

Zadanie Pre-selekcyjne

Wszyscy mówią o social mediach, analizują dostępne dane i wykorzystują je aby zyskać przewagę na rynku. Nie możemy być ostatni w tym wyścigu. Też chcemy z nich skorzystać, ale niestety nie wiemy jak... Dlatego potrzebujemy Twojej pomocy!

Chcemy zrozumieć specyfikę zgromadzonych danych oraz potencjalnych zależności między naszymi bankowymi danymi a danymi z social mediów. Wybraliśmy jedynie aktywnych klientów posiadających przynajmniej 1 aktywne konto w jednym z 4 najpopularniejszych portali społecznościowych.

Wśród danych bankowych mamy klasykę gatunku:

- 1) *id* [unikalny identyfikator klienta];
- 2) *data_start* [data zawiazania przyjaźni ☺ danego klienta z nami];
- 3) *depozyt* [1 – klient posiada rachunek depozytowy; 0 – klient nie posiada rachunku depozytowego];
- 4) *saldo* – [wartość salda na rachunku depozytowym w poprzednim miesiącu; w PLN];
- 5) *kredyt* [1 – klient posiada rachunek kredytowy; 0 – klient nie posiada rachunku kredytowego];
- 6) *typ_kredytu* [typ kredytu: 1 – karta kredytowa, 2 – limit w koncie, 3 – kredyt ratalny];
- 7) *wnioskowana_kwota* [wnioskowana kwota kredytu, w PLN] – „NA” oznacza brak wnioskowania;
- 8) *przyznana_kwota* [przyznana kwota kredytu, w PLN] – „NA” oznacza brak przyznanej kwoty kredytu;
- 9) *akcja* [1 – klient nabył ubezpieczenie na życie, 0 – nie nabył ubezpieczenia na życie].

A z social mediów udało nam się potajemnie pozyskać:

- 1) *SM_liczba_kont* [liczba kont na 4 najpopularniejszych portalach społecznościowych];
- 2) *SM_liczba_znajomych* [liczba unikalnych znajomych na 4 najpopularniejszych portalach społecznościowych];
- 3) *SM_liczba_tag* [liczba zdjęć z tagami miejsc poza granicami kraju; dane od założenia pierwszego konta na jednym z portali społecznościowych];
- 4) *SM_liczba_polubien* [średnia liczba polubień zdjęć publikowanych przez klienta; dane z ostatnich 12 miesięcy].

A teraz najważniejsze - nie mów nikomu, ale te nasze wyjątkowe, sekretne dane udostępniamy specjalnie dla Ciebie w załączonym pliku *dane.csv* abyś mogła/mógł nam pomóc lepiej zrozumieć specyfikę poszczególnych zmiennych i zależności między nimi:

- 1) Konkurencja (cichosza....!) twierdzi, że rozkład zmiennej *SM_liczba_znajomych* ma znaczenie. Dobrze by więc było zbadać jego charakterystykę, aby wskazać nam cechy rozkładu, którą mogą mieć istotne znaczenie w dalszych analizach (uwzględniających tę zmienną).
- 2) Tajemniczy Don Pedro, szpieg z krainy deszczowców donosi, że liczba zdjęć wykonanych za granicą może mieć może być powiązana z wysokość salda na rachunku depozytowym. Trzeba by więc zbadać taką zależność, niestety nie wiemy jaką miarę użyć do wykonania takiej analizy, więc musisz liczyć na siebie. Podobno warto

zastanowić nie tylko, czy istnieje współzależność między tymi zmiennymi, ale również jak jest ona silna? Chcielibyśmy też wiedzieć w jaki sposób uzasadnić taką zależność.

- 3) Czasami udaje nam się sprzedać ubezpieczenie na życie, ale niestety nie wiemy dlaczego. Potrzebujemy więc zrozumieć jakie cechy klienta determinują pozytywną odpowiedź na tę akcję, abyśmy mogli trochę więcej zarobić i Cię zatrudnić do naszego zespołu. Wszystko zależy więc od Ciebie – zaproponuj metodę analizy tego problemu oraz na podstawie uzyskanych wyników oceń jej poprawność, a my sprawdzimy ile da się na tym zarobić 😊

Kryteria oceny

Do konkursu mogą zgłaszać się **zespoły 1, 2 lub 3 osobowe**.

Wybór zespołów zakwalifikowanych do konkursu odbędzie się na podstawie **rozwiązań zadania preselekcyjnego** oraz oceny **listu motywacyjnego**.

Zadanie preselekcyjne składa się z **trzech analiz** do wykonania (trzech podpunktów). **Każda analiza** zostanie oceniona w **skali 0-5 punktów**. **Maksymalna liczba punktów** możliwa do zdobycia za zadanie preselekcyjne to **15 punktów**.

Ocena przyzna za rozwiązanie zadania preselekcyjnego będzie **jedna dla całego zespołu** i **nie będzie zależać o wielkości zgłoszonego zespołu**. W konkursie głównym, zespoły 3 osobowe będą występowały w składzie, w którym się zgłosiły. Zespoły 1 i 2 osobowe mogą zostać połączone w zespoły 3 osobowe, jednak nie ma możliwości by 2 osobowy zespół został rozdzielony.

Zespoły zostaną wybrane na podstawie **sumy punktów uzyskanych za rozwiązania** analiz przedstawionych w zadaniu (**punkty 1-3**). W sytuacji za dużej liczby zespołów, które uzyskały wynik równy wartości granicznej kwalifikacji, **dodatkowym kryterium** oceny będzie ocena przesłanego **listu motywacyjnego**.

Do każdej z 3 analiz opisanych w zadaniu preselekcyjnym (punkty 1-3) należy przygotować **osobną odpowiedź**, która powinna składać się **maksymalnie z 5-ciu punktów (Podsumowania rozwiązania zadania)**. Punkty te, powinny w sposób **syntetyczny** wskazywać **cechy opisujące rozwiązanie** (przykład poniżej).

Dodatkowo, każdy zespół powinien przygotować **maksymalnie 1 stronicowy** (~1800 znaków) **list motywacyjny**, w którym uzasadni dlaczego należy mu się miejsce w konkursie ING Banku Śląskiego. Zespoły mogą również przygotować krótki film prezentujący rozwiązanie punktu 3.

Poniżej przedstawiono **przykładowe Podsumowanie rozwiązania zadania**.

Przykładowe pytanie:

Analiza zależności między statusem zawodowym [2 grupy: tylko studiuje oraz pracuje (może studiować lub nie)] oraz typem kredytu. Jaką miarę należy użyć do wykonania takiej analizy? Czy istnieje współzależność między tymi zmiennymi, jeżeli tak to jaka? W jaki sposób taką zależność można uzasadnić?

Przykład Podsumowania rozwiązania zadania:

Podsumowanie rozwiązania zadania

1. Obie zmienne są nominalne, zmienna typ kredytu ma 3 poziomy, więc właściwą miarą asocjacji będzie **miara V Cramera**,
2. Wartość miary V Cramera wynosi **0,67** i wskazuje na **silną pozytywną asocjację** między zmiennymi
3. Asocjacja jest **istotna statystycznie** na poziomie istotności 5% – **p-value=0.003 (test G)**
4. Klienci pracujący częściej posiadają limity w koncie lub karty kredytowe, a tylko studiujący kredyty ratalne. Prawdopodobną przyczyną jest różnica w podstawowych celach kredytowania. Studiujący pożyczają na konkretne produkty, pracujący pożyczają w celu zarządzania ryzykiem płynności.