

Modelowanie statystyczne w zarządzaniu wierzytelnościami masowymi

Opis kursu:

W trakcie kursu zaprezentowane zostaną statystyczne metody budowania modeli prognostycznych. Omawiane będą główne etapy tworzenia modeli statystycznych zaczynając od zdefiniowania biznesowej potrzeby budowania modelu (celu) poprzez przygotowanie danych (przygotowanie zmiennych, czyszczenie, uzupełnianie) do wyboru konkretnego modelu oraz weryfikacji poprawności modelowania. Wszelkie przykłady oraz zadania będą wykonywane na rzeczywistych danych opisujących zachowania klientów („prosto” z bazy Kruk SA).

Zajmować się będziemy modelami klasyfikacyjnymi (modele bayessowskie, drzewa losowe) oraz regresyjnymi (modele regresji liniowej, wielokrotnej, logistycznej).

Na całość kursu składają się wykład (2 godziny tygodniowo) oraz zajęcia laboratoryjne (2 godziny tygodniowo). Zajęcia będą prowadzone przez:

- a) Wykład: Grzegorz Chłapiński (Ekspert ds. Rozwoju Metod Statystycznych w Kruk SA)
- b) Laboratorium: Bartłomiej Karaban (Analityk Danych w Kruk SA), Piotr Michalski (Ekspert ds. Rozwoju Metod Statystycznych w Kruk SA)

Przykłady na wykładzie oraz zadania laboratoryjne będą wykonywane z wykorzystaniem pakietu `data.table` (pakiet pozwalający na efektywną/szybką pracę z dużymi wolumenami danych).

<https://cran.r-project.org/web/packages/data.table/data.table.pdf>

Wymagania:

Od uczestników kursu będzie wymagana podstawowa wiedza z rachunku prawdopodobieństwa oraz statystyki matematycznej. Należy: rozumieć pojęcia zmiennej losowej i rozkładu prawdopodobieństwa, obciążenia i efektywności estymatora; wiedzieć czym są miary położenia, miary rozproszenia, kwantyle; znać metody estymacji częstościowej, przedziałowej, najmniejszych kwadratów, największej wiarygodności.

Na zajęciach (wykładach oraz laboratoriach) będzie wykorzystywany pakiet R. Tym samym od uczestników kursu wymagana będzie podstawowa znajomość R, gdzie przez podstawową znajomość rozumiemy:

- a) Znajomość podstawowych struktur danych: wektory, macierze, listy...
- b) Znajomość instrukcji sterujących: `if`, `for`, `while`...
- c) Umiejętność wizualizacji danych (np. `plot`)
- d) Umiejętność pisania własnych funkcji/skryptów
- e) Umiejętność używania dokumentacji
- f) Umiejętność instalacji pakietów

Jeżeli potrafisz napisać „cross join” dwóch tabel (wynikowa tabela powinna zawierać iloczyn kartezyjański wierszy obu tabel), to na pewno sobie poradzisz. Zwłaszcza jeśli potrafisz to zrobić w jednej linijce.

Zapraszamy

Kruk SA