kalibracji sprzętu, modyfikacji procedur konserwacyjnych lub zmian w protokołach leczenia [82].

Ponadto w sektorach takich jak transport, energetyka czy ochrona zdrowia obowiązują regulacje prawne oraz normy etyczne nakładające wymóg transparentności decyzji podejmowanych przez systemy automatyczne. Przykładem jest unijne ogólne rozporządzenie o ochronie danych (RODO, ang. *General Data Protection Regulation*, GDPR)[1], które w kontekście zautomatyzowanego podejmowania decyzji wymaga możliwości wyjaśnienia procesów decyzyjnych [88]. Brak objaśnialności w takich sytuacjach może prowadzić do utraty zaufania do modelu zarówno ze strony użytkowników, jak i organów regulacyjnych, a w skrajnych przypadkach skutkować naruszeniem prawa, pociągając za sobą konsekwencje finansowe i reputacyjne. Dlatego zapewnienie objaśnialności i interpretowalności staje się nie tylko wyzwaniem technicznym, lecz także priorytetem w projektowaniu i wdrażaniu modeli uczenia maszynowego w zastosowaniach wysokiego ryzyka.

4.1.3 Zależność między złożonością a interpretowalnością

Dobór modeli zależy od celu analizy. Gdy celem jest wnioskowanie statystyczne (estymacja efektów zmiennych i testowanie hipotez) oraz interpretowalność parametrów i efektów zmiennych, preferowane są modele o niskiej złożoności (np. liniowe). Gdy natomiast priorytetem jest wyłącznie predykcja, stosuje się modele o większej zdolności dopasowania do danych, zdolne do uchwycenia nieliniowości i interakcji [84, 87]. Modele liniowe (np. regresja liniowa czy model Coxa) umożliwiają

[1] Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2016/679 z dnia 27 kwietnia 2016 r. w sprawie ochrony osób fizycznych w związku z przetwarzaniem danych osobowych i w sprawie swobodnego przepływu takich danych oraz uchylenia dyrektywy 95/46/WE (ogólne rozporządzenie o ochronie danych)

bezpośrednią interpretację wpływu zmiennych na wynik, lecz mogą ustępować dokładnością modelom nieliniowym o większej złożoności (większej zdolności funkcji do dopasowania się do danych). Z kolei modele o wysokiej złożoności — takie jak boosting [77], lasy losowe [79] czy nieliniowe maszyny wektorów nośnych [89] — często osiągają niższy błąd generalizacji (np. wyższy indeks zgodności w analizie przeżycia, niższy błąd średniokwadratowy lub log-loss), kosztem zrozumiałości mechanizmu decyzyjnego [84].

W kontekście zależności między złożonością a interpretowalnością modele różnią się zakresem dopuszczalnych zależności (np. liniowych, nieliniowych, z interakcjami) między zmiennymi objaśniającymi a odpowiedzią [77, 90]. Modele liniowe charakteryzują się ograniczoną zdolnością dopasowania do danych i wysoką interpretowalnością — współczynniki mają jednoznaczną interpretację, wskazując kierunek i siłę efektu przy ustalonych pozostałych cechach. Podejścia promujące rzadkość rozwiązań (ang. *sparsity*, np. selekcja zmiennych lub regularyzacja L1) sprzyjają interpretacji, ograniczając liczbę predyktorów w modelu. Uogólnione modele addytywne (ang. *generalized additive models*, GAM) zwiększają zdolność dopasowania do danych, dopuszczając nieliniowe, addytywne efekty przy zachowaniu częściowej przejrzystości na poziomie funkcji składowych. Zespoły drzew i głębokie sieci oferują jeszcze większą swobodę funkcjonalną, ale ich interpretacja wymaga zwykle metod post-hoc [87, 84].

Większa zdolność dopasowania do danych nie gwarantuje jednak lepszej predykcji. Modele o wysokiej złożoności mogą nadmiernie dopasowywać się do danych uczących (ang. *overfitting*), co obniża zdolność uogólniania. W wielu zadaniach — w tym w analizie przeżycia — mniej złożone modele lub silniej regularyzowane procedury zapewniają korzystniejszy kompromis między obciążeniem (ang. *bias*) a wariancją (ang. *variance*), a w efekcie większą dokładność prognoz [91]. W kontekstach wysokiego ryzyka priorytetem jest przejrzystość decyzji, dlatego zaleca się stosowanie modeli interpretowalnych wszędzie tam, gdzie jest to możliwe [82]. Dobór modelu jest więc kompromisem między dokładnością prognozy a przejrzystością decyzji, zależnym od celu analizy i kontekstu zastosowania [92].

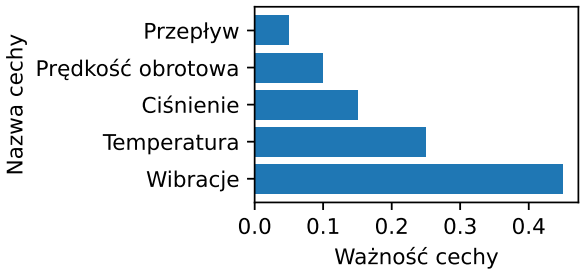
4.2 Metody objaśnialnego uczenia maszynowego

W analizie niezawodności i przeżycia stosowanie złożonych modeli uczenia maszynowego do modelowania danych typu czas-do-zdarzenia wymaga metod objaśnialnych, które umożliwiają zrozumienie procesów decyzyjnych związanych z przewidywaniem czasu do wystąpienia krytycznych zdarzeń [87]. Tradycyjne metryki oceny modeli przeżycia, takie jak indeks zgodności Harrella czy statystyka log-rank, koncentrują się na ocenie jakości predykcji. Współczesne systemy analizy niezawodności muszą jednak dodatkowo zapewniać wgląd w mechanizmy prowadzące do prognozowanych czasów awarii lub zdarzeń medycznych [82]. Brak transparentności w modelach przeżycia stanowi barierę dla akceptacji zaawansowanych technologii w zastosowaniach przemysłowych i medycznych [85].

Metody objaśnialnego uczenia maszynowego dzieli się na dwie główne kategorie: globalne i lokalne. Metody globalne pozwalają analizować ogólne zachowanie modelu na całym zbiorze danych, oferując wgląd w dominujące wzorce i zależności. Metody lokalne natomiast skupiają się na wyjaśnianiu pojedynczych predykcji, dostarczając szczegółowych informacji o decyzjach modelu w odniesieniu do konkretnych przypadków. Oba podejścia wzajemnie się uzupełniają. Metody globalne zapewniają szeroką perspektywę działania modelu, a lokalne umożliwiają wgląd w szczegóły konkretnych predykcji. Połączenie obu metod umożliwia pełną interpretację modeli uczenia maszynowego i stanowi podstawę współczesnych systemów wspomagania decyzji w przemyśle i medycynie [30].

4.2.1 Metody globalne

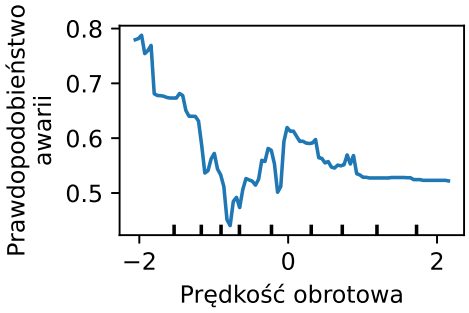
Metody globalne służą zrozumieniu zachowania modelu przeżycia na całym zbiorze danych, dostarczając wglądu w mechanizmy decyzyjne związane z przewidywaniem czasów do wystąpienia zdarzeń. Umożliwiają identyfikację wzorców oraz zależności między zmiennymi objaśniającymi a estymowanymi funkcjami przeżycia lub hazardu, co jest szczególnie przydatne w początkowych fazach projektowania i walidacji modelu [93]. Dzięki temu eksperci dziedzinowi, tacy jak inżynierowie niezawodności czy lekarze kliniczni, mogą ocenić, czy model odzwierciedla rzeczywiste procesy degradacji urządzeń lub progresji choroby, co sprzyja jego akceptacji w analizie niezawodności i praktyce klinicznej. Do najważniejsze metody w tej kategorii przedstawiono poniżej.



Rysunek 4.1: Przykładowa wizualizacja ważności cech w analizie predykcyjnego utrzymania ruchu dla systemu pomp przemysłowych. Ważność cech, wyznaczona np. poprzez permutację, wskazuje dominujący wpływ danych z czujników wibracji na predykcję awarii, co wspiera ukierunkowane działania konserwacyjne.

- Analiza ważności cech: Ta metoda umożliwia określenie, które zmienne objaśniające mają największy wpływ na estymowane funkcje przeżycia lub hazardu w analizie niezawodności maszyn oraz w zastosowaniach biomedycznych [94]. Ważność cech można wyznaczać poprzez permutację cech z oceną spadku indeksu zgodności Harrella modelu po losowym zaburzeniu danej zmiennej albo poprzez analizę współczynników w modelach proporcjonalnych hazardów. Wyniki najczęściej prezentuje się w formie rankingów lub wizualizacji graficznych, ułatwiając wskazanie głównych predyktorów czasu do awarii (np. temperatura, ciśnienie, wibracje) czy przeżycia pacjentów (np. biomarkery, wiek, historia choroby). W systemach monitorowania pomp przemysłowych analiza ważności cech może ujawnić, że najważniejszym predyktorem są dane z czujników drgań, natomiast w analizie medycznej może wskazać na biomarkery o decydującym znaczeniu dla prognozy. Umożliwia to ukierunkowanie działań konserwacyjnych w przemyśle lub interwencji terapeutycznych w medycynie. Przykładem zastosowania tej techniki jest wizualizacja ważności cech przedstawiona na Rysunku 4.1.

- Wykresy zależności częściowej (ang. *partial dependence plots*, PDP): PDP przedstawiają relację między wybraną cechą a wynikiem modelu, przy założeniu stałych wartości pozostałych zmiennych [77]. Technika ta umożliwia wizualizację wpływu zmian jednej zmiennej na predykcje i jest szczególnie przydatna w analizie nieliniowych zależności charakterystycznych dla złożonych modeli uczenia maszynowego. W praktyce przemysłowej PDP mogą



Rysunek 4.2: Przykładowy wykres zależności częściowej ilustrujący wpływ prędkości obrotowej silnika na prawdopodobieństwo predykcji awarii, wygenerowany dla modelu Random Forest. Wykres pokazuje nieliniową relację, co wspiera analizę ryzyka awarii w predykcyjnym utrzymaniu ruchu, przy założeniu niezależności cech.

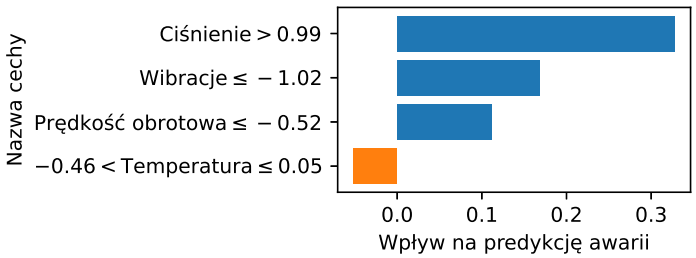
posłużyć do oceny, jak zmiany parametrów operacyjnych — np. prędkości obrotowej silnika — wpływają na ryzyko awarii. W medycynie natomiast pozwalają analizować zależności między wiekiem pacjenta czy poziomem biomarkera a prawdopodobieństwem wystąpienia zdarzenia, stanowiąc cenne narzędzie interpretacji modelu. Ograniczeniem tej metody jest założenie niezależności cech, które może prowadzić do zniekształceń w przypadku silnych korelacji między zmiennymi [95]. Przykładem zastosowania tej techniki jest wykres zależności częściowej przedstawiony na Rysunku 4.2, który ilustruje wpływ prędkości obrotowej silnika na ryzyko awarii.

Metody globalne znajdują zastosowanie w fazie eksploracyjnej analizy danych oraz podczas walidacji modelu przez specjalistów, umożliwiając ocenę, czy predykcje modelu są zgodne z wiedzą dziedzinową. Ich główną wadą jest jednak brak zdolności do uchwycenia lokalnych różnic w zachowaniu modelu, co może stanowić problem w sytuacjach, gdy predykcje dla poszczególnych przypadków znacząco odbiegają od ogólnych trendów. W takich sytuacjach konieczne jest sięgnięcie po metody lokalne, które uzupełniają analizę globalną.

4.2.2 Metody lokalne

Metody lokalne koncentrują się na wyjaśnianiu pojedynczych predykcji, oferując wgląd w decyzje modelu dla poszczególnych obserwacji. Stosuje się je w sytuacjach

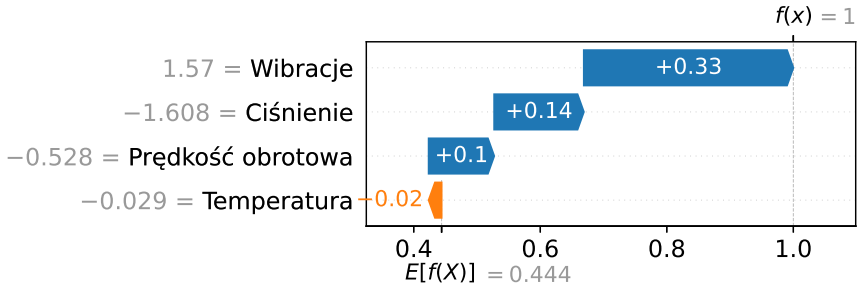
wymagających interpretacji konkretnych przewidywań czasów do awarii urządzeń przemysłowych lub zdarzeń medycznych, co wpływa na wybór strategii konserwacyjnej lub protokołu leczenia [10]. W odróżnieniu od metod globalnych, dających ogólny obraz funkcjonowania modelu, metody lokalne pozwalają analizować indywidualne przypadki, które wymagają szczególnej uwagi, np. gdy model przewiduje rzadkie, ale krytyczne zdarzenia przemysłowe lub medyczne. Najczęściej stosowane metody w tej kategorii przedstawiono poniżej.



Rysunek 4.3: Przykładowy wykres lokalnych wyjaśnień wygenerowany za pomocą LIME dla predykcji awarii. Wykres przedstawia ważność cech, takich jak wibracje czy temperatura, dla jednej obserwacji. Pokazuje ich pozytywny lub negatywny wpływ na predykcję klasy „Awaria" w kontekście predykcyjnego utrzymania ruchu. Demonstruje też niezależność LIME od typu modelu i jego zdolność do lokalnej interpretacji.

- LIME (Local Interpretable Model-agnostic Explanations): LIME działa poprzez lokalną aproksymację objaśnialnego modelu za pomocą prostego, interpretowalnego modelu (np. regresji liniowej) w sąsiedztwie wybranej obserwacji [86]. Metoda ta generuje wyjaśnienia oparte na ważności cech w lokalnym kontekście, wskazując, które zmienne miały największy wpływ na daną predykcję. W zastosowaniach przemysłowych LIME może ujawnić, że dla konkretnej maszyny predykcja awarii wynikała głównie z nieznacznie podwyższonych wartości wibracji, podczas gdy inne parametry, takie jak temperatura czy ciśnienie, mieściły się w granicach normy i sugerowały stabilność systemu. W analizie medycznej metoda ta może wskazać, że dla danego pacjenta predykcja wysokiego ryzyka wynikała przede wszystkim z podwyższonego poziomu konkretnego biomarkera, podczas gdy inne parametry, np. wiek czy BMI, pozostawały w normie. LIME jest niezależne od typu modelu (ang. *model-agnostic*), co czyni je uniwersalnym narzędziem.

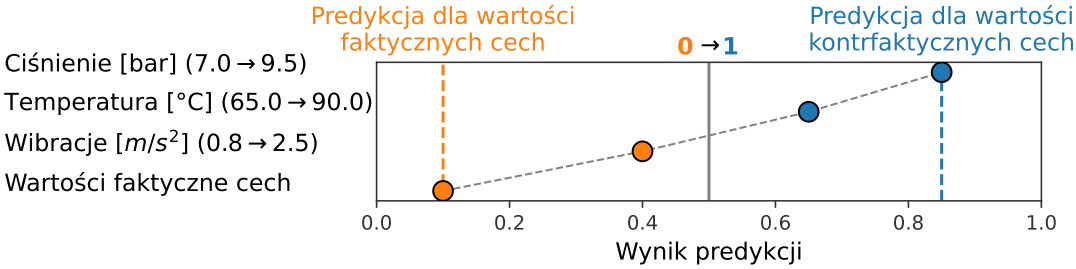
Skuteczność metody zależy jednak od odpowiedniego określenia lokalnego otoczenia analizowanej obserwacji i może być ograniczona w przypadku złożonych zależności między cechami [96]. Przykładem zastosowania tej techniki jest wykres lokalnych wyjaśnień przedstawiony na Rysunku 4.3, który ilustruje wpływ cech na predykcję awarii dla konkretnej maszyny.



Rysunek 4.4: Przykładowy wykres wodospadowy wygenerowany za pomocą SHAP dla predykcji awarii pompy. Wykres ilustruje, jak wkłady cech — wibracje (+1.57), ciśnienie (−1.608), prędkość obrotowa (−0.528) i temperatura (−0.029) — kumulują się od wartości bazowej predykcji ($E[f(x)] = 0.444$) do końcowej predykcji klasy „Awaria" ($f(x) = 1$), z oznaczeniem ich wpływu (niebieski — dodatni, pomarańczowy — ujemny). Prezentacja ta ilustruje podział predykcji na addytywne wkłady poszczególnych cech zgodnie z metodologią SHAP.

- SHAP (SHapley Additive exPlanations): SHAP opiera się na wartościach Shapleya pochodzących z teorii gier, przypisując każdej cesze wkład do konkretnej predykcji w sposób zgodny z aksjomatycznym podziałem wkładu [97]. Dodatkowo wartości SHAP można sumować dla wszystkich obserwacji, co umożliwia analizę globalnej ważności cech w całym zbiorze danych. W zastosowaniach przemysłowych SHAP może wskazać, że przewidywana awaria pompy wynikała w 60% z anomalii w ciśnieniu i w 30% z temperatury. Takie informacje pozwalają inżynierom planować działania prewencyjne w sposób bardziej precyzyjny. W medycynie SHAP może analogicznie określić, że przewidywane ryzyko zgonu pacjenta wynikało w 45% z poziomu kreatyniny, w 25% z wieku i w 20% z obecności chorób współistniejących, umożliwiając lekarzom dokładne planowanie interwencji medycznych. Choć SHAP jest bardziej wymagający obliczeniowo niż LIME, jego zaletą jest większa dokładność oraz możliwość interpretacji zarówno dodatnich, jak i ujemnych

wkładów cech. Przykładem zastosowania tej techniki jest wykres wodospadowy (ang. *waterfall plot*) przedstawiony na Rysunku 4.4, który ilustruje wkłady cech, takich jak ciśnienie i temperatura, w predykcję awarii pompy.



Rysunek 4.5: Wykres kontrfaktycznych wyjaśnień dla predykcji awarii turbiny wiatrowej, ilustrujący różnice między wartościami faktycznymi a kontrfaktycznymi cech. Kolejne punkty reprezentują stopniową zmianę wyniku predykcji wraz ze zmianą wartości trzech cech: ciśnienia (z 7.0 do 9.5 bar), temperatury (z 65.0 do 90.0 °C) i wibracji (z 0.8 do 2.5 m/s^2). Pierwszy pomarańczowy punkt (predykcja dla wartości faktycznych cech, $x = 0.1$) wskazuje pierwotną predykcję modelu na podstawie aktualnych wartości cech, natomiast ostatni niebieski punkt (predykcja dla wartości kontrfaktycznych cech, $x = 0.85$) reprezentuje alternatywną predykcję, która wynikałaby, gdyby wszystkie cechy osiągnęły kontrfaktyczne wartości. Przebieg linii przerywanej pokazuje hipotetyczną trajektorię zmiany predykcji, co pozwala zrozumieć, jakie modyfikacje cech mogłyby zapobiec przewidywanej awarii.

- Wyjaśnienia kontrfaktyczne (ang. *Counterfactual Explanations*): Ta metoda polega na generowaniu alternatywnych scenariuszy, które pokazują, jak zmiana wartości cech mogłaby doprowadzić do innej predykcji modelu [98]. Wyjaśnienia kontrfaktyczne dostarczają odpowiedzi na pytanie, jakie modyfikacje wartości cech byłyby konieczne, aby uzyskać odmienny wynik predykcji modelu. W zastosowaniach przemysłowych wyjaśnienie kontrfaktyczne może wskazać, że awaria turbiny wiatrowej nie zostałaby przewidziana, gdyby wartości wibracji łopat były o 10% niższe, przy zachowaniu pozostałych parametrów na obecnym poziomie. W medycynie metoda ta może określić, że dany pacjent nie byłby uznany za przypadek wysokiego ryzyka, gdyby jego poziom cholesterolu był o 20% niższy lub gdyby nie występowało u niego nadciśnienie. Wyjaśnienia kontrfaktyczne są szczególnie przydatne w sytuacjach,

gdzie celem jest nie tylko zrozumienie przyczyn predykcji, ale także identyfikacja działań zapobiegawczych. Ich zaletą jest intuicyjność i bezpośrednie powiązanie z interwencjami operacyjnymi, jednak skuteczne stosowanie wymaga precyzyjnego określenia dopuszczalnych zmian w cechach, co może być wyzwaniem w złożonych systemach. Przykładem zastosowania tej techniki jest wykres kontrfaktycznych wyjaśnień przedstawiony na Rysunku 4.5, który ilustruje różnice w wartościach cech między oryginalną predykcją awarii a kontrfaktycznym scenariuszem dla turbiny wiatrowej.

Metody lokalne znajdują zastosowanie w praktyce przemysłowej i medycznej, gdzie precyzyjne wyjaśnienie pojedynczych predykcji może bezpośrednio przełożyć się na decyzje operacyjne lub kliniczne [83]. Przykładowo, gdy model przewiduje awarię krytycznego komponentu, takiego jak turbina wiatrowa, lub wysokie ryzyko u pacjenta, techniki takie jak LIME, SHAP czy wyjaśnienia kontrfaktyczne dostarczają komplementarnych informacji: LIME wskazuje dominujące cechy, SHAP określa ich dokładny wkład, a wyjaśnienia kontrfaktyczne sugerują, jakie zmiany mogłyby zapobiec problemowi. Podstawową wartością metod lokalnych jest zdolność do dostarczania szczegółowych, kontekstowych wyjaśnień dla konkretnych przypadków. Takie podejście umożliwia uzasadnienie przewidywanego czasu do awarii lub ryzyka zdarzenia medycznego, w celu podejmowania działań prewencyjnych lub terapeutycznych.

4.3 Interpretowalne metody uczenia maszynowego

W analizie niezawodności i przeżycia metody interpretowalne odnoszą się do modeli, których mechanizmy przewidywania czasów do wystąpienia krytycznych zdarzeń są przejrzyste [90]. W odróżnieniu od złożonych modeli typu „black-box", takich jak głębokie sieci neuronowe, metody interpretowalne dostarczają wglądu w procesy estymacji funkcji przeżycia i hazardu. Są szczególnie przydatne w zastosowaniach wymagających wysokiego poziomu transparentności, gdzie zrozumienie przyczyn przewidywanych czasów awarii wspiera podejmowanie decyzji konserwacyjnych oraz klinicznych [91]. Niniejsza sekcja przedstawia cztery główne kategorie interpretowalnych metod uczenia maszynowego: drzewa decyzyjne, reguły logiczne, modele regresyjne oraz metody oparte na instancjach. Każda z tych metod wnosi specyficzne właściwości interpretacyjne do modelowania danych typu

czas-do-zdarzenia. Dzięki temu są przydatne zarówno w analizie niezawodności przemysłowej, jak i biomedycznej, gdzie precyzja predykcji czasów przeżycia oraz zrozumiałość modelu mają równorzędne znaczenie [84].

4.3.1 Drzewa decyzyjne

Drzewa decyzyjne to modele hierarchiczne stosowane zarówno w zadaniach klasyfikacji, jak i regresji, które odwzorowują proces decyzyjny za pomocą struktury drzewiastej [99]. Węzły wewnętrzne reprezentują testy warunków na dotyczących wartości atrybutów, gałęzie odpowiadają wynikom tych testów, a liście zawierają końcowe predykcje. Interpretowalność drzew decyzyjnych wynika z ich wizualnej formy, która umożliwia łatwe śledzenie ścieżki decyzyjnej dla dowolnej obserwacji.

Matematycznie, drzewo decyzyjne dzieli przestrzeń cech X na rozłączne podzbiory $R_1, R_2, \ldots, R_m$, którym odpowiadają predykcje $\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_m$ [60]. W klasyfikacji $\hat{y}_j$ jest klasą najczęściej występującą w danym obszarze, natomiast w regresji stanowi wartość średnią lub medianę zmiennej docelowej w R_j. Algorytm uczenia drzewa opiera się na rekurencyjnym podziale przestrzeni cech, minimalizując funkcję kosztu, taką jak entropia (dla klasyfikacji) lub średni błąd kwadratowy (ang. *mean square error, MSE*) (dla regresji) [100].

Drzewa decyzyjne wyróżniają się prostotą interpretacji, zdolnością do uchwycenia nieliniowych zależności oraz automatycznym modelowaniem interakcji między atrybutami. Ich ograniczeniem jest jednak tendencja do nadmiernego dopasowania (ang. *overfitting*), co można korygować poprzez techniki przycinania drzew (ang. *decision tree pruning*) lub stosowanie metod zespołowych, takich jak lasy losowe (ang. *random forests*), kosztem częściowej utraty interpretowalności [79]. W analizie przeżycia wykorzystuje się drzewa przeżyciowe (ang. *survival trees*), które stanowią adaptację drzew decyzyjnych do danych cenzurowanych. Ich liście zawierają estymatory funkcji przeżycia (np. krzywe Kaplana-Meiera), co umożliwia predykcję czasów do wystąpienia zdarzenia dla różnych podgrup pacjentów lub urządzeń [63, 59].

4.3.2 Reguły decyzyjne

Reguły decyzyjne to zbiory instrukcji logicznych w formie „jeśli-to", które określają predykcje na podstawie spełnienia określonych warunków [101]. W odróżnieniu

od hierarchicznej struktury drzew decyzyjnych, reguły mogą być zorganizowane w sposób nieuporządkowany lub częściowo uporządkowany, co zwiększa zakres ich zastosowania. Charakteryzują się podobieństwem do języka naturalnego, co ułatwia ich zrozumienie [102].

Formalnie, reguła decyzyjna r ma postać:

$$\text{jeśli } \phi \text{ to } \psi, \tag{4.1}$$

gdzie ϕ stanowi koniunkcję warunków na atrybutach (np. $x_1 > a \ \wedge x_2 = b$), a ψ jest predykcją — na przykład klasą, wartością liczbową lub, w analizie przeżycia, estymatą prawdopodobieństwa przeżycia bądź ryzyka wystąpienia zdarzenia w określonym czasie. Zbiór reguł $R = \{r_1, r_2, \ldots, r_k\}$ tworzy model predykcyjny, w którym dla obserwacji $\mathbf{x}$ predykcja wynika z reguły (lub reguł), dla której spełniony jest warunek $\phi(\mathbf{x})$ [103]. W przypadku nakładania się reguł, konflikty rozstrzyga się za pomocą mechanizmów takich jak głosowanie większościowe lub priorytetyzacja reguł.

Interpretowalność reguł wynika z ich prostej formy warunkowej „jeśli-to", która mapuje kombinacje wartości zmiennych objaśniających na predykcje. Reguły w ramach zbioru są niezależne, dzięki czemu można ją analizować konieczności rozumienia całego modelu. W predykcyjnym utrzymaniu ruchu reguły mogą określać konkretne kombinacje parametrów (np. „jeśli temperatura $>75\,^{\circ}\mathrm{C}$ i wibracje $> 5\,\mathrm{m/s^2}$, to prawdopodobieństwo awarii w ciągu 30 dni wynosi $\geqslant 70\%$."), oferując praktyczne wskazówki operacyjne [104]. W analizie przeżycia termin „reguły decyzyjne" odnosi się do reguł przeżyciowych (ang. *survival rules*), które określają warunki związane z czasem do wystąpienia zdarzenia, np. „jeśli wiek pacjenta > 65 lat i poziom kreatyniny $> 2.0\,\mathrm{mg/dL}$, to prawdopodobieństwo 5-letniego przeżycia wynosi $< 40\%$".

4.3.3 Modele regresyjne

Modele regresyjne stanowią klasę interpretowalnych metod uczenia maszynowego [53]. W kontekście predykcyjnego utrzymania ruchu można wyróżnić trzy główne kategorie modeli regresyjnych o wysokiej interpretowalności: regresję liniową, regresję Coxa oraz regresję logistyczną.

Regresja liniowa opisuje zależność między zmienną objaśnianą y a zmiennymi objaśniającymi $\mathbf{x} = (x_1, x_2, \ldots, x_p)$ za pomocą liniowej kombinacji:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \epsilon, \tag{4.2}$$

gdzie $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)$ to wektor parametrów, a ϵ reprezentuje składnik losowy [105]. Interpretowalność wynika z bezpośredniej relacji między współczynnikami $\beta_i, i = 0, \ldots, p$, a wpływem zmiennych na predykcję — każdy współczynnik określa oczekiwaną zmianę zmiennej objaśnianej przy jednostkowej zmianie odpowiedniej zmiennej objaśniającej. W analizie przeżycia rolę tej klasy modeli pełnią modele typu AFT (ang. *accelerated failure time*), które odnoszą się bezpośrednio do czasu do zdarzenia (szczegóły i przykłady znajdują się w Sekcji 3.4).

Model proporcjonalnych hazardów Coxa (szczegółowo opisany w Rozdziale 3) definiuje funkcję hazardu jako [1]:

$$h(t \mid \mathbf{x}) = h_0(t) e^{\boldsymbol{\beta}^T \mathbf{x}}, \tag{4.3}$$

gdzie h_0 to funkcja hazardu bazowego, a $\boldsymbol{\beta}^T \mathbf{x}$ to liniowy predyktor, czyli suma ważona wartości zmiennych objaśniających dla rozważanej obserwacji. Z postaci $e^{\boldsymbol{\beta}^T \mathbf{x}}$ wynika, że przy wzroście wartości danej cechy o 1 (przy pozostałych cechach stałych) hazard jest mnożony przez stały czynnik (tzw. iloraz hazardów), niezależny od czasu. Model ten jest semi-parametryczny — nie wymaga specyfikacji rozkładu czasów przeżycia i koncentruje się na relatywnym wpływie zmiennych objaśniających na ryzyko zdarzenia. Jeśli współczynnik przy danej zmiennej objaśniającej jest dodatni, czynnik jest większy od 1 (większe ryzyko); jeśli ujemny — mniejszy od 1 (mniejsze ryzyko).

Regresja logistyczna służy do klasyfikacji i znajduje zastosowanie w modelowaniu prawdopodobieństwa awarii w określonym przedziale czasowym. Prawdopodobieństwo wystąpienia zdarzenia wyraża się wzorem:

$$P(y = 1 \mid \mathbf{x}) = \frac{1}{1 + e^{-\boldsymbol{\beta}^T \mathbf{x}}}, \tag{4.4}$$

gdzie $\boldsymbol{\beta}$ to wektor parametrów modelu, $\mathbf{x}$ to wektor zmiennych objaśniających, a indeks górny T oznacza transpozycję. Interpretowalność wynika z liniowego predyktora $\boldsymbol{\beta}^T \mathbf{x}$ — dodatnie składowe wektora $\boldsymbol{\beta}$ zwiększają przewidywane prawdopodobieństwo wystąpienia zdarzenia, ujemne je zmniejszają. Większa wartość

bezwzględna składowej oznacza silniejszy wpływ (przy pozostałych cechach stałych).

Modele regresyjne umożliwiają kwantyfikację wpływu czynników operacyjnych na czas eksploatacji urządzeń. Przykładowo, w systemie monitorowania turbin wiatrowych regresja Coxa może wykazać, że wzrost prędkości wiatru o jedną jednostkę zwiększa ryzyko awarii łożysk proporcjonalnie do współczynnika $e^{\beta_{\text{wiatr}}}$, dostarczając precyzyjnych informacji do planowania konserwacji prewencyjnej [29].

4.3.4 Metody oparte na instancjach

Metody oparte na instancjach (ang. *instance-based learning*) zakładają, że obserwacje o podobnych wartościach w przestrzeni cech wykazują zbliżone wartości zmiennej docelowej [106]. Przykładem jest algorytm k-najbliższych sąsiadów (ang. *k-nearest neighbors*, k-NN), który dla nowej obserwacji $\mathbf{x}$ identyfikuje k najbardziej podobnych przykładów z danych treningowych i wyznacza predykcję na podstawie ich wartości zmiennej docelowej, np. poprzez głosowanie większościowe w zadaniach klasyfikacji lub obliczenie średniej w zadaniach regresji [107].

Matematycznie, dla metryki odległości d, zbiór k najbliższych sąsiadów $N_k(\mathbf{x})$ definiuje się jako:

$$N_k(\mathbf{x}) = \arg \min_{S \subseteq D, |S|=k} \sum_{\mathbf{x}_i \in S} d(\mathbf{x}, \mathbf{x}_i), \tag{4.5}$$

gdzie D to zbiór treningowy. Predykcja w klasyfikacji jest dana wzorem:

$$\hat{y} = \arg\max_{c} \sum_{\mathbf{x}_i \in N_k(\mathbf{x})} \mathbb{I}(y_i = c), \tag{4.6}$$

gdzie c oznacza etykietę klasy, a $\mathbb{I}$ to funkcja charakterystyczna zbioru (ang. *indicator function*) [108].

Interpretowalność tej metody wynika z jej prostoty — predykcja jest uzasadniona konkretnymi przykładami z danych, co pozwala użytkownikowi przeanalizować najbliższych sąsiadów i zrozumieć podstawy decyzji [109]. W analizie przeżycia metody oparte na instancjach mogą wykorzystywać podobieństwo w przestrzeni cech do identyfikacji przypadków o zbliżonych czasach do wystąpienia zdarzenia [106]. Algorytm k-NN może estymować czas do awarii na podstawie historycznych przypadków urządzeń o podobnych parametrach operacyjnych, a w medycynie może

identyfikować pacjentów o zbliżonych profilach klinicznych w celu prognozowania czasów przeżycia.

5 Analiza niezawodności i przeżycia za pomocą reguł logicznych

Niniejszy rozdział koncentruje się na regułach logicznych jako interpretowalnej metodzie analizy danych cenzurowanych. Reguły logiczne łączą interpretowalność klasycznych metod statystycznych analizy przeżycia ze zdolnością modelowania złożonych zależności, charakterystyczną dla zaawansowanych algorytmów uczenia maszynowego [102].

W kontekście analizy przeżycia reguły logiczne umożliwiają reprezentację odkrytej wiedzy w formie intuicyjnych warunków logicznych, zrozumiałych dla ekspertów dziedzinowych w medycynie i predykcyjnym utrzymaniu ruchu. W przeciwieństwie do tradycyjnych metod statystycznych, mogą modelować nieliniowe wielowymiarowe zależności, zachowując jednocześnie transparentność procesów decyzyjnych wymaganą w zastosowaniach krytycznych. Adaptacja algorytmów indukcji reguł do specyfiki danych cenzurowanych stanowi wyzwanie metodologiczne, wymagające uwzględnienia mechanizmów cenzurowania w procesach uczenia i predykcji.

Rozdział ten przedstawia proces indukcji reguł oraz rolę reguł klasyfikacyjnych w analizie przeżycia. Wprowadzenia teoretyczne dotyczące reguł akcji, reguł wyjątków oraz zespołów reguł zamieszczono na początku sekcji poświęconych autorskim metodom (odpowiednio: Sekcja 5.2, Sekcja 5.4, Sekcja 5.5). Sekcja 5.3 rozwija wątek reguł akcji, korzystając z podstaw teoretycznych omówionych w Sekcji 5.2. W dalszej części zaprezentowano autorskie algorytmy: pokryciowy algorytm indukcji przeżyciowych reguł akcji, algorytm rekomendacji przeżyciowych reguł akcji, algorytm indukcji przeżyciowych reguł wyjątków oraz interpretowalny zespół reguł przeżyciowych, które nie tylko rozszerzające istniejące podejścia, lecz także oferujące nowe narzędzia do interpretowalnego modelowania danych cenzurowanych.

5.1 Indukcja reguł

Reguły logiczne, dzięki interpretowalności i zdolności do modelowania złożonych zależności, znajdują zastosowanie w analizie przeżycia [102]. Indukcja reguł to metoda uczenia maszynowego polegająca na generowaniu reguł decyzyjnych na podstawie danych uczących [101]. W odróżnieniu od złożonych modeli, takich jak głębokie sieci neuronowe, których wewnętrzne reprezentacje pozostają nieinterpretowalne dla użytkownika, reguły logiczne zapewniają pełną transparentność procesu decyzyjnego poprzez formalne warunki typu „jeśli-to". Ma to znaczenie w obszarach takich jak diagnostyka medyczna czy systemy predykcyjnego utrzymania ruchu, gdzie decyzje oparte na modelach powinny umożliwiać specjalistom weryfikację podstawy predykcji [86].

Algorytm 1 Pokryciowy algorytm indukcji reguł

Wejście: D – zbiór danych
Wyjście: R – zbiór reguł

$R \leftarrow \varnothing$
$D_u \leftarrow D$ ▷ zbiór przykładów niepokrytych jeszcze przez żadną regułę
while $D_u \neq \varnothing$ **do**
    $r \leftarrow \textsc{GenerujRegułę}(D, D_u)$
    $D_u \leftarrow D_u \setminus \textsc{Pokrycie}(r, D_u)$
    $R \leftarrow R \cup \{r\}$
end while
return R

W niniejszej pracy przyjęto strategię sekwencyjnego pokrywania (ang. *separate-and-conquer*) jako metodę indukcji reguł. Polega ona na iteracyjnym generowaniu reguł pokrywających kolejne podzbiory danych [15]. Każda nowa reguła pokrywa część jeszcze niepokrytych przykładów, a proces trwa do momentu pokrycia całego zbioru lub niespełnienia kryteriów jakości. Zespoły reguł rozwijają to podejście, łącząc predykcje z wielu zbiorów reguł w celu zwiększenia dokładności przy jednoczesnym zachowaniu interpretowalności. Kolejne sekcje zaprezentują autorskie algorytmy wykorzystujące te strategie do analizy danych cenzurowanych. Przebieg pokryciowego algorytmu indukcji reguł przedstawiono w Algorytmie 1.

Reguły decyzyjne umożliwiają reprezentację wiedzy w formie interpretowalnych reguł implikacyjnych [101]. W analizie niezawodności oraz w predykcyjnym utrzymaniu ruchu służą do modelowania zależności w danych cenzurowanych [22]. Wyróżnia się trzy podstawowe typy reguł: klasyfikacyjne, akcji oraz wyjątków, różniące się celem, strukturą i zastosowaniem. W niniejszej sekcji omówiono reguły klasyfikacyjne, natomiast wprowadzenia teoretyczne do reguł akcji i reguł wyjątków przedstawiono odpowiednio w sekcjach 5.2 i 5.4, a do zespołów reguł w Sekcji 5.5.

W dalszej części rozdziału stosowane jest następujące nazewnictwo: *cecha* (*atrybut*) oznacza pojedynczą zmienną opisującą obiekt, a *wektor cech* $\mathbf{x} = (x_1, \ldots, x_p)$ zawiera wartości cech danej obserwacji. *Warunek elementarny* w_i rozumiany jest jako predykat zdefiniowany na cechach $\mathbf{x}$, przy czym przesłanka reguły ma postać koniunkcji $w_1 \wedge \cdots \wedge w_n$. *Zmienna decyzyjna* y przyjmuje wartości etykiet klas $c_k \in C$. W częściach dotyczących analizy przeżycia dodatkowo wykorzystywane są *czas przeżycia* T oraz *status cenzurowania* $\delta \in \{0, 1\}$.

Reguły klasyfikacyjne Reguły klasyfikacyjne przypisują obserwacje do jednej z wcześniej zdefiniowanych klas na podstawie spełnienia warunków logicznych, zapewniając prostotę i interpretowalność wymaganą w systemach diagnostycznych [110]. Formalnie, reguła klasyfikacyjna r definiowana jest jako:

$$\text{jeśli } w_1 \wedge w_2 \wedge \cdots \wedge w_n \text{ to } y = c_k, \tag{5.1}$$

gdzie:

- $\mathbf{x} = (x_1, x_2, \ldots, x_p)$ — wektor wartości cech opisujących obserwację (np. parametry operacyjne maszyny: temperatura, ciśnienie, wibracje),
- w_i — warunek elementarny na cechach $\mathbf{x}$; przesłanka reguły to koniunkcja $w_1 \wedge w_2 \wedge \cdots \wedge w_n$ (np. $x_1 > a \ \wedge x_2 \leqslant b$), gdzie $i = 1, \ldots, n$,
- n — liczba warunków w przesłance reguły,
- y — zmienna decyzyjna wskazująca klasę,
- $c_k \in \{c_1, c_2, \ldots, c_m\}$ — klasa, np. „sprawny” lub „uszkodzony”.

