

Prezentacja i wizualizacja danych [223100-0421]

dr Mariusz Rafało

<http://mariuszrafalo.pl>

mrafalo@sgh.waw.pl

Zasady zaliczenia przedmiotu

Zasady ogólne

Zaliczenie opiera się na dwóch elementach:

- Indywidualny projekt zaliczeniowy - należy zaliczyć na ocenę min. dostateczną
- Praca podczas zajęć: każda nieobecność skutkuje obniżeniem całkowitej oceny o 0.5 (pół oceny).

Projekt zaliczeniowy

Informacje ogólne

Należy przygotować projekt w technologii RMarkdowny lub R Shiny. Dokument RMarkdown eksportujemy do formatu HTML zaś aplikację Shiny publikujemy na <https://www.shinyapps.io/>. Celem raportu/aplikacji jest dokonanie analizy i wizualizacji danych zapomocą technik poznanych na zajęciach. Na wstępie określamy cel analizy a następnie dokumentujemy i wizualizujemy poszczególne kroki. W pracy koncentrujemy się na wizualizacji i poprawnej analizie danych. Pracujemy w języku *R* pracę dokumentujemy w *Markdown*.

Dane

- Dane pobieramy z ogólnodostępnego repozytorium zbiorów danych, przykładowo Kaggle: <https://www.kaggle.com/datasets>

- 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.

Przykładowa struktura raportu

1. Omówienie danych i problem badawczy

Wskazanie źródła danych, opis zmiennych oraz ogólnej problematyki. Wskazanie celu pracy.

2. 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

3. Analiza zaawansowana

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

4. Wnioski i podsumowanie

Opisanie wyników analizy, główne wnioski, ograniczenia analizy.

Finalny raport lub aplikacja

- Analiza może zostać przygotowana w *RMarkdown*, *Databricks* lub w *R Shiny*.
- Zaliczenie w formie RMarkdown wysyłamy mailem (plik RMD i plik HTML).
W przypadku technologii Databricks, wysyłamy tylko dokument HTML.
- Zaliczenie w formie aplikacj Shiny: aplikację w Shiny należy opublikować na platformie <https://www.shinyapps.io/> i przesłać link do aplikacji oraz kod źródłowy w R.