**WYŻSZA SZKOŁA INFORMATYKI**

**I ZARZĄDZANIA**

Z SIEDZIBĄ W RZESZOWIE

Sprawozdanie

Sztuczna inteligencja

**Metody generowania, oceny i optymalizacji reguł decyzji**

Prowadzący: dr inż. Mariusz Wrzesień Wykonawca: Kunigunda Orlowska w11111

Rzeszów 2019

# Opis problemu

Celem laboratorium jest nauczyć się parametrom oceny i optymalizacji zbioru reguł. W pierwszym zadaniu trzeba ocenić wcześniej wygenerowany zbiór reguł na podstawie parametrów oceny. Drugie zadanie polega na porównaniu dwóch metod generowania reguł, standardowej i algorytmu GTS, przed i po optymalizacji.

# Badane zbiory danych

W pierwszym zadaniu dane dotyczą problemu doboru soczewek. Są one przedstawione za pomocą tablicy decyzji, która składa się z pięciu kolumn i z dwudziestu dwóch wierszy. Wszystkie wiersze reprezentują poszczególne przypadki(obiekty). Pierwsza kolumna reprezentuje atrybut porządkowy o nazwie „wiek”. Zawiera trzy wartości atrybutów: młody, starczy i prestarczy. Druga kolumna reprezentuje atrybut binarny „wada wzroku”. Posiada dwie wartości, dalekowidz oraz krótkowidz. Trzecia kolumna reprezentuje atrybut binarny „astygmatyzm”, który także składa się z dwóch wartości: tak i nie. Przedostatnia kolumna reprezentuje atrybut binarny „łzawienie” wraz z dwoma wartościami, normalne i zmniejszone. Ostatnia kolumna ma nazwę atrybutu „soczewki” i jest to kolumna decyzji, która składa się z trzech klas: brak, miękkie, twarde.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| mlody | krotkowidz | nie | normalne | miekkie |
| mlody | dalekowidz | tak | normalne | twarde |
| mlody | dalekowidz | nie | zmniejszone | brak |
| mlody | krotkowidz | tak | zmniejszone | brak |
| prestarczy | krotkowidz | tak | zmniejszone | brak |
| prestarczy | krotkowidz | nie | normalne | miekkie |
| mlody | krotkowidz | tak | normalne | twarde |
| starczy | dalekowidz | tak | zmniejszone | brak |
| prestarczy | dalekowidz | nie | zmniejszone | brak |
| prestarczy | dalekowidz | tak | zmniejszone | brak |
| prestarczy | krotkowidz | nie | zmniejszone | brak |
| mlody | dalekowidz | tak | zmniejszone | brak |
| starczy | dalekowidz | tak | normalne | brak |
| starczy | krotkowidz | tak | zmniejszone | brak |
| starczy | krotkowidz | nie | normalne | brak |
| prestarczy | krotkowidz | tak | normalne | twarde |
| mlody | krotkowidz | nie | zmniejszone | brak |
| starczy | krotkowidz | nie | zmniejszone | brak |
| starczy | dalekowidz | nie | zmniejszone | brak |
| mlody | dalekowidz | nie | normalne | miekkie |
| starczy | dalekowidz | nie | normalne | miekkie |
| prestarczy | dalekowidz | tak | normalne | brak |

W drugim zadaniu wykorzystujemy zbiory danych: car, glass, iris, P250134, P250144, P410144, P548144, tea i wine. Każdy zbiór jest podzielony na dwie części, pierwszą wykorzystujemy do generowania reguł, drugą do testowania wygenerowanych reguł.

# Obliczenia

ZADANIE 1

A priori mamy wygenerowany zbiór reguł :

* R1 JEŻELI Łzawienie JEST zmniejszone TO SOCZEWKI JEST brak
* R2 JEŻELI Wiek JEST młody ORAZ Astygmatyzm JEST nie ORAZ Łzawienie JEST normalne TO SOCZEWKI JEST miękkie
* R3 JEŻELI Wada\_wzroku JEST krótkowidz ORAZ Astygmatyzm JEST tak ORAZ Łzawienie JEST normalne TO SOCZEWKI JEST twarde
* R4 JEŻELI Wiek JEST prestarczy ORAZ Astygmatyzm JEST nie ORAZ Łzawienie JEST normalne TO SOCZEWKI JEST miękkie
* R5 JEŻELI Wiek JEST młody ORAZ Astygmatyzm JEST tak ORAZ Łzawienie JEST normalne TO SOCZEWKI JEST twarde
* R6 JEŻELI Wiek JEST młody ORAZ Astygmatyzm JEST tak ORAZ Łzawienie JEST normalne TO SOCZEWKI JEST twarde .

