MSI – NLP – raport

Autor:  
Adam Szreter  
296694

# Trening modelu na dostarczonym zbiorze danych

## Model ROBERTA

|  |  |  |  |
| --- | --- | --- | --- |
| Liczba epok treningu | Rozmiar batcha | Accuracy | eval\_loss |
| 5 | 8 | 0.95 | 0.302 |

## Klasyfikator bayesowski

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | precision | recall | f1-score | support |
| 0 | 0.91 | 0.92 | 0.92 | 117 |
| 1 | 0.93 | 0.92 | 0.93 | 139 |

Total accuracy: 0.92

# Modyfikacja hiperparametrów

|  |  |  |  |
| --- | --- | --- | --- |
| Liczba epok treningu | Rozmiar batcha | Accuracy | eval\_loss |
| 5 | 16 | 0.91 | 0.354 |
| 10 | 16 | 0.93 | 0.504 |
| 5 | 4 | 0.94 | 0.302 |
| 10 | 4 | 0.92 | 0.683 |
| 3 | 1 | 0.46 | 1.234 |
| 2 | 1 | 0.46 | 1.401 |

# Zbiór danych

Opracowany zbiór danych dotyczy rozpoznawania rodzaju gry planszowej na podstawie jej krótkiego opisu. Każda z uwzględnionych gier należy do jednej z czterech kategorii: gry familijne, strategiczne, imprezowe, przygodowe.

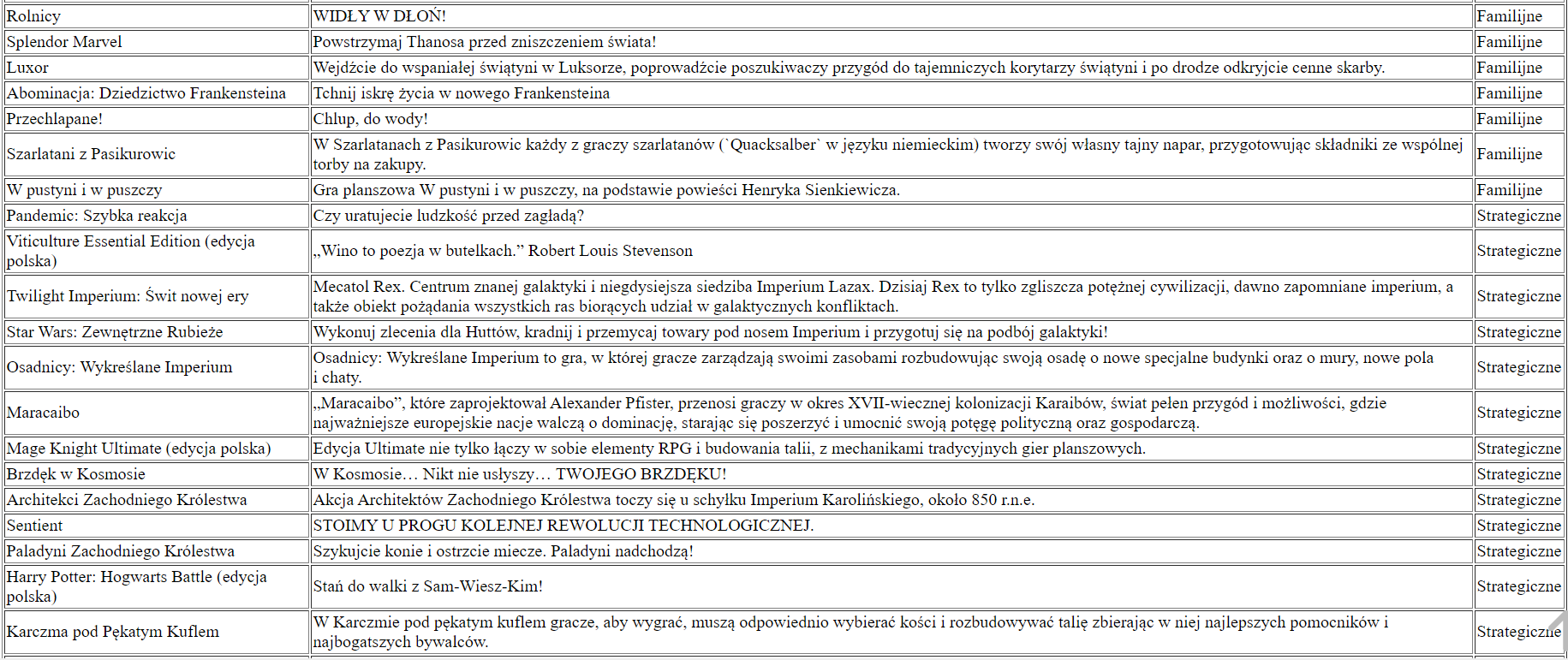
Zbiór ma postać pliku csv, kolumny to kolejno *tytuł*, *opis*, *kategoria*. Plik zawiera opisy 654 gier.

Rozkład gier należących do poszczególnych kategorii:

* Familijne: 322
* Strategiczne: 139
* Imprezowe: 179
* Przygodowe: 70

Zbiór danych można pobrać z repozytorium na GitHubie: <https://raw.githubusercontent.com/tyrrr-aj/NLP_lab/master/gry_planszowe.csv>

Wszystkie opisy pochodzą ze strony sklepu internetowego 3trolle: <https://3trolle.pl/>.

Przykładowy fragment zbioru:

# Wyniki klasyfikacji własnego zbioru danych

## Model ROBERTA

|  |  |  |  |
| --- | --- | --- | --- |
| Liczba epok treningu | Rozmiar batcha | Accuracy | eval\_loss |
| 5 | 8 | 0.58 | 0.679 |

## Klasyfikator bayerowski

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | precision | recall | f1-score | support |
| 0 (Familijne) | 0.52 | 0.83 | 0.64 | 29 |
| 1 (Strategiczne) | 0.50 | 0.33 | 0.40 | 12 |
| 2 (Imprezowe) | 0.80 | 0.27 | 0.40 | 15 |
| 3 (Przygodowe) | 0.57 | 0.40 | 0.47 | 10 |

Total accuracy: 0.55