

Seminarium licencjackie

## Metody probabilistyczne w analizie danych

dr Paweł Klinga

Seminarium przeznaczone jest dla studentów kierunku Modelowanie matematyczne i analiza danych.

Analiza danych ze względu na swoją naturę jest nauką interdyscyplinarną. Nie sposób jednak uprawiać ją bez odpowiedniego przygotowania w obszarze rachunku prawdopodobieństwa i statystyki. Analiza danych często używa narzędzi statystycznych w celu przewidywania lub badania trendów, podczas gdy narzędzia statystyczne oparte są na rozkładach prawdopodobieństwa zbiorów danych. Tym samym znajomość rozkładów prawdopodobieństwa jest ważna przy efektywnej pracy nad problematyką analizy danych.

Podczas seminarium przyjrzymy się takim tematom jak: ważne rozkłady prawdopodobieństwa (zarówno zmiennych losowych dyskretnych jak i ciągłych), znaczenie wartości oczekiwanej, błędu standardowego i innych parametrów statystycznych (wraz z obliczaniem ich w R), symulacja Monte Carlo, sieci bayesowskie, łańcuchy Markowa, Centralne Twierdzenie Graniczne (mawia się, że CTG to jedyne twierdzenie, które analitycy danych muszą znać) i inne.

**Wymagania:** podstawowa wiedza na temat rachunku prawdopodobieństwa i statystyki, a także znajomość języka angielskiego (ze względu na literaturę/materiały).

### Literatura:

1. Sanjiv Ranjan Das, *Data Science: Theories, Models, Algorithms, and Analytics*.
2. Carlos Fernandez-Granda, *Probability and Statistics for Data Science*