

# Metody wyznaczania zespołów klasyfikatorów z wykorzystaniem bireduktów decyzyjnych

(Ensembles of Classifiers Based on Decision Bireducts)

*Autoreferat rozprawy doktorskiej*

**Sebastian Stawicki**

Maj 2023

## 1 Wprowadzenie

Pojęcie bireduktów decyzyjnych jest kluczowym tematem niniejszej rozprawy. Jest to rozszerzenie pojęcia reduktów decyzyjnych z teorii zbiorów przybliżonych, wprowadzone w [23] i dalej rozwinięte w [17, 18, 19, 24, 6]. Biredukty decyzyjne zostały opracowane w celu zapewnienia prostego i interpretowalnego sposobu redukcji danych. Są one bardziej ogólne oraz pod wieloma względami bardziej elastyczne niż ich klasyczne odpowiedniki. Celem rozprawy jest zaprezentowanie, w spójny i przejrzysty sposób, kompletnego zestawienia wyników związanych z tym pojęciem.

Teoria zbiorów przybliżonych, która jest rozszerzeniem klasycznej teorii zbiorów, służy do modelowania pojęć w sytuacjach niepewności lub niekompletności informacji [11]. Jednym z podstawowych zagadnień tej teorii jest relacja nierozróżnialności – jest to relacja równoważności pozwalająca na grupowanie obiektów, których nie można od siebie odróżnić na podstawie określonego podzbioru cech. Relacja ta dzieli uniwersum obiektów na klasy abstrakcji, które mogą zostać użyte do opisu niedokładnych pojęć (rozumianych jako zbiór obiektów spełniających pewne pojęcie) za pomocą dolnych i górnych przybliżeń. W teorii zbiorów przybliżonych, pojęcie reduktów decyzyjnych jest uważane za jedno z podstawowych zagadnień. Jest ono zdefiniowane jako podzbiór atrybutów, który zachowuje charakterystykę relacji nierozróżnialności wyznaczonej przez pełny zbiór atrybutów [13]. Redukty decyzyjne mają szereg praktycznych zastosowań m.in. w zagadnieniach selekcji cech i odkrywaniu wiedzy [3].

Pojęcie reduktów decyzyjnych ma fundamentalne znaczenie w zastosowaniach teorii zbiorów przybliżonych. W praktyce często mamy jednak do czynienia z dużymi i zaszumionymi (rozumiejąc to sformułowanie jako dane, których fragmenty są bezużyteczne, zawierają niepoprawne wartości lub są uszkodzone) zbiorami danych. Aby sprostać takim wyzwaniom, zaproponowano kilka uogólnień dla reduktów decyzyjnych – wśród nich są m.in. pojęcia reduktów dynamicznych [1] oraz przybliżonych reduktów decyzyjnych [20].

Przybliżony redukt decyzyjny można zdefiniować jako nieredukowalny (w sensie zawierania się zbiorów) podzbiór atrybutów, który zgodnie z wybranym kryterium zachowuje określony stopień informacji decyzyjnej niesionej przez zbiór wszystkich atrybutów warunkowych. Przykładowo, może to być odzwierciedlenie szans, że klasyfikator skonstruowany przy użyciu wybranego podzbioru atrybutów, będzie poprawnie klasyfikował rozważane przez nas obiekty. Stosując takie podejście, można uzyskać podzbiory atrybutów, które są w kontrolowany sposób mniej dokładne niż standardowe redukty decyzyjne, ale mogą być preferowane w rzeczywistych zastosowaniach ze względu na to, że będą bardziej odporne na obecność szumu w danych oraz będą zawierały mniejszą liczbę atrybutów [12].

Przybliżone redukty decyzyjne wykorzystywane w zespołach klasyfikatorów mogą powtarzać błędne klasyfikacje na tych samych obiektach lub fragmentach danych [23]. Dzieje się tak, ponieważ wcze-

śniej omawiane funkcje oceniają podzbiory atrybutów za pomocą ogólnego podsumowania danych – bez skupiania się na poszczególnych obiektach. Jednym z rozwiązań tego problemu może być wykorzystanie przybliżonych reduktów decyzyjnych w połączeniu z innymi popularnymi technikami uczenia maszynowego, takimi jak boosting lub bagging [4].

Alternatywą do powyższego podejścia może być również wykorzystanie bireduktów decyzyjnych. Jest to rozszerzenie pojęcia reduktu decyzyjnego, koncentrujące się zarówno na podzbiorze atrybutów opisujących decyzję, jak i na podzbiorze obiektów, dla których ten opis jest prawidłowy. Możemy powiedzieć bardziej konkretnie, że biredukt decyzyjny jest reprezentowany jako para złożona z podzbiorem obiektów oraz podzbiorem atrybutów, gdzie podzbiór atrybutów możemy odnieść do podzbioru obiektów, dla których zapewnia on poprawną klasyfikację. Podzbiór obiektów dostarcza bardziej konkretnej i jednoznacznej informacji o odpowiadającym mu podzbiorze atrybutów i jego zdolności do skonstruowania dobrego klasyfikatora niż jakakolwiek funkcja oceny. Dodatkowo, informacja o obiektach, dla których decyzja jest wyrażona w sposób prawidłowy, pozwala zweryfikować, czy klasyfikatory wykorzystujące wybrane podzbiory atrybutów nie powtarzają błędów na tych samych fragmentach danych treningowych. Rozprawa ma na celu pokazanie analogii pomiędzy bireduktami decyzyjnymi a innymi metodami reprezentowania zależności w danych. Porównujemy biredukty decyzyjne ze standardowymi i przybliżonymi reduktami decyzyjnymi. Przedstawiamy także interpretacje stanowiące podstawę dla algorytmów, które w danych wyszukują biredukty decyzyjne o zadanych własnościach. Zwracamy szczególną uwagę na interpretację rozważanych typów reduktów i bireduktów decyzyjnych w kategoriach zbiorów reguł decyzyjnych, które są w stanie pominąć potencjalnie zaszumione fragmenty danych. Dzięki temu, bardziej prawdopodobne jest, że wynikowe klasyfikatory będą sobie dobrze radziły także przy klasyfikacji nowych danych.

W rozprawie skupiamy się na zespołach klasyfikatorów. W szczególności badamy sposoby i możliwości współdziałania bireduktów decyzyjnych, działających w ramach zespołu modeli, w rozwiązywaniu zadań decyzyjnych. W literaturze opisywanych jest wiele podejść tego typu m.in. w dziedzinach odkrywania wiedzy i klasyfikacji danych [4, 15]. Wykorzystując takie podejście, możemy liczyć na stabilność i odporność wynikowego zespołu, także w przypadkach, kiedy dane zawierają błędy lub są uszkodzone. Zwykle oczekuje się, że każdy pojedynczy model uczestniczący w zespole będzie prostszy niż zapewniający podobną jakość klasyfikacji, pojedynczy model, który nie został zbudowany w oparciu o metody zespołowe. W kategorii metod uczenia zespołów klasyfikatorów, istnieją pewne powszechnie uznane zasady i cele działania. Zgadza się na to, że każdy pojedynczy model może popełniać błędy, ale rozważając konkretny obiekt treningowy w ramach całego zespołu klasyfikatorów, chcemy, aby odpowiedź większości modeli była poprawna. Kolejnymi bardzo pożądanymi cechami są komplementarność i dywersyfikacja modeli składających się na zespół. W przypadku metod wykorzystujących teorię zbiorów przybliżonych odnosi się to do wykorzystania możliwie różnorodnych reduktów decyzyjnych. Jeśli chcielibyśmy rozszerzyć to podejście także na różnorodność obiektów, które są poprawnie/niepoprawnie klasyfikowane przez poszczególne modele składowe, możemy użyć do tego celu zespoły bireduktów decyzyjnych.

