

Metody Sztucznej Inteligencji – Raport z NLP

Kacper Korban
296592

1. Wyniki treningu modeli na dostarczonym zbiorze danych.

Wyniki dla sieci neuronowej:

INFO:simpletransformers.classification.classification_model: {'mcc': 0.8517603409302043, 'tp': 116, 'tn': 121, 'fp': 11, 'fn': 8, 'acc': 0.92578125, 'eval_loss': 0.3700453476049006}

Wyniki dla modelu bayesowskiego:

	precision	recall	f1-score	support
accuracy			0.87	256
macro avg	0.87	0.87	0.87	256
weighted avg	0.87	0.87	0.87	256

2. Wyniki eksperymentów związanych z modyfikacją hiperparametrów.

Dla parametrów {"num_train_epochs": 8, "eval_batch_size": 12, "train_batch_size": 12}:

INFO:simpletransformers.classification.classification_model: {'mcc': 0.8455522971652004, 'tp': 118, 'tn': 118, 'fp': 14, 'fn': 6, 'acc': 0.921875, 'eval_loss': 0.4962954245645299}

Dla parametrów {"num_train_epochs": 4, "eval_batch_size": 4, "train_batch_size": 4}:

INFO:simpletransformers.classification.classification_model: {'mcc': 0.8465584171413254, 'tp': 119, 'tn': 117, 'fp': 15, 'fn': 5, 'acc': 0.921875, 'eval_loss': 0.4497242777142674}

Dla parametrów {"num_train_epochs": 12, "eval_batch_size": 10, "train_batch_size": 10}:

INFO:simpletransformers.classification.classification_model: {'mcc': 0.8522488756826735, 'tp': 117, 'tn': 120, 'fp': 12, 'fn': 7, 'acc': 0.92578125, 'eval_loss': 0.6250196070272628}

3. Opis zbioru danych wraz z linkiem pozwalającym na jego pobranie.

Link do pobrania danych: <https://github.com/KacperFKorban/NLP-exercise/blob/master/dane.csv>

Zbiór jaki skonstruowałem składa się z tekstów z dwóch różnych źródeł:

- Wersów piosenek rapera Taco Hemingway
- Podzielone na fragmenty, wybrane wystąpienia prezydenta RP Andrzeja Dudy

Teksty piosenek zostały pobrane z portalu tekstowo.pl, natomiast teksty wystąpień prezydenta Andrzeja Dudy zostały pobrane ze strony prezydent.pl

Przykładowy fragment pliku z danymi:

'18 mnie wykończy, chyba w '19 przerwa, ej;taco

Ty przejrzysta jak woda Fiji (i chcę cię na hejnał);taco

Ale chciałbym, żeby z tego w ogromnym stopniu skorzystały także i samorządy;duda

Więc bardzo proszę o wzmożoną ostrożność;duda

4. Wyniki treningu modelu dla własnego zbioru danych.

Wyniki uczenia dla 5 epok i batch_size 8:

```
{'mcc': 0.8479565969392698, 'tp': 102, 'tn': 64, 'fp': 7, 'fn': 6, 'acc': 0.9273743016759777, 'eval_loss': 0.38592780532781035}
```

Wybrane fragmenty ze zbioru testowego, które zostały źle zaklasyfikowane:

Dzisiaj tańczą przy Wiśle

~ Andrzej Duda

Mój przyjaciel skończył skacowany w izbie wytrzeźwień

~ Andrzej Duda

Więc zaprasza do siebie, tu mało miejsca do tańców

~ Andrzej Duda

A WWA to NHL jest w tej kwestii

~ Andrzej Duda

Dla mnie takim momentem była wizyta w Moskwie kilka dni po katastrofie

~ Taco Hemingway

Ale burmistrz znał chłopców po imieniu

~ Taco Hemingway

...

Wyniki uczenia dla 3 epok i batch_size 10:

```
{'mcc': 0.909800957420005, 'tp': 94, 'tn': 77, 'fp': 4, 'fn': 4, 'acc': 0.9553072625698324, 'eval_loss': 0.16895643580291006}
```

Wybrane fragmenty ze zbioru testowego, które zostały źle zaklasyfikowane:

Powoduje stres, wzмага nerwowość i lęk

~ Taco Hemingway

Tyle dobrego, że znów jeżdżę sobie tym ZTM-em

~ Andrzej Duda

Mózg jak karabin, mógłbym zabić kogoś, zgładzić myślą

~ Andrzej Duda

A ja sobie stoję no i słucham o czym mowa, bo ich kocham, bo to w sumie jak nie patrzeć moi ludzie

~ Andrzej Duda

Bardzo proszę też o szczerość

~ Taco Hemingway

...

Wyniki uczenia dla 2 epok i batch_size 16:

{'mcc': 0.9210499059868467, 'tp': 95, 'tn': 77, 'fp': 4, 'fn': 3, 'acc': 0.9608938547486033, 'eval_loss': 0.12878633849322796}

Wybrane fragmenty ze zbioru testowego, które zostały źle zaklasyfikowane:

Powoduje stres, wzмага nerwowość i lęk

~ Taco Hemingway

Tyle dobrego, że znów jeżdżę sobie tym ZTM-em

~ Andrzej Duda

Wspaniali ludzie

~ Taco Hemingway

Boję się policji strasznie, ale wieście to wreszcie

~ Andrzej Duda

Wygrywa ten kto przegrał, to przecież jest jakiś obłęd

~ Andrzej Duda

...

Wyniki uczenia dla 1 epok i batch_size 6:

{'mcc': 0.8360126003851731, 'tp': 95, 'tn': 69, 'fp': 13, 'fn': 2, 'acc': 0.9162011173184358, 'eval_loss': 0.28408855355034274}

Wybrane fragmenty ze zbioru testowego, które zostały źle zaklasyfikowane:

Trwała 45 minut

~ Taco Hemingway

Chciałbym powiedzieć teraz, że uciekam przed policją

~ Andrzej Duda

Ze zbudowanym obok złobkiem – bardzo ładnym, eleganckim

~ Taco Hemingway

Dlatego bardzo proszę o wzmożoną ostrożność

~ Taco Hemingway

By rozwijać nasz kraj, dbać o naszą ojczyznę

~ *Taco Hemingway*

Ale teraz mówię adieu wszystkim którzy wyczekują aż upadnę!

~ *Andrzej Duda*

...

Wyniki uczenia dla 2 epok i batch_size 8:

{'mcc': 0.8740735810446687, 'tp': 98, 'tn': 70, 'fp': 6, 'fn': 5,
'acc': **0.9385474860335196**, 'eval_loss': 0.24162023307998542}

Wybrane fragmenty ze zbioru testowego, które zostały źle zaklasyfikowane:

Chciałbym powiedzieć teraz, że uciekam przed policją

~ *Andrzej Duda*

Tworzy ją pakiet działań na niespotykaną dotąd skalę

~ *Taco Hemingway*

Sądzą, że ten blant wielce pomocny

~ *Andrzej Duda*

Miałem tróję z podstaw przedsiębiorczości

~ *Andrzej Duda*

To była dla mnie także osobista tragedia

~ *Taco Hemingway*

Bardzo proszę też o szczerość

~ *Taco Hemingway*

...