Jakość reguł ocenia się za pomocą miar takich jak precyzja (ang. *precision*), czyli stosunek poprawnych predykcji do wszystkich obserwacji spełniających przesłankę reguły ($w_1 \wedge \cdots \wedge w_n$), pokrycie (ang. *coverage*), czyli odsetek obserwacji spełniających przesłankę reguły, oraz wsparcie (ang. *support*), wskazujące liczbę obserwacji spełniających zarówno przesłankę reguły, jak i $y = c_k$ [111]. Wysoka precyzja oznacza niski odsetek błędnych predykcji wśród przykładów spełniających przesłankę reguły, a duże pokrycie wskazuje, że reguła ma zastosowanie do znacznej części zbioru danych. Zwiększenie pokrycia często prowadzi jednak do obniżenia precyzji, ponieważ bardziej ogólne reguły mają tendencję do obejmowania przykładów z różnych klas.

Reguły klasyfikacyjne znajdują zastosowanie w prognozowaniu stanu technicznego maszyn i urządzeń. Przykładowa reguła „jeżeli temperatura $>75\,^{\circ}\mathrm{C}$ oraz wibracje > 4 m/s^2, to stan = uszkodzony” umożliwia jednoznaczną klasyfikację stanu maszyny, wspierając proces podejmowania decyzji konserwacyjnych [5]. Algorytmy indukcji reguł, takie jak CN2, oparte na strategii sekwencyjnego pokrywania, konstruują zbiory reguł o ograniczonej liczności i długości reguł poprzez usuwanie ze zbioru uczącego przykładów spełniających przesłankę aktualnie indukowanej reguły (tj. przykładów przez nią pokrywanych) [112]. Z kolei algorytm C4.5 przekształca drzewa decyzyjne w zbiory reguł, zwiększając możliwości modelowania złożonych zależności [110].

Literatura przedmiotu wskazuje na liczne zastosowania reguł klasyfikacyjnych. Tyagi i Sharma [113] zaproponowali metodę szacowania niezawodności systemów opartych na komponentach (ang. *Component-Based Software Systems*, CBSS) z wykorzystaniem reguł i logiki rozmytej, uwzględniając m.in. profil operacyjny oraz możliwość ponownego użycia komponentów. Wyniki eksperymentów potwierdziły skuteczność tej metody w modelowaniu warunków eksploatacyjnych. Avritzer i in. [114] przedstawili podejście do testowania niezawodności oparte na regułach, stosowane w monitorowaniu dużych systemów czasu rzeczywistego (np. oprogramowania przemysłowego), co podkreśla użyteczność modeli regułowych w zapewnianiu stabilności systemów.

Wariant przeżyciowy reguł klasyfikacyjnych dostosowuje się do modelowania danych cenzurowanych, umożliwiając predykcję czasu do zdarzenia, np. awarii lub zgonu pacjenta. Sikora i in. [115] zastosowali je do analizy danych pacjentów po przeszczepie szpiku kostnego, łącząc reguły z estymatorami Kaplana-Meiera,

co pozwoliło na identyfikację czynników przeżycia i tworzenie interpretowalnych wzorców. Wróbel i Sikora [116] rozwinęli metodę opartą na strategii *separate-and-conquer* z ważeniem danych cenzurowanych, osiągając wyższą dokładność predykcji niż klasyczne estymatory. Metody indukcji reguł przeżyciowych (akcji, wyjątków oraz zespołów), stanowiące wkład niniejszej pracy, zostaną szczegółowo omówione w dalszej części.

5.2 Pokryciowy algorytm indukcji przeżyciowych reguł akcji

W tej sekcji przedstawiono pokryciowy algorytm indukcji przeżyciowych reguł akcji, opisany w pracy *Separate-and-conquer Survival Action Rule Learning* [117]. Metoda ta stanowi podejście do indukcji przeżyciowych reguł akcji, umożliwiając identyfikację zmian wartości atrybutów prowadzących do określonej modyfikacji krzywej przeżycia w populacji pokrytej regułą. Jej opracowanie było motywowane potrzebą połączenie eksploracji akcji z analizą danych cenzurowanych oraz zaproponowania interpretowalnej procedury porównywania krzywych przeżycia grup źródłowych i docelowych. W niniejszej sekcji najpierw omówiono podstawy reguł akcji, a następnie przedstawiono założenia, sposób reprezentacji reguł oraz procedurę działania algorytmu.

5.2.1 Reguły akcji

Reguły akcji rozszerzają funkcjonalność reguł klasyfikacyjnych, wskazując konkretne zmiany wartości atrybutów prowadzące do przejścia z klasy źródłowej c_Z na docelową c_D. Dzięki temu możliwe jest formułowanie operacyjnych zaleceń dotyczących modyfikacji atrybutów, np. w systemach predykcyjnego utrzymania ruchu [18]. Niech $A = \{a_1, \ldots, a_p\}$ oznacza zbiór atrybutów, a dla każdego $a_i \in A$ niech $\mathcal{D}(a_i)$ będzie zbiorem dopuszczalnych wartości atrybutu a_i. Niech C oznacza zbiór klas decyzyjnych. Wówczas $w_{Zi}, w_{Di} \in \mathcal{D}(a_i)$ dla $i = 1, \ldots, n$, przy czym $n \leqslant p$, a $c_Z, c_D \in C$. Formalnie, regułę akcji zapisuje się jako:

$$\begin{aligned} &\textit{jeśli } (a_1 : w_{Z1} \to w_{D1}) \wedge (a_2 : w_{Z2} \to w_{D2}) \wedge \cdots \wedge (a_n : w_{Zn} \to w_{Dn}) \\ &\quad \textit{to } c_Z \to c_D, \end{aligned} \tag{5.2}$$

gdzie:

- $(a_i : w_{Zi} \to w_{Di})$ — elementarna akcja zmieniająca wartość atrybutu a_i z wartości źródłowej w_{Zi} na docelową w_{Di},
- $c_Z \to c_D$ — zmiana klasy, np. z „uszkodzony” na „sprawny”.

Regułę akcji można postrzegać jako kompozycję dwóch reguł: źródłowej (dla c_Z) i docelowej (dla c_D). Wyróżnia się trzy typy akcji elementarnych [118]: zmieniającą, $(a_i : w_{Zi} \to w_{Di})$, gdzie $w_{Zi} \neq w_{Di}$; dowolną, $(a_i : w_{Zi} \to \text{dowolna})$, gdy nie precyzuje się wartości docelowej; podtrzymującą, $(a_i : w_{Zi} \to w_{Zi})$, oznaczającą zachowanie wartości atrybutu.

W predykcyjnym utrzymaniu ruchu reguły akcji mogą wskazywać korekty parametrów procesu (np. ciśnienia, temperatury, obciążenia), które w danych historycznych dla grupy pokrytej przez regułę współwystępowały z obniżeniem ryzyka (np. spadkiem hazardu). Przykład reguły:

$$\begin{aligned} &\textit{jeśli } (\textit{ciśnienie} \in (60, 70]\,\text{bar} \to \textit{ciśnienie} \in [50, 55]\,\text{bar}) \\ &\quad \wedge\ (\textit{temperatura} \in (85, 90]^{\circ}\text{C} \to \textit{temperatura} \in [75, 80]^{\circ}\text{C}) \\ &\quad \textit{to}\ \text{kategoria ryzyka awarii} : \text{wysoka} \to \text{umiarkowana} \end{aligned} \tag{5.3}$$

Przykład obrazuje zmiany parametrów procesu, które w danych historycznych dla obserwacji pokrytych regułą korelowały z niższym ryzykiem awarii [61].

W dalszej części pracy stosowany jest równoważny, uproszczony zapis:

$$\begin{aligned} &\textit{jeśli } (\textit{ciśnienie}, (60, 70] \to [50, 55]) \\ &\quad \wedge\ (\textit{temperatura}, (85, 90] \to [75, 80]) \\ &\quad \textit{to}\ (\text{kategoria ryzyka awarii}, (\text{wysoka} \to \text{umiarkowana})) \end{aligned} \tag{5.4}$$

Prace nad regułami akcji koncentrują się na formalnych procedurach ich indukcji oraz zastosowaniach rekomendacyjnych. Raś i Tsay [118] zaproponowali system DEAR, który konstruuje reguły akcji z par reguł decyzyjnych odpowiadających klasie źródłowej i docelowej, rozróżniając atrybuty stabilne i zmienne. Elementarne akcje mają postać $(a : w_Z \to w_D)$, a ich trafność weryfikowana jest empirycznie. Wersja drzewiasta systemu ogranicza przestrzeń poszukiwań poprzez wykorzystanie struktury drzewa decyzyjnego, co redukuje koszt generowania reguł [118]. Sikora i in. [119] przedstawili algorytm SCARI, który integruje indukcję reguł akcji

(przejścia $c_Z \to c_D$) z mechanizmem wyboru działań dla pojedynczych obiektów. Algorytm wyznacza zestaw zmian atrybutów maksymalizujący prawdopodobieństwo przejścia do klasy docelowej przy zadanych ograniczeniach (np. kosztowych lub technologicznych).

W analizie przeżycia reguły akcji mogą sugerować interwencje poprawiające rokowania, np. zmianę harmonogramu konserwacji w celu wydłużenia czasu pracy maszyny lub dostosowanie terapii w medycynie. Ich potencjał w tym obszarze pozostaje jednak niedostatecznie zbadany, co stanowi lukę badawczą.

5.2.2 Opis metody

Metoda opiera się na strategii *separate-and-conquer* i działa na zbiorze danych $D(A, T, \delta)$, gdzie A to zbiór atrybutów, T to czas obserwacji (czas do zdarzenia), a $\delta \in \{0, 1\}$ to status przeżycia (1 — zdarzenie, 0 — cenzurowanie). Każdorazowo konstruowana jest reguła pokrywająca część dotychczas niepokrytych przykładów, które następnie usuwa się z puli pozostałej do pokrycia. Proces trwa aż do momentu, gdy zbiór niepokrytych przykładów jest pusty lub gdy nie da się wyindukować reguły spełniającej zadane minimalne progi pokrycia i jakości.

Centralnym elementem metody jest generowanie przeżyciowych reguł akcji, których konkluzja zawiera estymator funkcji przeżycia (Kaplana-Meiera). Reguły te mają postać:

$$\begin{aligned} &\textit{jeśli } (a_1 : w_{Z1} \to w_{D1}) \wedge (a_2 : w_{Z2} \to w_{D2}) \wedge \cdots \wedge (a_n : w_{Zn} \to w_{Dn}) \\ &\quad \textit{to } \hat{S}_Z \to \hat{S}_D, \end{aligned} \tag{5.5}$$

gdzie $(a_i : w_{Zi} \to w_{Di})$ reprezentuje zmianę wartości atrybutu a_i z zakresu źródłowego w_{Zi} na zakres docelowy w_{Di}, a $\hat{S}_Z$ i $\hat{S}_D$ to estymatory funkcji przeżycia wyznaczone dla przykładów pokrywanych, odpowiednio, przez część źródłową i część docelową reguły.

Metoda składa się z dwóch głównych faz: specjalizacji (wzrostu) reguły oraz jej generalizacji (przycinania). W fazie specjalizacji, rozpoczynając od pustej przesłanki, warunki są dodawane iteracyjnie w celu zmaksymalizowania różnicy w krzywych przeżycia między grupą źródłową a docelową, mierzonej testem log-rank. W fazie przycinania warunki są usuwane lub modyfikowane w celu poprawy

jakość reguły. Nadmiernemu dopasowaniu zapobiega się, pozostawiając wyłącznie modyfikacje, które nie pogarszają statystyki log-rank, natomiast nadmiernej ogólności przeciwdziała się poprzez ograniczenie maksymalnego pokrycia ρ oraz dopuszczalnego udziału przykładów wspólnych dla części źródłowej i docelowej reguły ξ.

Algorytm 2 Algorytm indukcji reguł akcji dla danych cenzurowanych

Wejście:
  $D(A, T, \delta)$ — zbiór danych opisany atrybutami A, czasem obserwacji T i statusem przeżycia δ
  μ — minimalna liczba niepokrytych przykładów, które nowa reguła musi pokrywać
  ξ — maksymalny procent przykładów wspólnych dla obu części reguły
  ρ — maksymalne pokrycie reguły
  τ — typ reguły: lepsza | gorsza | dowolna
  $A_{stabilny}$ — zbiór atrybutów stabilnych z A
Wyjście: R — zbiór przeżyciowych reguł akcji
procedure IndukujPrzeżycioweRegułyAkcji($D, \mu, \xi, \rho, \tau, A_{stabilny}$)
    $R \leftarrow \varnothing$
    $D_u \leftarrow D$ ▷ zbiór przykładów niepokrytych
    while $D_u \neq \varnothing$ **do**
        $r \leftarrow$ Specjalizuj($D, D_u, \mu, \tau, \rho, A_{stabilny}$)
        $r \leftarrow$ Uogólnij(D, r, τ, ξ, ρ)
        if $r \neq \varnothing$ **then** $R \leftarrow R \cup \{r\}$
        $D_c \leftarrow$ PokrytePrzykłady(r, D) ▷ przykłady z D pokrywane przez r
        $D_u \leftarrow D_u \backslash D_c$
    end while
    return R
end procedure

Pseudokod procedury indukcji zbioru przeżyciowych reguł akcji został przedstawiono w Algorytmie 2. Na wejściu przekazywany jest zbiór danych $D(A, T, \delta)$. Dodatkowo wykorzystywany jest zbiór parametrów:

- μ — minimalna liczba przykładów niepokrytych, które musi pokrywać indukowana reguła;
- ρ — dopuszczalne maksymalne pokrycie reguły względem D;
- ξ — dopuszczalny maksymalny udział przykładów wspólnych dla części źródłowej i docelowej reguły;
- $\tau \in \{\text{lepsza}, \text{gorsza}, \text{dowolna}\}$ — preferowany kierunek zmiany wyniku przeżycia implikowany przez regułę.

Zbiór $A_{stabilny} \subseteq A$ wyznacza atrybuty stabilne (niepoddające się działaniu), które nie mogą być modyfikowane przez część akcyjną reguły.

Zadaniem procedury jest zbudowanie zbioru reguł R, w którym każda reguła ma postać:

$$\text{jeśli } \varphi \text{ to } \hat{S}_Z \rightarrow \hat{S}_D$$

gdzie $\hat{S}_Z$ i $\hat{S}_D$ oznaczają nieparametryczne estymatory funkcji przeżycia (Kaplana-Meiera) odpowiednio dla części źródłowej i docelowej. Algorytm utrzymuje zbiór przykładów niepokrytych D_u i iteracyjnie konstruuje nową regułę r poprzez sekwencyjne uszczegóławianie i uogólnianie przesłanki. Procedura *Specjalizuj* buduje koniunkcję akcji złożoną z par warunek–kontrwarunek ograniczonych do atrybutów niestabilnych, respektując progi μ i ρ. Natomiast procedura *Uogólnij* upraszcza powstałą przesłankę przy zachowaniu ograniczeń ξ, ρ oraz zgodności z kierunkiem τ (tj. relacji $\hat{S}_D$ względem $\hat{S}_Z$ zgodnej z wymaganym kierunkiem). Po akceptacji reguły do R, z D_u usuwa się wszystkie przykłady pokryte przez jej część źródłową. Pętla kończy się, gdy $D_u = \varnothing$ lub nie istnieje reguła spełniająca jednocześnie ograniczenia μ, ξ, ρ oraz wymóg kierunku τ.

Faza specjalizacji reguły jest realizowana przez procedurę *Specjalizuj*, przedstawioną w Algorytmie 3. Polega ona na zachłannym dodawaniu akcji do przesłanki reguły, rozpoczynając od pustej przesłanki. W każdej iteracji poszukiwany jest najlepszy warunek elementarny w_{najl_Z} dla części źródłowej reguły, a następnie najlepszy kontrwarunek w_{najl_D} dla części docelowej. Akcja utworzona z tych dwóch warunków jest dodawana do przesłanki reguły. Proces ten trwa do momentu, gdy nie można znaleźć nowego warunku w_{najl_Z}.

Najlepszy warunek elementarny w_{najl_Z} wybierany jest tak, aby po dodaniu do części źródłowej reguły, statystyka log-rank między przykładami pokrywanymi

Algorytm 3 Specjalizacja przeżyciowej reguły akcji

Wejście:
  $D(A, T, \delta)$ — zbiór danych
  D_u — zbiór przykładów niepokrytych
  μ — minimalna liczba przykładów, które nowa reguła musi pokrywać
  ρ — maksymalne pokrycie reguły
  τ — typ reguły
  $A_{stabilny}$ — zbiór atrybutów stabilnych
Wyjście: r — reguła akcji
procedure SPECJALIZUJ($D, D_u, \mu, \tau, \rho, A_{stabilny}$)
    $\varphi, \varphi_Z, \varphi_D \leftarrow \varnothing$ ▷ przesłanka, warunki źródłowe i docelowe
    $W_{sprawdzone} \leftarrow \varnothing$ ▷ już sprawdzone warunki
    repeat
        $w_{najl_Z} \leftarrow$ ZNAJDŹNAJLEPSZYWARUNEKELEMENTARNY($D, D_u, \varphi_Z, \mu, \rho, W_{sprawdzone}$)
        $W_{sprawdzone} \leftarrow W_{sprawdzone} \cup \{w_{najl_Z}\}$
        if $w_{najl_Z} = \varnothing$ **then continue**
        $a \leftarrow$ ATRYBUT(w_{najl_Z})
        $w_{najl_D} \leftarrow$ ZNAJDŹKONTRWARUNEK($D, \varphi_Z \wedge w_{najl_Z}, \varphi_D, w_{najl_Z}, \mu, \tau$)
        $akcja \leftarrow$ BUDUJAKCJĘ(w_{najl_Z}, w_{najl_D})
        $\varphi \leftarrow \varphi \wedge akcja$
        $\varphi_Z \leftarrow \varphi_Z \wedge w_{najl_Z}$
        $\varphi_D \leftarrow \varphi_D \wedge w_{najl_D}$
    until $(w_{najl_Z} = \varnothing)$
    Oblicz $\hat{S}_Z$ dla POKRYTEPRZYKŁADY(φ_Z, D) (Kaplan-Meier)
    Oblicz $\hat{S}_D$ dla POKRYTEPRZYKŁADY(φ_D, D) (Kaplan-Meier)
    return $r \equiv$ jeśli φ to $\hat{S}_Z \rightarrow \hat{S}_D$
end procedure

a niepokrywanymi przez regułę była jak największa. W przypadku kilku warunków dających tę samą wartość statystyki, wybierany jest ten, który maksymalnie zwiększa liczbę dotychczas niepokrywanych przykładów. Najlepszy kontrwarunek w_{najl_D} poszukiwany jest dla tego samego atrybutu co w_{najl_Z} i wybierany tak, aby statystyka log-rank między przykładami pokrywanymi przez część źródłową z dodanym w_{najl_Z} a przykładami pokrywanymi przez część docelową z dodanym w_{najl_D} była maksymalna.

Po fazie specjalizacji następuje faza przycinania reguły, realizowana przez funkcję *Uogólnij*. W tej fazie iteracyjnie usuwane są akcje z przesłanki reguły, sprawdzając, czy ich usunięcie poprawia lub nie pogarsza wartości statystyki log-rank między krzywymi przeżycia części źródłowej i docelowej. Dodatkowo rozważana jest możliwość zamiany akcji na akcję dowolną, czyli taką, w której po stronie docelowej nie narzuca się konkretnej wartości atrybutu — dopuszcza się dowolną wartość $((a_i : w_{Zi} \rightarrow \text{dowolna}))$, jeśli prowadzi to do dalszej poprawy jakości reguły. Proces przycinania trwa do momentu, gdy nie ma już akcji, których usunięcie lub zamiana na akcję dowolną poprawiałoby jakość reguły. Usuwanie akcji podlega dwóm ograniczeniom. Akcja nie jest usuwana, jeśli jej usunięcie spowodowałoby, że udział przykładów pokrytych przez jedną regułę przekroczyłby ρ lub jeśli udział wspólnych przykładów dla części źródłowej i docelowej reguły przekroczyłby ξ.

Metoda obsługuje braki w danych poprzez strategię ignorowania brakujących wartości. Podczas poszukiwania możliwych warunków, brakujące wartości są pomijane, a reguły budowane są wyłącznie na podstawie znanych wartości. Przykłady z brakującymi wartościami dla atrybutów występujących w regule nie są uwzględniane przy ocenie pokrycia reguły. Strategia ta nie wymaga dodatkowych kroków, takich jak interpolacja danych.

Proponowana metoda jest unikalna, ponieważ jako pierwsza umożliwia indukcję reguł akcji dla danych cenzurowanych, łącząc techniki eksploracji akcji z analizą przeżycia. W przeciwieństwie do istniejących metod, zaprojektowanych dla danych klasyfikacyjnych, metoda ta obsługuje dane cenzurowane już na etapie indukcji i oceny jakości reguł. Umożliwia to formułowanie interpretowalnych rekomendacji zmian wartości atrybutów o kontrolowanym kierunku wpływu na funkcję przeżycia w zbiorach przykładów pokrywanych przez reguły.

5.2.3 Ilustracja działania metody

Niniejsza ilustracja przedstawia zastosowanie omawianej metody na zbiorze danych *Maintenance*, dotyczącym konserwacji urządzeń. Zbiór obejmuje 1000 przykładów opisanych trzema zmiennymi objaśniającymi:

- *pressureInd* — przepływ cieczy,
- *moistureInd* — względna wilgotność,
- *temperatureInd* — temperatura.

Ponadto zbiór zawiera dwa atrybuty związane z przeżyciem:

- *lifetime* — liczba tygodni, przez które maszyna była aktywna,
- *broken* — informacja o awarii maszyny w trakcie omawianego okresu aktywności.

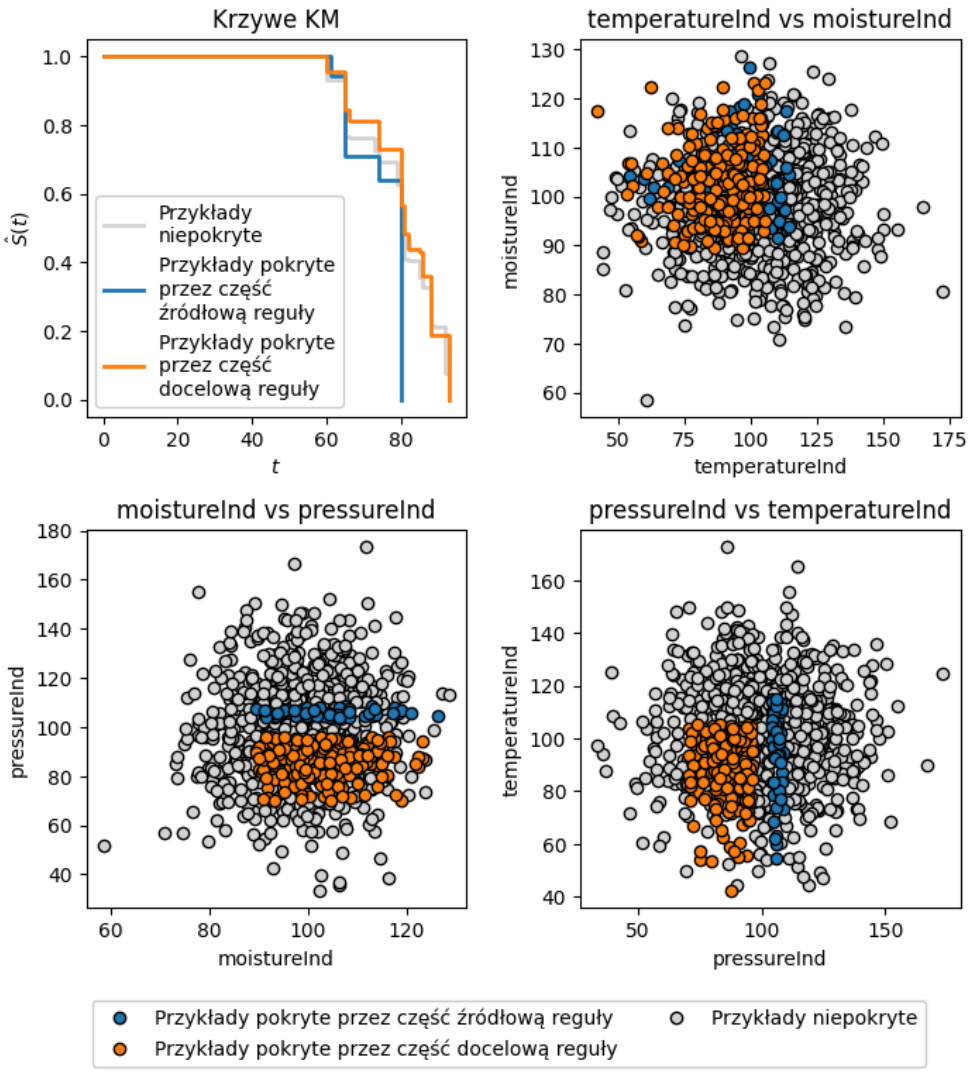
Celem analizy jest wygenerowanie przeżyciowych reguł akcji, które modyfikują krzywą przeżycia maszyn poprzez rekomendowanie zmian wartości atrybutów.

Na Rysunku 5.1 zaprezentowano wpływ pokrycia reguły na estymowaną funkcję przeżycia oraz rozkład przykładów w przestrzeni cech.

Przykład indukcji reguł

Poniżej przedstawiono szczegółowy przebieg indukcji reguł w trzeciej iteracji algorytmu. Iteracja rozumiana jest jako jeden cykl procesu *separate-and-conquer*, w którym generowana jest pojedyncza reguła pokrywająca część niepokrytych dotychczas przykładów. Trzecia iteracja została wybrana, ponieważ najlepiej ilustruje pełną funkcjonalność algorytmu, w tym obsługę warunków złożonych.

Inicjalizacja Na początku trzeciej iteracji w zbiorze pozostaje 937 przykładów niepokrytych przez żadną z wcześniej wygenerowanych reguł. Zadaniem algorytmu jest zidentyfikowanie reguły, której krzywa przeżycia przewyższa krzywą reguły źródłowej, tzn. zapewnia wyższe estymowane prawdopodobieństwo przeżycia dla przykładów spełniających tę regułę.



Rysunek 5.1: Krzywe przeżycia ilustrujące wpływ wygenerowanej reguły akcji na czas bezawaryjnej pracy maszyn. Niebieska krzywa reprezentuje przykłady spełniające warunki części źródłowej reguły, pomarańczowa — części docelowej, natomiast szara krzywa odnosi się do przykładów niepokrywanych przez tę regułę. Dodatkowo, trzy wykresy rozrzutu przedstawiają rozkład przykładów w przestrzeni cech (dla każdej pary atrybutów) w podziale na te trzy grupy. Rozważana reguła ma postać: jeśli $(\mathit{pressureInd}, (-\infty, 103.9) \rightarrow (-\infty, 81.0)) \wedge (\mathit{temperatureInd}, (-\infty, 115.4) \rightarrow (-\infty, 105.6)) \wedge (\mathit{moistureInd}, (-\infty, 89.5) \rightarrow (-\infty, 89.4))$ to $\hat{S}_Z \rightarrow \hat{S}_D$

Faza specjalizacji Specjalizacja realizowana jest zachłannie względem statystyki log-rank — dodawana akcja musi zwiększać różnicę (mierzoną testem log-rank) między krzywymi przeżycia części źródłowej i docelowej przy zachowaniu ograniczeń pokrycia i udziału przykładów wspólnych. Kontrwarunek dobierany jest dla tego samego atrybutu co warunek źródłowy.

1. **Znajdowanie najlepszego warunku elementarnego dla reguły źródłowej**: Algorytm identyfikuje warunek $(\mathit{pressureInd}, (-\infty, 103.7))$ jako najlepszy warunek elementarny dla reguły źródłowej. Kryterium wyboru była najwyższa wartości statystyki log-rank, równa 3.9, obliczona dla porównania krzywych przeżycia przykładów spełniających i niespełniających ten warunek.

2. **Dodawanie warunku do przesłanki reguły**: Zidentyfikowany warunek zostaje dodany do przesłanki reguły:

$$(\mathit{pressureInd}, (-\infty, 103.7))$$

3. **Znajdowanie najlepszego kontrwarunku dla reguły docelowej**: W odpowiedzi na ustaloną regułę źródłową algorytm wskazuje kontrwarunek $(\mathit{pressureInd}, (-\infty, 96.1))$ jako najlepszy. Wartość statystyki log-rank, obliczona dla porównania krzywych przeżycia przykładów spełniających część źródłową z przykładami spełniającymi część docelową reguły, wynosi 4.1.

4. **Tworzenie i dodawanie akcji do przesłanki reguły**: Akcja utworzona z reguły źródłowej i kontrwarunku jest włączana do przesłanki:

$$(\mathit{pressureInd}, (-\infty, 103.7) \rightarrow (-\infty, 96.1))$$

   Warunek źródłowy $(\mathit{pressureInd}, (-\infty, 103.7))$ stanowi nadzbiór warunku docelowego $(\mathit{pressureInd}, (-\infty, 96.1))$, ponieważ każdy przykład spełniający warunek docelowy spełnia również warunek źródłowy, natomiast nie zachodzi implikacja odwrotna, czyli:

$$\{x \in \mathbb{R} : x < 96.1\} \subset \{x \in \mathbb{R} : x < 103.7\}.$$

   Przejście $(\mathit{pressureInd}, (-\infty, 103.7) \rightarrow (-\infty, 96.1))$ należy interpretować jako zawężenie przesłanki z szerszego zbioru przypadków do węższego podzbioru określonego bardziej restrykcyjnym progiem.

5. **Iteracyjna modyfikacja warunków**: Algorytm kontynuuje specjalizację reguły, dodając kolejne warunki. Najlepszy kolejny warunek elementarny to $(\mathit{temperatureInd}, (-\infty, 115.4)$ (wartość log-rank 9.8). Następnie, najlepszy kontrwarunek został zidentyfikowany jako $(\mathit{temperatureInd}, (-\infty, 105.6))$ (wartość log-rank 13.0). Zestawienie warunków daje kolejną akcję, która zostaje włączona do przesłanki:

$$\begin{aligned} &(\mathit{pressureInd}, (-\infty, 103.7) \rightarrow (-\infty, 96.1)) \\ \wedge\ &(\mathit{temperatureInd}, (-\infty, 115.4) \rightarrow (-\infty, 105.6)) \end{aligned}$$

6. **Kontynuacja iteracyjnej modyfikacji warunków**: Proces jest powtarzany, co prowadzi do uzyskania kolejnych akcji:
   a) $(\mathit{moistureInd}, [89.5, +\infty) \rightarrow [89.4, +\infty))$ z wartością log-rank wynoszącą 13.7 dla reguły źródłowej i 19.0 dla reguły docelowej,
   b) $(\mathit{pressureInd}, (-\infty, 108.6) \rightarrow [70.1, +\infty))$ z wartością log-rank wynoszącą 15.9 dla reguły źródłowej i 26.1 dla reguły docelowej,
   c) $(\mathit{pressureInd}, [103.9, +\infty) \rightarrow [81.0, +\infty))$ z wartością log-rank wynoszącą 17.0 dla reguły źródłowej i 26.7 dla reguły docelowej,
   d) $(\mathit{pressureInd}, [104.7, +\infty) \rightarrow (-\infty, 96.1))$ z wartością log-rank wynoszącą 17.0 dla reguły źródłowej i 25.9 dla reguły docelowej.

   W efekcie przesłanka przyjmuje postać:

$$\begin{aligned} &(\mathit{pressureInd}, [104.7, 108.6) \rightarrow [81.0, 96.1)) \\ \wedge\ &(\mathit{temperatureInd}, (-\infty, 115.4) \rightarrow (-\infty, 105.6) \\ \wedge\ &(\mathit{moistureInd}, (-\infty, 89.5) \rightarrow (-\infty, 89.4) \end{aligned}$$

7. **Zakończenie specjalizacji**: Próba dodania kolejnego warunku elementarnego nie spełnia kryteriów jakości (minimalnej wartości statystyki log-rank lub minimalnego pokrycia), co oznacza zakończenie fazy specjalizacji w tej iteracji. Otrzymana reguła ma postać:

$$\begin{aligned} \text{jeśli } &(\mathit{pressureInd}, [104.7, 108.6)) \rightarrow [81.0, 96.1)) \\ \wedge\ &(\mathit{temperatureInd}, (-\infty, 115.4) \rightarrow (-\infty, 105.6) \\ \wedge\ &(\mathit{moistureInd}, (-\infty, 89.5) \rightarrow (-\infty, 89.4)) \\ \text{to } &\hat{S}_Z \rightarrow \hat{S}_D \end{aligned}$$

Faza uogólniania Uogólnianie polega na kontrolowanym upraszczaniu przesłanki — akcje usuwa się lub łagodzi, o ile nie pogarsza to statystyki log-rank, przy zachowaniu ograniczeń pokrycia oraz udziału przykładów wspólnych.

1. **Uogólnianie iteracja 1**:
   - **Akcje oznaczone do usunięcia**: Akcje
   $$(pressureInd, [103.7, +\infty) \rightarrow (-\infty, 96.1))$$
   $$(pressureInd, [104.7, +\infty) \rightarrow (-\infty, 96.0))$$
   zostały zidentyfikowane jako nieistotne statystycznie (ich usunięcie nie powoduje pogorszenia statystyki log-rank) i przeznaczone do usunięcia w celu uproszczenia reguły.
   - **Akcje oznaczone do złagodzenia**: Akcje
   $$(pressureInd, [103.7, +\infty) \rightarrow (-\infty, 96.1))$$
   $$(pressureInd, (-\infty, 108.6) \rightarrow [70.1, +\infty))$$
   zostały zakwalifikowane do złagodzenia, rozumianego jako poszerzenie odpowiedniego przedziału wartości (po stronie źródłowej lub docelowej) poprzez przesunięcie granicy progu tak, aby zwiększyć pokrycie bez istotnego spadku wartości statystyki log-rank, przy jednoczesnym zachowaniu statystycznie istotnej różnicy między krzywymi przeżycia.
   - **Wynik**: Mimo że wiele akcji zostało oznaczonych do modyfikacji, metoda uogólniania modyfikuje tylko jedną z nich, zgodnie z określonymi poniżej zasadami.
     - Akcja nie może być jednocześnie usunięta i złagodzona. Oznacza się ją do usunięcia, jeżeli jej eliminacja nie obniża wartości statystyki log-rank i nie narusza minimalnego pokrycia. Oznacza się ją do złagodzenia, jeżeli całkowite usunięcie pogarsza miarę jakości, natomiast poszerzenie przedziału zachowuje wymagane progi jakości i zwiększa pokrycie.
     - Akcje mogą być wzajemnie powiązane — usunięcie jednej akcji może wpływać na statystyczną istotność pozostałych. Algorytm uwzględnia te zależności, sprawdzając wpływ każdej modyfikacji na całościową jakość reguły.

– W każdej iteracji algorytm wybiera jedną modyfikację (usunięcie lub złagodzenie akcji), która najbardziej poprawia wartość statystyki log-rank między krzywymi przeżycia części źródłowej i docelowej reguły.

W rezultacie usunięto akcję dotyczącą atrybutu *pressureInd*, redukując liczbę akcji w regule bez pogorszenia wartości statystyki log-rank i przy zachowaniu minimalnego pokrycia.

2. **Uogólnianie iteracja 2**:
   - **Akcje oznaczone do złagodzenia**: Akcja
   $$(pressureInd, [103.9, +\infty) \rightarrow [81.0, +\infty))$$
   została wytypowana do złagodzenia, ponieważ całkowite usunięcie pogorszyłoby miarę jakości, natomiast poszerzenie odpowiednich przedziałów spełnia progi jakości i zwiększa pokrycie.
   - **Wynik**: W rezultacie akcja
   $$(pressureInd, [103.9, +\infty) \rightarrow [81.0, +\infty))$$
   została złagodzona do
   $$(pressureInd, [103.9, +\infty) \rightarrow dowolna)$$
   co zmniejszyło złożoność i poprawiło możliwość generalizacji reguły.
3. **Uogólnianie iteracja 3**:
   - **Brak akcji do usunięcia**: W tej iteracji algorytm dokonał oceny pozostałych akcji, lecz nie znalazł dalszych redundancji ani nieefektywności. Wszystkie istniejące akcje miały znaczący wkład w poprawę krzywej przeżycia.
4. **Zakończenie uogólniania**: Proces uogólniania zakończono, ponieważ dalsze usuwanie akcji skutkowałoby pogorszeniem jakości reguły. Zachowano równowagę między specyficznością a ogólnością, przy utrzymaniu wysokiego pokrycia oraz istotnego wpływu na krzywe przeżycia.

Uzyskana reguła Ostateczna reguła po specjalizacji i uogólnianiu ma postać:

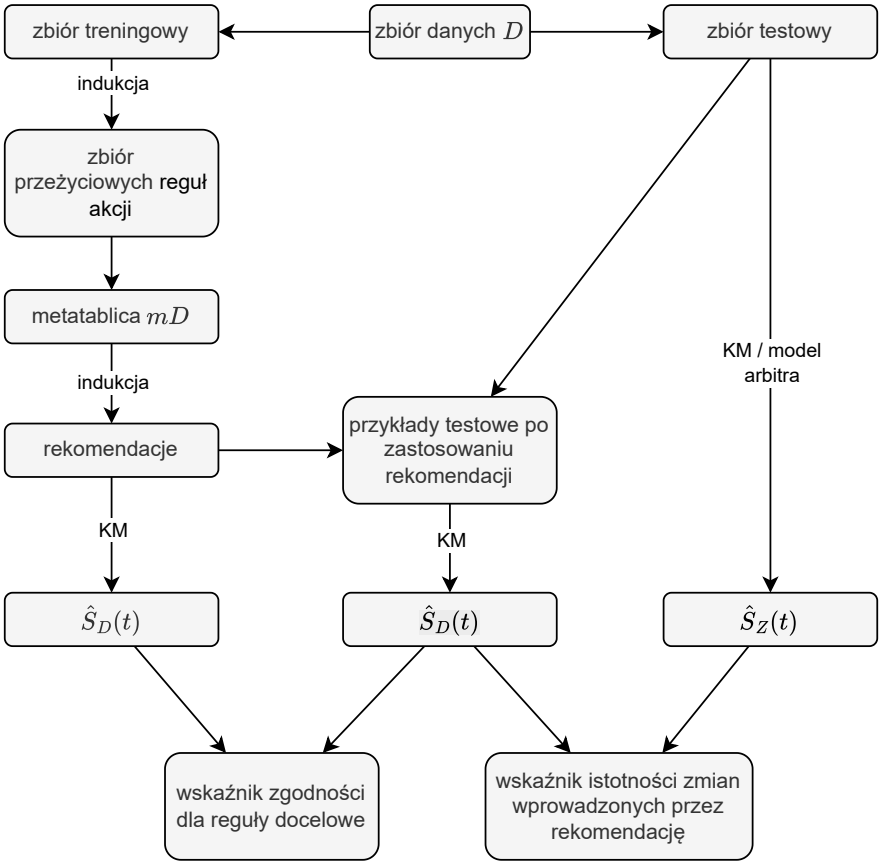
$$\begin{aligned}&\text{jeśli } (\mathit{pressureInd}, (-\infty, 103.9) \rightarrow (-\infty, 81.0))\\ &\quad \wedge\ (\mathit{temperatureInd}, (-\infty, 115.4) \rightarrow (-\infty, 105.6))\\ &\quad \wedge\ (\mathit{moistureInd}, (-\infty, 89.5) \rightarrow (-\infty, 89.4))\\ &\text{to } \hat{S}_Z \rightarrow \hat{S}_D\end{aligned}$$

Trzecia iteracja wygenerowała regułę pokrywającą 42 przykłady, zmniejszając liczbę niepokrytych przykładów z 937 do 895. Test log-rank potwierdził statystycznie istotną różnicę między krzywymi przeżycia przed i po zastosowaniu tej reguły. Wygenerowana reguła sugeruje obniżenie wartości ciśnienia (*pressureInd*), temperatury (*temperatureInd*) i wilgotności (*moistureInd*) w celu poprawy niezawodności.

Strategia *separate-and-conquer* kontynuuje proces indukcji do momentu, gdy wszystkie przykłady zostaną pokryte lub gdy nie będzie możliwe wygenerowanie kolejnych reguł spełniających kryteria jakości. W pierwszych trzech iteracjach wygenerowano trzy reguły, które łącznie pokryły 105 przykładów.

5.3 Algorytm rekomendacji przeżyciowych reguł akcji

Algorytm rekomendacji opisany w pracy *Recommendation Algorithm Based on Survival Action Rules* [120] stanowi rozszerzenie pokryciowego algorytmu indukcji przeżyciowych reguł akcji opisanego w Sekcji 5.2. Podczas gdy algorytm pokryciowy koncentruje się na generowaniu przeżyciowych reguł akcji na podstawie danych treningowych z wykorzystaniem strategii sekwencyjnego pokrywania, algorytm rekomendacji traktuje wygenerowane reguły jako dane wejściowe do procesu tworzenia spersonalizowanych zaleceń dla nowych obiektów. Oba algorytmy składają się na dwuetapową procedurę. W pierwszym etapie przeprowadzana jest indukcja wzorców w postaci reguł akcji, w drugim etapie reguły te wykorzystywane są w procesie generowania zaleceń dotyczących działań ukierunkowanych na poprawę estymowanej funkcji przeżycia. Schemat działania algorytmu przedstawiono na rysunku 5.2.