We współczesnych zastosowaniach uczenia maszynowego często definiowanym wymaganiem jest potrzeba dostarczenia interpretowalnych modeli decyzyjnych [5]. Jednym z aspektów interpretowalności jest wykorzystanie atrybutów (opisujących obiekty w danych), które są zrozumiałe dla ekspertów z danej dziedziny. Co więcej, także sposób wykorzystania atrybutów do przygotowania odpowiedzi modelu powinien być jasny i klarowny. W związku z tym zdolność modeli do określania istotności atrybutów przez nie wykorzystywanych jest istotnym aspektem działania i funkcjonalności dostarczanej przez model. W niektórych praktycznych zastosowaniach ranking atrybutów opracowany na podstawie ich istotności może być nawet bardziej przydatny niż same modele decyzyjne. Chcąc sprostać takim wymaganiom, również dla zespołów bireduktów decyzyjnych przygotowaliśmy metody oceny istotności atrybutów. Zaproponowaliśmy także procedurę służącą do oceny i porównywania istotności cech generowanych przez dowolne metody uczenia maszynowego.

Opracowane przez nas metody, zarówno w odniesieniu do klasyfikacji, jak i interpretowalności, poddaliśmy weryfikacji poprzez eksperymenty na danych syntetycznych i referencyjnych (benchmark data sets), a także na rzeczywistym problemie decyzyjnym odnoszącym się do branży HR.

## 2 Główne wyniki rozprawy

### 2.1 Biredukty decyzyjne

W rozprawie wprowadzamy pojęcie bireduktów decyzyjnych jako rozszerzenie klasycznych reduktów decyzyjnych z teorii zbiorów przybliżonych. Biredukty decyzyjne koncentrują się zarówno na podzbiórze atrybutów opisujących decyzję, jak i na podzbiórze obiektów, dla których ten opis jest prawidłowy. Bliskim odpowiednikiem tego pojęcia jest przybliżony redukt decyzyjny – zdefiniowany jako nieredukowalny podzbiór atrybutów, który zgodnie z wybranym kryterium zachowuje określony stopień informacji decyzyjnej niesionej przez zbiór wszystkich atrybutów warunkowych.

Różnica pomiędzy przybliżonymi reduktami decyzyjnymi a bireduktami decyzyjnymi polega na tym, w jaki sposób odnoszą się one do obiektów ze zbioru danych. W przypadku przybliżonych reduktów decyzyjnych odniesienie to jest pośrednie, określone poprzez kryterium użyte do ich konstrukcji. Z drugiej strony, biredukty decyzyjne używają odniesienia bezpośredniego, wskazując konkretne podzbiory obiektów.

Przyjmijmy standardową reprezentację danych tabelarycznych nazywaną tablicą decyzyjną [14, 13]. Tablica  $\mathbb{A} = (U, A \cup \{d\})$  jest parą niepustych zbiorów  $U$  oraz  $A \cup \{d\}$ , gdzie  $U$  nazywamy uniwersum obiektów, a  $A \cup \{d\}$  jest zbiorem atrybutów. Każdy atrybut  $a \in A \cup \{d\}$  jest funkcją  $a : U \rightarrow V_a$ , gdzie  $V_a$  jest nazywany zbiorem wartości atrybutu  $a$ . Wyróżniony atrybut  $d$ , taki że  $d \notin A$ , nazywany jest atrybutem decyzyjnym, a elementy  $A$  nazywane są atrybutami warunkowymi. Ponadto, niech  $IND(B) \subseteq U \times U$  (dla  $B \subseteq A$ ) będzie relacją nierozróżnialności (która jest relacją równoważności) wyrażającą, że dwa obiekty  $u, u' \in U$  są nierozróżnialne przez atrybuty z  $B$ :

$$u \text{ } IND(B) \text{ } u' \iff \forall_{a \in B} a(u) = a(u')$$

Rozważmy strategię indukcji reguł decyzyjnych, która aby wybrać wartość decyzji wykorzystuje głosowanie większościowe wśród obiektów znajdujących się każdej klasie abstrakcji (które są wyznaczone przez relację nierozróżnialności). Możemy także zdefiniować funkcję  $M : 2^A \rightarrow [0, 1]$  do obliczenia odsetka obiektów, które byłyby wybrane poprzez zastosowanie strategii większościowej:

$$M(B) = \frac{1}{|U|} \sum_{E \in U/IND(B)} \max_{k=1, \dots, |V_d|} |E \cap X^{(k)}|, \quad (1)$$

gdzie  $U/IND(B)$  jest zbiorem ilorazowym, a  $X^{(k)}$  oznacza klasę decyzyjną zawierającą obiekty z konkretną wartością decyzji  $v_{d_k} \in V_d$ , tzn.  $X^{(k)} = \{u \in U : d(u) = v_{d_k}\}$ . Widzimy, że  $M$  odpowiada strategii większościowej, ponieważ do sumy wybierane są moce zbiorów obiektów z najliczniej reprezentowaną decyzją w ramach każdej klasy abstrakcji.

	<i>Outlook</i>	<i>Temp.</i>	<i>Humid.</i>	<i>Wind</i>	<i>Play</i>
1	sunny	hot	high	weak	no
2	sunny	hot	high	strong	no
3	overcast	hot	high	weak	yes
4	rain	mild	high	weak	yes
5	rain	cool	normal	weak	yes
6	rain	cool	normal	strong	no
7	overcast	cool	normal	strong	yes
8	sunny	mild	high	weak	no
9	sunny	cool	normal	weak	yes
10	rain	mild	normal	weak	yes
11	sunny	mild	normal	strong	yes
12	overcast	mild	high	strong	yes
13	overcast	hot	normal	weak	yes
14	rain	mild	high	strong	no

Tablica 1: Przykładowa tablica decyzyjna. Widocznych jest 14 wierszy (lub obiektów), które reprezentują naszą bazę wiedzy. Obiekty opisywane są przez cztery atrybuty warunkowe – “Outlook” (prognoza pogody), “Temperature” (temperatura), “Humidity” (wilgotność), “Wind” (siła wiatru), oraz atrybut decyzyjny “Play”. W zależności od wartości atrybutów warunkowych, określona jest decyzja czy powinniśmy zagrać w golfa.

Dla  $\varepsilon \in [0, 1)$ , powiemy, że podzbiór  $B \subseteq A$  jest przybliżonym reduktem decyzyjnym dla miary  $M$  oraz przybliżenia  $\varepsilon$  (lub bardziej zwięźle, że jest  $(M, \varepsilon)$ -przybliżonym reduktem decyzyjnym) wtedy i tylko wtedy, gdy:

$$M(B) \geq (1 - \varepsilon)M(A),$$

oraz żaden z jego właściwych podzbiorów nie spełnia tego warunku ( $B$  jest nieredukowalny). Definicję można uogólnić do wykorzystania dowolnej, niemalejącej (w sensie zawierania się zbiorów) funkcji  $F : 2^A \rightarrow [0, 1]$ .

