

MSI – NLP – raport

Autor:
Adam Szreter
296694

1. Trening modelu na dostarczonym zbiorze danych

1.1. Model ROBERTA

Liczba epok treningu	Rozmiar batcha	Accuracy	eval_loss
5	8	0.95	0.302

1.2. 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

2. 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

3. 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 rodzinne, 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:

- Rodzinne: 322
- Strategiczne: 139
- Imprezowe: 179
- Przygodowe: 70

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

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

Przykładowy fragment zbioru:

Rolnicy	WIDLY W DŁOŃ!	Familijne
Splendor Marvel	Powstrzymaj Thanosa przed zniszczeniem świata!	Familijne
Luxor	Wejdźcie do wspaniałej świątyni w Luksorze, poprowadźcie poszukiwaczy przygód do tajemniczych korytarzy świątyni i po drodze odkryjcie cenne skarby.	Familijne
Abominacja: Dziedzictwo Frankensteina	Tchnij iskrę życia w nowego Frankensteina	Familijne
Przechłapane!	Chłup, do wody!	Familijne
Szarlatani z Pasikurowic	W Szarlatanach z Pasikurowic każdy z graczy szarlatanów ('Quacksalber' w języku niemieckim) tworzy swój własny tajny napar, przygotowując składniki ze wspólnej torby na zakupy.	Familijne
W pustyni i w puszczy	Gra planszowa W pustyni i w puszczy, na podstawie powieści Henryka Sienkiewicza.	Familijne
Pandemic: Szybka reakcja	Czy uratujecie ludzkosć przed zagładą?	Strategiczne
Viticulture Essential Edition (edycja polska)	„Wino to poezja w butelkach.” Robert Louis Stevenson	Strategiczne
Twilight Imperium: Świt nowej ery	Mecatol Rex. Centrum znanej galaktyki i niedysiejsza siedziba Imperium Lazax. Dzisiaj Rex to tylko zgłiszcza potężnej cywilizacji, dawno zapomniane imperium, a także obiekt pożądania wszystkich ras biorących udział w galaktycznych konfliktach.	Strategiczne
Star Wars: Zewnętrzne Rubieże	Wykonuj zlecenia dla Huttów, kradnij i przemycaj towary pod nosem Imperium i przygotuj się na podbój galaktyki!	Strategiczne
Osadnicy: Wykreślone Imperium	Osadnicy: Wykreślone Imperium to gra, w której gracze zarządzają swoimi zasobami rozbudowując swoją osadę o nowe specjalne budynki oraz o mury, nowe pola i chaty.	Strategiczne
Maracaibo	„Maracaibo”, które zaprojektował Alexander Pfister, przenosi graczy w okres XVII-wiecznej kolonizacji Karaibów, świat pełen przygód i możliwości, gdzie najważniejsze europejskie nacje walczą o dominację, starając się poszerzyć i umocnić swoją potęgę polityczną oraz gospodarczą.	Strategiczne
Mage Knight Ultimate (edycja polska)	Edycja Ultimate nie tylko łączy w sobie elementy RPG i budowania talii, z mechanikami tradycyjnych gier planszowych.	Strategiczne
Brzęk w Kosmosie	W Kosmosie... Nikt nie usłyszy... TWOJEGO BRZĘDKU!	Strategiczne
Architektury Zachodniego Królestwa	Akcja Architektów Zachodniego Królestwa toczy się u schyłku Imperium Karolińskiego, około 850 r.n.e.	Strategiczne
Sentient	STOIMY U PROGU KOLEJNEJ REWOLUCJI TECHNOLOGICZNEJ.	Strategiczne
Paladyni Zachodniego Królestwa	Szykujcie konie i ostrzcie miecze. Paladyni nadchodzą!	Strategiczne
Harry Potter: Hogwarts Battle (edycja polska)	Stań do walki z Sam-Wiesz-Kim!	Strategiczne
Karczna pod Pękatym Kuflem	W Karczmie pod pękatym kuflem gracze, aby wygrać, muszą odpowiednio wybierać kości i rozbudowywać talię zbierając w niej najlepszych pomocników i najbogatszych bywalców.	Strategiczne

4. Wyniki klasyfikacji własnego zbioru danych

4.1. Model ROBERTA

Liczba epok treningu	Rozmiar batcha	Accuracy	eval_loss
5	8	0.58	0.679

4.2. 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