

Inżynieria Uczenia Maszynowego

Etap 1

Michał Matak, Jakub Robaczewski

Polecenie od klienta:

“Mamy co prawda dodatkowe benefity dla naszych najlepszych klientów, ale może dałoby się ustalić kto potencjalnie jest skłonny wydawać u nas więcej?”

Zadanie biznesowe

Sugerowanie klientów, którzy mogą wrócić do serwisu.

Zadanie modelowania:

Model regresyjny, szacujący prawdopodobieństwo powrotu klienta do serwisu.

Założenia:

- Mamy dostęp do danych dotyczących:
 - przebiegu sesji użytkowników na serwerze
 - dostaw
 - produktów
 - użytkowników
- Klient poruszając się po stronie internetowej generuje pewne zdarzenia zapisane m.in. w logach dotyczących sesji
- Użytkownicy powracający do naszego sklepu posiadają pewne charakterystyczne cechy, które wskazują, że dla nich prawdopodobieństwo powrotu do sklepu jest większe

Testowanie:

Model będziemy testować poprzez wycięcie pewnego odcinka czasu z końca szeregu czasowego i potraktowanie go jako zbiór testowy

Kryteria sukcesu

- A) Efektywność na zbiorze testowym powyżej 50% (lepiej niż losowo)
- B) wartość krzywej ROC powyżej 0.5

Wymaganie techniczne

1. projekt realizowany w python-ie
2. wstępnie przetwarzanie batchowe

Dane:

Ewidentne błędy w danych (w samej tylko liście dotyczącej produktów):

- za duże ceny
- ceny ujemne
- ceny z wieloma miejscami po przecinku

Poinformowano o nich dostawcę, po czym otrzymano nowe dane z każdej z kategorii. Nie zawierają one tak ewidentnych błędów i nadają się do dalszego przetwarzania.