**Definicja 1.** Niech  $\mathbb{A} = (U, A \cup \{d\})$ ,  $B \subseteq A$  i  $\mathcal{X} \subseteq U$ . Mówimy, że “ $B$  wyznacza  $d$  wewnątrz  $\mathcal{X}$ ” (stosujemy oznaczenie  $B \Rightarrow_{\mathcal{X}} d$ ) wtedy i tylko wtedy, gdy  $B$  rozróżnia wszystkie pary  $u, u' \in \mathcal{X}$  takie, że  $d(u) \neq d(u')$ . Ponadto mówimy, że para  $(\mathcal{X}, B)$  jest bireduktem decyzyjnym wtedy i tylko wtedy, gdy zachodzą jednocześnie następujące warunki:

1. Zachodzi  $B \Rightarrow_{\mathcal{X}} d$ .
2. Nie istnieje właściwy podzbiór  $B' \subsetneq B$  taki, że  $B' \Rightarrow_{\mathcal{X}} d$ .
3. Nie istnieje właściwy nadzbiór  $\mathcal{X}' \supsetneq \mathcal{X}$  taki, że  $B \Rightarrow_{\mathcal{X}'} d$ .

Mówimy, że obiekty z  $\mathcal{X}$  są pokryte, a obiekty z  $U \setminus \mathcal{X}$  nie są pokryte przez biredukt decyzyjny  $(\mathcal{X}, B)$ .

Biredukt decyzyjny  $(\mathcal{X}, B)$  może być traktowany jako podstawa niedokładnej zależności funkcyjnej, łączącej podzbiór atrybutów  $B$  z decyzją  $d$  w stopniu  $\mathcal{X}$ -oznaczonej jako  $B \Rightarrow_{\mathcal{X}} d$  w Definicji 1. Każdy biredukt decyzyjny  $(\mathcal{X}, B)$  może być rozumiany jako para składająca się z nieredukowalnego podzbioru atrybutów, który można odnieść do nierozszerzalnego podzbioru obiektów, dla których zapewnia dobrą klasyfikację. Zbiór obiektów  $\mathcal{X}$  jest w rzeczywistości sumą zbiorów obiektów wspierających deterministyczne reguły decyzyjne wykorzystujące wartości atrybutów z  $B$  do opisanie wartości decyzji  $d$  [19]. Co więcej, obiekty w  $U \setminus \mathcal{X}$  mogą być traktowane jako wartości odstające (outliers) w odniesieniu do zależności funkcyjnej  $B \Rightarrow_{\mathcal{X}} d$ .

Można również sformułować pewne dodatkowe ograniczenie, na przykład, że dla każdego rozważanego bireduktu decyzyjnego zbiór niepokrytych obiektów nie może być zbyt duży. Taka forma ograniczenia jest do pewnego stopnia podobna do ograniczeń stosowanych w odniesieniu do zbiorów elementów częstych [10]. Można doszukać się także analogii do sformułowań podanych dla przybliżonych reduktów decyzyjnych.

**Definicja 2.** Niech  $\varepsilon \in [0, 1)$ . Powiemy, że para  $(\mathcal{X}, B)$ , gdzie  $\mathcal{X} \subseteq U$  i  $B \subseteq A$ , jest  $\varepsilon$ -bireduktem decyzyjnym, jeśli jest bireduktem decyzyjnym oraz zachodzi następująca nierówność:

$$|\mathcal{X}| \geq (1 - \varepsilon)|U|$$

Rozważmy zbiór danych z Tablicy 1 – mamy 9 obiektów z wartością decyzyjną ‘yes’ i 5 z wartością ‘no’. Szybko możemy stwierdzić, że dla  $\varepsilon \geq \frac{5}{14}$  będzie tylko jeden  $(M, \varepsilon)$ -przybliżony redukt decyzyjny, tj. pusty podzbiór atrybutów. Generując reguły decyzyjne z tego konkretnego  $(M, \varepsilon)$ -przybliżonego reduktu decyzyjnego, mielibyśmy tylko jedną możliwość, tzn. stworzyć regułę domyślną wskazującą na decyzję ‘yes’. Z perspektywy tworzenia zespołów klasyfikatorów, mniejsza wartość  $\varepsilon = \frac{4}{14}$ , dla której zbiór wszystkich przybliżonych reduktów decyzyjnych nie będzie już zbiorem jednoelementowym, będzie dużo bardziej interesująca. Lista przykładowych  $(M, \varepsilon)$ -przybliżonych reduktów i  $\varepsilon$ -bireduktów decyzyjnych dla  $\varepsilon = \frac{4}{14}$  jest przedstawiona w Tablicy 3. Możemy zauważyć, że liczba tych drugich jest znacznie większa. Ten wniosek będzie kluczowy przy omawianiu zespołów bireduktów decyzyjnych w Sekcji 2.2.

Tablica 2: Przykład wyboru wartości decyzji dla przyjętego podzbioru atrybutów warunkowych  $\{Outlook\}$  dla zbioru danych z Tablicy 1. (Po lewej) Wybór dokonany zgodnie ze strategią więkościową oraz funkcją  $M$  – dla każdej klasy abstrakcji wybierana jest najliczniej reprezentowana decyzja. (Po prawej) Wybór decyzji kontrolowany jest przez podzbiór obiektów pokrywanych przez biredukt decyzyjny  $(X, B) = (\{u_1, u_2, u_3, u_6, u_7, u_8, u_{12}, u_{13}, u_{14}\}, \{O\})$  – dla pierwszej i trzeciej klasy abstrakcji wybór decyzji jest taki sam jak dla funkcji  $M$ . Jednak w ramach drugiej klasy abstrakcji, dwa obiekty pokryte przez biredukt nie wskazują na decyzję więkościową. W przypadku bireduktów decyzyjnych wartość decyzji można wybrać dla każdej klasy abstrakcji oddzielnie.

	Outlook	Play
7	overcast	yes
12	overcast	yes
3	overcast	yes
13	overcast	yes
5	rain	yes
6	rain	no
4	rain	yes
10	rain	yes
14	rain	no
9	sunny	yes
1	sunny	no
2	sunny	no
8	sunny	no
11	sunny	yes

	Outlook	Play
7	overcast	yes
12	overcast	yes
3	overcast	yes
13	overcast	yes
5	rain	yes
6	rain	no
4	rain	yes
10	rain	yes
14	rain	no
9	sunny	yes
1	sunny	no
2	sunny	no
8	sunny	no
11	sunny	yes

Tablica 3: Przykłady  $(M, \varepsilon)$ -przybliżonych reduktów i  $\varepsilon$ -bireduktów decyzyjnych dla zbioru danych z Tablicy 1 dla  $\varepsilon = \frac{4}{14}$ . Każdy wiersz odpowiada podzbirowi atrybutów  $B \subseteq A$ . Wartości w drugiej kolumnie odpowiadają na pytanie, czy  $B$  jest  $(M, \varepsilon)$ -przybliżonym reduktem decyzyjnym. Trzecia kolumna zawiera podzbiory obiektów  $X$  takie, że  $(X, B)$  tworzą  $\varepsilon$ -biredukt decyzyjny. Dla większej czytelności, w przykładzie używana jest notacja  $[\dots]$ , która prezentuje jedynie indeksy obiektów ze zbioru danych, np.  $[[1, 4, 6, 8, 9, 11, 13]]$  odpowiada podzbirowi obiektów  $\{u_1, u_4, u_5, u_6, u_8, u_9, u_{11}, u_{12}, u_{13}\}$ .

