

ANALIZA ZBIORU DANYCH DOTYCZĄCYCH OCENY AKCEPTOWALNOŚCI SAMOCHODÓW ZA POMOCĄ TEORII ZBIORÓW PRZYBLIŻONYCH I PROGRAMU RSES

Mateusz Gorączka

Uniwersytet Kazimierza Wielkiego
Wydział Informatyki
ul. Mikołaja Kopernika 1, 85-074 Bydgoszcz
e-mail: mateusz.goraczka@student.ukw.edu.pl

Streszczenie: Artykuł skupia się na analizie danych z wykorzystaniem teorii zbiorów przybliżonych oraz różnych metod, takich jak algorytm genetyczny, klasyfikacja za pomocą zestawu reguł i metoda walidacji krzyżowej. Przedstawiono także kompletny proces analizy danych przy użyciu programu RSES. Wykorzystany zbiór danych oraz wyniki analizy zostały omówione w kontekście teorii zbiorów przybliżonych. Artykuł kończy się podsumowaniem i wnioskami skupiającymi się na aspekcie skuteczności wspomnianych metod w analizie zbioru danych oraz efektywności programu w kwestii przeprowadzania w nim analiz.

Słowa kluczowe: Teoria zbiorów przybliżonych, zbiór danych, analiza danych, klasyfikacja danych, RSES

Analysis of the car acceptability assessment dataset using rough set theory and the RSES program

Abstrakt: The article focuses on data analysis using rough set theory and various methods such as the genetic algorithm, rule set classification and the cross-validation method. The complete data analysis process using RSES is also presented. The dataset used and the results of the analysis are discussed in the context of rough set theory. The article concludes with a summary and conclusions focusing on the aspect of the effectiveness of aforementioned methods in analysing the dataset and the efficiency of the program in terms of performing analysis in it.

Keywords: Rough set theory, dataset, data analysis data classification, RSES

1. WSTĘP

W dzisiejszym dynamicznym środowisku rynkowym ocena akceptowalności produktów, w tym samochodów, jest kluczowym wyzwaniem. Aby skutecznie wspierać proces podejmowania decyzji, badacze coraz częściej sięgają po zaawansowane metody analizy danych. W kontekście oceny akceptowalności samochodów, teoria zbiorów przybliżonych stanowi potężne narzędzie, pozwalając na skomplikowaną analizę danych z uwzględnieniem niejednoznaczności i niepewności.

Celem niniejszego artykułu jest przedstawienie analizy zbioru danych dotyczącego oceny akceptowalności samochodów z wykorzystaniem teorii zbiorów przybliżonych oraz programu RSES (Rough Set Exploration System). Teoria zbiorów przybliżonych, wprowadzona

przez Zdzisława Pawlaka, stanowi potężne narzędzie do modelowania i analizy niepewności w danych. Program RSES natomiast umożliwia efektywną implementację tej teorii, wspierając badaczy w ekstrakcji istotnych informacji z danych oraz identyfikacji reguł decyzyjnych.

Artykuł skupi się na opisie procesu analizy danych związanych z oceną akceptowalności samochodów, począwszy od zebrania zbioru danych aż do uzyskania wniosków na podstawie zastosowanej metodyki. Przedstawione zostaną kroki podejmowane w trakcie analizy, w tym przygotowanie danych, identyfikacja cech istotnych dla oceny akceptowalności oraz konstrukcja modelu opartego na teorii zbiorów przybliżonych.

W efekcie, artykuł ma na celu nie tylko zaprezentowanie zastosowania teorii zbiorów przybliżonych i programu RSES w kontekście oceny akceptowalności samochodów, ale także udzielenie czytelnikowi praktycznych wskazówek

dotyczących analizy danych z wykorzystaniem tych narzędzi. Prezentowane podejście może stanowić wartościowy wkład w rozwój metodologii analizy danych w obszarze motoryzacji oraz być inspiracją dla kolejnych badań związanych z oceną preferencji konsumentów w innych dziedzinach.

2. TEORIA ZBIORÓW PRZYBLIŻONYCH

Teoria zbiorów przybliżonych, wprowadzona przez profesora Zdzisława Pawlaka, stanowi rozszerzenie klasycznej teorii zbiorów, z naciskiem na analizę zbiorów w warunkach niepewności. Zbiór przybliżony jest kluczowym pojęciem tej teorii. Składa się on z dwóch części: przybliżenia dolnego i górnego. Te dwie części pozwalają na określenie pewności przynależności elementów do zbioru.

