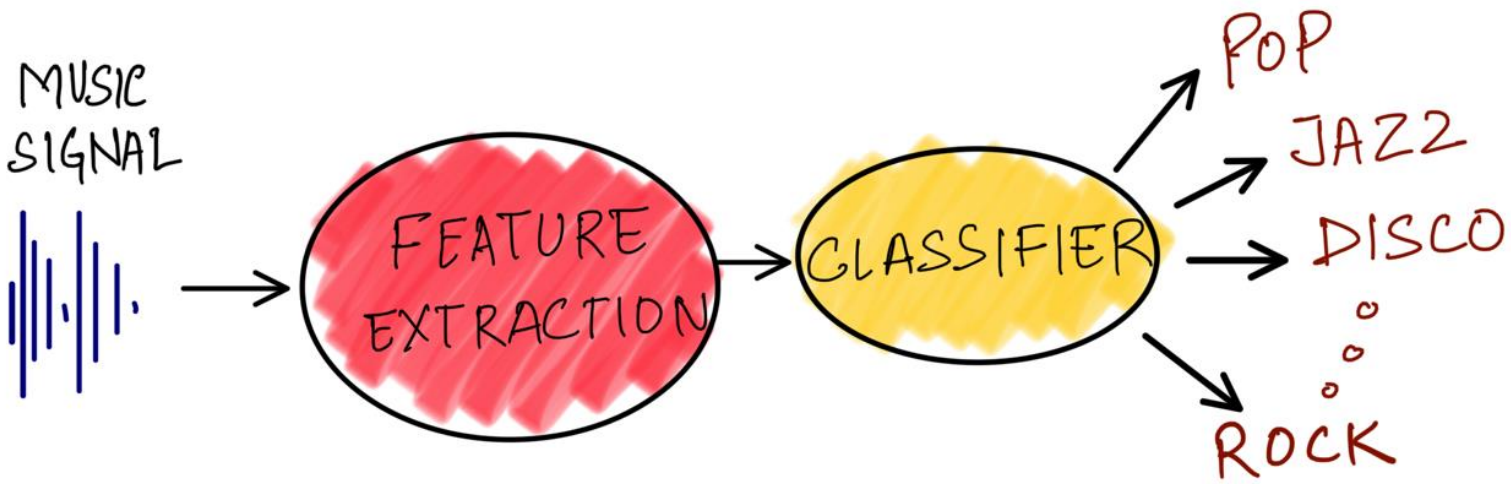
Study of different Machine Learning models for Music Genre Classification

MLSP Final Project 2021

Khadatkar Sameer Raju (17830)

Abhishek Kumar (18208)

- Music apps nowadays have millions of songs in their database. For the proper organization of these songs, their classification according to the genre is essential.
- In this project, we are trying to use different Machine Learning and Deep Learning models for doing this classification.



Technical Details

Timbral Texture Features:

- Spectral centroid
- Spectral flux
- Root Mean Square Energy (RMSE)
- Zero crossing rate
- Spectral contrast
- Spectral bandwidth
- Spectral flatness
- Spectral roll-off
- Mel-Frequency Cepstral Coefficients (MFCC) (20 Nos)

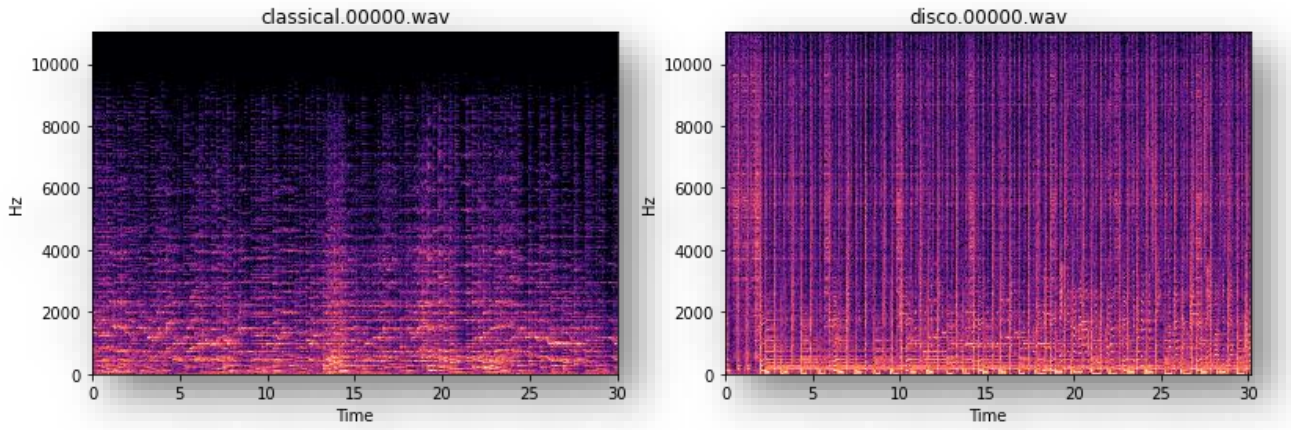
  Statistics used:
  - Maximum
  - Minimum
  - Mean
  - Standard deviation
  - Kurtosis
  - Skewness

28 (frequency domain features) × 6 (statistics) +
1 (the tempo of the music) = 169.

