

# Music genre classification based on song lyrics

BARTŁOMIEJ ELJASIAK  
ALEKSANDRA NAWROCKA  
DOMINIKA UMIASTOWSKA

# Presentation plan

1. Short project description
2. Final results of project 1
3. Improvements done in project 2
  1. New embeddings
  2. Sentiment
  3. Two-step classification
  4. Fine-tuning
  5. Custom dataset creation
4. Conclusions

# Purpose of the project

---

- Using different embeddings and classifiers we want to find the best way to predict genre based on song lyrics.
- In addition, we want to investigate the influence of sentiment analysis performed on the lyrics on the prediction.
- What is more, we want to create new lyrics dataset to see how our methods deal with the new data.

# First project final results

- Overall, ensemble of Base BERT and CNN gave the best results.
- Adding title and text preprocessing also increased the accuracy.

**Smaller BERT**

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

**CNN**

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

# Improvements in the second project

---

- New embeddings: DistilBERT and SentenceTransformerMPNET
- New classifiers: 2-step CNN classifier and NN classifier
- Fine tuning of DistilBERT
- Ensembling of sentiment pre-trained model with DistilBERT model
- Creation of new lyrics dataset

# New embeddings

- Based on DistilBERT and SentenceTransformerMPNET
- Significantly improved results for all classifiers
- Created from max. 400 words from lyrics
- Singular embedding represents the whole song

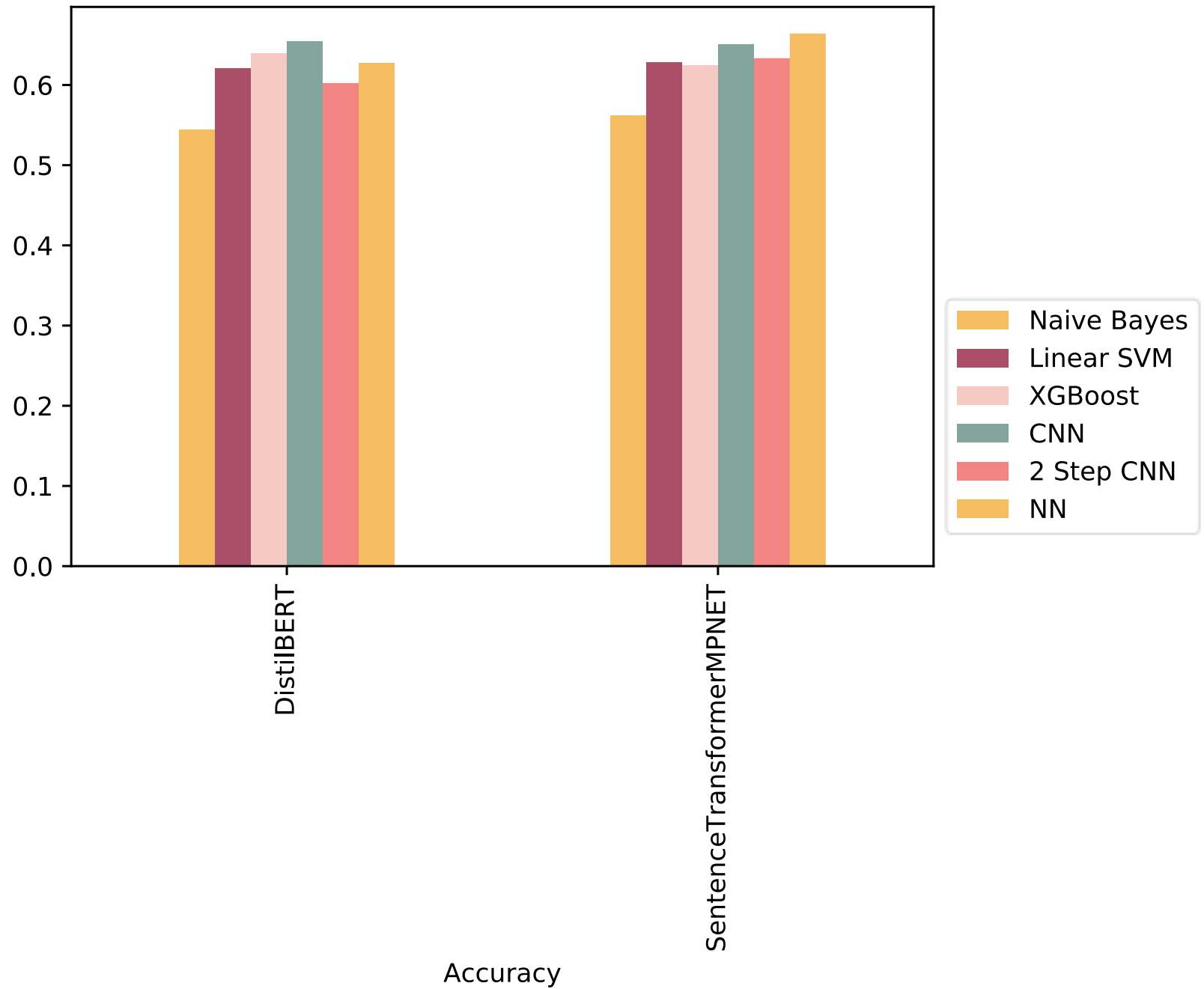
## DistilBERT

Classifier	Accuracy	F1-score	Acc. change
Naive Bayes	54,40%	53,16%	+ 11,16
Linear SVM	62,06%	61,07%	+ 11,45
XGBoost	63,93%	63,78%	<b>+ 18,46</b>
<b>CNN</b>	<b>65,47%</b>	<b>65,53%</b>	+ 8,99
NN	64,96%	64,99%	-
2StepCNN	62,21%	57,97%	-

