Recognize Genre of a Song Using Spotify Web API

MLWR Project SoSe - 2022
<u>Group: 07</u>

Group Members:

Others and
Rubaiya Kabir Pranti

Machine Learning and Scientific Computing

Technology
Arts Sciences
TH Köln

Contents

Introduction

- The aim of this project is to recognize genre of the song by applying different Machine Learning Models
- The dataset used for the project is retrieved from Spotify web API
- Data preprocessing and feature Engineering were performed on the data to get the final dataset, which was then split into training and testing dataset in the ratio of 8:2
- The Machine Learning Models used for our project are:
  - Logistic Regression
  - K-Nearest Neighbors
  - Random Forest Model
  - Decision Tree Model
  - Gradient Boosting Model

Methodology

- Spotify Web API and Data Extraction
- Data Pre-processing
- Feature Engineering
- Splitting Dataset
- Classification-Training Models
- Evaluation through Cross Validation
- Hyperparameter Optimization
- Final Prediction

Spotify Web API and Data Extraction

- Utilized the Spotify music data set using Spotify's [https://developer.spotify.com/documentation/web-api/ website](https://developer.spotify.com/documentation/web-api/)

- achieved a total of **150,550 songs** after first audio feature extraction

- after combining the extracted metadata set and extracted feature dataset, the combined dataset was of total **366,684 data points** or songs consisting of **126 genres** and each has **26 features/columns**

- for less complexity and because of high computation time, we decided to keep only **15 genres** in which each genre has equal or more than 850 songs.

Technology Arts Sciences TH Köln

Key audio dataset metadata and features

- Danceability
- Popularity
- Mode
- Speechiness
- Acousticness
- Instrumentalness
- Liveness
- Valence
- Tempo
- Duration
- Key
- Explicitness

```
df.head()
```

	genre	popularity	explicit	danceability	energy	key	mode	speechiness	acousticness	instrumentalness	liveness	valence	tempo	duration_s
0	salsa	57	0	0.463	0.876	1	1	0.0468	0.32200	0.000001	0.4090	0.569	92.092	261.0
1	detroit-techno	7	0	0.775	0.838	10	0	0.0616	0.00676	0.900000	0.0835	0.671	139.981	422.0
2	tango	12	0	0.534	0.274	11	1	0.0527	0.91800	0.594000	0.0798	0.687	126.114	185.0
3	detroit-techno	9	0	0.789	0.458	1	1	0.0783	0.00990	0.853000	0.1300	0.419	133.035	265.0
4	salsa	65	0	0.688	0.614	9	1	0.0378	0.58500	0.000000	0.0978	0.852	176.396	307.0

Technology Arts Sciences TH Köln

Preferred Genre List with 15 genres only

1. Study
2. Club
3. Comedy
4. Bluegrass
5. Black-metal
6. World-music
7. Iranian
8. Grindcore
9. Heavy-metal
10. Turkish
11. Afrobeat
12. Forro
13. Country
14. Happy
15. Chicago-house

Data Pre-processing

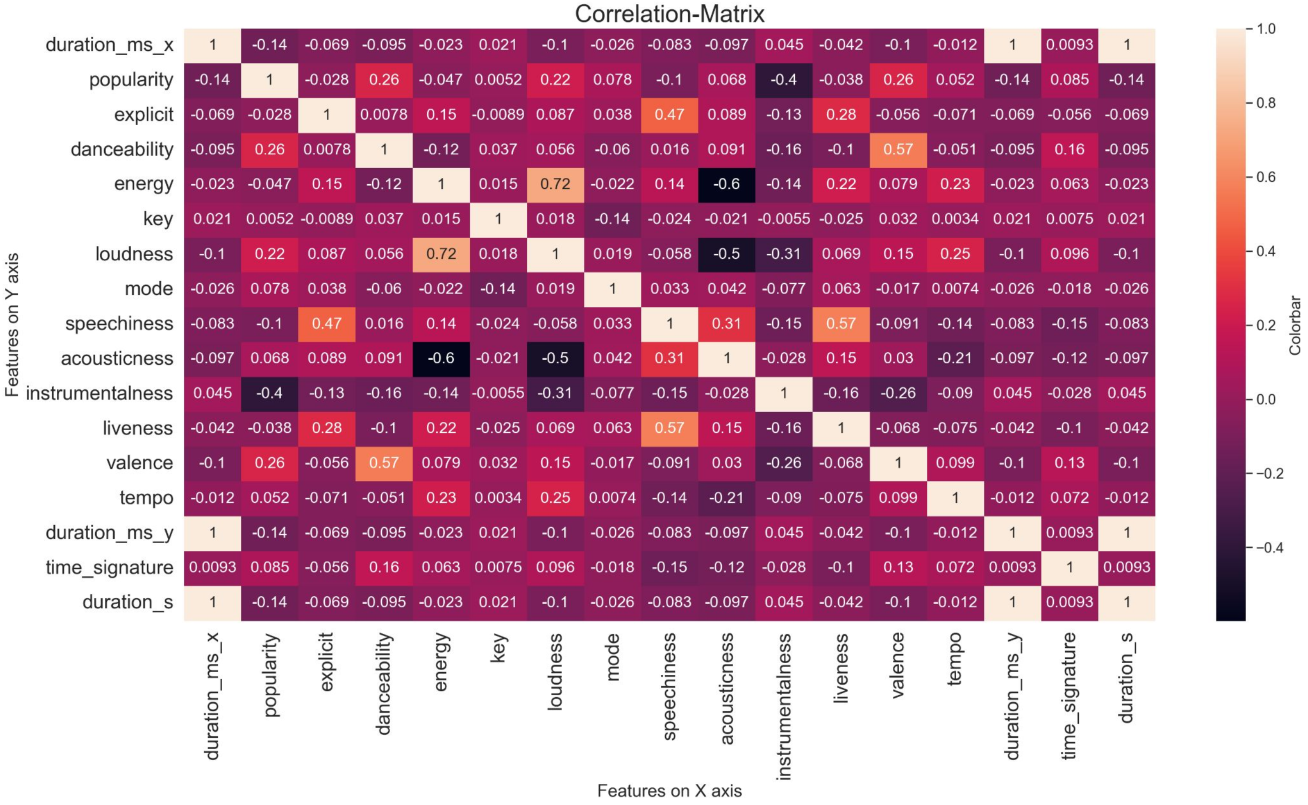
- Deleted rows with missing values
- Erased rows with duplicate data
- Replaced boolean values with numericals
- Took desired genres from 126 to 15 genres

Feature Engineering

- Dropped non-numeric or insignificant columns
- highly correlated feature loudness was dropped using correlation matrix
- After feature analysis, 15 features were kept finally

Technology
Arts Sciences
TH Köln

Correlation Matrix



Technology Arts Sciences
TH Köln

Splitting Dataset

- The dataset was splitted into training and testing dataset in the ratio of **8:2** respectively
- Training set is used to train the machine learning model to predict
- Testing set is used to measure the performance

Classification-Training Models

- Training is performed
- Data fitting is executed to fit the models

Evaluation through Cross Validation

- For better evaluation, Cross validation is performed to get the estimation of the performance of models through accuracy score function
- Specifically, K-fold cross validation on both training and test dataset is applied to get our initial and final predictions by shuffling the dataset randomly and by splitting into k number of folds

Hyperparameter optimization

- Manual search/investigation of the best possible hyperparameters
- Through Scikit-Learner's GridSearchCV and RandomizedSearchCV (only applied to Random Forest), models are optimized or retrained by applying optimal/best hyperparameters

