GTZAN dataset Genre Classification Project

Gaio
Bardelle

About me

I have been working with audio projects and sound reinformencent .

This project aim to help with playlist creation.
Eventually it could be integrate with Dj softwares to provide help

Music Genre Classification With audio features

introduction

Process

Model

Conclusion

About GTZAN Dataset

A collection of 10 genres with 100 audio files each, all having a length of 30 seconds (the famous GTZAN dataset, the MNIST of sounds)
images original - A visual representation for each audio file. the audio files were converted to Mel Spectrograms.
2 CSV files - Containing features of the audio files. One file has for each song (30 seconds long) a mean and variance computed over multiple features that can be extracted from an audio file. The other file has the same structure, but the songs were split before into 3 seconds audio files (this way increasing 10 times the amount of data we fuel into our classification models). With data, more is always better.

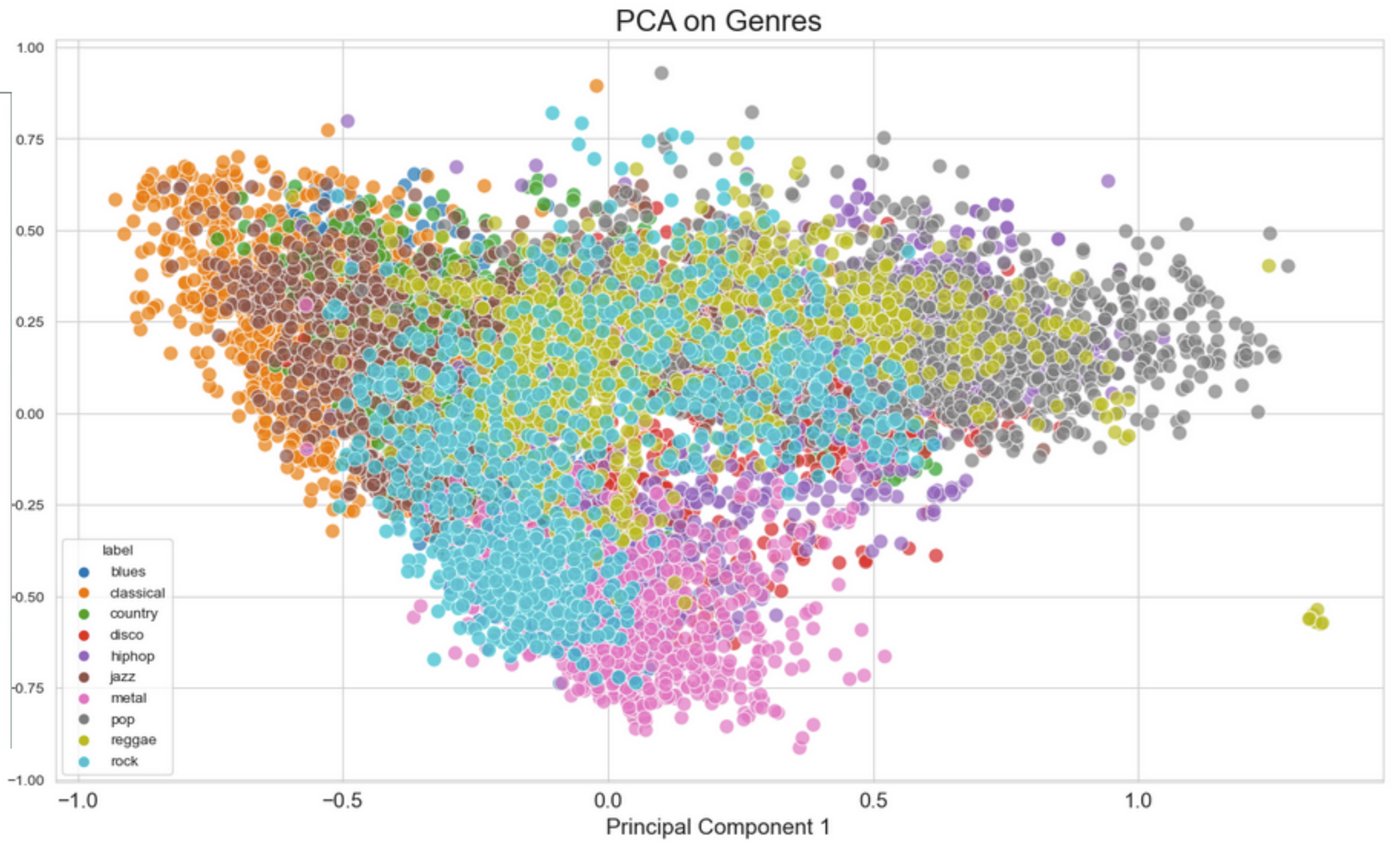
Introduction

Labels

(10 genres, 100 tracks per genre)

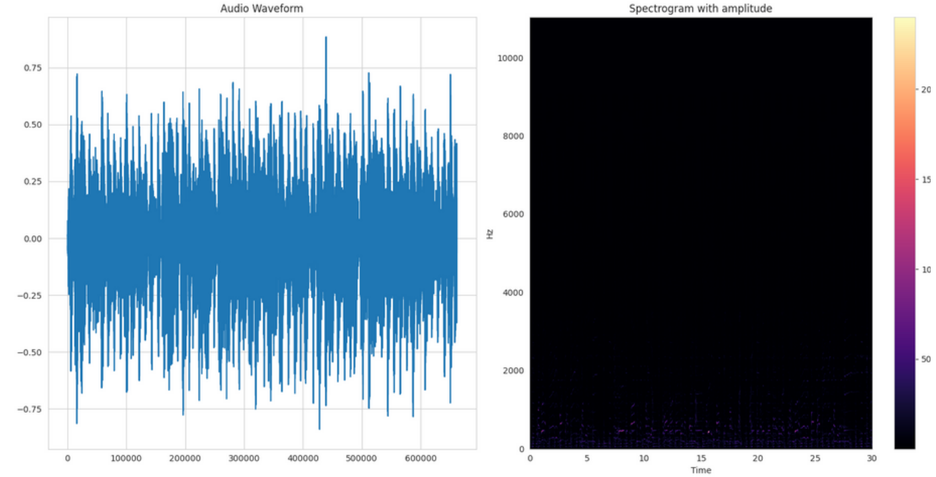
Genres

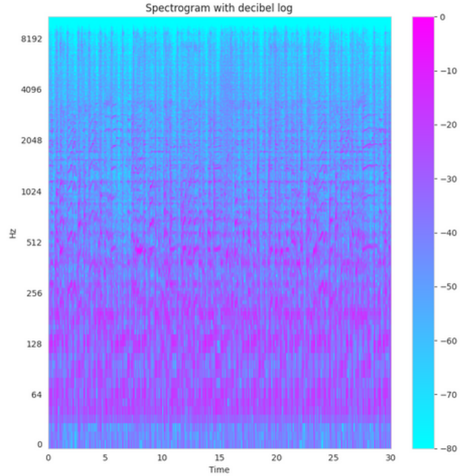
- Blues
- Classical
- Country
- Disco
- Hiphop
- Jazz
- Metal
- Pop
- Reggae
- Rock

