

Program szkolenia:

Analiza wielowymiarowa

Informacje:

Nazwa:	Analiza wielowymiarowa
Kod:	data-science-multi
Kategoria:	Analiza danych developerzy architekci analitycy
Grupa docelowa:	
Czas trwania:	2 dni
Forma:	40% wykłady / 60% warsztaty

Szkolenie przeznaczone dla odbiorców, którzy pragną zrozumieć i poczuć analizę wielowymiarową. Celem szkolenia jest świadome dobieranie modeli i reprezentacji. Uczestnicy będą w stanie stosować sprawdzone i wydajne narzędzia do analizy wielowymiarowej.

Analiza wielowymiarowa to zbiór metod i technik analizy danych zawierających informacje o wielu obiektach opisanych jednocześnie za pomocą wielu zmiennych. Jej celem jest redukcja dużego zbioru danych, uproszczenie ich struktury oraz zapewnienie przejrzystej interpretacji wyników. Analizę wielowymiarową można również zastosować do sortowania, grupowania, skupiania obiektów wykazujących podobne cechy. Dodatkowo za jej pomocą można badać zależności pomiędzy zmiennymi, ich siłę powiązań oraz wyciągać wnioski.

Zalety szkolenia:

- realne przykłady
- zbiory danych rzędu 10^{18}
- dobór modeli i reprezentacji pod kątem wydajności obliczeń

Szczegółowy program:

1. Typy zjawisk i adekwatne skale pomiaru zjawisk.

1.1. rodzaje zależności statystycznych

1.2. adekwatne współczynniki korelacyjne

1.3. metody obliczania

1.4. oprogramowanie i przykłady obliczeń wielkoskalowych

1.4.1. przegląd programowania

1.4.2. przykłady programów komercyjnych i bezpłatnych

1.4.3. przykłady obliczeń w programach okienkowych i obliczenia w konsoli

1.5. Macierze rzadkie jako przykład zapisu i przetwarzania danych wielkoskalowych (rzędu 10^{18} + rekordów)

2. Wielowymiarowa zmienna losowa

2.1. definicja i interpretacje

2.2. przykłady na rzeczywistych danych rynkowych i pomiarowych

2.3. typy danych

2.4. sposoby zapisu danych dla dużych i małych zbiorów danych

3. Graficzne metody prezentacji danych

3.1. przykłady oprogramowania

3.2. przykłady dla dużych i małych zbiorów danych

3.3. metody prezentacji graficznej danych dla wielkich zbiorów danych

3.4. praca z bibliotekami graficznymi na danych rynkowych i pomiarowych

4. Redukcja danych

4.1. analiza Głównych Składowych

4.1.1. krok po kroku - obliczanie wartości własnych i wektorów własnych oraz zredukowanej macierzy danych

4.2. pojęcie wariancji i jej interpretacja w zredukowanych przestrzeniach wymiarów latentnych

5. Regresja wieloraka

5.1. zależności pozorne i rzeczywiste

5.2. zmienne pierwotne i wtórne

5.3. kierunki zależności między zmiennymi

5.4. obliczanie współczynników równania regresji wielorakiej

5.5. przewidywanie zachowań nowych jednostek obserwacji