

Opis zadania

Cel zadania

Problem biznesowy

Na podstawie danych historycznych dotyczących podobnej, wcześniejszej komunikacji e-mail należy zbudować model predykcyjny, którego zadaniem będzie „określanie” komu w przyszłości wysłać newsletter tak, aby za miesiąc - dwa osoba, która będzie ten newsletter wysyłać, mogła z niego skorzystać i szybko dobrać grupę do wysyłki.

Zbiór

- target - zmienna celu, jest to zmienna binarna 0-1, sukces oznaczony jest jako 1 (kliknięcie w odnośniki zawarte w wiadomości),
- us_id - id obserwacji,
- pozostałe zmienne należy potraktować jako potencjalne zmienne objaśniające.

Wynik zadania

1. Plik z kodem

Prosimy o przesłanie pliku z kodem twojego rozwiązania (może to być plik .R, notebook jupyterowy lub plik .py) ogólnie w takiej postaci, żebyśmy stosunkowo łatwo mogli to odtworzyć u siebie. Pamiętaj, że jeżeli korzystasz z jakichś specyficznych bibliotek R-owych lub Pythonowych, to nie zapomnij, żeby w pliku z kodem była ich deklaracja, tak, żebyśmy wiedzieli jaki pakiet / biblioteka został wykorzystany i w razie potrzeby zainstalowali go u siebie.

W pliku z kodem, mogą też być polecenia do wykresów, metryk klasyfikacji które będą printować te informacje, czyli wszystko co będzie dla nas informacyjne i pokaże Twój tok myślenia o problemie.

Prośba by każdy znaczący krok krótko wykomentować co robi dany blok kodu plus opcjonalnie dlaczego takie podejście.

Zasadniczo zakładamy, że to co otrzymamy będzie w R lub Pythonie.

2. Wynikowy plik submission_form.csv

Dodatkowa rzecz do przesłania, to plik submission_form.csv, gdzie trzeba dla wybranych przez Ciebie us_id, dołożyć kolumnę z aktualnym targetem (zmienna target ze zbioru danych) plus z Twoją predykcją targetu (**jako prawdopodobieństwo**, a nie binarnie, wartość samej prognozy 0-1). To ile us_id zawrzesz w tym pliku zależy od Ciebie i jest to twoja decyzja, zasadniczo mają to być us_id dla zbioru testowego, który musisz sobie wygenerować z otrzymanych danych. Z naszej strony tylko prośba, aby pilnować formatu i wysłać nam to w takiej formie i nie pomylić kolumny us_id. Załączony plik submission_form.csv stanowi tylko przykład co do oczekiwanej formy tego pliku.

POWODZENIA! ☺