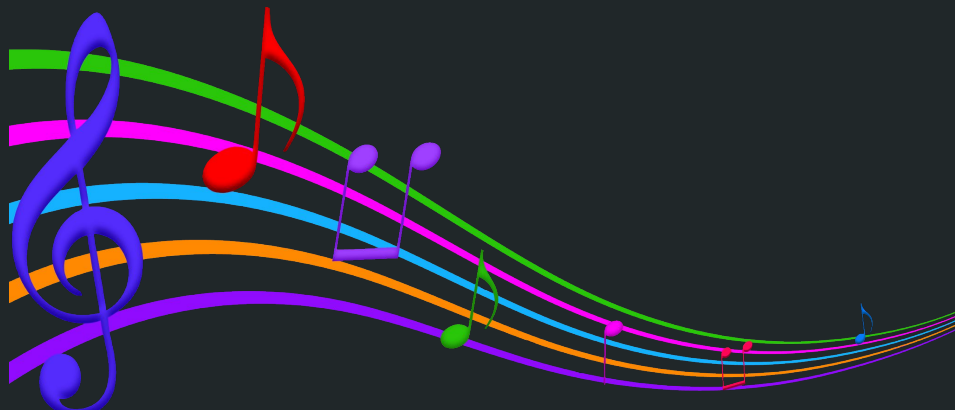
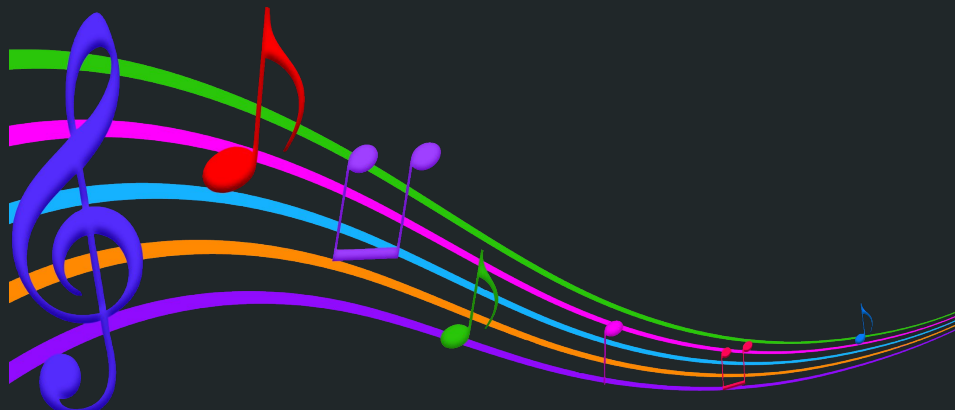
Using Neural Nets to Predict Music Genre from Waveforms

Dayv Doberne



Background



Background

- Genres
  - Loosely divided categories of music
  - Change over time
  - Eg. ‘Pop’

Background

- Genres
  - Loosely divided categories of music
  - Change over time
  - Eg. ‘Pop’
- Fashion is cyclical!

Background

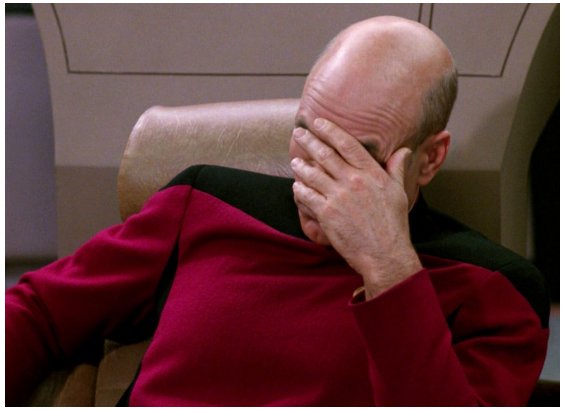
- Can we train NNs to classify genre?

Background

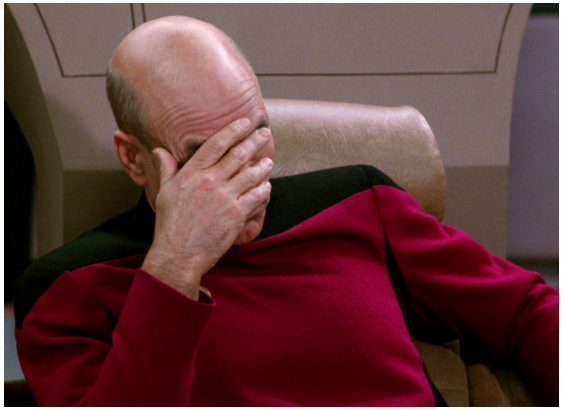
- Can we train NNs to classify genre?
- Then we can use that model to forecast trends!

