

Project 1

Anna Niżnik & Piotr Krzeszewski

October 9, 2017

Wynik

Otrzymany wynik to **81.8%** uzyskany za pomocą klasyfikatora XG-Boost.

Metodologia

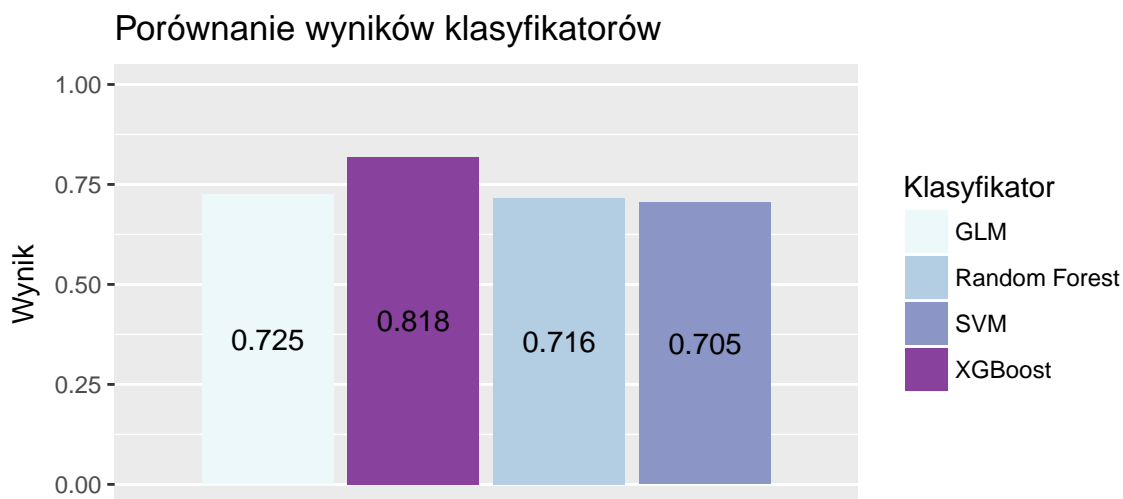
Proces znalezienia optymalnego rozwiązania składał się z dwóch etapów.

Najpierw została przeprowadzona selekcja zmiennych, która pozwoliła na ograniczenie wymiarów przestrzeni danych oraz zmniejszenie szumu wynikającego z istnienia nieznaczących zmiennych. W tym celu wykorzystano kryterium BIC, które zwróciło następujące zmienne W1, U2, E1, P2, Q1, T2, M1.

Drugim krokiem analizy było wybranie klasyfikatorów, które wykorzystano budowy testowanych modeli.

Wypróbowano następujące metody:

- **GLM** - regresja logistyczna,
- **XG-Boost**,
- **lasy losowe**,
- **SVM** - klasyfikacja przy pomocy maszyny wektorów nośnych.



Poprawność wyniku

Dane treningowe zostały podzielone na dwie części w stosunku 70/30. Liczniejsza część została wykorzystana do trenowania modeli, które następnie testowano na pozostałych 30% danych.

Z klasyfikowanych wyników wybrano 20% rekordów, które otrzymały najwyższe prawdopodobieństwo uzyskania “klasy +”. Skuteczność klasyfikatora została zdefiniowana jako procent wybranych rekordów, które rzeczywiście należały do “klasy +”. W ten sposób otrzymano zaprezentowany powyżej wynik: **81.8%**.