**SPRAWOZDANIE**

Zajęcia: Analiza Procesów Uczenia

Prowadzący: prof. dr hab. Vasyl Martsenuk

**Laboratorium** 5

Data: 20.04.2020r

**Temat**: Modelowanie procesów uczenia maszynowego w pakecie mlr. Trenowanie, ocena i porównywanie modeli w pakecie mlr.

**Wariant** 1

Zadanie 1

Aleksander Słodczyk

Informatyka II stopień

Stacjonarne

1 semestr

Grupa 2

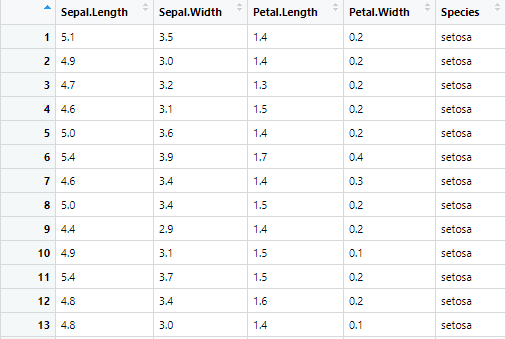
1. **Polecenie**:

Zadanie dotyczy konstruowania drzew decyzyjnych oraz reguł klasyfikacyjnych na podstawie zbioru danych (library(MASS lub datasets)).

Wariant 1 – iris.

1. **Wprowadzane dane**

Dane to zbiór 150 kwiatów „iris”. Został przedstawiony na poniższym obrazku. Na podstawie długości płatków należy określić gatunek Irysa.



1. **Wykorzystane komendy:**

**Drzewo decyzyjne - rpart**

* Komenda **rpart** tworzy model drzewa decyzyjnego. Na wyjściu jest model rpart

model <- **rpart**(Species ~., data = iris)

* Komenda **rpart.plot(model)** rysuje model drzewa rpart z wypisanymi poszczególnymi właściwościami drzewa.
* Komenda **predict** przewiduje na podstawie modelu gatunek irysa

model %>% **predict**(newdata, "class")

**Reguły klasyfikacyjne – C5.0**

* Komenda **C5.0** tworzy model drzewa decyzyjnego. Na wyjściu jest model C5.0

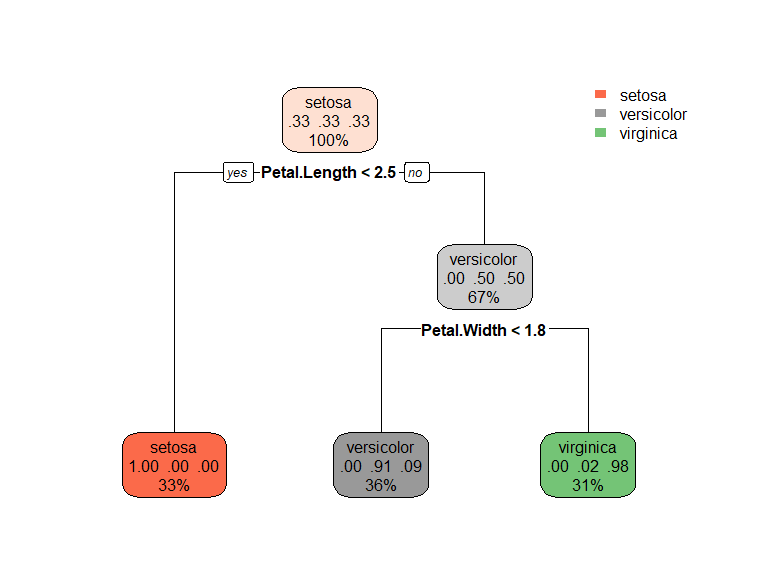
mod1 <- **C5.0**(Species ~ ., data = iris, rules = TRUE)

* Komenda **plot(model)** rysuje model drzewa C5.0 z wypisanymi poszczególnymi wartościami gałęzi drzewa.
* Komenda **summary(model)** tworzy podsumowanie wyników drzewa

1. **Wynik działania**

**Drzewo decyzyjne - rpart**

Dla drzewa decyzyjnego rpart otrzymałem następujący model drzewa:



Po utworzeniu przykładowego kwiatu, na podstawie modelu wybrany został gatunek irysa *virginica,* co odpowiada modelowi drzewa decyzyjnego.