42%

Model prediction accuracy



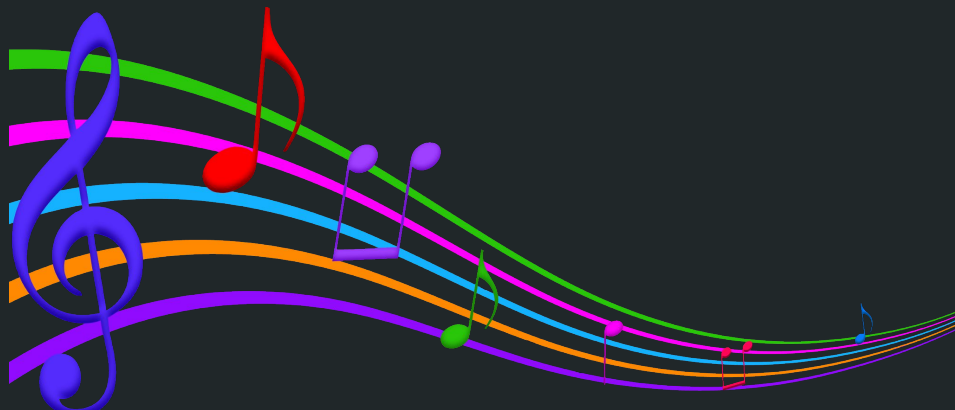
42%



Model prediction accuracy

(not enough for forecasting)

