BO·TT·EGA
IT minds

Program szkolenia:

Analiza czynnikowa i analizy wielowymiarowe w środowisku Julia, segmentacja dużych liczb zmiennych.

Informacje:

Nazwa:	**Analiza czynnikowa i analizy wielowymiarowe w środowisku Julia, segmentacja dużych liczb zmiennych.**
Kod:	**julia-analysis**
Kategoria:	Julia
Odbiorcy:	architekci, analitycy, developerzy
Czas trwania:	2 dni
Forma:	40% wykłady, 60% warsztaty

Spojrzenie na dane w nowy, uniwersalny sposób, który jest wejściem do dosłownie wielowymiarowego świata. Metoda ta pozwoliła w XX wieku fizykom na przekroczenie granicy poznania świata kwantowego, jest również podstawą w rodzącym się świecie big data.

Redukcja danych i odkrywanie ukrytych zależności miedzy dużą liczbą zmiennych jest kluczowym problemem w big data. Środowisko Julia posiada bardzo wydajne biblioteki pozwalające przetwarzać wielkie macierze. Szkolenie ma na celu naukę, krok po kroku, poprzez przeprowadzenie i pokazanie całego procesu oraz jego zrozumienie i pełną interpretację wyników.

Szkolenie zostanie przeprowadzone na danych o reakcjach użytkowników na reklamy w sieci. Szkolenie może być przeprowadzone na danych klienta lub innych danych np. z katalogu klasycznych danych używanych w ośrodkach akademickich.

Szkolenie krok po kroku przedstawia proces Analizy Głównych Składowych oraz praktyczne wykorzystanie automatycznego zrównoleglenia obliczeń dużych macierzy danych. Jest także warsztatem z wykorzystania Julii, jednego z najwydajniejszych programów do analizy numerycznej. Uczestniczy szkolenia będą potrafili przeprowadzić cały proces w każdym innym środowisku programistycznym. Szkolenie obejmuje również dogłębną praktyczną interpretację procesu i wyników analizy i transformacji.

Zalety szkolenia:

- Dobór modelu do klasy problemu
- Zrównoleglenie obliczeń dużych zbiorów danych
- Realne przykłady

BO·TT·EGA
IT minds

Szczegółowy program:

1. Wprowadzenie do Języka i środowiska programistycznego Julia

1.1. Instalacja środowiska Julia

1.2. Instalacja pakietów

1.3. Czytanie i zapis danych, wykorzystanie formatów danych adekwatnych do postaci danych roboczych

1.4. Transformacja danych

1.5. Pętle i funkcje w Julii, makra

1.6. Dane rzadkie (Sprase arrays), skuteczna metoda eliminowania zer z danych

2. Wprowadzenie do analizy danych

2.1. Analiza zależność miedzy zmiennymi

2.2. Miary zależności

2.3. Wariancja i jej znaczenie w opisie zbiorów

2.4. Kowariancja a korelacja, praktyczne skutki użycia i interpretacja.

2.5. Skalowanie i standaryzacja, ich wpływ na interpretację wyników

3. Analiza głównych składowych i ich obliczanie

3.1. Obliczanie macierzy korelacji i kowariancji dla dużych zbiorów danych

3.2. Pojęcia wartości i wektorów własnych i ich obliczanie

3.3. Kryteria wyboru liczby zredukowanych wymiarów

3.4. Wyliczanie Głównych Składowych i budowanie nowego modelu danych

3.5. Interpretacja modelu opisanego w układzie Głównych Składowych

3.6. Analiza związków statystycznych Głównych Składowych z danymi pierwotnymi

3.7. Interpretacja i opis danych pierwotnych przy użyciu nowego modelu

4. Przegląd możliwość analizy i klasyfikacji przy użyciu innych rotowanych układów odniesienia.

5. Analiza Głównych Składowych na tle innych metod

BO·TT·EGA
IT minds

5.1. Podstawowa, kanoniczna metoda przekształcenia dużych ilości danych

5.2. Łatwość interpretacji

5.3. Zmniejszenie ilości danych (przy zachowaniu w nich prawie pełnej informacji)

5.4. Zastosowanie w procesach big data oraz tam gdzie krytycznym czynnikiem jest czas.