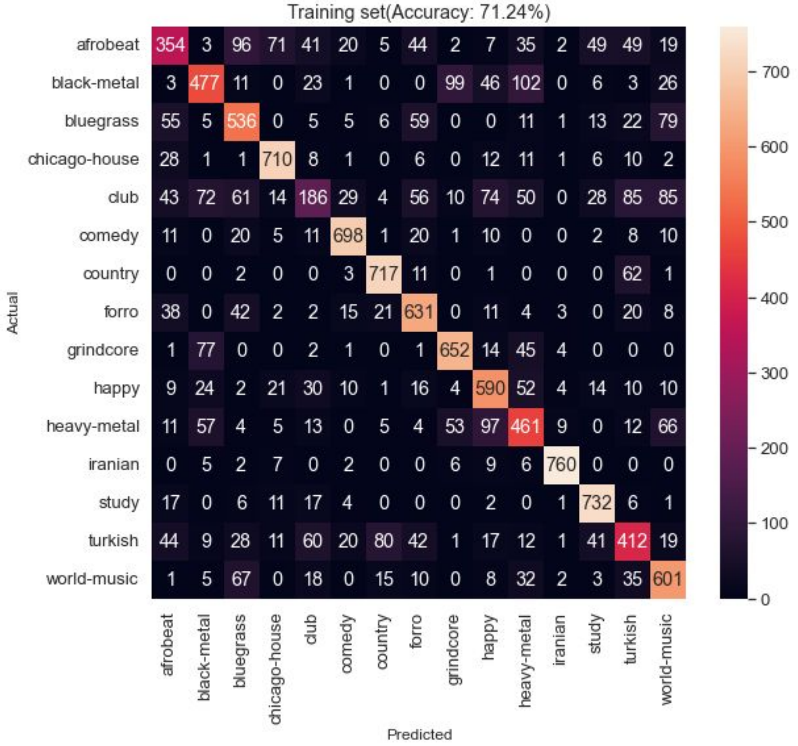
Logistic Regression

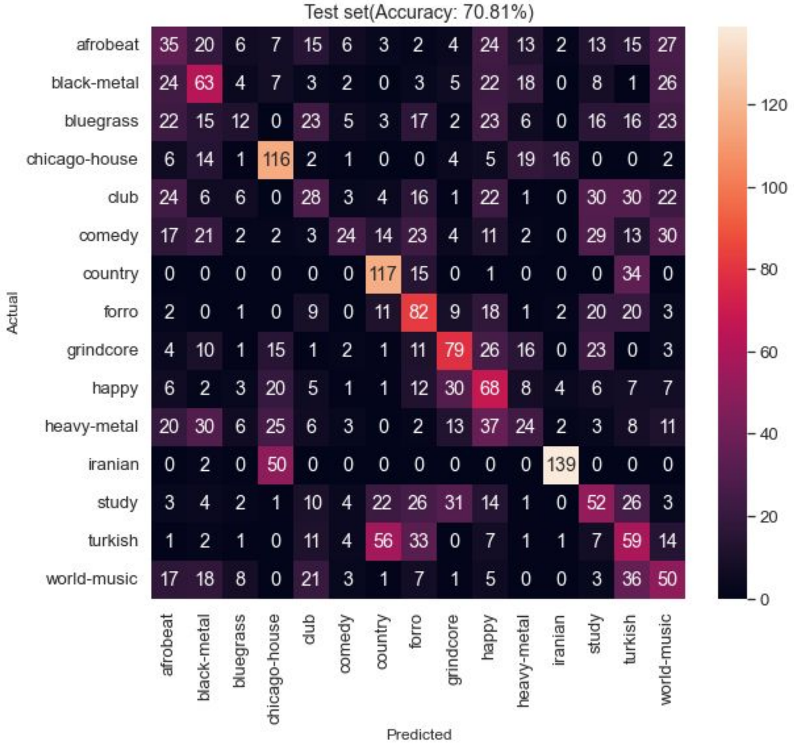
<u>Steps performed to get optimized LR model:</u>

- used **SMOTE()** function
- "**popularity**" and "**tempo**" having large integer values were scaled **StandardScaler()** function
- included **PCA()** function
- incorporated grid search optimization also
- cross-validation parameter along with pipeline as estimators and after fine tuning model, the LR model was retrained by providing test set accuracy as **70.81%** which was initially **33.87%.**

Best Hyperparameters	**Values**
C	100
Penalty	l2
multiclass	ovr
solver	newton-cg
pipeline	pca,standard scalar

Technology
Arts Sciences
TH Köln





Technology Arts Sciences TH Köln

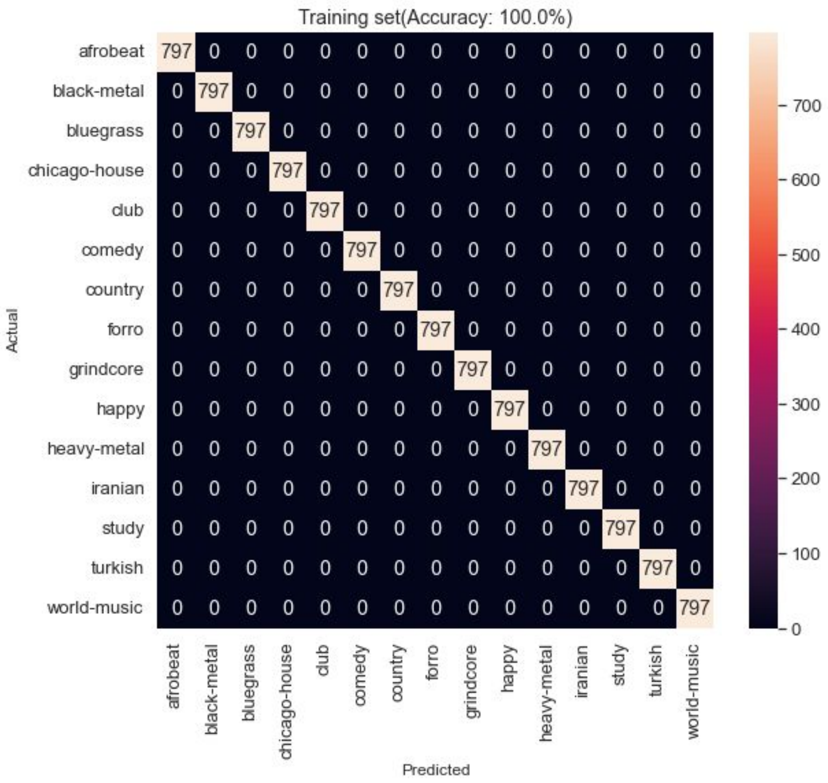
K-Nearest Neighbor

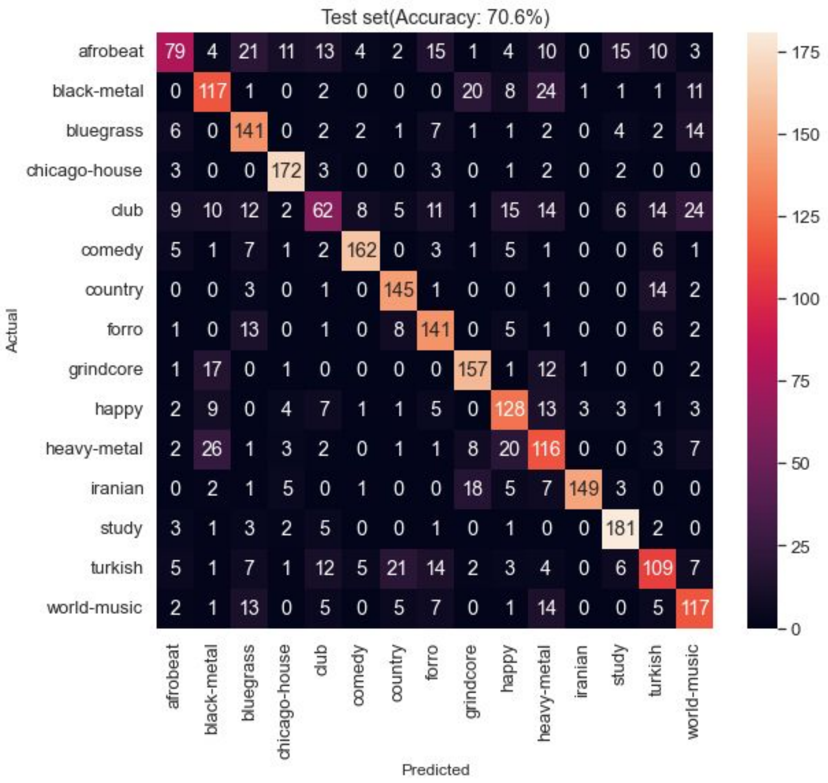
Steps performed to get optimized KNN model:

- If the number of nearest neighbors=**k is too small**, then it results in the decrease of accuracy rate
- the experimental results show that initial value of k is 21, where accuracy rate reached up to **44.68%**.
- after applying smote function, respective pipeline same as LR classification model and
- applying tuning of model through grid search;the K-NN model gave accuracy of about **70.6%**.

