

Moduł prognostyczny

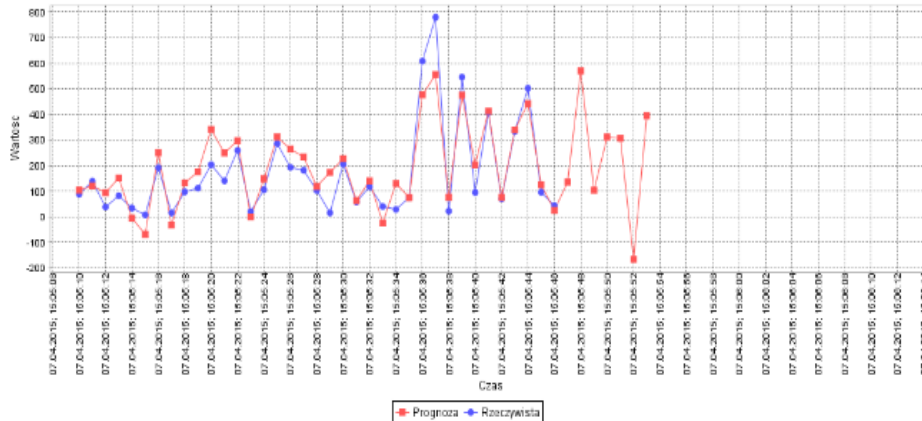
- Bazuje na modelach prognostycznych (np. utworzonych w module analitycznym)
- Uruchamia usługi prognozowania on-line (wirtualny czujnik)
- Monitoruje jakość prognoz (docelowo: FP, FN)
- Dostosowuje parametry modelu prognozowania

Moduł prognostyczny

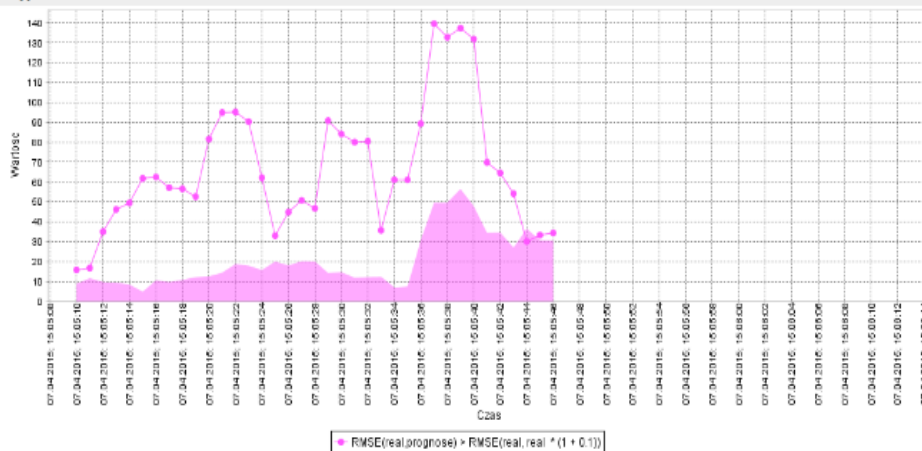
DISESOR: Wizualizacja usługi

Wyniki klasyfikacji/regresji

usługa-regresja, predykcja atrybutu label



Miary jakości

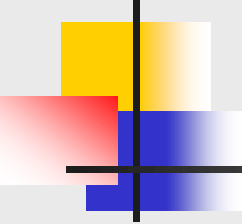


Reguły monitorowania

Reguła	Pokaz miarę	Pokaz obszar alarmowy
$RMSE(real.prognose) < 50.0$	☐	☐
$RMSE(real.prognose) > RMSE(real.real * 0.1)$	☐	☐
$RMSE(real.prognose) > RMSE(real.real * 0.1)$	☐	☐
$RMSE(real.prognose) > RMSE(real.real * (1 + 0.1))$	☐	☐
$RMSE(real.prognose) > RMSE(real.real * (1 + 0.1))$	☒	☒
$RMSE(real.prognose) < RMSE(real.real * (1 + 0.1))$	☐	☐