Zadaniem jest wykonanie dla tych reguł obliczeń ustalających jaką siłę, dokładność, ogólność, specyficzność I wsparcie ma każda z nich.

Parametr siła określa liczbę poprawnie klasyfikowanych przypadków.

Siła = Ep.

Parametr dokładność jest stosunkiem liczby poprawnie klasyfikowanych przypadków (Ep) z bazy informacyjnej do sumy przypadków klasyfikowanych poprawnie (Ep) oraz błędnie (Eb).

Dokładność =

Parametr ogólność jest stosunkiem sumy przypadków klasyfikowanych poprawnie (Ep) oraz błędnie (Eb) do całkowitej liczby przypadków w zbiorze uczącym (E).

Ogólność =

Parametr specyficzność jest stosunkiem liczby poprawnie klasyfikowanych przypadków (Ep) z bazy informacyjnej do liczby przypadków w tej bazie, należących do danej klasy decyzyjnej (Eclass).

Specyficzność =

Parametr wsparcie to stosunek liczby poprawnie klasyfikowanych przypadków (Ep) z bazy informacyjnej do całkowitej liczby przypadków w tym zbiorze (E).

Wsparcie =

Dla pierwszej reguły

Siła = Ep = 12

Dokładność = = 12/12 = 1

Ogólność = = 12/22 = 0.54

Specyficzność = = 12/15 = 0.8

Wsparcie = = 12/22 = 0.54

I tak samo liczymy dla wszystkich reguł

ZADANIE 2

W drugim zadaniu do obliczeń został wykorzystany program RuleSEEKER, za pomocą którego został obliczony błąd klasyfikacji dla poszczególnych zbiorów dwoma metodami generowania reguł: algorytmem GTS i metodą standardową. Później liczymy średni błąd dla wszystkich zbiorów dla danej metody. Potem optymizujemy nasze reguły i liczymy wszystko ponownie.

# Wyniki

ZADANIE 1

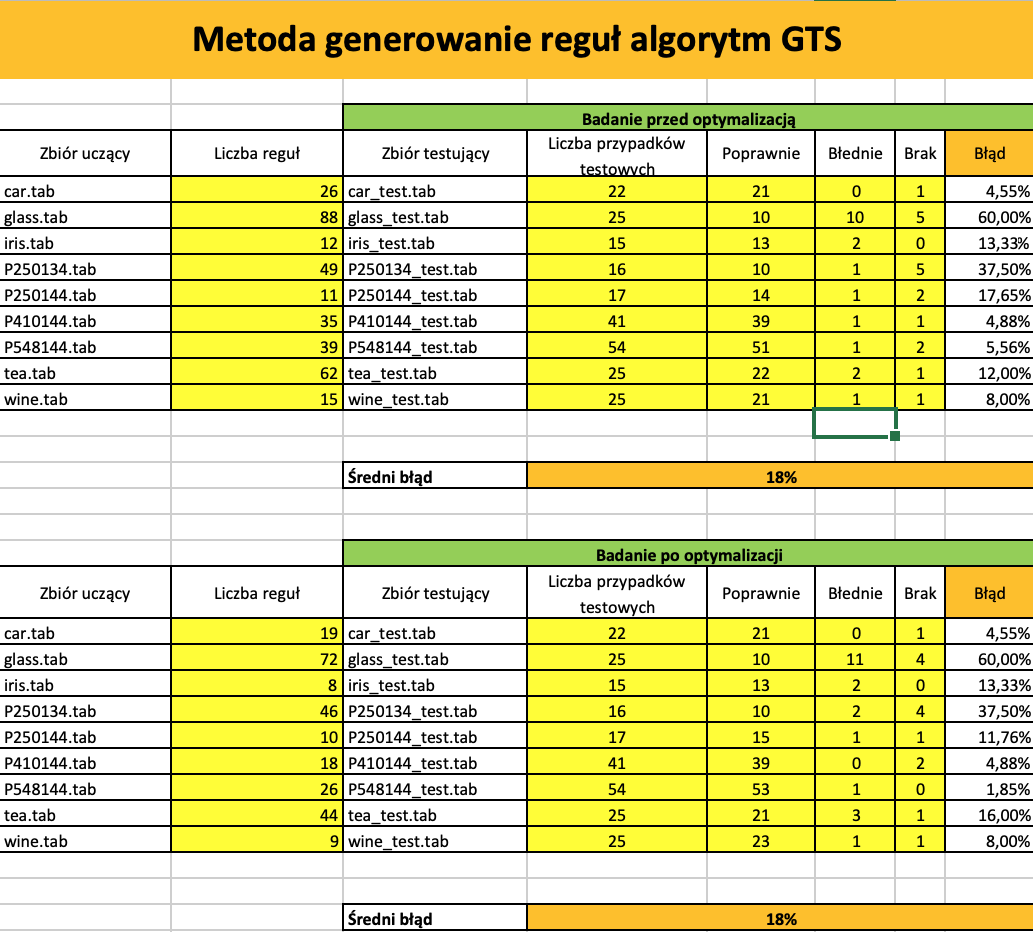
Po wyliczeniu wszystkich parametrów oceny, dostajemy takie wyniki:

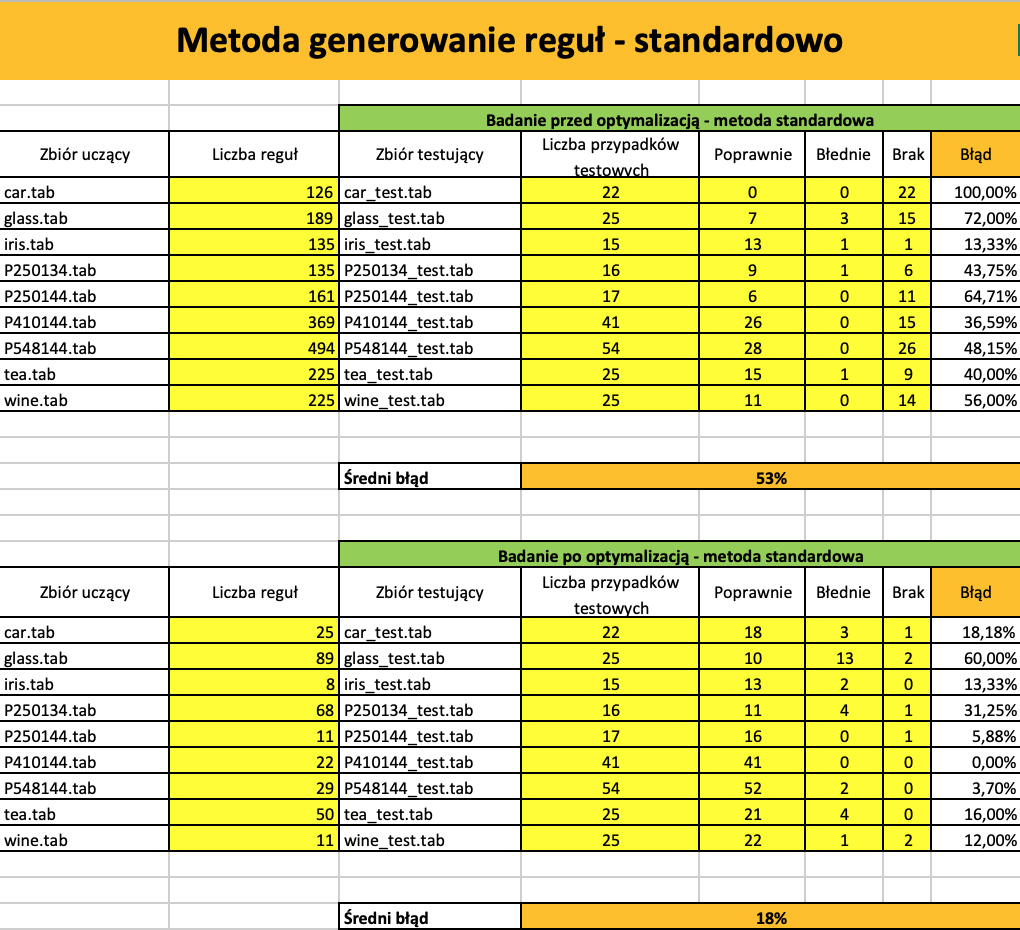
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Siła | Dokładność | Ogólność | Specyficzność | Wsparcie |
| REGULA 1 | 12 | 1 | 0,545454545 | 0,8 | 0,545454545 |
| REGULA 2 | 2 | 1 | 0,090909091 | 0,5 | 0,090909091 |
| REGULA 3 | 2 | 1 | 0,090909091 | 0,666666667 | 0,090909091 |
| REGULA 4 | 1 | 1 | 0,045454545 | 0,25 | 0,045454545 |
| REGULA 5 | 2 | 1 | 0,090909091 | 0,666666667 | 0,090909091 |
| REGULA 6 | 2 | 1 | 0,090909091 | 0,666666667 | 0,090909091 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Wiek | Wada\_wzrok | Astygmatyzm | Lzawienie | SOCZEWKI |
| mlody | krotkowidz | nie | normalne | miekkie |
| mlody | dalekowidz | tak | normalne | twarde |
| mlody | dalekowidz | nie | zmniejszone | brak |
| mlody | krotkowidz | tak | zmniejszone | brak |
| prestarczy | krotkowidz | tak | zmniejszone | brak |
| prestarczy | krotkowidz | nie | normalne | miekkie |
| mlody | krotkowidz | tak | normalne | twarde |
| starczy | dalekowidz | tak | zmniejszone | brak |
| prestarczy | dalekowidz | nie | zmniejszone | brak |
| prestarczy | dalekowidz | tak | zmniejszone | brak |
| prestarczy | krotkowidz | nie | zmniejszone | brak |
| mlody | dalekowidz | tak | zmniejszone | brak |
| starczy | dalekowidz | tak | normalne | brak |
| starczy | krotkowidz | tak | zmniejszone | brak |
| starczy | krotkowidz | nie | normalne | brak |
| prestarczy | krotkowidz | tak | normalne | twarde |
| mlody | krotkowidz | nie | zmniejszone | brak |
| starczy | krotkowidz | nie | zmniejszone | brak |
| starczy | dalekowidz | nie | zmniejszone | brak |
| mlody | dalekowidz | nie | normalne | miekkie |
| starczy | dalekowidz | nie | normalne | miekkie |
| prestarczy | dalekowidz | tak | normalne | brak |