Best Hyperparameters	Values
Leaf size	30
N neighbors	16
weights	distance
metric	minkowski
pipeline	pca,standard scalar

Technology Arts Sciences
TH Köln



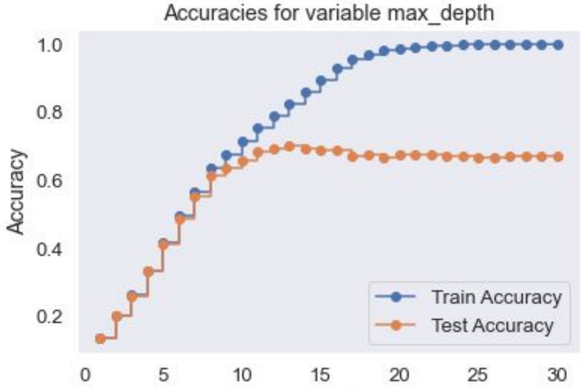


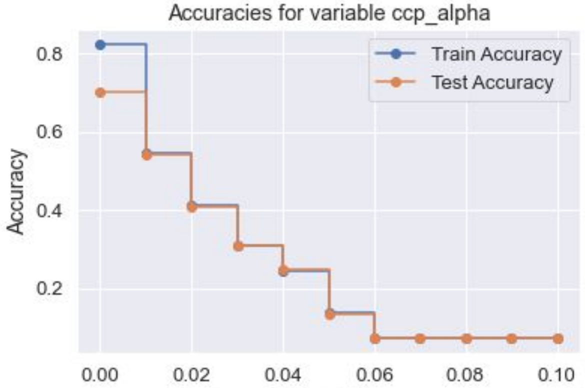
Technology Arts Sciences TH Köln

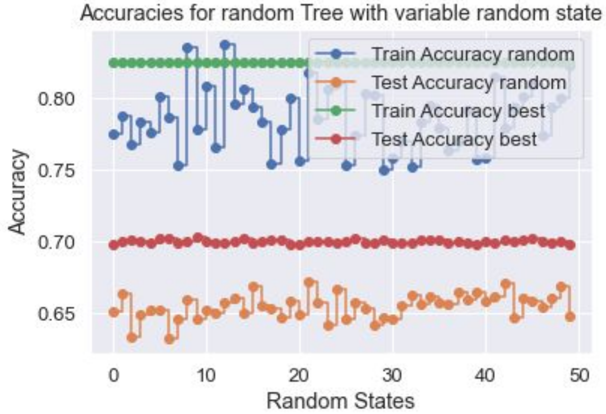
Decision Tree

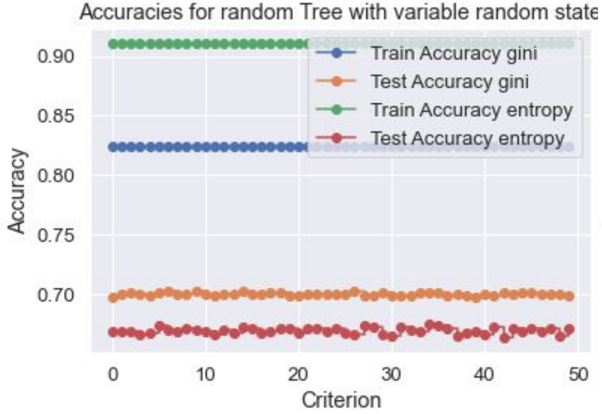
- Observation of DTs performance on our case.
- Initial Tree accuracy: 67%
- For optimization: Hyperparameter tuning (manually)
    - Maximum depth
    - Ccp_alpha (pruning)
    - Minimum samples per split
    - Criterion (Entropy vs. Gini-impurity)
    - Splitter ("Best" vs. Random)
- GridSearch

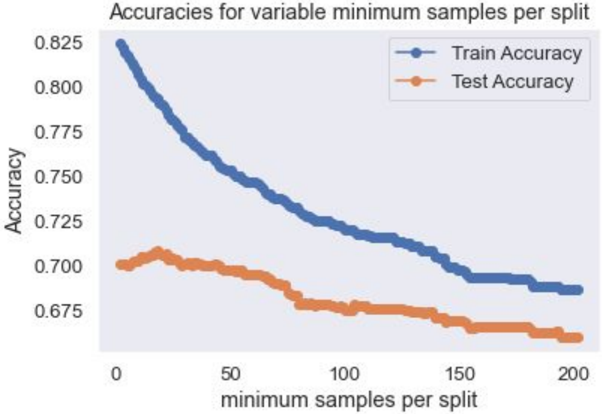
DT Hyperparameter tuning











Technology Arts Sciences
TH Köln

Optimal values for observed hyperparameters

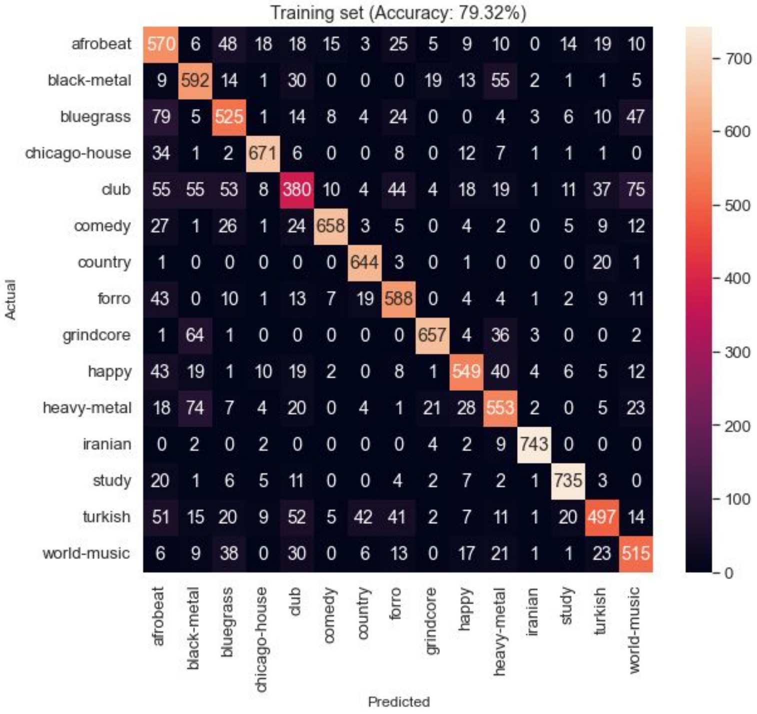
Hyperparameter	Value (tuning)	Value (gridSearch)
Max. depth	13	13
Ccp_alpha	0 (no pruning)	0
Min. samples / split	18	33
splitter	best	best
criterion	gini	gini