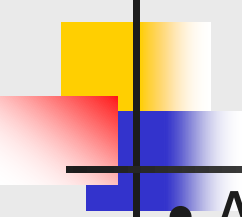
Alerty monitorowania

Zdarzenie	Naruszona reguła	Czas
07.04.2016 15:05:10	$RMSE(real.prognose) < RMSE(real.real * (1 + 0.1))$	wyzwolony
07.04.2016 15:05:15	$RMSE(real.prognose) < 50.0$	wyzwolony
07.04.2016 15:05:26	$RMSE(real.prognose) < 50.0$	zakonczone
07.04.2016 15:05:29	$RMSE(real.prognose) < 50.0$	wyzwolony
07.04.2016 15:05:44	$RMSE(real.prognose) > RMSE(real.real * (1 + 0.1))$	wyzwolony
07.04.2016 15:05:44	$RMSE(real.prognose) > RMSE(real.real * (1 + 0.1))$	wyzwolony
07.04.2016 15:05:44	$RMSE(real.prognose) < RMSE(real.real * (1 + 0.1))$	zakonczone
07.04.2016 15:05:45	$RMSE(real.prognose) < 50.0$	zakonczone
07.04.2016 15:05:45	$RMSE(real.prognose) > RMSE(real.real * (1 + 0.1))$	zakonczone
07.04.2016 15:05:45	$RMSE(real.prognose) > RMSE(real.real * (1 + 0.1))$	zakonczone
07.04.2016 15:05:45	$RMSE(real.prognose) < RMSE(real.real * (1 + 0.1))$	wyzwolony



Moduł prognostyczny

- „Przeuczanie” – pobranie najnowszych danych i sprawdzenie czy spełnione są kryteria minimalnej jakości
- W razie niepowodzenia – komunikat o zaprzestaniu generowania prognoz



Moduł analityczny

- Analiza trendu
 - Klasyfikacja
 - Regresja
 - Przeżycie
-
- Meta-operatory tematyczne (prognozowanie metanu, trendy zmian CO, CO₂)
 - Predefiniowane schematy klasyfikacji i regresji – tzw. raporty analityczne – asystent analityka (użytkownik niezaawansowany)
 - **Problem otwarty:** jak szacować czas obliczeń dla ustalonego zbioru i metody?

Moduł analityczny

- Asystent „analityka” (m.in. gridowe przeszukiwanie przestrzeni parametrów)

Main Process

Data (out) → Select Attrib... (exa, ori) → Analytics (In, exa, mod, rep)

Algorithm Configuration

Algorytm ma działać	Nazwa algorytmu	Ustawienie parametrów
☒	knn	Ustaw parametry
☒	svm	Ustaw parametry
☒	lasz losowe	Ustaw parametry

Parameter Setting Dialog

sigma	c
1	0.5
1	0.6
1	0.7

Ok Dodaj parametry Usuń parametry ☐ Domyślne parametry

Analytics (Analytics Reports)

Plik wyjściowy: C:\Users\demo\Desktop\output.html

☐ Asocjacja
☐ Szereg czasowy
☒ Statystyki opisowe
☒ Wykresy zmiennych
☒ Korelacja

Ustawienie parametrów algorytmów

Sposób oceny: podział na zbiór treningowy i testowy

zbiór treningowy (%): 70

Metoda strojenia parametrów: boot

Parametr number w train: 10

Kryterium strojenia parametrów klasyfikacja: Accuracy

Moduł analityczny

• Asystent „analitika” - raporty

- Ogólne informacje o raporcie
- Dane
 - Statystyki opisowe
- Analiza korelacji
 - Macierz korelacji
 - Korelacje ze zmienną prognozowaną
- Wykresy zmiennych
 - Histogramy i wykresy słupkowe
 - Wykres rozrzutu zmiennych numerycznych
- Reguły decyzyjne
 - Model
 - Postać końcowego modelu
 - Ocena istotności zmiennych
 - Ocena dokładności
 - Współczynniki dokładności
 - Macierz pomyłek
 - Krzywa ROC
- Drzewa decyzyjne
 - Model
 - Optymalne parametry
 - Raport ze strojenia
 - Krzywa ROC
 - Ocena modelu
 - Macierz pomyłek
 - Istotność zmiennych
- Gradient boost machine
 - Raport ze strojenia
 - Optymalne parametry
 - Ocena istotności zmiennych
 - Ocena dokładności modelu
 - Krzywa ROC
- K-najbliższych sąsiadów
 - Model
 - Wykres

