MUSIC GENRE CLASSIFICATION

Bartłomiej Eljasiak

Aleksandra Nawrocka

Dominika Umiastowska

LYRICS PREPROCESSING

- Initial preprocessing:
  - Removing punctuation, numbers and special strings (e.g. 'Verse', 'Chorus', '2x')
  - Expanding contractions
  - Converting to lower case
- Text Normalization
  - Tokenization
  - Lemmatization
  - Removing stop words

DATASETS PREPROCESSING

- Creation of multiple datasets out of data that we had
  - Merging both datasets
  - Leaving only 5 genres – Rock (125 230), Pop (47 896), Metal (34 706), Hip-hop (31 066), Country (23 224)
  - Making final dataset balanced
- Creation of an additional dataset that has both lyrics and the title of the song out of *Song lyrics from 79 musical genres* dataset
  - 5 genres - Rock (25 177), Pop (13 759), Metal (13 496), Hip-hop (8 412), Country (7 377)

FIRST EXPERIMENTS - GLOVE

Classifier	Accuracy	Balanced accuracy	F1-score
Naive Bayes	15,43%	16,48%	9,66%
Linear SVM	46,18%	20,55%	42,78%
XGBoost	30,39%	13,20%	31,17%
CNN	50,91%	25,84%	48,72%

BALANCED LYRICS DATASET RESULTS

SMALLER BERT - METHODS COMPARISON

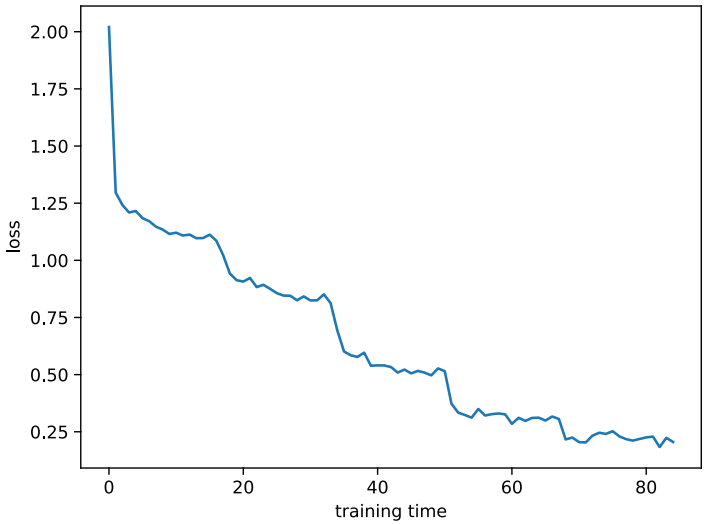
UNNORMALIZED

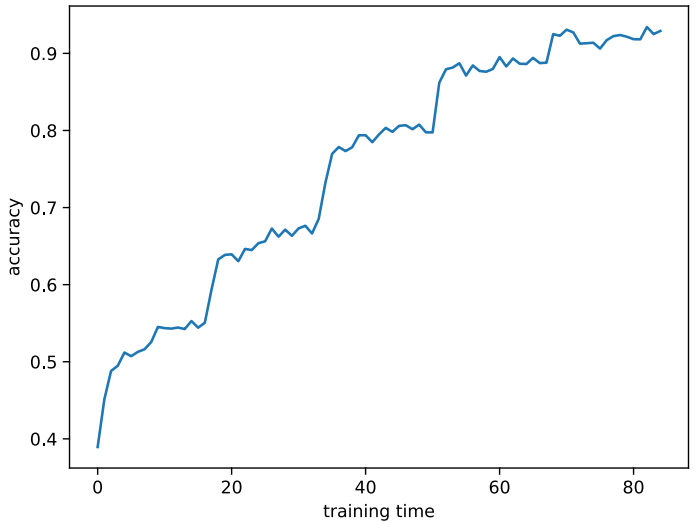
Classifier	Accuracy	F1-score
Naive Bayes	43,24%	39,18%
Linear SVM	43,60%	40,03%
XGBoost	42,38%	42,30%
CNN	51,31%	51,05%

NORMALIZED

Classifier	Accuracy	F1-score
Naive Bayes	40,33%	33,69%
Linear SVM	50,61%	46,12%
XGBoost	45,47%	45,68%
CNN	53,54%	53,02%