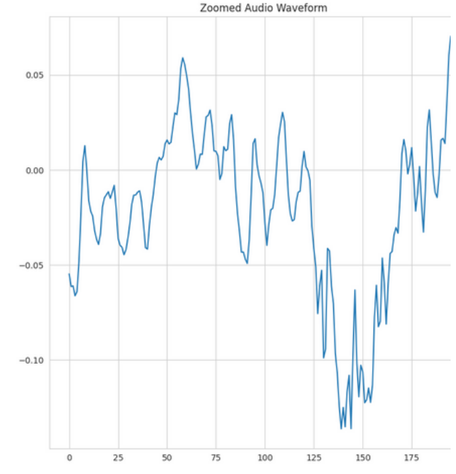


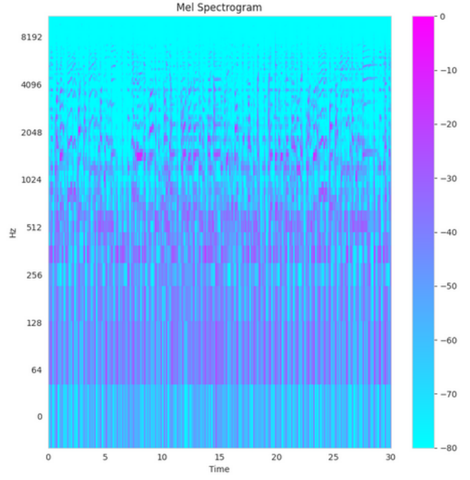
Introduction

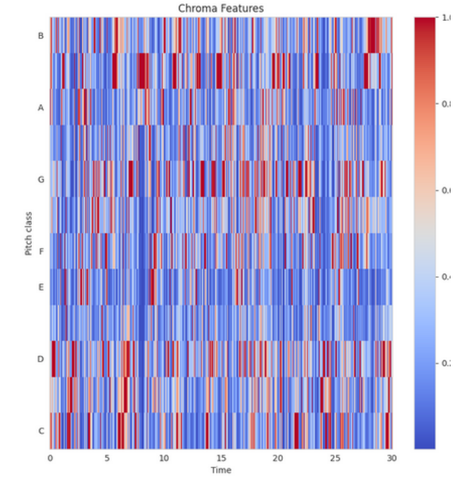
blues.00000.wav



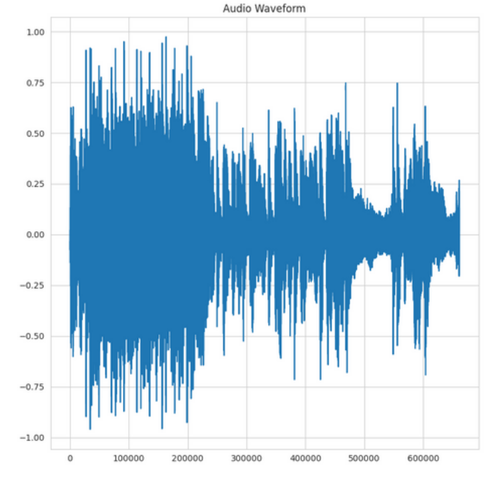


