

# Surrogate Assisted Feature Extraction

*Aleksandra Gacek, Piotr Luboń*

*26/10/2018*

## Cel projektu

Celem projektu jest utworzenie biblioteki umożliwiającej ekstrakcję zmiennych przy pomocy modelu pomocniczego, używając partial dependency plot.

## Opis projektu

Aplikacja zostanie zaimplementowana w języku Python. Efektem pracy będzie moduł pythonowy, z możliwością zaimportowania go jako biblioteki.

Udostępni on klasę pozwalającą na ekstrakcję cech. Klasa ta będzie implementować interfejs Transformer-Mixin, biblioteki scikit-learn. Będzie to wymagało implementacji metod fit oraz transform. Interfejs metody fit pozwoli na przekazanie następujących argumentów:

- macierz zmiennych
- wektor zmiennej celu
- dopasowany model

Interfejs metody transform przyjmować będzie macierz zmiennych, a zwróci nowe, już przekształcone zmienne.

Implementacja w języku Python umożliwi przekazanie jako argument dowolnego modelu z biblioteki scikit-learn. Klasa będzie implementowała metody fit oraz transform, zatem będzie możliwe jej użycie jako elementu Pipeline biblioteki scikit-learn.

Dodatkowo w klasie będzie przechowywana informacja na temat tego, jak uzyskane zostały nowe zmienne.

Bibliotekę będzie można uruchomić na wszystkich systemach z zainstalowanym językiem Python.

## Opis metody

Dla każdej zmiennej zostaną wyznaczony partial dependency plot, poprzez podawanie do modelu danych wejściowych z uśrednionymi pozostałymi zmiennymi i szukaną zmieniającą się o pewną małą wartość. Nowe zmienne zostaną utworzone na podstawie podziału krzywej tego wykresu według punktów zmiany.