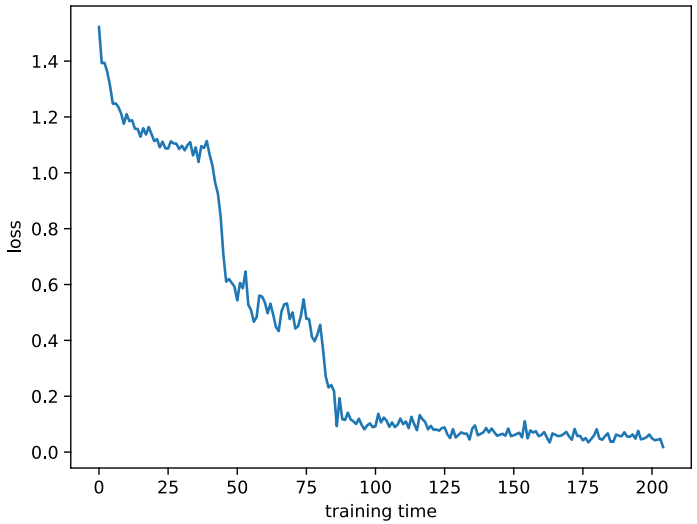
LYRICS - GLOVE + CNN

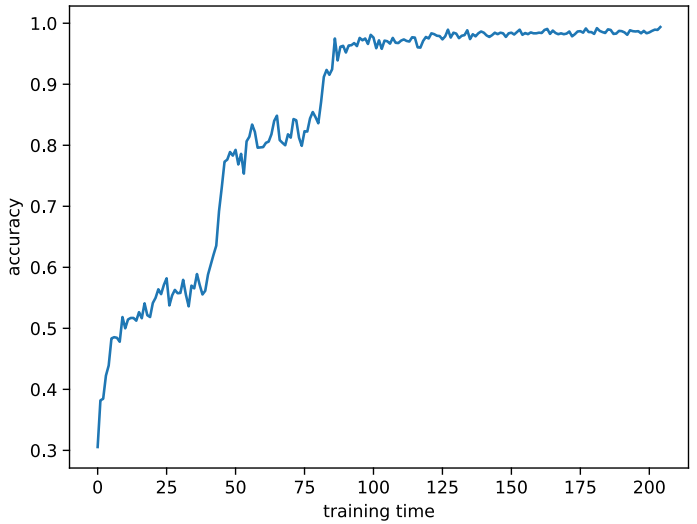




Merged dataset (5 genres, balanced)	
Accuracy	53,73%
F1-score	53,25%

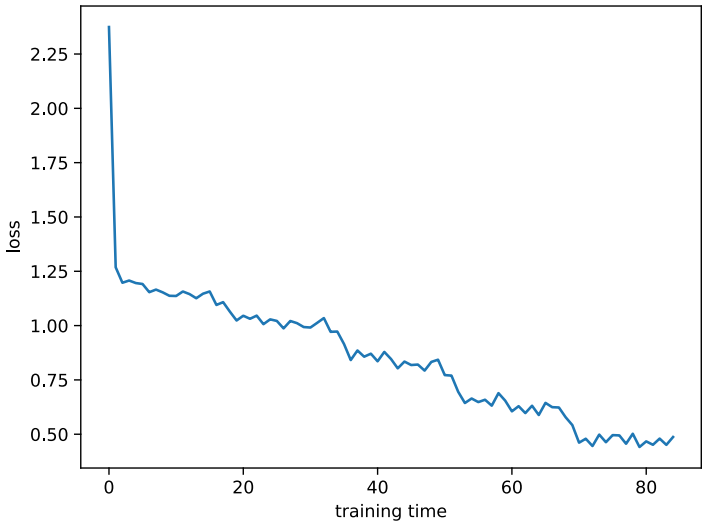
LYRICS - WORD2VEC + CNN

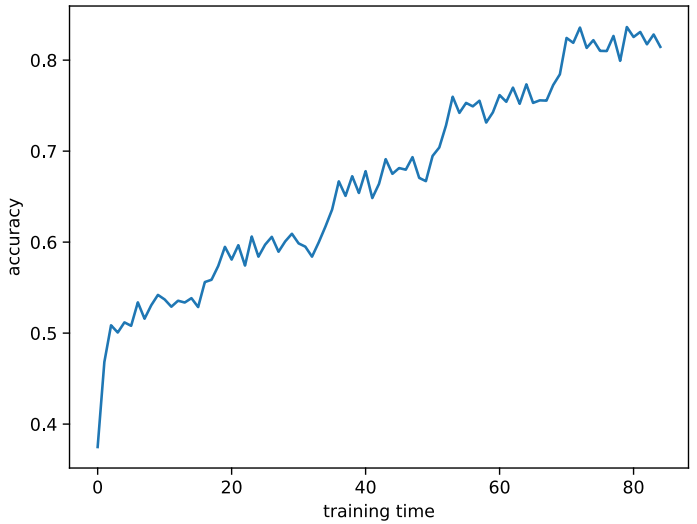




Merged dataset (5 genres, balanced)	
Accuracy	52,61%
F1-score	52,65%

