

Wytyczne do projektu nr 2

Temat: Praktyczne zastosowanie modeli oceny wartości relacji z klientem w czasie

Termin złożenia: 2022-06-11

Za projekt można zdobyć max 20 punktów.

Projekt realizowany w zespołach 6 osobowych (skład może być jak przy projekcie nr 1).

Złożenie pracy po terminie będzie wiązało się z ujemnymi punktami.

Wymagania edytorskie:

- standard pracy dyplomowej (czcionka Times New Roman, 12 pkt, interlinia 1,5 itd.),
- na stronie tytułowej powinien znaleźć się rozpisany skład osobowy grupy projektowej, ze wskazaniem osoby kontaktowej,
- w pracy należy następnie wskazać część, za którą odpowiada każdy członek zespołu,
- limit do 15 stron (w tym tabele, wykresy + jednostronicowy kod programu z opisem, dotyczący kluczowych procedur).

Należy sporządzić dokumentację elektroniczną, na którą składają się:

- praca projektowa w formacie PDF,
- zbiór danych w formacie SAS,
- program SAS z pełnym kodem.

Dokumentacja elektroniczna powinna zostać przekazana poprzez zadanie w kanale zespołu MS Teams w dniu oddania projektu w trakcie zajęć. Dokumentacja powinna mieć format .zip i odpowiednią nazwę jak przy projekcie nr 1, np. nm12345

Punktacja:

6 pkt. - struktura i analiza właściwa

4 pkt. - jakość pracy (tabele, wykresy, zachowanie standardów edytorskich)

5 pkt. - kodu programu: poprawność i czytelność

5 pkt. - elementy własne, dodatkowe

łącznie 20 pkt

Struktura projektu (raportu):

- Wprowadzenie
- Krótkie omówienie tematyki podjętej w pracy + przedstawienie pytań badawczych
- Analiza struktury danych, ich przedstawienie, wyniki analizy eksploracyjnej
- Analiza właściwa (dotycząca modelu segmentacji lub retencji, szczegóły poniżej)
- Podsumowanie
- Kod programu z opisem, obejmujący najistotniejsze procedury

WERSJA I - MODEL SEGMENTACJI

Elementy wymagane:

1. Transformacja zmiennych przed budową modelu
2. Wygenerowanie wielokrotne rozwiązań klastrowych dla różnych ziaren początkowych
1. oraz wybór najlepszego na podstawie uzyskanych wyników
2. Szczegółowa interpretacja wyników dla najlepszego modelu segmentacji + przedstawienie wizualne klastrów (jak na zajęciach - rozkłady zmiennych w klastrach itd. + interpretacja)
3. Wnioski biznesowe, które można wyciągnąć z wyników modelu. Propozycje działań, które można podjąć w oparciu o wyróżnione segmenty

Przykłady wkładu własnego:

- opisanie przyjętej procedury wyboru zmiennych do modelu (jeśli liczba zmiennych w modelu jest mniejsza niż w zbiorze danych)
- optymalizacja liczby klastrów w modelu (propozycja metody wyboru optymalnej liczby klastrów w modelu)
- inny sposób przedstawienia klastrów niż było pokazane na zajęciach
- inny typ modelu segmentacyjnego niż metoda K-średnich

WERSJA II - MODELE RETENCJI

Elementy wymagane:

1. Budowa prostego modelu retencji (w tym oszacowanie stopy retencji, średniego czasu do odejścia $E(T)$, medianę czasu do odejścia, oczekiwaną wartość CLV przy założeniu płatności „z góry” oraz „z dołu”, sporządzenie wykresu PMF, funkcji przeżycia + odpowiednia interpretacja/opisanie wszystkich uzyskanych wyników)
2. Budowa uogólnionego modelu retencji z wykorzystaniem metody Kaplana-Meiera (przedstawienie wyników oraz interpretacja)
3. Porównanie oraz omówienie różnicy oszacowań uzyskanych dwoma powyższymi modelami
4. Budowa uogólnionego modelu retencji z uwzględnieniem zmiennej stratyfikującej (przedstawienie wyników oraz interpretacja)

Przykłady wkładu własnego:

- użycie innego modelu do estymacji uogólnionego modelu retencji niż model Kaplana-Meiera (np. model dla czasu dyskretnego oparty na regresji logistycznej)
- uwzględnienie w analizie cech zmiennych w czasie
- uwzględnienie w analizie również innego rodzaju ocenowania poza prawostronnym
- opisanie próby dopasowania konkretnego rozkładu do rozkładu czasu do odejścia uzyskanego na danych (można wykorzystać modele parametryczne ANZ – proc lifereg)

WAŻNE! Jako wkład własny definiujemy to co jest zawarte w projekcie, a nie było przedstawiane podczas zajęć i stanowi Państwa dodatkową pracę nad projektem.

Przykładowe zbiory danych:

<https://bigml.com/gallery/datasets>

<https://www.kaggle.com/>

WAŻNE! Dane muszą dotyczyć zagadnienia biznesowego.