## SentenceTransformerMPNET

Classifier	Accuracy	F1-score	Acc. change
Naive Bayes	56,23%	55,97%	+ 12,99
Linear SVM	62,83%	61,66%	+ 12,22
XGBoost	62,45%	62,31%	<b>+ 16,98</b>
CNN	65,06%	64,30%	+ 8,58
<b>NN</b>	<b>66,42%</b>	<b>66,08%</b>	-
2StepCNN	63,34%	60,44%	-

# Methods comparison



# Sentiment

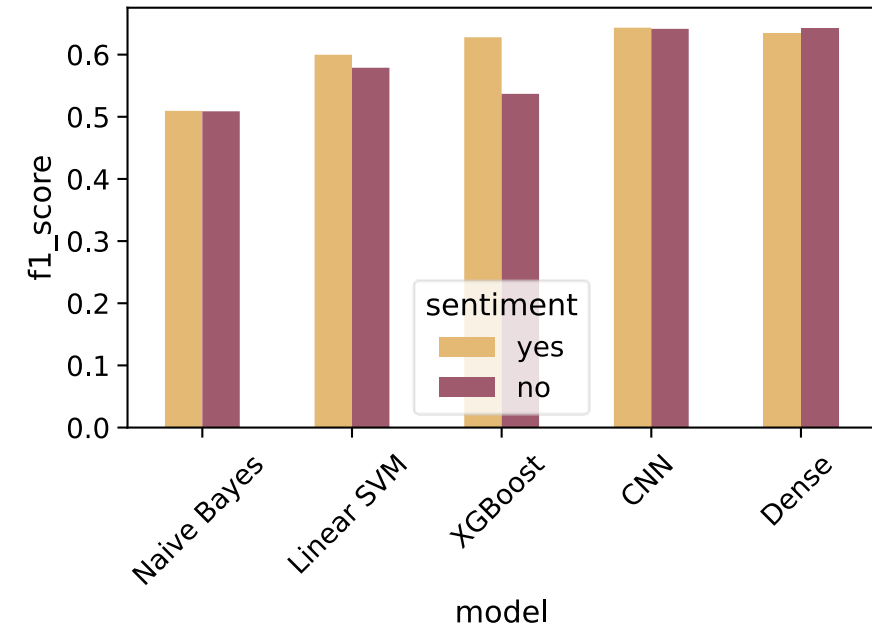
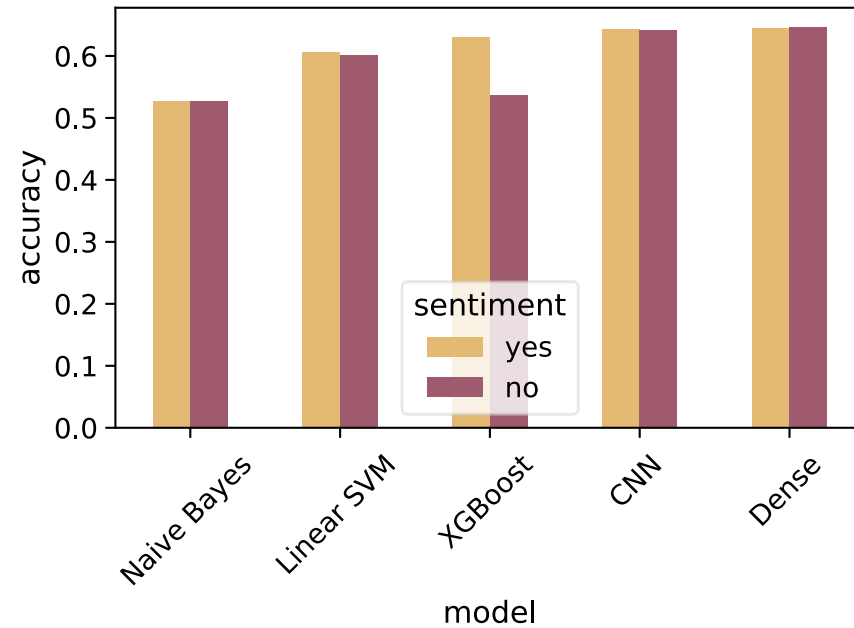
---

- Based on *bert\_uncased\_L-2\_H-128\_A-2* (*tiny BERT*)
- Relatively small model due to GPU memory limitations
- Sentiment defined as 6 dimensional vector
  - Happy, sad, neutral, angry, excited, frustrated