$B \subseteq A$	czy $B$ jest $(M, \varepsilon)$ -przybliżonym reduktem decyzyjnym	podzbiory obiektów $X$ , dla których $(X, B)$ jest $\varepsilon$ -bireduktem decyzyjnym
$\{O\}$	tak	$[[1..5, 7, 8, 10, 12, 13]]$
$\{O, T\}$	nie	$[[1..5, 7..10, 12, 13], [1..5, 7, 9..13], [1..4, 6..10, 12, 13], [1..4, 6, 7, 9..13], [1..3, 5, 7..9, 12..14], [1..3, 5, 7, 9, 11..14], [1..3, 6..9, 12..14], [1..3, 6, 7, 9, 11..14]]$
$\{O, H\}$	nie	$[[1..5, 7..13], [1..4, 6..9, 11..13], [1..3, 5, 7..14], [1..3, 6..9, 11..14]]$
$\{O, W\}$	nie	$[[1..8, 10, 12..14], [1, 3..8, 10..14], [2..7, 9, 10, 12..14], [3..7, 9..14]]$
$\{T, W\}$	tak	$[[2..6, 9..13], [2..5, 7, 9..13]]$
$\{H, W\}$	nie	$[[2..5, 7, 9..11, 13, 14]]$
$\{O, T, W\}$	nie	$[[1..14]]$
$\{T, H, W\}$	nie	$[[1, 2, 4..6, 9..13], [1, 2, 4..6, 9..11, 13, 14], [1, 2, 4, 5, 7, 9..11, 13, 14], [1, 2, 5, 6, 8..13], [1, 2, 5, 6, 8..11, 13, 14], [1, 2, 5, 7..13], [2..6, 9..11, 13, 14], [2, 3, 5, 6, 8..13], [2, 3, 5, 6, 8..11, 13, 14], [2, 3, 5, 7..13], [2, 3, 5, 7..11, 13, 14]]$

Poniżej wymieniamy główne wyniki rozprawy (podzielone na kategorie) odnoszące się do pojęcia bireduktów decyzyjnych:

- Definicje i formalizacje: Zdefiniowaliśmy pojęcie bireduktu decyzyjnego oraz wprowadziliśmy jego warianty, np.  $\varepsilon$ -biredukty decyzyjne; biredukty  $\gamma$ -decyzyjne związane z pojęciem obszaru pozytywnego z teorii zbiorów przybliżonych.
- Charakterystyka i właściwości: W klasycznym podejściu prezentowanym w teorii zbiorów przybliżonych konstruowana jest “funkcja rozróżnialności” (formuła rachunku zdań) [16, 13]. Przedstawia ona warunki, które muszą być spełnione dla wszystkich par rozróżnialnych obiektów. W rezultacie, korzystając z takiej reprezentacji, zbiór reduktów decyzyjnych odpowiada zbiorowi wszystkich implikantów pierwszych funkcji rozróżnialności. W rozprawie prezentujemy analogiczne konstrukcje, które pozwalają wyrazić warunki związane z istnieniem bireduktów i  $\gamma$ -bireduktów decyzyjnych za pomocą odpowiedniej formuły rachunku zdań, wykorzystującej przy tym zmienne zdaniowe określone zarówno dla atrybutów (tak jak w klasycznym podejściu), jak i dla obiektów.
- Ocena i porównanie: Zaprezentowaliśmy analogie i porównaliśmy biredukty decyzyjne z ich klasycznymi odpowiednikami – standardowymi i przybliżonymi reduktami decyzyjnymi.
- Złożoność obliczeniowa: Wykazaliśmy NP-trudność problemu optymalizacyjnego polegającego na poszukiwaniu minimalnego  $\varepsilon$ -bireduktu decyzyjnego.
- Podejście algorytmiczne: Zaproponowaliśmy algorytmy i heurystyki, których celem jest efektywne wyszukiwanie bireduktów decyzyjnych oraz ich wariantów.
- Uogólnienia i rozszerzenia: Wprowadziliśmy specjalny rodzaj bireduktów decyzyjnych odpowiadający warunkom istniejącym w dziedzinie strumieni danych. Zaprezentowaliśmy nasze podejście, koncentrując się na sytuacjach, w których kompletny zestaw danych nie jest dostępny podczas procesu obliczeniowego. Zamiast tego przyjęliśmy scenariusz, w którym obiekty są przetwarzane przyrostowo. W praktycznych zastosowaniach może to odpowiadać sytuacji przetwarzania zdarzeń przychodzących w sposób ciągły, tzn. przychodzących do odpowiedniego systemu w momencie kiedy zaistnieją w świecie rzeczywistym.

## 2.2 Zespoły bireduktów decyzyjnych

Przykład przedstawiony w Tabeli 3 pokazuje nam, że  $\varepsilon$ -biredukty mogą być bardziej elastyczne w konstruowaniu zespołów klasyfikatorów niż przybliżone redukty decyzyjne. W szczególności, takie zespoły będą wykorzystywały mniejszą liczbę atrybutów, dając jednocześnie prostsze i bardziej ogólne modele. We wskazanym przykładzie, dla  $\varepsilon = \frac{4}{14}$ , sytuacja jest jeszcze bardziej intuicyjna. Jeśli skupimy się na zadaniu poszukiwania zespołu składającego się z trzech modeli, tak aby każdy obiekt ze zbioru danych był pokryty przez co najmniej dwa z nich (tj. każdy obiekt z danych treningowych będzie poprawnie klasyfikowany przez co najmniej dwa modele), wówczas niemożliwe jest uzyskanie takiego zespołu dla  $(M, \varepsilon)$ -przybliżonych reduktów decyzyjnych. Natomiast wykorzystując  $\varepsilon$ -biredukty decyzyjne, można to wykonać na 637 (wartość obliczona za pomocą skryptu) różnych sposobów. Kilka przykładów takich zespołów przedstawiono w Tabeli 4.

W powyższym przykładzie rozważane zespoły trzech  $\varepsilon$ -bireduktów pozwalają na prawidłową klasyfikację znanych obiektów oraz – ze względu na prostotę odpowiednich reguł decyzyjnych – mogą zapewnić dobrą jakość klasyfikacji również dla nowych obiektów. Poprawność klasyfikacji na danych treningowych jest gwarantowana przez fakt, że w prostej metodzie głosowania, dwa z trzech klasyfikatorów opartych na regułach decyzyjnych dają zawsze poprawną odpowiedź, a tylko jeden z nich może popełnić błąd (jeśli klasyfikowany obiekt nie jest akurat pokryty przez odpowiedni  $X \subseteq U$ ).