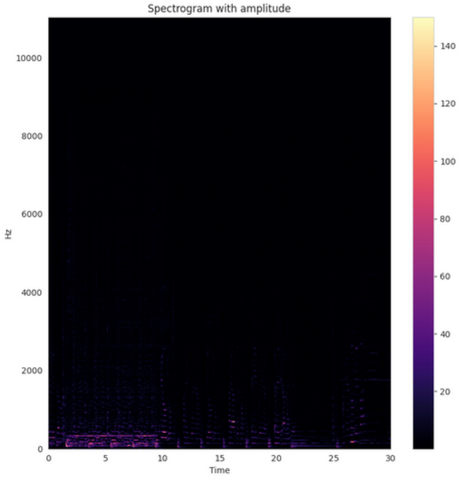


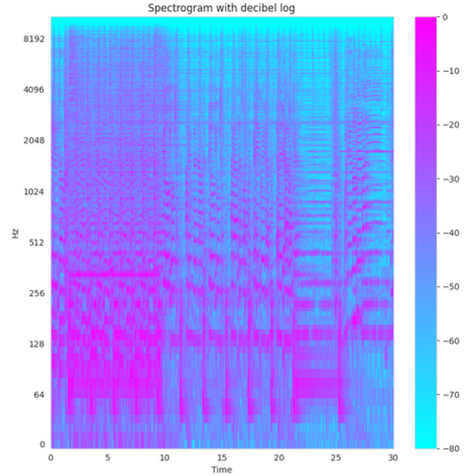


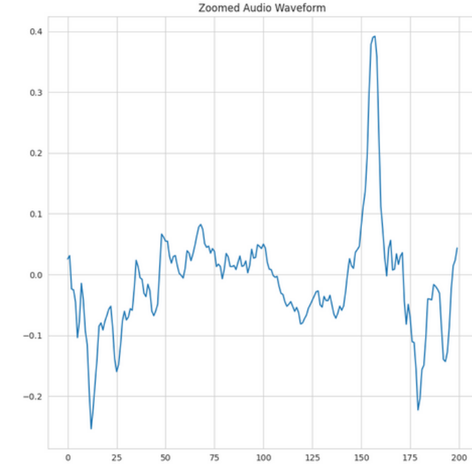


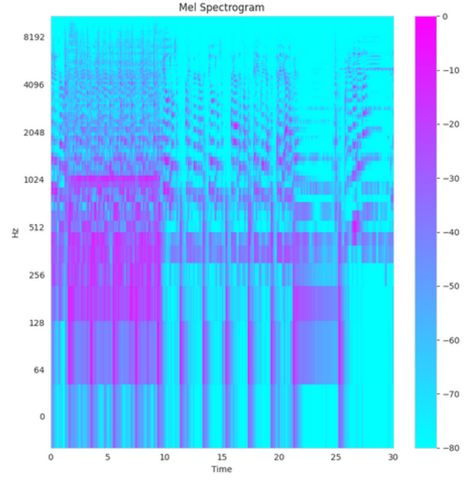
rock.00000.wav