Ogólne informacje o raporcie

- **Data i czas utworzenia:** 2015-10-21 11:45:06
- **Typ predykcji:** Klasyfikacja
- **Wybrane algorytmy:** regresja logistyczna, drzewa decyzyjne, reguły decyzyjne, kNN, naiwny klasyfikator bayesowski, svm, lasy losowe, gbm
- **Metoda oceny algorytmów:** na całym zbiorze danych
- **Metoda strojenia parametrów:** boot
- **Strojenie parametrów na podstawie oceny błędu:** BAC
- **Dane traktowane jako szereg czasowy:** nie

Dane

- Liczba przypadków: 150
- Liczba atrybutów: 5
- Zmienna prognozowana: label

Statystyki opisowe

Min - wartość minimalna

Q1 - pierwszy kwartył

Me - mediana

Mean - średnia

Q3 - trzeci kwartył

Max - wartość maksymalna

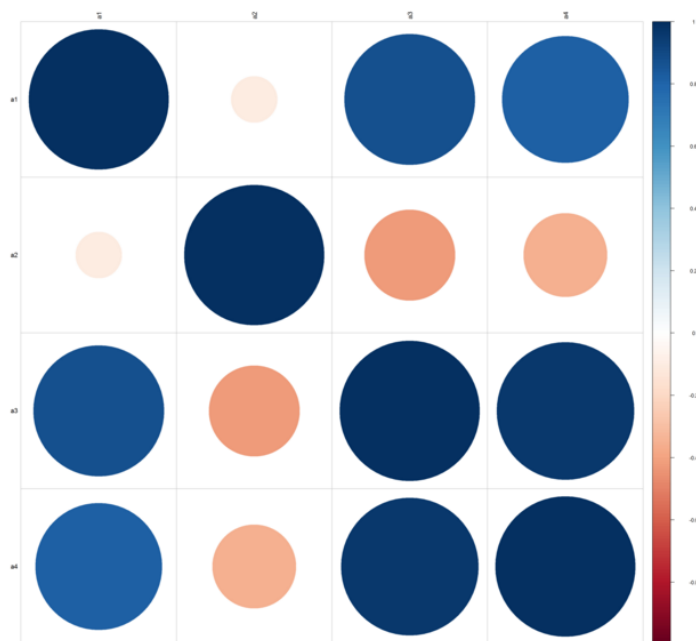
	Min	Q1	Me	Mean	Q3	Max
a1	4.3	5.1	5.80	5.84	6.4	7.9
a2	2.0	2.8	3.00	3.05	3.3	4.4
a3	1.0	1.6	4.35	3.76	5.1	6.9
a4	0.1	0.3	1.30	1.20	1.8	2.5

Moduł analityczny

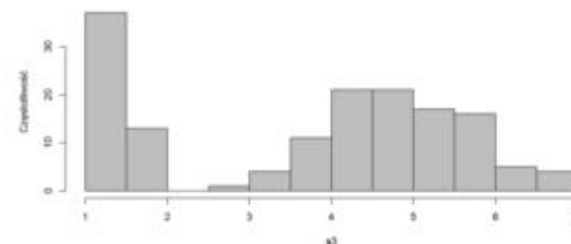
- Asystent „analitika”

Analiza korelacji

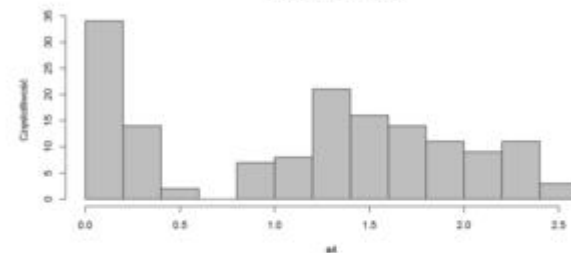
Macierz korelacji



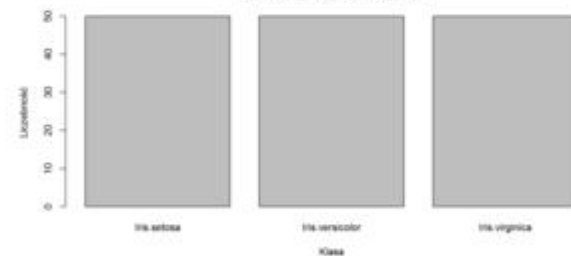
Histogram zmiennej a3



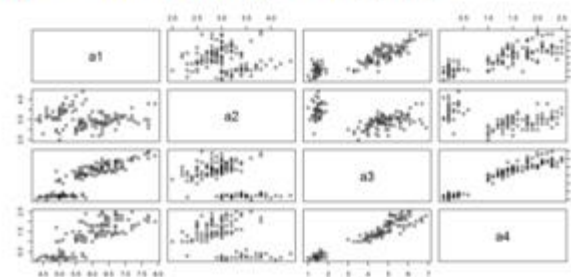
Histogram zmiennej a4



Wykres słupkowy zmiennej label



Wykres rozrzutu zmiennych numerycznych



Moduł analityczny

- Asystent „analityka”