Pojęcie bireduktu decyzyjnego pozwala nam operować na podzbiorach atrybutów warunkowych traktowanych jako opisy informacji decyzyjnej dla powiązanych z nimi podzbiorów obiektów, dla których te opisy są poprawne. Pozwala nam to w prosty sposób zweryfikować komplementarność bireduktów

Tablica 4: Kilka przykładów trójelementowych zespołów modeli – każdy z użytych modeli to  $\varepsilon$ -biredukt decyzyjny uzyskany dla  $\varepsilon = \frac{4}{14}$ , cf. Tabela 3. Każdy obiekt ze zbioru danych jest pokrywany przez co najmniej dwa modele z zespołu. W przykładach prezentujemy także szczegółową charakterystykę pokrycia obiektów przez biredukty stanowiące zespół. Do obiektów odnosimy się poprzez ich indeks w ramach zbioru danych, tzn. “1” oznacza  $u_1$ , “2” oznacza  $u_2$ , itd.

zespoły modeli złożone z trzech $\varepsilon$ -bireduktów decyzyjnych	Pokrycie obiektów przez zespół modeli														
$(\llbracket 1..5, 7, 8, 10, 12, 13 \rrbracket, \{O\})$ $(\llbracket 1..3, 6..9, 11..14 \rrbracket, \{O, H\})$ $(\llbracket 3..7, 9..14 \rrbracket, \{O, W\})$	obiekt:	1	2	3	4	5	6	7	8	9	10	11	12	13	14
	pokrycie:	2	2	3	2	2	2	3	2	2	2	2	3	3	2
$(\llbracket 2..5, 7, 9..11, 13, 14 \rrbracket, \{H, W\})$ $(\llbracket 1..4, 6..9, 11..13 \rrbracket, \{O, H\})$ $(\llbracket 1..8, 10, 12..14 \rrbracket, \{O, W\})$	obiekt:	1	2	3	4	5	6	7	8	9	10	11	12	13	14
	pokrycie:	2	3	3	3	2	2	3	2	2	2	2	2	3	2
$(\llbracket 1..5, 7, 8, 10, 12, 13 \rrbracket, \{O\})$ $(\llbracket 1..14 \rrbracket, \{O, T, W\})$ $(\llbracket 2, 3, 5, 6, 8..11, 13, 14 \rrbracket, \{T, H, W\})$	obiekt:	1	2	3	4	5	6	7	8	9	10	11	12	13	14
	pokrycie:	2	3	3	2	3	2	2	3	2	3	2	2	3	2
$(\llbracket 1..3, 5, 7..9, 12..14 \rrbracket, \{O, T\})$ $(\llbracket 1, 3..8, 10..14 \rrbracket, \{O, W\})$ $(\llbracket 2..6, 9..13 \rrbracket, \{T, W\})$	obiekt:	1	2	3	4	5	6	7	8	9	10	11	12	13	14
	pokrycie:	2	2	3	2	3	2	2	2	2	2	2	3	3	2

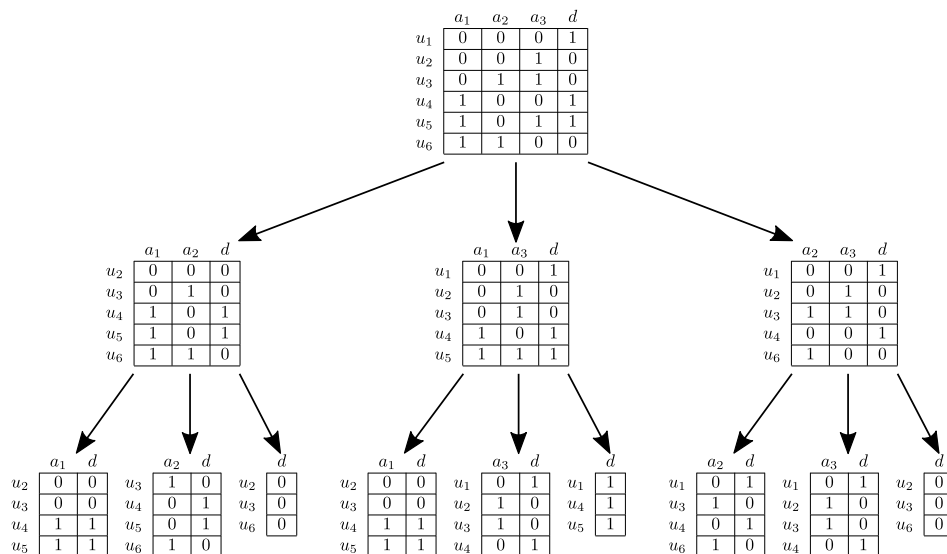
interpretowanych jako klasyfikatory współdziałające w zespole. Poniższe sformułowanie pozwala na wyrażenie idei głosowania większościowego między modelami użytymi w zespole, które – jeśli zostanie odpowiednio dostrojone do danych treningowych – daje nam szansę na dobrą jakość klasyfikacji, również na wcześniej niewidzianych obiektach.

**Definicja 3.** Niech  $\mathbb{A} = (U, A \cup \{d\})$  oraz zespół bireduktów decyzyjnych  $\mathcal{B} = \{(X_1, B_1), \dots, (X_m, B_m)\}$  będą dane. Powiemy, że  $\mathcal{B}$  jest “prawidłowy” wtedy i tylko wtedy, gdy

$$\forall_{u \in U} |\{i \in \{1, \dots, m\} : u \in X_i\}| > \frac{m}{2} \quad (2)$$

Powyższa nierówność (2) oznacza, że ponad 50% reguł decyzyjnych pasujących do obiektu  $u \in U$  wskazuje na poprawną decyzję  $d(u)$ . Rysunek 1 ilustruje rodzaj hierarchii “prawidłowych” zespołów bireduktów decyzyjnych dla  $m = 3$ .

Przykład zaprezentowany w ramach Rysunku 1 sugeruje, że metoda “od góry do dołu” może być bardzo pomocna w odpowiednim podejściu do procesu konstruowania “poprawnych” zespołów lub nawet ich hierarchii. Wszystkie  $m$  bireduktów, które zostaną włączone do zespołu mogą być konstruowane jednocześnie, zapewniając odpowiednią synchronizację, tak aby w całości pokryły zbiór danych w “poprawny” sposób. Proces ten może być kontynuowany aby uzyskane elementy zdekomponować na jeszcze mniejsze części. Podobne podejście było już wcześniej rozważane w [21] dla innego typu (bi)reduktów – uogólnionych reduktów decyzyjnych. Można także w odpowiedni sposób dostosować metody wykorzystujące permutacje [23], w taki sposób aby każdy kolejny krok obliczenia uwzględniał, które obiekty są w danym momencie w najmniejszym stopniu pokryte przez konstruowany zespół bireduktów.



Rysunek 1: Przykład hierarchii “prawidłowych” zespołów bireduktów decyzyjnych dla tablicy decyzyjnej  $\mathbb{A} = (\{u_1, \dots, u_6\}, \{a_1, a_2, a_3\} \cup \{d\})$ , złożonej z sześciu obiektów  $\{u_1, \dots, u_6\}$ , trzech atrybutów warunkowych  $\{a_1, a_2, a_3\}$  oraz atrybutu decyzyjnego  $d$ . Trzy biredukty w środkowej warstwie tworzą “prawidłowy” zespół dla  $\mathbb{A}$  (każdy obiekt jest poprawnie klasyfikowany przez co najmniej dwa z trzech bireduktów). Biredukty te wyznaczają odpowiednie podtablice decyzyjne, dla których cały proces może zostać powtórzony. W ostatniej warstwie widzimy trzy “prawidłowe” zespoły, każdy obliczony dla odpowiedniej podtablicy.