Rysunek 5.2: Diagram przedstawiający zasadę działania algorytmu rekomendacyjnego opartego na przeżyciowych regułach akcji.

5.3.1 Opis metody

Pojedynczy obiekt może spełniać przesłanki wielu przeżyciowych reguł akcji, co prowadzi do konfliktów. Konflikty te obejmują przede wszystkim wielokrotne pokrycie (jednoczesna aktywacja kilku reguł oferujących różne zbiory akcji lub różne docelowe krzywe przeżycia $\hat{S}_D$) oraz kolizje akcji (rozłączne wartości dla tego samego atrybutu). Na etapie indukcji konflikty są częściowo ograniczane poprzez strategię *separate-and-conquer*, progi ρ i ξ oraz niedopuszczanie modyfikowania atrybutów stabilnych. Rozstrzyganie konfliktów zapewnia metoda rekomendacyjna opisana w niniejszej sekcji, oparta na indukcji meta-reguł akcji, generowanych na podstawie meta-tablicy. Podejście to umożliwia ocenę alternatywnych zestawów działań i wybór rozwiązania maksymalizującego zmianę krzywej przeżycia w zadanym kierunku.

Meta-tablica mD jest elementem metody umożliwiającym generowanie rekomendacji dla nowych obiektów poprzez reprezentację danych w postaci zbioru meta-przykładów. Każdy meta-przykład w mD odzwierciedla unikalną kombinację wartości atrybutów, określoną na podstawie reguł akcji uzyskanych w procesie indukcji. Dla atrybutów symbolicznych meta-atrybuty ma odpowiadają bezpośrednio wartościom atrybutów ze zbioru wejściowego D, na przykład „lokalizacja = sekcja A”. W przypadku atrybutów numerycznych zbiór wartości danego atrybutu dzielony jest na przedziały wyznaczone przez wartości progowe występujące w regułach akcji. Przykładowo, dla atrybutu a_1 mogą to być przedziały: $(-\infty, 50], (50, 70], (70, +\infty)$. Każdy meta-przykład w mD jest powiązany z podzbiorem przykładów ze zbioru treningowego D, co umożliwia estymację krzywej przeżycia $\hat{S}$ dla danego meta-przykładu za pomocą estymatora Kaplana-Meiera. W ten sposób meta-tablica pełni rolę mechanizmu dyskretyzacji i generalizacji danych treningowych, co znacząco przyspiesza proces generowania rekomendacji i zwiększa jego efektywność obliczeniową.

Algorytm rekomendacji rozwiązuje problem konfliktów wynikających z pokrywania obiektu przez wiele reguł akcji poprzez ocenę proponowanych działań oraz wybór zestawu modyfikacji atrybutów maksymalizującego zmianę krzywej przeżycia. Proces ten opiera się na meta-tablicy mD i obejmuje cztery etapy:

1. **Identyfikacja meta-przykładu**: Obiekt x, opisany zbiorem wartości atrybutów $A = \{a_1, a_2, \ldots, a_m\}$, jest mapowany na odpowiadający mu meta-

przykład $mx \in mD$. Mapowanie polega na przypisaniu wartości atrybutów x do odpowiednich przedziałów zdefiniowanych w meta-tablicy. Dla atrybutów symbolicznych wartości są porównywane bezpośrednio z wartościami meta-atrybutów (np. „typ usterki = mechaniczna" w x odpowiada tej samej wartości w mx). Natomiast dla atrybutów numerycznych wartości są przypisywane do przedziałów wyznaczonych przez reguły akcji (np. jeśli $a_1 = 65$ dla temperatury, a meta-tablica definiuje przedziały $(-\infty, 50], (50, 70], (70, +\infty)$, to a_1 trafia do przedziału $(50, 70]$). Meta-przykład mx pełni rolę reprezentanta podzbioru przestrzeni atrybutów, co pozwala na uogólnienie analizy na podobne obiekty.

2. **Estymacja krzywej przeżycia**: Obliczana jest krzywa przeżycia $\hat{S}_Z$ dla obiektu x, odzwierciedlająca jego stan przed wprowadzeniem jakichkolwiek zmian. Domyślnie $\hat{S}_Z$ estymowana jest metodą Kaplana-Meiera na podstawie wszystkich przykładów treningowych ze zbioru D, które są pokrywane przez meta-przykład mx. Aby estymacja była statystycznie istotna, meta-przykład musi obejmować wystarczającą liczbę przykładów. Jeśli liczba ta jest zbyt mała (mniejsza niż zadany próg μ), co obniża wiarygodność estymacji krzywej przeżycia, stosuje się model arbitra. XGBoost z funkcją straty Coxa, trenowany na całym zbiorze D, generuje przewidywaną krzywą $\hat{S}_Z$ dla x. Estymator Kaplana-Meiera stosuje się, gdy meta-przykład zawiera wystarczającą liczbę próbek ($\geqslant \mu$), natomiast model arbitra — gdy liczba przykładów jest niewystarczająca.

3. **Indukcja meta-reguły akcji**: W meta-tablicy mD indukowana jest meta-reguła akcji r_D, której konkluzja $\hat{S}_D$ reprezentuje docelową krzywą przeżycia po zastosowaniu rekomendowanych zmian. Celem jest znalezienie reguły, dla której $\hat{S}_D$ maksymalnie różni się od $\hat{S}_Z$, mierzone testem Kołmogorowa-Smirnowa dla dwóch próbek. Test ten ocenia maksymalną odległość między empirycznymi funkcjami dystrybuanty przeżycia, umożliwiając porównanie krzywych przeżycia bez przyjmowania założeń o rozkładzie czasu do zdarzenia. Algorytm wykorzystuje strategię wspinaczki (ang. *hill climbing*), iteracyjnie przeszukując przestrzeń meta-przykładów w mD, aby zidentyfikować zestaw zmian atrybutów prowadzący do pożądanej modyfikacji krzywej przeżycia. Minimalizacja liczby zmian atrybutów pozwala na po-

tencjalne ograniczenie kosztów wdrożenia zaleceń — zamiast modyfikować wiele parametrów maszyny, algorytm może zasugerować tylko jedną zmianę (np. obniżenie temperatury), jeśli pojedyncza modyfikacja wystarcza do osiągnięcia pożądanego efektu. Proces indukcji działa analogicznie do specjalizacji reguł w algorytmie głównym, ale operuje na meta-przykładach zamiast na surowych danych, co zmniejsza złożoność obliczeniową.

4. **Generowanie rekomendacji**: Na podstawie indukowanej reguły r_D generowane są konkretne zalecenia dla obiektu x. Rekomendacje przyjmują formę zestawu działań zmieniających wartości atrybutów ze stanu źródłowego na docelowy, zgodnie z przesłanką reguły r_D. Przykładowo, jeśli r_D sugeruje zmianę ($a_1 > 70\,^\circ\mathrm{C} \rightarrow a_1 \leqslant 70\,^\circ\mathrm{C}$), zalecenie brzmi: „zmniejsz temperaturę z wartości powyżej $70\,^\circ$C do wartości równej lub mniejszej niż $70\,^\circ$C". Działania są wyprowadzane z różnic między mx (reprezentującym obecny stan x) a meta-przykładem (lub grupą sąsiadujących meta-przykładów) odpowiadającym $\hat{S}_D$. Jeśli kilka meta-przykładów jednocześnie prowadzi do lepszej $\hat{S}_D$, algorytm może łączyć ich zakresy wartości, o ile są one sąsiadujące i poprawiają jakość krzywej przeżycia w porównaniu do pojedynczych meta-przykładów.

Strategia wspinaczki ogranicza złożoność obliczeniową poprzez przeszukiwanie lokalnego otoczenia bieżącego meta-przykładu zamiast pełnej przestrzeni meta-przykładów, co ma znaczenie w przypadku dużych meta-tablic zawierających wiele kombinacji meta-przykładów. Rozpoczynając od meta-przykładu mx, algorytm iteracyjnie eksploruje sąsiednie meta-przykłady, oceniając ich wpływ na $\hat{S}_D$, i zatrzymuje się, gdy dalsze zmiany nie poprawiają różnicy między $\hat{S}_Z$ a $\hat{S}_D$ lub gdy osiągnięto minimalny zestaw zmian atrybutów. Jeśli meta-przykład zawiera mało próbek treningowych (poniżej progu μ), model arbitra (np. XGBoost z funkcją straty Coxa) zapewnia stabilną estymację $\hat{S}_Z$ na podstawie wszystkich danych treningowych.

Jakość rekomendacji weryfikowana jest za pomocą niezależnego modelu arbitra, który estymuje krzywą przeżycia po zastosowaniu zaleceń ($\hat{S}_{D_{est}}$). Jeśli $\hat{S}_{D_{est}}$ nie różni się statystycznie od $\hat{S}_D$ (test log-rank lub Kołmogorowa-Smirnowa), rekomendacje uznaje się za skuteczne.

5.3.2 Ilustracja działania metody

Niniejsza ilustracja przedstawia zastosowanie przedstawionego algorytmu rekomendacyjnego opartego na przeżyciowych regułach akcji. W tym przykładzie metodę zastosowano na zbiorze danych *Maintenance*, opisanego w Sekcji 5.2.3. Celem jest zilustrowanie działania metody oraz skuteczności w generowaniu możliwych do wdrożenia rekomendacji, które zwiększają wskaźniki przeżywalności poprzez modyfikację konkretnych wartości atrybutów.

Zbiór treningowy posłużył do indukcji przeżyciowych reguł akcji za pomocą algorytmu opisanego w Sekcji 5.2, który zidentyfikował wzorce i warunki wpływające na wskaźniki przeżywalności. Reguły te obejmują warunki, w których określone modyfikacje atrybutów mogą prowadzić do poprawy przeżywalności.

Indukcja przeżyciowych reguł akcji

Podczas fazy generowania reguł z wykorzystaniem zbioru treningowego uzyskano łącznie 22 przeżyciowych reguły akcji. Dla ilustracji poniżej przedstawiono trzy pierwsze z nich:

- **Reguła 1**:

$$\begin{aligned}&\text{jeśli } (\mathit{temperatureInd}, (-\infty, 73.0) \rightarrow (124.1, +\infty))\\ &\quad \wedge\ (\mathit{moistureInd}, (-\infty, 82.1) \rightarrow (94.9, +\infty))\\ &\quad \text{to } \hat{S}_Z \rightarrow \hat{S}_D\end{aligned}$$

- **Reguła 2**:

$$\begin{aligned}&\text{jeśli } (\mathit{pressureInd}, [37.9, 64.9) \rightarrow [81.1, 85.7))\\ &\quad \wedge\ (\mathit{temperatureInd}, [66.2, +\infty) \rightarrow [84.8, +\infty))\\ &\quad \text{to } \hat{S}_Z \rightarrow \hat{S}_D\end{aligned}$$

- **Reguła 3**:

$$\begin{aligned}&\text{jeśli } (\mathit{pressureInd}, [69.2, 113.5) \rightarrow (+\infty, 89.0)))\\ &\quad \wedge\ (\mathit{moistureInd}, [85.8, 88.4) \rightarrow [94.9, 104.1))\\ &\quad \text{to } \hat{S}_Z \rightarrow \hat{S}_D\end{aligned}$$

Budowa meta-tablicy

Meta-tablica zawiera meta-przykłady, czyli zbiory warunków dotyczących atrybutów. W każdym meta-przykładzie każdy uwzględniony atrybut ma przypisany dokładnie jeden przedział wartości (dla atrybutów liczbowych) lub jeden zbiór dopuszczalnych kategorii (dla atrybutów kategorycznych), wyznaczony przez reguły akcji. Atrybuty nieobecne w meta-przykładzie pozostają nieograniczone.

Ekstrakcja warunków Pierwszym krokiem jest ekstrakcja warunków z indukowanych reguł. Dla trzech pierwszych reguł warunki te przedstawiają się następująco:

- **Reguła 1**:
  - $(temperatureInd, [73.0, 124.1))$
  - $(moistureInd, (-\infty, 82.1))$
  - $(temperatureInd, [118.0, 122.3))$
  - $(moistureInd, [94.9, +\infty))$
- **Reguła 2**:
  - $(pressureInd, [37.9, 64.9))$
  - $(temperatureInd, [66.2, +\infty))$
  - $(pressureInd, [81.1, 85.7))$
  - $(temperatureInd, [84.8, +\infty))$
- **Reguła 3**:
  - $(pressureInd, [69.2, 113.5))$
  - $(moistureInd, [85.8, 88.4))$
  - $(pressureInd, (-\infty, 89.0))$
  - $(moistureInd, [94.9, 104.1))$

Powtarzanie się atrybutów w ramach pojedynczej reguły jest zamierzone i wynika ze struktury reguł akcji oraz sposobu konstrukcji meta-tablicy. W regułach akcji ten sam atrybut pojawia się wielokrotnie — w części opisującej część źródłową i docelową — co wskazuje, jak zmienia się zakres dopuszczalnych wartości danego

atrybutu. Na etapie ekstrakcji, dla każdego atrybutu gromadzone są wszystkie końce przedziałów występujące w regułach (zarówno w warunkach źródłowych, jak i docelowych), porządkuje je rosnąco i na ich podstawie dzieli dziedzinę atrybutu na rozłączne przedziały. Do meta-tablicy wprowadza się następnie meta-wartości odpowiadające tym przedziałom, tak aby każdy meta-przykład wskazywał dopuszczalny zakres wartości danego atrybutu.

Przetwarzanie atrybutów Algorytm przetwarza każdy atrybut niezależnie. Dla atrybutów numerycznych, takich jak *temperatureInd*, obliczane są wszystkie możliwe przecięcia między przedziałami określonymi w regułach, z wykorzystaniem wcześniej wyodrębnionych przedziałów (por. przykłady dla Reguł 1–3 powyżej). Dla atrybutu *temperatureInd* w pierwszych trzech regułach zidentyfikowano następujące rozłączne przedziały:

- $[66.2, +\infty)$
- $[73.0, 124.1)$
- $[84.8, +\infty)$
- $[118.0, 122.3)$

Na podstawie granic tych przedziałów wyznaczono rozłączny podział dziedziny atrybutu *temperatureInd* na siedem sąsiadujących przedziałów:

- $(-\infty, 66.2]$
- $(66.2, 73.0]$
- $(73.0, 84.8]$
- $(84.8, 118.0]$
- $(118.0, 122.3]$
- $(122.3, 124.1]$
- $(124.1, +\infty)$

Takie zdefiniowanie podziału umożliwia precyzyjne ukierunkowanie modyfikacji atrybutów.

Generowanie rekomendacji

Na etapie wprowadzania zmian, czyli dostosowania wartości atrybutów do rekomendowanych meta-wartości, algorytm aktualizuje wartości atrybutów przykładu testowego zgodnie z rekomendacjami pochodzącymi z meta-reguły akcji.

1. **Analiza rekomendacji** Dla każdego przykładu w zbiorze testowym algorytm identyfikuje odpowiednie meta-wartości z meta-tablicy. Poniżej przedstawiono wybrany przykład ze zbioru testowego, dla którego zostanie zastosowana rekomendacja:

$$\begin{aligned} \mathit{temperatureInd} &= 108.9 \\ \mathit{moistureInd} &= 94.0 \\ \mathit{pressureInd} &= 76.1 \end{aligned}$$

   Algorytm zidentyfikował trzy pasujące meta-wartości dla tego przykładu:

   - *temperatureInd*: $(107.2, 114.8)$
   - *moistureInd*: $(93.7, 94.9)$
   - *pressureInd*: $(75.8, 77.0)$

   Przedziały te stanowią meta-wartości w meta-tablicy zbudowanej na podstawie progów występujących w wyindukowanych przeżyciowych regułach akcji. Ich granice wyznaczono poprzez uporządkowanie wszystkich progów dla danego atrybutu i obliczenie rozłącznych przecięć sąsiednich przedziałów (części wspólnych tych przedziałów). Następnie każdy przykład jest odwzorowywany do jednego z tak powstałych przedziałów.

   Podział na 7 przedziałów dla *temperatureInd* pokazany powyżej został wyznaczony wyłącznie na podstawie progów z Reguł 1–3 i ma charakter ilustracyjny. Meta-wartości używane w dalszej części (np. $(107.2, 114.8)$) pochodzą z meta-tablicy zbudowanej na podstawie wszystkich progów ze wszystkich wyindukowanych reguł akcji, przez co stanowią uszczegółowienie tamtego podziału. W szczególności $(107.2, 114.8) \subset (84.8, 118.0]$, więc jest to fragment jednego z siedmiu przedziałów. Gdyby ograniczyć się do progów z Reguł 1–3, ten sam obiekt zostałby przypisany do przedziału $(84.8, 118.0]$.

Dodatkowo zidentyfikowano trzy główne meta-wartości do potencjalnych modyfikacji:

- *temperatureInd*: $(84.8, 118.0)$
- *moistureInd*: $(88.4, 94.9)$
- *pressureInd*: $(69.2, 81.1)$

2. **Wyznaczanie rekomendowanych zmian** Algorytm generuje rekomendacje poprzez porównanie aktualnych wartości atrybutów z potencjalnymi ulepszeniami. Dla wspomnianego przykładu zalecane są następujące zmiany:
   - *temperatureInd*: Przejście z $(84.8, 118.0)$ do $(122.3, 124.1)$
   - *moistureInd*: Przejście z $(88.4, 94.9)$ do $(85.776, 88.4)$
   - *pressureInd*: Przejście z $(69.2, 81.1)$ do $(-\infty, 37.9)$

   Te rekomendacje mają na celu dostosowanie wartości atrybutów do zakresów, które są statystycznie powiązane z poprawionymi wskaźnikami przeżywalności.

3. **Ocena jakości:** Jakość zestawu modyfikacji definiuje się jako wartość statystyki Kołmogorowa-Smirnowa (KS) obliczonej dla pary krzywych przeżycia: przed i po zastosowaniu zmian. Wyższa wartość oznacza większą poprawę. Równocześnie raportowana jest wartość p testu KS w celu weryfikacji istotności (próg $p \leqslant 0.05$). Dla przykładowego obiektu:
   - wartość początkowa (bez zmian): 0.000
   - po zmianie *pressureInd*: 0.372
   - po zmianie *moistureInd*: 0.551
   - wartość końcowa (po wszystkich zmianach): 0.654

   Wartości te odzwierciedlają skumulowany wpływ kolejnych modyfikacji na rozkład przeżycia oceniany za pomocą statystyki KS.

4. **Proces przycinania:** Aby zachować jedynie modyfikacje istotnie wpływające na poprawę wyniku, algorytm stosuje procedurę przycinania. Polega ona na tymczasowym usunięciu każdej modyfikacji i ponownym obliczeniu statystyki KS (oraz wartość p) dla pozostałego zestawu zmian. Modyfikacja

zostaje odrzucana, jeśli jej usunięcie nie powoduje pogorszenia jakości lub prowadzi do niespełnienia progu istotności. Przykładowo:

- Usunięcie zmiany *pressureInd*: Jakość spada do 0.090
- Usunięcie zmiany *moistureInd*: Jakość spada do 0.462
- Usunięcie zmiany *temperatureInd*: Jakość spada do 0.551

Znaczący spadek jakości po usunięciu każdej modyfikacji wskazuje na konieczność zachowania wszystkich proponowanych zmian.

5. **Końcowe zastosowanie zmian:** Na podstawie wyników przycinania algorytm wybiera optymalny zestaw zmian. Dla danego przykładu końcowe, zmienione wartości atrybutów są następujące:

   pressureInd: 35.8
   moistureInd: 87.1
   temperatureInd: 123.2

   Te modyfikacje dostosowują atrybuty przykładu do zalecanych meta-wartości, zwiększając tym samym jego wskaźnik przeżywalności zgodnie z indukowanymi regułami akcji.

Proces generowania rekomendacji zapewnia, że każdy przykład przechodzi statystycznie znaczące modyfikacje, przyczyniając się do poprawy wyników przeżywalności.

5.4 Algorytm indukcji przeżyciowych reguł wyjątków

W poniższej sekcji przedstawiono algorytm indukcji przeżyciowych reguł wyjątków w danych cenzurowanych. Analiza wyjątku opiera się na układzie trzech reguł (CR, RR, ER). Celem metody jest wykrywanie lokalnych odstępstw od typowego wzorca przeżycia (CR) wraz z regułą referencyjną (RR), wyjaśniającą kontekst odstępstwa, oraz regułą wyjątku (ER), łączącą warunki CR i RR. Algorytm bazuje na strategii *separate-and-conquer*, w której proces wzrostu reguły bazowej jest rozszerzony o wyszukiwanie wyjątków. Dzięki temu możliwa jest identyfikacja statystycznie istotnych wyjątków. Podobieństwo między regułami ocenia się za pomocą testu log-rank stosowanego do krzywych Kaplana-Meiera wyznaczonych dla przykładów pokrywanych przez dane reguły. Reguły CR i RR nie różnią się

statystycznie, natomiast pary CR i ER oraz RR i ER wykazują istotne różnice statystyczne.

5.4.1 Reguły wyjątków

Reguły wyjątków służą do identyfikacji nietypowych przypadków odstępujących od ogólnych wzorców w danych, co umożliwia wykrywanie anomalii [121]. Ich struktura opiera się na parze reguł: bazowej (ang. *commonsense rule*, CR), opisującej typowe zależności w danych, oraz wyjątku (ang. *exception rule*, ER), wskazującej odstępstwa. Hussain i in. [121] proponują również rozszerzoną strukturę potrójną (CR, ER, RR), w której reguła referencyjna (ang. *reference rule*, RR) określa kontekst występowania wyjątku. Formalnie, regułę wyjątku można zapisać jako:

- CR: jeśli $w_1^{CR} \wedge \cdots \wedge w_m^{CR}$ to $y = c_1$,
- ER: jeśli $w_1^{CR} \wedge \cdots \wedge w_m^{CR} \wedge w_1^{RR} \wedge \cdots \wedge w_n^{RR}$ to $y = c_2$,

gdzie m to liczba warunków w regule bazowej (CR), a n to liczba warunków reguły referencyjnej (RR). Warunki $w_1^{RR}, \ldots, w_n^{RR}$ stanowią dodatkowy kontekst wyróżniający wyjątek, a $c_1 \neq c_2$. Jakość reguł wyjątków ocenia się na podstawie ich zdolności do wykrywania rzadkich i nietypowych przypadków, często z wykorzystaniem miar odchyleń lub testów statystycznych [122].

Literatura wskazuje na różnorodne podejścia do odkrywania reguł wyjątków (ang. exception rules), koncentrujące się głównie na problemach klasyfikacyjnych. Klasyczna definicja opiera się na parze (CR, ER), czyli regule bazowej i regule wyjątku. Aby uniknąć „pozornych wyjątków", wprowadzono dodatkowo regułę referencyjną (RR), która nadaje kontekst — CR i RR powinny prowadzić do podobnych wniosków, natomiast dopiero ich koniunkcja kształtuje ER o odmiennym wniosku [121]. Taki trójelementowy układ (CR, ER, RR) ogranicza generowanie reguł o znikomym wsparciu i niskiej wartości poznawczej.

Ocena jakości wyjątków nie sprowadza się do oceny pojedynczej reguły. W literaturze stosuje się m.in. miary informacyjne oraz miary względnej atrakcyjności (ang. *relative interestingness measures*). W podejściu stochastycznym [122] ocenia jednocześnie jakość CR i ER, gdzie wyjątek uznaje się za interesujący, gdy ma wysoką jakość oraz kontrastuje z CR o wysokim wsparciu i wysokiej ufności. Z kolei miara względnej atrakcyjności łączy komponenty ufności i wsparcia wszystkich

trzech reguł (CR, RR, ER), co pozwala zrównoważyć rzadkość i wiarygodność [121].

Metody generowania wyjątków dzieli się na podejścia bezpośrednie i pośrednie. W podejściu bezpośrednim użytkownik dostarcza reguły CR, a algorytm wyszukuje dla nich wyjątki (ER) oraz odpowiadające im konteksty (RR). W podejściu pośrednim wzorzec bazowy (CR) nie jest narzucony — algorytmy samodzielnie konstruują i oceniają kandydatów (pary lub trójki reguł) względem zadanych kryteriów. W nurcie pośrednim zaproponowano m.in. podejście stochastyczne [122], jak również metody oparte na logice rozmytej, umożliwiające uchwycenie nieostrych granic wyjątków w danych technicznych [123]. Ujednolicony algorytm odkrywania wyjątków bez uprzednio wskazanej wiedzy (CR) przedstawili Suzuki i Żytkow, demonstrując jego skuteczność na 15 rzeczywistych zbiorach danych [124].

Badania przedstawione w literaturze koncentrują się głównie na problemach klasyfikacyjnych, podczas gdy zagadnienie wyjątków w regresji, a w szczególności w analizie przeżycia, jest reprezentowane słabo lub nie pojawia się wcale. Adaptacja koncepcji (CR, RR, ER) do danych cenzurowanych wymaga użycia miar i procedur statystycznych specyficznych dla analizy przeżycia (krzywe Kaplana-Meiera, test log-rank, kryteria istotności i wielkości efektu). W niniejszej pracy przyjęto takie ujęcie — wyjątki w danych przeżyciowych definiuje się poprzez trójkę (CR, RR, ER) oraz weryfikuje testem log-rank w taki sposób, aby CR i RR nie różniły się statystycznie, natomiast ER istotnie różniła się od obu. Dzięki temu możliwe jest wykrycie podgrup o nietypowo dłuższym lub krótszym czasie przeżycia względem CR, co ułatwia identyfikację czynników prognostycznych.

5.4.2 Opis metody

Opracowane na potrzeby metody kryteria statystyczne dla ważności reguły wyjątku są następujące:

1. $H_0 : S_{CR}(t) = S_{RR}(t)$ nie może być odrzucona na poziomie istotności $\alpha = 0.05$ (test log-rank),
2. $H_0 : S_{CR}(t) = S_{ER}(t)$ musi być odrzucona na poziomie istotności $\alpha = 0.05$,
3. $H_0 : S_{RR}(t) = S_{ER}(t)$ musi być odrzucona na poziomie istotności $\alpha = 0.05$,

gdzie $S_{CR}(t)$, $S_{RR}(t)$ i $S_{ER}(t)$ oznaczają funkcje przeżycia odpowiednio dla reguł CR, RR i ER.

Dodatkowo wymaga się występowania przeciwnego wzorca przeżycia względem reguły bazowej. Weryfikacja odbywa się poprzez porównanie median czasu przeżycia, statystyk log-rank lub indeksów zgodności Harrella. Wprowadza się warunek kierunkowości efektu — ER ma reprezentować populację o trendzie przeżycia przeciwnym do CR. Warunek ten sprawdzany jest na trzy równoważne sposoby, z których spełnienie co najmniej jednego przy istotności statystycznej stanowi wymóg:

- kryterium mediany — mediana(ER) istotnie różni się od mediana(CR) z niepokrywającymi się w 95% przedziałami ufności; dla wyjątku „gorszego” mediana(ER) < mediana(CR), a dla „lepszego” odwrotnie;
- kryterium log-rank — znak statystyki porównania CR z całym zbiorem oraz ER z CR jest przeciwny przy $p \leqslant 0.05$;
- kryterium indeksu zgodności Harrella — gdy $c_{CR,all} < 0.5$, wymaga się $c_{ER,CR} > 0.5$; gdy $c_{CR,all} > 0.5$, wymaga się $c_{ER,CR} < 0.5$.

Ponadto wymagana jest minimalna wielkość efektu między ER a CR, rozumiana jako różnica wartości przyjętej miary (np. różnica median czasu przeżycia), która musi przekraczać ustalony próg istotności. Przykładowo wymaga się, aby bezwzględna różnica median $|\,\text{mediana}(ER) - \text{mediana}(CR)\,|$ była nie mniejsza niż ustalony próg. Dodatkowo egzekwowane są minimalne liczności grup (np. |CR|, |RR|, |ER| $\geqslant \mu$), aby wyniki nie były zniekształcone przez bardzo małe grupy. W przypadku spełnienia wielu kryteriów priorytet ma test log-rank, natomiast kryteria mediany i indeksu zgodności Harrella pełnią rolę potwierdzenia kierunku i wielkości efektu.

Walidacja statystyczna reguł wyjątków opiera się na teście log-rank, który porównuje rozkłady czasu przeżycia między różnymi grupami. Algorytm sprawdza trzy pary porównań statystycznych: między regułą bazową a regułą wyjątków (CR-ER), między regułą referencyjną a regułą wyjątków (RR-ER) oraz między regułą zdroworozsądkową a regułą referencyjną (CR-RR). Reguła wyjątków jest uznawana za statystycznie istotną, gdy porównania CR-ER i RR-ER wykazują istotne różnice ($p \leqslant 0.05$), podczas gdy CR-RR nie wykazuje istotnych różnic ($p > 0.05$). Takie

podejście gwarantuje, że reguły wyjątków reprezentują rzeczywiście odmienne wzorce przeżycia w stosunku do reguł bazowych i referencyjnych.

Algorytm sekwencyjnego wyszukiwania przeżyciowych reguł wyjątków Proponowany algorytm opiera się na strategii sekwencyjnego pokrywania z wyszukiwaniem wyjątków w trakcie procesu indukcji reguł. Podczas konstruowania (wzrostu) reguły bazowej (CR) na każdym etapie sprawdzane jest, czy w jej kontekście można wyodrębnić regułę referencyjną (RR) oraz odpowiadającą jej regułę wyjątku (ER). Jeśli wyjątek zostanie potwierdzony, wzrost CR zostaje przerwany, a do zbioru wynikowego dołączane są reguły RR i ER. W przeciwnym razie algorytm kontynuuje wzrost CR.

Poszukiwanie wyjątku sprowadza się do identyfikacji sytuacji, w której krzywe przeżycia dla CR i RR nie różnią się istotnie statystycznie, podczas gdy krzywa przeżycia dla ER różni się istotnie statystycznie zarówno od CR, jak i od RR. Podejście to odróżnia się od istniejących metod, w których najpierw generuje się wiele reguł, a dopiero w następnej kolejności sprawdza, czy ich kombinacje spełniają określone kryteria dla wyjątków.

Proponowana metoda ukierunkowuje wzrost reguły poprzez iteracyjne dodawanie warunków maksymalizujących wartość statystyki log-rank. Po każdym rozszerzeniu reguły sprawdzane jest, czy można dla niej znaleźć odpowiednią regułę referencyjną. Algorytm rejestruje historię jakości dla kolejnych rozszerzeń reguły, a następnie wybiera optymalną długość reguły poprzez przycięcie do warunku o najwyższej jakości. Schemat całego procesu został przedstawiony w Algorytmie 4.

Indukcja warunków Procedura IndukujWarunek implementuje algorytm zachłanny, który w każdej iteracji wybiera warunek maksymalizujący jakość reguły. Dla każdego kandydującego warunku c:

1. tworzona jest tymczasowa reguła r' przez dodanie warunku c do reguły r,
2. obliczane jest pokrycie reguły r',
3. estymowana jest funkcja przeżycia dla przykładów pokrytych przez r' za pomocą estymatora Kaplana-Meiera,
4. obliczana jest jakość reguły jako wartość statystyki log-rank między przykładami pokrytymi a niepokrytymi,

Algorytm 4 Wzrost reguły bazowej (CR)

procedure WZROST($CR, D, U, maks\_wzrost$)
  Wejście:
    CR — bieżąca reguła bazowa
    $D(A, T, \delta)$ — zbiór danych opisany atrybutami A, czasem obserwacji T i statusem przeżycia δ
    U — lista niepokrytych przykładów
    $maks\_wzrost$ — maksymalna liczba warunków w regule
  Wyjście: wartość logiczna określająca czy reguła została rozszerzona
  q $\leftarrow \varnothing$ ▷ historia jakości (wektor)
  cont $\leftarrow$ true ▷ flaga kontynuacji
  while cont **do**
    $(w, q, \text{cov}) \leftarrow$ INDUKUJWARUNEK(CR, D, U) ▷ warunek, jakość, pokrycie
    if $w \neq$ None **then**
      Dodaj w do CR ▷ Rozszerz regułę o najlepszy warunek
      Dodaj q do **q** ▷ aktualizacja historii ocen jakości
      ex $\leftarrow$ WYSZUKAJWYJĄTKI(CR, D) ▷ flaga znalezienia wyjątku
      if ex **then**
        cont $\leftarrow$ false ▷ Znaleziono wyjątek, przerwij wzrost
      end if
    else
      cont $\leftarrow$ false ▷ Nie znaleziono kandydującego warunku
    end if
    if liczba warunków w CR $\geqslant maks\_wzrost$ **then**
      cont $\leftarrow$ false ▷ Osiągnięto maksymalną długość reguły
    end if
  end while
  if q nie jest pusta **then**
    $i_{\text{najl}} \leftarrow$ ARGMAX(**q**) ▷ indeks maksymalnej jakości
    Ogranicz CR do pierwszych $i_{\text{najl}} + 1$ warunków
    return true
  else
    return false
  end if
end procedure

5. sprawdzane jest kryterium minimalnego pokrycia: $|\text{PokrytePrzykłady}(r')| \geqslant \mu$

Spośród kandydujących wybierany jest warunek, który maksymalizuje jakość reguły i jednocześnie spełnienia kryterium minimalnego pokrycia.

Wyszukiwanie wyjątków Procedura WyszukajWyjątki (Algorytm 5) wywoływana jest po każdym rozszerzeniu CR i służy do weryfikacji, czy w aktualnym kontekście istnieje para (RR, ER) spełniająca definicję wyjątku. Wejściem procedury są bieżąca reguła CR, macierz cech i wektor danych przeżycia. W pierwszym kroku identyfikowane są przykłady pokryte i niepokryte przez CR, przy czym niepokryte przykłady stanowią przestrzeń poszukiwań dla reguły referencyjnej RR.

W kolejnym kroku uruchamiana jest procedura WzrostReguluReferencyjnej, która indukuje regułę referencyjną w oparciu o kryterium jakości specyficznym dla analizy przeżycia. Po uzyskaniu kandydata RR konstruowany jest kandydat wyjątku ER jako koniunkcja warunków z CR i RR. Następnie przeprowadzana jest weryfikacja statystyczna z wykorzystaniem testu log-rank dla trzech par: CR-ER, RR-ER oraz CR-RR. Wyjątek uznawany jest za statystycznie istotny, gdy różnice CR-ER i RR-ER są istotne ($p \leqslant 0.05$), natomiast różnica CR-RR nie jest istotna ($p > 0.05$). W przypadku spełnienia tych kryteriów proces wzrostu CR zostaje przerwany, w przeciwnym razie wzrost CR jest kontynuowany.

Indukcja reguły referencyjnej Procedura IndukujRR implementuje zachłanną strategię indukcji warunków dla reguły referencyjnej, której głównym celem jest maksymalizacja p-wartości testu log-rank między CR a RR (im wyższa p-wartość, tym mniejsza różnica statystyczna). Dla każdego kandydującego warunku algorytm konstruuje tymczasową regułę RR, oblicza pokrycie reguły wyjątku ER jako przecięcie pokryć CR i RR, a następnie przeprowadza testy log-rank dla trzech par: CR-RR, CR-ER oraz RR-ER. Ocena kandydatów jest prowadzona pod warunkiem spełnienia kryteriów minimalnych (minimalnego pokrycia μ oraz minimalnej liczby niecenzurowanych zdarzeń ν w RR) przy poziomie istotności $\alpha = 0.05$. W przypadku równych p-wartości porównania CR-RR preferowane są warunki prowadzące do większego pokrycia w RR, a następnie do krótszej przesłanki (mniejszej liczby warunków).

Algorytm 5 Wyszukiwanie wyjątków

procedure WyszukajWyjątki(CR, D)
  Wejście:
    CR — reguła bazowa
    $D(A, T, \delta)$ — zbiór danych opisany atrybutami A, czasem obserwacji T i statusem przeżycia δ
  Wyjście: wartość logiczna określająca czy znaleziono statystycznie istotny wyjątek
  $\mathcal{C}_{CR} \leftarrow \{i \mid CR$ pokrywa przykład $i\}$ ▷ zbiór przykładów pokrytych przez CR
  $\overline{\mathcal{C}}_{CR} \leftarrow \{i \mid CR$ nie pokrywa przykładu $i\}$ ▷ zbiór przykładów niepokrytych przez CR
  $RR \leftarrow$ WzrostReguluReferencyjnej$(RR, D, \overline{\mathcal{C}}_{CR}, \mathcal{C}_{CR})$
  if $RR \neq \varnothing$ **then**
    $ER \leftarrow$ nowa reguła wyjątku
    $ER \leftarrow CR \wedge RR$ ▷ koniunkcja przesłanek (warunków) CR i RR
    $\mathcal{C}_{ER} \leftarrow \{i \mid ER$ pokrywa przykład $i\}$ ▷ zbiór przykładów pokrytych przez ER
    $\mathcal{C}_{RR} \leftarrow \{i \mid RR$ pokrywa przykład $i\}$ ▷ zbiór przykładów pokrytych przez RR
    $p_{CR,ER} \leftarrow$ LogRank$(S_{CR}(t), S_{ER}(t))$
    $p_{RR,ER} \leftarrow$ LogRank$(S_{RR}(t), S_{ER}(t))$
    $p_{CR,RR} \leftarrow$ LogRank$(S_{CR}(t), S_{RR}(t))$
    if $p_{CR,ER} \leqslant 0.05$ **and** $p_{RR,ER} \leqslant 0.05$ **and** $p_{CR,RR} > 0.05$ **then**
      return true ▷ Znaleziono statystycznie istotny wyjątek
    else
      return false
    end if
  else
    return false ▷ Nie znaleziono reguły referencyjnej
  end if
end procedure

Warunek jest akceptowany jako najlepszy kandydat, gdy spełnia łącznie następujące trzy kryteria:

(1) statystyczna istotność wyjątku ($p_{CR,RR} > 0.05 \wedge p_{CR,ER} \leqslant 0.05 \wedge p_{RR,ER} \leqslant 0.05$),

(2) minimalne pokrycie wyjątku ($|ER\_pokryte| \geqslant \mu_{wyjtek}$),

(3) poprawa p-wartości CR-RR w stosunku do dotychczasowego najlepszego wyniku.

Algorytm priorytetowo wybiera warunki prowadzące do jak najmniejszej różnicy statystycznej między CR a RR, przy jednoczesnym zachowaniu istotnych różnic między regułami wyjątku a obiema regułami bazowymi.

Procedura kończy działanie, gdy żaden z pozostałych kandydatów nie spełnia wymaganych kryteriów, nie poprawia aktualnej najlepszej p-wartości CR-RR lub gdy osiągnięto maksymalną długość reguły. Zwracany jest najlepszy znaleziony warunek wraz z wartościami miar jakości, które następnie są wykorzystywane w procedurze wzrostu reguły referencyjnej.

Weryfikacja kandydata na wyjątek W algorytmie wyszukiwania wyjątków (Algorytm 5) weryfikacja statystyczna kandydata na regułę wyjątku opiera się na trzech krokach:

(1) konstrukcja reguły wyjątku ER jako koniunkcji wszystkich warunków z reguły bazowej CR i reguły referencyjnej RR,

(2) obliczenie pokryć dla każdej z trzech reguł na zbiorze danych,

(3) wykonanie dwóch testów log-rank porównujących rozkłady czasu przeżycia między grupami: CR z ER oraz RR z ER.

Kandydat jest uznawany za statystycznie istotną regułę wyjątku, jeśli oba testy wykazują istotne różnice na poziomie $\alpha = 0.05$ ($p \leqslant 0.05$). Po pozytywnej weryfikacji reguły RR i ER są przypisywane do reguły CR, tworząc trójkę (CR, RR, ER), która następnie zostaje dodana do zbioru wynikowego. Procedura zwraca wartość logiczną informującą o wyniku weryfikacji, co determinuje dalsze kroki algorytmu wzrostu reguły bazowej.

