

Automatic cyberbullying detection

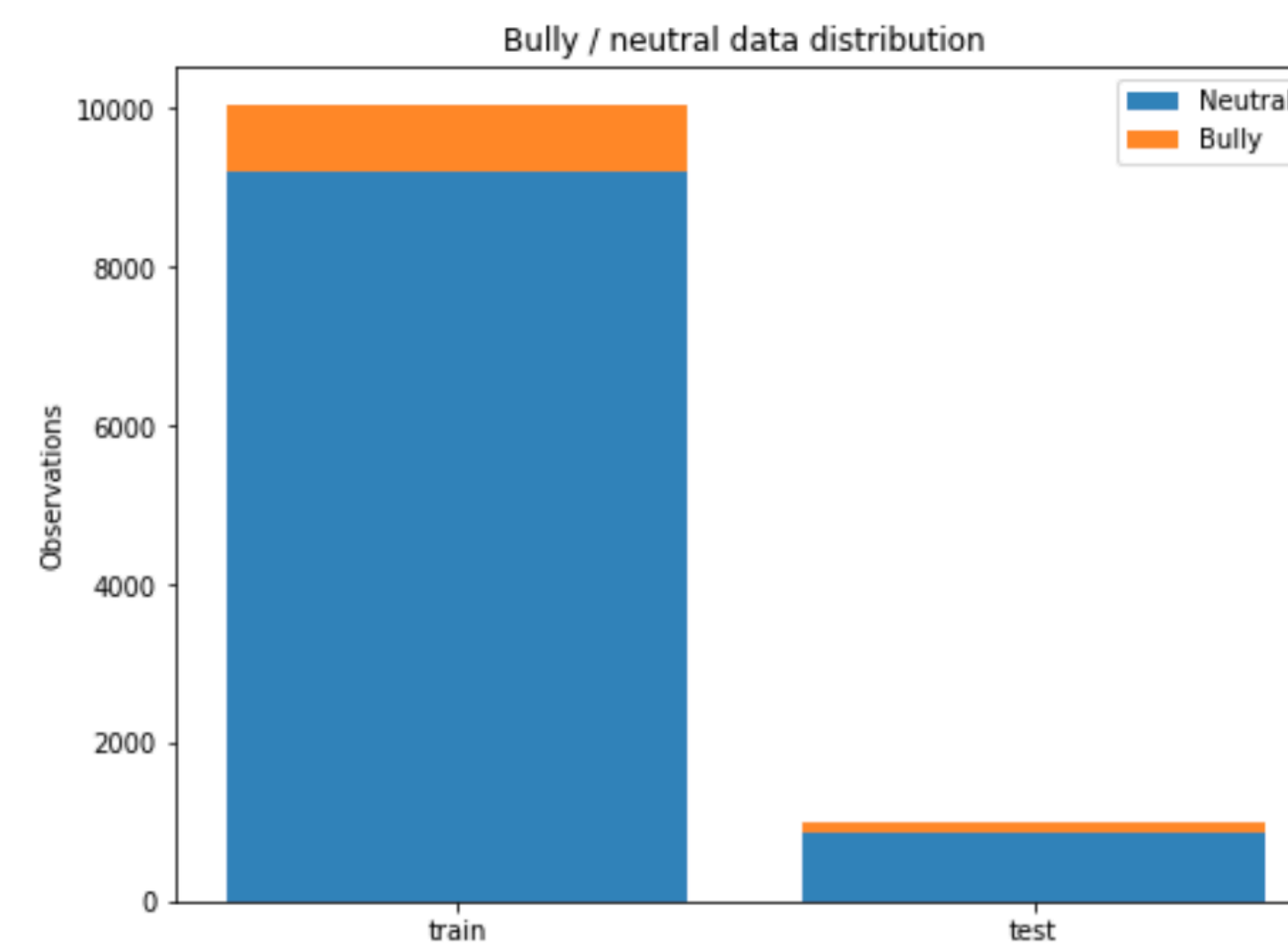
