

Prezentacja i wizualizacja danych [223100-0421]

Prowadzący: dr Mariusz Rafało – <http://mariuszrafalo.pl>

Plan zajęć

1. Wstępna analiza i eksploracja danych oraz wprowadzenie do *RMarkdown*
2. Pakiet *graphics*
3. Warsztat (praca z danymi)
4. Regresja liniowa
5. Drzewa decyzyjne
6. Analiza jakości modeli i porównywanie modeli
7. Analizowanie sekwencji zdarzeń
8. Pakiet *ggplot2*
9. Analizowanie danych geograficznych
10. Inetraktywna analiza danych (R Shiny)

Literatura

1. N. Yau, Data points: visualization that means something, Indianapolis, Ind. Wiley, 2013;
2. N. Yau, Visualize this the FlowingData guide to design, visualization, and statistics, Indianapolis, Ind. Wiley 2011;
3. J. Maindonald, Data analysis and graphics using R': an example-based approach, Cambridge UK, New York: Cambridge University Press, 2003.

Zasady zaliczenia przedmiotu

Zaliczenie opiera się na dwóch elementach:

Praca podczas zajęć: każde 2 nieobecności skutkują obniżeniem oceny o 1.

Zasady oceniania projektu: zgodnie z poniższą tabelą:

Od	Do	Ocena
0	59	niedostateczny
60	67	dostateczny
68	75	dostateczny plus
76	84	dobry
85	92	dobry plus
93	100	bardzo dobry

Projekt zaliczeniowy

Dane

- Dane pobieramy z ogólnodostępnego repozytorium zbiorów danych, przykładowo UCI: <http://mlr.cs.umass.edu/ml/datasets.html>
- Najlepiej gdyby dane dotyczyły działalności biznesowej, życia społecznego lub podobnych
- W oparciu o dane definiujemy problem, który chcemy zbadać, np.: analiza bankowych transakcji marketingu bezpośredniego, analiza uwarunkowań zarobków pracowników w różnych krajach, itp.

Analiza eksploracyjna

- Dla postawionego problemu, przeprowadzenie (wraz z opisem) analizy wstępnej interesujących nas zmiennych. W szczególności powinny się znaleźć takie narzędzia jak:
 - Histogramy/wykresy gęstości
 - Zależności pomiędzy zmiennymi
 - Rozkłady zmiennych w czasie/przestrzeni
 - Grupowania zmiennych względem kategorii

Analiza zaawansowana

- Przeprowadzenie bardziej złożonej analizy, z wykorzystaniem (przynajmniej) jednej z technik:
 - Regresja liniowa/logistyczna
 - Drzewo decyzyjne
 - Segmentacja

Raport lub aplikacja

- Analiza może zostać przygotowana w RMarkdownu lub w Shiny.
- Raport RMarkdown wysyłamy mailem (plik rmd i plik html). Struktura dokumentu:
 - Źródło danych (opis, informacje o źródle)
 - Określenie problemu badawczego (stawiane pytania)
 - Analiza wstępna danych (eksploracja)
 - Analiza służąca realizacji celu (odpowiedzi na postawione pytania)
 - Podsumowanie
- Aplikację w Shiny należy opublikować na platformie <https://www.shinyapps.io/>