Algorytm 6 Indukcja reguły referencyjnej (RR)

procedure INDUKUJRR($RR, D, p_{najl}, \overline{C}, C_{CR}$)
  Wejście:
    RR — bieżąca reguła referencyjna
    $D(A, T, \delta)$ — zbiór danych opisany atrybutami A, czasem obserwacji T i statusem przeżycia δ
    p_{najl} — najlepsza p-wartość CR-RR
    $\overline{C}$ — indeksy niepokryte przez żadną regułę
    C_{CR} — indeksy pokryte przez CR
  Wyjście: najlepszy warunek, jakość, pokrycie, p-wartość
  $w_{najl} \leftarrow \varnothing$, $q_{najl} \leftarrow -\infty$
  $C_{najl} \leftarrow \varnothing$, $p_{\text{najl}} \leftarrow p_{najl}$
  $W' \leftarrow$ POBIERZMOŻLIWEWARUNKI(RR, D)
  Odfiltruj kandydatów już obecnych w RR
  for all warunek w w W' **do**
    $RR' \leftarrow RR.add(w)$ ▷ Dodaj warunek do RR
    $C_{RR} \leftarrow$ pokrycie RR'
    $C_{ER} \leftarrow C_{CR} \cap C_{RR}$ ▷ Przecięcie CR i RR
    if $|C_{ER}| = 0$ **then**
      continue ▷ Brak przecięcia, pomiń
    end if
    $p_{CR,ER} \leftarrow$ LOGRANK(S_{CR}, S_{ER})
    $p_{RR,ER} \leftarrow$ LOGRANK(S_{RR}, S_{ER})
    $p_{CR,RR} \leftarrow$ LOGRANK(S_{CR}, S_{RR})
    sig $\leftarrow$ ($p_{CR,RR} > 0.05$ **and** $p_{CR,ER} \leqslant 0.05$ **and** $p_{RR,ER} \leqslant 0.05$)
    $cov_{wyst} \leftarrow |C_{ER}| \geqslant \mu_{wyjtek}$ ▷ minimalne pokrycie ER spełnione
    p_lepsze_od_najl $\leftarrow p_{CR,RR} > p_{\text{najl}}$ ▷ czy aktualne $p_{CR,RR}$ jest większe niż dotychczasowe p_{najl}
    if sig **and** cov_{wyst} **and** p_lepsze_od_najl **then**
      $w_{najl} \leftarrow w$, $p_{\text{najl}} \leftarrow p_{CR,RR}$
      Aktualizuj q_{najl} i C_{najl}
    end if
  end for
  return w_{najl}, q_{najl}, C_{najl}, p_{najl}
end procedure

Metody porównywania wzorców przeżycia Niezależnie od testu log-rank, można rozważyć trzy alternatywne metody porównywania wzorców przeżycia w celu weryfikacji, czy reguła wyjątku wykazuje przeciwny wzorzec przeżycia względem reguły bazowej.

Metoda mediany. Jeśli mediana czasu przeżycia w grupie CR jest nie większa niż w całym zbiorze danych, to ER uznawana jest za istotną, gdy jej mediana jest większa niż w CR. W przeciwnym razie (gdy mediana w CR jest większa niż w całym zbiorze) ER uznawana jest za istotną, gdy jej mediana jest mniejsza niż w CR.

Metoda log-rank. Wyznacza się statystyki dla porównań CR z całym zbiorem oraz ER z CR. Warunek jest spełniony, gdy znaki tych statystyk są przeciwne.

Metoda indeksu zgodności Harrella. Oblicza się wartości dla porównań CR z całym zbiorem oraz ER z CR. Jeśli dla porównania CR z całym zbiorem otrzymana wartość jest mniejsza niż 0.5 (gorsze przeżycie), to ER uznaje się za istotną, gdy dla porównania ER z CR otrzymana wartość jest większa niż 0.5 (lepsze przeżycie). Zastosowanie indeksu zgodności weryfikuje kierunek efektu na poziomie porządkowania par obserwacji, a nie tylko przesunięcie rozkładów: wartość 0.5 odpowiada losowej zgodności, a wartości powyżej/poniżej 0.5 wskazują odpowiednio lepszą/gorszą zdolność dyskryminacyjną. Metoda ta jest odporna na cenzorowanie i komplementarna względem median oraz testu log-rank. W tym kontekście wykorzystuje się jedynie próg 0.5 do potwierdzenia odwrócenia wzorca przeżycia.

5.4.3 Ilustracja działania metody

W poniższej sekcji przedstawiono przebieg indukcji reguł wyjątków w zbiorze *LEDLife*. Zbiór obejmuje 167 obserwacji i zawiera dwie zmienne objaśniające: *DegreesC* (temperatura) oraz *Current* (natężenie prądu). Opis skoncentrowano na przebiegu generowania reguł, od wzrostu reguły bazowej (CR), przez wyznaczenie reguły referencyjnej (RR), po konstrukcję i weryfikację wyjątku (ER). Przedstawiony wyjątek jest jedynym wyjątkiem zidentyfikowanym w tym zbiorze danych.

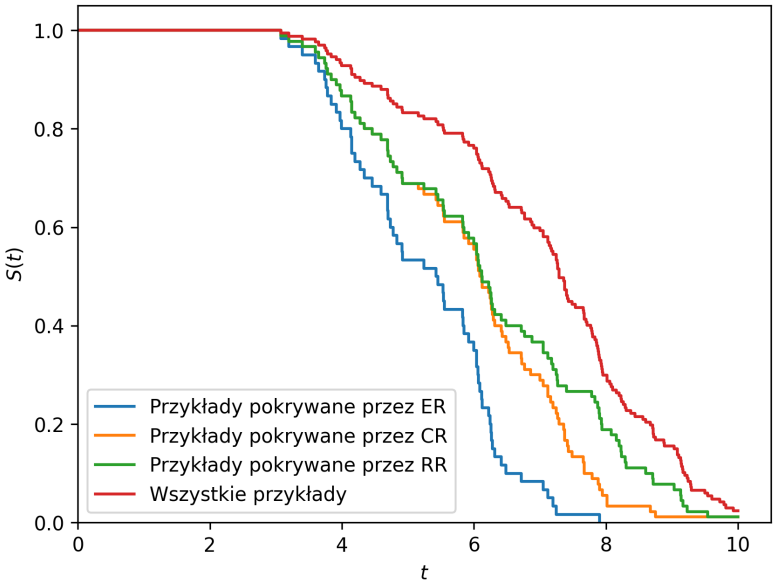
Krok 1: Wzrost reguły bazowej (CR) W pierwszym etapie przeprowadzono wzrost reguły bazowej, stanowiącej punkt odniesienia dla dalszych porównań. Trójkę (CR, RR, ER) uzyskano w drugiej próbie wzrostu reguły CR. W tej próbie w pierwszej iteracji specjalizacji dodano warunek $DegreesC \geqslant 102.00$, który wyodrębnia podzbiór obserwacji spełniających przesłankę reguły. Dla tej grupy wyznaczono medianę przeżycia równą 6.09, przy pokryciu 90 ze 167 przykładów i liczbie zdarzeń 89 (pozostałe obserwacje w tej grupie są cenzurowane). Reguła CR pełni rolę wzorca odniesienia w porównaniach CR z RR i CR z ER. We wcześniejszej oraz późniejszej próbie wzrostu CR nie wyznaczono RR. W konsekwencji nie wyznaczono również ER, gdyż w tych kontekstach nie zostało jednocześnie spełnione kryterium minimalnego pokrycia wyjątku oraz kryteria istotności statystycznej.

Krok 2: Wyznaczenie reguły referencyjnej (RR) W drugiej próbie wzrostu CR, w pierwszej iteracji specjalizacji RR, wybrano warunek referencyjny $Current \geqslant 35.00$. Został on wybrany, ponieważ uzyskano dla niego największą p-wartość w teście log-rank dla porównania CR z RR (najmniejsza różnica między krzywymi przeżycia CR i RR), przy jednoczesnym spełnieniu kryterium minimalnego pokrycia wyjątku (liczba przykładów $CR \wedge RR$) oraz kryterium istotności dla porównań CR z ER i RR z ER. W kolejnym kroku nie dodano dalszych warunków, gdyż żaden kandydat nie poprawiał p-wartości w porównaniu CR z RR ani nie spełniał równocześnie wymagań pokrycia i istotności. W konsekwencji zakończono wzrost RR. Dla RR uzyskano medianę przeżycia równą 6.12, przy pokryciu 90 ze 167 przykładów i liczbie zdarzeń 89. Porównanie CR z RR testem log-rank dało p-wartość równą 0.07 (powyżej progu istotności 0.05), co spełnia założenie o braku istotnej różnicy między CR i RR i umożliwia konstrukcję wyjątku.

Krok 3: Konstrukcja i weryfikacja wyjątku (ER) Wyjątek zdefiniowano jako koniunkcję CR $\wedge$ RR: $DegreesC \geqslant 102.00 \wedge Current \geqslant 35.00$. Dla ER otrzymano medianę przeżycia równą 5.45, przy pokryciu 60 ze 167 przykładów i liczbie zdarzeń 60. Testem log-rank porównano krzywe przeżycia w dwóch parach: dla CR z ER uzyskano p-wartość w przybliżeniu równą $9.4 \cdot 10^{-5}$, a dla RR z ER $p = 4.0 \cdot 10^{-6}$ (obie wartości poniżej progu istotności 0.05). Zatem warunki wyjątkowości zostały spełnione: CR i RR nie różnią się istotnie, natomiast ER różni się istotnie od obu.

Odpowiadająca trójka reguł została przedstawiona poniżej (w nawiasach podano liczbę obserwacji spełniających przesłankę reguły k oraz liczebność zbioru obserwacji n).

$$\begin{aligned}
&\text{CR: jeśli } (\mathit{DegreesC} \geqslant 102.00) \\
&\qquad \text{to } \mathit{mediana\ przeżycia} = 6.0865 \quad (k = 90, n = 167) \\
&\text{RR: jeśli } (\mathit{Current} \geqslant 35.00) \\
&\qquad \text{to } \mathit{mediana\ przeżycia} = 6.1193 \quad (k = 90, n = 167) \\
&\text{ER: jeśli } (\mathit{DegreesC} \geqslant 102.00 \wedge \mathit{Current} \geqslant 35.00) \\
&\qquad \text{to } \mathit{mediana\ przeżycia} = 5.4502 \quad (k = 60, n = 167)
\end{aligned} \tag{5.6}$$



Rysunek 5.3: Krzywe KM dla zbioru *LEDLife*: reguły CR, RR oraz ER. Widoczny jest brak różnic między CR i RR oraz istotnie gorsza krzywa ER.

Na rysunku 5.3 przedstawiono krzywe KM dla trzech grup (CR, RR i ER). Krzywe CR i RR mają zbliżony przebieg, co potwierdza test log-rank ($p \approx 0.07$). Dla ER obserwuje się natomiast niższe wartości funkcji przeżycia S oraz istotne statystycznie różnice względem CR i RR. Szczegółowe metryki (pokrycie, liczba zdarzeń, mediana) zestawiono w tabeli 5.1.

Podsumowując, przy jednoczesnym spełnieniu warunków CR i RR wyodrębnia się grupa o istotnie gorszym przebiegu przeżycia (ER). Wynik ten wskazuje na

Reguła	pokrycie (k/n)	zdarzenia	mediana
CR	90/167	89	6.09
RR	90/167	89	6.12
ER	60/167	60	5.45

Tabela 5.1: Zestawienie wartości metryk dla trójki (CR, RR, ER): pokrycie k/n (k — liczba obserwacji spełniających przesłankę reguły; n — liczba obserwacji w zbiorze), *zdarzenia* (liczba niecenzurowanych obserwacji w grupie) oraz *mediana* (mediana czasu przeżycia w grupie).

interakcję wysokiej temperatury i wysokiego natężenia prądu, której nie obserwuje się przy analizie każdego z warunków oddzielnie.

5.5 Interpretowalny zespół reguł przeżyciowych

Metoda interpretowalnego zespołu reguł przeżyciowych (ang. *survival rules ensemble*) stanowi podejście do analizy przeżycia, które łączy interpretowalność modeli opartych na regułach z dokładnością technik uczenia zespołowego. Metoda ta rozszerza klasyczny paradygmat lasów losowych, zastępując drzewa decyzyjne modelami reguł przeżyciowych, dostosowanymi do analizy danych cenzurowanych. W tej sekcji najpierw zdefiniowano pojęcie zespołu reguł i sposób agregacji predykcji, a następnie opisano architekturę metody dostosowaną do danych cenzurowanych oraz procedurę budowy i ewaluacji zespołu.

5.5.1 Zespoły reguł

Zespół modeli (ang. *ensemble*) definiuje się jako zbiór estymatorów bazowych, których predykcje są łączone według ustalonej reguły agregacji. Celem jest redukcja wariancji lub obciążenia oraz zwiększenie stabilności względem pojedynczego estymatora. Wyróżnia się m.in. zespoły równoległe (np. bagging, gdzie estymatory trenuje się niezależnie) oraz zespoły sekwencyjne (np. boosting, gdzie estymatory trenuje się kolejno, korygując błędy poprzedników).

W zespołach reguł estymatorami bazowymi są zbiory reguł (modele regułowe). Wewnątrz estymatora bazowego predykcja wyznaczana jest przez agregację wkła-

dów reguł, których przesłanki są spełnione dla obserwacji $\mathbf{x}$ (np. głosowanie większościowe, uśrednianie, suma ważona). Następnie, na poziomie zespołu, agreguje się predykcje poszczególnych estymatorów bazowych. Taka konstrukcja zachowuje interpretowalność zarówno na poziomie reguł, jak i na poziomie estymatorów bazowych.

Formalnie, funkcję predykcji $\hat{f}(\mathbf{x})$ aproksymuje się za pomocą kombinacji liniowej:

$$\hat{f}(\mathbf{x}) = \beta_0 + \sum_{m=1}^{M} \beta_m r_m(\mathbf{x}), \tag{5.7}$$

gdzie β_0 oznacza wyraz wolny, $r_m(\mathbf{x})$ — predykcję m-tego estymatora bazowego (modelu regułowego) dla obserwacji $\mathbf{x}$ (w szczególnym przypadku konstrukcji typu RuleFit może to być funkcja indykatorowa aktywacji pojedynczej reguły), β_m — wagę przypisaną m-temu estymatorowi bazowemu (uczoną np. metodą regresji z regularyzacją lub ustalaną heurystycznie), a M — liczbę estymatorów bazowych w zespole. W niniejszej sekcji symbolem $r_m(\mathbf{x})$ oznaczono funkcję bazową: predykcję modelu regułowego lub, jak w RuleFit, indykator aktywacji pojedynczej reguły.

Taka reprezentacja stanowi podstawę metody RuleFit [125], która generuje reguły poprzez ekstrakcję ścieżek z lasów losowych, a następnie optymalizuje wagi β_m za pomocą regularyzowanej regresji (L1 lub lasso). Model osiąga wysoką precyzję predykcji zachowując interpretowalność — każda reguła r_m opisuje lokalny wzorzec w danych, a waga β_m wskazuje jej znaczenie.

W analizie przeżycia, zespoły reguł dostosowuje się do specyficznych wymagań, takich jak estymacja funkcji przeżycia $\hat{S}(t|\mathbf{x})$ lub modelowanie hazardu. Przykładem jest metoda LR-Rules [126], która dla danej obserwacji estymuje funkcję przeżycia poprzez uśrednianie (np. średnią nieważoną) krzywych Kaplana-Meiera reguł ją pokrywających, co umożliwia modelowanie danych cenzurowanych z zachowaniem interpretowalności. Innym przykładem jest SURVFIT [127], który wykorzystuje liniowy model reguł z funkcją straty dostosowaną do cenzurowania oraz regularyzacją indukującą tzw. podwójną rzadkość: wybór niewielkiej liczby reguł oraz niewielkiej liczby zmiennych występujących w regułach. Dzięki temu metoda zachowuje interpretowalność i jest skalowalna obliczeniowo.

Podsumowując, przedstawione podejście do zespołów reguł opiera się na ich matematycznej definicji jako zbioru funkcji indykatorowych, agregowanych w celu

uzyskania predykcji. Wprowadzenie notacji, takiej jak $r_m(\mathbf{x})$ czy $\hat{f}(\mathbf{x})$, oraz wzorów, takich jak liniowa kombinacja reguł czy estymacja przeżycia, umożliwia precyzyjne opisanie ich konstrukcji i działania, stanowiąc podstawę teoretyczną dla dalszych rozważań.

5.5.2 Opis metody

Podstawowe elementy metody to indukcja przeżyciowych reguł decyzyjnych w estymatorach bazowych, losowe próbkowanie danych i cech ograniczające korelację między estymatorami oraz agregacja predykcji estymatorów bazowych dostosowana do analizy przeżycia. Każdy estymator indukuje zbiór reguł w postaci koniunkcji predykatów na atrybutach, z przypisaną estymowaną funkcją przeżycia dla pokrywanego podzbioru danych. Agregacja polega na wyznaczeniu wspólnej dziedziny czasowej obejmującej wszystkie punkty predykcji, interpolacji każdej krzywej przeżycia do tej dziedziny oraz obliczeniu średniej arytmetycznej wartości w każdym punkcie. Wynikową krzywą koryguje się tak, aby była nierosnąca.

Algorytm przebiega w trzech fazach:

(1) próbkowanie *bootstrap* i losowy wybór podzbioru cech (ang. *random subspace*, *feature subsampling*) (wejście: $\mathbf{X}, \mathbf{y}$; wyjście: zbiory $\{(\mathbf{X}_i, \mathbf{y}_i, \mathcal{F}_i)\}_{i=1}^{n}$),

(2) indukcja reguł i trening (wejście: $(\mathbf{X}_i, \mathbf{y}_i)$; wyjście: estymatory bazowe $\mathcal{M}_i$ oraz zbiory reguł $\mathcal{R}_i$ z estymatorami $\hat{S}_i(t)$),

(3) agregacja specyficzna dla przeżycia (wejście: $\{\mathcal{M}_i, \mathcal{R}_i\}_{i=1}^{n}$; wyjście: zespołowe krzywe przeżycia).

Konstrukcja faz 1-3 ogranicza wariancję predykcji i zachowuje interpretowalność na poziomie reguł oraz całego zespołu.

Na potrzeby implementacji wyróżnia się dwa zbiory parametrów: zespołowe (n, θ, α) oraz indukcji reguł $(\sigma, \gamma, \pi, \iota, \mu)$. Na poziomie zespołu: n to liczba estymatorów (domyślnie 100), θ kontroluje maksymalną liczbę cech w zbiorze treningowym dla pojedynczego estymatora bazowego (domyślnie $\lfloor\sqrt{|A|}\rfloor$, gdzie $|A|$, to liczba atrybutów w zbiorze; alternatywnie strategie typu $\log_2 |A|$ lub udział w $(0, 1]$), a α wyznacza rozmiar próbki bootstrap (domyślnie 1.0). Na poziomie indukcji reguł: σ określa minimalne wsparcie, γ ogranicza długość przesłanki, π włącza poinduk-

cyjne przycinanie, ι definiuje sposób traktowania braków danych (pomijanie lub interpolacja), a μ ogranicza maksymalny odsetek przykładów niepokrytych.

Faza 1: Próbkowanie bootstrap i losowy wybór podzbioru cech Dla każdego estymatora bazowego (tj. pojedynczego modelu reguł przeżyciowych trenowanego niezależnie na własnej próbce bootstrap i podzbiorze cech) próbkowanie bootstrap oraz losowy wybór podzbioru cech (ang. *random subspace*, *feature subsampling*) zapewniają różnorodność modeli, generując różne podzbiory danych treningowych, a także redukują korelację między modelami dzięki losowej selekcji cech.

Implementacja wykorzystuje agregację bootstrapową (ang. *bagging, bootstrap aggregating*), w której każdy estymator trenowany jest na próbce bootstrap pochodzącej z oryginalnego zbioru danych. Rozmiar tej próbki kontrolowany jest przez parametr α, który można ustawić jako ułamek całkowitej liczby obserwacji lub jako konkretną wartość liczbowa.

Dodatkowym elementem różnicującym estymatory jest losowy wybór cech (ang. *feature selection*) dla każdego z nich. Implementacja obsługuje różne strategie selekcji cech, w tym $\sqrt{|A|}$ (pierwiastek kwadratowy z liczby cech), $\log_2 |A|$ (logarytm binarny) oraz udział całkowitej liczby cech. Takie podejście dodatkowo zwiększa różnorodność zespołu i zmniejsza ryzyko przeuczenia, ograniczając korelację między estymatorami.

Faza 2: Indukcja reguł i trening Każdy estymator bazowy generuje reguły przeżycia. Proces ich wzrostu przebiega zachłannie (*separate-and-conquer*) — w każdej iteracji dodawany jest warunek maksymalizujący statystykę log-rank między przykładami pokrytymi a niepokrytymi. Jednocześnie egzekwowane są ograniczenia minimalnego wsparcia σ oraz maksymalnej długości reguły γ. Po zakończeniu wzrostu stosowane jest przycinanie π, które usuwa lub łagodzi warunki, o ile nie pogarsza jakości ocenianej testem log-rank. Estymator bazowy generuje funkcję przeżycia poprzez kombinację reguł z wykorzystaniem estymatorów Kaplana-Meiera, co umożliwia późniejszą agregację na poziomie zespołu.

Estymatory bazowe trenowane są równolegle, każdy niezależnie na odrębnej próbce bootstrap. W wyniku tego procesu każdy z nich generuje zbiór reguł decyzyjnych, które następnie konwertowane są do formatu umożliwiającego interpretację i analizę. Dodatkowo obliczane są metryki predykcyjne dla poszczególnych

estymatorów, co pozwala monitorować ich jakość. Faza ta odpowiada etapowi trenowania estymatorów w Algorytmie 7.

Faza 3: Agregacja predykcji zespołu Predykcja zespołu wymaga połączenia wyników wszystkich estymatorów bazowych poprzez uśrednianie funkcji przeżycia, zgodnie z Algorytmem 8. Każdy estymator może generować funkcje przeżycia zdefiniowane w różnych punktach czasowych. Agregacja rozpoczyna się od konstrukcji wspólnej dziedziny czasowej poprzez zebranie wszystkich unikalnych punktów czasowych ze wszystkich estymatorów i uporządkowania ich w porządku rosnącym. Następnie każda funkcja przeżycia jest interpolowana na wspólną dziedzinę za pomocą funkcji schodkowej z obsługą wartości brzegowych. Na końcu obliczana jest średnia arytmetyczna wszystkich interpolowanych funkcji. Jeśli funkcja przeżycia nie spada poniżej poziomu 0.5, mediana czasu przeżycia nie może być wyznaczona. Implementacja używa wówczas ostatniego dostępnego czasu z krzywej przeżycia. Natomiast gdy estymator nie jest w stanie wygenerować predykcji dla określonej próbki (np. z powodu braku pokrycia przez reguły), algorytm wykorzystuje predykcje z pozostałych estymatorów, co zwiększa stabilność zespołu.

Model raportuje szereg artefaktów wspierających interpretację i walidację ekspercką. Obejmują one: ranking atrybutów oparty na częstości ich występowania w regułach ważonej pokryciem i wkładem do predykcji, ranking reguł według pokrycia i wpływu na krzywą zespołu, rozkład długości reguł i udział niepokrytych przykładów, wizualizacje krzywych — średniej zespołu oraz zbioru krzywych składowych po interpolacji. Interpretowalność metody wynika z jej fundamentu opartego na regułach — każda reguła decyzyjna w zespole reprezentuje warunek logiczny w postaci „jeśli $w_1 \wedge w_2 \wedge \cdots \wedge w_n$ to $\hat{S}(t)$", który jest zrozumiały i możliwy do zweryfikowania przez ekspertów dziedzinowych.

Skuteczność metody ocenia się miarą specyficzną dla analizy przeżycia — zintegrowanym wskaźnikiem Briera liczonym na horyzoncie $[0, \tau]$, gdzie τ oznacza maksymalny czas obserwacji. Ewaluację prowadzi się w schemacie k-krotnej walidacji krzyżowej, a porównania wykonuje względem klasycznych metod (np. Cox, estymator Kaplana-Meiera), przy zachowaniu tego samego horyzontu czasowego i identycznego schematu walidacji.

Algorytm 7 Trenowanie przeżyciowego zespołu reguł

procedure TrenujZespół(D, M)
  Wejście:
    $D(A, T, \delta)$ — zbiór danych opisany atrybutami A, czasem obserwacji T i statusem przeżycia δ
    M — liczba estymatorów
  Wyjście: zespół estymatorów $\mathcal{E}$
  $\mathcal{E} \leftarrow \varnothing$
  for $i = 1$ **to** M **do**
    $(D_i) \leftarrow$ PróbkaBootstrap(D)
    $\mathcal{M}_i \leftarrow$ RegułyPrzeżycia()
    $\mathcal{M}_i$.trenuj(D_i)
    Dodaj $\mathcal{M}_i$ do $\mathcal{E}$ (jeśli trenowanie się powiodło)
  end for
  return $\mathcal{E}$
end procedure
procedure Predykcja($\mathbf{x}$)
  Wejście:
    $\mathbf{x}$ — próbka
  Wyjście: agregowana funkcja przeżycia $\hat{S}(t)$
  Zbierz predykcje: $\{S_1(t), S_2(t), ..., S_k(t)\} \leftarrow \{\mathcal{M}_i.\text{predykuj}(\mathbf{x})\}$
  $\hat{S}(t) \leftarrow$ AgregujFunkcjePrzeżycia($\{S_i(t)\}$)
  return $\hat{S}(t)$
end procedure

Algorytm 8 Agregacja funkcji przeżycia

procedure AgregujFunkcjePrzeżycia($\{S_i(t)\}_{i=1}^{k}$)
  Wejście: funkcje przeżycia $\{S_1(t), S_2(t), \ldots, S_k(t)\}$
  Wyjście: agregowana funkcja przeżycia $\hat{S}(t)$
  Utwórz wspólną dziedzinę czasową ze wszystkich funkcji
  $\mathcal{T} \leftarrow$ sortuj($\bigcup_{i=1}^{k}$ czasy(S_i))
  for każda funkcja $S_i(t)$ **do**
    Interpoluj $S_i(t)$ na punkty $\mathcal{T}$ (funkcja schodkowa)
  end for
  $\hat{S}(t) \leftarrow$ średnia($\{S_1(t), S_2(t), \ldots, S_k(t)\}$) na $\mathcal{T}$
  return $\hat{S}(t)$
end procedure

5.5.3 Ilustracja działania metody

W poniższej sekcji przedstawiono zastosowanie interpretowalnego zespołu reguł przeżyciowych na zbiorze danych *NiCdBattery*, dotyczącym żywotności akumulatorów niklowo-kadmowych. Zbiór obejmuje 87 obserwacji (78 treningowych, 9 testowych), opisanych za pomocą ośmiu cech numerycznych: czas ładowania (*charge_time*), czas rozładowania (*discharge_time*), głębokość rozładowania (*discharge_depth*), czas wstępnego ładowania (*precharge_time*), temperatura (*degrees_c*), stężenie KOH (*koh_concentration*), objętość KOH (*koh_volume*) oraz poziom doładowania (*recharge_level*). Odsetek zdarzeń w zbiorze treningowym wynosi 92.3%, a w testowym 100%. Odpowiadające mediany czasu przeżycia to odpowiednio 3042 i 1810 cykli.

Analizę przeprowadzono dla próbki testowej o indeksie 0, charakteryzującej się następującymi wartościami atrybutów: *discharge_depth* = 80.0, *discharge_time* = 1.0, *charge_time* = 2.0, *recharge_level* = 140.0, *koh_concentration* = 34.0, *koh_volume* = 20.5, *precharge_time* = 3.0 oraz *degrees_c* = 50.0. Dla tej próbki zaobserwowano zdarzenie (awarię akumulatora) po 964 cyklach pracy.

Zespół utworzono ze 100 estymatorów bazowych, wykorzystując strategię próbkowania *bootstrap* oraz losowy wybór podzbioru cech. Parametry indukcji reguł ustawiono następująco: minimalne wsparcie $\sigma = 5.0$, włączone przycinanie oraz ignorowanie braków danych.

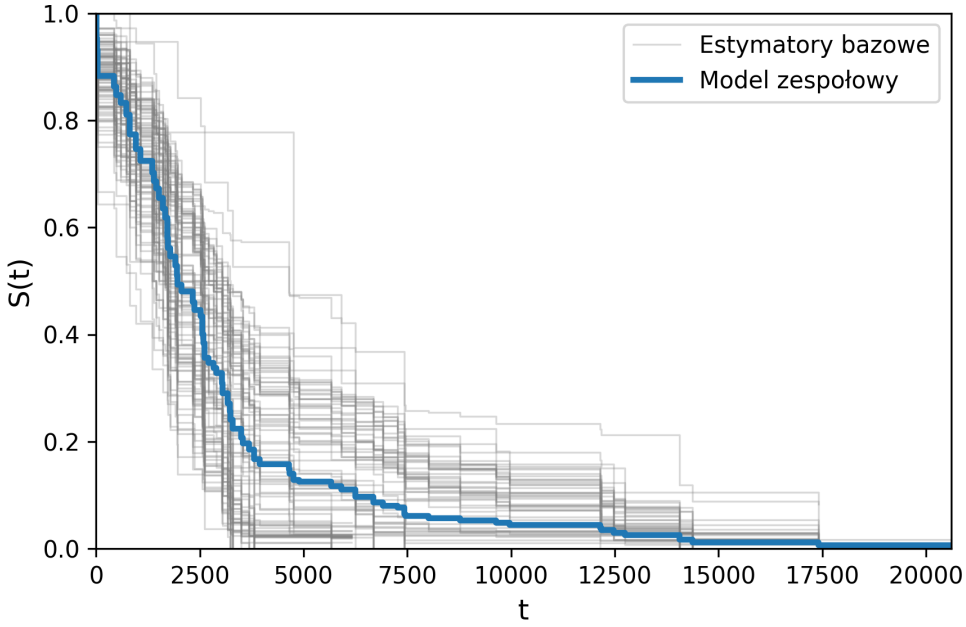
Faza 1: Próbkowanie bootstrap i losowy wybór podzbioru cech Dla każdego ze 100 estymatorów bazowych przeprowadzono niezależne próbkowanie *bootstrap* ze zbioru treningowego liczącego 78 obserwacji oraz losowy wybór 4 cech spośród 8 dostępnych. Przykładowo, estymator 1 otrzymał cechy: *charge_time*, *discharge_time*, *koh_concentration*, *degrees_c*, podczas gdy estymator 2: *charge_time*, *degrees_c*, *koh_volume*, *recharge_level*. Każda próbka bootstrap zawierała 78 obserwacji z powtórzeniami. Strategia ta zapewniła różnorodność estymatorów przy jednoczesnym ograniczeniu korelacji między nimi.

Faza 2: Indukcja reguł i trening Każdy estymator bazowy indukował reguły przeżyciowe za pomocą algorytmu *separate-and-conquer*. Proces wzrostu reguł przebiegał zachłannie — w każdej iteracji dodawano warunek maksymalizujący statystykę log-rank między przykładami pokrytymi a niepokrytymi przez regułę. Po zakończeniu procesu wzrostu stosowano przycinanie, usuwając warunki, których eliminacja nie pogarszała jakości reguły.

Trenowanie przeprowadzono równolegle dla wszystkich 100 estymatorów bazowych. Wszystkie estymatory zostały pomyślnie wytrenowane, generując łącznie 221 unikalnych reguł. Średnia liczba reguł na estymator wyniosła 2.21, przy czym rozkład wahał się od 1 do 5 reguł. Każdą regułę powiązano z krzywą przeżycia Kaplana-Meiera, estymowaną na podstawie pokrywanych przez nią przykładów.

Faza 3: Agregacja predykcji zespołu Predykcja zespołu dla próbki testowej wymagała agregacji funkcji przeżycia ze wszystkich estymatorów bazowych. Proces rozpoczęto od konstrukcji wspólnej dziedziny czasowej, zbierając wszystkie unikalne punkty czasowe ze wszystkich funkcji przeżycia, co dało 73 punkty w zakresie od 12 do 20605 cykli. Następnie każdą funkcję interpolowano na tę dziedzinę za pomocą funkcji schodkowej z obsługą wartości brzegowych.

Końcową krzywą przeżycia uzyskano poprzez obliczenie średniej arytmetycznej wszystkich interpolowanych funkcji w każdym punkcie czasowym. Na Rysunku 5.4 zaprezentowano zarówno indywidualne funkcje przeżycia ze wszystkich 100 estymatorów bazowych, jak i ich zagregowaną funkcję zespołu.



Rysunek 5.4: Funkcje przeżycia dla przykładu ze zbioru *NiCdBattery* wygenerowane przez zespół 100 zbiorów reguł przeżyciowych. Wykres przedstawia indywidualne funkcje przeżycia ze wszystkich estymatorów bazowych (oznaczone kolorem szarym) oraz model zespołowy (oznaczony kolorem niebieskim). Różnorodność krzywych indywidualnych ilustruje wariancję predykcji poszczególnych estymatorów, podczas gdy model zespołowy stanowi ich uśrednioną predykcję.

6 Eksperymenty i przypadki użycia

Niniejszy rozdział przedstawia empiryczną walidację zaproponowanych metod. Eksperymenty koncentrują się na weryfikacji skuteczności podejść zastosowanych do danych medycznych i przemysłowych, ukazując użyteczność interpretowalnych algorytmów indukcji reguł w analizie danych cenzurowanych.

W eksperymentach oceniono cztery opracowane algorytmy: pokryciowy algorytm indukcji przeżyciowych reguł akcji, algorytm rekomendacji przeżyciowych reguł akcji, algorytm indukcji przeżyciowych reguł wyjątków oraz interpretowalny zespół reguł przeżyciowych. Zostały one przetestowane na zróżnicowanych zbiorach danych, odzwierciedlających typową dla analizy przeżycia różnorodność zastosowań — od danych przemysłowych związanych z funkcjonowaniem maszyn w predykcyjnym utrzymaniu ruchu, po zbiory medyczne pochodzące z badań klinicznych i epidemiologicznych. Skuteczność oceniono na podstawie kryteriów powszechnie stosowanych w analizie przeżycia, takich jak test log-rank oraz wskaźnik Briera.

Treść rozdziału obejmuje metodykę testowania, kryteria oceny algorytmów oraz charakterystykę wykorzystanych zbiorów danych. Następnie przedstawiono szczegółowe analizy eksperymentów dla każdego z opracowanych algorytmów. Poprzez porównanie z metodami referencyjnymi oraz analizę przypadków użycia, rozdział prezentuje wyniki badań i ilustruje zastosowanie interpretowalnych metod analizy przeżycia w zadaniach wymagających transparentności procesów decyzyjnych.

6.1 Kryteria oceny

Ocena skuteczności algorytmów w analizie przeżycia wymaga zastosowania kryteriów uwzględniających specyfikę danych cenzurowanych. W eksperymentach wykorzystano zestaw uzupełniających się metryk, które pozwalają zarówno na ocenę dokładności prognozowania funkcji przeżycia, jak i na analizę jakości odkrywanej wiedzy w postaci reguł.

W badaniach wykorzystano wskaźnik Briera do oceny metod generujących bezpośrednie predykcje, takich jak algorytm rekomendacji przeżyciowych reguł akcji oraz interpretowalny zespół reguł przeżyciowych. Dla metod eksploracyjnych, którymi są pokryciowy algorytm indukcji przeżyciowych reguł akcji oraz algorytm indukcji przeżyciowych reguł wyjątków, zastosowano statystykę log-rank oraz metryki opisujące strukturę i jakość wygenerowanych reguł. Dodatkowo, dla wszystkich typów reguł przeprowadzono jakościową weryfikację działania algorytmu w oparciu o wizualizację krzywych przeżycia, przy czym ocena końcowa opiera się na metrykach ilościowych.

6.1.1 Wskaźnik Briera

Wskaźnik Briera [128] stanowi jedną z miar oceny dokładności modeli prognostycznych w analizie przeżycia. Jest rozszerzeniem błędu średniokwadratowego na dane cenzurowane i umożliwia ocenę predykcji ryzyka w obecności niepełnych obserwacji. Przyjmuje wartości z przedziału $[0, 1]$, gdzie 0 oznacza perfekcyjną predykcję, a 1 najgorszą możliwą jakość predykcji. Wskaźnik Briera ocenia kalibrację modelu (ang. *calibration*), czyli zgodność między predykowanymi prawdopodobieństwami a rzeczywistymi częstotliwościami zdarzeń. Znajduje zastosowanie głównie w ocenie metod generujących bezpośrednie predykcje funkcji przeżycia. Dla i-tej obserwacji w chwili t obliczany jest zgodnie ze wzorem:

$$\mathrm{BS}_i(t) = (S_i(t) - I(T_i > t))^2 \tag{6.1}$$

gdzie $S_i(t)$ oznacza predykowane prawdopodobieństwo przeżycia dla i-tej obserwacji w chwili t, a $I(T_i > t)$ to rzeczywisty status przeżycia (1, jeśli dla obserwacji nie zanotowano zdarzenia do czasu t, 0 w przeciwnym przypadku).

W przypadku danych z obserwacjami cenzurowanymi, Graf i in. [129] zaproponowali modyfikację wskaźnika Briera, która uwzględnia obecność danych cenzurowanych poprzez wprowadzenie wagi odwrotnie proporcjonalnej do prawdopodobieństwa nieocenzurowania:

$$\mathrm{BS}_i^c(t) = \frac{(S_i(t) - I(T_i > t))^2}{\hat{G}(T_i)} \tag{6.2}$$

gdzie $\hat{G}(T_i)$ to estymator Kaplana-Meiera funkcji przeżycia dla rozkładu cenzurowania, obliczany na zbiorze treningowym z wykorzystaniem odwróconego

wskaźnika zdarzenia, tj. przyjmując $1 - \delta_i$ jako wskaźnik cenzurowania, gdzie δ_i oznacza standardowy wskaźnik zdarzenia ($\delta_i = 1$ dla obserwacji z zaobserwowanym zdarzeniem, $\delta_i = 0$ dla obserwacji cenzurowanych).

Definiuje się również całkowy wskaźnik Briera (ang. *Integrated Brier Score*, IBS), który stanowi uogólnienie wskaźnika Briera na cały przedział czasowy i jest obliczany jako średnia ważona wskaźników Briera dla wszystkich czasów obserwacji:

$$\mathrm{IBS} = \frac{1}{n \cdot |T|} \sum_{i=1}^{n} \sum_{t \in T} \mathrm{BS}_i^c(t) \tag{6.3}$$

gdzie n oznacza liczbę obserwacji, T to zbiór wszystkich unikalnych czasów obserwacji, a $|T|$ oznacza liczbę elementów tego zbioru. Wartość IBS równa 0 oznacza perfekcyjny model, natomiast wartość 0.25 odpowiada modelowi losowemu, który zawsze zwraca prawdopodobieństwo 0.5. Model uznaje się za użyteczny, gdy jego wartość IBS jest niższa od 0.25.

Ograniczeniem wskaźnika Briera jest fakt, że może być stosowany wyłącznie do modeli estymujących funkcję przeżycia. Nie nadaje się więc do oceny metod, które nie generują bezpośrednich estymatów prawdopodobieństwa przeżycia, takich jak maszyny wektorów nośnych dla analizy przeżycia (ang. *Survival Support Vector Machines*) [129].

6.1.2 Kryteria oceny jakości reguł

Ocena modeli eksploracyjnych, takich jak pokryciowy algorytm indukcji przeżyciowych reguł akcji oraz algorytm indukcji przeżyciowych reguł wyjątków, wymaga użycia kryteriów dostosowanych do faktu, że wynikiem są zbiory reguł, a nie bezpośrednie predykcje. Z tego względu stosuje się miary odnoszące się do struktury, złożoności i pokrycia reguł, zamiast klasycznych metryk predykcyjnych opartych na estymatach funkcji przeżycia. W porównaniu ze standardowymi metrykami analizy przeżycia, metody eksploracyjne wymagają kryteriów uwzględniających zarówno jakość odkrywanej wiedzy, jak i interpretowalność generowanych reguł.

Złożoność i rozmiar zbioru reguł Podstawowym kryterium oceny modelu regułowego jest jego wielkość, wyrażona jako liczba reguł wchodzących w jego skład — ozn. $|\mathcal{R}|$, gdzie $\mathcal{R}$ to zbiór wygenerowanych reguł. Mniejsza liczba reguł jest pożądana, gdyż zwiększa interpretowalność modelu.

Dodatkowym kryterium, mierzącym złożoność generowanych reguł, jest średnia długość reguł w zbiorze reguł, rozumiana jako średnia liczba warunków elementarnych w przesłankach reguł:

$$\overline{w} = \frac{1}{|\mathcal{R}|} \sum_{r \in \mathcal{R}} |r|, \tag{6.4}$$

gdzie $|r|$ oznacza liczbę warunków elementarnych w przesłance reguły r (por. Rozdział 5). Krótsze reguły są na ogół preferowane, ponieważ są bardziej zrozumiałe i mniej podatne na przeuczenie.

Średnie pokrycie zbioru treningowego przez reguły stanowi kolejne kryterium oceny:

$$\overline{C} = \frac{1}{|\mathcal{R}|} \sum_{r \in \mathcal{R}} \frac{|X_r|}{n}, \tag{6.5}$$

gdzie X_r oznacza zbiór obserwacji pokrywanych przez regułę r, a n to liczba obserwacji w zbiorze treningowym. Wskaźnik $\overline{C}$ określa, jaką część zbioru treningowego pokrywa średnio każda reguła. Wysokie wartości świadczą o dużym uogólnieniu reguł, natomiast wartości bardzo bliskie 1 mogą wskazywać na nadmierną ogólność.

Struktura akcji W ocenie reguł akcji wykorzystuje się również miary opisujące strukturę akcji. Stosunek liczby akcji do liczby warunków ($|\mathcal{A}|/|\mathcal{W}|$) określa relację częstości akcji do liczby elementarnych warunków w regule. Udział akcji dowolnych ($|\mathcal{A}_d|/|\mathcal{A}|$) wskazuje, jaki odsetek wszystkich akcji stanowią akcje dowolne, tj. niewyznaczające konkretnego kierunku zmiany wartości atrybutu. Średnia liczba warunków i akcji na regułę ($\overline{w}$ i $\overline{a}$) oraz ich zakresy (minimum m_w, maksimum M_w dla warunków; minimum m_a, maksimum M_a dla akcji) służą do oceny złożoności reguł. Niższe wartości $\overline{w}$ i $\overline{a}$ oznaczają mniejszą złożoność, natomiast skrajnie wysokie M_w lub M_a mogą wskazywać na obecność pojedynczych nadmiernie złożonych reguł.

