

Natural Language Processing

Andrzej Ratajczak 296658

1. Wynik treningu modeli na dostarczonym zbiorze danych

Dla sieci neuronowej

	Accuracy score	Loss
"num_train_epochs": 5	0.90625	0.4745

Dla Bayes'a

Hiperparametry, najlepsze z parameters przekazanych do GridSearch'a

max_df=0.5, ngram_range=(1, 1), alpha=0.01

	precision	recall	f1-score	support
0	0.91	0.95	0.93	131
1	0.95	0.90	0.92	125
accuracy			0.93	256
macro avg	0.93	0.93	0.93	256
weighted avg	0.93	0.93	0.93	256

2. Wynik eksperymentu związany z modyfikacją hiperparametrów sieci neuronowej

	Accuracy score	Loss
"num_train_epochs": 7, "eval_batch_size": 10	0.9140625	0.5675
"num_train_epochs": 4, "eval_batch_size": 15	0.90625	0.4126
"num_train_epochs": 6, "eval_batch_size": 4	0.91796875	0.4849

3. Opis zbioru danych

Przygotowany przeze mnie zbiór danych to 857 wersów piosenek Comy oraz 753 wersów piosenek Czadomana. Dane zostały zebrane z serwisu tekstowo.pl, zaprezentowane są w pliku CSV z kolumnami „description” oraz „label”, gdzie label to wartość z dyskretnego zbioru {coma, czadoman}.

Przykładowe kilka rekordów ze zbioru danych:

Biorę głęboki wdech;coma

które od bólu łzami pękną;coma

Hej, hej śmiejmy się rozwalmy te budę!;czadoman

Wyświecone pod nosem;coma

jestem dziś;czadoman

bo ta impreza to nie byle co;czadoman

DJ wrzuca płytę balet rozpoczyna;czadoman

Co Blondi ze mną robi, jak jej chodzą nogi;czadoman

Dane można znaleźć na poniższym repozytorium:

<https://github.com/BarkingBad/NLP-exercise>

4. Wyniki treningu dla własnego zbioru danych

	Accuracy score	Loss
"num_train_epochs": 5	0.8819875776397516	0.6027269281801724
"num_train_epochs": 7, "eval_batch_size": 10	0.8819875776397516	0.7017024538055172
"num_train_epochs": 4, "eval_batch_size": 15	0.8695652173913043	0.6391457819633863
"num_train_epochs": 6, "eval_batch_size": 4	0.8757763975155279	0.5151640896390124

Poniżej znajdują się wybrane wersy, które zostały zakwalifikowane błędnie

on tu nas ochładza

coma

Dobry bit króluje w żyłach krew buzuje

coma

Za nim zacznę z nią przygodę

coma

Bateria w telefonie wyzionęła ducha

coma

jeszcze tylko zapach typu Bond

coma

Ma zielone kocie oczy, tak samo jak ty

czadoman

Tylko ja się dziś odważę

coma

Ogłaszam zbiorową dymisję, wszyscy musimy ustąpić

czadoman

Na skraju miast, na skraju dnia

czadoman

Będę z tobą wobec tego szczerzy

czadoman

Niech nam będzie wspaniale zabiorę Cię z sobą gdzieś

czadoman

DJ wrzuca płytę balet rozpoczyna

coma

Chciałbym ocalić, co się da.

czadoman

Na skraju miast, na skraju dnia

czadoman

Czy byłem z nią w piekle,

coma

rozkładam na piasku, przy świecach koc,

coma

Nie byłeś pewny gdy to się zdarzyło

czadoman

Czuję jak można z naturą blisko

czadoman