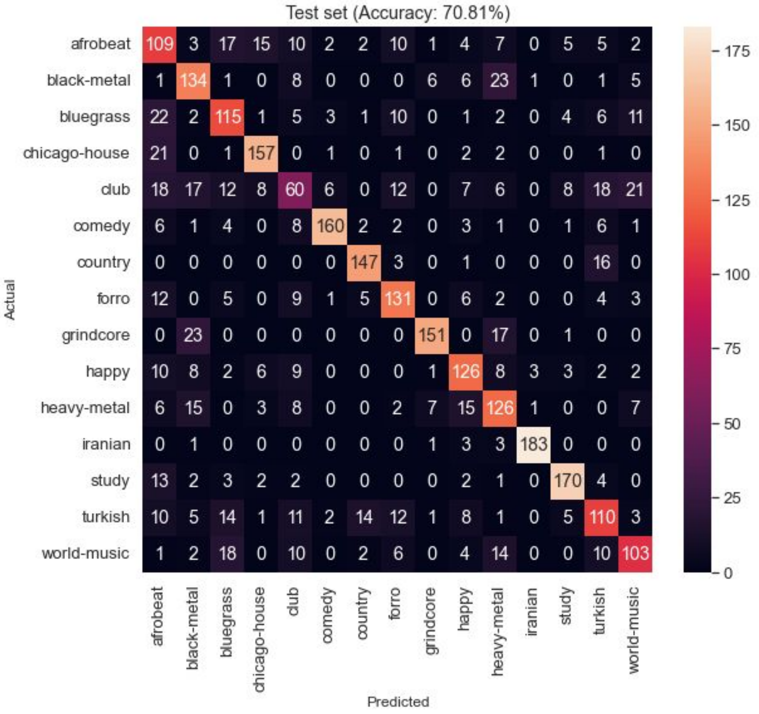
Tuning accuracy: 70.81%

Gridsearch accuracy: 70.03%

Technology Arts Sciences TH Köln

Manual tuning





Technology Arts Sciences TH Köln

GridSearch

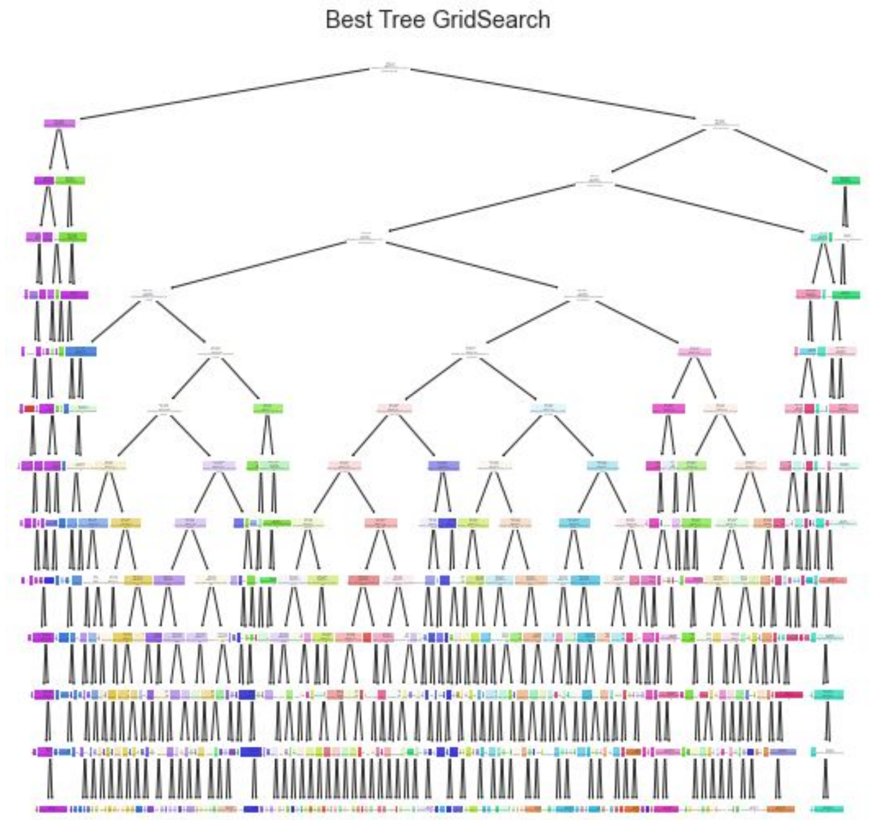
Training set (Accuracy: 76.92%)

Actual \ Predicted	afrobeat	black-metal	bluegrass	chicago-house	club	comedy	country	forro	grindcore	happy	heavy-metal	iranian	study	turkish	world-music
afrobeat	570	6	48	18	18	15	3	25	5	9	10	0	14	19	10
black-metal	9	592	14	1	30	0	0	0	19	13	55	2	1	1	5
bluegrass	79	5	525	1	14	8	4	24	0	0	4	3	6	10	47
chicago-house	34	1	2	671	6	0	0	8	0	12	7	1	1	1	0
club	55	55	53	8	380	10	4	44	4	18	19	1	11	37	75
comedy	27	1	26	1	24	658	3	5	0	4	2	0	5	9	12
country	1	0	0	0	0	0	644	3	0	1	0	0	0	20	1
forro	43	0	10	1	13	7	19	588	0	4	4	1	2	9	11
grindcore	1	64	1	0	0	0	0	0	657	4	36	3	0	0	2
happy	43	19	1	10	19	2	0	8	1	549	40	4	6	5	12
heavy-metal	18	74	7	4	20	0	4	1	21	28	553	2	0	5	23
iranian	0	2	0	2	0	0	0	0	4	2	9	743	0	0	0
study	20	1	6	5	11	0	0	4	2	7	2	1	735	3	0
turkish	51	15	20	9	52	5	42	41	2	7	11	1	20	497	14
world-music	6	9	38	0	30	0	6	13	0	17	21	1	1	23	515

Test set (Accuracy: 70.03%)

Actual \ Predicted	afrobeat	black-metal	bluegrass	chicago-house	club	comedy	country	forro	grindcore	happy	heavy-metal	iranian	study	turkish	world-music
afrobeat	109	3	17	15	10	2	2	10	1	4	7	0	5	5	2
black-metal	1	134	1	0	8	0	0	0	6	6	23	1	0	1	5
bluegrass	22	2	115	1	5	3	1	10	0	1	2	0	4	6	11
chicago-house	21	0	1	157	0	1	0	1	0	2	2	0	0	1	0
club	18	17	12	8	60	6	0	12	0	7	6	0	8	18	21
comedy	6	1	4	0	8	160	2	2	0	3	1	0	1	6	1
country	0	0	0	0	0	0	147	3	0	1	0	0	0	16	0
forro	12	0	5	0	9	1	5	131	0	6	2	0	0	4	3
grindcore	0	23	0	0	0	0	0	0	151	0	17	0	1	0	0
happy	10	8	2	6	9	0	0	0	1	126	8	3	3	2	2
heavy-metal	6	15	0	3	8	0	0	2	7	15	126	1	0	0	7
iranian	0	1	0	0	0	0	0	0	1	3	3	183	0	0	0
study	13	2	3	2	2	0	0	0	0	2	1	0	170	4	0
turkish	10	5	14	1	11	2	14	12	1	8	1	0	5	110	3
world-music	1	2	18	0	10	0	2	6	0	4	14	0	0	10	103

DT Visualization



Technology Arts Sciences TH Köln

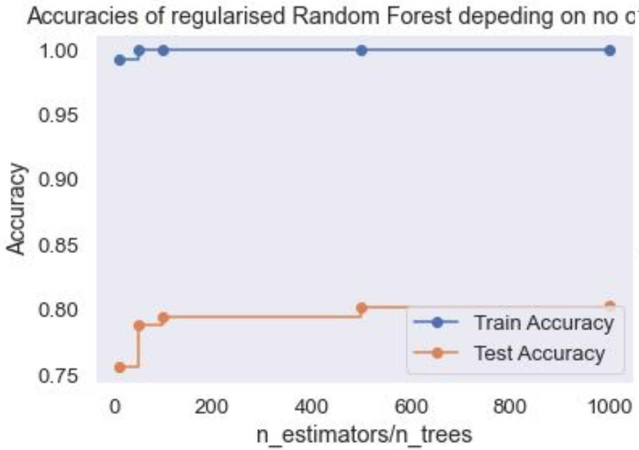
Random Forest

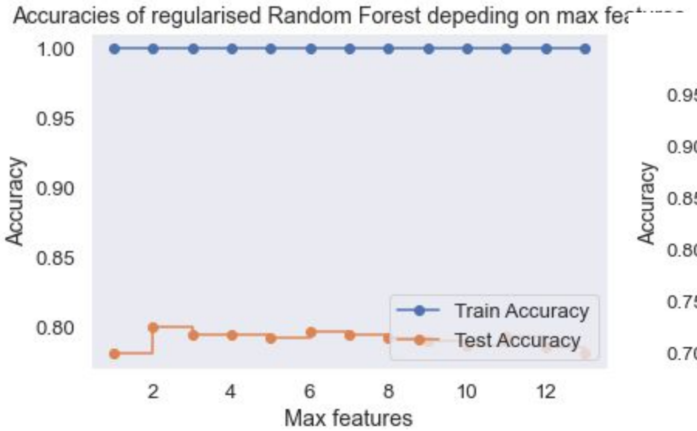
- final parameters used are named as **number of estimators**, **maximum depth**, **maximum features**, **minimum samples per split**, **minimum samples per leaf** etc.
- grid search optimization always comes up with the best possible combination of hyperparameters
- delivers an accuracy of about **79.59%** without hyperparameter tuning. Later accuracy increases up to **79.67%** after tuning.