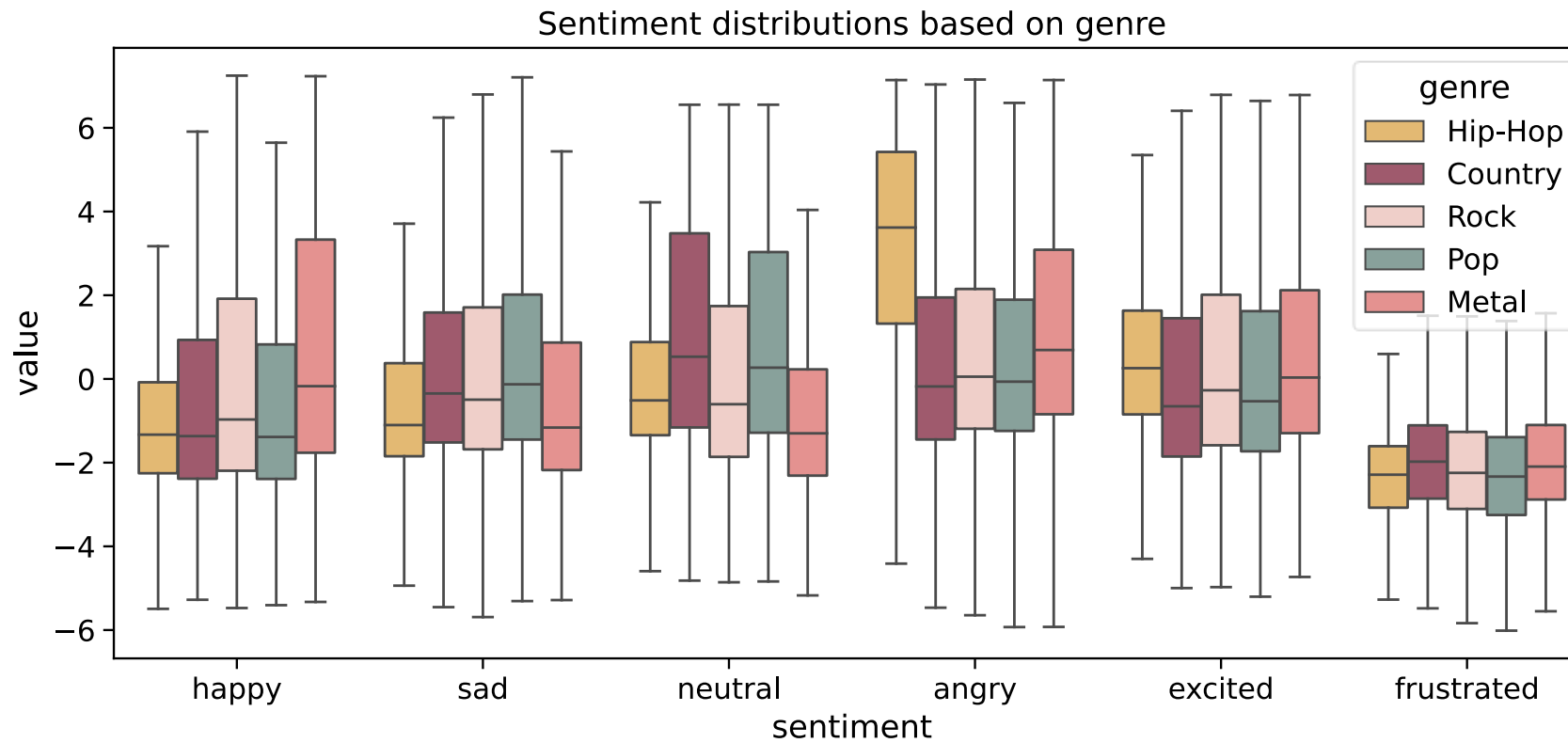


# Sentiment - results

---

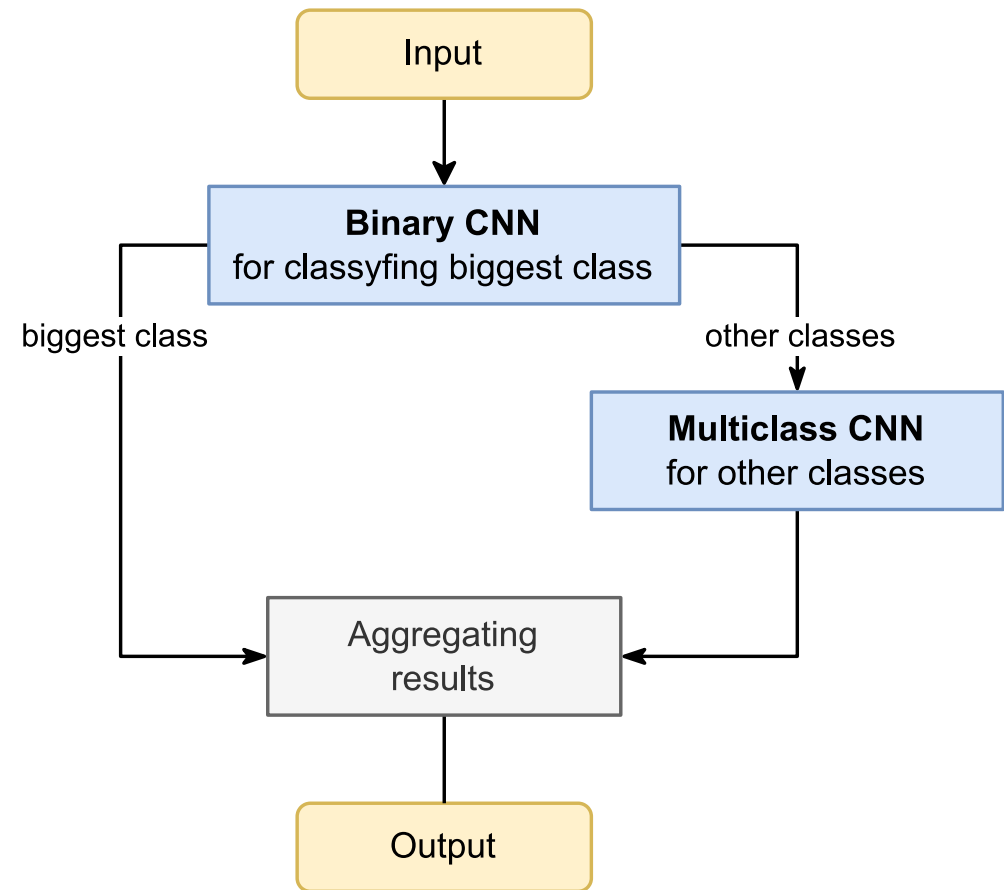


# Sentiment - further inspections

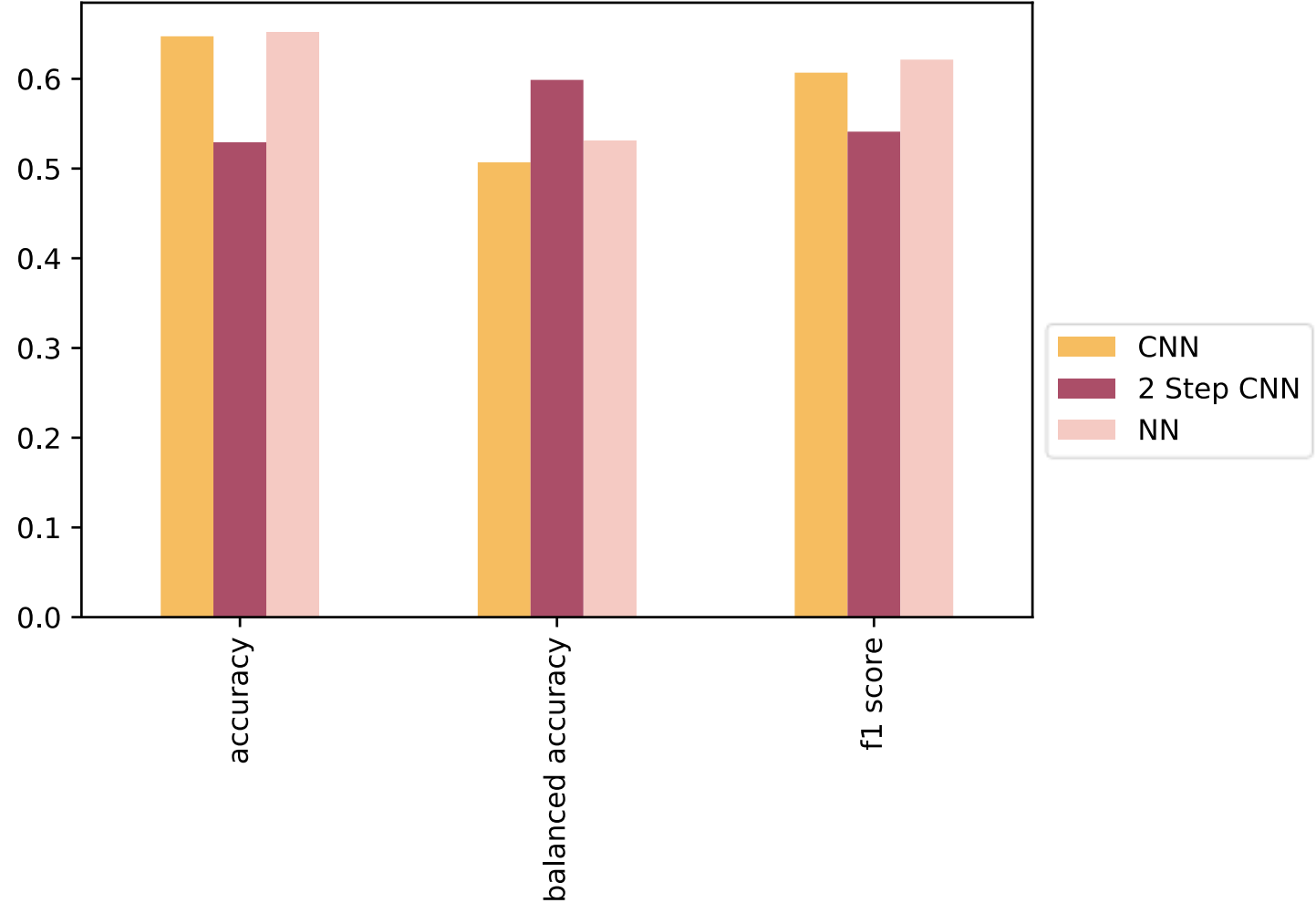


# Two-step classification

- In previous tests we had problems with highly unbalanced main dataset
- *Rock* genre has 2 times more cases than all other genres altogether
- Therefore, we created a new classifier, which concentrates on separating this one genre