Features used for CNN model:

- Convert 1D shape of time-series data of a sample to a 2D form using a window size = 33000 and window offset = 16500. Due to which shape of each data sample became (39,33000) for a 30 seconds .wav file.
- Calculate 128 points mel-spectrogram for each 33000 size data for hop length of 256, which change the size 33000 to (128, 129)

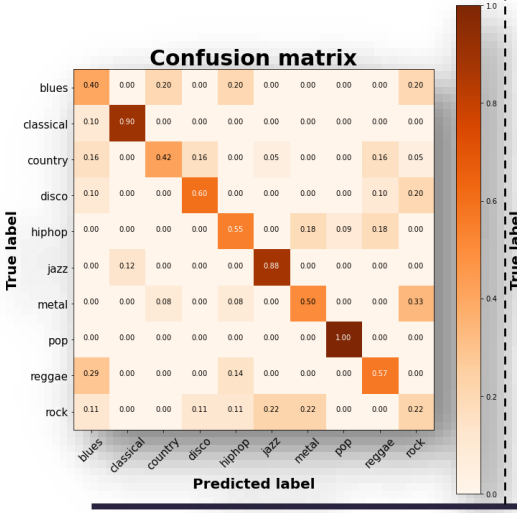
Therefore, final size of the features for each sample became (39,128,129).



Technical Details

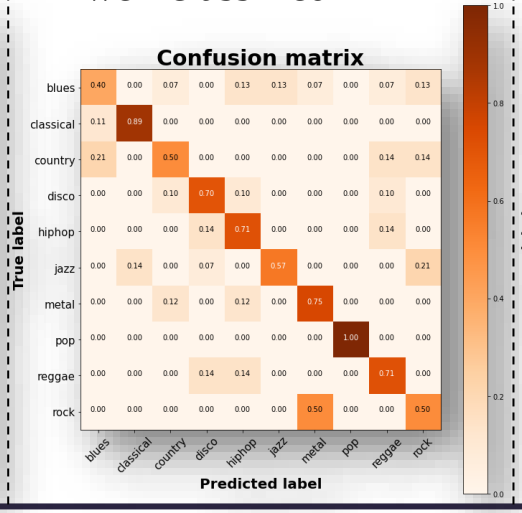
KNN

- PCA = 15 - 20 components
- Accuracy: 59 ± 8 %
- Classical, Jazz and Pop were well classified



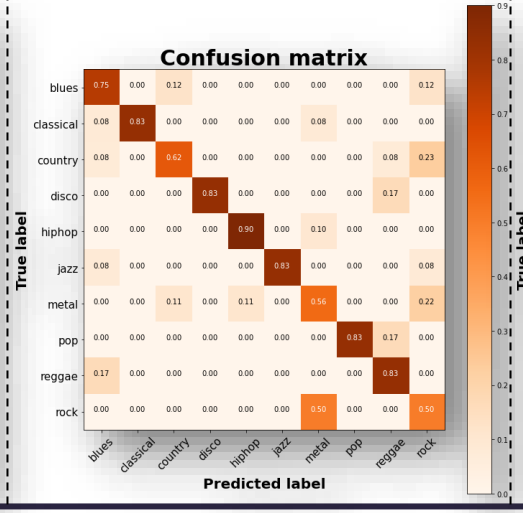
XGBoost

- Boosters: **gbtree**, dart
- Max. depth of tree: [4, 6, 10, 20]
- Objective function: “multi:softmax”
- Accuracy: 65 ± 3 %
- Classical and Pop were well classified



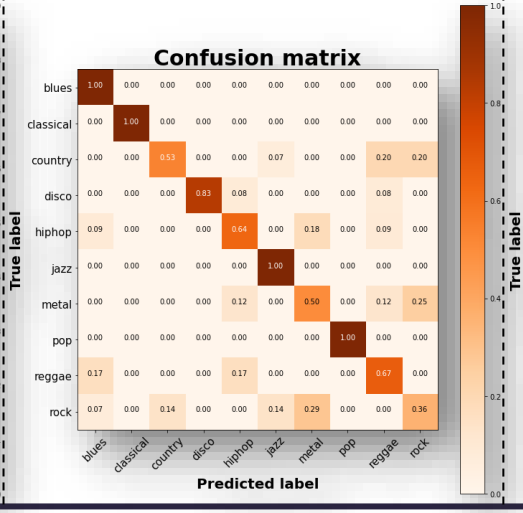
SVM

- C: [0.5, ***1***, 5, 10]
- kernel: ***rbf***, sigmoid, poly
- Accuracy: 76 ± 3 %
- Except Country, Metal and Rock all were well classified