W kontekście analizy zbiorów, teoria zbiorów przybliżonych umożliwia elastyczne modelowanie relacji między elementami a zbiorami, zwłaszcza gdy pewność decyzji jest ograniczona. Dzięki zastosowaniu logiki trójwartościowej ta teoria pozwala na uwzględnienie różnych stopni przynależności, gdzie elementy mogą należeć do zbioru, nie należeć wcale lub pozostawać w stanie nieokreślonym.

Wariant tej teorii, znany jako zbiór przybliżony rozmyty, dodaje jeszcze jeden poziom elastyczności, umożliwiając stopniowanie przynależności elementów do zbioru. Istnieje także możliwość, że dany element może należeć zarówno do przybliżenia dolnego, jak i górnego, co wyraża pewną niepewność.

Stosując teorię zbiorów przybliżonych w analizie zbiorów, badacze i praktycy mogą skutecznie radzić sobie z sytuacjami, w których brak pełnej pewności co do przynależności elementów do określonych zbiorów. Ta teoria znalazła zastosowanie w dziedzinach takich jak eksploracja danych, sztuczna inteligencja i systemy decyzyjne, gdzie ważne jest uwzględnienie niepewności przy podejmowaniu decyzji.

3. METODY

Do przeprowadzania analizy zbioru danych wykorzystano opisane poniżej metody.

I. Algorytm genetyczny:

Algorytm genetyczny to heurystyczna metoda optymalizacyjna oparta na zasadach ewolucji biologicznej. Reprezentuje potencjalne rozwiązania problemu za pomocą genotypów, które ewoluują w populacji. Proces ewolucji obejmuje etapy selekcji, krzyżowania i mutacji. Selekcja ocenia jakość genotypów i faworyzuje te najbardziej obiecujące, które mają większą szansę na przekazanie swojego materiału genetycznego do kolejnej generacji. Krzyżowanie polega na wymianie informacji genetycznej między rodzicami, tworząc potomstwo, które może posiadać kombinację cech obu rodziców. Mutacja wprowadza niewielkie losowe zmiany w genotypie, zwiększając różnorodność populacji.

Algorytm genetyczny zaczyna od losowej populacji, a następnie iteracyjnie powtarza cykle selekcji, krzyżowania i mutacji, aby osiągnąć lepsze rozwiązania. W każdej iteracji oceniane są genotypy pod kątem ich dopasowania do określonych kryteriów. Algorytm ma zdolność przeszukiwania przestrzeni rozwiązań, adaptacji do zmiennych warunków i znajdowania rozwiązań zbliżonych do optymalnych. Ostateczny wynik to genotyp reprezentujący znalezione optymalne rozwiązanie lub zestaw rozwiązań. Wykorzystanie algorytmu genetycznego w generowaniu reduktów i reguł jest ściśle związane z teorią zbiorów przybliżonych. Może być on efektywnym narzędziem do ekstrakcji istotnych informacji z danych w kontekście tej teorii.

W przypadku generowania reduktów, algorytm genetyczny może być używany do identyfikacji zbiorów cech, które są minimalne (redundantne) i jednocześnie wystarczające do opisanego zbioru danych. Redukt w teorii zbiorów przybliżonych jest pewnym podzbiorem atrybutów, który minimalizuje utratę informacji przy klasyfikacji obiektów. Algorytm genetyczny może być wykorzystany do optymalizacji tego procesu, przeszukując przestrzeń możliwych reduktów.

W kontekście generowania reguł, algorytm genetyczny może pomóc w znalezieniu zestawu reguł decyzyjnych, które opisują dane zgodnie z koncepcją zbiorów przybliżonych. Reguły te są zazwyczaj konstruowane na podstawie relacji między atrybutami, a algorytm genetyczny może optymalizować ich skuteczność, eliminując niepotrzebne warunki lub redundancje.

II. Klasyfikacja za pomocą zestawu reguł:

Klasyfikacja za pomocą zestawu reguł polega na wykorzystaniu logicznych reguł decyzyjnych do

przyporządkowywania obiektów do określonych klas. Każda reguła w zestawie opisuje warunek lub kombinację warunków, które muszą być spełnione przez dany obiekt, aby został on zaklasyfikowany do konkretnej klasy. W przypadku klasyfikacji, reguły te są zazwyczaj konstruowane na podstawie cech lub atrybutów obiektów oraz związanych z nimi wartości.

W procesie klasyfikacji za pomocą zestawu reguł, obiekt jest analizowany na podstawie warunków opisanych w kolejnych regułach. W momencie spełnienia warunku danej reguły, obiekt zostaje przyporządkowany do odpowiadającej mu klasy. W przypadku, gdy obiekt nie spełnia warunków żadnej reguły, może być przypisany do domyślnej klasy lub pozostać niezaklasyfikowany.