W zestawieniach tabelarycznych stosowane są następujące oznaczenia: $|\mathcal{R}|$ — liczba reguł w zbiorze $\mathcal{R}$, $|\mathcal{W}|$ — całkowita liczba warunków w zbiorze reguł, $|\mathcal{A}|$ — całkowita liczba akcji, $|\mathcal{A}_d|$ — liczba akcji dowolnych. Średnia liczba akcji dowolnych na regułę oznaczana jest jako $\overline{a}_d$. Pokrycie części źródłowej i docelowej opisują wskaźniki $|X_{r_Z}|$ i $|X_{r_D}|$, definiowane jako średnie (po regułach)

odsetki obserwacji spełniających odpowiednio przesłankę oraz część docelową reguły względem liczebności zbioru.

Istotność statystyczna Kolejnym kryterium oceny jakości reguł jest odsetek reguł istotnych statystycznie, wyznaczany na podstawie testu log-rank. W przypadku reguł akcji test porównuje krzywe przeżycia części źródłowej i docelowej danej reguły, natomiast dla reguł wyjątków — krzywą przeżycia obserwacji pokrywanych przez regułę z krzywą dla pozostałych obserwacji. Test log-rank polega na porównaniu obserwowanych i oczekiwanych liczebności zdarzeń w grupach definiowanych przez regułę. Statystyka testowa ma rozkład chi-kwadrat z jednym stopniem swobody. Procent reguł istotnych na poziomie $\alpha = 0.05$ oblicza się według wzoru:

$$\%_{\text{istotnych}} = \frac{|\{r \in \mathcal{R} : p < \alpha\}|}{|\mathcal{R}|} \times 100\% \tag{6.6}$$

Wyższe wartości tego wskaźnika oznaczają większy udziale reguł, dla których test log-rank wykazuje istotne statystycznie różnice między porównywanymi krzywymi przeżycia.

Na tej podstawie raportowane są wskaźniki $P_{0.05}$ i $P_{0.01}$, czyli odpowiednio procent reguł w zbiorze spełniających warunek istotności $p < 0.05$ lub $p < 0.01$. Uogólniając, P_α oznacza odsetek reguł z wartością $p < \alpha$.

Metryki dla algorytmu rekomendacji W ocenie algorytmu rekomendacji wykorzystano dodatkowo wskaźnik spójności S_α, definiowany jako odsetek przykładów testowych, dla których różnica między krzywymi przeżycia przed i po zastosowaniu rekomendowanych zmian jest istotna statystycznie na poziomie α (test log-rank). Metryka MAE (ang. *mean absolute error*) mierzy średni bezwzględny błąd między estymacją czasu przeżycia uzyskaną przez algorytm rekomendacji a estymacją niezależnego modelu walidacyjnego, gdzie niższe wartości wskazują na większą zgodność. Pokrycie, definiowane jako odsetek przykładów testowych, dla których wygenerowano co najmniej jedną rekomendację, informuje, dla jakiej części przypadków algorytm zwraca rekomendację. Wysokie pokrycie oznacza, że rekomendacje są generowane dla większości przypadków, z kolei niskie pokrycie oznacza, że rekomendacja pojawia się tylko dla niewielkiej ich części. W tej części wyników symbol $\overline{a}$ odnosi się do średniej liczby akcji na rekomendację (a nie

na regułę), natomiast wskaźnik $\overline{q}$ oznacza średni odsetek atrybutów poddanych modyfikacji w pojedynczej rekomendacji względem liczby dostępnych atrybutów.

Kierunek wpływu i mediana czasu przeżycia W przypadku reguł akcji dodatkowym kryterium oceny jest liczba reguł polepszających ($|R_+|$) i pogarszających ($|R_-|$), definiowanych jako reguły, dla których krzywa przeżycia części docelowej znajduje się odpowiednio powyżej lub poniżej krzywej części źródłowej. Informacja ta pozwala ocenić kierunek wpływu generowanych akcji na funkcję przeżycia. Przewaga $|R_+|$ nad $|R_-|$ wskazuje na dominację akcji poprawiających przeżycie. Relacja ta zależy od ustawienia parametru τ, który określa preferowany typ reguł („lepsza", „gorsza" lub „dowolna").

W ocenie reguł wykorzystywana jest również mediana czasu przeżycia. Stosowane są dwie uzupełniające się definicje. Pierwsza określa medianę jako czas t, dla którego funkcja przeżycia S przyjmuje wartość 0.5, o ile istnieje czas t spełniający $S(t) = 0.5$. W przeciwnym razie mediana pozostaje niezdefiniowana. Druga, stosowana dla estymatora Kaplana-Meiera, definiuje medianę jako najmniejszy czas t, dla którego $\hat{S}_{\mathrm{KM}}(t) \leqslant 0.5$, co uwzględnia skokowy charakter tego estymatora. Obie definicje dotyczą danych cenzurowanych, przy czym druga precyzuje sposób wyznaczania mediany na podstawie empirycznej krzywej Kaplana-Meiera. Mediana czasu przeżycia pełni funkcję pomocniczą w ocenie, pozwalając na porównanie charakterystyki przeżycia między różnymi grupami obserwacji. Wyższa mediana dla części docelowej względem źródłowej wskazuje na wyższe prawdopodobieństwo przeżycia w grupie docelowej.

Wizualizacja krzywych przeżycia stanowi dodatkowe narzędzie oceny, szczególnie przydatne w przypadku reguł akcji. Porównanie krzywych przeżycia części źródłowej i docelowej pozwala na intuicyjną ocenę skuteczności generowanych akcji. Choć nie jest to formalne kryterium, wizualizacja odgrywa ważną rolę w weryfikacji działania algorytmów oraz jakościowej analizie metod, ułatwiając identyfikację potencjalnych problemów w generowaniu reguł.

6.2 Zbiory danych

W badaniach eksperymentalnych wykorzystano dwie kategorie zbiorów danych dotyczących przeżycia. Pierwszą kategorię stanowią zbiory przemysłowe, związane

z funkcjonowaniem maszyn i urządzeń, drugą natomiast — zbiory medyczne pochodzące z badań klinicznych i epidemiologicznych.

nazwa zbioru	$\|A\|$	$\|A_n\|$	$\|A_k\|$	n	cenz (%)
AdhesiveBondB [29]	2	2	0	104	5.8
Aircraft [1]	8	5	3	3135	99.3
KevlarVessel [29]	2	2	0	108	10.2
LaminatePanel [29]	1	1	0	125	8.0
LEDLife [29]	2	2	0	167	9.4
Maintenance [130]	5	3	2	1000	60.3
NewSpring [29]	3	2	1	108	32.4
NiCdBattery [29]	8	8	0	87	6.9
PM [2]	9	9	0	1274	91.7
Tantalum [29]	2	2	0	2204	98.2
ZelenCap [29]	2	2	0	64	50.0

Tabela 6.1: Charakterystyka zbiorów przemysłowych wykorzystanych w badaniach eksperymentalnych. W kolejnych kolumnach przedstawiono: nazwę zbioru, liczbę atrybutów ($|A|$), liczbę atrybutów numerycznych ($|A_n|$), liczbę atrybutów kategorycznych ($|A_k|$), liczbę obserwacji (n), udział obserwacji cenzurowanych (cenz, w %). We wszystkich zbiorach z tej grupy odsetek wartości brakujących wynosi 0%.

W badaniach wykorzystano 11 zbiorów danych dotyczących niezawodności maszyn i urządzeń, których charakterystykę przedstawiono w Tabeli 6.1. Analizowane zbiory są znacząco zróżnicowane pod względem liczby obserwacji, liczby atrybutów warunkowych oraz proporcji obserwacji cenzurowanych, przy czym w żadnym z nich nie występują wartości brakujące. Liczba obserwacji waha się od 64 do 3135, liczba atrybutów od 1 do 9, a udział obserwacji cenzurowanych wynosi od 5.8% do 99.3%. Niektóre zmienne, takie jak identyfikator maszyny, nie mają wartości prognostycznej, dlatego zostały wykluczone z analizy i nie są uwzględnione w Tabeli 6.1.

Zbiory przemysłowe wybrane do analizy pochodzą z trzech głównych kategorii źródeł. Najliczniejszą grupę stanowią dane pochodzące z publikacji [29], obejmujące osiem zbiorów: *AdhesiveBondB*, *KevlarVessel*, *LaminatePanel*, *LEDLife*,

nazwa zbioru	temat
AdhesiveBondB	przyspieszone testy degradacji spoin klejowych
Aircraft	symulacje IoT urządzeń wirujących
KevlarVessel	próby *creep-rupture* naczyń Kevlar/epoksyd
LaminatePanel	testy wytrzymałości paneli laminowanych
LEDLife	testy żywotności diod LED
Maintenance	awarie maszyn na podstawie sensorów IoT
NewSpring	testy wytrzymałości sprężyn
NiCdBattery	testy żywotności akumulatorów NiCd
PM	predykcja awarii z danych sensorów
Tantalum	testy żywotności kondensatorów tantalowych
ZelenCap	testy żywotności kondensatorów ceramicznych

Tabela 6.2: Tematyka zbiorów przemysłowych wykorzystanych w badaniach eksperymentalnych. W kolejnych kolumnach przedstawiono: nazwę zbioru oraz tematykę zbioru.

NewSpring, *NiCdBattery*, *Tantalum* oraz *ZelenCap*. Zbiory te reprezentują klasyczne przypadki analizy niezawodności w środowisku przemysłowym i dotyczą zróżnicowanych materiałów i komponentów — od klejów strukturalnych i kompozytów kevlarowych, przez półprzewodniki i baterie, po komponenty metalurgiczne.

Drugą kategorię tworzą zbiory pochodzące z platform udostępniania danych. Zbiór *Aircraft*[1] został udostępniony na platformie Kaggle i stanowi przykład otwartych danych przemysłowych związanych z predykcyjnym utrzymania ruchu komponentów lotniczych. Charakteryzuje się on złożoną strukturą telemetryczną, obejmującą pomiary napięcia, obrotów, ciśnienia oraz wibracji rejestrowane w czasie rzeczywistym. Zbiór *PM*[2], również dostępy na Kaggle, bazuje na danych symulacyjnych.

Trzecią kategorię reprezentuje zbiór *Maintenance* [130], pochodzący z biblioteki *PySurvival* dedykowanej do analizy przeżycia.

Wszystkie analizowane zbiory danych są publicznie dostępne, co sprzyja reprodukowalność wyników i umożliwia empiryczne porównania z alternatywnymi meto-

[1] `https://www.kaggle.com/datasets/arnabbiswas1/microsoft-azure-predictive-maintenance`
[2] `https://www.kaggle.com/datasets/hiimanshuagarwal/predictive-maintenance-dataset`

zbiór	$\|A\|$	$\|A_n\|$	$\|A_k\|$	n	brak (%)	cenz (%)
actg320 [57]	11	4	7	1151	0.00	91.7
BMT-Ch [131]	35	7	28	187	1.24	54.5
cancer [132]	7	6	1	228	4.14	27.6
follic [133]	4	2	2	541	0.00	35.7
GBSG2 [134]	8	5	3	686	0.00	56.4
hd [133]	6	1	5	865	0.00	50.8
lung [135]	7	0	7	1032	2.60	26.0
Melanoma [136]	7	2	5	205	0.00	65.4
mgus [137]	9	7	2	241	19.59	23.7
pbc [138]	17	10	7	418	14.54	61.5
std [139]	21	2	19	877	0.00	60.4
uis [57]	13	7	6	575	0.00	19.3
wcgs [140]	10	7	3	3154	0.04	91.9
whas1 [57]	7	2	5	481	0.00	48.2
whas500 [57]	13	5	8	500	0.00	57.0
zinc [141]	55	8	47	431	57.17	81.2

Tabela 6.3: Charakterystyka medycznych zbiorów danych wykorzystanych w badaniach eksperymentalnych. W kolejnych kolumnach przedstawiono: nazwę zbioru, liczbę atrybutów ($|A|$), liczbę atrybutów numerycznych ($|A_n|$), liczbę atrybutów kategorycznych ($|A_k|$), liczbę obserwacji (n), udział wartości brakujących (brak, w %), udział obserwacji cenzurowanych (cenz, w %).

zbiór	temat
actg320	pacjenci zakażeni HIV
BMT-Ch	przeszczep szpiku kostnego
cancer	zaawansowany rak płuc
follic	chłoniak z komórek pęcherzykowych
GBSG2	rak piersi
hd	choroba Hodgkina
lung	wczesne wykrywanie raka płuc
Melanoma	złośliwy czerniak
mgus	łagodna gammapatia monoklonalna
pbc	pierwotne zapalenie dróg żółciowych
std	choroby przenoszone drogą płciową
uis	leczenie uzależnień
wcgs	choroba wieńcowa
whas1	zawał mięśnia sercowego (wersja 1)
whas500	zawał mięśnia sercowego (wersja 2)
zinc	rak przełyku

Tabela 6.4: Tematyka medycznych zbiorów danych wykorzystanych w badaniach eksperymentalnych. W kolejnych kolumnach przedstawiono: nazwę zbioru oraz tematykę zbioru.

dami. W odróżnieniu od standardowych zbiorów przeżyciowych, które w większości dotyczą badań medycznych, wybrane zbiory danych odnoszą się do niezawodności systemów przemysłowych.

Dodatkowo w eksperymentach wykorzystano 16 zbiorów danych medycznych, których charakterystykę przedstawiono w Tabeli 6.3. Zbiory te pochodzą z różnych dziedzin medycyny, w tym onkologii, kardiologii, transplantologii oraz badań nad HIV. Liczba obserwacji w zbiorach medycznych wahała się od 187 do 3154, liczba atrybutów od 4 do 55, a udział obserwacji cenzurowanych wynosił od 19.3% do 91.9%. W przeciwieństwie do zbiorów przemysłowych, niektóre zbiory medyczne zawierały brakujące wartości, których udział wynosił od 0.04% do 57.17%.

Zbiory medyczne wybrano ze względu na ich różnorodność pod względem rodzaju schorzeń, wielkości próby oraz struktury danych. Większość z nich stanowi standardowe zbiory referencyjne używane w literaturze dotyczącej analizy przeżycia, co umożliwia porównanie wyników z innymi metodami. Pochodzą głównie z publikacji naukowych oraz specjalistycznych pakietów statystycznych.

Przygotowanie zbiorów danych do analizy przeżycia wymagało zastosowania procedur przetwarzania, dostosowanych do specyfiki poszczególnych źródeł danych oraz ich pierwotnej struktury. Można wyróżnić cztery główne kategorie takich procesów, stosowane w zależności od charakterystyki źródłowych zbiorów danych.

Dla ośmiu zbiorów pochodzących z repozytorium danych niezawodnościowych (*AdhesiveBondB*, *KevlarVessel*, *LEDLife*, *LaminatePanel*, *NewSpring*, *NiCdBattery*, *Tantalum*, *ZelenCap*) zastosowano jednolitą procedurę przetwarzania. Polegała ona na przekształceniu kategorycznego wskaźnika cenzurowania na binarny status przeżycia: zdarzenia końcowe zakodowano jako 1, a obserwacje cenzurowane jako 0. W przypadku zbioru *AdhesiveBondB* wartość „Exact" wskazywała na wystąpienie zdarzenia, natomiast dla pozostałych zbiorów stosowano oznaczenia „Failed" lub „Failure". Dodatkowo, w zbiorach zawierających kolumnę „Count", przeprowadzono ekspansję danych poprzez powielenie wierszy zgodnie z liczbą powtórzeń, po czym usunięto pierwotną kolumnę licznikową.

Zbiór *Aircraft* wymagał najbardziej złożonej procedury przetwarzania ze względu na wielowarstwową strukturę telemetryczną. Proces obejmował integrację czterech źródeł danych: pomiarów telemetrycznych, zdarzeń konserwacyjnych, awarii oraz charakterystyk maszyn. Utworzono interwały czasowe dla każdej kombinacji maszyny i komponentu. Początek interwału stanowiło rozpoczęcie obserwacji lub

poprzednie zdarzenie konserwacyjne, a koniec — awaria lub konserwacja prewencyjna. Dla każdego interwału obliczono czas przeżycia w godzinach oraz dokonano agregacji statystycznej zmiennych telemetrycznych (napięcie, obroty, ciśnienie, wibracje) poprzez uśrednienie wartości w ramach interwału. Zastosowano filtry jakości danych, eliminując interwały krótsze niż 2 godziny lub nieposiadające odpowiadających pomiarów telemetrycznych.

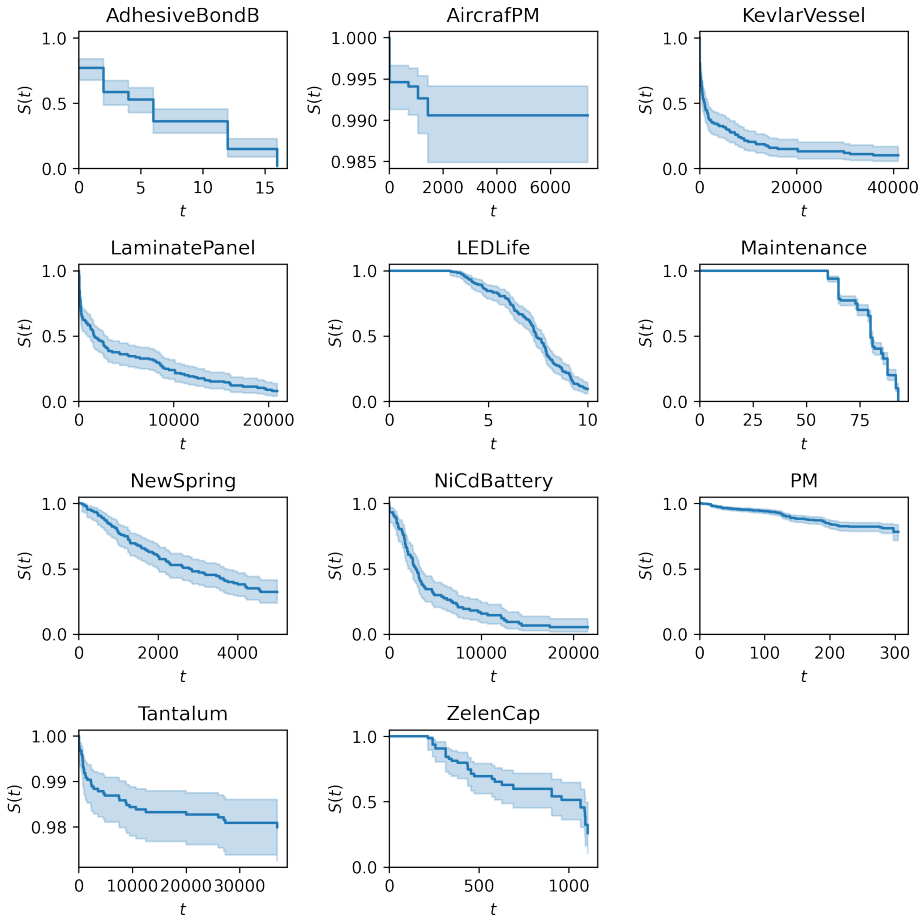
Zbiór *PM* przetworzono za pomocą agregacji czasowej. Dla każdej jednostki identyfikowano wszystkie dostępne obserwacje, określano pierwszy i ostatni moment pomiaru, a następnie obliczano czas przeżycia jako różnicę między tymi punktami wyrażoną w dniach. W przypadku wystąpienia wielokrotnych zdarzeń awaryjnych dla tej samej jednostki przyjęto ostatnie zdarzenie jako definitywny punkt końcowy obserwacji. Jednostki bez zdarzeń awaryjnych otrzymały status cenzurowany,a czas przeżycia liczono od pierwszej do ostatniej obserwacji.

Dodatkowo, dla wszystkich zbiorów przeprowadzono weryfikację danych, eliminując obserwacje z ujemnymi lub zerowymi czasami przeżycia oraz identyfikatory przemysłowe nieistotne dla analizy statystycznej.

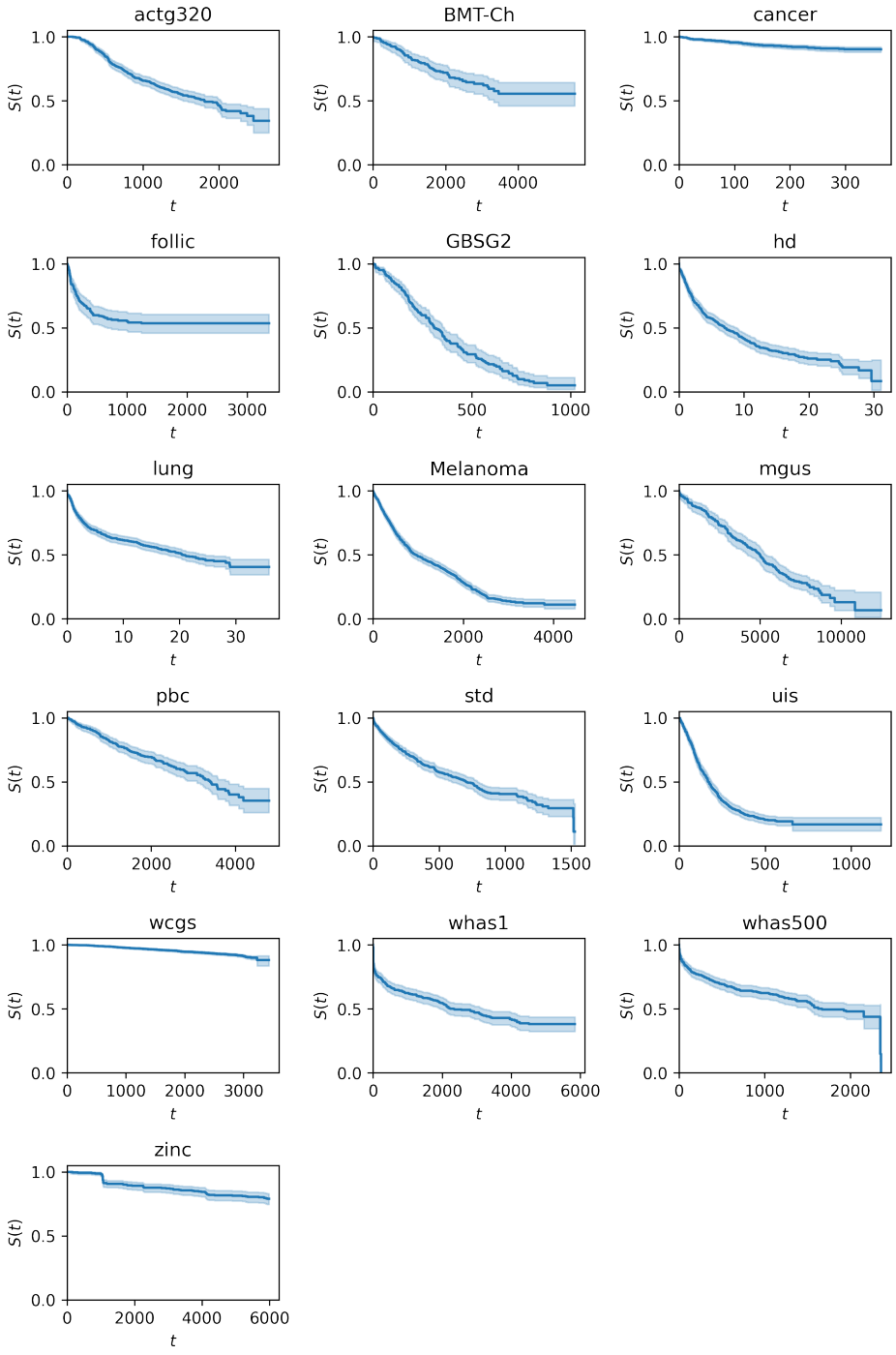
W celu lepszego zobrazowania charakterystyki analizowanych zbiorów danych, na Rysunku 6.1 i 6.2 przedstawiono empiryczne krzywe przeżycia dla wszystkich wykorzystanych zbiorów. Wizualizacja ta umożliwia identyfikację różnic w rozkładach czasów przeżycia między poszczególnymi zbiorami oraz porównanie charakterystyk między domenami przemysłową i medyczną.

Krzywe przeżycia dla zbiorów przemysłowych (Rysunek 6.1) pokazują zróżnicowaną dynamikę procesów awaryjności. Niektóre zbiory, np. *AdhesiveBondB* i *KevlarVessel*, wykazują wysoką wczesną awaryjność, podczas gdy inne, jak *Tantalum* czy *Aircraft*, charakteryzują się długotrwałą niezawodnością. Dla zbiorów *Tantalum* i *Aircraft* zakres wartości na osi rzędnych został ograniczony, aby poprawić czytelność wykresów, gdyż wartości funkcji przeżycia pozostają bliskie jedności przez większość okresu obserwacji.

Krzywe przeżycia dla zbiorów medycznych (Rysunek 6.2) wykazują odmienną charakterystykę w porównaniu ze zbiorami przemysłowymi. Większość z nich wykazuje bardziej stopniowy spadek funkcji przeżycia w czasie, co odzwierciedla naturę procesów biologicznych. Niektóre zbiory, takie jak *actg320* czy *wcgs*, charakteryzują się wysokim stopniem cenzurowania, co przekłada się na wolniejszy spadek wartości funkcji przeżycia szacowanej estymatorem Kaplana-Meiera



Rysunek 6.1: Krzywe KM dla przemysłowych zbiorów danych z Tabeli 6.1 wykorzystanych w badaniach eksperymentalnych (przedziały ufności na poziomie 95% obliczone metodą Greenwooda).



Rysunek 6.2: Krzywe KM dla medycznych zbiorów danych z Tabeli 6.3 wykorzystanych w badaniach eksperymentalnych (przedziały ufności na poziomie 95% obliczone metodą Greenwooda).

w dłuższym horyzoncie obserwacji. Inne, np. *cancer* i *uis*, wykazują bardziej dynamiczne zmiany, odpowiadające różnej agresywności schorzeń lub różnym długościom okresów obserwacji.

6.3 Pokryciowy algorytm indukcji przeżyciowych reguł akcji

W niniejszej sekcji przedstawiono wyniki eksperymentów przeprowadzonych z wykorzystaniem pokryciowego algorytmu indukcji przeżyciowych reguł akcji opisanego w Sekcji 5.2. Eksperymenty przeprowadzono na wszystkich zbiorach danych przedstawionych w Tabeli 6.1 i 6.3. Nie stosowano walidacji krzyżowej ze względu na eksploracyjny charakter metody.

W konfiguracji eksperymentów przyjęto następujące parametry: $\tau =$ „dowolna” (brak preferencji kierunku zmian krzywych przeżycia), $\mu = 30$ (minimalna liczba niepokrytych przykładów), $\xi = 0.1$ (maksymalny procent przykładów wspólnych), $\rho = 0.5$ (maksymalne pokrycie reguły).

Wybór parametru $\mu = 30$ zapewnia statystyczną istotność generowanych reguł przy jednoczesnym zachowaniu zdolności do wykrywania wzorców w mniejszych grupach. Wartość $\xi = 0.1$ ogranicza nakładanie się reguł, co zwiększa różnorodność odkrywanych wzorców. Parametr $\rho = 0.5$ zapobiega dominacji pojedynczych reguł nad całym zbiorem danych. Założono, że wszystkie atrybuty są zmienne, tzn. mogą występować w części akcyjnej reguł i podlegać zmianom. Dla zmiennych ciągłych oznacza to zawężanie lub przesuwanie przedziału wartości, a dla kategorycznych — zmianę kategorii. Nie rozróżniano atrybutów sterowalnych i niesterowalnych. Takie ujednolicenie upraszcza analizę i zapewnia porównywalność wyników między zbiorami.

Tabele 6.5 i 6.6 przedstawiają wyniki eksperymentów dla zbiorów przemysłowych, wskazujące na znaczną różnorodność zarówno w liczbie generowanych reguł, jak i w ich charakterystykach.

W podsumowaniu przedstawionym w tabelach 6.5 i 6.6 liczba warunków ($|\mathcal{W}|$) odnosi się do liczby warunków w części źródłowej reguły, a liczba akcji ($|\mathcal{A}|$) to liczba modyfikacji zdefiniowanych w regule dla atrybutów, w których część docelowa różni się od części źródłowej. Akcje podtrzymujące są traktowane jako warunki

zbiór	$\lvert\mathcal{R}\rvert$	$\lvert\mathcal{W}\rvert$	$\lvert\mathcal{A}\rvert$	$\lvert\mathcal{A}_d\rvert$	$\lvert\mathcal{A}_d\rvert/\lvert\mathcal{A}\rvert$	$\lvert\mathcal{A}\rvert/\lvert\mathcal{W}\rvert$
AdhesiveBondB	2	4	2	0	0.00	0.50
Aircraft	80	170	103	6	0.06	0.61
KevlarVessel	2	4	2	0	0.00	0.50
LaminatePanel	2	4	2	0	0.00	0.50
LEDLife	4	8	5	1	0.20	0.62
Maintenance	7	15	13	1	0.08	0.87
NewSpring	2	4	2	0	0.00	0.50
NiCdBattery	2	4	2	0	0.00	0.50
PM	31	109	104	16	0.15	0.95
Tantalum	5	10	6	1	0.17	0.60
ZelenCap	1	2	1	0	0.00	0.50

Tabela 6.5: Podstawowe statystyki wygenerowanych przeżyciowych reguł akcji dla zbiorów danych z Tabeli 6.1. W kolejnych kolumnach przedstawiono: nazwę zbioru danych, liczbę reguł ($|\mathcal{R}|$), całkowitą liczbę: warunków ($|\mathcal{W}|$), akcji ($|\mathcal{A}|$), akcji dowolnych ($|\mathcal{A}_d|$), udział akcji dowolnych ($|\mathcal{A}_d|/|\mathcal{A}|$), oraz stosunek liczby akcji do liczby warunków ($|\mathcal{A}|/|\mathcal{W}|$).

i nie uwzględnia się ich przy liczeniu akcji. Akcje podtrzymujące to takie, które nie zmieniają wartości względem części źródłowej (utrzymanie stanu), tj. pozostawienie wartości bez zmian, zarówno dla zmiennych ciągłych, jak i kategorycznych. Pełnią one rolę dodatkowych ograniczeń definiujących kontekst reguły, a nie działań modyfikujących.

Najmniejszą liczbę reguł w eksperymencie uzyskano dla zbioru danych *ZelenCap*, który charakteryzował się również najmniejszą liczbą obserwacji (64 przykładów) spośród wszystkich analizowanych zbiorów z Tabeli 6.1. Najwięcej reguł (80) wygenerowano dla zbioru danych *Aircraft*, który zawierał 3135 obserwacji. Liczba warunków dla poszczególnych zbiorów w większości przypadków istotnie różniła się od całkowitej liczby akcji. Można zatem wnioskować, że proponowany algorytm tworzy reguły, które w większości zawierają akcje, a nie elementarne warunki (akcje podtrzymujące), choć te ostatnie również występowały. Założenie o braku tzw. atrybutów stabilnych — rozumianych jako atrybuty, których wartości nie mogą być zmieniane w ramach akcji — pozwoliło na większą różnorodność typów akcji w regułach. Jeżeli w zbiorze występują atrybuty stabilne, różnica między

zbiór	$\overline{w}$	$\overline{a}$	m_w	M_w	$r_{0.01}$ (%)	$\lvert R_+\rvert$	$\lvert R_-\rvert$
AdhesiveBondB	2.0	1.0	2	2	100.0	1	1
Aircraft	2.1	1.3	2	5	100.0	43	37
KevlarVessel	2.0	1.0	2	2	100.0	1	1
LaminatePanel	2.0	1.0	2	2	100.0	1	1
LEDLife	2.0	1.2	2	2	100.0	3	1
Maintenance	2.1	1.9	2	3	100.0	4	3
NewSpring	2.0	1.0	2	2	100.0	1	1
NiCdBattery	2.0	1.0	2	2	100.0	1	1
PM	3.5	3.4	2	7	100.0	26	5
Tantalum	2.0	1.2	2	2	100.0	4	1
ZelenCap	2.0	1.0	2	2	0.0	1	0

Tabela 6.6: Szczegółowe statystyki i wyniki wygenerowanych przeżyciowych reguł akcji dla zbiorów danych z Tabeli 6.1. W kolejnych kolumnach przedstawiono: nazwę zbioru danych, średnią liczbę warunków/akcji ($\overline{w}/\overline{a}$) na regułę, minimalną i maksymalną liczbę warunków (m_w i M_w) w pojedynczej regule, procent reguł z wartością $p < 0.01$ ($r_{0.01}$) oraz liczbę reguł polepszających/pogarszających ($|R_+|/|R_-|$).

liczbą warunków a liczbą akcji wzrasta, ponieważ dla takich atrybutów można tworzyć tylko akcje podtrzymujące. Najwyższy udział dowolnych akcji w stosunku do całkowitej liczby akcji (0.95) uzyskano dla zbioru danych *PM*. Średnia liczba warunków i akcji na regułę wahała się od 1.0 do 3.5. Minimalna liczba warunków w pojedynczej regule była jednakowa dla wszystkich zbiorów danych, w Tabeli 6.6 oznaczona jako m_w. Analogicznie, maksymalną liczbę warunków oznaczono jako M_w. Maksymalna liczba akcji w pojedynczej regule wynosiła 7 i w każdym przypadku była mniejsza niż liczba atrybutów warunkowych w zbiorze danych. W konsekwencji akcje pojawiały się w przesłankach jedynie dla wybranych atrybutów, tj. tych o największym znaczeniu w momencie zdarzenia.

Wartości $r_{0.01}$ w Tabeli 6.6 wskazują, że uzyskane reguły charakteryzują się dobrymi wartościami p w teście log-rank między krzywymi KM części źródłowej i docelowej. Oznacza to, że wygenerowane reguły opisują akcje prowadzące do istotnych zmian w krzywych przeżycia. Najniższy odsetek reguł z $p < 0.01$ uzyskano dla zbioru *ZelenCap*, gdzie jedyna wygenerowana reguła nie spełniła tego kryterium.

Ostatnie dwie kolumny Tabeli 6.6 pokazują liczbę reguł polepszających, gdzie krzywa przeżycia reguły docelowej jest powyżej krzywej reguły źródłowej ($|R_+|$), oraz liczbę reguł pogarszających, gdzie krzywa przeżycia reguły docelowej jest poniżej krzywej reguły źródłowej ($|R_-|$). W większości zbiorów wygenerowano więcej reguł polepszających, natomiast dla części zbiorów liczba reguł polepszających i pogarszających była taka sama. Wskazuje to, że proponowany algorytm, w którym τ = dowolna, w przypadku niektórych zbiorów faworyzuje jeden typ reguły akcji. Ostateczne wnioski zależą jednak od specyfiki analizowanego zbioru danych.

Analogiczne eksperymenty przeprowadzono dla medycznych zbiorów danych przedstawionych w Tabeli 6.3. Wyniki zaprezentowano w Tabeli 6.7 i 6.8. W tym przypadku zaobserwowano większą różnorodność liczby generowanych reguł w porównaniu ze zbiorami przemysłowymi. Najmniejszą liczbę reguł uzyskano dla zbiorów *mgus* i *Melanoma* (po 4 reguły), natomiast największą (57) wygenerowano dla zbioru *wcgs*, który charakteryzował się największą liczbą obserwacji (3154) spośród medycznych zbiorów danych. Średnia liczba warunków na regułę wahała się od 2.1 (*hd*, *follic*) do 5.2 (*std*), co wskazuje na większą złożoność reguł niż w przypadku zbiorów przemysłowych. Podobnie, średnia liczba akcji na regułę

zbiór	$\lvert\mathcal{R}\rvert$	$\lvert\mathcal{W}\rvert$	$\lvert\mathcal{A}\rvert$	$\lvert\mathcal{A}_d\rvert$	$\lvert\mathcal{A}_d\rvert/\lvert\mathcal{A}\rvert$	$\lvert\mathcal{A}\rvert/\lvert\mathcal{W}\rvert$
actg320	20	86	81	17	0.21	0.94
BMT-Ch	5	14	14	2	0.14	1.00
cancer	6	18	16	2	0.13	0.89
follic	8	17	15	3	0.20	0.88
GBSG2	16	49	48	5	0.10	0.98
hd	12	25	20	7	0.35	0.80
lung	9	23	22	4	0.18	0.96
Melanoma	4	12	12	2	0.17	1.00
mgus	4	15	14	2	0.14	0.93
pbc	8	23	23	3	0.13	1.00
std	17	89	72	16	0.22	0.81
uis	8	30	30	7	0.23	1.00
wcgs	57	281	280	34	0.12	1.00
whas1	7	18	18	0	0.00	1.00
whas500	9	39	38	2	0.053	0.97
zinc	10	29	26	5	0.19	0.90

Tabela 6.7: Podstawowe statystyki wygenerowanych przeżyciowych reguł akcji dla zbiorów danych medycznych z Tabeli 6.3. W kolejnych kolumnach przedstawiono: nazwę zbioru danych, liczbę reguł ($|\mathcal{R}|$), całkowitą liczbę: warunków ($|\mathcal{W}|$), akcji ($|\mathcal{A}|$), akcji dowolnych ($|\mathcal{A}_d|$), udział akcji dowolnych ($|\mathcal{A}_d|/|\mathcal{A}|$), oraz stosunek liczby akcji do liczby warunków ($|\mathcal{A}|/|\mathcal{W}|$).

zbiór	$\overline{w}$	$\overline{a}$	m_w	M_w	$r_{0.01}$ (%)	$\lvert R_+\rvert$	$\lvert R_-\rvert$
actg320	4.3	4.0	2	7	100.0	18	2
BMT-Ch	2.8	2.8	2	4	100.0	3	2
cancer	3.0	2.7	2	6	83.3	3	3
follic	2.1	1.9	2	3	87.5	5	3
GBSG2	3.1	3.0	2	5	100.0	13	3
hd	2.1	1.7	2	3	100.0	9	3
lung	2.6	2.4	2	4	100.0	4	5
Melanoma	3.0	3.0	2	4	100.0	3	1
mgus	3.8	3.5	3	4	100.0	3	1
pbc	2.9	2.9	2	4	100.0	6	2
std	5.2	4.2	3	10	100.0	10	7
uis	3.8	3.8	3	4	100.0	7	1
wcgs	4.9	4.9	2	8	100.0	51	6
whas1	2.6	2.6	2	3	100.0	5	2
whas500	4.3	4.2	2	5	100.0	8	1
zinc	2.9	2.6	2	5	100.0	1	1

Tabela 6.8: Szczegółowe statystyki i wyniki wygenerowanych przeżyciowych reguł akcji dla zbiorów danych medycznych z Tabeli 6.3. W kolejnych kolumnach przedstawiono: nazwę zbioru danych, średnią liczbę warunków/akcji ($\overline{w}/\overline{a}$) na regułę, minimalną i maksymalną liczbę warunków (m_w, M_w) w pojedynczej regule, procent reguł z wartością $p < 0.01$ ($r_{0.01}$) oraz liczbę reguł polepszających/pogarszających ($|R_+|/|R_-|$).

mieściła się w przedziale od 1.7 (*hd*) do 4.9 (*wcgs*, *std*), również przekraczając zakresy obserwowane dla zbiorów przemysłowych.

Stosunek liczby akcji do liczby warunków ($|\mathcal{A}|/|\mathcal{W}|$) dla medycznych zbiorów danych charakteryzował się większą zmiennością, osiągając wartości od 0.80 (*hd*, *std*) do 1.00 (*BMT-Ch*, *Melanoma*, *pbc*, *uis*, *whas1*, *wcgs*). Wysokie wartości tego wskaźnika wskazują na dominację akcji nad warunkami elementarnymi w generowanych regułach. Udział akcji dowolnych ($|\mathcal{A}_d|/|\mathcal{A}|$) mieścił się w przedziale od 0.00 (*whas1*) do 0.35 (*hd*), przy czym średnia wartość dla zbiorów medycznych przewyższała tę uzyskaną dla zbiorów przemysłowych. Największy udział akcji dowolnych odnotowano dla zbioru *hd* (0.35), co może świadczyć o większej złożoności zależności występujących w danych medycznych.

Podobnie jak w przypadku zbiorów przemysłowych, wszystkie reguły wygenerowane dla zbiorów medycznych okazały się istotne statystycznie w teście log-rank ($\%r = 100.0\%$ dla wszystkich zbiorów), z wyjątkiem zbioru *cancer*, dla którego 83.3% reguł spełniało kryterium $p < 0.01$. Analiza zbiorów reguł polepszających (R_+) i pogarszających (R_-) wykazała przewagę reguł polepszających w większości medycznych zbiorów danych. W części zbiorów liczba reguł polepszających i pogarszających była taka sama, natomiast dla zbioru *lung* przeważały reguły pogarszające. Wyraźna przewaga reguł polepszających występowała dla zbiorów *actg320* (18 vs 2), *wcgs* (51 vs 6) oraz *uis* (7 vs 1).