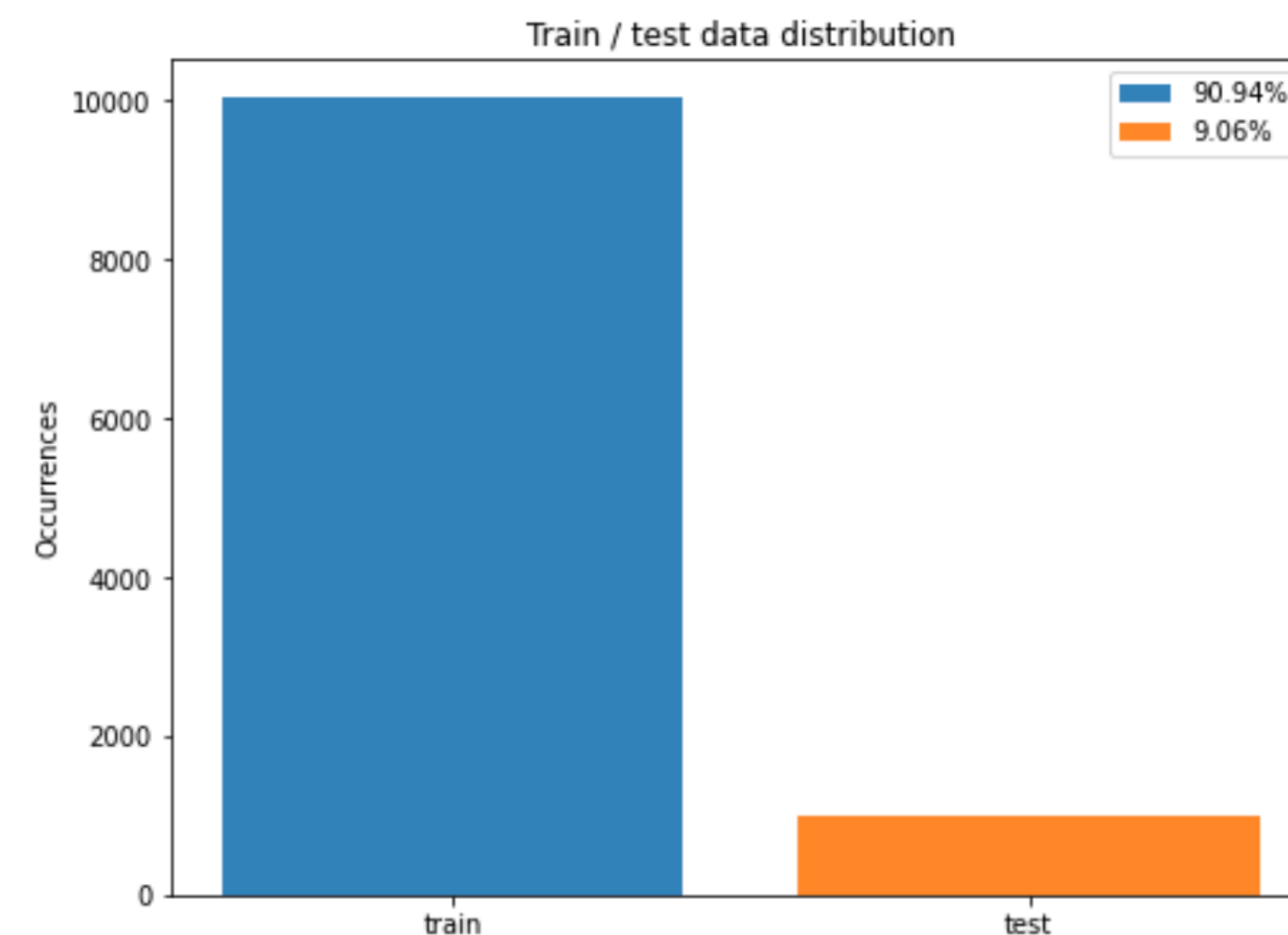
Harmful vs non-harmful

