

# Wybrane zagadnienia *data mining*

Propozycja seminarium licencjackiego dla studentów kierunku  
Modelowanie matematyczne i analiza danych

dr Marta Frankowska

*Data mining*, czyli eksploracja danych jest relatywnie nową dziedziną matematyki i informatyki, prężnie rozwijającą się w odpowiedzi na potrzebę automatycznej i zaawansowanej analizy dużych zbiorów danych. W ostatnich latach obserwujemy dynamiczny rozwój metod i algorytmów, w szczególności metod *machine learning* oraz *sztucznej inteligencji*.

Podczas seminarium ugruntujemy swoją wiedzę oraz wzbogacimy umiejętności przetwarzania danych oraz przyjrzymy się bliżej zagadnieniu wzbogacania danych oraz uzupełniania braków danych. Ten etap badania data science jest najbardziej twórczy, a od jego wyników zależą wyniki całego eksperymentu, w szczególności dokładność i wiarygodność modeli predykcyjnych. Omówimy również podstawowe zagadnienia *machine learning* takie jak: klasyfikacja, grupowanie oraz regresja. Zapoznamy się bliżej z nieliniowymi modelami regresji oraz odpowiemy na pytanie: jak zastosować sieci neuronowe do regresji.

**Wymagania:** podstawowa wiedza ze statystyki i rachunku prawdopodobieństwa, znajomość **podstaw** języka R i/lub Python oraz podstawowa znajomość języka angielskiego (ze względu na literaturę).

# Bibliografia

- [1] Tadeusz Morzy, *Eksploracja danych. Metody i algorytmy*, Wydawnictwo Naukowe PWN, Warszawa 2013.
- [2] Marcin Szeliga, *Data Science i uczenie maszynowe*, Wydawnictwo Naukowe PWN, Warszawa 2017.
- [3] Jacek Koronacki, Jan Ćwik, *Statystyczne systemy uczące się*, Akademicka Oficyna Wydawnicza EXIT, Warszawa 2008.
- [4] Sanjiv Ranjan Das, *Data Science: Theories, Models, Algorithms, and Analytics*, <https://srdas.github.io/MLBook/>
- [5] Andreas Ruckstuhl, *Introduction to Nonlinear Regression*, [https://www.researchgate.net/publication/228785279\\_Introduction\\_to\\_Nonlinear\\_Regression](https://www.researchgate.net/publication/228785279_Introduction_to_Nonlinear_Regression).