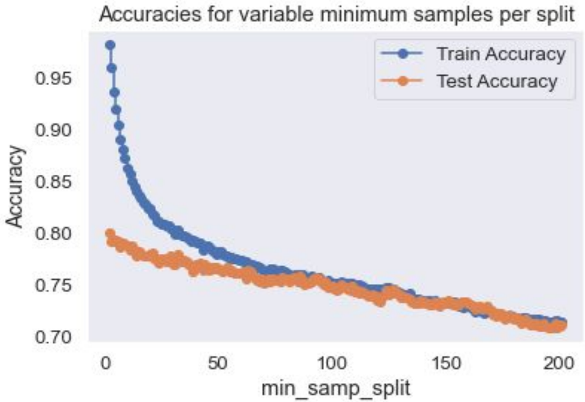
Best Hyperparameters	**Values**
maximum depth	19
maximum features	3
n estimators	1000
criterion	entropy
min. samples leaf	2
min. samples split	3
class weight	balanced_subsample

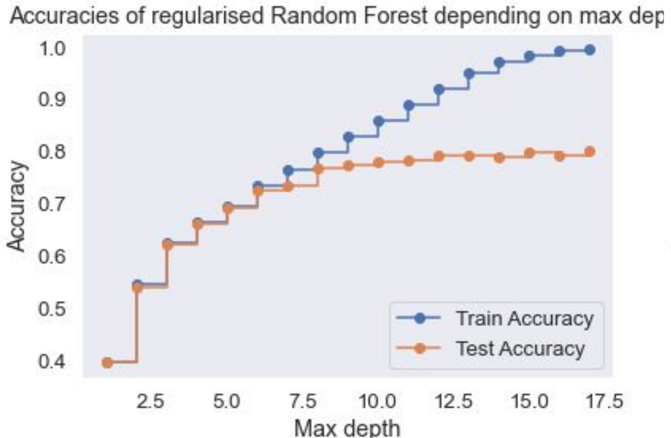
Technology
Arts Sciences
TH Köln

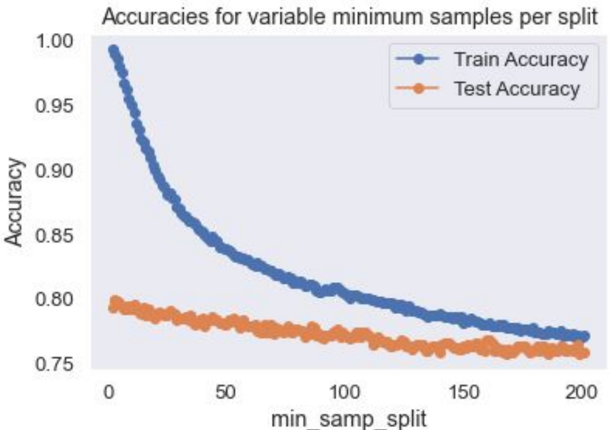
Random Forest Hyperparameter Tuning(individual)



