

PROJEKT Z PRZEDMIOTU EKSPLOACJA DANYCH

DRUGI ETAP: PRZYGOTOWANIE DANYCH + MODELOWANIE

SPEED DATING EXPERIMENT

13.06.2025 r.

Spis treści

1. Charakterystyka zbioru danych	2
1.1. Pochodzenie	2
1.2. Format	2
1.3. Liczba przykładów	2
1.4. Ilość zbiorów danych	2
2. Cel eksploracji i kryteria sukcesu	2
3. Dyskusja kroków dalszego postępowania	2
3.1. Założenia wstępne	2
3.2. Testowanie wyników	2
4. Przygotowanie danych	2
4.1. Dane brakujące i dane do ujednolicenia	2
4.2. Zamiana na nominalne/numeryczne	2
4.3. Podzbiór danych	2
5. Wyniki i model	2
5.1. Krótki opis modelu	2
5.2. Wyniki osiągnięte przez model	3
6. Optymalizacja modelu	3
7. Wnioski	3

1. Charakterystyka zbioru danych

1.1. Pochodzenie

<https://www.kaggle.com/datasets/annavictoria/speed-dating-experiment>

1.2. Format

.csv

1.3. Liczba przykładów

8378 rekordów

1.4. Ilość zbiorów danych

1

2. Cel eksploracji i kryteria sukcesu

Celem eksploracji danych ze zbioru „Speed Dating Experiment” jest znalezienie odpowiedzi na pytania:

- Czy ludzie potrafią dokładnie przewidzieć swoją postrzeganą wartość na rynku randkowym?
- Sprawdzenie, jaki atrybut najmocniej wpływa na dobór partnera przeciwnej płci.

Kryteria sukcesu, które zostaną przyjęte w celu oceny skuteczności eksploracji danych, obejmują:

- wysoka korelacja ($\geq 0,6$) między przewidywaną a rzeczywistą wartością uczestników na rynku randkowym
- zidentyfikowanie cech, które mają największy wpływ na postrzeganą wartość uczestników
- przeprowadzenie analizy istotności atrybutów ze wskazaniem najistotniejszego

3. Dyskusja kroków dalszego postępowania

3.1. Założenia wstępne

< opisać, jeśli są lub brak >

3.2. Testowanie wyników

< cośtam >

4. Przygotowanie danych

4.1. Dane brakujące i dane do ujednolicenia

< usunąć, jeśli nie występują >

4.2. Zamiana na nominalne/numeryczne

< usunąć, jeśli nie występuje >

4.3. Podzbiór danych

< krótki opis podzioru danych, który analizowaliśmy >

5. Wyniki i model

5.1. Krótki opis modelu

< do opisanie, uwzględnić parametry modelu >

5.2. Wyniki osiągnięte przez model

< wyniki, może być w formie tabeli >
< jeśli drzewo, wstawić obrazek drzewa >

6. Optymalizacja modelu

< opis zmian i optymalizacji: co, dlaczego, na ile ustawiono, inny podzbiór danych >
< dodatkowe podpunkty w miarę potrzeb >

7. Wnioski

< wnioski końcowe, osiągnięte wyniki, komentarz, jakie parametry były najlepsze >