MUSIC GENRE CLASSIFICATION BASED ON SONG LYRICS

COMPARISON BETWEEN DIFFERENT WORD EMBEDDING TECHNIQUES AND CLASSIFIERS

Authors: Bartłomiej Eljasiak, Aleksandra Nawrocka, Dominika Umiastowska
Supervisor: Anna Wróblewska

Goal of the project

We want to test different architectures for music genre classification task by:

- testing different word embedding methods,
- testing various classification models,
- researching influence of emotions contained in lyrics on this task.

RELATED WORKS

Bibliography can be found in our project proposal

Datasets

Song lyrics from 79 musical genres

- 379 893 song lyrics from 4239 artists
- around 50% of the song lyrics in English
- preprocessing needed

MetroLyrics

- 362 237 song lyrics from 18231 artists
- around 60% of the song lyrics in English
- preprocessing already done by students of the University of California

Preparation of our own dataset

- Spotify API
- Genius API

Embeddings

- Continuous Bag-of-Words
- GloVe
- word2vec
- BERT

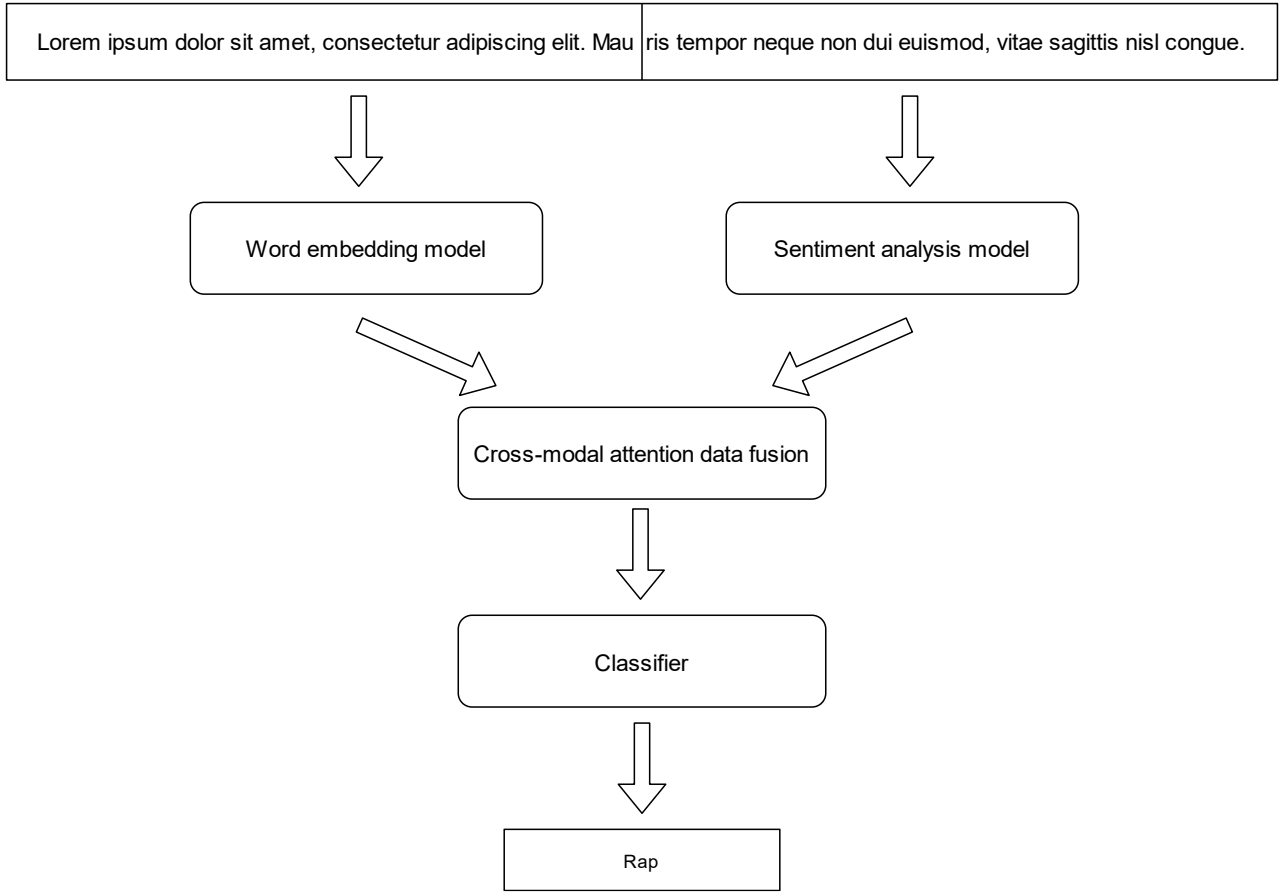
Classification models

- Naive Bayes
- Support Vector Machine
- XGBoost
- Convolutional Neural Networks

Emotion English DistilRoBERTa-base

- emotion detection pre-trained model
- based on DistilRoBERTa-base
- fine-tuned on balanced dataset with 20k observations in total
- predicts Ekman's six basic emotions and a neutral class:
  - anger 😠
  - disgust 🤢
  - fear 😨
  - joy 😀
  - sadness 😭
  - surprise 😲
  - neutral 😐

Modified approach using sentiment analysis model and data fusion technique



Risk analysis

Risk	Consequence	Mitigation
not enough time for performing the project because of other obligations	late submission of the project or not finishing it	reducing the project scope by testing smaller numbers of models/techniques
not big enough computational resources to conduct proper experiments	lower quality of conducted experiments	-

THANK YOU FOR ATTENTION!