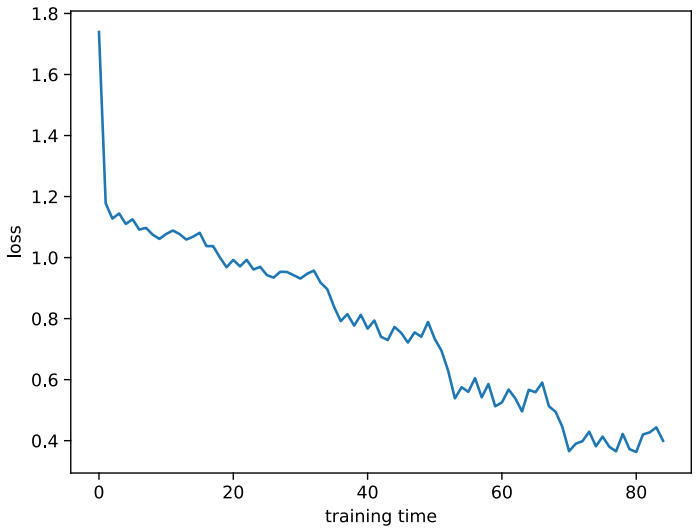
LYRICS – SMALLER BERT + CNN

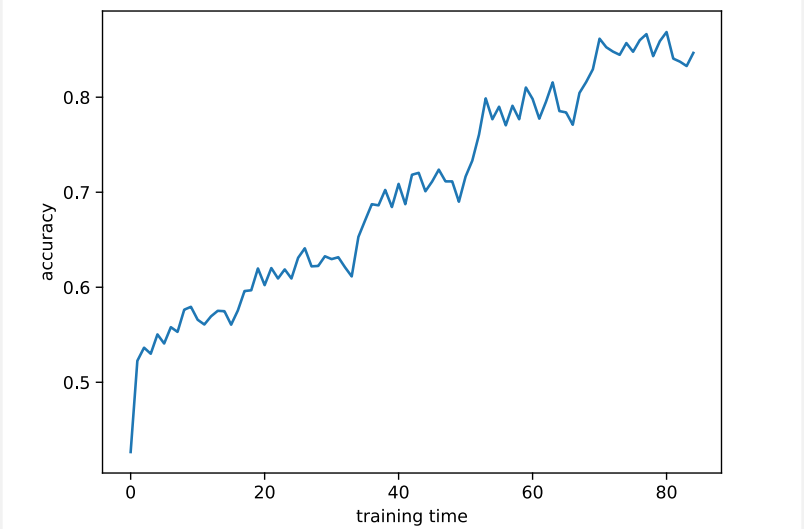




Merged dataset (5 genres, balanced)	
Accuracy	51,31%
F1-score	51,05%

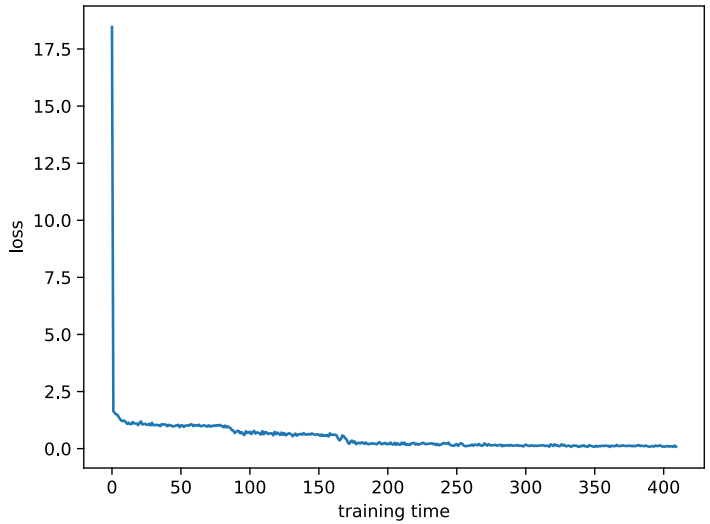
TOKENIZED LYRICS – SMALLER BERT + CNN

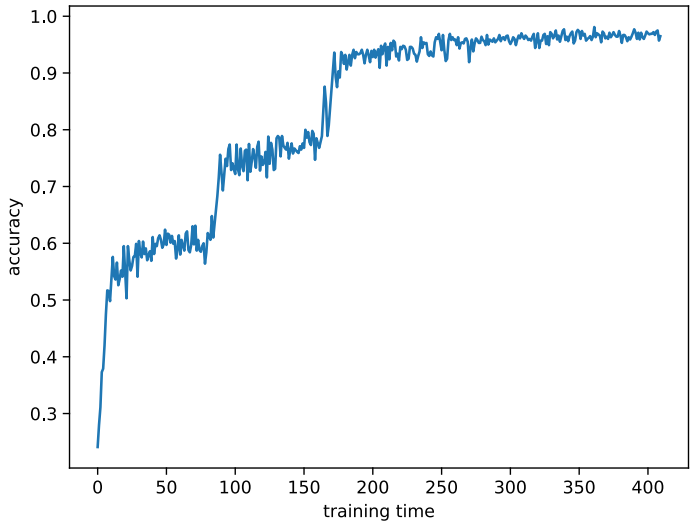




Merged dataset (5 genres, balanced)	
Accuracy	53,54%