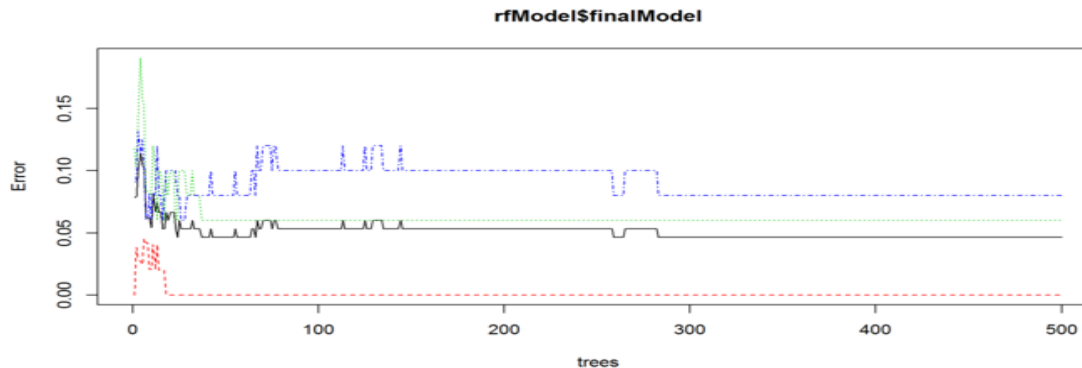
Lasy losowe

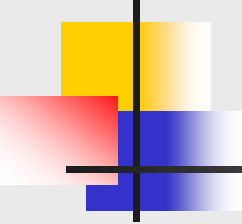
Optymalne parametry modelu:

```
mtry  
1
```

Postać końcowa modelu:

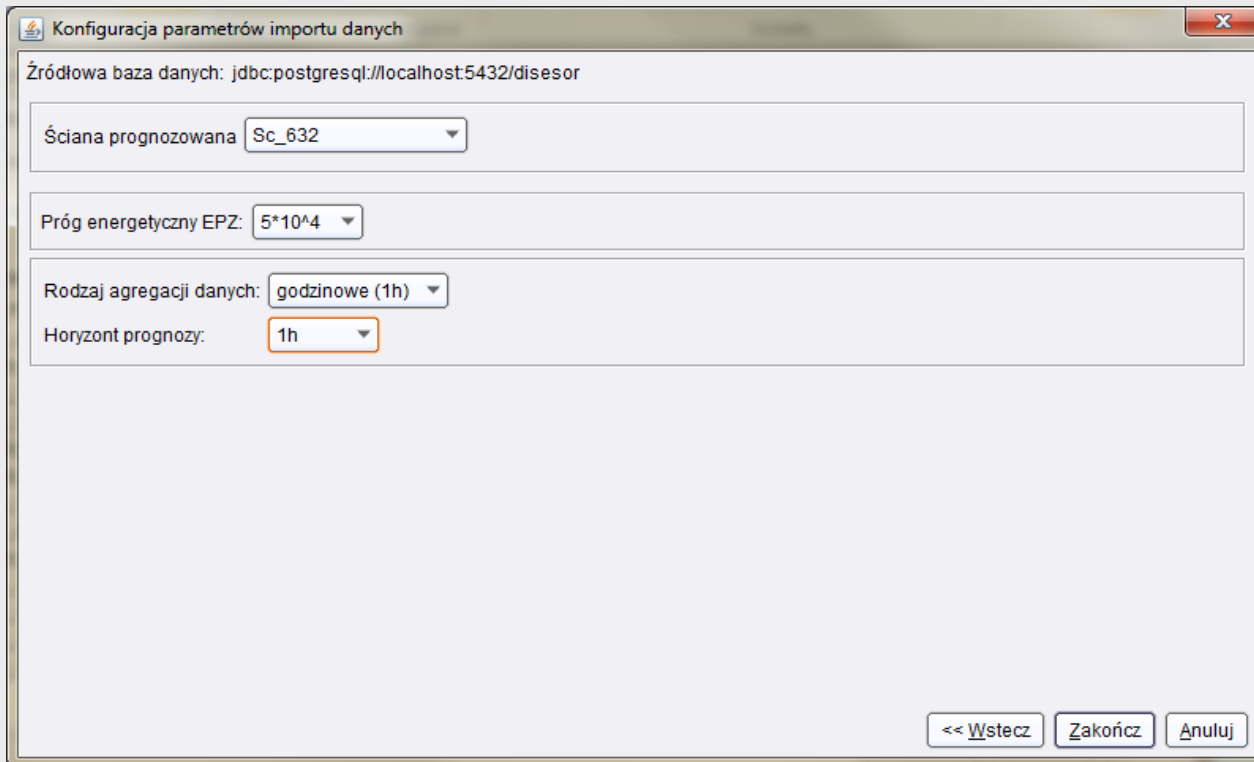
```
Call:  
randomForest(x = x, y = y, mtry = param$mtry, importance = TRUE)  
Type of random forest: classification  
Number of trees: 500  
No. of variables tried at each split: 1  
  
OOB estimate of error rate: 4.67%  
Confusion matrix:  
      Iris.setosa Iris.versicolor Iris.virginica class.error  
Iris.setosa      50           0           0         0.00  
Iris.versicolor   0          47           3         0.06  
Iris.virginica    0           4          46         0.08
```





