

WWW14 Data science

Mateusz Sieniawski

15 maja 2018

Rozwiązania zadań proszę wysyłać na adres mateusz@sieniawski.net z tagiem [WWW14]. W rozwiązaniu należy umieścić odpowiedzi na pytania oraz kod źródłowy. Można pisać w dowolnym języku programowania, choć mocno preferowany jest Python.

Zadanie 1

Czy grałeś/gralaś w Hearthstona? Jeśli tak, opisz krótko swoje doświadczenia. Jeśli nie, zapoznaj się z zasadami gry.

Jakie talie wg Ciebie mają największe szanse na wygraną? Jakie czynniki mają najistotniejszy wpływ na to, czy talia jest dobra? Podane przez Ciebie czynniki posortuj wg istotności.

Zadanie 2

Pod adresem <https://bit.ly/2IgrX0G> znajdziesz logi pochodzące z wielu gier Herthstona rozegranych pomiędzy botami (archiwum z grami waży po rozpakowaniu 100Gb, dlatego dla wygody ich obróbki zostały umieszczone także w 20 osobnych mniejszych archiwach). Twoim zadaniem jest na podstawie tych danych odpowiedzieć na następujące pytania:

1. Ile decków jest w tych danych?
2. Który bot ma największy winrate? (winrate to stosunek liczby gier wygranych do wszystkich rozegranych gier)
3. Który deck ma największy winrate? Ilukrotnie był grany?
4. Która klasa bohatera ma największy winrate? Która najmniejszy? Czy jest to znaczące odchylenie od średniej?
5. Jakie 10 kart jest najczęściej granych? A które 10 ma największy winrate? Jakiego typu są to karty?
6. Ile procent wszystkich kart stanowią zaklęcia? A ile stwory?

Zadanie 3

Na podstawie tych samych danych co w poprzednim zadaniu, narysuj:

1. Rozkład winratów decków dla każdej klasy bohatera. Czy te winraty rozkładają się równomiernie?
2. Histogram, ile jest kart o danym koszcie.
3. Rozkład kosztów kart dla każdej klasy bohatera. Czy jakaś klasa widocznie używa kart o mniejszym/większym koszcie niż średnia? Czy do tych rozkładów można dopasować jakąś krzywą? Jeśli tak, określ parametry tej krzywej dla poszczególnych klas.

Zadanie 4

Założmy, że chcesz stworzyć funkcję oceniającą, jak dobry jest dany deck. Przyjmowałaby ona jako argument deck, a zwracałaby liczbę z zakresu $[0;1]$. Zaproponuj (nie implementuj jej), jak można by stworzyć taką funkcję.