Poniżej wymieniamy główne wyniki rozprawy (podzielone na kategorie) odnoszące się do zespołów bireduktów decyzyjnych:

- Definicje i formalizacje: Przedstawiliśmy ideę zespołów bireduktów decyzyjnych i zbadaliśmy ich właściwości.
- Złożoność obliczeniowa: Wykazaliśmy NP-trudność problemu optymalizacji polegającego na poszukiwaniu najprostszego “prawidłowego” (cf. Definicja 3) zespołu bireduktów decyzyjnych. Pojęcie “prostoty” zespołu zostało zdefiniowane analogicznie do podejścia przedstawionego dla uogólnionych reduktów decyzyjnych [22] – tzn. “prostota” uwzględnia maksymalną moc podzbiorów atrybutów wykorzystywanych przez modele składające się na zespół.
- Ocena i porównanie: Zaproponowane metody zostały zweryfikowane eksperymentalnie na zestawach danych syntetycznych i referencyjnych.
- Zastosowania praktyczne: Kompleksowo zaprezentowaliśmy studium przypadku demonstrujące zastosowanie zespołów bireduktów do problemu decyzyjnego, który został zdefiniowany podczas realizacji projektu dla firmy HR specjalizującej się w rekrutacji ekspertów IT.
- Interpretowalność: Przygotowaliśmy metody umożliwiające ocenę istotności atrybutów użytych w zespołach klasyfikatorów składających się z bireduktów decyzyjnych.
- Implementacja: Dla języka Python przygotowaliśmy otwartoźródłową bibliotekę programistyczną *scikit-rough* (<https://github.com/sebov/scikit-rough>) udostępnioną w ramach platformy GitHub.

### 2.3 Procedura służąca do oceny i porównywania wyników istotności cech

W celu zwiększenia praktycznej wartości opracowanych metod, zaproponowaliśmy ogólny sposób oceny i porównywania wyników istotności cech, niezależny od użytej metody uczenia maszynowego, która była



wykorzystana do ich przygotowania. Dzięki temu mogliśmy porównać wyniki istotności atrybutów uzyskiwane dla zespołów bireduktów decyzyjnych do innych metod stosowanych w praktyce.

Procedurę rozpoczynamy od skonstruowania rozszerzonej tablicy decyzyjnej, dla której chcemy przeprowadzić porównanie wybranych metod zwracających istotności cech. Dla oryginalnej reprezentacji danych  $\mathbb{A} = (U, A \cup \{d\})$ , konstruujemy wariant tabeli  $\mathbb{A}^\circ = (U, A \cup A^\circ \cup \{d\})$ , gdzie  $A^\circ$  jest zbiorem “przetasowanych” atrybutów, zdefiniowanych w następujący sposób:

$$a_i^\circ : U \rightarrow V_{a_i}, \quad a_i^\circ(u_j) = a_i(u_{\delta_i(j)}),$$

gdzie  $\delta_i$  odpowiada losowej permutacji uniwersum obiektów  $U$ . Atrybuty “przetasowane”, stworzone jako losowe permutacje oryginalnych wartości, zachowują oryginalny rozkład. Takie podejście do randomizacji jest dość popularne w uczeniu maszynowym [2, 8, 7].

Przyjmujemy, że procedura porównywania wyników istotności cech powinna odpowiednio uwzględnić następujące dwa postulaty:

- Dobra metoda oceny istotności cech powinna pozwolić na rozróżnienie atrybutów oryginalnych od “przetasowanych”.
- Nie wszystkie oryginalne atrybuty muszą być uznawane przez metodę oceny istotności cech za dobre, np. niektóre z nich mogą być nieistotne lub zaszumione.

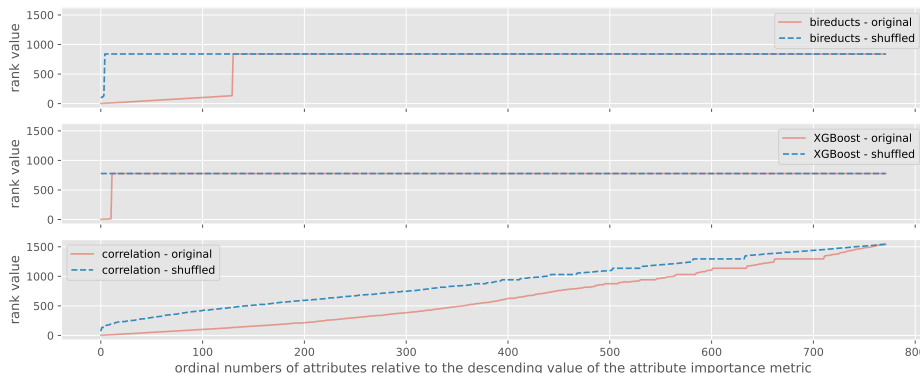
Metoda oceny istotności cech przypisuje atrybutom pewne wagi (wartości numeryczne), które wyrażają ocenę jak ważny jest dany atrybut w odniesieniu do konkretnej metody – przyjmujemy, że im wyższa ocena, tym atrybut jest uznawany za bardziej istotny. W dalszej kolejności, procedura dokonuje zbiorczego przypisania rang do atrybutów (oryginalnych i “przetasowanych”) na podstawie wyników istotności cech dostarczonych przez badaną metodę. Przyjmujemy, że remisy w ocenach atrybutów są obsługiwane w prosty sposób, poprzez przypisanie tej samej rangi atrybutom z tym samym wynikiem. Przyglądamy się teraz oddzielnie rangom atrybutów oryginalnych i “przetasowanych”, wybierając spośród nich tylko  $top\_k$  najlepszych (tzn. najniższych) wartości. Wartość  $top\_k$  powinna odzwierciedlać przybliżoną liczbę prawdziwie istotnych atrybutów w ramach rozważanego problemu decyzyjnego – niestety, w praktyce często taka informacja nie jest dostępna. Rangi wybranych atrybutów z  $A$  i  $A^\circ$  są następnie uśredniane, dając końcowy wynik złożony z dwóch liczb - średnich rangi  $top\_k$  oryginalnych i “przetasowanych” atrybutów. Takie dane wyjściowe można zinterpretować w następujący sposób:

- Im pierwsza liczba jest mniejsza, tym mniejsza jest liczba atrybutów “przetasowanych”, które występują przed  $top\_k$  oryginalnych.
- Im druga liczba jest większa, tym mniejsza jest liczba oryginalnych atrybutów, które występują po  $top\_k$  “przetasowanych”.
- Im większa jest różnica pomiędzy tymi dwiema liczbami, tym lepsza jest dana metoda istotności cech w rozróżnianiu  $A$  i  $A^\circ$ .

Rysunek 2 przedstawia wyniki zaproponowanej procedury dla wybranych metod oceny istotności cech, które (poza metodą referencyjną) były również wykorzystane w studium przypadku związanym z branżą HR, kompleksowo przedstawionym w rozprawie.