Cel

**Rozróżnienie komentarzy
neutralnych od obraźliwych**

Dane

- PolEval2019 - Task6-1
- Liczba obserwacji: 11041 Tweetów:
 - Obrażliwych: 985 (8.92%)
 - Neutralnych: 10056 (91,08%)



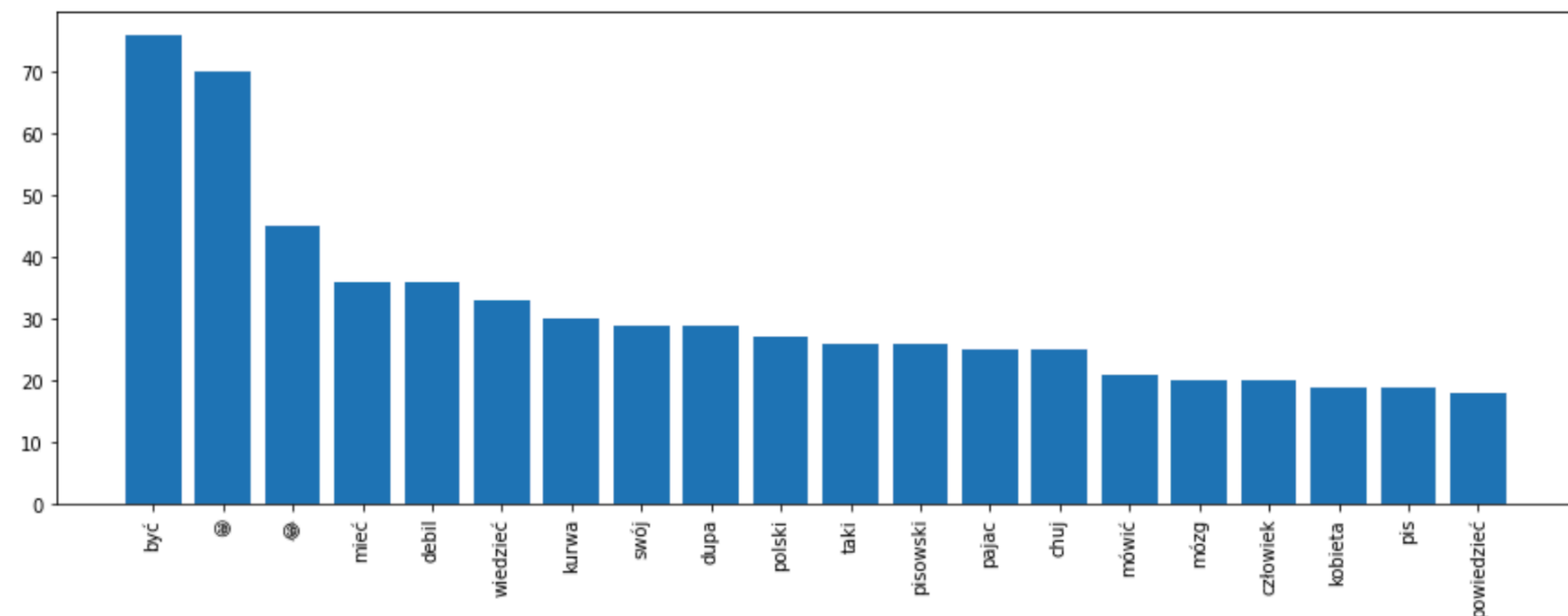
Najczęściej występujące słowa w komentarzach obraźliwych



- **Brak znaków interpunkcyjnych**
- **Brak słów przystankowych**
- **Brak tagów**
charakterystycznych dla
Twittera
- **Lematy**
- **Lowercase**

Found 6497 words.

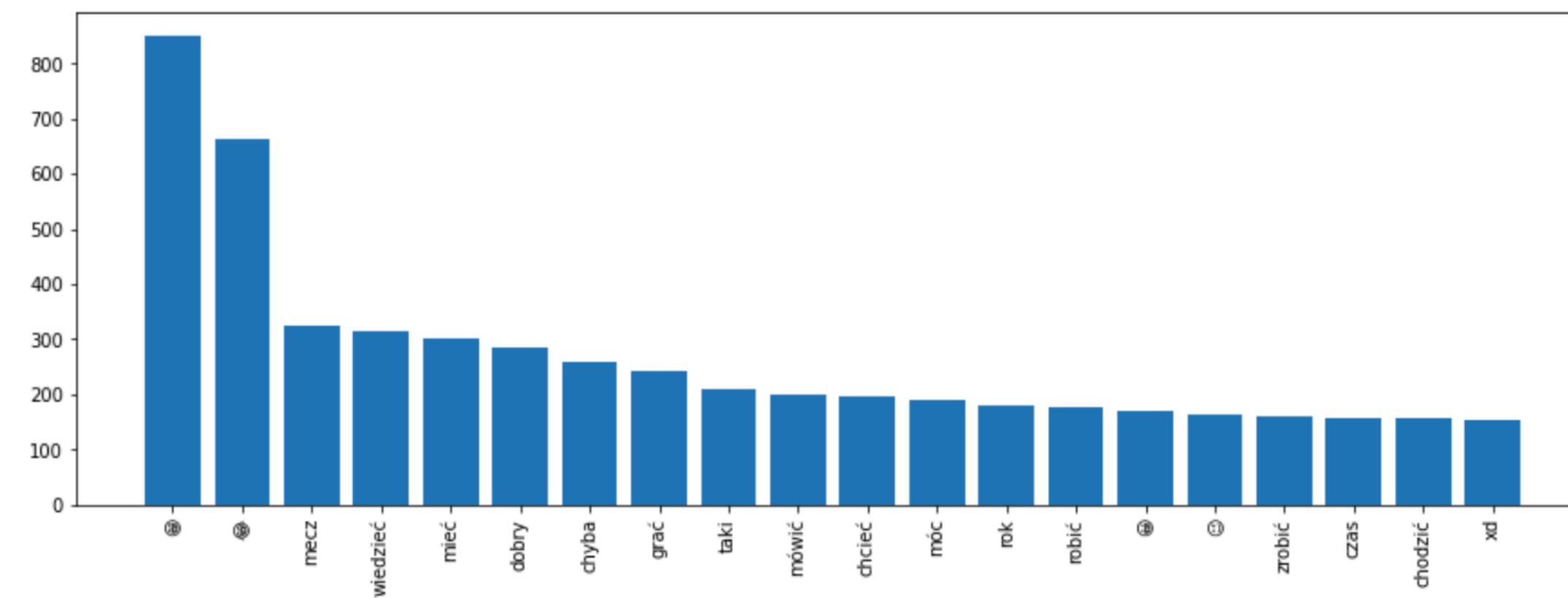
Top 30



**Oraz w
kometarzach
neutralnych...**



Found 62369 words.