Zaletą klasyfikacji za pomocą zestawu reguł jest ich zrozumiałość i interpretowalność. Reguły decyzyjne mogą być czytelnie sformułowane, co ułatwia zrozumienie procesu klasyfikacji nawet dla osób bez specjalistycznej wiedzy. Ponadto, takie podejście pozwala na łatwą modyfikację reguł w miarę pojawiania się nowych danych lub potrzeb związanych z analizą.

Klasyfikacja za pomocą zestawu reguł w kontekście teorii zbiorów przybliżonych umożliwia elastyczne podejście do rozpoznawania wzorców decyzyjnych, które mogą zawierać pewną stopę niepewności lub przybliżenia. Wykorzystując teorię zbiorów przybliżonych w procesie konstrukcji reguł, klasyfikacja staje się bardziej odporna na zmienność danych, a reguły są bardziej elastyczne w odzwierciedlaniu realnych zależności między atrybutami a decyzją. W rezultacie klasyfikacja oparta na zestawie reguł staje się potężnym narzędziem w modelowaniu skomplikowanych relacji decyzyjnych w obecności niepewności i przybliżeń.

III. Klasyfikacja za pomocą metody walidacji krzyżowej:

Klasyfikacja za pomocą metody walidacji krzyżowej to podejście do oceny wydajności modelu, które minimalizuje wpływ losowości podczas podziału danych na zbiór treningowy i testowy. W metodzie tej dane są dzielone na kilka podzbiorów, nazywanych foldami. Następnie model jest wielokrotnie trenowany i testowany na różnych kombinacjach tych podzbiorów, zawsze używając jednego jako zbioru testowego i pozostałych jako zbiorów treningowych. To pozwala na bardziej stabilne i reprezentatywne oszacowanie skuteczności modelu.

Podczas każdej iteracji, model jest trenowany na jednym zbiorze treningowym i testowany na zbiorze testowym.

Proces ten powtarza się tyle razy, ile jest foldów. Ostateczna skuteczność modelu to średnia wyników uzyskanych we wszystkich iteracjach. Walidacja krzyżowa pomaga w lepszym zrozumieniu, jak dobrze model generalizuje się do nowych danych, minimalizując ryzyko przetrenowania na konkretnym zestawie danych treningowych.

Metoda ta jest szczególnie przydatna w sytuacjach, gdzie ograniczone ilości danych mogą wprowadzić pewną losowość w procesie oceny modelu. Klasyfikacja za pomocą walidacji krzyżowej jest powszechnie stosowana w uczeniu maszynowym i statystyce, aby dostarczyć bardziej obiektywnej i rzetelnej oceny wydajności modelu, co przyczynia się do jego lepszej adaptacji do różnych warunków.

W kontekście teorii zbiorów przybliżonych, klasyfikacja za pomocą walidacji krzyżowej pozwala uwzględnić różnorodność danych treningowych i testowych, co ma szczególne znaczenie w sytuacjach, gdzie ograniczone zasoby danych mogą prowadzić do wprowadzenia losowości. Poprzez wielokrotne trenowanie i testowanie modelu na różnych podzbiórach danych, walidacja krzyżowa umożliwia bardziej wiarygodne oszacowanie skuteczności modelu, co przyczynia się do bardziej stabilnej oceny jego zdolności do dokładnej klasyfikacji nowych, nieznanych danych. W ten sposób, klasyfikacja za pomocą walidacji krzyżowej, zbliżając się do filozofii teorii zbiorów przybliżonych, stara się uwzględnić niepewność i przybliżenia w procesie oceny modelu, co prowadzi do bardziej obiektywnych i rzetelnych wyników.