F1-score	53,02%

LYRICS – BASE BERT + CNN





Merged dataset (5 genres, balanced)	
Accuracy	56,48%
F1-score	56,55%

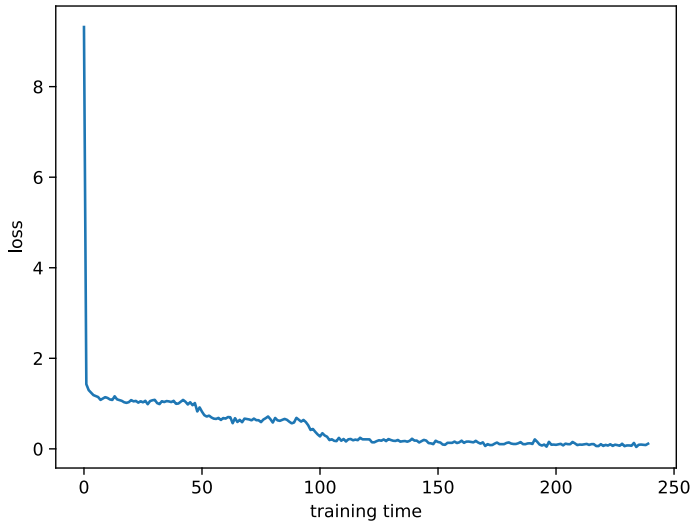
EMBEDDING COMPARISON

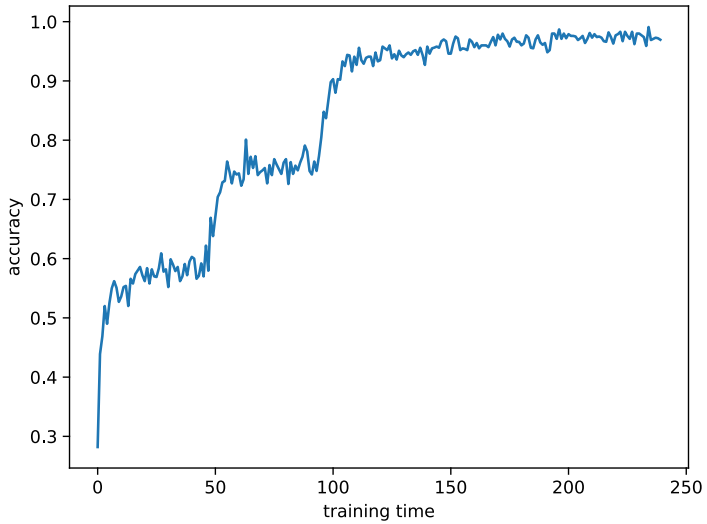
Embedding	Accuracy	F1-score
GloVe (100)	53,73%	53,25%
Smaller BERT (128)	51,31%	51,05%
Base BERT (768)	56,48%	56,55%
Word2vec (300)	52,61%	52,65%