W eksperymentach bierzemy pod uwagę zespoły bireduktów decyzyjnych oraz XGBoost. Jako punkt odniesienia sprawdzamy również prostą metodę oceny istotności cech obliczającą bezwzględne wartości współczynników korelacji Spearmana między każdym atrybutem warunkowym a atrybutem decyzyjnym. Ustawienia eksperymentów dobieraliśmy tak, aby warunki pracy metod wykorzystujących bireduktów oraz XGBoost były porównywalne. W szczególności, ustaliliśmy liczbę 1,000 drzew decyzyjnych (odpowiadającą liczbie iteracji algorytmu XGBoost) oraz 1,000 bireduktów decyzyjnych tworzących zespół. Ponadto, zrównaliśmy szanse obu metod w odniesieniu do liczby atrybutów

średnia ranga top_k najlepszych atrybutów							
top_k	biredukty decyzyjne		XGBoost		korelacja		
	oryginalne	przetasowane	oryginalne	przetasowane	oryginalne	przetasowane	
10	5.50	548.70	5.50	778.00	5.50	149.50	
20	10.50	694.10	353.40	778.00	10.50	181.45	
30	15.50	742.57	494.93	778.00	15.50	201.17	
50	25.50	781.34	608.16	778.00	25.52	230.94	
100	50.55	810.42	693.08	778.00	50.80	296.93	
all	709.27	835.73	767.00	778.00	642.58	902.42	



Rysunek 2: Średnie rangi istotności dla atrybutów oryginalnych ( $A$ ) i “przetasowanych” ( $A^\odot$ ) obliczone dla porównywanych metod oceny istotności cech – tutaj: zespoły bireduktów decyzyjnych, XGBoost oraz metoda wykorzystująca wartość współczynnika korelacji (jako referencyjna) – na zbiorze danych właściwym dla prezentowanego w rozprawie studium przypadku z dziedziny HR. Wykresy pod tabelą przedstawiają rzeczywiste rangi atrybutów (oryginalnych i “przetasowanych”) wyświetlone w porządku rosnącym.

używanych w pojedynczych modelach składających się na zespoły klasyfikatorów – ustawiliśmy maksymalną głębokość drzewa w XGBoost na 2 oraz maksymalną liczbę atrybutów w każdym biredukcie decyzyjnym na 3. Wynika to z faktu, że binarne drzewa decyzyjne o głębokości 2 zwykle wykorzystują wewnętrznie trzy atrybuty. Dodatkowo, zastosowaliśmy miary oceny istotności atrybutów, które można interpretować podobnie zarówno dla XGBoost, jak i dla zespołów bireduktów decyzyjnych.

Wykorzystując proponowaną procedurę jako narzędzie do profilowania metod oceny istotności atrybutów, możemy wyciągnąć następujące wnioski (bez zagłębiania się w specyfikę problemu decyzyjnego lub samych danych):

1. Metody wykorzystujące biredukty oraz korelację są równie dobre w wyborze dużej liczby oryginalnych atrybutów i nie koncentrują się na stosunkowo niewielkim podzbiore najlepszych atrybutów. W niektórych zastosowaniach może być wskazane wykorzystanie większego i bardziej zróżnicowanego podzbioru atrybutów, nawet mimo tego, że nie wszystkie atrybuty będą najwyższej jakości.
2. Algorytm XGBoost ma tendencję do wykorzystywania tylko niewielkiej liczby najbardziej informatywnych atrybutów oryginalnych, a zatem nie jest w stanie odróżnić mniej istotnych atrybutów oryginalnych od tych “przetasowanych”.
3. Metoda wykorzystująca biredukty decyzyjne pozwalają na znaczenie lepsze odróżnianie oryginalnych atrybutów od “przetasowanych” niż metody oparte o wartości korelacji Spearmana lub oceny istotności obliczone przy użyciu XGBoost.
4. Największa różnica między średnimi rangami  $top_k$  atrybutów (oryginalnych i “przetasowanych”), dla prawie wszystkich badanych wartości  $top_k$ , została osiągnięta przy metodzie wykorzystującej biredukty. Jedynym wyjątkiem była najmniejsza wartość  $top_k$  – w tym przypadku najwyższą różnicę odnotowano dla XGBoost, jednak wynik dla bireduktów był tylko nieznacznie gorszy.

### 3 Podsumowanie i dalsze kierunki badań

W rozprawie doktorskiej przedstawiliśmy wyzwania związane z wyszukiwaniem bireduktów decyzyjnych w zbiorach danych i omówiliśmy kilka skutecznych metod ich obliczania. Nakreśliśmy kilka intuicji dotyczących możliwych praktycznych zastosowań bireduktów decyzyjnych. Warto zauważyć, że jedną z największych zalet pojęcia bireduktów decyzyjnych jest jego prosty, elastyczny i łatwy do interpretacji sposób reprezentacji wiedzy. Dodatkowo, chociaż pojęcie bireduktów decyzyjnych różni się koncepcyjnie od przybliżonych reduktów decyzyjnych, nasze badania pokazały obecność pewnych analogii i związków między nimi.

Pokazaliśmy, w jaki sposób biredukty decyzyjne mogą być wykorzystywane do konstruowania zespołów klasyfikatorów. Wprowadziliśmy pojęcie zespołu “prawidłowego”, które oznacza, że każdy obiekt treningowy musi być poprawnie klasyfikowany (wykorzystując odpowiednie reguły decyzyjne) przez więcej niż połowę klasyfikatorów. Omówiliśmy, w jaki sposób określić kryterium “prostoty” dla takich zespołów i sformułowaliśmy przykładowy problem optymalizacyjny związany z wyodrębnianiem możliwie najprostszych prawidłowych zespołów bireduktów decyzyjnych z danych wejściowych oraz pokazaliśmy, że problem ten należy do kategorii problemów NP-trudnych.

Zwróciliśmy także uwagę na interpretowalność opracowanych modeli. Zadanie polegające na ocenie istotności cech może być postrzegane nie tylko jako efekt uboczny procesu budowania wydajnych modeli, ale także jako odrębny, dobrze zdefiniowany problem. Zaproponowaliśmy kilka metryk służących ocenie istotności atrybutów wykorzystywanych przez zespoły bireduktów decyzyjnych. Przedstawiliśmy również ogólną procedurę służącą do oceny i porównywania różnych metod określania ważności cech, niezależną od użytego algorytmu uczenia maszynowego. Ponadto wykorzystaliśmy tę procedurę do porównania wyników istotności cech dostarczonych – przez zespoły bireduktów decyzyjnych, XGBoost oraz metodę wykorzystującą korelację – na syntetycznych, referencyjnych i rzeczywistych zbiorach danych.

#### 3.1 Dalsze kierunki badań

W przyszłości planujemy kontynuować badania nad innymi rodzajami bireduktów. Chcielibyśmy przeprowadzić eksperymenty obejmujące różne typy bireduktów i różne strategie konstruowania zespołów klasyfikatorów, a także przeprowadzić testy w szerszym zakresie na zbiorach danych referencyjnych.

Chcielibyśmy dalej badać właściwości bireduktów, aby lepiej wykorzystać ich zalety do opracowania w celu opracowania bardziej intuicyjnych sposobów wizualizacji i interaktywnej eksploracji złożonych danych. Pragniemy rozwinąć nasze dotychczasowe metody algorytmiczne [23, 19] do wyszukiwania i stosowania bireduktów decyzyjnych również dla innych przykładów danych rzeczywistych.

W odniesieniu do złożoności obliczeniowej, korzystne byłoby wykorzystanie aparatu matematycznego opracowanego w [9], w celu wzmocnienia otrzymanych wyników w kierunku twierdzeń o nieaprosymowalności.

Procedura służąca porównywaniu metod oceny istotności atrybutów wymaga dalszych badań. Istnieją alternatywne podejścia, które mogłyby zostać wykorzystane. Przykładowo, możemy rozważyć problem decyzyjny polegający na klasyfikacji typu atrybutu – czy jest to atrybut oryginalny, czy “przetasowany”. Dla tak zdefiniowanego zadania można wykorzystać metody z dziedziny wyszukiwania informacji (information retrieval), np. krzywą ROC (receiver operating characteristic) lub AUC (area under the ROC curve). Możemy również zastosować bardziej zaawansowane podejścia statystyczne.