4. WYKORZYSTANY ZBIÓR DANYCH

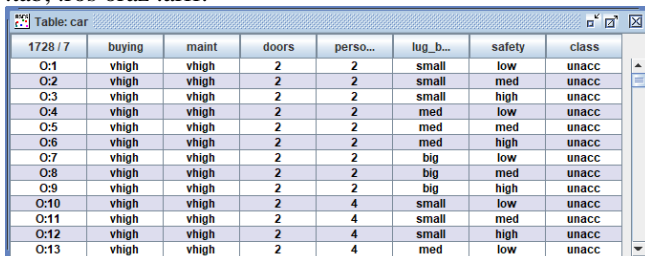
„Car Evaluation Database” to zbiór danych oceniający akceptowalność samochodów stworzony w 1997 roku. Powstał na bazie modelu decyzyjnego, który analizuje samochody pod kątem takich kategorii jak cechy techniczne czy komfort. Zawiera wpisy dotyczące 1728 samochodów, z 6 atrybutami wejściowymi wśród których znajdują się m.in. cena oraz liczba drzwi. Samochody są klasyfikowane jako „unacc” (70.023% całego zbioru), „acc” (22.222% całego zbioru), „good” (3.993% całego zbioru) lub „v-good” (3.762% całego zbioru). Zbiór ten jest przydatny do testowania metod konstruktywnej indukcji i odkrywania struktury, ze względu na znane hierarchie koncepcji.

5. ANALIZA ZBIORU ZA POMOCĄ PROGRAMU RSES

Program RSES zapewnia funkcjonalności umożliwiające przeprowadzenie analizy zbioru danych za pomocą wcześniej wspomnianych metod. Procedura postępowania w celu realizacji tej analizy obejmuje zestaw kroków, które zostały opisane poniżej:

I. Utworzenie projektu, dodanie obiektu tabeli i wczytanie do niego danych

Projekt należy utworzyć poprzez wybranie opcji „New Project” z zakładki „File”, następnie należy nacisnąć przycisk odpowiedzialny za dodanie tabeli z menu znajdującego się po lewej stronie. W celu wczytania danych do tabeli należy wybrać opcję „Load” z menu kontekstowego uprzednio dodanej tabeli. Program jest w stanie zaimportować dane z plików zapisanych w formatach .tab, .ros oraz .arff.

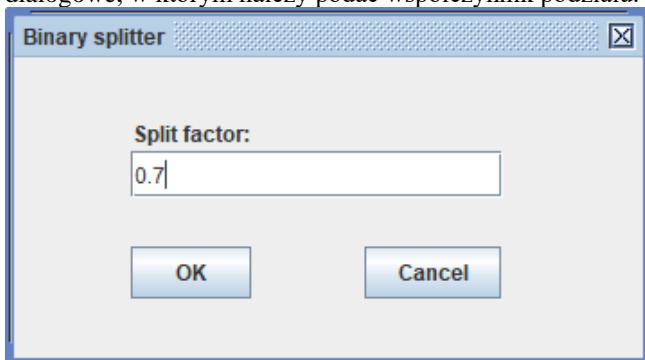


	buying	maint	doors	perso...	lug_b...	safety	class
O:1	vhigh	vhigh	2	2	small	low	unacc
O:2	vhigh	vhigh	2	2	small	med	unacc
O:3	vhigh	vhigh	2	2	small	high	unacc
O:4	vhigh	vhigh	2	2	med	low	unacc
O:5	vhigh	vhigh	2	2	med	med	unacc
O:6	vhigh	vhigh	2	2	med	high	unacc
O:7	vhigh	vhigh	2	2	big	low	unacc
O:8	vhigh	vhigh	2	2	big	med	unacc
O:9	vhigh	vhigh	2	2	big	high	unacc
O:10	vhigh	vhigh	2	4	small	low	unacc
O:11	vhigh	vhigh	2	4	small	med	unacc
O:12	vhigh	vhigh	2	4	small	high	unacc
O:13	vhigh	vhigh	2	4	med	low	unacc

Rysunek. 1 Fragment wczytanego do tabeli zbioru danych

II. Podział tabeli na część trenującą i testującą

Podziału można dokonać poprzez wybór opcji „Split in Two” z menu kontekstowego tabeli. Pojawi się wtedy okno dialogowe, w którym należy podać współczynnik podziału.

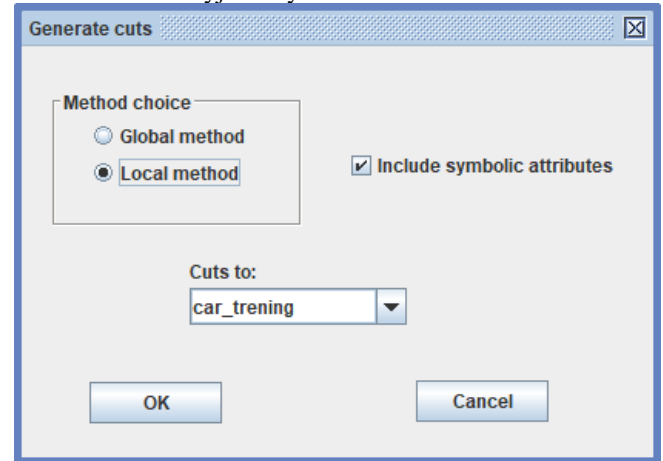


Rysunek. 2 Określanie współczynnika podziału tabeli

III. Generowanie cięć na części treningowej

Generowanie cięć umożliwia opcja „Generate cuts” dostępna w menu kontekstowym tabeli w kategorii