Porównanie wyników uzyskanych dla zbiorów przemysłowych i medycznych ujawnia różnice charakteryzujące specyfikę obu domen. Do porównań wykorzystano medianę liczby reguł na zbiór, aby ograniczyć wpływ wartości odstających (np. *wcgs* z 57 regułami). Mediana dla zbiorów przemysłowych wynosi 2, natomiast dla zbiorów medycznych 8.5. Należy jednak zauważyć, że liczba wygenerowanych reguł zależy od wielkości zbioru danych, a zbiory medyczne były zazwyczaj większe od przemysłowych. Niezależnie od tego, reguły w zbiorach medycznych charakteryzowały się większą złożonością pod względem liczby warunków i akcji, co może odzwierciedlać bardziej skomplikowaną naturę procesów biologicznych w porównaniu z deterministycznymi mechanizmami awaryjności przemysłowej. Wyższy procent akcji dowolnych w zbiorach medycznych (średnio 16.8%) względem przemysłowych (średnio 8.1%) sugeruje większą niepewność w określaniu konkretnych kierunków zmian parametrów w kontekście medycznym, co jest zgodne z większą złożonością i nieprzewidywalnością procesów biologicznych.

Podsumowując, pokryciowy algorytm indukcji przeżyciowych reguł akcji okazał się skuteczny w odkrywaniu interpretowalnych wzorców akcji w obu analizowanych domenach. Wysoki odsetek reguł istotnych statystycznie (powyżej 95% w większości przypadków) potwierdza zdolność algorytmu do identyfikacji znaczących zależności w danych przeżyciowych. Różnice w charakterystykach reguł między domenami przemysłową a medyczną wskazują na adaptowalność metody do specyfiki różnych obszarów zastosowań, co stanowi zaletę w kontekście uniwersalności podejścia.

Studium przypadku: Interpretowalne reguły akcji w konserwacji predykcyjnej

Poniżej przedstawiono szczegółową analizę reguły akcji wygenerowanej przez algorytm pokryciowy dla zbioru danych *Maintenance*, stosowanego do ilustracji konserwacji predykcyjnej maszyn przemysłowych. Jest to syntetyczny i ogólny zbiór danych pochodzących z czujników IoT (m.in. wskaźniki ciśnienia, wilgotności i temperatury) oraz metadanych organizacyjnych (takich jak zespół czy dostawca). Nie odnosi się on do konkretnego typu urządzenia, a w dokumentacji źródłowej nie określono jednostek dla atrybutów IoT. Ze względu na największą zaobserwowaną poprawę czasu przeżycia, analiza koncentruje się na regule R4, która dobrze ilustruje potencjał algorytmu w generowaniu praktycznych zaleceń operacyjnych.

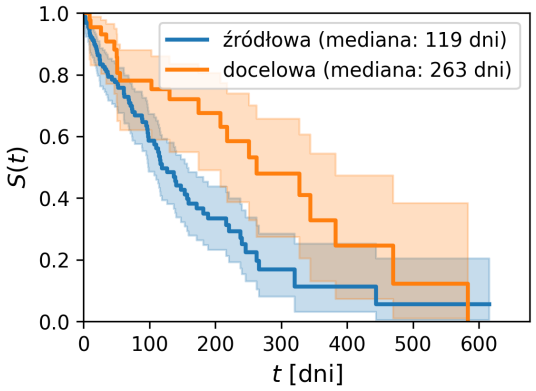
Wybrana reguła akcji ma złożoną strukturę, obejmującą trzy różne typy parametrów operacyjnych:

$$\begin{aligned} &\text{jeśli } (\mathit{moisture\_ind}, (-\infty, 120.41) \rightarrow (-\infty, 108.24)) \\ &\quad \wedge\ (\mathit{temperature\_ind}, [95.98, 157.55) \rightarrow [116.98, 138.52)) \\ &\quad \wedge\ (\mathit{team}, \{\text{TeamC}\} \rightarrow \{\text{TeamA}\}) \\ &\quad \text{to } \mathit{mediana\_przeżycia} : 74 \rightarrow 92 \text{ dni} \end{aligned}$$

Reguła pokrywa 17.2% przykładów w części źródłowej oraz 5.2% w części docelowej, wykazując wysoką istotność statystyczną ($p = 3.09 \cdot 10^{-8}$) oraz wydłużenie przewidywanego czasu bezawaryjnej pracy o 144 dni, co odpowiada 55% poprawie względem stanu wyjściowego.

Interpretacja reguły w kontekście konserwacji predykcyjnej wskazuje trzy główne obszary optymalizacji:

- Zarządzanie wilgocią — reguła sugeruje dalsze obniżenie wilgotności z poziomu już niskiego (<120.4) do jeszcze niższego (<108.2). Kierunek ten wskazuje na istnienie progu wilgotności, poniżej którego niezawodność urządzeń wyraźnie wzrasta.
- Optymalizacja temperatury operacyjnej — rekomendowane jest zawężenie przedziału z szerokiego zakresu 95.98–157.55°C do bardziej kontrolowanego 116.98–138.52°C. Wynika stąd, że utrzymanie temperatury w węższym, wyższym zakresie może sprzyjać długoterminowej niezawodności sprzętu.
- Reorganizacja zespołowa — przypisanie zadań konserwacyjnych zespołowi A zamiast zespołu C, co może odzwierciedlać różnice w kompetencjach, doświadczeniu lub stosowanych procedurach.

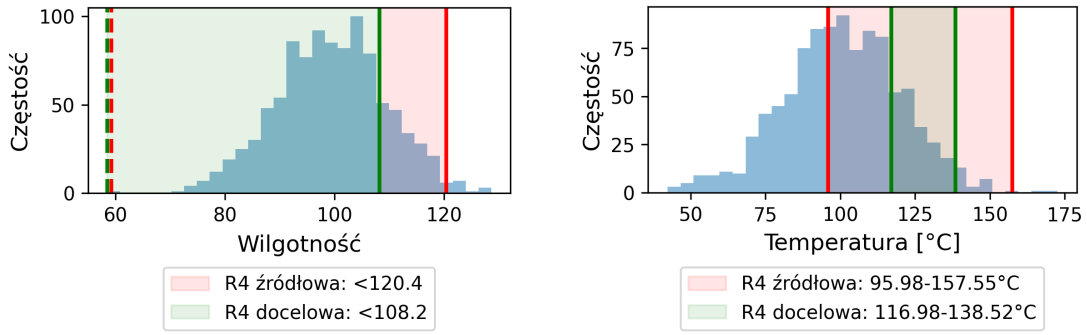


Rysunek 6.3: Krzywe przeżycia Kaplana-Meiera dla reguły R4 w zbiorze danych *Maintenance*. Niebieska krzywa reprezentuje grupę źródłową (przed zastosowaniem akcji), pomarańczowa — grupę docelową (po zastosowaniu akcji).

Analiza krzywych przeżycia przedstawiona na Rysunku 6.3 wizualnie potwierdza istotność statystyczną reguły. Krzywa reprezentująca grupę docelową (po zastosowaniu akcji) wykazuje wyższe prawdopodobieństwo przeżycia w porównaniu z grupą źródłową, przy czym wyraźna separacja krzywych widoczna jest już od początku okresu obserwacji. Mediana czasu przeżycia wzrasta ze 119 dni w grupie źródłowej do 263 dni w grupie docelowej.

Wdrożenie zaleceń wynikających z reguły R4 wymaga podjęcia działań w trzech wspomnianych obszarach. Kontrola wilgotności może wymagać inwestycji w systemy osuszania lub modyfikacji procedur operacyjnych w celu utrzymania niższych

poziomów wilgotności. Optymalizacja temperatury może obejmować dostrojenie systemów grzewczych lub chłodzących oraz modyfikację procedur operacyjnych w celu utrzymania węższego zakresu temperatur. Reorganizacja zespołowa może wymagać przeszkolenia personelu, redystrybucji zadań lub modyfikacji harmonogramów pracy.



(a) **(b)**

Rysunek 6.4: Histogramy rozkładów atrybutów wilgotności (a) i temperatury (b) w zbiorze danych *Maintenance* z zaznaczonymi zakresami źródłowymi (czerwone) i docelowymi (zielone) reguły R4. Pionowe linie wskazują precyzyjne granice zakresów, natomiast przezroczyste obszary pokazują całe zakresy reguł.

Analiza rozkładów atrybutów przedstawiona na Rysunku 6.4 obrazuje charakterystykę danych źródłowych oraz kontekst dla reguły R4. Histogram wilgotności (Rysunek 6.4a) wskazuje na rozkład zbliżony do normalnego z medianą 99.4. Zakresy reguły R4 obejmują znaczną część populacji — warunki źródłowe (<120.4) dotyczą 98.6% przykładów, natomiast bardziej restrykcyjne warunki docelowe (<108.2) obejmują 81.3% obserwacji. Histogram temperatury (Rysunek 6.4b) również wykazuje rozkład zbliżony do normalnego z medianą 100.6°C, ale zakresy reguły R4 są tu bardziej selektywne — szeroki zakres źródłowy (95.98–157.55°C) obejmuje 59.2% przykładów, podczas gdy wąski zakres docelowy (116.98–138.52°C) dotyczy jedynie 17.7% obserwacji.

Reguła R4 dobrze ilustruje zalety pokryciowego algorytmu indukcji przeżyciowych reguł akcji w zastosowaniach przemysłowych. Algorytm generuje konkretne zalecenia operacyjne z jasno określonymi zakresami parametrów, co ułatwia ich implementację w praktyce. Uwzględnia przy tym wielowymiarowe interakcje między różnymi aspektami operacji (środowisko, technologia, zasoby ludzkie) i dostarcza

ilościowej oceny spodziewanych korzyści (18 dni wydłużenia czasu przeżycia). Dzięki temu umożliwia oszacowanie opłacalności zainwestowania w proponowane zmiany. Wynik testu log-rank ($p < 10^{-7}$) dodatkowo potwierdza istotność statystyczną zaobserwowanego efektu.

6.4 Algorytm rekomendacji przeżyciowych reguł akcji

W niniejszej sekcji przedstawiono wyniki eksperymentów dotyczących algorytmu rekomendacji przeżyciowych reguł akcji opisanego w Sekcji 5.3. Eksperymenty przeprowadzono w trybie 10-krotnej stratyfikowanej walidacji krzyżowej na zbiorach danych z Tabeli 6.1 i 6.3.

Konfiguracja parametrów była zgodna z ustawieniami opisanymi w Sekcji 6.3. Ocena obejmowała zarówno jakość generowanych reguł akcji, jak i skuteczność wygenerowanych rekomendacji. Charakterystyki reguł zaprezentowane w tej sekcji odpowiadają ujęciu z Sekcji 6.3 (Tabela 6.6, 6.8). Różnica polega na tym, że tam raportowano statystyki dla pełnych zbiorów (bez walidacji krzyżowej), a tutaj przedstawiono wyniki uzyskane w konfiguracji 10-krotnej walidacji krzyżowej.

Z zestawień usunięto zbiory, dla których nie uzyskano stabilnych wyników w walidacji krzyżowej. Spośród zbiorów przemysłowych wykluczono *Aircraft*, *PM* i *ZelenCap*, ze względu na trudności z modelem arbitra XGBoost — prawdopodobnie spowodowane wysokim poziomem cenzurowania (odpowiednio 99.3%, 91.7% i 50.0%) oraz niewystarczającą liczbą zdarzeń. Wśród zbiorów medycznych pominięto *cancer*, *wcgs* i *zinc*, z analogicznych powodów (m.in. wysokie cenzurowanie: 91.9% i 81.2%), które utrudniały stabilną estymację hazardu bazowego i trenowanie arbitra w trybie 10-krotnej walidacji krzyżowej.

Tabele 6.9 i 6.10 zestawiają wyniki algorytmu rekomendacji. Kolumny $S_{0.05}$ i $S_{0.01}$ przedstawiają wskaźnik spójności (por. uzasadnienie testu KS w Sekcji 5.3). MAE oznacza średni błąd bezwzględny między krzywą uzyskaną na podstawie rekomendacji a krzywą dla przykładu zmutowanego przez model arbitra. Pokrycie to odsetek przykładów testowych spełniających kryterium minimalnego pokrycia, umożliwiającego estymację krzywej docelowej (KM). $\overline{a}$ to średnia liczba akcji na rekomendację, a $\overline{q}$ — średni odsetek atrybutów poddanych modyfikacji. Krzywa docelowa KM_D jest estymowana metodą Kaplana-Meiera na zbiorze uczącym

zbiór	$S_{0.05}$	$S_{0.01}$	MAE	Pokrycie (%)	$\overline{a}$	$\overline{q}$ (%)
AdhesiveBondB	1.00	1.00	0.24	1.00	1.0	49.0
KevlarVessel	1.00	1.00	0.28	1.00	1.0	50.0
LaminatePanel	0.76	0.40	0.14	1.00	1.0	100.0
LEDLife	0.88	0.88	0.25	1.00	1.3	65.0
Maintenance	0.82	0.12	0.07	1.00	2.0	40.4
NewSpring	0.90	0.78	0.25	1.00	0.8	27.3
NiCdBattery	0.86	0.82	0.21	1.00	0.9	11.2
Tantalum	1.00	1.00	0.04	1.00	1.4	70.0

Tabela 6.9: Wyniki eksperymentów dla zbiorów przemysłowych. Kolumny przedstawiają: nazwę zbioru danych, metryki spójności ($S_{0.05}$ i $S_{0.01}$), MAE, pokrycie (w %), średnią liczbę akcji na rekomendację ($\overline{a}$) oraz średni procent atrybutów poddanych mutacji ($\overline{q}$, w %).

zbiór	$S_{0.05}$	$S_{0.01}$	MAE	Pokrycie (%)	$\overline{a}$	$\overline{q}$ (%)
actg320	1.00	0.98	0.09	1.00	1.6	14.9
BMT-Ch	1.00	1.00	0.32	1.00	1.9	5.4
follic	0.84	0.82	0.16	1.00	1.6	40.0
GBSG2	0.96	0.94	0.22	1.00	2.5	31.5
hd	0.70	0.58	0.13	1.00	1.0	16.7
lung	0.88	0.80	0.16	1.00	3.0	43.4
Melanoma	0.98	0.96	0.18	1.00	1.8	26.3
mgus	0.97	0.95	0.35	1.00	2.2	24.2
pbc	0.94	0.86	0.32	0.98	2.0	11.5
std	0.90	0.80	0.20	0.90	1.9	9.0
uis	0.98	0.96	0.19	1.00	2.7	20.8
whas1	1.00	0.98	0.12	1.00	1.3	18.9
whas500	0.98	0.96	0.26	0.98	2.0	15.5

Tabela 6.10: Wyniki eksperymentów dla zbiorów medycznych. Kolumny przedstawiają: nazwę zbioru danych, metryki spójności ($S_{0.05}$ i $S_{0.01}$), MAE, pokrycie (w %), średnią liczbę akcji na rekomendację ($\overline{a}$) oraz średni procent atrybutów poddanych mutacji ($\overline{q}$, w %).

ograniczonym do obserwacji pokrywanych przez prawą stronę reguły. Stanowi ona referencję dla rekomendacji (por. Sekcję 5.3).

Dodatkowo, w trakcie eksperymentów zastosowano zintegrowany wskaźnik Briera (IBS) jako pomocniczą miarę błędu prognozy do wewnętrznej kontroli jakości modeli. Wartości IBS nie są w tym przypadku raportowane w wynikach. Definicję, ograniczenia i kontekst użycia omówiono w Sekcji 5.3.

Walidację przeprowadzono z wykorzystaniem modelu arbitra XGBoost z funkcją straty opartą na proporcjonalnych hazardach Coxa. Arbiter był trenowany wyłącznie na zbiorze uczącym w każdej części walidacji. Krzywą przeżycia dla zmutowanego przykładu x' wyznaczano zgodnie z modelem Coxa jako

$$S_{x'}(t) = S_0(t)^{\exp(\eta(x'))},$$

gdzie $S_0(t)$ oznacza krzywą przeżycia bazowego estymowaną na czasach zdarzeń w zbiorze uczącym, a $\eta(x')$ to liniowa kombinacja atrybutów x' wyznaczona przez model. Zgodność krzywej arbitra $KM_{x'}$ z krzywą docelową KM_D (KM na danych uczących pokrywanych przez prawą stronę reguły) oceniano testem Kołmogorowa-Smirnowa dla dwóch prób. Wskaźnik spójności S_α definiowano jako odsetek przypadków spełniających warunek $p \geqslant \alpha$.

Ilościowa analiza indukowanych reguł (por. Tabela 6.6, 6.8) potwierdziła ich statystyczną istotność w większości przypadków. Ze względu na brak powtórnych pomiarów dla tych samych przypadków nie było możliwe bezpośrednie sprawdzenie, czy sugerowane modyfikacje faktycznie wpływają na czasu przeżycia w danych testowych. Z tego powodu zastosowano model arbitra, aby zweryfikować zgodności alternatywnego modelu przeżycia z estymacjami algorytmu rekomendacji.

Wyniki walidacji wskazują na wysoką precyzję algorytmu. Dla zbiorów przemysłowych $S_{0.05}$ mieści się w zakresie od 76% (*LaminatePanel*) do 100% (*AdhesiveBondB*, *KevlarVessel*, *Tantalum*). Dla $S_{0.01}$ skuteczność również pozostaje wysoka, z wyjątkiem *Maintenance* (12%). MAE wynosi 0.04–0.28 przy pełnym pokryciu. W przypadku zbiorów medycznych większość z nich osiąga spójność powyżej 90% dla obu poziomów istotności. Wyjątek stanowi *hd* (70% i 58%). MAE mieści się w zakresie 0.09–0.35, a pokrycie wynosi 90–100%.

Algorytm generuje średnio od 4 do 22 reguły na zbiór danych, z przewagą reguł polepszających. Ponad 90% reguł jest istotnych statystycznie na obu poziomach istotności, z wyjątkiem *zinc* (67% dla $p < 0.05$).

Wskaźniki z Tabeli 6.9 i 6.10 wskazują na niewielką liczbę akcji na rekomendację oraz ograniczony odsetek modyfikowanych atrybutów, co sprzyja zastosowaniu rekomendacji w praktyce. W zbiorach przemysłowych średnia liczba akcji na rekomendację wynosi 0.8–2.0, przy modyfikacji 11.2–100% atrybutów. W zbiorach medycznych to odpowiednio 1.0–3.0 akcji i 5.4–43.4% cech. Taka selektywność jest szczególnie istotna w zastosowaniach, gdzie jednoczesna zmiana wielu parametrów jest kosztowna.

Ponadto wysokie wartości spójności (średnio $> 80\%$ dla 14 z 17 zbiorów), niskie MAE (< 0.3 dla 15 z 17) oraz pełne pokrycie (100% dla 15 z 17) potwierdzają, że rekomendacje prowadzą do statystycznie istotnych różnic w krzywych przeżycia. Połączenie wysokiej zgodności, niewielkiej liczby modyfikowanych atrybutów i wysokiego pokrycia dowodzi potencjalnej użyteczność metody.

Studium przypadku: Rekomendacje dla pacjenta w terapii antyretrowirusowej

Analiza działania algorytmu rekomendacji przeżyciowych reguł akcji została przeprowadzona na zbiorze danych *actg320* z badania *AIDS Clinical Trials Group Study 320*. Wybór tego zbioru wynika z najwyższego wskaźnika spójności ($S_{0.05} = 100\%$) oraz najniższej wartości MAE (0.09) spośród wszystkich zbiorów medycznych. Zbiór charakteryzuje się pełnym pokryciem (100%) i zawiera 19 reguł akcji.

Studium przypadku dotyczy pacjenta o indeksie 38 z pierwszego podzbioru walidacji krzyżowej. Profil kliniczny pacjenta obejmuje następujące parametry: wiek 36 lat ($age = 36$), płeć męska ($sex = 1$), rasa biała ($raceth = 1$), brak hemofilii ($hemophil = 0$), brak historii używania narkotyków dożylnie ($ivdrug = 1$), wynik 80 punktów w skali Karnofsky'ego ($karnof = 80$), liczba komórek CD4 wynosząca 3.0 komórek/µL ($cd4 = 3.0$) oraz 6-dniowa wcześniejsza ekspozycja na azydotymidynę ($priorzdv = 6$). Początkowa predykcja wskazuje, że mediana czasu przeżycia przekracza 364 dni (maksymalny czas obserwacji w badaniu), przy prawdopodobieństwie przeżycia 364 dni wynoszącym 0.82.

Wskazane parametry odzwierciedlają zaawansowany stan kliniczny pacjenta z HIV. Liczba komórek CD4 wynosząca 3 komórek/µL świadczy o ciężkiej immunosupresji (norma: 500–1200 komórek/µL), charakterystyczną dla AIDS [142]. Wynik 80 punktów w skali Karnofsky'ego oznacza normalną aktywność z wysił-

kiem przy obecności objawów choroby, co odpowiada umiarkowanemu ograniczeniu sprawności funkcjonalnej [143]. Krótka, 6-dniowa ekspozycja na azydotymidynę może wskazywać na wczesny etap terapii antyretrowirusowej lub profilaktykę poekspozycyjną [144, 145].

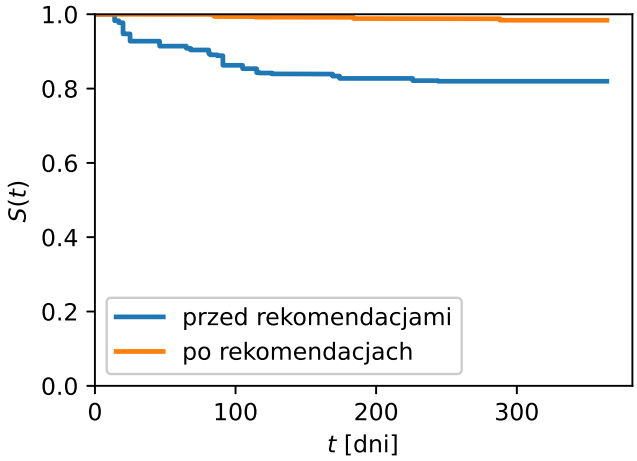
Proces generowania rekomendacji rozpoczęto od mapowania atrybutów pacjenta na strukturę meta-danych zawierającą dyskretyzowane przedziały wartości dla 11 atrybutów klinicznych. Na tej podstawie utworzono reprezentację meta-obiektu, co umożliwiło indukcję meta-reguły obejmującej trzy atrybuty. Otrzymana reguła akcji wskazuje na modyfikację trzech parametrów klinicznych:

$$\begin{aligned} \text{jeśli } & (\textit{cd4}, (2.5, 5.75] \to (64.25, 65.0]) \wedge \\ & (\textit{priorzdv}, (5.5, 9.5] \to (9.5, 19.0]) \wedge \\ & (\textit{karnof}, (-\infty, 85.0] \to (95.0, +\infty)) \\ \text{to } & S_Z(t) \to S_D(t) \end{aligned}$$

Reguła sugeruje zwiększenie liczby komórek CD4 z 3.0 do 64.5 (+61.5), wydłużenie wcześniejszej ekspozycji na ZDV z 6 do 14 dni (+8 dni) oraz poprawę wyniku w skali Karnofsky'ego z 80 do 100 punktów (+20 punktów). Modyfikacje te dotyczą 27.3% dostępnych atrybutów (3 z 11 cech) i obejmują trzy aspekty terapii HIV: poprawę stanu immunologicznego (*cd4*), optymalizację profilaktyki antyretrowirusowej (*priorzdv*) oraz zwiększenie ogólnej wydolności funkcjonalnej pacjenta (*karnof*).

Rysunek 6.5 przedstawia krzywe przeżycia przed i po zastosowaniu rekomendacji. Krzywa po wprowadzeniu zmian przebiega wyżej niż krzywa dla stanu wyjściowego, co wskazuje na wzrost estymowanego prawdopodobieństwa przeżycia.

Analiza kliniczna wygenerowanych rekomendacji ujawnia trzy kierunki interwencji terapeutycznej. Sugerowany wzrost liczby komórek CD4 z 3.0 do 64.5 odpowiada przejściu z zaawansowanego niedoboru immunologicznego do poziomu zbliżonego do dolnej granicy normy, zgodnie z celem terapii antyretrowirusowej, jakim jest odtworzenie funkcji immunologicznej i trwała supresji wiremii [142, 146, 147]. Wskazanie na wydłużenie wcześniejszej ekspozycji na ZDV z 6 do 14 dni odnosi się do zmiennej opisowej *priorzdv* (historia stosowania zydowudyny) i ma charakter nieakcyjny na etapie podejmowania decyzji terapeutycznych — nie należy interpretować tego jako zalecenia wydłużania ani rozpoczynania PEP w ramach bieżącego leczenia. Model w tym przypadku wykorzystuje asocjacje



Rysunek 6.5: Krzywe przeżycia Kaplana-Meiera dla pacjenta przed i po zastosowaniu rekomendacji w zbiorze danych *actg320*. Wykres przedstawia porównanie funkcji przeżycia dla oryginalnego profilu pacjenta oraz po modyfikacji zgodnej z wygenerowaną regułą akcji.

zaobserwowane w danych (np. różnice wynikające z wcześniejszego stosowania ZDV, w tym monoterapii), a nie sugeruje modyfikowalnej interwencji. W cytowanych pracach [144, 145] zależność dotyczyła stosowania zydowudyny (ZDV) w ramach profilaktyki poekspozycyjnej. Wzrost wyniku w skali Karnofsky'ego z 80 do 100 punktów odpowiada przejściu od ograniczonej aktywności życiowej do pełnej sprawności funkcjonalnej [143].

Studium przypadku ilustruje zdolność algorytmu do identyfikacji spójnych modyfikacji parametrów terapeutycznych w leczeniu HIV. Wygenerowana reguła akcji integruje trzy niezależne domeny kliniczne: status immunologiczny, historię farmakoterapii oraz sprawność funkcjonalną pacjenta, tworząc wielowymiarowy profil interwencji. Zbieżność z modelem arbitra (MAE = 0) potwierdza wewnętrzną spójność obliczeń, a jawność składników reguły (konkretne przedziały dla *cd4*, *priorzdv* i *karnof*) ułatwia weryfikację kliniczną oraz ocenę wykonalności proponowanych modyfikacji.

6.5 Algorytm indukcji przeżyciowych reguł wyjątków

W niniejszej sekcji przedstawiono wyniki eksperymentów dla algorytmu indukcji przeżyciowych reguł wyjątków opisanego w Sekcji 5.4. Ze względu na eksploracyjny charakter metody, eksperymenty przeprowadzono na pełnych zbiorach danych, bez stosowania walidacji krzyżowej. Wykorzystanie całego zbioru danych zwiększa liczbę przykładów spełniających warunki reguł, co ułatwia wykrywanie wyjątków i zmniejsza ryzyko, że rzadkie wzorce nie spełnią wymagań pokrycia po podziale na zbiory treningowe i testowe.

Konfiguracja eksperymentów była następująca: $mincov = 1$ (minimalna liczba niepokrytych przykładów), $max\_growing = 5$ (maksymalna liczba warunków w regule), $\alpha = 0.05$ (próg istotności), a miarą oceny jakości był test log-rank. Celem eksperymentów była identyfikacja trójek (CR, RR, ER), w których ER definiowana jest względem CR i RR jako $CR \wedge RR$, reprezentujących nietypowe wzorce przeżycia.

Reguły wyjątków wykryto w 16 z 27 analizowanych zbiorów danych (59%), zgodnie z ich definicją — małe pokrycie i wysoka specyficzność. W pozostałych zbiorach ich brak zwykle wynika z niewystarczającej liczby obserwacji spełniających jednocześnie warunki reguł CR i RR oraz z przyjętych wartości progowych (*mincov*, próg istotności).

Tabele 6.11 i 6.12 przedstawiają liczbę reguł CR — interpretowanych jako próby identyfikacji wyjątków — oraz liczbę wyjątków ER wykrytych w ramach tych prób, dla każdego analizowanego zbioru danych. Łącznie w zbiorach z dziedziny diagnostyki predykcyjnej maszyn przeprowadzono 40 prób, z czego 16 zakończyło się identyfikacją wyjątków, natomiast w zbiorach medycznych przeprowadzono 99 prób, z czego 37 zakończyło się wykryciem wyjątków. Najwyższy udział wyjątków odnotowano w zbiorach *Maintenance* i *cancer*: odpowiednio 10/20 (50.0%) oraz 12/23 (52.2%).

Tabele 6.13 i 6.14 prezentują wybrane reguły wyjątków dla obu grup zbiorów danych. W zbiorach przemysłowych przeważają reguły opisujące skrajne wartości parametrów operacyjnych (np. wysoka temperatura w połączeniu z wysokim prądem w zbiorze *LEDLife*), natomiast w zbiorach medycznych częste są złożone interakcje czynników demograficznych i klinicznych (np. wiek wraz z parametrami morfologii krwi w zbiorze *mgus*). Zidentyfikowano również wyjątki o niewielkim

zbiór	CR	ER
AdhesiveBondB	2	0
Aircraft	1	0
KevlarVessel	2	0
LaminatePanel	2	0
LEDLife	2	1
Maintenance	20	10
NewSpring	2	1
NiCdBattery	2	1
PM	4	1
Tantalum	2	1
ZelenCap	2	1
Razem	41	16

Tabela 6.11: Liczba CR i ER dla zbiorów danych z Tabeli 6.1

zbiór	CR	ER
actg320	7	1
BMT-Ch	2	1
cancer	23	12
follic	2	2
GBSG2	11	2
hd	2	0
lung	5	2
Melanoma	8	1
mgus	9	5
pbc	2	1
std	11	2
uis	2	1
wcgs	8	5
whas1	2	0
whas500	2	1
zinc	3	1
Razem	99	37

Tabela 6.12: Liczba CR i ER dla zbiorów danych z Tabeli 6.3

zbiór	Reguła [p, P]
LEDLife	$\underline{DegreesC \geqslant 102.00} \wedge Current \geqslant 35.00$ [50, 133]
Maintenance	$\underline{team = \text{TeamB}} \wedge pressureInd \leqslant 87.09 \wedge moistureInd \leqslant 123.46$ [93, 800]
NewSpring	$\underline{stroke \geqslant 65.00} \wedge method = \text{Old}$ [16, 86]
NiCdBattery	$\underline{discharge\_depth \geqslant 50.00} \wedge degrees\_c \geqslant 45.00$ [16, 69]
Tantalum	$\underline{volts \leqslant 49.00} \wedge degrees\_c \geqslant 25.00$ [1403, 1763]
ZelenCap	$\underline{volts \geqslant 225.00} \wedge degrees\_c \geqslant 175.00$ [19, 51]

Tabela 6.13: Przykłady zidentyfikowanych wyjątków dla zbiorów danych z Tabeli 6.1. Dla każdego zbioru wybrano jedną regułę o największym pokryciu k. W części ER podkreślono fragment odpowiadający CR, a fragment niepodkreślony odpowiada RR. Każda reguła jest przedstawiona wraz z jej pokryciem w formacie $[k, n]$, gdzie k oznacza liczbę przykładów pokrytych przez ER, a n — całkowitą liczbę przykładów w zbiorze danych.

pokryciu. Przykładowo, w zbiorze *Melanoma* jedna z reguł obejmuje 3 ze 164 obserwacji, co potwierdza identyfikację wzorców o wysokiej specyficzności prognostycznej.

Algorytm indukcji przeżyciowych reguł wyjątków wykazał zdolność do identyfikacji znaczących wzorców odchyleń w 16 z 27 analizowanych zbiorów danych. Na szczególną uwagę zasługują wyniki dla zbiorów *Maintenance* i *cancer*, które charakteryzują się najwyższą koncentracją wyjątków (odpowiednio 50.0% i 52.2%). Brak publicznie dostępnych implementacji alternatywnych metod w dziedzinie indukcji reguł wyjątków w analizie przeżycia uniemożliwia przeprowadzenie porównawczej walidacji empirycznej.

Studium przypadku: Analiza reguł wyjątków w zbiorze danych onkologicznych

Poniżej przedstawiono szczegółową analizę reprezentatywnej trójki (CR, RR, ER) zidentyfikowanej przez algorytm w zbiorze danych *cancer*, dotyczącym przeżycia pacjentów z zaawansowanym rakiem płuc. Zbiór ten charakteryzuje się najwyższą liczbą wykrytych wyjątków — 12 z 23 reguł (52% wszystkich wygenerowanych reguł).

zbiór	Reguła [p, P]
Melanoma	$\underline{thick \geqslant 2.17 \wedge age \geqslant 13.00} \wedge age \leqslant 21.50$ [3, 164]
GBSG2	$\underline{pnodes \leqslant 7.50 \wedge estrec \geqslant 5.50 \wedge progrec \geqslant 1.50 \wedge estrec \leqslant 825.50} \wedge horTh = \text{yes}$ [115, 548]
cancer	$\underline{sex \leqslant 1.50 \wedge ph.karno \geqslant 55.00 \wedge age \geqslant 60.50 \wedge wt.loss \leqslant 27.50} \wedge ph.ecog \geqslant 1.50$ [14, 182]
follic	$\underline{age \leqslant 56.75} \wedge hgb \leqslant 121.50 \wedge hgb \geqslant 102.50$ [24, 432]
lung	$\underline{Operated = 1} \wedge Stage3 \leqslant 0.50$ [211, 825]
mgus	$\underline{age \geqslant 62.50} \wedge hgb \leqslant 12.85$ [48, 192]
pbc	$\underline{bili \geqslant 1.35} \wedge chol \leqslant 198.50$ [8, 334]
std	$\underline{age \leqslant 17.50 \wedge os30d \leqslant 0.50} \wedge race = \text{W}$ [43, 701]
uis	$\underline{FRAC \leqslant 0.88} \wedge NDT \geqslant 5.50 \wedge FRAC \leqslant 1.51$ [68, 460]
whas500	$\underline{age \geqslant 67.50} \wedge diasbp \leqslant 65.50$ [74, 400]
zinc	$\underline{sevdysp = \text{Severe Dysplasia}} \wedge cacent \geqslant 2.39$ [1, 344]

Tabela 6.14: Przykłady zidentyfikowanych wyjątków dla zbiorów danych z Tabeli 6.3. Dla każdego zbioru wybrano jedną regułę o największym pokryciu k. W części ER podkreślono fragment odpowiadający CR, a fragment niepodkreślony odpowiada RR. Każda reguła przedstawiona jest wraz z jej pokryciem w formacie $[k, n]$, gdzie k oznacza liczbę przykładów pokrytych przez ER, a n — całkowitą liczbę przykładów w zbiorze danych.

Algorytm zidentyfikował trójkę reguł ujawniającą nieoczekiwany wzorzec przeżycia u starszych pacjentów płci męskiej z pozornie korzystnymi czynnikami prognostycznymi. Pierwsza z reguł opisuje wzorzec bazowy (w nawiasach podano liczbę obserwacji spełniających przesłankę reguły k oraz liczebność zbioru obserwacji n).:

$$\begin{aligned}\text{CR: jeśli } & (\textit{płeć} \leqslant 1.5 \wedge \textit{ECOG} \geqslant 0.5 \wedge \\ & \quad \textit{Karnofsky} \geqslant 55 \wedge \textit{utrata\_wagi} \leqslant 27.5) \\ & \text{to } \textit{mediana\_przeżycia} = 196.5 \text{ dni} \quad (k = 78, n = 228)\end{aligned}$$

Druga reguła pełni rolę punktu odniesienia dla populacji starszych pacjentów:

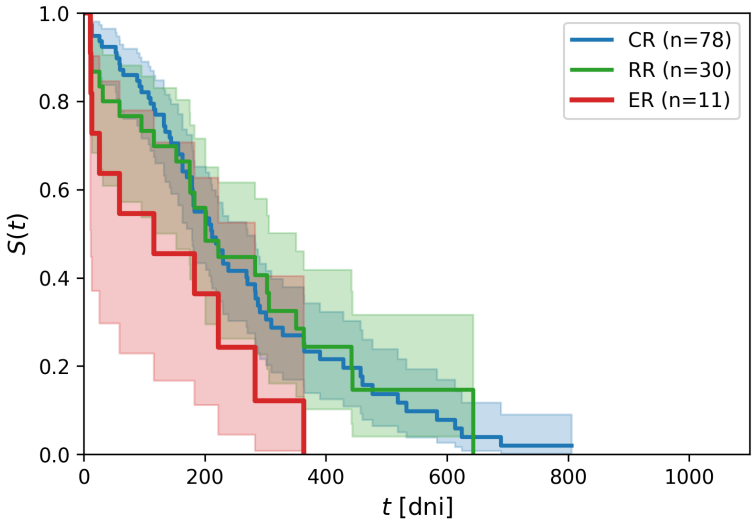
$$\begin{aligned}\text{RR: jeśli } & (\textit{wiek} \geqslant 72.5 \wedge \textit{Karnofsky} \geqslant 55) \\ & \text{to } \textit{mediana\_przeżycia} = 194.5 \text{ dni} \quad (k = 30, n = 228)\end{aligned}$$

Trzecia reguła definiuje wyjątek jako kombinację warunków z dwóch poprzednich reguł:

$$\begin{aligned}\text{ER: jeśli } & (\textit{płeć} \leqslant 1.5 \wedge \textit{ECOG} \geqslant 0.5 \wedge \textit{Karnofsky} \geqslant 55 \wedge \\ & \quad \textit{utrata\_wagi} \leqslant 27.5 \wedge \textit{wiek} \geqslant 72.5) \\ & \text{to } \textit{mediana\_przeżycia} = 116.0 \text{ dni} \quad (k = 11, n = 228)\end{aligned}$$

Reguła bazowa (CR) obejmuje 78 pacjentów (84.6% współczynnik zdarzeń) i dotyczy mężczyzn z umiarkowanym pogorszeniem wydolności według skali ECOG, dobrą oceną w skali Karnofsky'ego oraz umiarkowaną utratą masy ciała. Reguła referencyjna (RR) pokrywa 30 pacjentów (80.0% współczynnik zdarzeń) i dotyczy starszych pacjentów z zachowaną dobrą wydolnością według skali Karnofsky'ego. Wyjątek (ER) obejmuje 11 pacjentów należących zarówno do populacji CR, jak i RR. Charakteryzuje się on istotnie wyższym współczynnikiem śmiertelności (90.9%) oraz znacząco krótszym czasem przeżycia.

Walidacja statystyczna potwierdza poprawność identyfikacji wyjątku zgodnie z kryteriami algorytmu. Test log-rank między regułami CR i RR wykazał brak istotnej różnicy ($p = 0.91$), co wskazuje, że obie reguły opisują podobne wzorce przeżycia. Natomiast porównanie CR z ER ujawniło istotną różnicę statystyczną ($p = 0.03$), podczas gdy porównanie RR z ER dało wartość $p = 0.07$, bliską progu istotności statystycznej.



Rysunek 6.6: Empiryczne krzywe przeżycia dla trójki reguł (CR, RR, ER) zidentyfikowanej w zbiorze danych *cancer*.

Obserwowana jest istotna redukcja mediany przeżycia — z 196.5 dni (CR) i 194.5 dni (RR) do 116.0 dni w grupie wyjątku, co stanowi spadek o około 40–41%. Obserwowany wzorzec wskazuje, że starsi pacjenci płci męskiej łączący tradycyjnie „dobre" czynniki prognostyczne — zachowaną wydolność według skali Karnofsky'ego i umiarkowaną utratę masy ciała — wykazują znacząco gorszą prognozę niż każda z grup referencyjnych analizowana niezależnie. Różnice te są wyraźnie widoczne na krzywych Kaplana-Meiera przedstawionych na Rysunku 6.6.

Szczegółowa analiza indywidualnych przypadków ujawnia wewnętrzne zróżnicowanie pacjentów spełniających kryteria reguły wyjątku, z wyodrębnieniem dwóch wyraźnych wzorców. Pierwsza grupa (5 pacjentów) charakteryzuje się bardzo krótkim przeżyciem (11–26 dni), znacznym ograniczeniem wydolności (ECOG = 2.0) oraz często znaczną utratą masy ciała. Druga grupa (6 pacjentów) wykazuje dłuższe przeżycie (116–363 dni) przy mieszanych parametrach wydolności.

Identyfikacja tego wyjątku ma istotne implikacje kliniczne. Po pierwsze, wskazuje na ograniczenia standardowych skal prognostycznych (Karnofsky'ego, ECOG) w populacji starszych pacjentów płci męskiej z nowotworem, które mogą maskować subtelne, lecz krytyczne różnice w stanie ogólnym[148]. Po drugie, sugeruje potrzebę opracowania dedykowanych modeli prognostycznych uwzględniających in-

terakcje między wiekiem, płcią a innymi czynnikami prognostycznymi. W praktyce klinicznej może to oznaczać konieczność rozszerzenia diagnostyki o kompleksową ocenę geriatryczną (*comprehensive geriatric assessment*, CGA) oraz ostrożniejszego podejścia do kwalifikacji tej grupy pacjentów do agresywnych terapii[149, 150, 151, 152].

Odkrycie to ilustruje potencjalną użyteczność algorytmu indukcji reguł wyjątków w identyfikacji nietypowych wzorców, które mogą pozostać niezauważone przy stosowaniu tradycyjnych metod analizy. Wyjątek ten może prowadzić do sformułowania nowych hipotez badawczych dotyczących mechanizmów biologicznych odpowiedzialnych za obserwowany wzorzec oraz wskazywać potencjalne kierunki badań w geriatroonkologii. Z punktu widzenia praktyki klinicznej obserwacja ta uzasadnia włączenie oceny geriatrycznej do postępowania diagnostyczno-terapeutycznego oraz ostrożniejszą kwalifikację do terapii o wysokiej intensywności w tej podgrupie (por. Rysunek 6.6).

