



Sprawozdanie z laboratorium
Przedmiot : Inżynieria wiedzy i symboliczne uczenie
maszynowe
Lab : 01
Data : 2022.11.12
Autor : Grzegorz Kuliński
Mail : gkulinski@student.agh.edu.pl

1. Zbiór danych

Utworzony zbiór danych opisuje pojazdy. Zawiera w sobie takie atrybuty takie jak masa własna, pojemność baku paliwa czy maksymalna predkość. Dane każdego z pojazdów sa oparte na faktycznych statystykach, które udało mi sie znaleźć w sieci. Jeśli dany pojazd miał wiele generacji/konkretnych wersji, to starałem sie wybrać statystyki tej najbardziej popularnej wersji.

Na podstawie zbioru przeprowadzałem klasyfikacje jakim typem pojazdu spośród {*auto sportowe*, *motocykl*, *suv*, *ciężarówka*, *sedan*} jest dany środek transportu

Wszystkie cechy zbioru wypisane sa poniżej:

```
@attribute max_speed numeric
@attribute mass numeric
@attribute num_of_doors numeric
@attribute fuel_tank_cap numeric
@attribute max_passangers numeric
@attribute num_of_tires numeric
@attribute avg_fuel_coms numeric
@attribute make {'porsche', 'jaguar', 'Bmw', 'suzuki'...}
@attribute fuel_type {petrol,diesel}
@attribute type {sport_car,motorcycle,suv,lorry,edan}
```

2. Eksperymenty

Do eksperymentów wybrałem 5 różnych algorytmów.

- (1) `trees.BFTree`
- (2) `bayes.NaiveBayes`
- (3) `bayes.BayesNet`

- (4) `rules.ZeroR`
- (5) `trees.J48`

Badany zestaw danych został podzielony na 80 % treningowego i 20 % zbioru testowego. Pod względem skuteczności klasyfikacji najlepiej sprawdziły się *BFTree* osiągając 96 % skuteczności (tab. 1).

trees.BFTree	bayes.Naive	bayes.BayesN	rules.ZeroR	trees.J48
96.33	91.00	93.33	18.00	88.00

Cuadro 1: Zestawienie wyników dla *Percent_correct*

W przypadku badania *Cross-validation* zbiór testujący składał się z 10 % elementów. Wyniki poszczególnych algorytmów były bardziej zbliżone do siebie (około 88 %) wyłączając *rules.Zero* (tab. 2).

trees.BFTree	bayes.Naive	bayes.BayesN	rules.ZeroR	trees.J48
88.67	88.00	89.00	0.00	88.67

Cuadro 2: Zestawienie wyników dla *Percent_correct*

W obu przypadkach najgorzej wypadł wynik przy klasyfikacji z użyciem algorytmu *rules.Zero*. Najlepiej sorawdzają się do przedstawionego problemu drzewa (lasy) losowe, co nie jest raczej zaskoczeniem. Zwłaszcza algorytmu *trees.BFTree* oraz *trees.J48*.

Szczegółowe wyniki są zawarte w pliku *results.txt* oraz *auto_out.arff*.