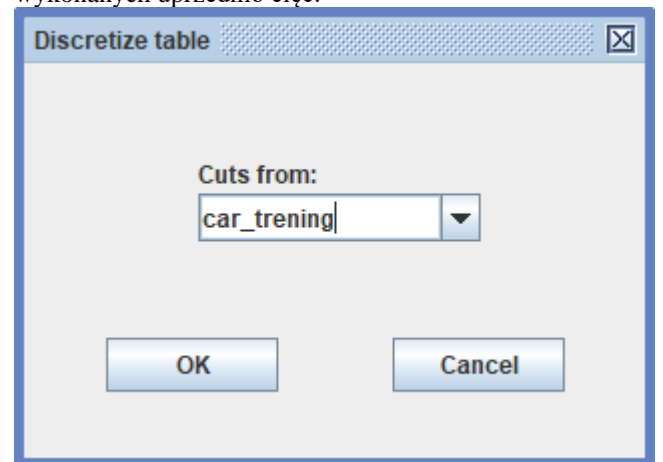
„Discretize”. Po jej wyborze należy jeszcze wybrać typ metod, określić czy uwzględnić atrybuty symboliczne oraz wskazać obiekt wyjściowy.



Rysunek. 3 Określanie opcji generowania cięć

IV. Dyskretyzacja części treningowej

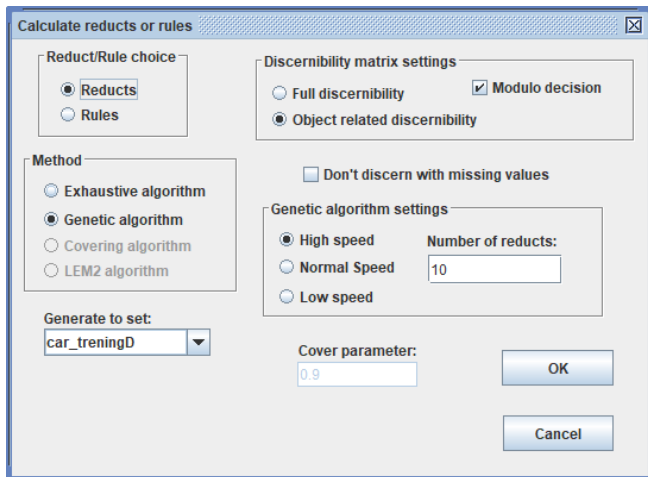
Opcja dyskretyzacji, podobnie jak opcja generowania cięć, również jest umieszczona w kategorii „Discretize”. Do przeprowadzenia tej operacji konieczne jest wskazanie wykonanych uprzednio cięć.



Rysunek. 4 Wskazywanie wygenerowanych cięć przy dyskretyzacji tabeli

V. Wyliczenie reduktu na zdyskretyzowanej części treningowej

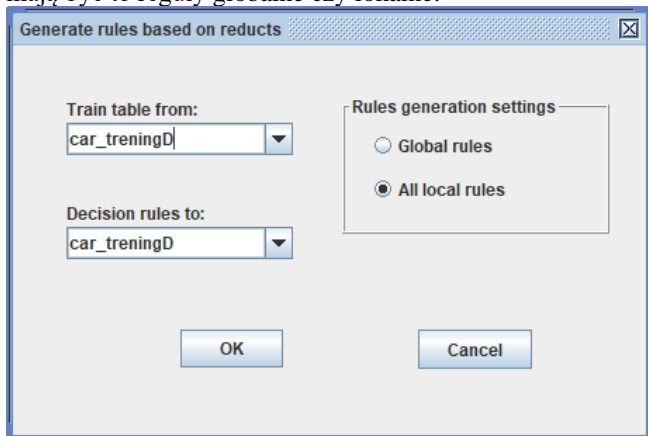
Redukty mogą zostać obliczone poprzez wybranie opcji „Calculate reducts or rules” znajdującej się w kategorii „Reducts/Rules”. Aby tego dokonać należy wskazać metodę, za pomocą której zostaną przeprowadzone obliczenia, jej szczegółowe ustawienia (jeśli są one możliwe do określenia) oraz ustawienia macierzy rozróżnialności.