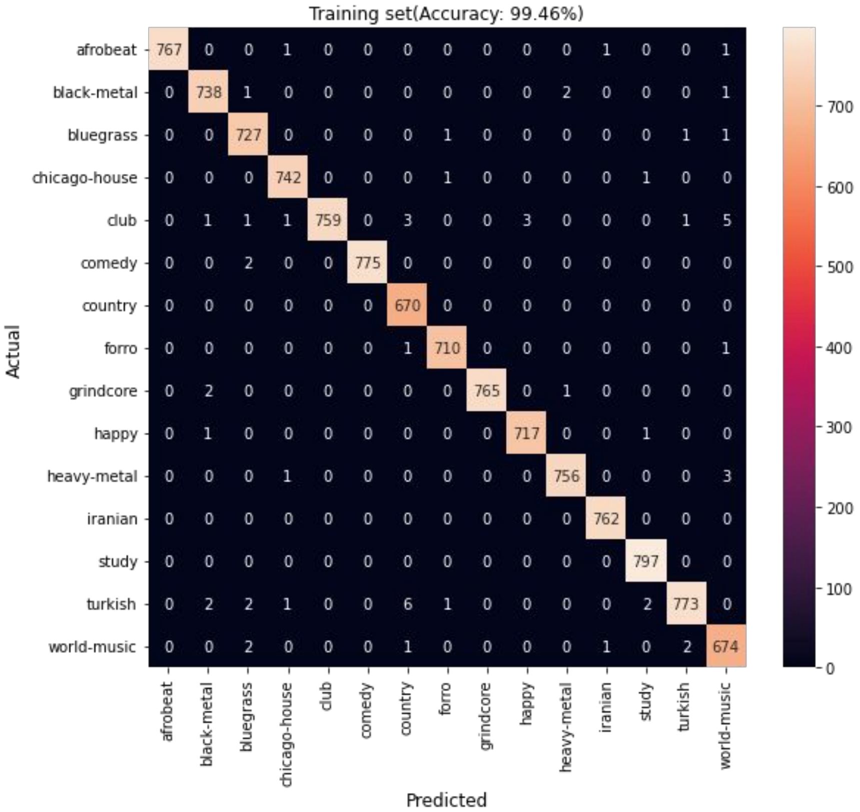


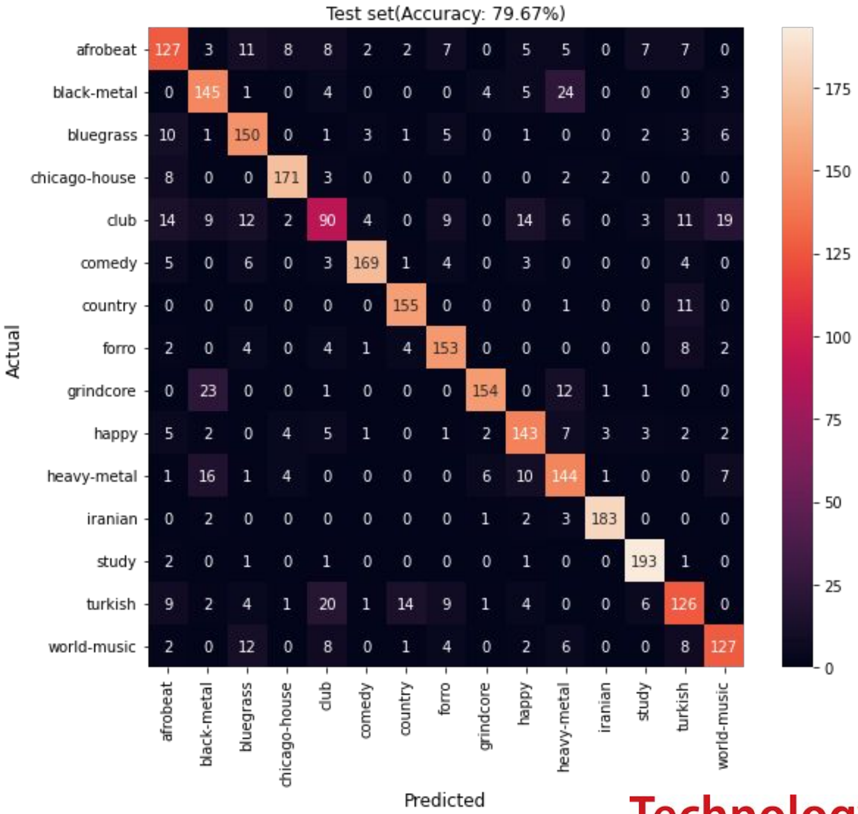






Technology Arts Sciences
TH Köln





Technology Arts Sciences
TH Köln

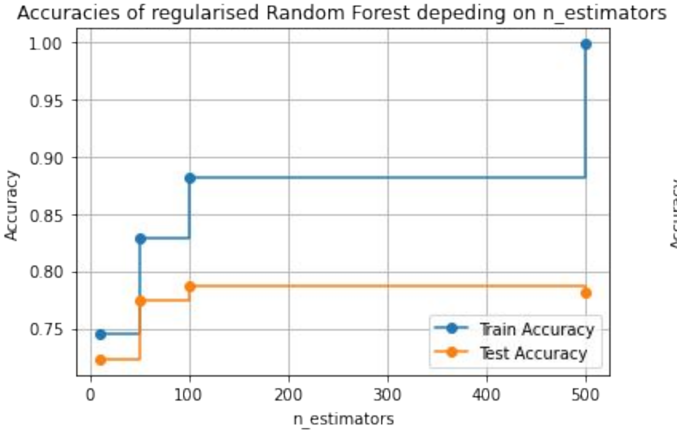
Gradient Boosting

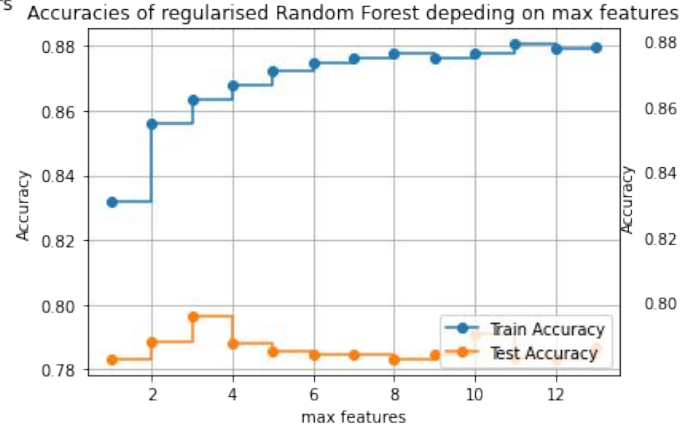
- GB approach uses **multi-classifiers**, which create hundreds of trees.
- Therefore,the design of each classifier is simple and speed up the progress of our training.
- To find the optimal parameters for GB we implemented following parameters are as mentioned for which we get an accuracy score of **80.46%** on the test set which was initially **78.67%** without hyperparameter tuning.

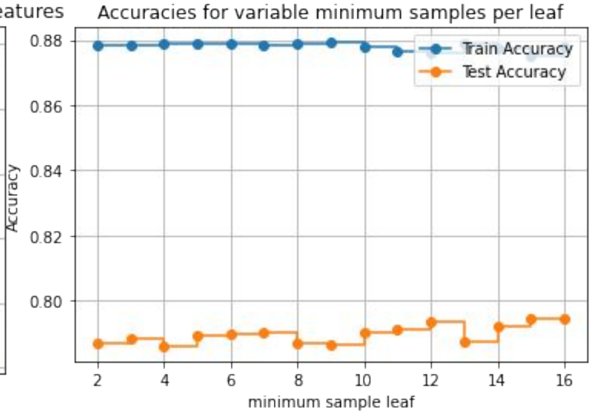
Best Hyperparameters	**Values**
Maximum depth	8
Maximum features	2
Learning rate	0.1
N estimators	500
subsample	1.0
Min. samples split	7
Min. samples leaf	15
criterion	friedman_msc

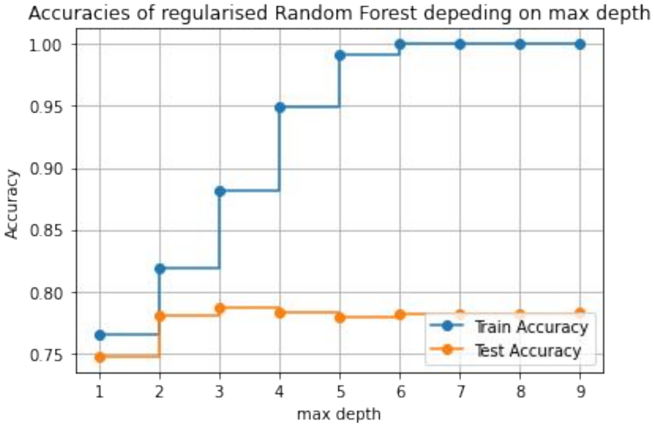
Technology
Arts Sciences
TH Köln

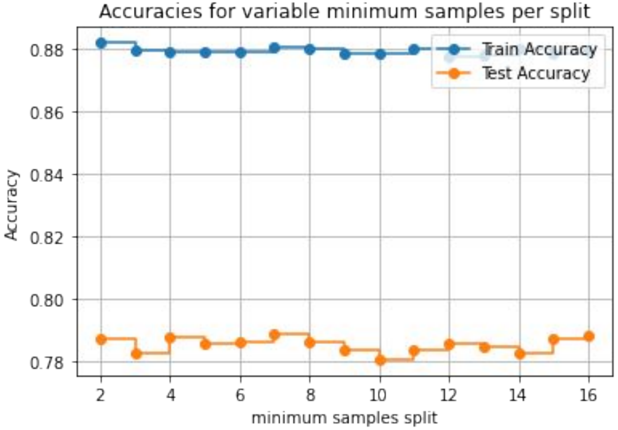
Gradient Boosting Hyperparameter Tuning(individual)