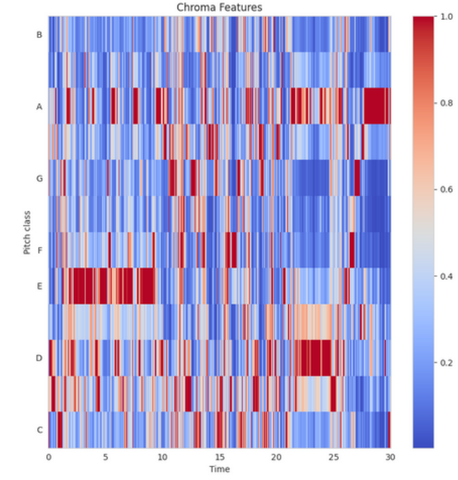










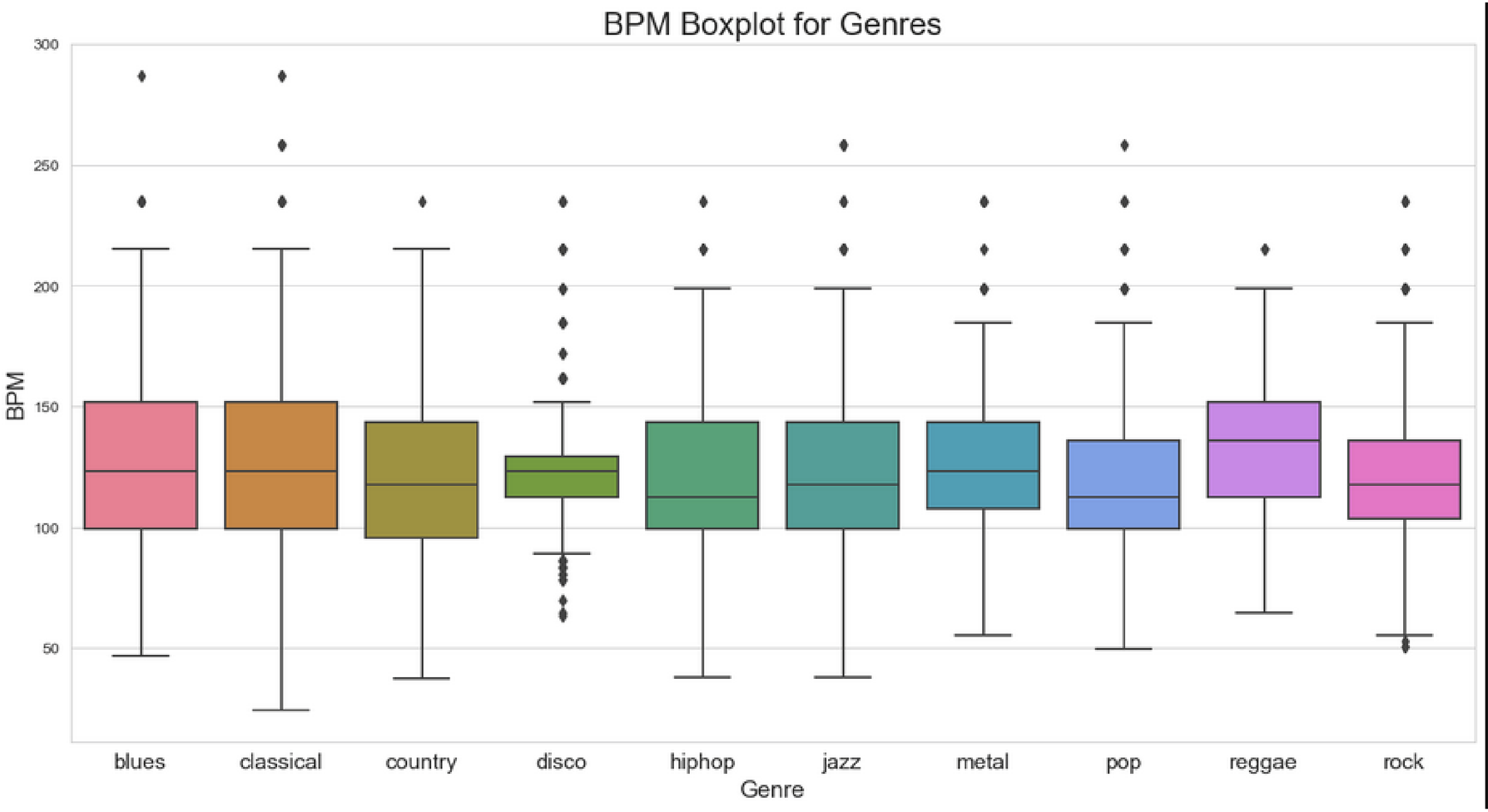


Introduction

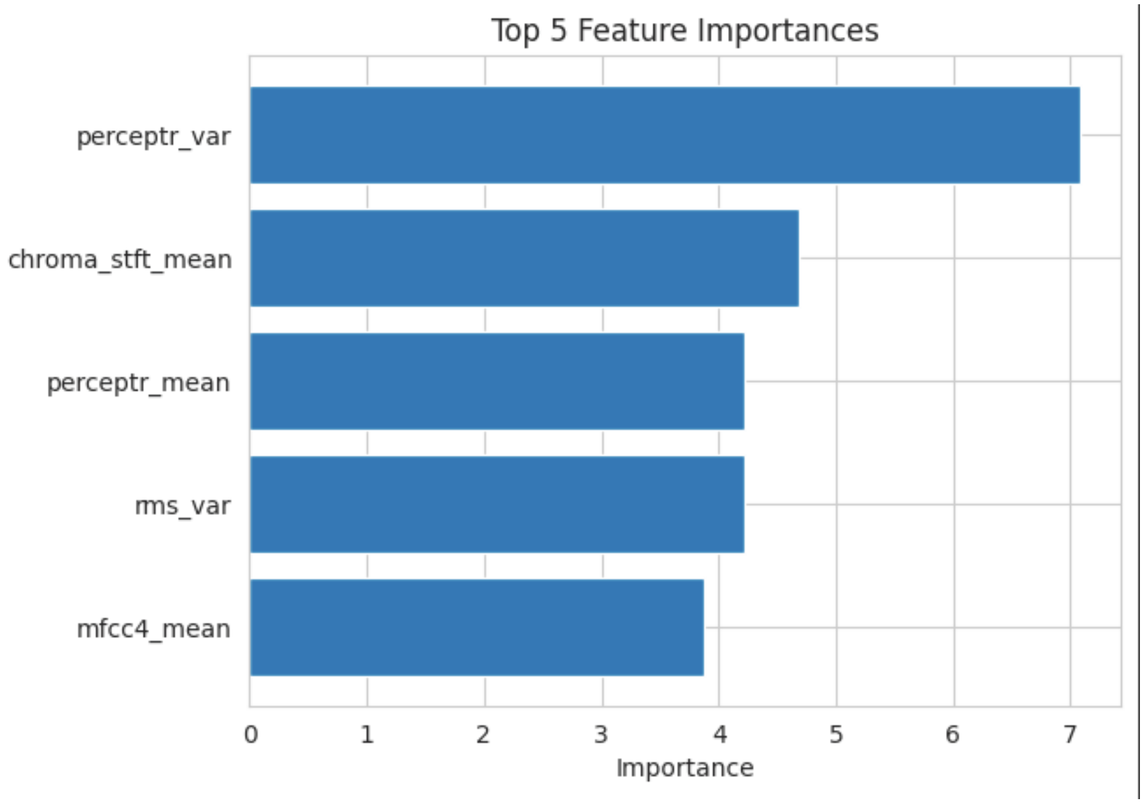
Features

(57 features in total)