# Main dataset results using DistilBERT

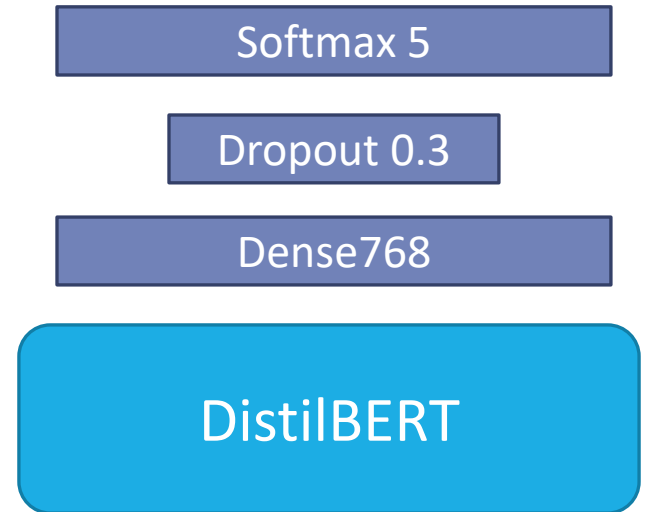


Classifier	Accuracy	Balanced acc.	F1 score
CNN	64,74%	50,71%	60,88%
NN	<b>65,23%</b>	53,08%	<b>62,14%</b>
2-step CNN	52,94%	<b>59,88%</b>	54,11%

# Fine-tuning – process

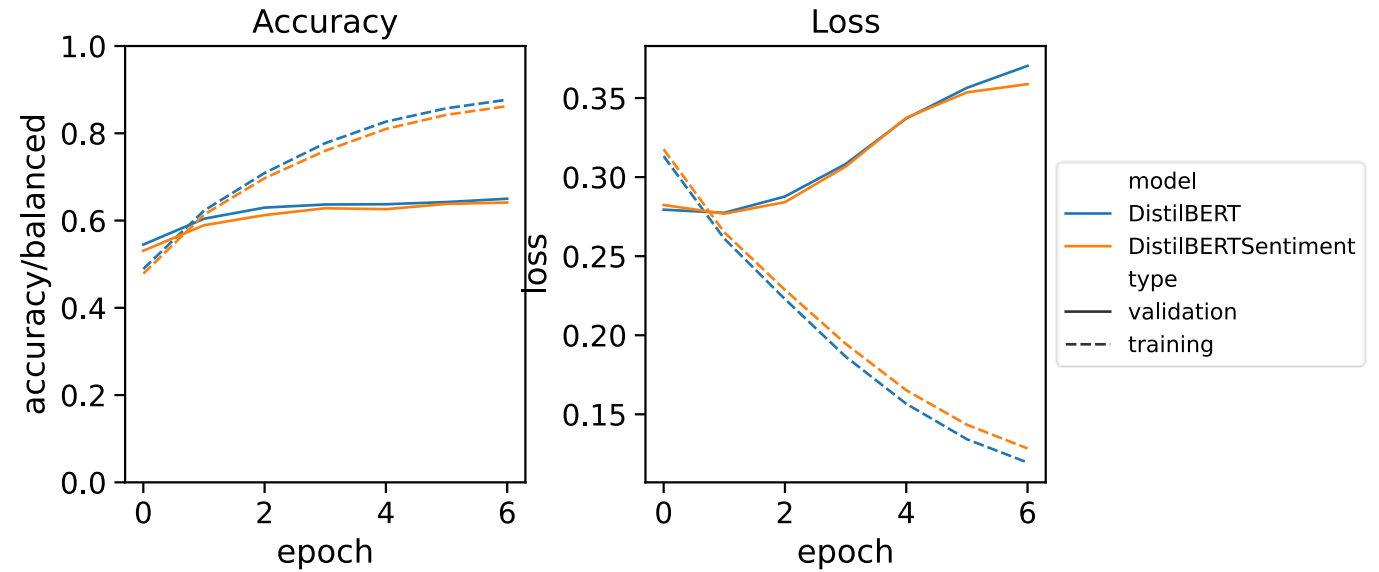
---

- We tested DistilBERT with 2 dense layers for classification
- Same model and training parameters
- Embeddings from one model were concatenated with sentiment vector



# Fine-tuning – results

- Almost the same models
- Same training parameters



Sentiment	Accuracy	F1 score
yes	0.6412	0.6638
no	0.65	0.6682

# Our dataset

---

- Created using Spotify API and Genius API
- Number of observations: 4092
- 5 genres: pop, rock, metal, country, hip-hop
- Quite balanced