Data

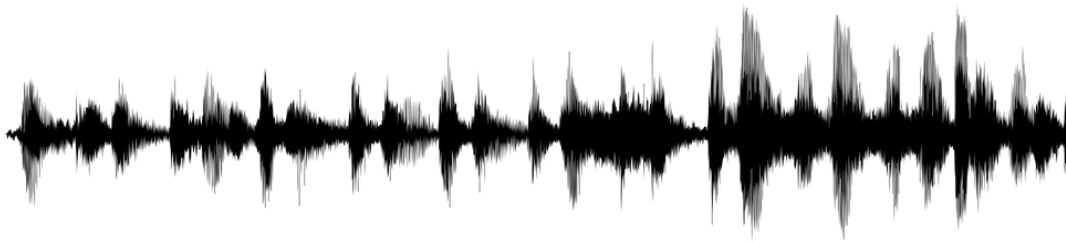


Data

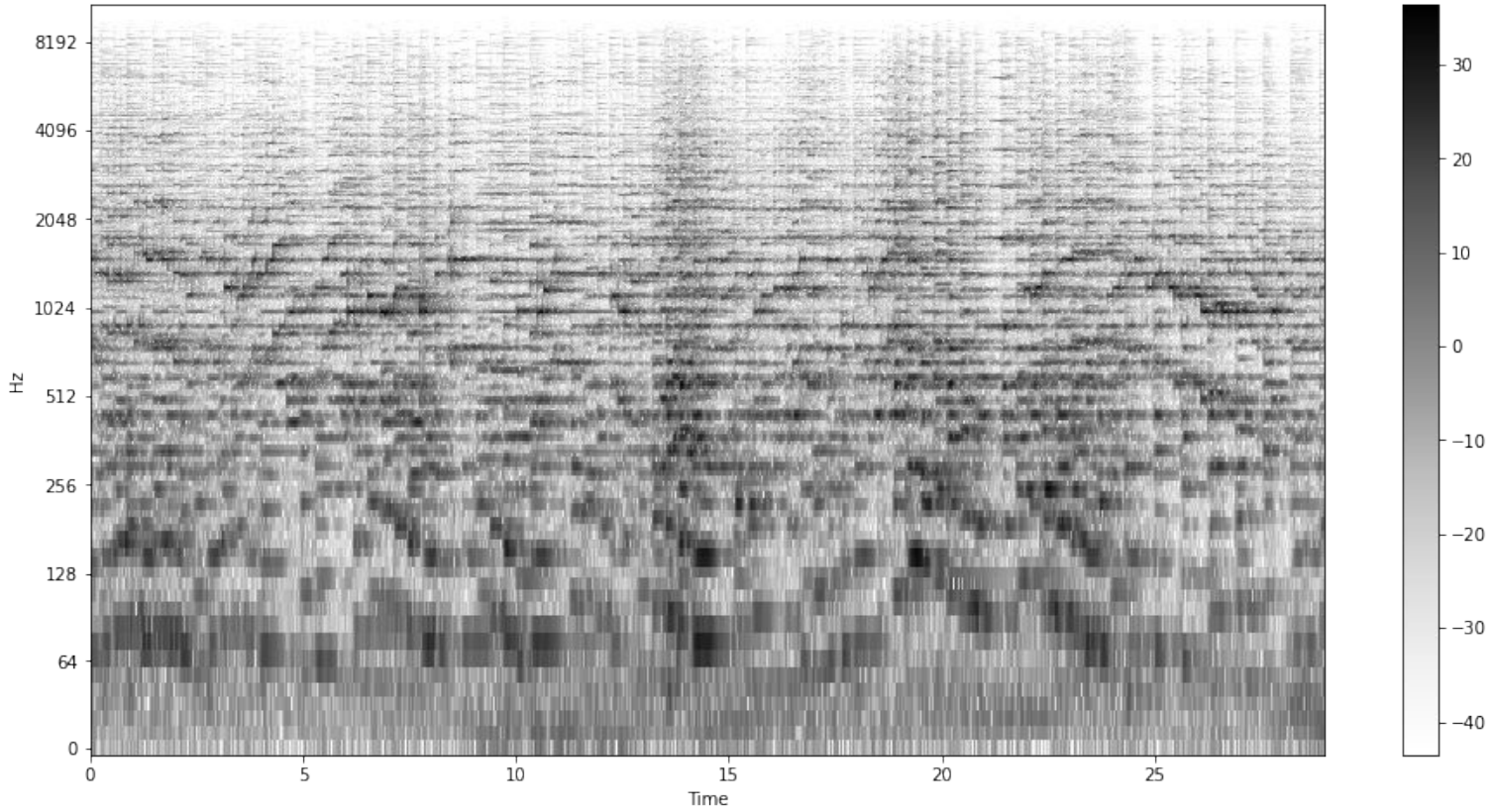
- GTZAN data set
  - Blues, Classical, Country, Disco, Hiphop, Jazz, Metal, Pop, Reggae, Rock
  - 100 30s clips of each

Data

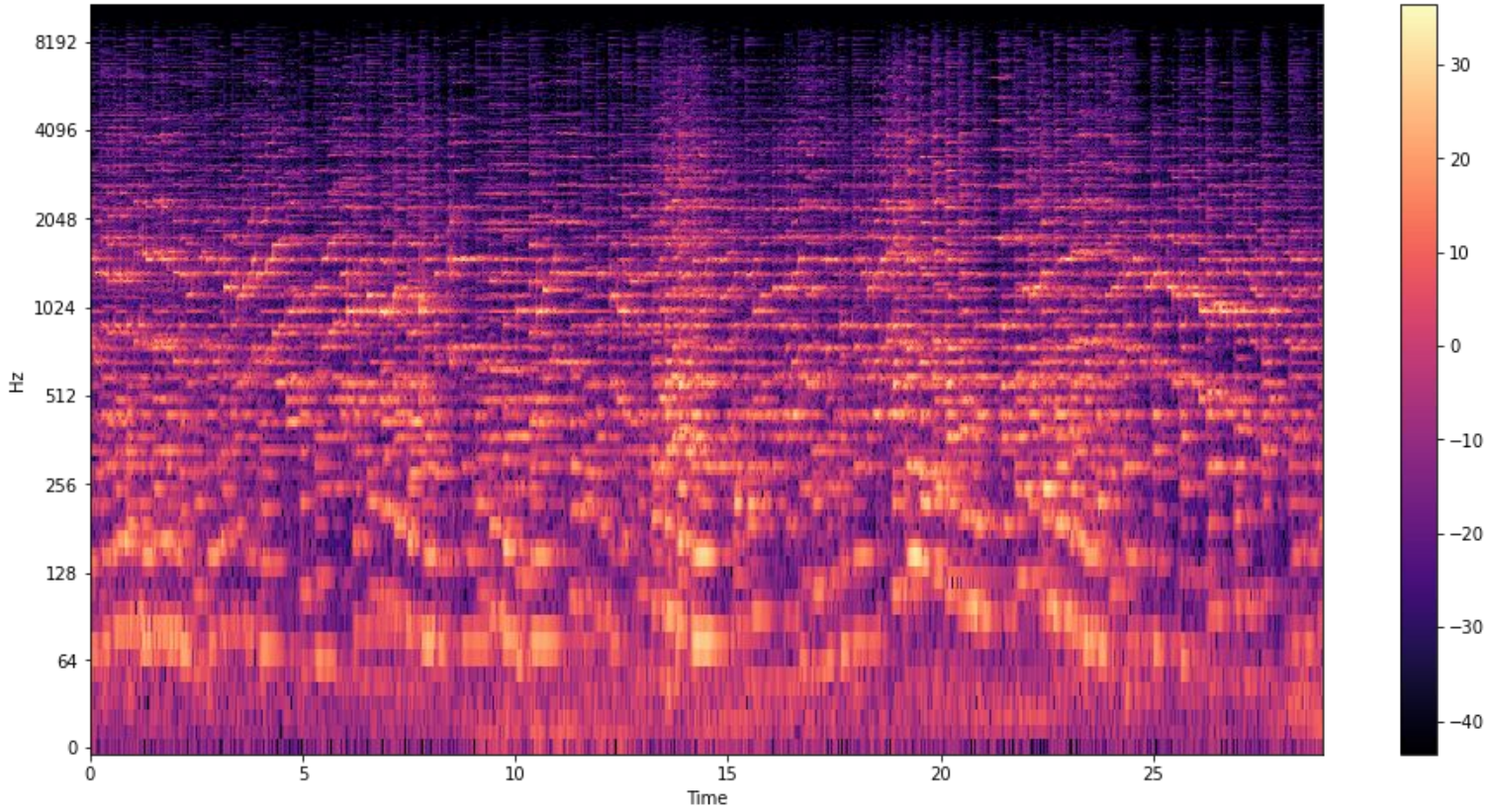
- Librosa library
  - Converts waveform to vector
  - Can transform vector into 2D array