6.6 Interpretowalny zespół reguł przeżyciowych

W niniejszej sekcji przedstawiono wyniki eksperymentów dla interpretowalnego zespołu reguł przeżyciowych (ang. *interpretable rule-based survival ensemble*, RSE) opisanego w Sekcji 5.5. Eksperymenty przeprowadzono z użyciem 10-krotnej walidacji krzyżowej, stratyfikowanej względem statusu zdarzenia, na zbiorach z Tabeli 6.1 i 6.3. Zastosowano następujące parametry: $n = 100$ (liczba estymatorów), $\theta = \lfloor\sqrt{|A|}\rfloor$ (maksymalna liczba cech), $\alpha = 1.0$ (próbkowanie bootstrap) oraz $\sigma = 5.0$ (minimalny próg wsparcia). Główną metryką oceny był całkowity wskaźnik Briera (IBS) mierzący dokładność estymacji funkcji przeżycia. Jako estymator bazowy w architekturze zespołowej wykorzystano metodę Survival Rules (SR) [12, 126]. Ze względu na brak możliwości generowania skalarnych wskaźników ryzyka dla metod RSE i SR, ograniczono ocenę zdolności dyskryminacyjnej do metryki IBS.

Wyniki dla metody RSE porównano z metodami referencyjnymi: SR (ang. *survival rules* — indukcja reguł przeżyciowych z estymacją funkcji przeżycia [126]), ST (ang. *survival tree* — pojedyncze drzewo przeżycia z podziałami opartymi na statystyce log-rank i estymacją Kaplana-Meiera w liściach), RSF (ang. *random survival forest* — losowy las drzew przeżycia z bootstrapowaniem próbek i losowym

wyborem cech), Cox — regularyzowana regresja Coxa z karą *elastic net*) oraz KM (estymator Kaplana-Meiera — nieparametryczny estymator funkcji przeżycia).

zbiór	RSE	Cox	KM	RSF	SR	ST
AdhesiveBondB	.16±.04	.19±.02	.21±.03	.08±.03	.16±.08	.11±.05
KevlarVessel	.10±.04	.06±.02	.13±.02	.04±.02	.09±.06	.05±.03
LaminatePanel	—	.07±.02	.17±.03	.07±.02	.04±.03	.07±.02
LEDLife	.11±.02	.07±.01	.20±.02	.06±.01	.11±.02	.06±.01
Maintenance	.03±.00	.05±.01	.05±.01	.00±.00	.01±.02	.00±.00
NewSpring	.16±.03	.12±.04	.22±.01	.14±.05	.17±.07	.14±.05
NiCdBattery	.11±.02	.06±.02	.18±.03	.10±.02	.13±.05	.08±.03
PM	.08±.01	.08±.01	.08±.01	.07±.01	.08±.01	.11±.02
Tantalum	.01±.00	.01±.00	.02±.00	.02±.00	.01±.00	.02±.00
ZelenCap	.19±.03	.16±.04	.20±.04	.19±.06	.20±.06	.18±.05

Tabela 6.15: Wartości metryki IBS (średnie wartości obliczone na podzbiorach w ramach walidacji krzyżowej) dla zbiorów z Tabeli 6.1. Objaśnienia skrótów modeli: RSE = interpretowalny zespół reguł przeżyciowych, SR = *survival rules*, ST = *survival tree*, RSF = *random survival forest*, Cox = regularyzowana regresja Coxa z karą *elastic net*, KM = estymator Kaplana-Meiera.

Empiryczne wyniki dla metryki IBS przedstawiono w Tabeli 6.15 i 6.16. Analiza predykcyjna wskazuje zróżnicowaną skuteczność metody RSE, zależną od charakterystyk statystycznych badanych zbiorów danych. Najniższą wartość IBS uzyskano dla zbioru *Tantalum* (0.01 ± 0.00), co daje wynik porównywalny z metodami Cox i RandomSurvivalForest. Zbliżony poziom błędu odnotowano dla zbioru *Maintenance* (0.03 ± 0.00). Dla pozostałych 19 zbiorów metoda RSE osiąga umiarkowaną jakość predykcyjną, z wartościami IBS w zakresie 0.08–0.23.

Zbiór *Aircraft* został wykluczony z tabeli wartości IBS. Bardzo wysoki stopień cenzurowania (99.3%) oraz złożona struktura uniemożliwiły stabilne wytrenowanie wymaganej liczby estymatorów bazowych w ramach zespołu, co skutkowało brakiem wyników dla tego zbioru danych.

Analiza porównawcza z metodami referencyjnymi wskazuje na systematyczną przewagę RSE nad estymatorem Kaplana-Meiera (KM) w 96% przypadków (24 z 25). W porównaniu z metodami takimi jak *random survival forest* (RSF) i regre-

zbiór	RSE	Cox	KM	RSF	SR	ST
GBSG2	0.19±0.03	0.18±0.03	0.20±0.02	0.18±0.04	0.19±0.07	0.31±0.07
Melanoma	0.18±0.03	0.17±0.04	0.21±0.04	0.18±0.04	0.17±0.04	0.25±0.08
actg320	0.06±0.00	0.06±0.00	0.06±0.00	0.06±0.00	0.06±0.01	0.09±0.01
BMT-Ch	0.23±0.03	0.22±0.03	0.26±0.04	0.25±0.07	0.20±0.02	0.33±0.10
cancer	0.19±0.02	0.19±0.03	0.18±0.02	0.20±0.02	0.17±0.03	0.29±0.06
follic	0.21±0.04	0.19±0.04	0.21±0.04	0.22±0.04	0.19±0.04	0.29±0.05
hd	0.22±0.02	0.21±0.02	0.23±0.02	0.22±0.02	0.21±0.02	0.28±0.02
lung	0.16±0.02	0.15±0.02	0.18±0.02	0.15±0.02	0.15±0.02	0.17±0.02
mgus	0.18±0.01	0.18±0.02	0.20±0.01	0.16±0.02	0.16±0.03	0.26±0.06
pbc	0.16±0.01	0.15±0.03	0.20±0.02	0.14±0.02	0.15±0.02	0.24±0.06
std	0.22±0.03	0.22±0.03	0.23±0.03	0.23±0.03	0.23±0.04	0.34±0.02
uis	0.19±0.03	0.18±0.04	0.20±0.03	0.19±0.03	0.15±0.04	0.24±0.05
wcgs	0.04±0.01	0.04±0.01	0.05±0.00	0.04±0.00	0.04±0.01	0.07±0.01
whas1	0.21±0.05	0.24±0.04	0.24±0.04	0.20±0.08	0.21±0.05	0.32±0.12
whas500	0.19±0.02	0.18±0.07	0.23±0.02	0.18±0.04	0.20±0.04	0.31±0.07
zinc	0.10±0.02	0.12±0.04	0.11±0.02	0.09±0.02	0.10±0.02	0.15±0.04

Tabela 6.16: Wartości metryki IBS (średnie wartości obliczone na podzbiorach w ramach walidacji krzyżowej) dla zbiorów z Tabeli 6.3. Objaśnienia skrótów modeli: RSE = interpretowalny zespół reguł przeżyciowych, SR = *survival rules*, ST = *survival tree*, RSF = *random survival forest*, Cox = regularyzowana regresja Coxa z karą *elastic net*, KM = estymator Kaplana-Meiera.

sja Coxa, metoda RSE wykazuje zbliżoną skuteczność predykcyjną, szczególnie w zbiorach o wysokim stopniu cenzurowania (powyżej 80%): *Tantalum* (98.2%) oraz *PM* (91.7%).

Warto zwrócić szczególną uwagę na relację między metodą RSE a estymatorem bazowym Survival Rules (SR). W 18 z 21 porównywanych przypadków (85.7%) metoda zespołowa RSE nie daje statystycznie istotnej poprawy względem pojedynczego modelu SR. Zjawisko to wynika ze specyfiki algorytmów indukcji reguł. W przeciwieństwie do zespołów modeli drzewiastych, które — mimo własnych mechanizmów regularyzacji (ograniczanie głębokości, minimalna liczba próbek w liściu, przycinanie) — dodatkowo zmniejszają wariancję dzięki mniejszej współzależności modeli bazowych (próbkowanie bootstrap, losowe podzbiory cech) i uśrednianiu, reguły przeżycia charakteryzują się na ogół niższą wariancją już na poziomie pojedynczego modelu (ograniczanie złożoności i długości reguł, minimalne progi wsparcia/liczby zdarzeń, kryteria zatrzymania, przycinanie). W efekcie, łączenie wielu modeli rzadko prowadzi do istotnego wzrostu trafności predykcyjnej, ale zwiększa stabilność i ogólność modelu poprzez agregację zróżnicowanych reguł.

W tabelach wyników niektóre wartości oznaczono jako brakujące (—) ze względu na ograniczenia obliczeniowe lub specyfikę algorytmów. W zbiorze *LaminatePanel* brak wyników wynika z jednoatrybutowej struktury danych — w przypadku RSE uniemożliwia ona wytrenowanie wymaganej liczby modeli składowych zespołu, natomiast w przypadku RandomSurvivalForest prowadzi do błędów numerycznych podczas trenowania w implementacji pakietu *scikit-survival*.

Aby potwierdzić różnice między metodami zastosowano nieparametryczny test post hoc Dunna, oparty na różnicach średnich rang (średnich pozycji wyników w grupach odpowiadających porównywanym metodom po uporządkowaniu wszystkich obserwacji). Korekta Bonferroniego dostosowuje poziom istotności do liczby porównań, kontrolując łączny błąd I rodzaju (odrzucenie hipotezy zerowej, która w rzeczywistości jest prawdziwa). Analiza metryki IBS dla wszystkich zbiorów medycznych i przemysłowych (Tabela 6.17) wskazuje na istotnie gorsze wyniki estymatorów Kaplana-Meiera oraz SurvivalTree względem RSE ($p < 0.05$). Różnice względem RandomSurvivalForest, regresji Coxa i SurvivalRules nie osiągnęły istotności statystycznej ($p > 0.05$).

RSE uzyskuje wyniki porównywalne z najlepszymi metodami referencyjnymi w zbiorach o wysokim stopniu cenzurowania, czego przykładem są zbiory *Tantalum*

Model	Średnia (kontrolny)	Średnia (testowany)	Różnica	p-wartość	Istotna
RSF	0.1464	0.1351	-0.0113	0.4678	Nie
KME	0.1464	0.1694	0.0230	0.0000	Tak
SR	0.1464	0.1374	-0.0090	0.6618	Nie
Cox	0.1464	0.1413	-0.0051	2.7422	Nie
ST	0.1464	0.1899	0.0435	0.0004	Tak

Tabela 6.17: Test istotności różnic dla metryki IBS (test Dunna z korektą Bonferroniego). Model kontrolny: RuleSurvivalEnsemble. Poziom istotności: $\alpha = 0.05$. Dane połączone ze wszystkich zbiorów. Objaśnienia skrótów modeli: RSE = interpretowalny zespół reguł przeżyciowych, SR = *survival rules*, ST = *survival tree*, RSF = *random survival forest*, Cox = regularyzowana regresja Coxa z karą *elastic net*, KM = estymator Kaplana-Meiera.

(98.2% cenzurowania) i *Maintenance* (60.3% cenzurowania). W takich scenariuszach niewielka liczba obserwowanych zdarzeń utrudnia ocenę modeli, a własności RSE — bagging na próbkach bootstrap, losowy dobór cech oraz uśrednianie krzywych KM reguł — wspierają stabilną estymację funkcji przeżycia.

Ograniczeniem metody RSE jest brak systematycznej poprawy względem estymatora bazowego SurvivalRules, co pokazuje, że agregacja reguł nie redukuje obciążenia modelu bazowego. Zaleta podejścia zespołowego jest natomiast redukcja wariancji i większa stabilność predykcji między powtórzeniami uczenia (różne próbkowania bootstrap i losowy dobór cech). W efekcie uzyskiwana jest bardziej powtarzalna estymacja funkcji przeżycia przy porównywalnej średniej jakości predykcyjnej.

Dodatkowym ograniczeniem RSE jest brak możliwości bezpośredniej oceny zdolności dyskryminacyjnej. Wynika to z charakteru predykcji — RSE (podobnie jak SR) estymuje dla każdej obserwacji krzywą przeżycia, a nie skalarny wskaźnik ryzyka umożliwiający jednoznaczne porządkowanie par obserwacji. W efekcie typowe miary dyskryminacji oparte na rangowaniu (porządkowaniu par obserwacji według wartości przewidywanego wskaźnika ryzyka) nie są tu stosowalne, a porównania ogranicza się do metryk kalibracji, takich jak zintegrowany wskaźnik Briera. W zastosowaniach, w których priorytetem jest interpretowalność i stabilna

estymacja funkcji przeżycia, ograniczenie to może być akceptowalne ze względu na większą przejrzystość procesu decyzyjnego.

Interpretowalny zespół reguł przeżyciowych (RSE) uzupełnia zestaw metod, zapewniając kompromis między interpretowalnością a dokładnością predykcji. W zbiorach o wysokim cenzurowaniu (np. *Tantalum*, *Maintenance*) utrzymuje niską wariancję estymacji funkcji przeżycia i osiąga wyniki porównywalne z metodami referencyjnymi. Nie wykazuje jednak systematycznej przewagi nad estymatorem bazowym SurvivalRules. Wartość zespołu polega na redukcji wariancji poprzez agregację niezależnych estymatorów oraz zwiększeniu stabilności wyników.

Wyniki mogą zostać uzupełnione o analizę interpretowalności dla wszystkich zbiorów danych. Dla każdego z nich można wyznaczyć globalny ranking ważności cech na podstawie spadku wartości IBS po permutacji. Jako reprezentatywny przykład wybrano zbiór *NiCdBattery* (umiarkowany poziom cenzurowania, mieszane typy cech, wystarczająca liczba cech), a szczegółową analizę przedstawiono w dalszej części rozdziału. Dla pozostałych zbiorów procedura przebiega analogicznie.

Studium przypadku: Analiza stabilności zespołu dla testów żywotności akumulatorów NiCd

Analiza działania interpretowalnego zespołu reguł przeżyciowych została przeprowadzona na zbiorze danych *NiCdBattery*, dla którego metoda RSE osiągnęła wynik IBS $= 0.11 \pm 0.02$. Wybrano go ze względu na umiarkowany poziom cenzurowania oraz mieszaną reprezentacją atrybutów (numeryczne i kategoryczne). Zespół obejmował 20 estymatorów bazowych trenowanych na próbkach bootstrap z losową selekcją $\lfloor\sqrt{8}\rfloor = 2$ cech na model.

Zbiór *NiCdBattery* zawiera 8 cech objaśniających: *discharge_depth*, *discharge_time*, *charge_time*, *recharge_level*, *koh_concentration*, *koh_volume*, *precharge_time*, *degrees_c*. Cechy te opisują odpowiednio: poziom rozładowania ogniwa, czas rozładowania, czas ładowania, docelowy poziom doładowania po ładowaniu, stężenie roztworu KOH (elektrolitu), objętość roztworu KOH, czas wstępnego ładowania oraz temperaturę procesu (w stopniach Celsjusza).

Łącznie uzyskano 69 reguł, przy czym na estymator przypadało od 2 do 10 reguł (mediana = 3).

Analiza pokrycia na reprezentatywnych próbkach testowych wskazuje, że liczba reguł aktywowanych dla pojedynczej obserwacji zależy od konfiguracji uczenia (m.in. progów wsparcia i liczby estymatorów) i różni się między uruchomieniami. Ze tego względu nie podaje się jednej stałej wartości.

Przykładowe reguły (rzeczywiste, z bieżącego uruchomienia):

- jeśli $degrees\_c \in [35, +\infty)$ to S(t)
- jeśli $discharge\_depth \in (-\infty, 50)$ to S(t)
- jeśli $(degrees\_c \in (-\infty, 35) \wedge koh\_concentration \in [32, +\infty) \wedge charge\_time \in (-\infty, 1.25) \wedge discharge\_time \in (-\infty, 2.50))$ to S(t)
- jeśli $degrees\ c \in [35, +\infty)$ to S(t)
- jeśli $discharge\ depth \in (-\infty, 50)$ to S(t)
- jeśli $(degrees\ c \in (-\infty, 35) \wedge koh\ concentration \in [32, +\infty) \wedge charge\ time \in (-\infty, 1.25) \wedge discharge\ time \in (-\infty, 2.50))$ to S(t)

Analizę globalnej ważności cech oparto na ważności permutacyjnej [79, 59]. Wpływ danej cechy mierzono jako spadek jakości (IBS) po losowym przemieszaniu je wartości w zbiorze testowym, przy niezmienionych pozostałych zmiennych. Różnica względem jakości bazowej definiowała ważność cechy.

Do obliczania IBS stosowano siatkę czterech punktów czasowych wybieranych automatycznie i równomiernie w zakresie obserwowanych czasów w zbiorze testowym $[t_{\min}, t_{\max})$. Punkt $t_{\max}$ pominięto, aby zapewnić zgodność z wymaganiami metryki. W razie potrzeby siatkę przycinano do dopuszczalnego zakresu obserwacji.

W literaturze najczęściej stosuje się dwa warianty wyboru siatki czasów do obliczania zintegrowanego wskaźnika Briera: wszystkie unikalne czasy zdarzeń w zbiorze testowym lub gęstą, regularną siatkę (np. 50–200 punktów) ograniczoną do przedziału $[t_{\min}, t_{\max})$. W niniejszym studium przyjęto uproszczoną, równomierną siatkę czterech punktów jako kompromis obliczeniowy dla niewielkich zbiorów, przy zachowaniu zgodności z wymaganiami metryki (wykluczenie punktu $t_{\max}$ oraz przycinanie do rzeczywistego horyzontu obserwacji). Wybór ten nie wpływa na interpretację metody, a w eksperymentach potwierdzono stabilność kolejności najważniejszych cech także po zagęszczeniu siatki.

Cecha	Spadek IBS (po permutacji)
discharge_depth	0.0282 ± 0.0217
degrees_c	0.0112 ± 0.0060
precharge_time	0.0037 ± 0.0009
koh_concentration	0.0012 ± 0.0003
koh_volume	0.0005 ± 0.0007

Tabela 6.18: Wyniki interpretowalności dla zbioru *NiCdBattery*: ważność permutacyjna (spadek IBS po permutacji cechy).

Zgodnie z miarą permutacyjną dla zbioru *NiCdBattery* (Tabela 6.18), najwyższą ważność uzyskała cecha *discharge_depth* (spadek IBS po permutacji = 0.0282 ± 0.0217). Kolejne w rankingu to: *degrees_c* (0.0112 ± 0.0060), *precharge_time* (0.0037 ± 0.0009), *koh_concentration* (0.0012 ± 0.0003) oraz *koh_volume* (0.0005 ± 0.0007). Ranking wskazuje, że główne determinanty ryzyka w tym zadaniu związane są z głębokością rozładowania, temperaturą oraz parametrami procesu ładowania/rozładowania ogniw.

Bezwzględne spadki IBS są niewielkie, co wynika z faktu, że typowe wartości tej metryki mieszczą się w przedziale około 0-0.25. Dla zbioru *NiCdBattery* wynik bazowy wynosi IBS ≈ 0.11, zatem spadek 0.0282 oznacza pogorszenie rzędu 25% względem wartości bazowej, co wskazuje na istotny wpływ analizowanej cechy na jakość predykcji.

Porównanie z metodami referencyjnymi potwierdza charakterystyczny profil interpretowalnego zespołu. Przy wyniku IBS $= 0.03 \pm 0.00$ metoda RSE osiąga skuteczność porównywalną z regresją Coxa (IBS $= 0.05 \pm 0.01$), przewyższając estymator Kaplana-Meiera (IBS $= 0.05 \pm 0.01$). Random Survival Forest uzyskuje lepszy wynik (IBS $= 0.00 \pm 0.00$), ale kosztem utraty interpretowalności na poziomie poszczególnych decyzji.

Empiryczna walidacja czterech autorskich metod dostarcza kompleksowej charakterystyki możliwości i ograniczeń interpretowalnych podejść do analizy przeżycia. Każdy algorytm wykazuje unikalne cechy i znajduje zastosowanie w różnych scenariuszach, takich jak eksploracja danych, generowanie rekomendacji predykcyjnych, identyfikacja anomalii oraz stabilna estymacja zespołowa. Obserwowane zróżnicowanie charakterystyk predykcyjnych między domeną przemysłową (predykcyjne

utrzymanie ruchu) a medyczną (badania kliniczne) potwierdza możliwość zastosowania proponowanych metod w różnych kontekstach, czyli ich uniwersalność w interpretowalnej analizie przeżycia.

7 Podsumowanie

Analiza niezawodności i przeżycia obejmuje zbiór metod służących do modelowania danych typu czas-do-zdarzenia, charakteryzujących się występowaniem cenzurowania. Wymaga to stosowania podejść uwzględniających tę specyfikę. Pomimo dostępności takich metod oraz zaawansowanych algorytmów uczenia maszynowego typu „czarna skrzynka", istnieje luka badawcza w zakresie interpretowalnych metod uczenia maszynowego dedykowanych do analizy danych cenzurowanych. W praktyce wykorzystuje się m.in. modele proporcjonalnych hazardów Coxa, parametryczne modele czasu życia (np. wykładniczy, Weibulla, log-normalny) oraz drzewa przeżyciowe i ich zespoły. Modele Coxa ograniczają możliwość uchwycenia nieliniowości i interakcji, pojedyncze drzewa często ustępują dokładnością podejściom zespołowym, natomiast zespoły (np. przeżyciowe lasy losowe) cechują się ograniczoną transparentnością procesu podejmowania decyzji. Podobna uwaga dotyczy rozbudowanych modeli regułowych — wraz ze wzrostem liczby reguł i liczby warunków w regułach maleje ich transparentność. Jednocześnie w literaturze dominują zastosowania o charakterze predykcyjnym, podczas gdy inne zastosowania reguł, takie jak reguły akcji i reguły wyjątków, pozostają słabiej rozpoznane w kontekście danych cenzurowanych. Brakuje zatem metod interpretowalnych, które wprost uwzględniają cenzurowanie, zapewniają wgląd w mechanizm decyzyjny i jednocześnie oferują dobrą skuteczność predykcyjną. Problem ten nabiera szczególnego znaczenia w zastosowaniach o podwyższonych wymaganiach, gdzie decyzje modeli mogą wpływać na bezpieczeństwo operacyjne, efektywność ekonomiczną lub ciągłość działania. Ograniczona interpretowalność utrudnia w takich przypadkach walidację i wdrażanie modeli.

Celem niniejszej rozprawy doktorskiej było opracowanie interpretowalnych metod analizy niezawodności i przeżycia, wykorzystujących algorytmy indukcji reguł logicznych dostosowanych do specyfiki danych cenzurowanych, oraz wykazanie ich skuteczności i potencjalnej użyteczności w zastosowaniach medycznych oraz

w predykcyjnym utrzymaniu ruchu. Praca koncentrowała się na rozwoju nowych algorytmów łączących interpretowalność z dobrą skutecznością predykcyjną, umożliwiających ekspertom dziedzinowym zrozumienie i weryfikację mechanizmów decyzyjnych.

W ramach realizacji głównego celu opracowano cztery oryginalne algorytmy, które stanowią zasadniczy wkład naukowy pracy. Pokryciowy algorytm indukcji przeżyciowych reguł akcji adaptuje strategię sekwencyjnego pokrywania do danych cenzurowanych, odkrywając wzorce regułowe. Algorytm rekomendacji przekłada przeżyciowe reguły akcji na zalecenia dla pojedynczych obiektów, wykorzystując m.in. meta-tablicę oraz weryfikację statystyczną. Algorytm przeżyciowych reguł wyjątków (CR, RR, ER) stosuje sekwencyjne wyszukiwanie wyjątków dostosowane do danych przeżyciowych. Z kolei interpretowalny zespół reguł przeżyciowych agreguje modele regułowe, zachowując interpretowalność. Metody te łączy wspólna reprezentacja decyzyjna w postaci reguł typu „jeśli-to", możliwość prześledzenia uzasadnień decyzji na poziomie reguł oraz dostosowanie do specyfiki danych cenzurowanych. Na tle istniejącej literatury wprowadzają one nowe elementy, w tym adaptację strategii sekwencyjnego pokrywania do analizy przeżycia, procedurę rekomendacyjną opartą na meta-tablicy z mechanizmem rozstrzygania konfliktów reguł, formalne ujęcie wyjątków (CR, RR, ER) dla danych przeżyciowych oraz interpretowalny zespół reguł umożliwiający odtworzenie podstaw decyzji. W konsekwencji zaproponowane algorytmy stanowią rozwiązania nowe bądź rozwijające istniejące nurty metod regułowych.

Pokryciowy algorytm indukcji przeżyciowych reguł akcji i algorytm rekomendacji są komplementarne — pierwszy odkrywa reguły wyodrębniające podgrupy o odmiennych krzywych przeżycia, a drugi przekłada je na zalecenia dla pojedynczych obiektów. Zastosowanie sekwencyjnego pokrywania do danych cenzurowanych pozwala maksymalizować wartość statystyki log-rank mierzącej różnicę między krzywymi przeżycia dla obserwacji spełniających regułę a pozostałymi obserwacjami (wyższa wartość wskazuje na większą rozbieżność między krzywymi), kontrolować nakładanie się i pokrycie reguł oraz ograniczać liczbę warunków w regułach. W konsekwencji prowadzi to do zbioru reguł o krótkich, zrozumiałych przesłankach i niskim nakładaniu pokryć (niewielkim odsetku obserwacji spełniających wiele reguł). Cenzurowanie uwzględniono poprzez zastosowanie miar specyficznych dla analizy przeżycia (KM, log-rank), a interpretowalność uzyskuje

się dzięki regułom. W eksperymentach (Rozdział 6) reguły akcji osiągały wysoki odsetek istotności w teście log-rank dla obu domen, a rekomendacje cechowały się wysoką spójnością z niezależnym modelem arbitra (wysokie wartości $S_{0.05}$ i niskie MAE), przy jednocześnie niewielkiej liczbie modyfikowanych atrybutów. Ograniczeniem metody jest asocjacyjny charakter rekomendacji — reguły opisują współwystępowanie określonych konfiguracji atrybutów (i sugerowanych akcji) z odmiennym kształtem krzywych przeżycia, ale nie dowodzą istnienia związku przyczynowo-skutkowego pomiędzy tymi konfiguracjami. Potencjalne zastosowania obejmują predykcyjne utrzymanie ruchu (harmonogramy i parametry eksploatacji), medycynę (scenariusze modyfikacji terapii) oraz analizy typu „jeśli-to".

Algorytm przeżyciowych reguł wyjątków umożliwia odkrywanie rzadkich, nieoczywistych wzorców w danych cenzurowanych, przy zachowaniu interpretowalności wynikającej ze struktury reguł. Podejście opiera się na wyznaczeniu zbioru trzech reguł — reguły bazowej (CR), reguły referencyjnej (RR) oraz reguły wyjątku (ER). Reguła bazowa opisuje typowe relacje obecne w danych. Regułę referencyjną (RR) definiuje się tak, aby krzywa KM nie różniła się istotnie od CR. Pozwala to odfiltrować pozorne wyjątki — gdyby sama RR odpowiadała innej krzywej KM niż CR, różnica zostałaby wykryta w porównaniu CR-RR i RR nie pełniłaby roli referencyjnej. W konsekwencji istotne różnice obserwuje się dopiero dla reguły wyjątku, tj. koniunkcji CR $\wedge$ RR, dla której krzywa przeżycia różni się istotnie zarówno od CR, jak i od RR. Porównania krzywych KM w parach CR-RR, CR-ER oraz RR-ER przeprowadza się testem log-rank. Takie podejście ogranicza liczbę pozornych wyjątków (sytuacji, w których sama RR różni się istotnie od CR, więc ich kombinacja nie stanowi rzeczywistego wyjątku) i dostarcza uzasadnienia, dlaczego dana reguła jest wyjątkiem. Wyjątki pokrywają zwykle niewielkie zbiory obserwacji, co może prowadzić do mało precyzyjnych oszacowań (szersze przedziały ufności KM). Stąd uzyskane rezultaty wymagają ostrożnej interpretacji i potwierdzenia w ocenie eksperckiej. Potencjalne zastosowania obejmują wczesne ostrzeganie o nieoczywistych ryzykach w utrzymaniu ruchu, identyfikację fenotypów o odmiennym rokowaniu w medycynie oraz generowanie hipotez badawczych.

Interpretowalny zespół reguł przeżyciowych to model predykcyjny dla danych cenzurowanych, łączący interpretowalność modeli regułowych z jakością predykcji technik zespołowych. W założeniu podejście redukuje wariancję predykcji poprzez agregację predykcji wielu modeli regułowych i zachowuje interpretowalność na

poziomie reguł dzięki możliwości śledzenia ich aktywacji i oceny ich wkładów do estymowanej funkcji przeżycia zespołu. Estymator bazowy to zbiór reguł przeżyciowych trenowany metodą *separate-and-conquer* na próbkach bootstrap w losowym podzbiorze cech. Dla każdej reguły w zbiorze estymowana jest krzywa KM dla pokrywanych obserwacji. Predykcja estymatora powstaje przez uśrednianie krzywych KM aktywnych reguł, zaś predykcja zespołu przez agregację krzywych dla wszystkich estymatorów. W eksperymentach uzyskano porównywalne wartości metryki IBS względem *random survival forest* i modelu Coxa w części zbiorów, szczególnie przy wysokim cenzurowaniu, zachowując jednocześnie transparentność procesu decyzyjnego. Ograniczeniem jest brak skalarnego wskaźnika ryzyka, co uniemożliwia ocenę zdolności dyskryminacyjnej. Przewaga nad pojedynczym modelem regułowym nie zawsze jest istotna statystycznie, a złożoność rośnie wraz z liczbą estymatorów bazowych. Potencjalne zastosowania obejmują scenariusze wymagające stabilnych, powtarzalnych krzywych przeżycia z możliwością weryfikacji uzasadnień na poziomie reguł.

Przeprowadzono walidację empiryczną zaproponowanych metod na łącznie 27 zbiorach danych z domen przemysłowej oraz medycznej, zróżnicowanych pod względem liczby cech, liczby obserwacji i poziomu cenzurowania. Procedura oceny była dostosowana do charakteru metod — dla modeli zespołowych wyznaczano zintegrowany wskaźnik Briera, dla reguł akcji i wyjątków stosowano test log-rank oraz metryki opisujące strukturę i pokrycie reguł, natomiast dla algorytmu rekomendacji raportowano spójność z niezależnym modelem walidacyjnym i błąd MAE. Wyniki wskazują, że zbiory reguł akcji w większości przypadków dostarczały statystycznie istotnych wzorców, a generowane rekomendacje były spójne z modelem arbitra przy niewielkiej liczbie modyfikowanych atrybutów. Reguły wyjątków umożliwiały wykrywanie podgrup o nietypowych wzorcach w większości zbiorów. Zespół reguł przeżyciowych osiągał wyniki porównywalne z metodami referencyjnymi i przewyższał estymator Kaplana-Meiera, w szczególności przy wysokim udziale obserwacji cenzurowanych.

Przykładowe obszary zastosowań obejmują medycynę oraz predykcyjne utrzymanie ruchu. W medycynie metody te mogą wspierać personalizację terapii, np. poprzez identyfikację podgrup wymagających odmiennego postępowania oraz formułowanie hipotez dotyczących modyfikacji leczenia. W predykcyjnym utrzymaniu ruchu mogą być wykorzystywane przy projektowaniu strategii konserwacyjnych

poprzez identyfikację urządzeń o nietypowych wzorcach degradacji i analizę scenariuszy interwencji prewencyjnych. W ujęciu ogólnym opracowane metody wpisują się w nurt interpretowalnego uczenia maszynowego, kładąc nacisk na przejrzystość procesu podejmowania decyzji przez modele.

Opracowane metody wnoszą również wkład teoretyczny. Sformalizowano strategię sekwencyjnego pokrywania dla danych cenzurowanych, definiując funkcję celu jako maksymalizację statystyki log-rank między krzywymi przeżycia części źródłowej i docelowej oraz kryteria wzrostu i przycinania oparte na statystyce log-rank i estymatorze Kaplana-Meiera. Zdefiniowano trójskładnikową strukturę reguł (CR, RR, ER) dla analizy przeżycia wprowadzając procedurę weryfikacji hipotez opartą na teście log-rank dla par CR-RR, CR-ER i RR-ER oczekując braku istotnej różnicy między CR a RR, przy jednocześnie istotnych różnicach między CR a ER oraz RR a ER. Ponadto sformułowano schemat rekomendacyjny jako problem rozwiązywania konfliktów reguł oraz zaproponowano konstrukcję interpretowalnego zespołu reguł przeżyciowych wraz z metodą agregacji krzywych oraz estymacją ważności cech (m.in. metodą permutacyjną) i śledzeniem aktywacji reguł. Elementy te porządkują terminologię i ramy formalne łączące indukcję reguł z analizą przeżycia i mogą stanowić podstawę do dalszych analiz teoretycznych oraz rozszerzeń.

Wyniki przedstawione w Rozdziale 6 obejmują: istotność zidentyfikowanych wzorców w teście log-rank dla reguł akcji, spójność rekomendacji z niezależnym modelem walidacyjnym (wysokie wartości $S_{0.05}$ i niskie MAE), wykrywanie odrębnych podgrup przez reguły wyjątków oraz porównywalne wartości IBS zespołu reguł względem *random survival forest* i Cox. Zespół reguł uzyskuje niższe wartości IBS niż estymator Kaplana-Meiera, zwłaszcza przy wysokim cenzurowaniu. Wyniki te łącznie potwierdzają tezę pracy o możliwości opracowania skutecznych i transparentnych metod analizy niezawodności i przeżycia w oparciu o interpretowalne algorytmy regułowe dostosowane do danych cenzurowanych.

Ograniczeniem przeprowadzonych badań jest m.in. koncentracja na wybranych klasach reguł: regułach akcji, regułach wyjątków oraz zespole reguł. Inne paradygmaty regułowe (np. reguły asocjacyjne, zbiory kontrastowe) nie były przedmiotem analizy. Analiza złożoności obliczeniowej nie była przedmiotem niniejszej rozprawy. Wiarygodna charakterystyka wymagałaby parametrycznej, etapowej analizy zależności od struktury danych, sposobu dyskretyzacji, kryteriów stopu i przyjętych

heurystyk wyszukiwania, co zaplanowano jako kierunek dalszych prac. Dodatkowo, walidację przeprowadzono na ograniczonym zestawie danych. Liczba publicznie dostępnych zbiorów benchmarkowych dla analizy przeżycia jest ograniczona w porównaniu z klasyfikacją i regresją. W niniejszej pracy wykorzystano publicznie dostępne zbiory medyczne oraz syntetyczne dane z obszaru predykcyjnego utrzymania ruchu, co należy uwzględnić przy uogólnianiu wniosków na inne domeny.

Kierunki dalszych badań obejmują m.in. rozszerzenie metod na inne typy reguł. Dotyczy to zarówno reguł asocjacyjnych dostosowanych do danych przeżyciowych jak i zbiorów kontrastowych, rozumianych jako wzorce istotnie różnicujące krzywą przeżycia wyodrębnionych podzbiorów względem zbioru odniesienia. Kierunek ten nawiązuje do wyników [153], które wykazały skuteczność heurystyki *separate-and-conquer* w wyznaczaniu zbiorów kontrastowych także dla danych przeżyciowych. Z perspektywy kosztów obliczeń kierunkiem rozwojowym są algorytmy przyrostowe (ang. *online*) oraz metody przybliżone (próbkowanie, ograniczanie przestrzeni wyszukiwania), które znajdują zastosowanie w analizie danych strumieniowych i w warunkach ograniczonych zasobów.

Spis skrótów

AFT	model przyspieszonego czasu do awarii (ang. *accelerated failure time*)
AUC	pole pod krzywą (ang. *Area Under the Curve*)
CBM	strategia utrzymania ruchu na podstawie stanu technicznego (ang. *Condition-Based Maintenance*)
CDF	dystrybuanta skumulowana (ang. *cumulative distribution function*)
CR	reguła bazowa (ang. *commonsense rule*)
ER	reguła wyjątku (ang. *exception rule*)
GAM	uogólnione modele addytywne (ang. *generalized additive models*)
IBS	zintegrowany błąd Briera (ang. *Integrated Brier Score*)
IoT	Internet Rzeczy (ang. *Internet of Things*)
k-**NN**	algorytm k-najbliższych sąsiadów (ang. *k-nearest neighbors*)
MAE	średni bezwzględny błąd (ang. *mean absolute error*)
MSE	średni błąd kwadratowy (ang. *mean square error*)
PdM	predykcyjne utrzymanie ruchu (ang. *predictive maintenance*)
PDP	wykresy zależności częściowej (ang. *partial dependence plots*)
PM	prewencyjne utrzymanie ruchu (ang. *Preventive Maintenance*)
RR	reguła referencyjna (ang. *reference rule*)
RSE	interpretowalny zespół reguł przeżyciowych (ang. *Rule Survival Ensemble*)
RTF	strategia do wystąpienia uszkodzenia (ang. *Run to Failure*)
TTF	czas do awarii (ang. *time-to-failure*)
Cox	regularyzowana regresja Coxa z karą *elastic net*
KM	estymator Kaplana-Meiera
RSF	*random survival forest*
SR	Survival Rules
ST	Survival Tree

Bibliografia

[1] Cox, D. R. „Regression Models and Life-Tables". W: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 34.2 (1 sty. 1972), s. 187–202. ISSN: 1369-7412, 1467-9868. DOI: 10.1111/j.2517-6161.1972.tb00899.x. (Term. wiz. 23. 02. 2025).

[2] Hrnjica, B. i Softic, S. „The Survival Analysis for a Predictive Maintenance in Manufacturing". W: *Advances in Production Management Systems. Artificial Intelligence for Sustainable and Resilient Production Systems*. Red. Dolgui, A. i in. T. 632. Springer International Publishing, 2021, s. 78–85. ISBN: 978-3-030-85905-3. DOI: 10.1007/978-3-030-85906-0_9. (Term. wiz. 24. 02. 2025).

[3] Mobley, R. K. *An Introduction to Predictive Maintenance*. 2nd ed. Butterworth-Heinemann, 2002. 438 s. ISBN: 978-0-7506-7531-4.

[4] Collett, D. *Modelling Survival Data in Medical Research*. 3 wyd. Chapman and Hall/CRC, 4 maj. 2015. 548 s. ISBN: 978-0-429-19629-4. DOI: 10.1201/b18041.

[5] O'Connor, P. D. T. i Kleyner, A. *Practical Reliability Engineering*. 1 wyd. Wiley, 23 grud. 2011. ISBN: 978-0-470-97982-2. DOI: 10.1002/9781119961260. (Term. wiz. 23. 02. 2025).

[6] Modarres, M., Amiri, M. i Jackson, C. *Probabilistic Physics of Failure Approach to Reliability: Modeling, Accelerated Testing, Prognosis and Reliability Assessment*. 1 wyd. Wiley, 12 czer. 2017. ISBN: 978-1-119-38863-0. DOI: 10.1002/9781119388692. (Term. wiz. 03. 03. 2025).

[7] Allison, P. D. „Discrete-Time Methods for the Analysis of Event Histories". W: *Sociological Methodology* 13 (1982), s. 61. ISSN: 00811750. DOI: 10.2307/270718. JSTOR: 270718. (Term. wiz. 02. 08. 2025).

[8] Blossfeld, H.-P., Hamerle, A. i Mayer, K. U. *Event History Analysis: Statistical Theory and Application in the Social Sciences*. Psychology Press, 24 lut. 2014. 298 s. ISBN: 978-1-315-80816-1. DOI: 10.4324/9781315808161.

[9] Yamaguchi, K. *Event History Analysis*. Applied Social Research Methods Series v. 28. Sage Publications, 1991. 182 s. ISBN: 978-0-8039-3323-1.

[10] Carvalho, T. P. i in. „A Systematic Literature Review of Machine Learning Methods Applied to Predictive Maintenance". W: *Computers & Industrial Engineering* 137 (1 list. 2019), s. 106024. ISSN: 0360-8352. DOI: 10.1016/j.cie.2019.106024. (Term. wiz. 24. 02. 2025).

[11] LeBlanc, M. i Crowley, J. „Relative Risk Trees for Censored Survival Data". W: *Biometrics. Journal of the International Biometric Society* 48.2 (1992), s. 411–425.

[12] Gudyś, A., Sikora, M. i Wróbel, Ł. „RuleKit: A Comprehensive Suite for Rule-Based Learning". W: *Knowledge-Based Systems* 194 (kw. 2020), s. 105480. ISSN: 09507051. DOI: 10.1016/j.knosys.2020.105480. (Term. wiz. 23. 09. 2024).

[13] Katzman, J. L. i in. „DeepSurv: Personalized Treatment Recommender System Using a Cox Proportional Hazards Deep Neural Network". W: *BMC Medical Research Methodology* 18.1 (26 lut. 2018), s. 24. ISSN: 1471-2288. DOI: 10.1186/s12874-018-0482-1. (Term. wiz. 09. 07. 2025).