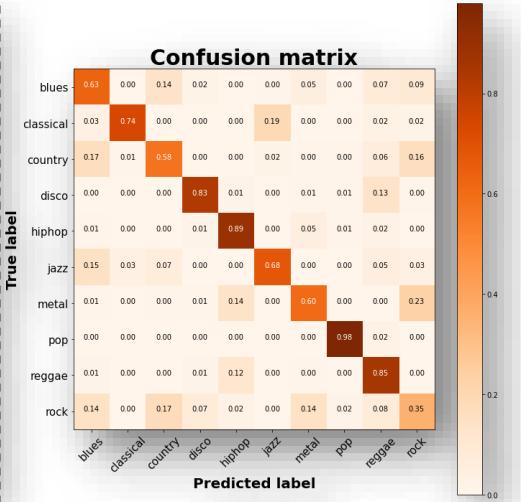
DNN

- Hidden layer : 6
- Activation Function: ReLU, Softmax
- Optimizer: Adam
- Loss Function: Cross Entropy
- Accuracy: 72 ± 4 %
- Blues, Classical, Jazz and Pop were exactly classified



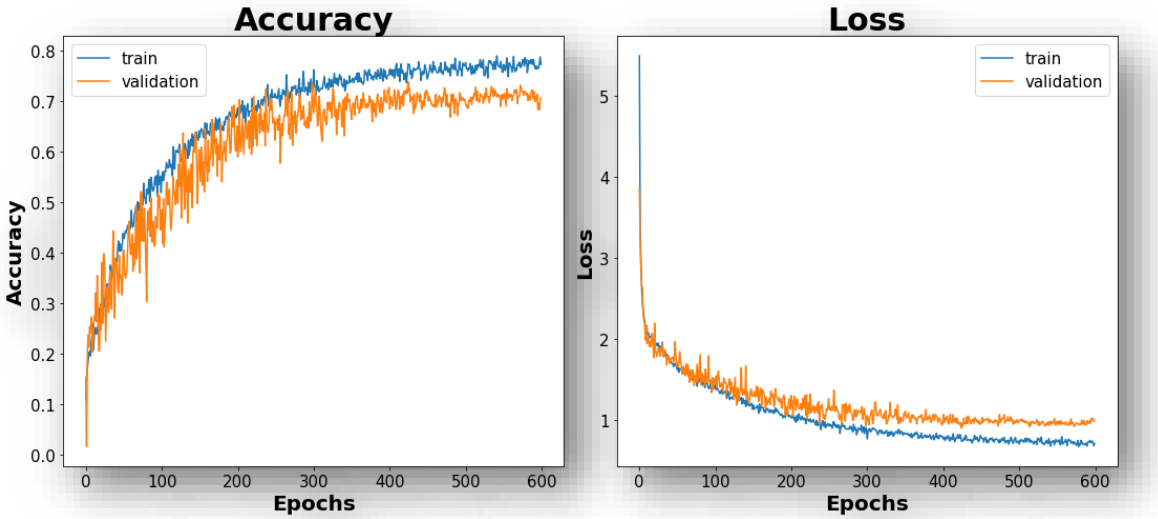
CNN

- Convolution layer: 4
- Maxpooling: (2,2)
- Dense Layer: 1
- Activation Function: ReLU, Softmax
- Optimizer: Adam
- Loss Function: Cross Entropy
- Accuracy: 69 ± 4 %
- Disco, Hip-hop, Pop and Reggae were well classified