ZADANIE 2

Mamy takie wyniki:





# Wnioski

ZADANIE 1

Badany zbiór reguł okazał się dobrej jakości. Dokładność każdej reguły wynosi 100%, co oznacza, że żadna z tych reguł nie doprowadziłaby do błędnego wyniku przy klasyfikacji danego zbioru przypadków. Natomiast trzeba powiedzieć, że wygenerowany wcześniej zbiór reguł, jak widać, pokrywa nasz zbiór przypadków nie do końca, 4 przypadki z 22 są niepokryte regułami, dlatego ten zbiór reguł trzeba dopracować generacją niedostatecznych reguł.

ZADANIE 2

Mając wyniki błędów przy klasyfikacji przed i po optymalizacji dla dwóch metod generowania reguł, możemy zrobić pewne wnioski. Algorytm GTS od razu dobrze poradził sobie z generowaniem zbioru reguł, liczba reguł dla poszczególnych zbiorów i błąd klasyfikacji przed optymalizacją są dosyć małe. Po optymalizacji błąd klasyfikacji prawie się nie zmienił, natomiast w dużym stopniu zmniejszyła się liczba reguł.

Metoda standardowa przed optymalizacją pokazała znacznie gorszy wynik niż GTS. Wygenerowany zbiór reguł był przerażająco duży. Ale po optymalizacji wszystko się zmieniło metoda standardowa pokazała wynik średniego błędu na poziomie GTS, a liczba reguł była tylko trochę większa.

Na podstawie tych wszystkich wyników można stwierdzić, że optymalizacja jest ważną częścią tworzenia zbioru reguł i pomaga w znacznym stopniu polepszyć istniejący zbiór reguł.