Rysunek. 5 Wybór i ustawienia metody stosowanej do obliczenia reduktów oraz macierzy rozróżnialności

VI. Generowanie reguł na wyliczonym redukcje

Program umożliwia generowanie reguł bezpośrednio z poziomu reduktu. W tym celu z menu kontekstowego reduktu należy wybrać opcję „Generate rules” i określić czy mają być to reguły globalne czy lokalne.



Rysunek. 6 Ustawienia dotyczące generowania reguł

VII. Dyskretyzacja części testowej zbiorem cięć wygenerowanym na części treningowej

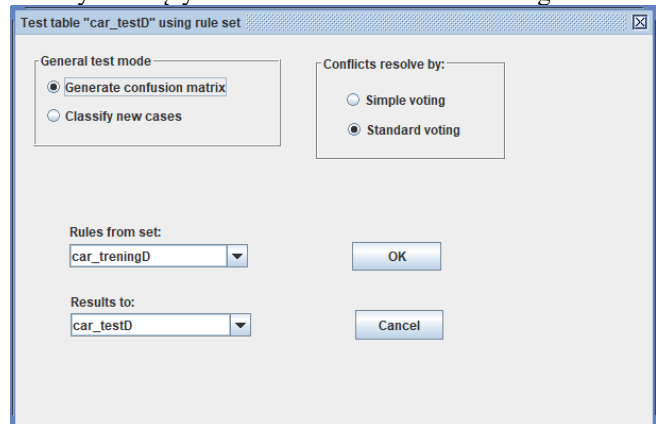
Krok ten jest analogiczny do kroku czwartego. Jediną różnicą jest to, że operacja przeprowadzana jest na części testowej. Poniżej przedstawiono fragment tych danych po dokonaniu dyskretyzacji.

519 / 7	buying	maint	doors	perso...	lug_b...	safety	class
O:1	{vhigh}	{high,vhigh...	{(2,3)}	{4}	{small}	"<call rest"	unacc
O:2	{vhigh}	{high,vhigh...	{(2,3)}	{4}	"<call rest"	{low}	unacc
O:3	{vhigh}	{high,vhigh...	{(2,3)}	{4}	"<call rest"	{med}	unacc
O:4	{vhigh}	{high,vhigh...	{(2,3)}	{4}	"<call rest"	"<call rest"	unacc
O:5	{vhigh}	{high,vhigh...	{(2,3)}	"<call rest"	"<call rest"	"<call rest"	unacc
O:6	{vhigh}	{high,vhigh...	{(2,3)}	{2}	{small}	{low}	unacc
O:7	{vhigh}	{high,vhigh...	{(2,3)}	{2}	{small}	{med}	unacc
O:8	{vhigh}	{high,vhigh...	{(2,3)}	{2}	{small}	"<call rest"	unacc
O:9	{vhigh}	{high,vhigh...	{(2,3)}	{2}	{med}	{med}	unacc
O:10	{vhigh}	{high,vhigh...	{(2,3)}	"<call rest"	{med}	{med}	unacc

Rysunek. 7 Fragment danych części testowej po dyskretyzacji

VIII. Klasyfikacja zdyskretyzowanych danych testujących z użyciem wygenerowanych reguł

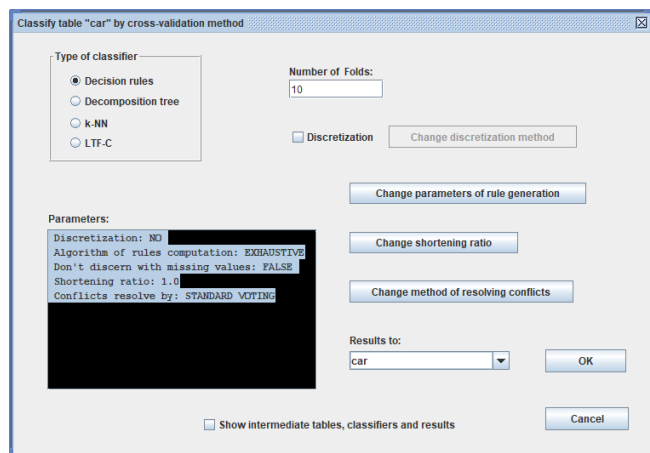
Opcje umożliwiające klasyfikację danych umieszczone są w kategorii „Classify” dostępnej z poziomu menu kontekstowego tabel. W celu przeprowadzenia tego typu klasyfikacji należy dokonać wyboru trybu testowania, metody rozwiązywania konfliktów oraz zbioru reguł.