Genre	Number of observations
country	896
metal	887
pop	815
rock	767
hip-hop	727

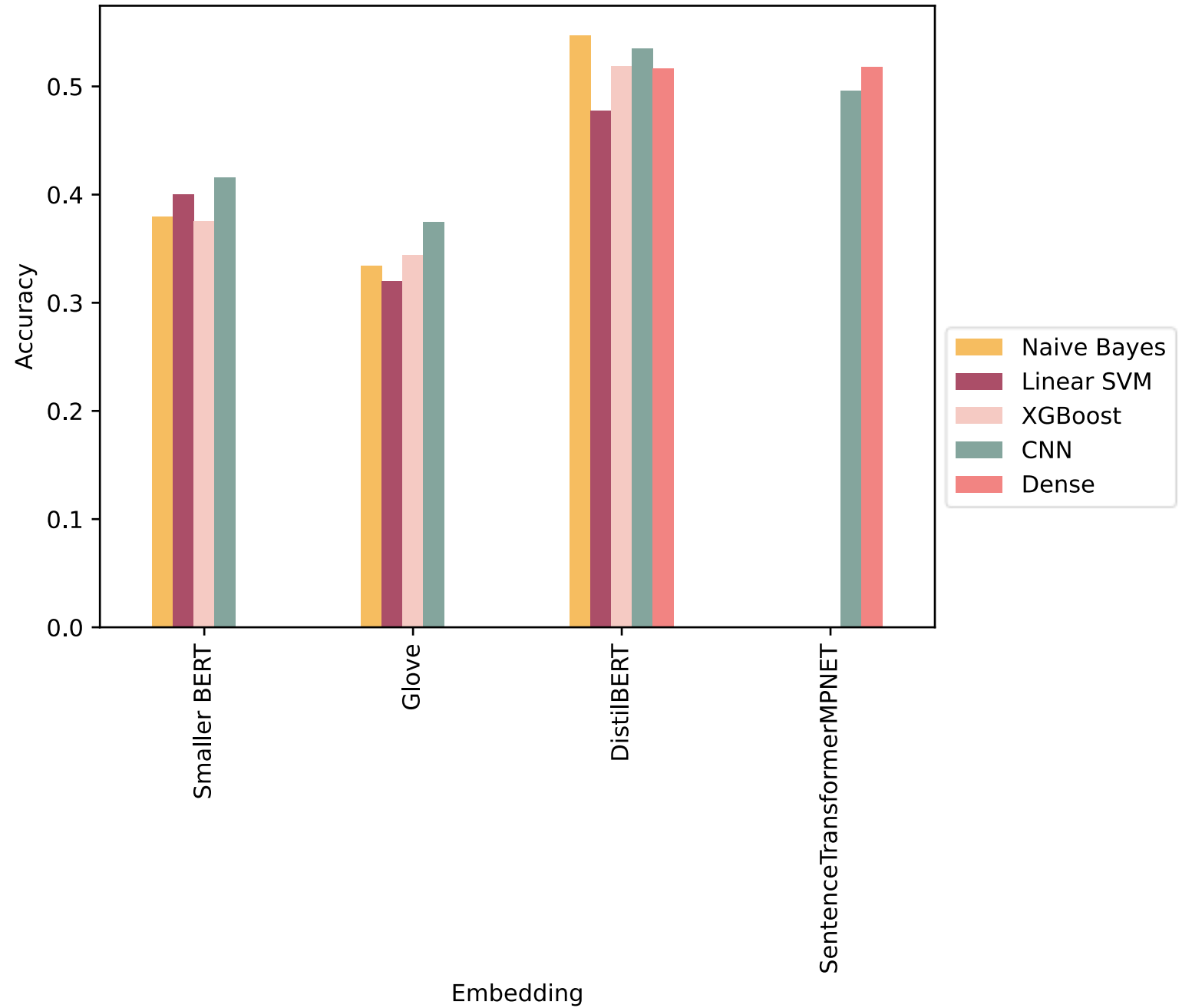
# Our dataset creation – problems

---

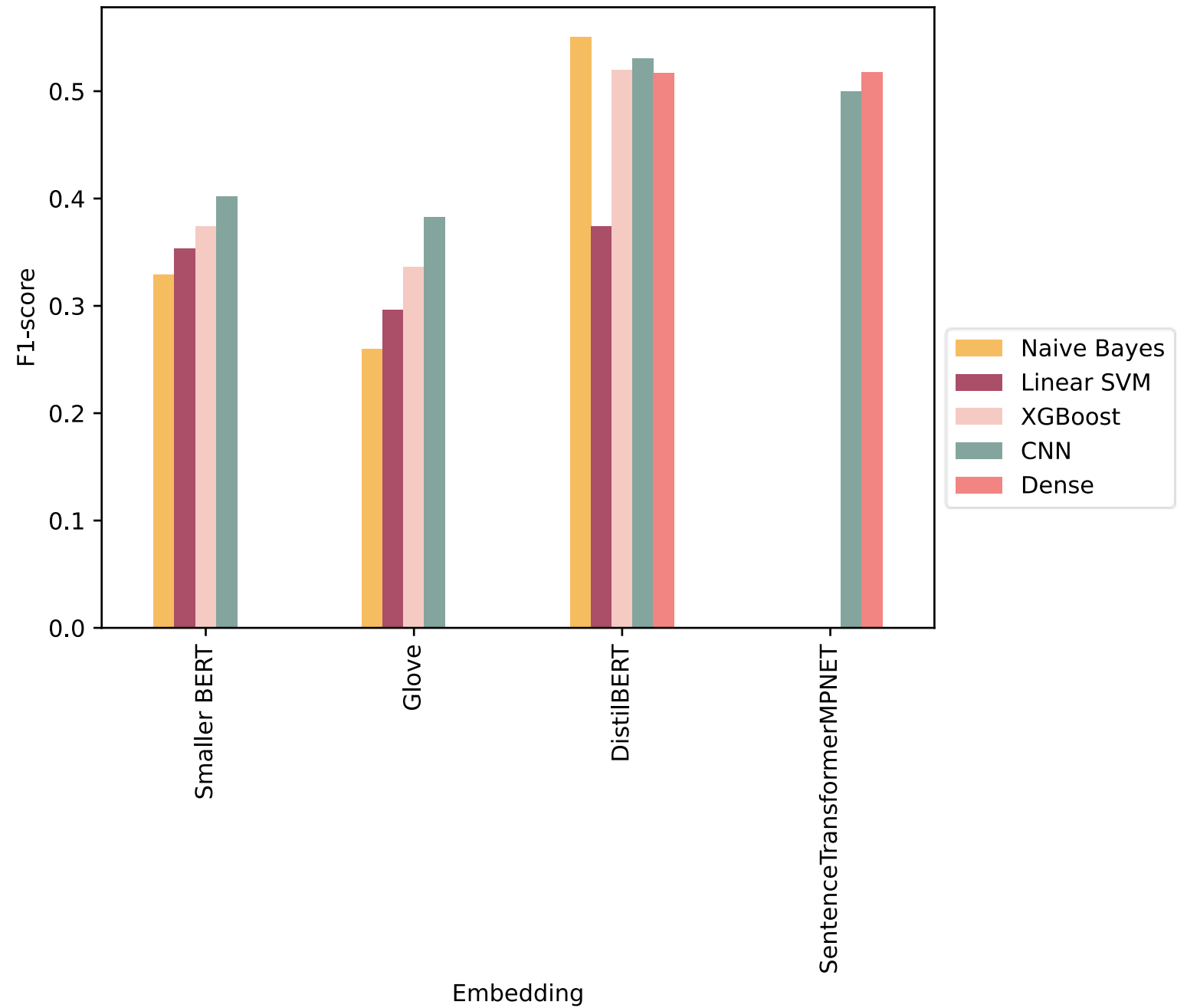
- Recommendations given by Spotify repeat a lot
- By searching for a specific genre only 1000 songs can be returned
- Language of the lyrics cannot be chosen, only the localization that the songs have to be available in
- Genius does not have all of the lyrics
- Even if Genius returns the output there are a lot of cases that not lyrics but strange lists of song titles or artists are returned
- Manual data cleaning is needed