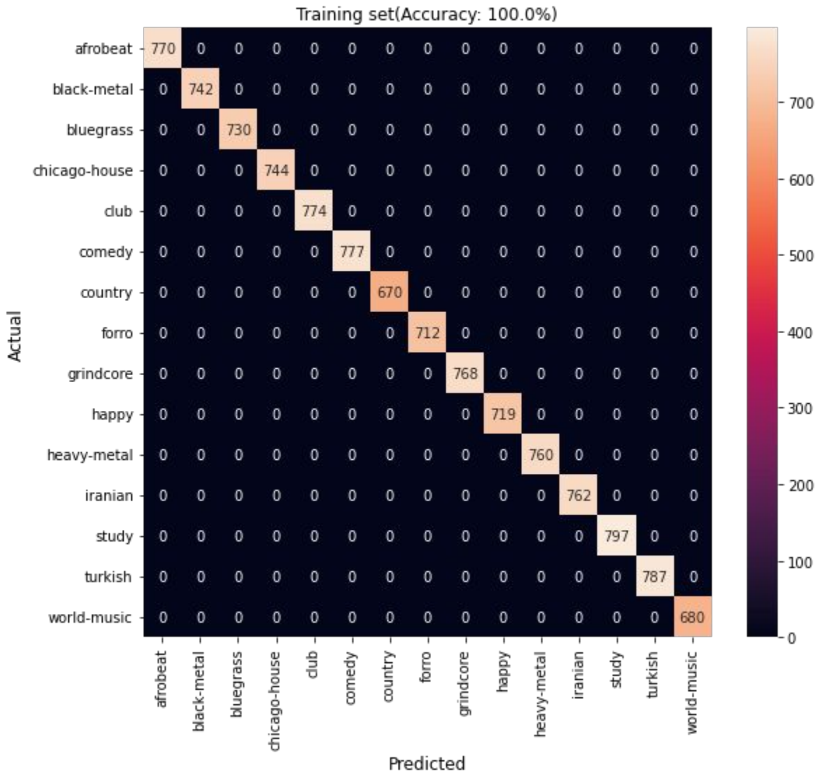


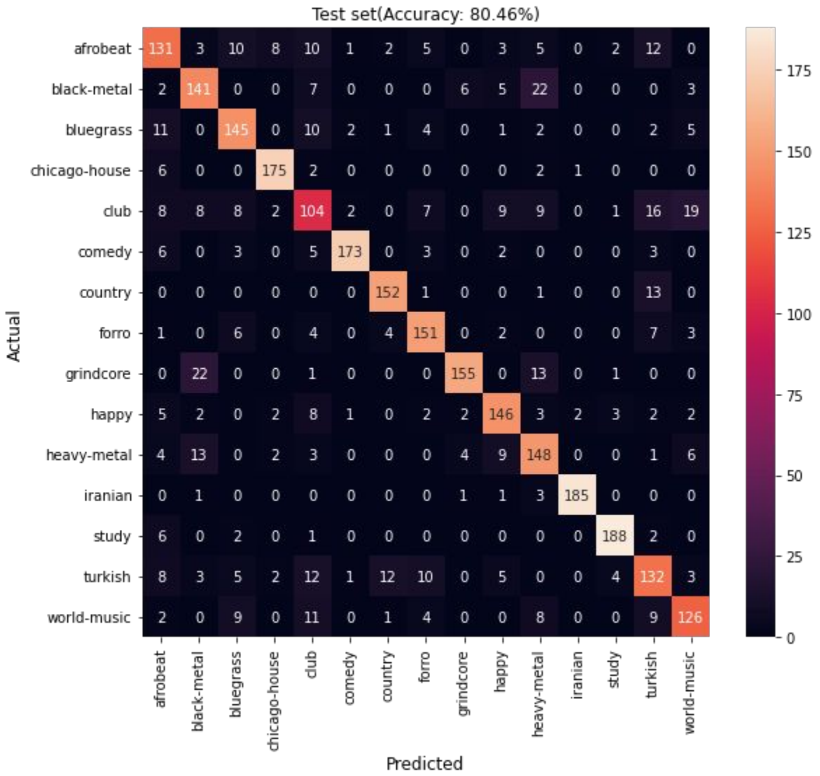




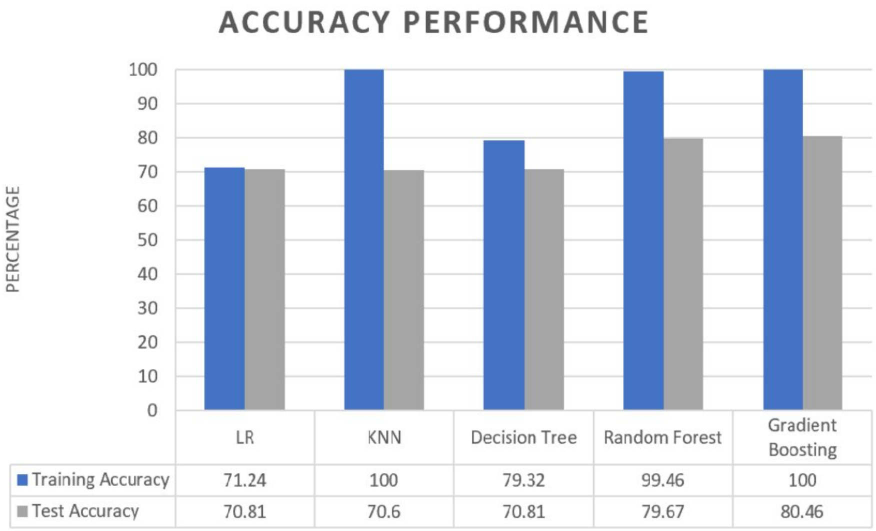


Technology Arts Sciences
TH Köln


Training set(Accuracy: 100.0%)
Actual
Predicted
afrobeat 770 0 0 0 0 0 0 0 0 0 0 0 0 0 0
black-metal 0 742 0 0 0 0 0 0 0 0 0 0 0 0 0
bluegrass 0 0 730 0 0 0 0 0 0 0 0 0 0 0 0
chicago-house 0 0 0 744 0 0 0 0 0 0 0 0 0 0 0
club 0 0 0 0 774 0 0 0 0 0 0 0 0 0 0
comedy 0 0 0 0 0 777 0 0 0 0 0 0 0 0 0
country 0 0 0 0 0 0 670 0 0 0 0 0 0 0 0
forro 0 0 0 0 0 0 0 712 0 0 0 0 0 0 0
grindcore 0 0 0 0 0 0 0 0 768 0 0 0 0 0 0
happy 0 0 0 0 0 0 0 0 0 719 0 0 0 0 0
heavy-metal 0 0 0 0 0 0 0 0 0 0 760 0 0 0 0
iranian 0 0 0 0 0 0 0 0 0 0 0 762 0 0 0
study 0 0 0 0 0 0 0 0 0 0 0 0 797 0 0
turkish 0 0 0 0 0 0 0 0 0 0 0 0 0 787 0
world-music 0 0 0 0 0 0 0 0 0 0 0 0 0 0 680
afrobeat black-metal bluegrass chicago-house club comedy country forro grindcore happy heavy-metal iranian study turkish world-music
700 600 500 400 300 200 100 0


Test set(Accuracy: 80.46%)
Actual
Predicted
afrobeat 131 3 10 8 10 1 2 5 0 3 5 0 2 12 0
black-metal 2 141 0 0 7 0 0 0 6 5 22 0 0 0 3
bluegrass 11 0 145 0 10 2 1 4 0 1 2 0 0 2 5
chicago-house 6 0 0 175 2 0 0 0 0 0 2 1 0 0 0
club 8 8 8 2 104 2 0 7 0 9 9 0 1 16 19
comedy 6 0 3 0 5 173 0 3 0 2 0 0 0 3 0
country 0 0 0 0 0 0 152 1 0 0 1 0 0 13 0
forro 1 0 6 0 4 0 4 151 0 2 0 0 0 7 3
grindcore 0 22 0 0 1 0 0 0 155 0 13 0 1 0 0
happy 5 2 0 2 8 1 0 2 2 146 3 2 3 2 2
heavy-metal 4 13 0 2 3 0 0 0 4 9 148 0 0 1 6
iranian 0 1 0 0 0 0 0 0 1 1 3 185 0 0 0
study 6 0 2 0 1 0 0 0 0 0 0 0 188 2 0
turkish 8 3 5 2 12 1 12 10 0 5 0 0 4 132 3
world-music 2 0 9 0 11 0 1 4 0 0 8 0 0 9 126
afrobeat black-metal bluegrass chicago-house club comedy country forro grindcore happy heavy-metal iranian study turkish world-music
175 150 125 100 75 50 25 0

Model Comparison



- In the plot, it is visible that both highest performing classifiers are respectively GB with an accuracy of 80.46% and RF with 79.67%.
- The lowest accuracy was achieved by the K-NN classifier with 70.6%. Besides, DT and LR had the same accuracy about 70.81% which is only a small amount of higher than K-NN.

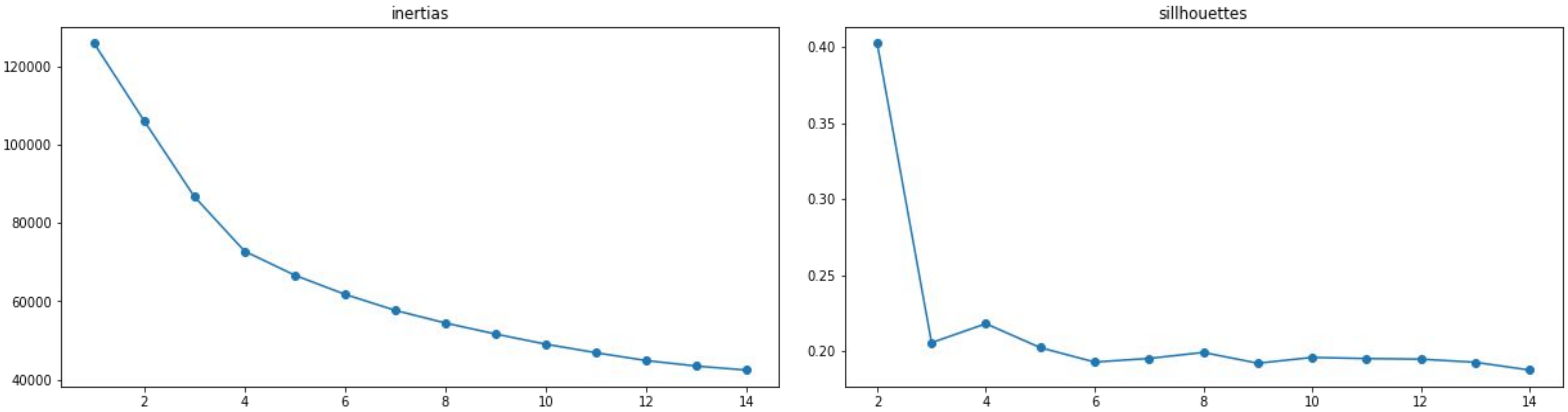
Technology Arts Sciences
TH Köln

Additional Task: K-Means Clustering

- Used numerical parameters, which give informations about specific character of a song.
  - Popularity
  - Danceability
  - Energy
  - Speechiness
  - Acousticness
  - Instrumentalness
  - Liveness
  - Tempo
  - valence

