

# The Genre Factor

## Project Presentation - ML Seminar 2023

---

Henry Krämerkämper

Christopher Breitfeld

13.07.2023

Technische Universität Dortmund

# Description of the Data Set

## Dataset from Kaggle: Spotify and YouTube

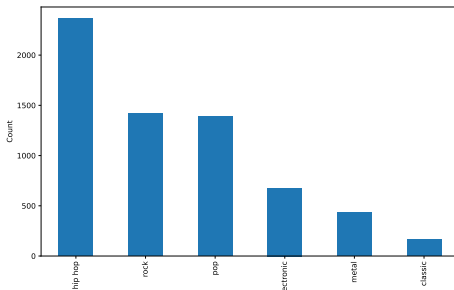
- Contains statistics of songs on Spotify and YouTube
- 20,7 k entries by 2 k artists
- Does NOT include the genre

## Wikidata Query for the Top-Genre of the Artist

- Query artist's Wikidata page for genre names
- Assign artists/album genre to song

## Selection

- Grouped Subgenres into Overcategories
- Select sample of 6 Genres
- Remaining Songs: 6446



## Model

- 4 Dense Layers with Dropout
- Loss function: categorical crossentropy
- Optimizer: adam

Layer (type)	Output Shape	Param #
dense_4 (Dense)	(None, 64)	1280
dropout_3 (Dropout)	(None, 64)	0
dense_5 (Dense)	(None, 256)	16640
dropout_4 (Dropout)	(None, 256)	0
dense_6 (Dense)	(None, 256)	65792
dropout_5 (Dropout)	(None, 256)	0
dense_7 (Dense)	(None, 6)	1542
Total params: 85,254		
Trainable params: 85,254		
Non-trainable params: 0		

## Training

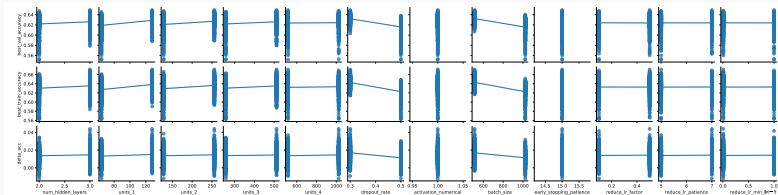
- Early stopping: Stops training when the validation loss function no longer improves
- Reduce learning rate: Decreases learning rate if validation loss function stagnates → better convergence
- Train the model using the training data with the defined set of hyperparameters.

## Method

- Grid Search
- Train models with all combinations of hyperparameters

## Validation

- $k = 3$  Cross Validation
- save train/validation Acc. and Loss



## Methods to prevent Overtraining

- Dropout
- Early stopping
- minimize Training Acc. - Validation Acc.  
but maximize Validation Acc.

Try different values and decide after  
Hyperparameter Optimization