

1. Wyniki treningu na dostarczonym zbiorze danych

1.1. Sieć neuronowa

Output:

```
INFO:simpletransformers.classification.classification_model:
{'mcc': 0.8984277314260114,
 'tp': 119,
 'tn': 124,
 'fp': 7,
 'fn': 6,
 'acc': 0.94921875,
 'eval_loss': 0.2754532750695944}
```

Stąd precyzja wynosi około 95%.

1.2. Klasyfikator Bayesowski

| Applying best classifier on test data: | | | | |
|--|-----------|--------|----------|---------|
| | precision | recall | f1-score | support |
| 0 | 0.90 | 0.90 | 0.90 | 131 |
| 1 | 0.90 | 0.90 | 0.90 | 125 |
| accuracy | | | 0.90 | 256 |
| macro avg | 0.90 | 0.90 | 0.90 | 256 |
| weighted avg | 0.90 | 0.90 | 0.90 | 256 |

2. Modyfikacja hiper-parametrów

Testowano zmiany liczby epok i rozmiaru batcha, nie testowano rozmiarów batcha mniejszych od 10, gdyż trenowanie modelu dla większej niż 1 liczby epok trwało zbyt długo.

| Batchsize/Epochs | 1 | 2 | 5 |
|------------------|----------|----------|----------|
| 10 | 0.914062 | 0.9375 | 0.917969 |
| 15 | 0.921875 | 0.929688 | 0.9375 |

W badanym zakresie nie obserwowano drastycznych zmian, wszystkie wartości precyzji są powyżej 90%, a różnica między największą i najmniejszą precyzją to około 2%. Dodatkowo wszystkie wartości są niższe niż precyzja

uzyskana dla parametrów użytych w pierwszej części ćwiczenia (liczba epok: 5, rozmiar batcha: 8 (domyślny)).

3. Własny zbiór danych

Wybrano fragmenty pierwszego tomu „Ogniem i mieczem” oraz „Odyseji” z tekstów dostępnych w portalu *wolnelektury.pl*. Kryterium doboru tych pozycji były różne epoki w jakich powstawały oraz fakt, że jedna z nich jest tekstem oryginalnym a druga tłumaczeniem. Ze względu na zbyt długi czas treningu modelu dla całego tekstu, z każdego z nich wyciągnięto po 1000 zdań rozpoczynając od 50 z kolei (aby pominąć numery ISBN, tytuł i inne nieistotne w tym ćwiczeniu informacje zamieszczane na początku utworu). Zbiór danych jest dostępny pod adresem: https://raw.githubusercontent.com/pwrobel5/msi/master/nlp/text_set.csv

4. Wyniki treningu modelu dla własnego zbioru danych

Zastosowano liczbę epok równą 5 i domyślny rozmiar batcha (8), rozmiar zbioru uczącego wynosił 90% całego zbioru.

Output:

```
{ 'mcc': 0.8015425832202753,
  'tp': 88,
  'tn': 92,
  'fp': 7,
  'fn': 13,
  'acc': 0.9,
  'eval_loss': 0.3862661376595497 }
```

Stąd precyzja wynosi 90%.