> newdata <- data.frame(

+ Sepal.Length = 6.5, Sepal.Width = 3.0,

+ Petal.Length = 5.2, Petal.Width = 2.0

+ )

> model %>% predict(newdata, "class")

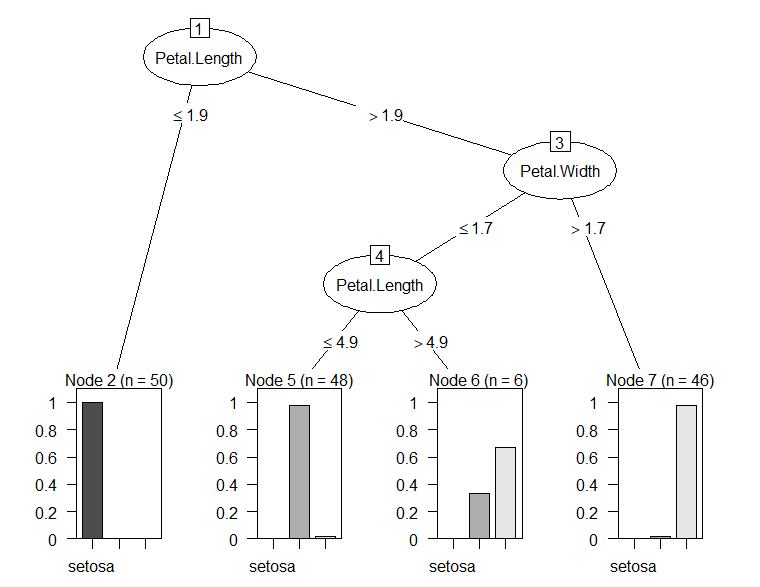
1

virginica

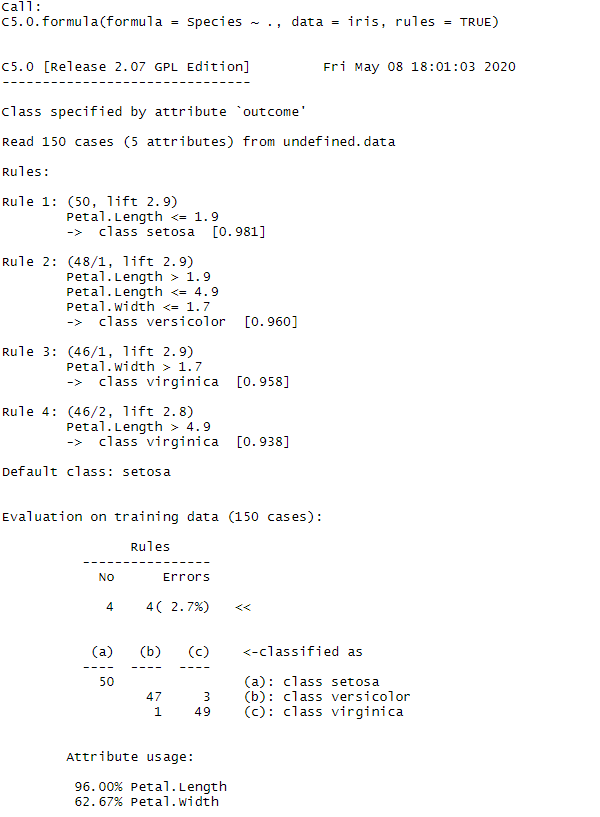
Levels: setosa versicolor virginica

Z drzewa wynika, że najbardziej znaczące są zmienne Petal.Length i Petal.Width, na podstawie których da się określić już gatunek irysa.  **Reguły klasyfikacyjne – C5.0**

Dla reguł klasyfikacyjnych otrzymałem następujący model drzewa:



Po wykonaniu metody C5.0 oraz summary dostałem następujące wyniki:



Dla reguł klasyfikacyjnych również okazuje się, że największy wpływ na gatunek irysa mają zmienne odpowiadające za szerokość i długość płatka irysa czyli Petal.Width i Petal.Length. Różnią się tu jednak wartościami oraz jest większy podział drzewa. Błąd w tym przypadku wyniósł 2.7%.

1. **Wnioski:**

* Oba drzewa przedstawiają podobny model oceny gatunku irysa.
* Reguły klasyfikacyjne zawierają więcej rozgałęzień, więc są przez to dokładniejsze

1. **Link do githuba:**

https://github.com/allo97/Analiza-procesow-uczenia-Programming-in-R/tree/master/lab5

Skrypty do drzewa decyzyjnego i reguł klasyfikacyjnych znajdują się odpowiednio w pliku zadanie/**skrypt1.R i** zadanie/**skrypt2.R**

Obraz sieci znajduje się w pliku zadanie/ **DrzewoDecyzyjneC50.png** i zadanie/**DrzewoDecyzyjneRpart.png**

Wyniki z consoli znajdują się w pliku zadanie/**skrypt1\_console\_log**  i zadanie/**skrypt2\_console\_log**