SpaCy TextCategorizer

TextCategorizer

Próba #1

- 50 epok
- Dane podzielone oryginalnie

Training the model...

Epoch	Loss	Prec	Recall	F-score	Time (s)	pred_bully	true_bully
0	4.997	0.0	0.0	0.0	36.71	2	134
1	0.1524	0.0	0.0	0.0	21.71	2	134
2	0.0271	0.3333	0.0149	0.0286	17.68	6	134
3	0.013	0.4286	0.0224	0.0426	17.16	7	134
4	0.0086	0.1667	0.0149	0.0274	17.25	12	134
5	0.0048	0.1429	0.0149	0.027	17.04	14	134
■■■							
43	0.0007	0.0909	0.0149	0.0256	17.97	22	134
44	0.0007	0.0952	0.0149	0.0258	17.7	21	134
45	0.0004	0.0909	0.0149	0.0256	17.69	22	134
46	0.0007	0.0952	0.0149	0.0258	18.35	21	134
47	0.0012	0.05	0.0075	0.013	17.91	20	134
48	0.0005	0.0476	0.0075	0.0129	17.83	21	134
49	0.0006	0.05	0.0075	0.013	18.08	20	134

> Saved model as nlp_spacy_model_161183666686_unbalanced

y_true: Counter({0: 866, 1: 134})

y_pred: Counter({0: 950, 1: 50})

Accuracy	Precision	F1-Score	Recall
83.2%	16.0%	8.7%	5.97%

TextCategorizer

Próba #2

- 50 epok
- W danych treningowych znajdują się komentarze neutralne i obraźliwe w proporcji 1:1

```
Loaded model pl_core_news_sm-2.3.0/pl_core_news_sm/pl_core_news_sm-2.3.0
Training the model...
```

Epoch	Loss	Prec	Recall	F-score	Time (s)	pred_bully	true_bully
0	10.3244	0.5882	0.4598	0.5161	8.44	68	87
1	1.838	0.4944	0.5057	0.5	6.66	89	87
2	0.422	0.4545	0.4598	0.4571	5.72	88	87
3	0.0959	0.4699	0.4483	0.4588	4.94	83	87
4	0.0242	0.4667	0.4023	0.4321	4.33	75	87

...

44	0.0001	0.4845	0.5402	0.5109	3.07	97	87
45	0.0001	0.4898	0.5517	0.5189	3.01	98	87
46	0.0001	0.4804	0.5632	0.5185	3.01	102	87
47	0.0001	0.4757	0.5632	0.5158	2.99	103	87
48	0.0001	0.4762	0.5747	0.5208	2.98	105	87
49	0.0003	0.4811	0.5862	0.5285	2.98	106	87

```
> Saved model as nlp_spacy_model_161184770176_balanced
```



```
y_true: Counter({0: 866, 1: 134})
y_pred: Counter({0: 591, 1: 409})
```

Accuracy	Precision	F1-Score	Recall
60.7%	18.34%	27.62%	55.97%

Logistic Regression

Logistic Regression with GridSearchCv

- Oryginalny podział danych
- Parametry: {'C': 1, 'solver': 'saga'}

```
Took 21.85 second(s).  
Best parameter (CV score=0.917):  
{'C': 1, 'solver': 'saga'}
```

```
Accuracy: 87.9%  
Precision: 88.24%  
f1_score: 19.87%  
Recall: 11.19%
```

Top Bully:

```
1) ('pajac', 2.859591009483846)  
2) ('gnój', 2.5714211556821867)  
3) ('mózg', 2.345975341635868)  
4) ('debil', 2.24125208421774)  
5) ('kutas', 2.1918653866857105)
```

Top Neutral:

```
1) ('sędzia', -1.2354121672847465)  
2) ('mecz', -1.227042980365117)  
3) ('nhttps', -1.2209669882544243)  
4) ('pytanie', -1.0602664994571964)  
5) ('liga', -1.0421198854035345)
```

Regresja Logistyczna

Wygrywa

Patryk Laskowski
s21887