* Fine-tuned Smaller BERT (10 epochs)
  Accuracy: 48,49%

TITLE + LYRICS DATASET RESULTS

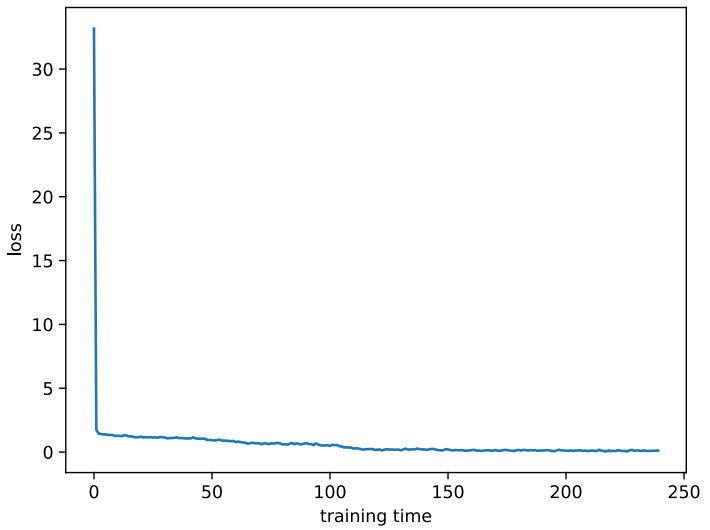
LYRICS – BASE BERT + CNN

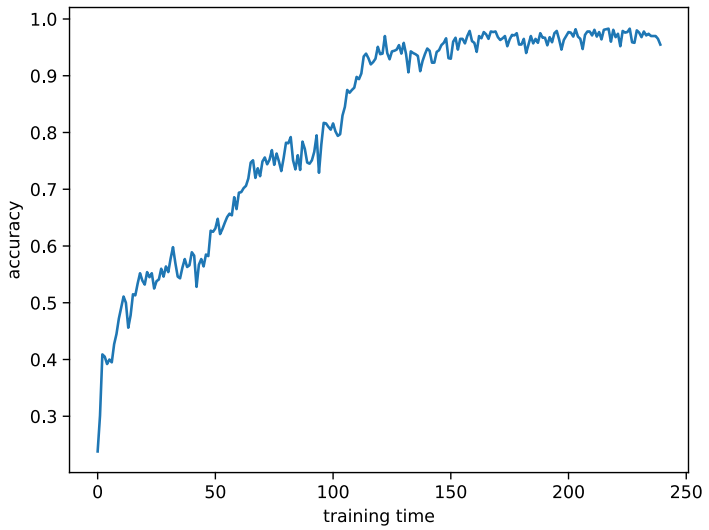




Song lyrics from 79 musical genres (5 genres)	
Accuracy	54,27%
Balanced accuracy	55,36%
F1-score	54,80%

TITLE + LYRICS – BASE BERT + CNN





Song lyrics from 79 musical genres (5 genres)	
Accuracy	57,17%
Balanced accuracy	52,65%
F1-score	56,10%

WHAT HAVE WE NOT DONE AND WHY NOT?

- Adding sentiment analysis
  - Problems with Tensorflow library as required layer cannot be used in the most common way (2 input vectors)
  - Cross-attention not available in Pytorch library
  - Not enough time

PLAN FOR THE SECOND PROJECT

- Adding sentiment analysis
- Adding neural network classifier (a few dense layers)
- Droping classifiers other than CNN and NN
- Creating our own dataset
- New architecture consisting of two networks connected sequentially:
  - First one deciding whether song belongs to Rock genre or not
  - Second for genres other than Rock

THANK YOU FOR ATTENTION!