Rysunek. 8 Ustawienia klasyfikacji danych za pomocą zbioru reguł

IX. Klasyfikacja zbioru danych za pomocą metody walidacji krzyżowej

Metodę klasyfikacji za pomocą metody walidacji krzyżowej można zastosować na bezpośrednio wczytanym zbiorze danych bez uprzedniego przeprowadzania na nim jakichkolwiek operacji. Ta opcja również jest dostępna z poziomu kategorii „Classify” w menu kontekstowym tabel. Po jej wybraniu użytkownikowi ukazuje się kompleksowe okno dialogowe ze wszystkimi kluczowymi ustawieniami powiązanych ze wspomnianą metodą.



Rysunek. 9 Okno dialogowe z ustawieniami metody walidacji krzyżowej

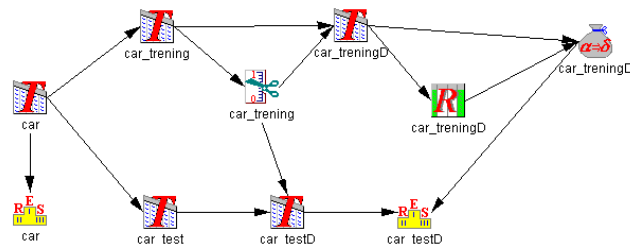
Program umożliwia przedstawienie wyników klasyfikacji w postaci macierzy pomyłek, która na skutek działania wybranej metody klasyfikującej zostaje zapisana jako osobny obiekt. Przykład prezentacji takiej macierzy w programie umieszczono poniżej.

		Predicted				No. of obj.	Accuracy	Coverage
		unacc	acc	vgood	good			
Actual	unacc	331	0	29	4	364	0.909	1
	acc	54	13	36	9	112	0.116	1
	vgood	0	0	23	0	23	1	1
	good	0	0	7	13	20	0.65	1
True positive rate		0.86	1	0.24	0.5			

Total number of tested objects: 519
Total accuracy: 0.732
Total coverage: 1

Rysunek. 10 Macierz pomyłek dla klasyfikacji zbioru danych dotyczących oceny akceptowalności samochodów za pomocą zbioru reguł na podstawie dziesięciu reduktów

Przeprowadzenie analizy zbioru danych dotyczących oceny akceptowalności samochodów za pomocą opisanych powyżej kroków przyczyniło się do powstania zamieszczonego poniżej schematu projektu w programie.



Rysunek. 11 Schemat projektu przeprowadzonej analizy zbioru danych dotyczących oceny akceptowalności samochodów w programie RSES

6. WYNIKI

Analiza została przeprowadzona według kroków opisanych w poprzednim rozdziale dla różnej liczby reduktów (10, 8, 6, 4, 2). Przy klasyfikacji za pomocą metody walidacji krzyżowej, oprócz liczby reduktów, w analogiczny sposób zmieniano również liczbę foldów. Wyniki zostały przedstawione w poniższych tabelach.

Tabela. 1 Wyniki klasyfikacji za pomocą zbioru reguł dla różnej liczby reduktów

Klasyfikacja za pomocą zbioru reguł		
Liczba reduktów	Całkowita dokładność	Całkowite pokrycie
10	73,2%	100%
8	76,7%	100%
6	74,6%	100%
4	69,4%	100%
2	70,9%	100%

Przeprowadzając klasyfikację za pomocą zbioru reguł dla 10 reduktów, uzyskano całkowitą dokładność na poziomie 73,2%, co oznacza, że zredukowane przypadki testowe zachowały wysoką reprezentatywność w porównaniu do pełnej puli. Całkowite pokrycie wyniosło 100%, oznacza, że wszystkie obiekty zostały uwzględnione w procesie testowania.

W przypadku 8 reduktów, całkowita dokładność wzrosła do 76,7%, co wskazuje na skuteczność redukcji przy minimalnej utracie reprezentatywności. Pokrycie utrzymało swoją wartość.

Dla 6 reduktów osiągnięto całkowitą dokładność na poziomie 74,6%, co utrzymuje trend wysokiej skuteczności

redukcji, a pokrycie w dalszym ciągu utrzymało poziom 100%.

W przypadku 4 reduktów dokładność spadła do 69,4%, co może wskazywać na większą utratę reprezentatywności przy bardziej znaczącej redukcji przypadków testowych, mimo że całkowite pokrycie zostało utrzymane.

Redukcja do 2 przypadków testowych przyniosła całkowitą dokładność na poziomie 70,9% bez zmian w całkowitym pokryciu.