Technology Arts Sciences TH Köln

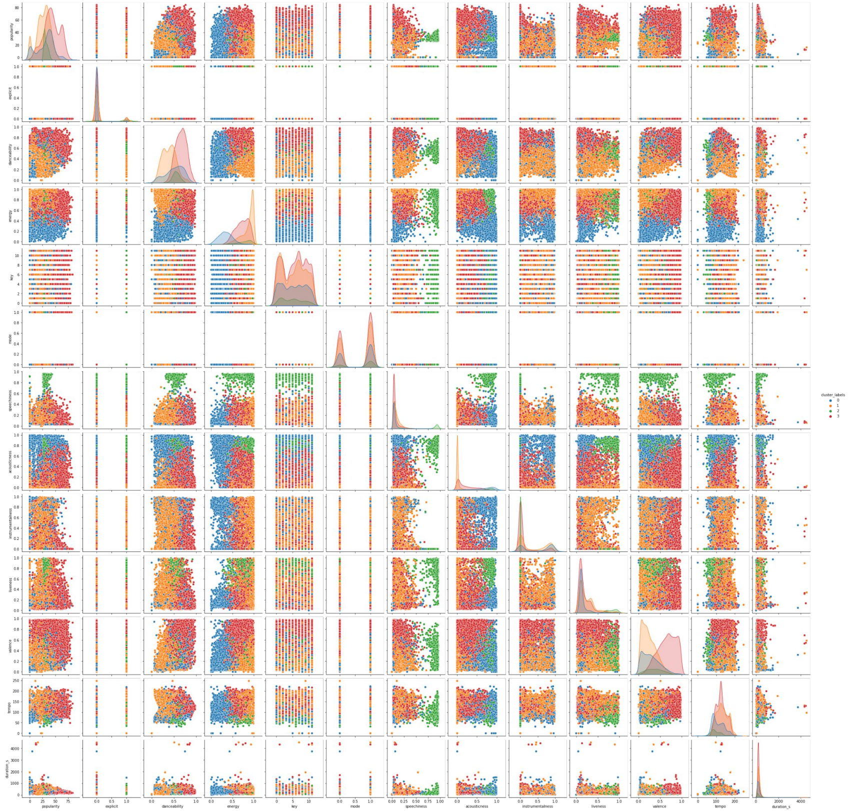
Amount of clusters



Both say: take 4

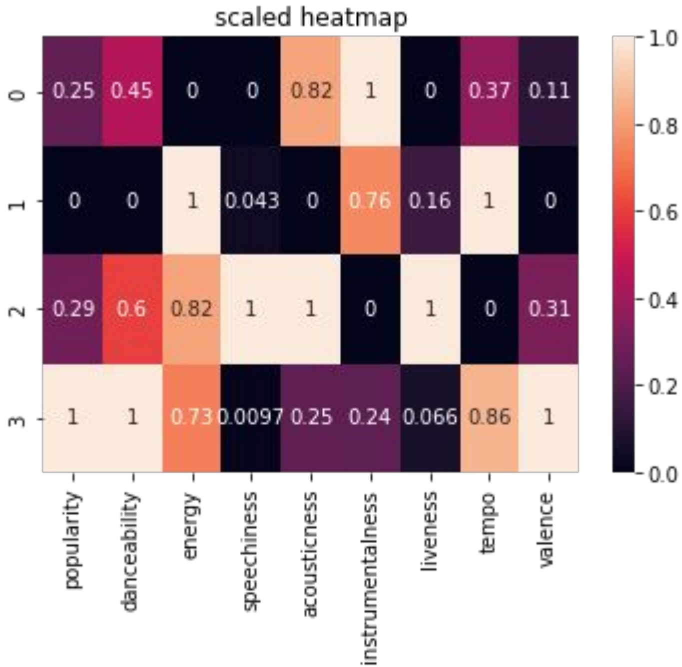
Technology Arts Sciences TH Köln

Pairplots



popularity
explicit
danceability
energy
key
mode
speechiness
acousticness
instrumentalness
liveness
valence
tempo
duration_s
cluster_labels
0
1
2
3

Heatmap

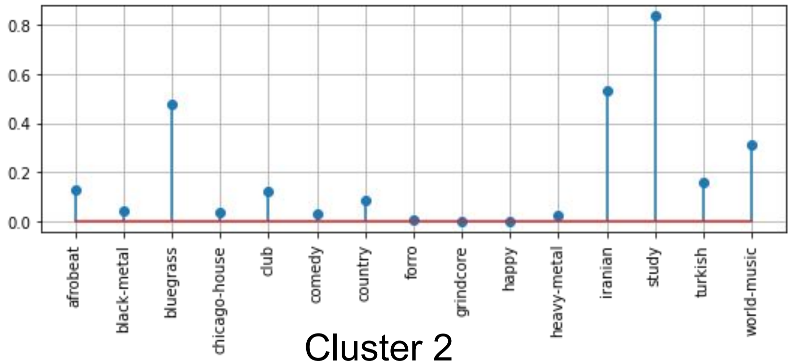


Technology Arts Sciences TH Köln

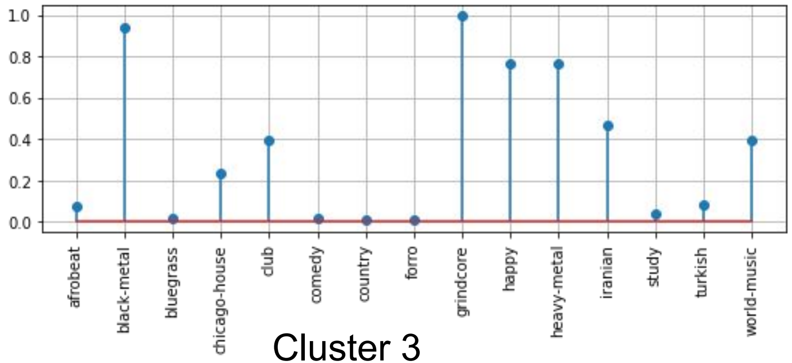
The Clusters

- Cluster 0: Acoustic Instruments
- Cluster 1: energetic fast instrumentals
- Cluster 2: speech live acoustic
- Cluster 3: Pop-Dance

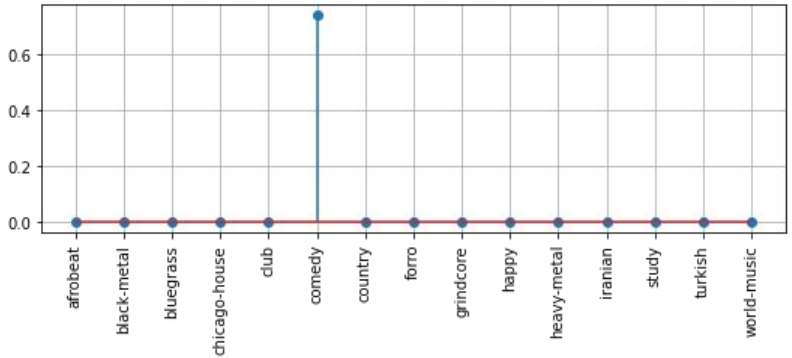
Cluster 0



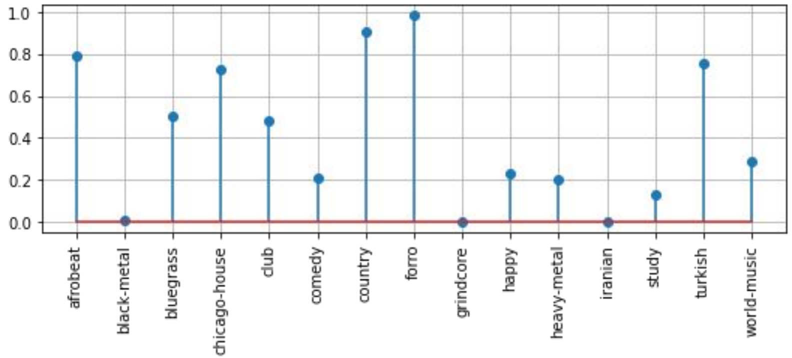
Cluster 1



Cluster 2



Cluster 3



- **Acoustic Instruments:**
  - Bluegrass 50%
  - Iranian 50%
  - Study
  - World-music 33%
- **Energetic fast Instrumentals:**
  - Black-metal
  - Club 50%
  - Grindcore
  - Happy
  - Heavy-metal
  - Iranian 50%
  - World-music
- **Speech live acoustic:**
  - Comedy
- **Pop-Dance:**
  - Afrobeat
  - Bluegrass 50%
  - chicago-House
  - Club 50%
  - Country
  - Forro
  - Turkish
  - World-music 33%

Result and Conclusion

- Self explaining: Hyperparameter leads to accuracy improvement.
- GB and RF are performing way better, than the rest.
- KNN and GB are overfitting.

Thank you !