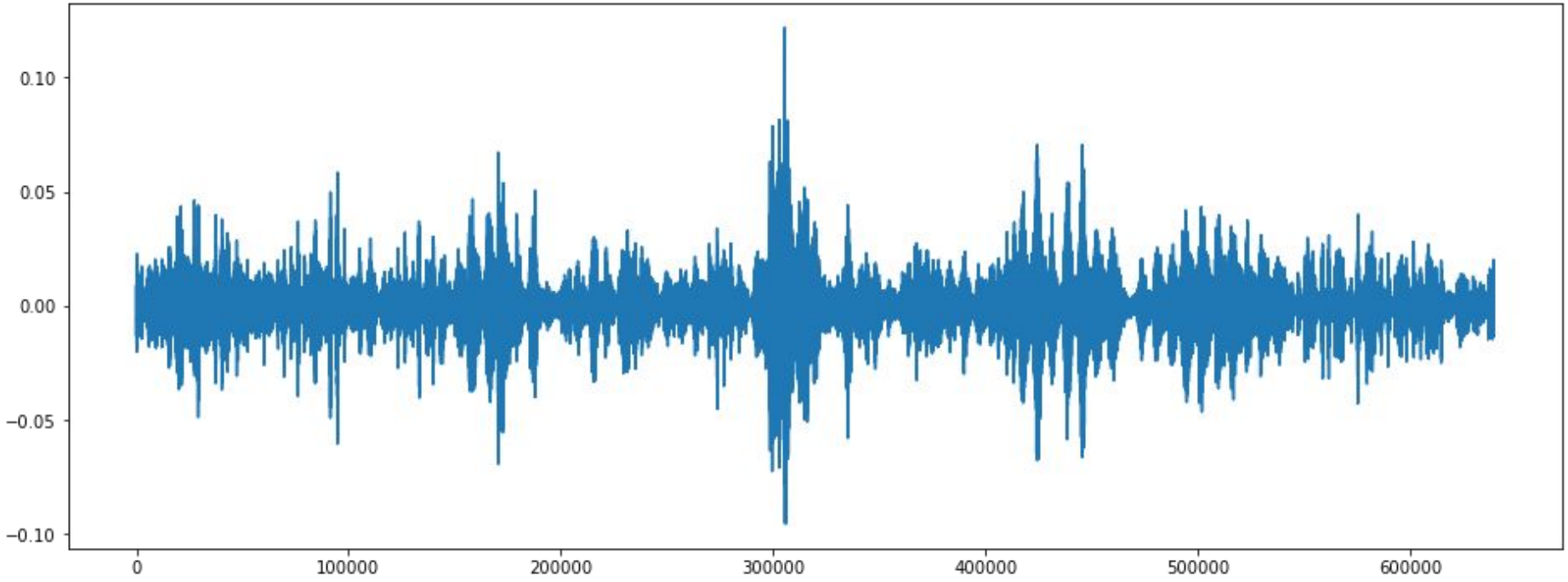
Spectogram


8192
4096
2048
1024
512
256
128
64
0
Hz
0
5
10
15
20
25
Time
30
20
10
0
−10
−20
−30
−40