[14] Molnar, C. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. 3:e upplagan. Christoph Molnar, 2025. ISBN: 978-3-911578-03-5.

[15] Fürnkranz, J. „Separate-and-Conquer Rule Learning". W: *Artificial Intelligence Review* 13.1 (1999), s. 3–54. ISSN: 02692821. DOI: 10.1023/A:1006524209794. (Term. wiz. 17. 03. 2025).

[16] Wrobel, S. „An Algorithm for Multi-Relational Discovery of Subgroups". W: *Principles of Data Mining and Knowledge Discovery*. Red. Komorowski, J. i Zytkow, J. Springer, 1997, s. 78–87. ISBN: 978-3-540-69236-2. DOI: 10.1007/3-540-63223-9_108.

[17] Atzmueller, M. i Puppe, F. „SD-Map – A Fast Algorithm for Exhaustive Subgroup Discovery". W: *Knowledge Discovery in Databases: PKDD 2006*. Red. Fürnkranz, J., Scheffer, T. i Spiliopoulou, M. Springer, 2006, s. 6–17. ISBN: 978-3-540-46048-0. DOI: 10.1007/11871637_6.

[18] Ras, Z. W. i Wieczorkowska, A. „Action-Rules: How to Increase Profit of a Company". W: *Principles of Data Mining and Knowledge Discovery*. Red. Zighed, D. A., Komorowski, J. i Żytkow, J. Zred. Goos, G., Hartmanis, J. i Van Leeuwen, J. T. 1910. Springer Berlin Heidelberg, 2000, s. 587–592. ISBN: 978-3-540-41066-9. DOI: 10.1007/3-540-45372-5_70. (Term. wiz. 18. 03. 2025).

[19] Tzacheva, A. A. i Raś, Z. W. „Action Rules Mining". W: *International Journal of Intelligent Systems* 20.7 (2005), s. 719–736.

[20] Suzuki, E. „Undirected Discovery of Interesting Exception Rules". W: *International Journal of Pattern Recognition and Artificial Intelligence* 16.08 (grud. 2002), s. 1065–1086. ISSN: 0218-0014. DOI: 10.1142/S0218001402002155. (Term. wiz. 31. 08. 2025).

[21] Bay, S. D. i Pazzani, M. J. „Detecting Group Differences: Mining Contrast Sets". W: *Data Mining and Knowledge Discovery* 5.3 (1 lip. 2001), s. 213–246. ISSN: 1573-756X. DOI: 10.1023/A:1011429418057. (Term. wiz. 31. 08. 2025).

[22] Stefanowski, J. i Vanderpooten, D. „Induction of Decision Rules in Classification and Discovery-Oriented Perspectives". W: *International Journal of Intelligent Systems* 16.1 (sty. 2001), s. 13–27. ISSN: 0884-8173, 1098-111X. DOI: 10.1002/1098-111X(200101)16:1<13::AID-INT3>3.0.CO;2-M. (Term. wiz. 17. 03. 2025).

[23] Bazan, J. G. i in. „Rough Set Approach to the Survival Analysis." W: *Rough Sets and Current Trends in Computing*. Red. Alpigini, J. J. i in. T. 2475. Lecture Notes in Computer Science. Springer, 2002, s. 522–529. ISBN: 3-540-44274-X.

[24] Pattaraintakorn, P. i Cercone, N. „A Foundation of Rough Sets Theoretical and Computational Hybrid Intelligent System for Survival Analysis". W: *Computers & Mathematics with Applications* 56.7 (1 paź. 2008), s. 1699–

1708. ISSN: 0898-1221. DOI: 10.1016/j.camwa.2008.04.030. (Term. wiz. 09. 07. 2025).

[25] Kagermann, H. „Recommendations for Implementing the Strategic Initiative INDUSTRIE 4.0". W: 1 kw. 2013. (Term. wiz. 09. 07. 2025).

[26] Ahmed Murtaza, A. i in. „Paradigm Shift for Predictive Maintenance and Condition Monitoring from Industry 4.0 to Industry 5.0: A Systematic Review, Challenges and Case Study". W: *Results in Engineering* 24 (1 grud. 2024), s. 102935. ISSN: 2590-1230. DOI: 10.1016/j.rineng.2024.102935. (Term. wiz. 23. 02. 2025).

[27] *PN-EN 13306:2018 — Maintenance — Maintenance Terminology*. 12 sty. 2013.

[28] Wyczółkowski, R., Mazurkiewicz, D. i Jasiulewicz-Kaczmarek, M., red. *Strategie i Metody Utrzymania Ruchu*. Przemysł 4.0 : Ryszard Knosala Poleca. Polskie Wydawnictwo Ekonomiczne, 2023. 285 s. ISBN: 978-83-208-2509-1.

[29] Meeker, W. Q., Escobar, L. A. i Pascual, F. G. *Statistical Methods for Reliability Data*. Second edition. Wiley Series in Probability and Statistics. Wiley, 2022. 659 s. ISBN: 978-1-118-11545-9.

[30] Dalzochio, J. i in. „Machine Learning and Reasoning for Predictive Maintenance in Industry 4.0: Current Status and Challenges". W: *Computers in Industry* 123 (1 grud. 2020), s. 103298. ISSN: 0166-3615. DOI: 10.1016/j.compind.2020.103298. (Term. wiz. 24. 02. 2025).

[31] Chen, C. i in. „Predictive Maintenance Using Cox Proportional Hazard Deep Learning". W: *Advanced Engineering Informatics* 44 (1 kw. 2020), s. 101054. ISSN: 1474-0346. DOI: 10.1016/j.aei.2020.101054. (Term. wiz. 23. 02. 2025).

[32] Feng, K. i in. „A Review of Vibration-Based Gear Wear Monitoring and Prediction Techniques". W: *Mechanical Systems and Signal Processing* 182 (1 sty. 2023), s. 109605. ISSN: 0888-3270. DOI: 10.1016/j.ymssp.2022.109605. (Term. wiz. 23. 02. 2025).

[33] Ullah, I. i in. „Predictive Maintenance of Power Substation Equipment by Infrared Thermography Using a Machine-Learning Approach". W: *Energies* 10.12 (12 grud. 2017), s. 1987. ISSN: 1996-1073. DOI: 10.3390/en10121987. (Term. wiz. 07. 08. 2025).

[34] Huda, A. S. N. i Taib, S. „Application of Infrared Thermography for Predictive/Preventive Maintenance of Thermal Defect in Electrical Equipment". W: *Applied Thermal Engineering* 61.2 (3 list. 2013), s. 220–227. ISSN: 1359-4311. DOI: 10.1016/j.applthermaleng.2013.07.028. (Term. wiz. 07. 08. 2025).

[35] Marinescu, A.-D. i in. „Assessing the Opportunity to Use the Infrared Thermography Method for Predictive Maintenance of Hydrostatic Pumps". W: *2017 International Conference on ENERGY and ENVIRONMENT (CIEM)*. 2017 International Conference on ENERGY and ENVIRONMENT (CIEM). Paź. 2017, s. 270–274. DOI: 10.1109/CIEM.2017.8120790. (Term. wiz. 07. 08. 2025).

[36] Keartland, S. i Van Zyl, T. L. „Automating Predictive Maintenance Using Oil Analysis and Machine Learning". W: *2020 International SAUPEC/RobMech/PRASA Conference*. 2020 International SAUPEC/RobMech/PRASA Conference. IEEE, sty. 2020, s. 1–6. ISBN: 978-1-7281-4162-6. DOI: 10.1109/SAUPEC/RobMech/PRASA48453.2020.9041003. (Term. wiz. 07. 08. 2025).

[37] Higham, E. H. i Perovic, S. „Predictive Maintenance of Pumps Based on Signal Analysis of Pressure and Differential Pressure (Flow) Measurements". W: *Transactions of the Institute of Measurement and Control* 23.4 (1 paź. 2001), s. 226–248. ISSN: 0142-3312. DOI: 10.1177/014233120102300402. (Term. wiz. 07. 08. 2025).

[38] Velarde-Suárez, S., Ballesteros-Tajadura, R. i Hurtado-Cruz, J. P. „A Predictive Maintenance Procedure Using Pressure and Acceleration Signals from a Centrifugal Fan". W: *Applied Acoustics* 67.1 (1 sty. 2006), s. 49–61. ISSN: 0003-682X. DOI: 10.1016/j.apacoust.2005.05.006. (Term. wiz. 07. 08. 2025).

[39] Bonaldi, E. i in. „Predictive Maintenance by Electrical Signature Analysis to Induction Motors". W: *Induction Motors - Modelling and Control*. 1 wyd.

InTech, paź. 2012, s. 487–520. ISBN: 978-953-51-0843-6. DOI: 10.5772/48045.

[40] Hedkvist, A. *Predictive Maintenance with Machine Learning on Weld Joint Analysed by Ultrasound*. 2019. (Term. wiz. 07. 08. 2025).

[41] Tareen, S., Herzau, J. i Tianshu, W. „Predictive Maintenance Oriented Pattern Recognition System Based on Ultrasound Data Analysis". W: *2019 14th IEEE International Conference on Electronic Measurement & Instruments (ICEMI)*. 2019 14th IEEE International Conference on Electronic Measurement & Instruments (ICEMI). List. 2019, s. 1208–1214. DOI: 10.1109/ICEMI46757.2019.9101708. (Term. wiz. 07. 08. 2025).

[42] Khan, Z. J. „Predictive Maintenance and Internet of Things". W: *2021 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube)*. 2021 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube). Paź. 2021, s. 1–5. DOI: 10.1109/ICECube53880.2021.9628262. (Term. wiz. 07. 08. 2025).

[43] Civerchia, F. i in. „Industrial Internet of Things Monitoring Solution for Advanced Predictive Maintenance Applications". W: *Journal of Industrial Information Integration*. Enterprise Modelling and System Integration for Smart Manufacturing 7 (1 wrz. 2017), s. 4–12. ISSN: 2452-414X. DOI: 10.1016/j.jii.2017.02.003. (Term. wiz. 07. 08. 2025).

[44] Zhong, D. i in. „Overview of Predictive Maintenance Based on Digital Twin Technology". W: *Heliyon* 9.4 (1 kw. 2023). ISSN: 2405-8440. DOI: 10.1016/j.heliyon.2023.e14534. (Term. wiz. 07. 08. 2025).

[45] Van Dinter, R., Tekinerdogan, B. i Catal, C. „Predictive Maintenance Using Digital Twins: A Systematic Literature Review". W: *Information and Software Technology* 151 (1 list. 2022), s. 107008. ISSN: 0950-5849. DOI: 10.1016/j.infsof.2022.107008. (Term. wiz. 07. 08. 2025).

[46] Abidi, M. H., Mohammed, M. K. i Alkhalefah, H. „Predictive Maintenance Planning for Industry 4.0 Using Machine Learning for Sustainable Manufacturing". W: *Sustainability* 14.6 (6 sty. 2022), s. 3387. ISSN: 2071-1050. DOI: 10.3390/su14063387. (Term. wiz. 07. 08. 2025).

[47] LeCun, Y., Bengio, Y. i Hinton, G. „Deep Learning". W: *Nature* 521.7553 (maj 2015), s. 436–444. ISSN: 1476-4687. DOI: 10.1038/nature14539. (Term. wiz. 09. 07. 2025).

[48] Hung, Y.-H. „Improved Ensemble-Learning Algorithm for Predictive Maintenance in the Manufacturing Process". W: *Applied Sciences* 11.15 (15 sty. 2021), s. 6832. ISSN: 2076-3417. DOI: 10.3390/app11156832. (Term. wiz. 01. 04. 2025).

[49] Medeiros, A. i in. „Failure Analysis of Gear Using Continuous Wavelet Transform Applied in the Context of Wind Turbines". W: *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology* 238.7 (1 lip. 2024), s. 860–868. ISSN: 1350-6501. DOI: 10.1177/13506501241235726. (Term. wiz. 07. 08. 2025).

[50] Elsamanty, M., Ibrahim, A. i Saady Salman, W. „Principal Component Analysis Approach for Detecting Faults in Rotary Machines Based on Vibrational and Electrical Fused Data". W: *Mechanical Systems and Signal Processing* 200 (1 paź. 2023), s. 110559. ISSN: 0888-3270. DOI: 10.1016/j.ymssp.2023.110559. (Term. wiz. 07. 08. 2025).

[51] Chen, C. i in. „Data-Driven Predictive Maintenance Strategy Considering the Uncertainty in Remaining Useful Life Prediction". W: *Neurocomputing* 494 (14 lip. 2022), s. 79–88. ISSN: 0925-2312. DOI: 10.1016/j.neucom.2022.04.055. (Term. wiz. 07. 08. 2025).

[52] Omol, E., Mburu, L. i Onyango, D. „Anomaly Detection In IoT Sensor Data Using Machine Learning Techniques For Predictive Maintenance In Smart Grids". W: *International Journal of Science, Technology & Management* 5.1 (29 sty. 2024), s. 201–210. ISSN: 2722-4015. DOI: 10.46729/ijstm.v5i1.1028. (Term. wiz. 07. 08. 2025).

[53] Klein, J. P. i Moeschberger, M. L. *Survival Analysis: Techniques for Censored and Truncated Data*. 2nd ed. Statistics for Biology and Health. Springer, 2003. 536 s. ISBN: 978-0-387-95399-1.

[54] Blischke, W. R., Karim, M. R. i Murthy, D. N. P. *Warranty Data Collection and Analysis*. Springer Series in Reliability Engineering. Springer London, 2011. ISBN: 978-0-85729-646-7. DOI: 10.1007/978-0-85729-647-4. (Term. wiz. 20. 02. 2025).

[55] Leung, K. M., Elashoff, R. M. i Afifi, A. A. „Censoring Issues in Survival Analysis". W: *Annual Review of Public Health* 18 (1997), s. 83–104. ISSN: 0163-7525. DOI: 10.1146/annurev.publhealth.18.1.83. PMID: 9143713.

[56] Kaplan, E. L. i Meier, P. „Nonparametric Estimation from Incomplete Observations". W: *Journal of the American Statistical Association* 53.282 (1958), s. 457–481. ISSN: 0162-1459. DOI: 10.2307/2281868. JSTOR: 2281868. (Term. wiz. 23. 02. 2025).

[57] Hosmer, D. W., Lemeshow, S. i May, S. *Applied Survival Analysis: Regression Modeling of Time to Event Data*. Wiley-Interscience, 2008.

[58] Tessoni, V. i Amoretti, M. „Advanced Statistical and Machine Learning Methods for Multi-Step Multivariate Time Series Forecasting in Predictive Maintenance". W: *Procedia Computer Science*. 3rd International Conference on Industry 4.0 and Smart Manufacturing 200 (1 sty. 2022), s. 748–757. ISSN: 1877-0509. DOI: 10.1016/j.procs.2022.01.273. (Term. wiz. 07. 08. 2025).

[59] Ishwaran, H. i in. „Random Survival Forests". W: *Ann. Appl. Statist.* 2.3 (2008), s. 841–860.

[60] Breiman, L. i in. *Classification and Regression Trees*. Chapman and Hall/CRC, 1984. 368 s. ISBN: 978-1-315-13947-0. DOI: 10.1201/9781315139470.

[61] Hwang, J. *Reliability Analysis Using MINITAB and Python*. John Wiley & Sons, 2023. 1 s. ISBN: 978-1-119-87078-4.

[62] Djeddi, A. Z. i in. „Gas Turbine Reliability Estimation to Reduce the Risk of Failure Occurrence with a Comparative Study between the Two-Parameter Weibull Distribution and a New Modified Weibull Distribution". W: *Diagnostyka* 23.1 (2022), s. 1–18. ISSN: 1641-6414, 2449-5220. (Term. wiz. 23. 02. 2025).

[63] Hothorn, T. i in. „Survival Ensembles". W: *Biostatistics* 7.3 (1 lip. 2006), s. 355–373. ISSN: 1465-4644. DOI: 10.1093/biostatistics/kxj011. (Term. wiz. 01. 04. 2025).

[64] Kleinbaum, D. G. i Klein, M. *Survival Analysis: A Self-Learning Text*. Statistics for Biology and Health. Springer New York, 2012. ISBN: 978-1-4419-6645-2. DOI: 10.1007/978-1-4419-6646-9. (Term. wiz. 20. 10. 2024).

[65] *ISO 8402:1986, Quality — Vocabulary*. 1986.

[66] Lawless, J. F. *Statistical Models and Methods for Lifetime Data*. 1 wyd. Wiley Series in Probability and Statistics. Wiley, 13 list. 2002. ISBN: 978-0-471-37215-8. DOI: 10.1002/9781118033005. (Term. wiz. 22. 02. 2025).

[67] *International Electrotechnical Vocabulary (IEV) Chapter 191 - Dependability and Quality of Service*. 1990.

[68] Aalen, O. „Nonparametric Inference in Connection with Multiple Decrement Models". W: *Scandinavian Journal of Statistics* 3.1 (1976), s. 15–27. ISSN: 0303-6898. JSTOR: 4615603. (Term. wiz. 23. 02. 2025).

[69] Zacks, S. *Introduction to Reliability Analysis: Probability Models and Statistical Methods*. Springer Texts in Statistics. Springer New York, 1992. 212 s. ISBN: 978-1-4612-7697-5. DOI: 10.1007/978-1-4612-2854-7.

[70] Ba, L. i in. „Survival Analysis and Prognostic Factors of Palliative Radiotherapy in Patients with Metastatic Colorectal Cancer: A Propensity Score Analysis". W: *Journal of Gastrointestinal Oncology* 12.5 (paź. 2021), s. 2211–2222. ISSN: 2078-6891. DOI: 10.21037/jgo-21-540. PMID: 34790386. (Term. wiz. 24. 02. 2025).

[71] Buckley, J. i James, I. „Linear Regression with Censored Data". W: *Biometrika* 66.3 (1979), s. 429–436. ISSN: 0006-3444. DOI: 10.2307/2335161. JSTOR: 2335161. (Term. wiz. 21. 09. 2025).

[72] Li, J. i Ma, J. „On Hazard-Based Penalized Likelihood Estimation of Accelerated Failure Time Model with Partly Interval Censoring". W: *Statistical Methods in Medical Research* 29.12 (1 grud. 2020), s. 3804–3817. ISSN: 0962-2802. DOI: 10.1177/0962280220942555. (Term. wiz. 21. 09. 2025).

[73] Swindell, W. R. „Accelerated Failure Time Models Provide a Useful Statistical Framework for Aging Research". W: *Experimental gerontology* 44.3 (mar. 2009), s. 190–200. ISSN: 0531-5565. DOI: 10.1016/j.exger.2008.10.005. PMID: 19007875. (Term. wiz. 24. 02. 2025).

[74] Sutar, S. S. i Naik-Nimbalkar, U. V. „Accelerated Failure Time Models For Load Sharing Systems". W: *IEEE Transactions on Reliability* 63.3 (wrz. 2014), s. 706–714. ISSN: 0018-9529, 1558-1721. DOI: 10.1109/TR.2014.2313793. (Term. wiz. 24. 02. 2025).

[75] Bradburn, M. J. i in. „Survival Analysis Part II: Multivariate Data Analysis – an Introduction to Concepts and Methods". W: *British Journal of Cancer* 89.3 (sierp. 2003), s. 431–436. ISSN: 1532-1827. DOI: 10.1038/sj.bjc.6601119. (Term. wiz. 23. 02. 2025).

[76] Wang, P., Li, Y. i Reddy, C. K. „Machine Learning for Survival Analysis: A Survey". W: *ACM Computing Surveys (CSUR)* 51.6 (2019), s. 1–36.

[77] Friedman, J. H. „Greedy Function Approximation: A Gradient Boosting Machine." W: *The Annals of Statistics* 29.5 (paź. 2001), s. 1189–1232. ISSN: 0090-5364, 2168-8966. DOI: 10.1214/aos/1013203451. (Term. wiz. 25. 02. 2025).

[78] Chen, T. i Guestrin, C. „XGBoost: A Scalable Tree Boosting System". W: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. The 22nd ACM SIGKDD International Conference. Sierp. 2016, s. 785–794. DOI: 10.1145/2939672.2939785.

[79] Breiman, L. „Random Forests". W: *Machine Learning* 45.1 (1 paź. 2001), s. 5–32. ISSN: 1573-0565. DOI: 10.1023/A:1010933404324. (Term. wiz. 27. 07. 2025).

[80] Van Belle, V. i in. „Support Vector Machines for Survival Analysis". W: *Proceedings of the Third International Conference on Computational Intelligence in Medicine and Healthcare (Cimed2007)*. 2007, s. 1–8.

[81] Lee, C., Yoon, J. i Schaar, M. V. D. „Dynamic-DeepHit: A Deep Learning Approach for Dynamic Survival Analysis With Competing Risks Based on Longitudinal Data". W: *IEEE Transactions on Biomedical Engineering* 67.1 (sty. 2020), s. 122–133. ISSN: 0018-9294, 1558-2531. DOI: 10.1109/TBME.2019.2909027. (Term. wiz. 02. 08. 2025).

[82] Rudin, C. *Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead*. 22 wrz. 2019. DOI: 10.48550/arXiv.1811.10154. arXiv: 1811.10154 [stat]. (Term. wiz. 24. 02. 2025). Przedw.

[83] Susto, G. A. i in. „Machine Learning for Predictive Maintenance: A Multiple Classifier Approach". W: *IEEE Transactions on Industrial Informatics* 11.3

(czer. 2015), s. 812–820. ISSN: 1941-0050. DOI: 10.1109/TII.2014.2349359. (Term. wiz. 24. 02. 2025).

[84] Lipton, Z. C. *The Mythos of Model Interpretability*. 6 mar. 2017. DOI: 10.48550/arXiv.1606.03490. arXiv: 1606.03490 [cs]. (Term. wiz. 24. 02. 2025). Przedw.

[85] Miller, T. *Explanation in Artificial Intelligence: Insights from the Social Sciences*. 15 sierp. 2018. DOI: 10.48550/arXiv.1706.07269. arXiv: 1706.07269 [cs]. (Term. wiz. 24. 02. 2025). Przedw.

[86] Ribeiro, M. T., Singh, S. i Guestrin, C. *"Why Should I Trust You?": Explaining the Predictions of Any Classifier*. 9 sierp. 2016. DOI: 10.48550/arXiv.1602.04938. arXiv: 1602.04938 [cs]. (Term. wiz. 24. 02. 2025). Przedw.

[87] Doshi-Velez, F. i Kim, B. *Towards A Rigorous Science of Interpretable Machine Learning*. 2 mar. 2017. DOI: 10.48550/arXiv.1702.08608. arXiv: 1702.08608 [stat]. (Term. wiz. 24. 02. 2025). Przedw.

[88] Goodman, B. i Flaxman, S. „European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation"". W: *AI Magazine* 38.3 (wrz. 2017), s. 50–57. ISSN: 0738-4602, 2371-9621. DOI: 10.1609/aimag.v38i3.2741. arXiv: 1606.08813 [stat]. (Term. wiz. 25. 02. 2025).

[89] Cortes, C. i Vapnik, V. „Support-Vector Networks". W: *Machine Learning* 20.3 (1 wrz. 1995), s. 273–297. ISSN: 1573-0565. DOI: 10.1007/BF00994018. (Term. wiz. 27. 07. 2025).

[90] Molnar, C. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. Second edition. Christoph Molnar, 2022. 1 s. ISBN: 979-8-4114-6333-0.

[91] Breiman, L. „Statistical Modeling: The Two Cultures". W: *Statistical Science* 16.3 (sierp. 2001), s. 199–231. ISSN: 0883-4237, 2168-8745. DOI: 10.1214/ss/1009213726. (Term. wiz. 26. 02. 2025).

[92] Hastie, T., Tibshirani, R. i Friedman, J. H. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. Springer Series in Statistics. Springer, 2009. 745 s. ISBN: 978-0-387-84857-0 978-0-387-84858-7.

[93] Goldstein, A. i in. „Peeking Inside the Black Box: Visualizing Statistical Learning With Plots of Individual Conditional Expectation". W: *Journal of Computational and Graphical Statistics* 24.1 (2 sty. 2015), s. 44–65. ISSN: 1061-8600, 1537-2715. DOI: 10.1080/10618600.2014.907095. (Term. wiz. 25. 02. 2025).

[94] Louppe, G. i in. „Understanding Variable Importances in Forests of Randomized Trees". W: *Advances in Neural Information Processing Systems*. T. 26. Curran Associates, Inc., 2013. (Term. wiz. 25. 02. 2025).

[95] Apley, D. W. i Zhu, J. „Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models". W: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 82.4 (1 wrz. 2020), s. 1059–1086. ISSN: 1369-7412. DOI: 10.1111/rssb.12377. (Term. wiz. 25. 02. 2025).

[96] Zafar, M. R. i Khan, N. „Deterministic Local Interpretable Model-Agnostic Explanations for Stable Explainability". W: *Machine Learning and Knowledge Extraction* 3.3 (3 wrz. 2021), s. 525–541. ISSN: 2504-4990. DOI: 10.3390/make3030027. (Term. wiz. 26. 02. 2025).

[97] Lundberg, S. i Lee, S.-I. *A Unified Approach to Interpreting Model Predictions*. 25 list. 2017. DOI: 10.48550/arXiv.1705.07874. arXiv: 1705.07874 [cs]. (Term. wiz. 26. 02. 2025). Przedw.

[98] Wachter, S., Mittelstadt, B. i Russell, C. *Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR*. 21 mar. 2018. DOI: 10.48550/arXiv.1711.00399. arXiv: 1711.00399 [cs]. (Term. wiz. 26. 02. 2025). Przedw.

[99] Quinlan, J. R. „Induction of Decision Trees". W: *Machine Learning* 1.1 (1 mar. 1986), s. 81–106. ISSN: 1573-0565. DOI: 10.1007/BF00116251. (Term. wiz. 26. 02. 2025).

[100] Loh, W.-Y. „Classification and Regression Trees". W: *WIREs Data Mining and Knowledge Discovery* 1.1 (2011), s. 14–23. ISSN: 1942-4795. DOI: 10.1002/widm.8. (Term. wiz. 26. 02. 2025).

[101] Fürnkranz, J., Gamberger, D. i Lavrač, N. *Foundations of Rule Learning*. Springer Science & Business Media, 2012.

[102] Lakkaraju, H., Bach, S. H. i Leskovec, J. „Interpretable Decision Sets: A Joint Framework for Description and Prediction". W: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 13 sierp. 2016, s. 1675–1684. ISBN: 978-1-4503-4232-2. DOI: 10.1145/2939672.2939874. (Term. wiz. 26. 02. 2025).

[103] Holte, R. C. „Very Simple Classification Rules Perform Well on Most Commonly Used Datasets". W: *Machine Learning* 11.1 (1 kw. 1993), s. 63–90. ISSN: 1573-0565. DOI: 10.1023/A:1022631118932. (Term. wiz. 26. 02. 2025).

[104] Liu, H., Gegov, A. i Cocea, M. *Rule Based Systems for Big Data: A Machine Learning Approach*. T. 13. Studies in Big Data. Springer International Publishing, 2016. ISBN: 978-3-319-23695-7. DOI: 10.1007/978-3-319-23696-4. (Term. wiz. 26. 02. 2025).

[105] Montgomery, D. C. *Introduction to Statistical Quality Control*. Eighth edition. John Wiley & Sons, Inc., 2020. ISBN: 978-1-118-98915-9.

[106] Aha, D. W., Kibler, D. i Albert, M. K. „Instance-Based Learning Algorithms". W: *Machine Learning* 6.1 (1 sty. 1991), s. 37–66. ISSN: 1573-0565. DOI: 10.1007/BF00153759. (Term. wiz. 26. 02. 2025).

[107] Cover, T. i Hart, P. „Nearest Neighbor Pattern Classification". W: *IEEE Transactions on Information Theory* 13.1 (sty. 1967), s. 21–27. ISSN: 1557-9654. DOI: 10.1109/TIT.1967.1053964. (Term. wiz. 26. 02. 2025).

[108] Hart, P. „The Condensed Nearest Neighbor Rule". W: *IEEE Transactions on Information Theory* 14.3 (maj 1968), s. 515–516. ISSN: 1557-9654. DOI: 10.1109/TIT.1968.1054155. (Term. wiz. 26. 02. 2025).

[109] Duda, R. O., Stork, D. G. i Hart, P. E. *Pattern Classification*. 2nd ed. Wiley, 2001. 654 s. ISBN: 978-0-471-05669-0.

[110] Quinlan, J. R. *C4.5: Programs for Machine Learning*. The Morgan Kaufmann Series in Machine Learning. Morgan Kaufmann Publishers, 1993. 302 s. ISBN: 978-1-55860-238-0.

[111] Fürnkranz, J. i Flach, P. A. „ROC 'n' Rule Learning -Towards a Better Understanding of Covering Algorithms". W: *Machine Learning* 58.1 (sty. 2005), s. 39–77. ISSN: 0885-6125, 1573-0565. DOI: 10.1007/s10994-005-5011-x. (Term. wiz. 14. 03. 2025).

[112] Clark, P. i Niblett, T. „The CN2 Induction Algorithm". W: *Machine Learning* 3.4 (mar. 1989), s. 261–283. ISSN: 0885-6125, 1573-0565. DOI: 10.1007/BF00116835. (Term. wiz. 17. 03. 2025).

[113] Tyagi, K. i Sharma, A. „A Rule-Based Approach for Estimating the Reliability of Component-Based Systems". W: *Advances in Engineering Software* 54 (1 grud. 2012), s. 24–29. ISSN: 0965-9978. DOI: 10.1016/j.advengsoft.2012.08.001. (Term. wiz. 20. 03. 2025).

[114] Avritzer, A., Ros, J. i Weyuker, E. „Reliability Testing of Rule-Based Systems". W: *IEEE Software* 13.5 (wrz. 1996), s. 76–82. ISSN: 1937-4194. DOI: 10.1109/52.536461. (Term. wiz. 20. 03. 2025).

[115] Sikora, M. i in. „Application of Rule Induction to Discover Survival Factors of Patients after Bone Marrow Transplantation". W: *Journal of Medical Informatics & Technologies* Vol. 22 (2013). ISSN: 1642-6037. (Term. wiz. 20. 03. 2025).

[116] Wróbel, Ł. i Sikora, M. „Censoring Weighted Separate-and-Conquer Rule Induction from Survival Data". W: *Methods of Information in Medicine* 53.2 (2014), s. 137–148. ISSN: 2511-705X. DOI: 10.3414/ME13-01-0046. PMID: 24573170.

[117] Badura, J. i in. „Separate-and-Conquer Survival Action Rule Learning". W: *Knowledge-Based Systems* 280 (list. 2023), s. 110981. ISSN: 09507051. DOI: 10.1016/j.knosys.2023.110981. (Term. wiz. 22. 09. 2024).

[118] Ras, Z. W. i Tsay, L.-S. „Discovering Extended Action-Rules (System DEAR)". W: *Intelligent Information Processing and Web Mining*. Red. Kłopotek, M. A., Wierzchoń, S. T. i Trojanowski, K. Springer, 2003, s. 293–300. ISBN: 978-3-540-36562-4. DOI: 10.1007/978-3-540-36562-4_31.

[119] Sikora, M., Matyszok, P. i Wróbel, Ł. *SCARI: Separate and Conquer Algorithm for Action Rules and Recommendations Induction*. 9 czer. 2021. arXiv: 2106.05348 [cs]. (Term. wiz. 22. 09. 2024). Przedw.

[120] Hermansa, M. i in. „Recommendation Algorithm Based on Survival Action Rules". W: *Applied Sciences* 14.7 (30 mar. 2024), s. 2939. ISSN: 2076-3417. DOI: 10.3390/app14072939. (Term. wiz. 22. 09. 2024).

[121] Hussain, F. i in. „Exception Rule Mining with a Relative Interestingness Measure". W: *Knowledge Discovery and Data Mining. Current Issues and New Applications*. Red. Terano, T., Liu, H. i Chen, A. L. P. Springer, 2000, s. 86–97. ISBN: 978-3-540-45571-4. DOI: 10.1007/3-540-45571-X_11.

[122] Suzuki, E. „Discovering Unexpected Exceptions: A Stochastic Approach". W: *Proc. Fourth Int'l Workshop on Rough Sets, Fuzzy Sets, and Machine Discovery (RSFD)*. 1996, s. 225–232. (Term. wiz. 31. 03. 2025).

[123] Delgado, M., Ruiz, M. D. i Sánchez, D. „Mining Exception Rules". W: *Foundations of Reasoning under Uncertainty*. Red. Bouchon-Meunier, B. i in. Springer, 2010, s. 43–63. ISBN: 978-3-642-10728-3. DOI: 10.1007/978-3-642-10728-3_3. (Term. wiz. 31. 03. 2025).

[124] Suzuki, E. i Żytkow, J. M. „Unified Algorithm for Undirected Discovery of Exception Rules". W: *International Journal of Intelligent Systems* 20.7 (2005), s. 673–691. ISSN: 1098-111X. DOI: 10.1002/int.20090. (Term. wiz. 20. 03. 2025).

[125] Friedman, J. H. i Popescu, B. E. „Predictive Learning via Rule Ensembles". W: *The Annals of Applied Statistics* 2.3 (1 wrz. 2008). ISSN: 1932-6157. DOI: 10.1214/07-AOAS148. arXiv: 0811.1679 [stat]. (Term. wiz. 01. 04. 2025).

[126] Wróbel, Ł., Gudyś, A. i Sikora, M. „Learning Rule Sets from Survival Data". W: *BMC Bioinformatics* 18.1 (grud. 2017), s. 285. ISSN: 1471-2105. DOI: 10.1186/s12859-017-1693-x. (Term. wiz. 07. 12. 2024).

[127] Shakur, A. H. i in. „SURVFIT: Doubly Sparse Rule Learning for Survival Data". W: *Journal of Biomedical Informatics* 117 (1 maj. 2021), s. 103691. ISSN: 1532-0464. DOI: 10.1016/j.jbi.2021.103691. (Term. wiz. 01. 04. 2025).

[128] Brier, G. W. „Verification of Forecasts Expressed in Terms of Probability". W: *Monthly Weather Review* 78.1 (sty. 1950), s. 1–3. ISSN: 0027-0644, 1520-0493. DOI: 10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2. (Term. wiz. 03. 07. 2025).

[129] Graf, E. i in. „Assessment and Comparison of Prognostic Classification Schemes for Survival Data". W: *Statistics in Medicine* 18.17–18 (15 wrz. 1999–30), s. 2529–2545. ISSN: 0277-6715. DOI: 10.1002/(sici)1097-0258(19990915/30)18:17/18<2529::aid-sim274>3.0.co;2-5. PMID: 10474158.

[130] Fotso, S. *PySurvival: Open Source Package for Survival Analysis Modeling*. 2019.

[131] Kałwak, K. i in. „Higher CD34(+) and CD3(+) Cell Doses in the Graft Promote Long-Term Survival, and Have No Impact on the Incidence of Severe Acute or Chronic Graft-versus-Host Disease after in Vivo T Cell-Depleted Unrelated Donor Hematopoietic Stem Cell Transplantation in Children." W: *Biology of blood and marrow transplantation : journal of the American Society for Blood and Marrow Transplantation* 16.10 (paź. 2010), s. 1388–1401. DOI: 10.1016/j.bbmt.2010.04.001. PMID: 20382248.

[132] Loprinzi, C. L. i in. „Prospective Evaluation of Prognostic Variables from Patient-Completed Questionnaires. North Central Cancer Treatment Group." W: *Journal of Clinical Oncology* 12.3 (1994), s. 601–607.

[133] Pintilie, M. *Competing Risks: A Practical Perspective*. T. 58. John Wiley & Sons, 2006.

[134] Schumacher, M. i in. „Randomized 2 x 2 Trial Evaluating Hormonal Treatment and the Duration of Chemotherapy in Node-Positive Breast Cancer Patients. German Breast Cancer Study Group." W: *Journal of clinical oncology: official journal of the American Society of Clinical Oncology* 12.10 (1994), s. 2086.

[135] Lange, N. i in. *Case Studies in Biometry*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. Wiley, 1994. ISBN: 978-0-471-58925-9.

[136] Andersen, P. K. i in. *Statistical Models Based on Counting Processes*. Springer Science & Business Media, 2012.

[137] Kyle, R. A. „„Benign" Monoclonal Gammopathy—after 20 to 35 Years of Follow-Up". W: *Mayo Clinic Proceedings*. T. 68. Elsevier, 1993, s. 26–36.

[138] Fleming, T. R. i Harrington, D. P. *Counting Processes and Survival Analysis*. T. 169. John Wiley & Sons, 2011.

[139] Klein, J. P. i Moeschberger, M. L. *Survival Analysis: Techniques for Censored and Truncated Data*. Springer Science & Business Media, 2005.

[140] Rosenman, R. H. i in. „Clinically Unrecognized Myocardial Infarction in the Western Collaborative Group Study". W: *The American journal of cardiology* 19.6 (1967), s. 776–782.

[141] Abnet, C. C. i in. „Zinc Concentration in Esophageal Biopsy Specimens Measured by X-Ray Fluorescence and Esophageal Cancer Risk". W: *Journal of the National Cancer Institute* 97.4 (2005), s. 301–306.

[142] Battegay, M. i in. „Immunological Recovery and Antiretroviral Therapy in HIV-1 Infection". W: *The Lancet. Infectious Diseases* 6.5 (maj 2006), s. 280–287. ISSN: 1473-3099. DOI: 10.1016/S1473-3099(06)70463-7. PMID: 16631548.

[143] Schag, C. C., Heinrich, R. L. i Ganz, P. A. „Karnofsky Performance Status Revisited: Reliability, Validity, and Guidelines". W: *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 2.3 (mar. 1984), s. 187–193. ISSN: 0732-183X. DOI: 10.1200/JCO.1984.2.3.187. PMID: 6699671.

[144] Cardo, D. M. i in. „A Case-Control Study of HIV Seroconversion in Health Care Workers after Percutaneous Exposure. Centers for Disease Control and Prevention Needlestick Surveillance Group". W: *The New England Journal of Medicine* 337.21 (20 list. 1997), s. 1485–1490. ISSN: 0028-4793. DOI: 10.1056/NEJM199711203372101. PMID: 9366579.

[145] Tsai, C. C. i in. „Prevention of SIV Infection in Macaques by (R)-9-(2-Phosphonylmethoxypropyl)Adenine". W: *Science (New York, N.Y.)* 270.5239 (17 list. 1995), s. 1197–1199. ISSN: 0036-8075. DOI: 10.1126/science.270.5239.1197. PMID: 7502044.

[146] INSIGHT START Study Group i in. „Initiation of Antiretroviral Therapy in Early Asymptomatic HIV Infection". W: *The New England Journal of Medicine* 373.9 (27 sierp. 2015), s. 795–807. ISSN: 1533-4406. DOI: 10.1056/NEJMoa1506816. PMID: 26192873.

[147] TEMPRANO ANRS 12136 Study Group i in. „A Trial of Early Antiretrovirals and Isoniazid Preventive Therapy in Africa". W: *The New England Journal of Medicine* 373.9 (27 sierp. 2015), s. 808–822. ISSN: 1533-4406. DOI: 10.1056/NEJMoa1507198. PMID: 26193126.

[148] Hamaker, M. E. i in. „The Value of Geriatric Assessments in Predicting Treatment Tolerance and All-Cause Mortality in Older Patients With Cancer". W: *The Oncologist* 17.11 (list. 2012), s. 1439–1449. ISSN: 1083-7159. DOI: 10.1634/theoncologist.2012-0186. PMID: 22941970. (Term. wiz. 16. 09. 2025).

[149] Mohile, S. G. i in. „Practical Assessment and Management of Vulnerabilities in Older Patients Receiving Chemotherapy: ASCO Guideline for Geriatric Oncology". W: *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 36.22 (1 sierp. 2018), s. 2326–2347. ISSN: 1527-7755. DOI: 10.1200/JCO.2018.78.8687. PMID: 29782209.

[150] Wildiers, H. i in. „International Society of Geriatric Oncology Consensus on Geriatric Assessment in Older Patients with Cancer". W: *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 32.24 (20 sierp. 2014), s. 2595–2603. ISSN: 1527-7755. DOI: 10.1200/JCO.2013.54.8347. PMID: 25071125.

[151] Hurria, A. i in. „Validation of a Prediction Tool for Chemotherapy Toxicity in Older Adults With Cancer". W: *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 34.20 (10 lip. 2016), s. 2366–2371. ISSN: 1527-7755. DOI: 10.1200/JCO.2015.65.4327. PMID: 27185838.

[152] Extermann, M. i in. „Predicting the Risk of Chemotherapy Toxicity in Older Patients: The Chemotherapy Risk Assessment Scale for High-Age Patients (CRASH) Score". W: *Cancer* 118.13 (1 lip. 2012), s. 3377–3386. ISSN: 1097-0142. DOI: 10.1002/cncr.26646. PMID: 22072065.

[153] Gudyś, A., Sikora, M. i Wróbel, Ł. „Separate and Conquer Heuristic Allows Robust Mining of Contrast Sets in Classification, Regression, and Survival Data". W: *Expert Systems with Applications* 248 (15 sierp. 2024), s. 123376. ISSN: 0957-4174. DOI: 10.1016/j.eswa.2024.123376. (Term. wiz. 23. 09. 2025).

Spis rysunków

Spis tabel

Spis algorytmów