Means and Variances of Music Features



Process



❑ Remove irrelevant columns

❑ Feature Preprocessing

Standardization

StandardScaler()

❑ Target Transformation

Multi-Class Label Encoding

LabelEncoder()

0: 'blues'	1: 'classical'	2: 'country'	3: 'disco'	4: 'hiphop'	5: 'jazz'	6: 'metal'	7: 'pop'	8: 'reggae'	9: 'rock'

Models

Fully connected Neural Network

```
model = k.models.Sequential([
k.layers.Dense(1024, activation='relu',),
k.layers.Dropout(0.3),
k.layers.Dense(512, activation='relu', ),
k.layers.Dropout(0.4),
k.layers.Dense(256, activation='relu',
k.layers.Dropout(0.4),
k.layers.Dense(128, activation='relu',
k.layers.Dropout(0.4),
k.layers.Dense(64, activation='relu', ,
k.layers.Dropout(0.4),
k.layers.Dense(10, activation='softmax',
```

Catboost

```
(iterations=500,
learning_rate=0.1,
depth=5,
loss_function='MultiClass'
```

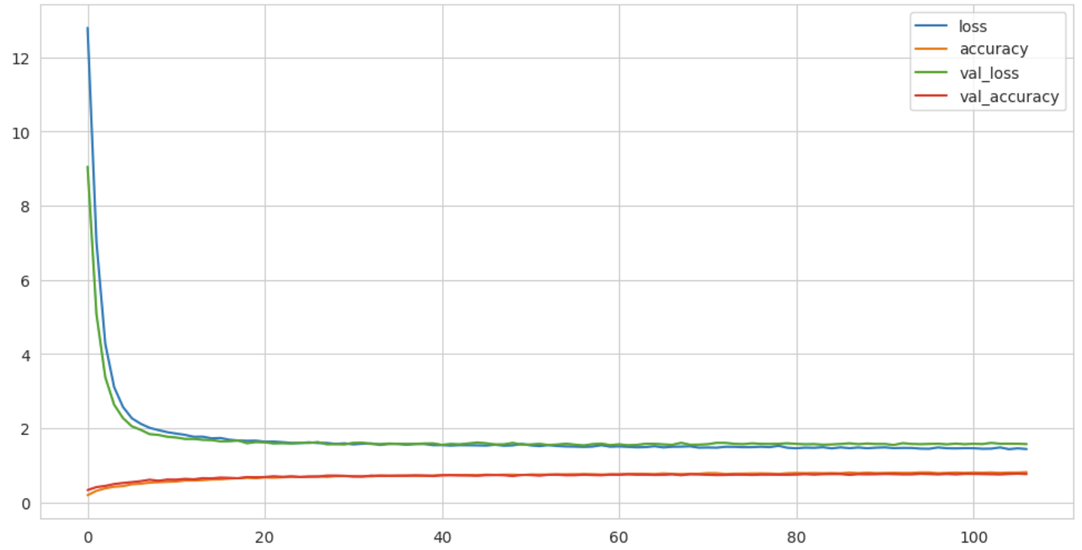
Conclusion

Catboost

Fully connected Neural Network

Mean CV Score	Accuracy	Precision	Recall	F1 Score	AUC
0.822394	0.749428	0.749428	0.749428	0.746941	0.982343
0.770056	0.70546	0.70634	0.70465	0.712942	0.952343

Neural Network training



Conclusion

GTZAN is a great Dataset to work on Genre classification,I learnt a lot about Music features extraction methods.

A Fully connected Neural Network would require more data to be able to perform better

With some adaptation we could use a similar model to raccomand similar music to a listener

My next project will be working on a different datasetand obtain and process the features myself.