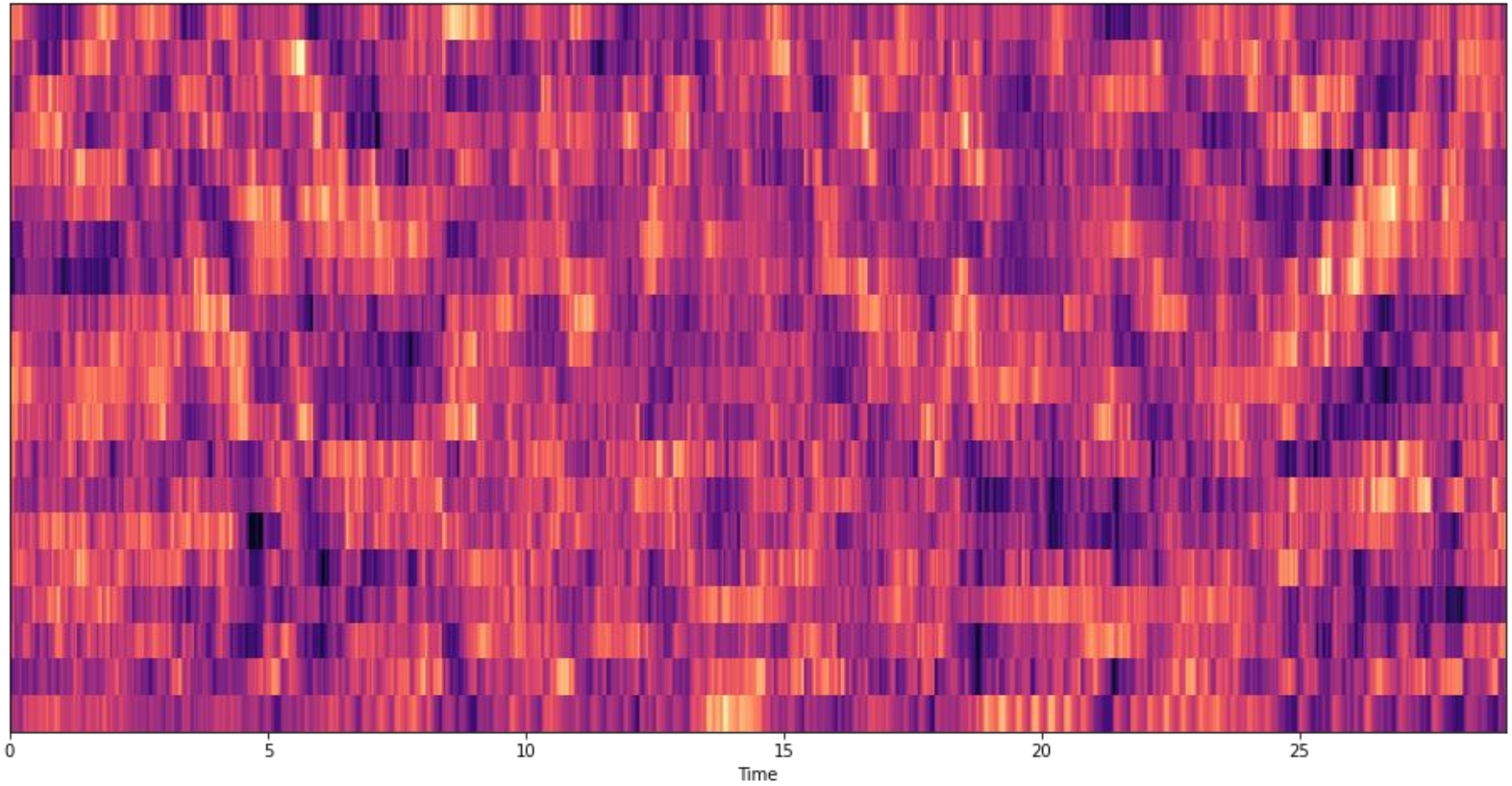
Spectogram


8192
4096
2048
1024
512
256
128
64
0
Hz
0
5
10
15
20
25
Time
30
20
10
0
−10
−20
−30
−40