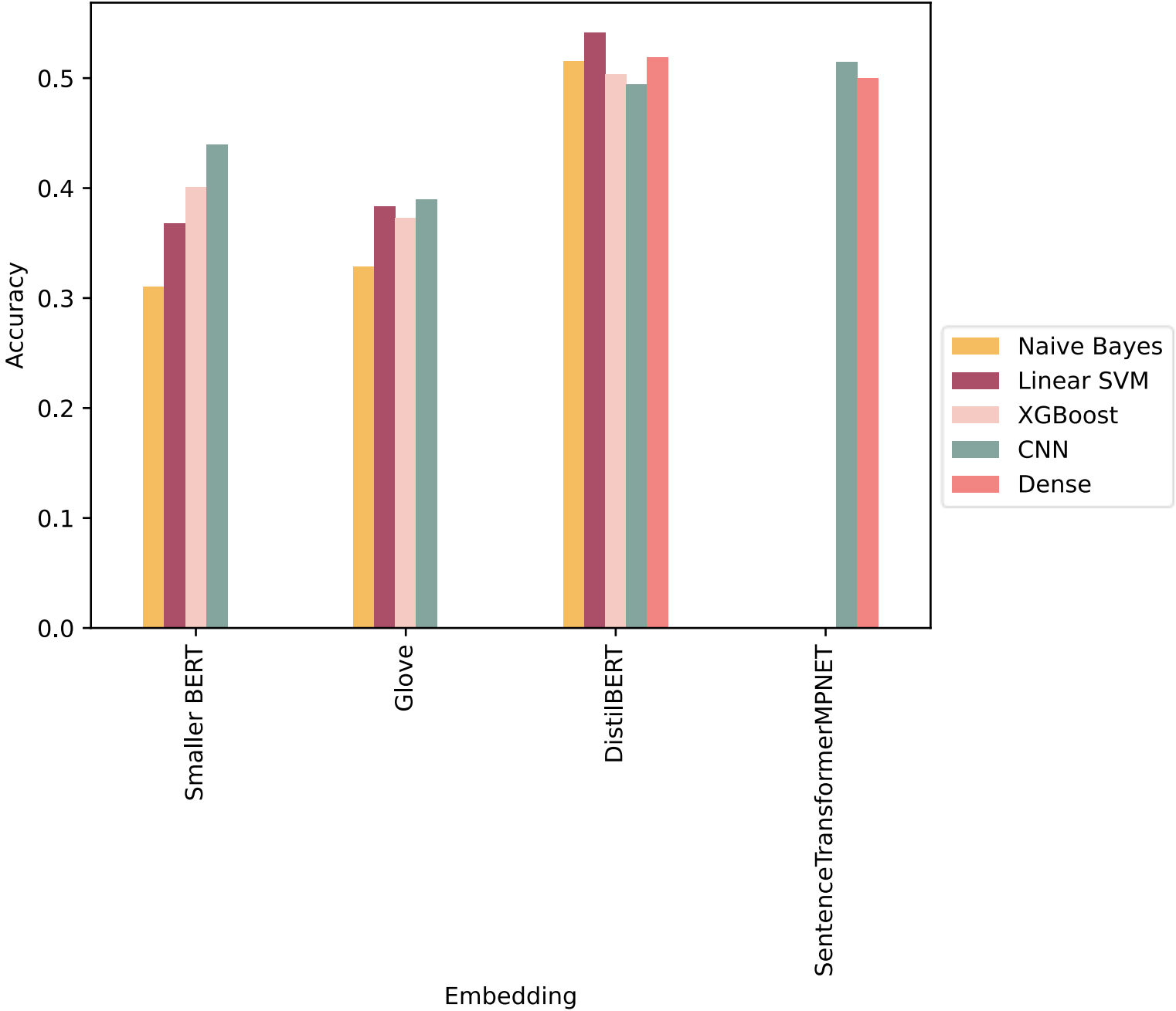
## Accuracy for unnormalized lyrics



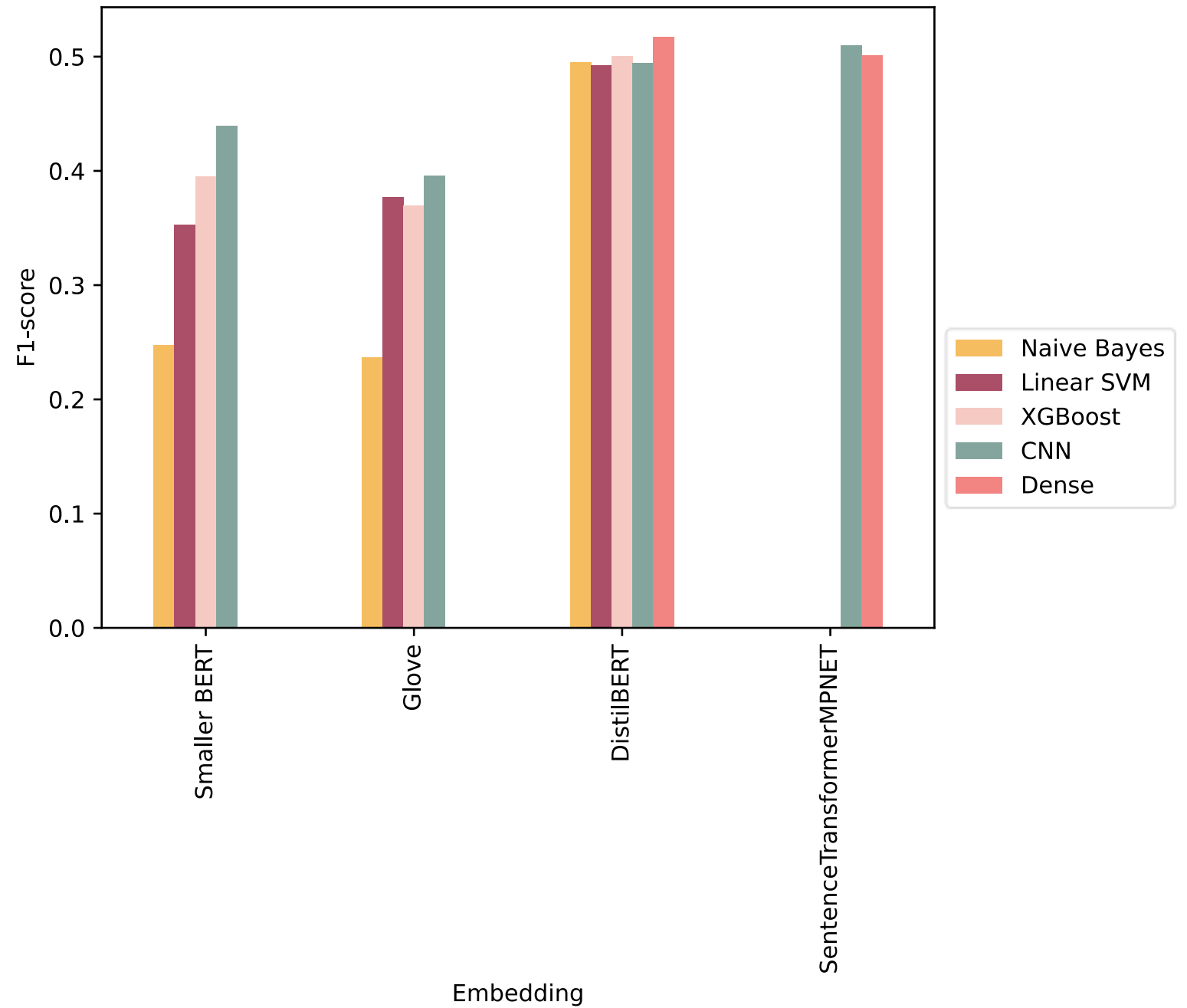
# F1-score for unnormalized lyrics



# Accuracy for normalized lyrics



# F1-score for normalized lyrics



# Conclusions

---

- More complex embedding models seem to be the key for improving classification results
- Surprisingly fine-tuning pretrained model does not bring significant improvements
- 2-step classifier gave better results when it comes to balanced accuracy for highly unbalanced dataset but besides that provided worse results in most cases
- In the case of small datasets simpler classifiers may give better results and they are faster than training e.g. CNNs for multiple epochs

Thank You For Attention!

---