Kod szkolenia: **XAI**

Tytuł szkolenia: **Interpretowalne uczenie maszynowe (XAI)**

Dni: 2

Opis:

Adresaci szkolenia

Szkolenie przeznaczone jest dla analityków danych, którzy czują potrzebę poznania metod oraz narzędzi wspomagających interpretację modeli uczenia maszynowego. Na początku warsztatu przypomnimy zagadnienia dotyczące złożonych modeli predykcyjnych. Uczestnicy będą mogli zapoznać się zarówno z teoretycznymi, jak i praktycznymi aspektami interpretacji modeli. Do przyswojenia praktycznej strony kursu konieczna będzie znajomość podstaw programowania w języku R, strona teoretyczna wymagać będzie podstawowej orientacji w tematyce modeli predykcyjnych.

Cel szkolenia

Uczestnicy szkolenia zapoznają się z najbardziej aktualnymi metodami interpretacji modeli uczenia maszynowego. Skupimy się na technikach pozwalających na zrozumienie dowolnego modelu (w szczególności tzw. czarnych skrzynek, czyli modeli trudnych w interpretacji). Zaprezentowane narzędzia pozwalają na identyfikację kluczowych czynników wpływających na predykcję modelu, lepsze zrozumienie jego globalnej struktury oraz lokalnego zachowania. Przedstawimy implementacje metod i narzędzi dostępnych w języku R. Przedstawione rozwiązania zastosować będą mogli również analitycy pracujący w języku Python.

Mocne strrony szkolenia

Szkolenie prowadzone jest przez osoby na co dzień zajmujące się analizą danych, które nie tylko korzystają, ale również tworzą narzędzia służące do wyjaśniania modeli uczenia maszynowego. Szkolenie składa się z kilku części dotyczących różnych aspektów wyjaśniania modeli uczenia maszynowego. Każda z nich podzielona jest na blok teoretyczny oraz praktyczny.

Wymagania

Umiejętność programowania w języku R, podstawowa umiejętność modelowania predykcyjnego

Parametry szkolenia

2 * 8 godzin (7 godzin netto) wykładów połączonych z warsztatami i ćwiczeniami.

sages

Program szkolenia:

1. Problemy uczenia maszynowego
    - Podstawowe pojęcia
    - Przegląd modeli uczenia maszynowego i specyficznych dla nich metod interpretacji
2. Wprowadzenie do zagadnienia wyjaśniania modeli
    - Czym jest interpretowalność?
    - Podstawowe typy metod interpretacji
    - Problemy z interpretowalnością modeli
    - Korzyści płynące z interpretacji modeli uczenia maszynowego
3. Globalne metody wyjaśniania
    - Audyt i diagnostyka modelu
    - Jakość modelu (performance)
    - Ważność zmiennych
    - Wpływ pojedynczej zmiennej na odpowiedź modelu
        - Individual Conditional Expectation Plots
        - Partial Dependence Plots
        - Accumulated Local Effects Plots
        - Merging Path Plots
4. Lokalne metody wyjaśniania modelu
    - Dekompozycja predykcji
        - Break Down Plots
        - Shapley Values
        - Inne metody dekompozycji
    - Lokalne przybliżenie modelu
        - LIME
        - live
        - Local Surrogates
    - Wykresy typu what-if
        - Ceteris Paribus Plots
        - Wangkardu Plots
5. Podsumowanie
    - Alternatywne metody
    - Alternatywne implementacje, w tym w języku Python

Sages sp. z o.o., ul. Nowogrodzka 62c, 02-002 Warszawa
tel. +48 22 203 56 00 fax: +48 22 203 56 01
e-mail: biuro@sages.com.pl web: www.sages.com.pl