

DRZEWA DECYZYJNE

LAB 1 – Eksploracja zbiorów danych

Zadanie 1. (2 pkt.)

Wczytaj dane pobrane z serwisu kaggle.com do obiektu DataFrame.

Link do zbioru danych CSV:

https://www.kaggle.com/shivamb/netflix-shows?select=netflix_titles.csv

Jako wynik działania programu wypisz następujące informacje:

1. Ilość wczytanych wierszy danych
2. Wymiar wczytanych danych (kolumny x indeksy)
3. Zlicz ilość wartości pustych (jeśli w są) w każdej z kolumn

Zadanie 2. (3 pkt.)

Wczytaj dane o nazwie titanic pobrane z biblioteki seaborn do obiektu DataFrame.

Na podstawie danych wykonaj serie poniższych operacji

1. Zlicz ile jest wartości pustych w zbiorze NaN
2. Zlicz ilość wartości pustych (null) w każdej kolumnie, a wynik zapisz w postaci sumy skumulowanej
3. Usuń te kolumny, jeśli takie istnieją dla których liczba wartości pustych jest większa niż 30% wielkości pobranego zbioru danych
4. Zamień dane katagoryczne w kolumnie sex {female,male} na wartości liczbowe {0,1}

Zadanie 3. (3 pkt.)

Na podstawie danych publikowanych w serwisie GitHub za pomocą GitHub API wykonaj następujące operacje:

1. Odszukaj profil o nazwie użytkownika: MikiKru
2. Oblicz statystykę języków programowania publikowanych w powyższym repozytorium, tj. jaki udział procentowy ma danych język programowania względem wszystkich repozytoriów użytkownika MikiKru
3. Wynik zaprezentuj ja pomocą wykresu kołowego

Zadanie 4. (2 pkt.)

Wykorzystując biblioteki Requests oraz BeautifulSoup pobierz informacje o temperaturze w dniu wykonywania żądania oraz prognozę temperatury na kolejny dzień (dzień / noc). Jako wynik należy zwrócić napis:

Data: 09.03.2021 10:10

Temperatura: 12 C

Jutro: 15C / 2C