

Ranking filmów

Celem ćwiczenia jest wykonanie modelu, który będzie oceniał czy recenzja jest pozytywna czy negatywna.

Dane:

Pliki positive.txt oraz negative.txt zawierają recenzje filmów.

Zadanie:

1. Wczytaj dane z plików

```
# Wczytaj dane z plików
with open('file.txt', 'r', encoding='utf-8') as file:
    positive_data = file.readlines()
```

2. Przydziel do nich klasę „class” (positive – 0, negative – 1)
3. Utwórz dataframe z nagłówkami „text” i „class”

```
# Przydziel klasy (0 - positive, 1 - negative)
dataframe_1= pd.DataFrame({'text': pozytywne_recenzje , 'class': 0})
dataframe_2 = pd.DataFrame({'text': negatywne_recenzje, 'class': 1})
# Połącz ramki danych
df = pd.concat([lista sklepanych df], ignore_index=True)
```

4. Zmieszaj (shuffle) zestaw danych
5. Przedstaw dane za pomocą df.head
6. Utwórz słownik „Bag od Words”
7. Pokaż, które słowa występują najczęściej dla całego zbioru
8. Pokaż, które słowa występują najczęściej dla zbioru negative a które dla positive
9. CDN