Wyniki te sugerują, że dobór odpowiedniej liczby reduktów wpływa na efektywność procesu testowania, gdzie wysoka dokładność zachowuje się proporcjonalnie do stopnia redukcji, choć przy większej redukcji istnieje ryzyko utraty reprezentatywności.

Tabela. 2 Wyniki klasyfikacji za pomocą metody walidacji krzyżowej dla różnej liczby foldów i reduktów

Klasyfikacja za pomocą metody walidacji krzyżowej			
Liczba foldów	Liczba reduktów	Całkowita dokładność	Całkowite pokrycie
10	10	91,8%	98,7%
8	8	92,1%	98,8%
6	6	91,1%	98,9%
4	4	89,9%	99,5%
2	2	88,4%	99,6%

Koncentrując się na klasyfikacji z wykorzystaniem metody walidacji krzyżowej dla 10 foldów i 10 reduktów osiągnięto całkowitą dokładność na poziomie 91,8%, przy przetestowaniu 172 obiektów. Model uzyskał wysokie pokrycie na poziomie 98,7%, co sugeruje, że różnorodność przypadków testowych była dobrze reprezentowana.

W przypadku 8 foldów i 8 reduktów uzyskano podobnie wysoką dokładność wynoszącą 92,1% na próbie 216 obiektów. Pokrycie utrzymała się na wysokim poziomie 98,8%.

Dla 6 foldów i 6 reduktów całkowita dokładność wyniosła 91,1%, a liczba przetestowanych obiektów zwiększyła się do 288. Pokrycie pozostało na bardzo wysokim poziomie 98,9%.

Zmniejszenie liczby foldów do 4 i reduktów do 4 spowodowało spadek dokładności do 89,9%, ale pokrycie minimalnie wzrosło, osiągając 99,5% przy przetestowaniu 432 obiektów.

Dla 2 foldów i 2 reduktów dokładność wyniosła 88,4% na próbie 864 obiektów, a pokrycie osiągnęło najwyższy poziom 99,6%.

Podsumowując, wyniki wskazują na skuteczność modelu w klasyfikacji, z uwzględnieniem różnych konfiguracji foldów i reduktów. Warto zauważyć, że większa liczba foldów i reduktów może prowadzić do lepszej generalizacji, ale może jednocześnie wpływać na rozmiar próby testowej, co należy uwzględnić przy interpretacji wyników. Wysokie pokrycie oznacza, że różnorodność danych jest dobrze reprezentowana w procesie walidacji krzyżowej.

7. PODSUMOWANIE I WNIOSKI

Artykuł dotyczący oceny akceptowalności samochodów za pomocą teorii zbiorów przybliżonych i programu RSES przedstawia kompleksową analizę z użyciem różnych metod, obejmujących algorytm genetyczny, klasyfikację za pomocą zestawu reguł oraz poprzez metodę walidacji krzyżowej. Zastosowanie teorii zbiorów przybliżonych w połączeniu z programem RSES pozwala na wszechstronną analizę zbioru danych oraz generowanie reguł klasyfikacyjnych.

Wyniki uzyskane z analizy wskazują na wysoką skuteczność zastosowanych metod w ocenie akceptowalności samochodów. Zarówno algorytm genetyczny, jak i klasyfikacja za pomocą zestawu reguł prezentują obiecujące rezultaty, co potwierdza użyteczność tych podejść w kontekście analizy wspomnianych danych.

Program RSES okazuje się być efektywnym narzędziem do przeprowadzania analizy zbioru danych, generowania reguł klasyfikacyjnych oraz oceny skuteczności klasyfikacji za pomocą metody walidacji krzyżowej. Wyniki uzyskane na różnych etapach analizy potwierdzają poprawność zastosowanych procedur oraz skuteczność narzędzi użytych w badaniach.

Literatura

1. Skowron A., Suraj Z., „Rough Sets and Intelligent Systems - Professor Zdzisław Pawlak in Memoriam”, Springer-Verlag Berlin Heidelberg 2013

2. Pawlak Z., „Rough Sets: Theoretical Aspects of Reasoning about Data”, Springer Science+Business Media Dordrecht 1991
3. Han J., Kamber M., Pei J. “ Data Mining: Concepts and Techniques”, Elsevier, 2011
4. Provost F., Fawcett T., “Data Science for Business. What You Need to Know about Data Mining and Data-Analytic Thinking”, O'Reilly Media, 2013
5. RSES 2.2. Rough Set Exploration System. Podręcznik Użytkownika. Publikacja elektroniczna, [Data dostępu: 19.01.2024] https://www.mimuw.edu.pl/~szczuka/rses/RSES_doc.pdf