Jesteśmy zainteresowani także rozwinięciem metod budowy hierarchicznych modeli wykorzystujących zespoły bireduktów decyzyjnych (zob. Rysunek 1). Konieczne jest opracowanie wydajnych algorytmów ułatwiających proces konstruowania takich modeli. Co więcej, należy przyjrzeć się, jak różne strategie głosowania stosowane w takich hierarchicznych podejściach będą wpływały na osiągnięte wyniki klasyfikacji.

## Literatura

- [1] Jan G. Bazan, Andrzej Skowron, and Piotr Synak. Dynamic Reducts as a Tool for Extracting Laws from Decisions Tables. In Zbigniew W. Raś and Maria Zemankova, editors, *Methodologies for Intelligent Systems, 8th International Symposium, ISMIS '94, Charlotte, North Carolina, USA, October 16-19, 1994, Proceedings*, volume 869 of *Lecture Notes in Computer Science*, pages 346–355. Springer, 1994.
- [2] Leo Breiman. Random Forests. *Machine Learning*, 45(1):5–32, 2001.
- [3] Igor Chikalov, Vadim V. Lozin, Irina Lozina, Mikhail J. Moshkov, Hung Son Nguyen, Andrzej Skowron, and Beata Zielosko. *Three Approaches to Data Analysis – Test Theory, Rough Sets and Logical Analysis of Data*, volume 41 of *Intelligent Systems Reference Library*. Springer, 2013.
- [4] Thomas G. Dietterich. An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization. *Machine Learning*, 40(2):139–157, 2000.
- [5] Kevin Fauvel, Élixa Fromont, Véronique Masson, Philippe Faverdin, and Alexandre Termier. XEM: An Explainable-by-Design Ensemble Method for Multivariate Time Series Classification. *Data Mining and Knowledge Discovery*, 36(3):917–957, 2022.
- [6] Andrzej Janusz, Dominik Ślęzak, Sebastian Stawicki, and Krzysztof Stencel. A Practical Study of Methods for Deriving Insightful Attribute Importance Rankings Using Decision Bireducts. *Information Sciences*, 2023.
- [7] Miron B. Kursa, Aleksander Jankowski, and Witold R. Rudnicki. Boruta - A System for Feature Selection. *Fundamenta Informaticae*, 101(4):271–285, 2010.
- [8] Miron B. Kursa and Witold R. Rudnicki. Feature Selection with the Boruta Package. *Journal of Statistical Software*, 36(11):1–13, 2010.
- [9] Mikhail J. Moshkov, Marcin Piliszczuk, and Beata Zielosko. *Partial Covers, Reducts and Decision Rules in Rough Sets – Theory and Applications*, volume 145 of *Studies in Computational Intelligence*. Springer, 2008.
- [10] Sinh Hoa Nguyen and Hung Son Nguyen. Pattern Extraction from Data. *Fundamenta Informaticae*, 34(1-2):129–144, 1998.
- [11] Zdzisław Pawlak. *Rough Sets: Theoretical Aspects of Reasoning about Data*, volume 9 of *Theory and decision library : series D*. Kluwer, 1991.
- [12] Zdzisław Pawlak. Rough Set Elements. In Lech Polkowski and Andrzej Skowron, editors, *Rough Sets in Knowledge Discovery 1 – Methodology and Applications*, volume 18 of *Studies in Fuzziness and Soft Computing*, pages 10–30. Physica-Verlag, 1998.
- [13] Zdzisław Pawlak and Andrzej Skowron. Rough Sets and Boolean Reasoning. *Information Sciences*, 177(1):41–73, 2007.
- [14] Zdzisław Pawlak and Andrzej Skowron. Rudiments of Rough Sets. *Information Sciences*, 177(1):3–27, 2007.
- [15] Robi Polikar, Joseph DePasquale, Hussein Syed Mohammed, Gavin Brown, and Ludmilla I. Kuncheva. Learn++.MF: A Random Subspace Approach for the Missing Feature Problem. *Pattern Recognition*, 43(11):3817–3832, 2010.
- [16] Andrzej Skowron and Cecylia Rauszer. *The Discernibility Matrices and Functions in Information Systems*, pages 331–362. Springer Netherlands, Dordrecht, 1992.

- [17] Sebastian Stawicki and Sebastian Widz. Decision Bireducts and Approximate Decision Reducts: Comparison of Two Approaches to Attribute Subset Ensemble Construction. In Maria Ganzha, Leszek A. Maciaszek, and Marcin Paprzycki, editors, *Federated Conference on Computer Science and Information Systems - FedCSIS 2012, Wroclaw, Poland, 9-12 September 2012, Proceedings*, pages 331–338, 2012.
- [18] Sebastian Stawicki and Dominik Ślęzak. Recent Advances in Decision Bireducts: Complexity, Heuristics and Streams. In Pawan Lingras, Marcin Wolski, Chris Cornelis, Sushmita Mitra, and Piotr Wasilewski, editors, *Rough Sets and Knowledge Technology - 8th International Conference, RSKT 2013, Halifax, NS, Canada, October 11-14, 2013, Proceedings*, volume 8171 of *Lecture Notes in Computer Science*, pages 200–212. Springer, 2013.
- [19] Sebastian Stawicki, Dominik Ślęzak, Andrzej Janusz, and Sebastian Widz. Decision Bireducts and Decision Reducts – A Comparison. *International Journal of Approximate Reasoning*, 84:75–109, 2017.
- [20] Dominik Ślęzak. Approximate Reducts in Decision Tables. In *Proceedings of IPMU 1996*, volume 3, pages 1159–1164, 1996.
- [21] Dominik Ślęzak. Decomposition and Synthesis of Decision Tables with Respect to Generalized Decision Functions. In Sankar K. Pal and Andrzej Skowron, editors, *Rough Fuzzy Hybridization – A New Trend in Decision Making*, pages 110–135. Springer, 1999.
- [22] Dominik Ślęzak. On Generalized Decision Functions: Reducts, Networks and Ensembles. In Yiyu Yao, Qinghua Hu, Hong Yu, and Jerzy W. Grzymala-Busse, editors, *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing - 15th International Conference, RSFDGrC 2015, Tianjin, China, November 20-23, 2015, Proceedings*, volume 9437 of *Lecture Notes in Computer Science*, pages 13–23. Springer, 2015.
- [23] Dominik Ślęzak and Andrzej Janusz. Ensembles of Bireducts: Towards Robust Classification and Simple Representation. In Tai-Hoon Kim, Hojjat Adeli, Dominik Ślęzak, Frode Eika Sandnes, Xiaofeng Song, Kyo-Il Chung, and Kirk P. Arnett, editors, *Future Generation Information Technology - Third International Conference, FGIT 2011 in Conjunction with GDC 2011, Jeju Island, Korea, December 8-10, 2011. Proceedings*, volume 7105 of *Lecture Notes in Computer Science*, pages 64–77. Springer, 2011.
- [24] Dominik Ślęzak and Sebastian Stawicki. The Problem of Finding the Simplest Classifier Ensemble is NP-Hard - A Rough-Set-Inspired Formulation Based on Decision Bireducts. In Rafael Bello, Duoqian Miao, Rafael Falcon, Michinori Nakata, Alejandro Rosete, and Davide Ciucci, editors, *Rough Sets - International Joint Conference, IJCRS 2020, Havana, Cuba, June 29 - July 3, 2020, Proceedings*, volume 12179 of *Lecture Notes in Computer Science*, pages 204–212. Springer, 2020.