Zabrze, 21.05.2019r.

Politechnika Śląska w Gliwicach Wydział Automatyki, Elektroniki i Informatyki



**Biologically Inspired Artificial Intelligence**

Classification of the genre of the films based on the description of the plot

autorzy Aleksandra Musiał

Benedykt Waszkiewicz

prowadzący dr inż. Grzegorz Baron

rok akademicki 2018/2019

kierunek informatyka

rodzaj studiów SSI

semestr 6

termin spotkań środa, 8.30 - 10.00

grupa GKIO3

# Opis zestawu danych

## Kontekst

Opis fabuły zawarty na stronie Wikipedii.

## Zawartość zestawu danych

Zestaw danych zawiera opis 34 886 filmów z całego świata. Opisy kolumn są wymienione poniżej:

* Realease Year - rok wydania
* Title – tytuł filmu
* Origin/Ethnicity – pochodzenie filmu (American, Bollywood, etc.)
* Director – reżyser
* Cast – obsada
* Genre – gatunek(i) filmu
* Wiki Page – adres URL do strony Wikipedii
* Plot – długi opis fabuły filmu

# Wstępna analiza danych wejściowych

*Listing 2.1. Sprawdzenie liczb wierszy i kolumn*

|  |  |
| --- | --- |
| 1  2  3 | movies = pd.read\_csv('../input/wiki\_movie\_plots\_deduped.csv', ',')  nRow, nCol = movies.shape  **print**(f'There are {nRow} rows and {nCol} columns') |

*Wynik*

|  |  |
| --- | --- |
|  | There are 34886 rows and 8 columns |

*Listing 2.2. Sprawdzenie liczb różnych gatunków*

|  |  |
| --- | --- |
| 1  2 | 1. mmovies['Count'] = 1 2. **p** 3. **rprint**(movies[['Genre', 'Count']].groupby(['Genre']).count().shape[0]) |

*Wynik*

|  |  |
| --- | --- |
|  | 2265 |

W zbiorze danych znajduje się 2265 różnych gatunków filmów.

Wykorzystanie algorytmu klasyfikującego o dobrej dokładności dla tak wielkiej liczby klas nie jest możliwa. Po przejrzeniu zestawu danych można zauważyć, że wiele gatunków jest przypisanych do pojedynczych filmów.