

# 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

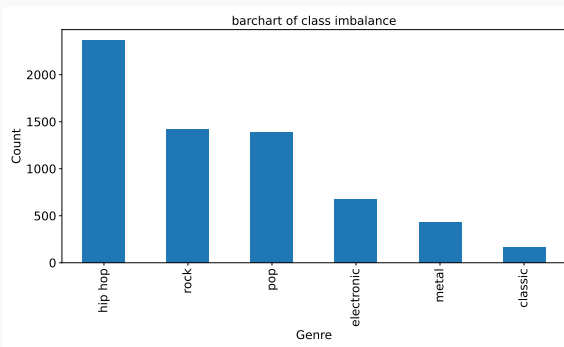
- Contains statistics of 20,7 k songs on Spotify and YouTube
- Does **NOT** include genre information.

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

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

## Selection

- Groupe Subgenres into Supercategories
- Select sample of 6 Genres
- Remaining Songs: 6446

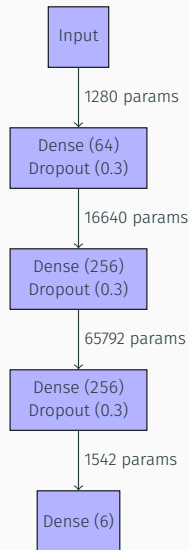


## Model

- 4 Dense Layers with Dropout
- Trainable parameters: 85 254
- Loss function: categorical crossentropy
- Optimizer: adam

## 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.



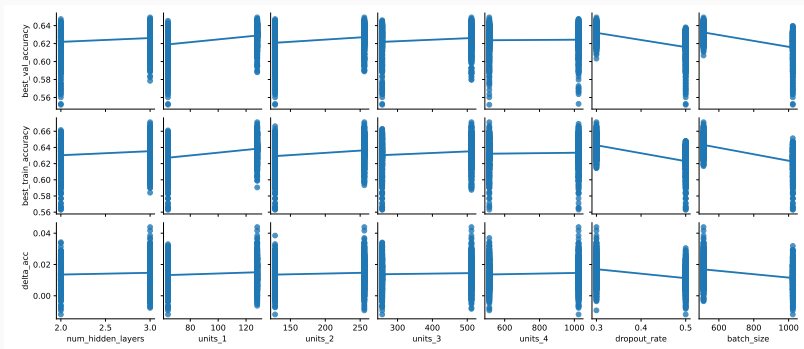
# Hyperparameter Optimization

## Method

- Grid Search
- Train models with all combinations of hyperparameters

## Validation

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

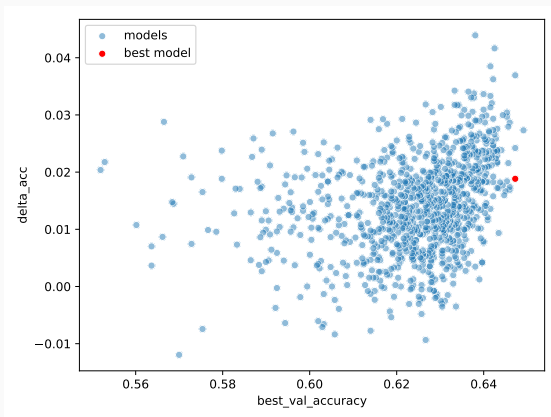


# Overtraining Checks

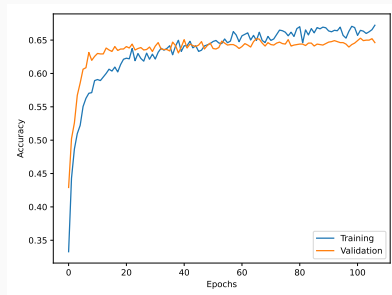
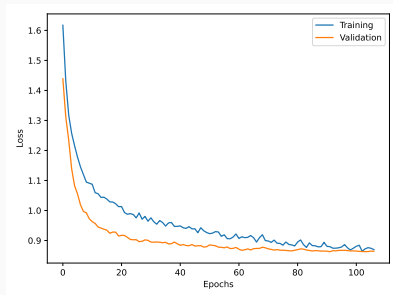
## Methods to prevent Overtraining

- Dropout
- Early stopping
- minimize (Training Acc. - Validation Acc.) **but** maximize Validation Acc.

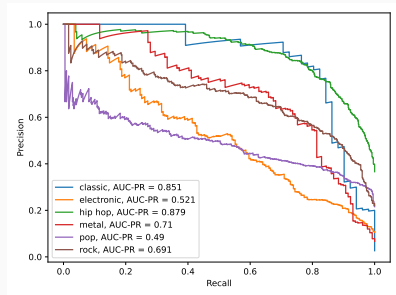
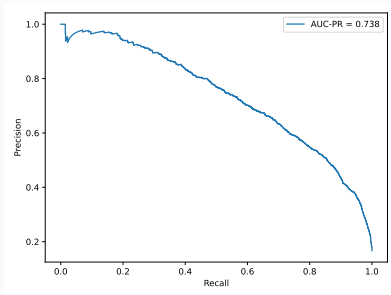
Try different values and decide after Hyperparameter Optimization



# Appendix: Accuracy and Loss



# Appendix: Precision-Recall Curve



## Appendix: Substructure of Hip Hop