Moduł analityczny

- Bloczki tematyczne (metan, sejsmologia)
 - metoda analizy jest określona (rezultaty badań i konkursów 😊)
 - wystarczy wybrać dane i sparametryzować zadanie



Konfiguracja parametrów importu danych

Źródłowa baza danych: jdbc:postgresql://localhost:5432/disesor

Ściana prognozowana: Sc_632

Próg energetyczny EPZ: $5 \cdot 10^4$

Rodzaj agregacji danych: godzinowe (1h)

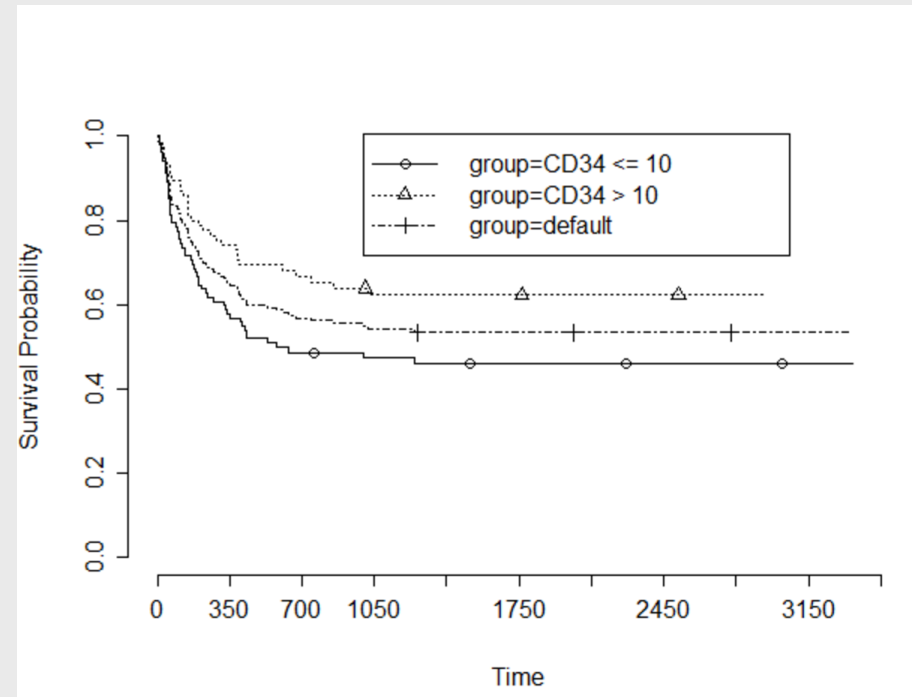
Horyzont prognozy: 1h

<< Wstecz Zakończ Anuluj

Moduł analityczny

- Nowe blozki analityczne
 - RapidRoughSet (MIMUW)
 - RuleInduction (MY 😊)
 - classification
(+ expert-driven)
 - regression
 - survival

CD34 $\leq$ 10,
CD34 $>$ 10



- **Problem otwarty:** inkrementacyjna indukcja i ocena reguł