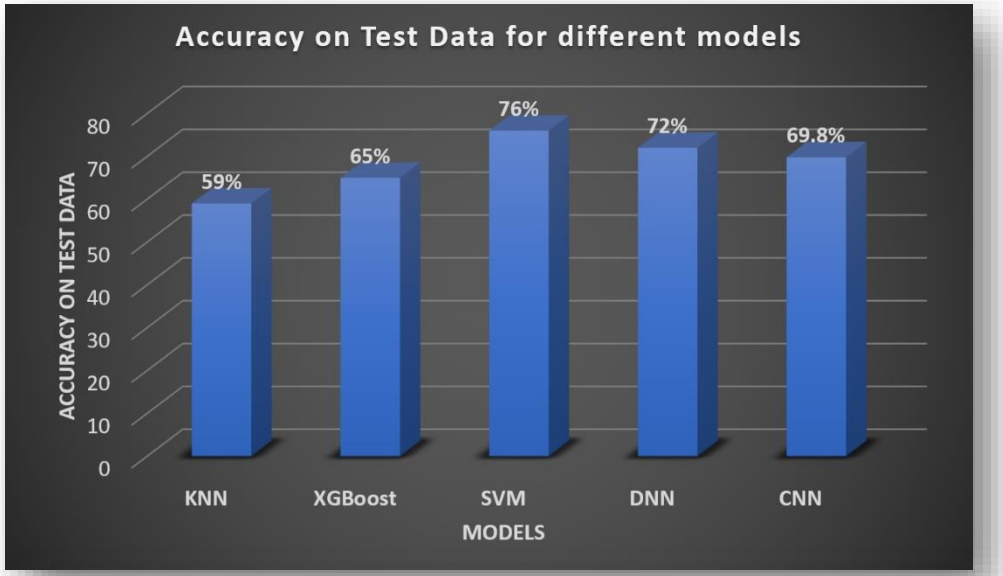


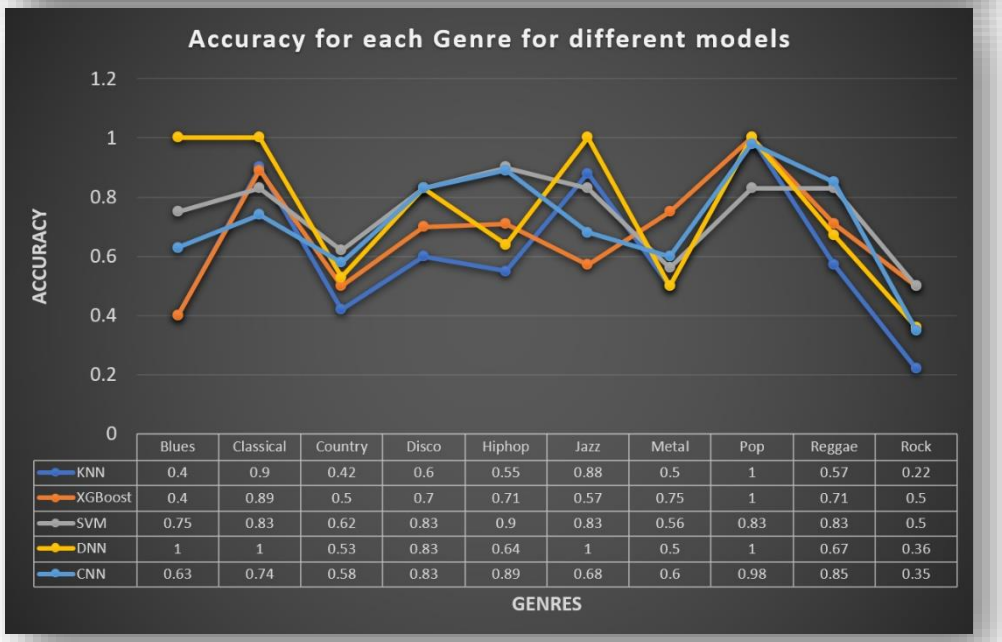
Results

- SVM is the best performing model for us.
- The pop and the classical genre were easy to classify, whereas rock was comparatively difficult to classify.
- We were expecting good results from CNN, but somehow we were not getting the desired results, as we couldn't make enough trials for the model's training because it took a lot of time on our system.



Change in accuracy and loss with number of epochs of CNN





	Blues	Classical	Country	Disco	Hiphop	Jazz	Metal	Pop	Reggae	Rock
KNN	0.4	0.9	0.42	0.6	0.55	0.88	0.5	1	0.57	0.22
XGBoost	0.4	0.89	0.5	0.7	0.71	0.57	0.75	1	0.71	0.5
SVM	0.75	0.83	0.62	0.83	0.9	0.83	0.56	0.83	0.83	0.5
DNN	1	1	0.53	0.83	0.64	1	0.5	1	0.67	0.36
CNN	0.63	0.74	0.58	0.83	0.89	0.68	0.6	0.98	0.85	0.35

Novel Contributions

Research papers read: Tzanetakis & Cook, 2002 and Ghildiyal et al., 2020.
Through which we got some insight about the new approaches to extract features of the audio (.wav format) file, which are more appropriate for the music genre classification.

Python Libraries that helped us to get the results are :

- Numpy
- Pandas
- Librosa: Extracting features
- Scikit-learn: KNN, SVM
- xgboost
- Keras: DNN,CNN
- Matplotlib

We distributed our work among us in the following manner:

- Abhishek Kumar : DNN, KNN, XGBoost and report writing.
- Khadatkar Sameer Raju : Feature Extraction, SVM, CNN and PPT.