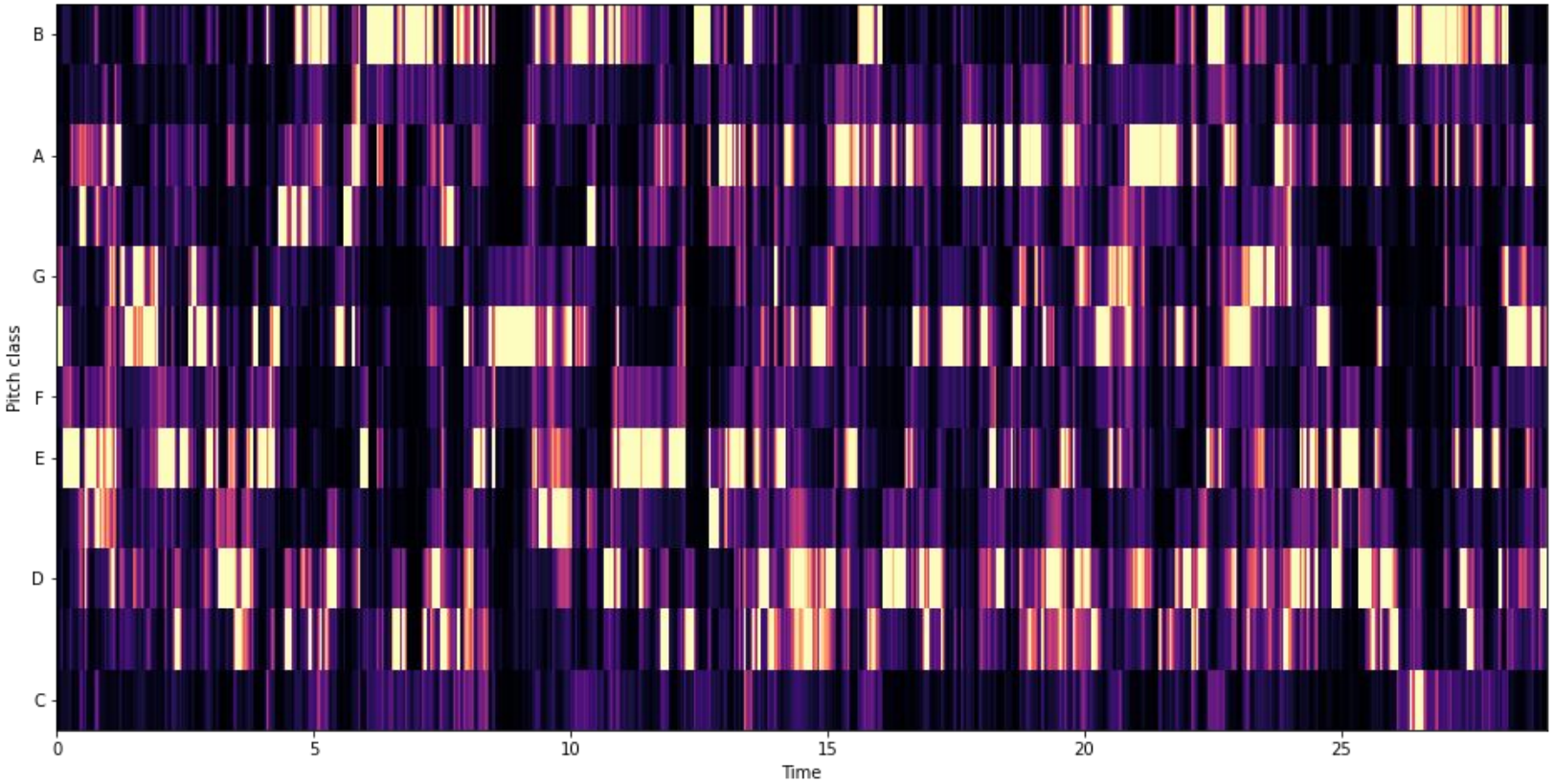
Percussion


0.10
0.05
0.00
−0.05
−0.10
0
100000
200000
300000
400000
500000
600000

Mel-Frequency Cepstral Coefficients

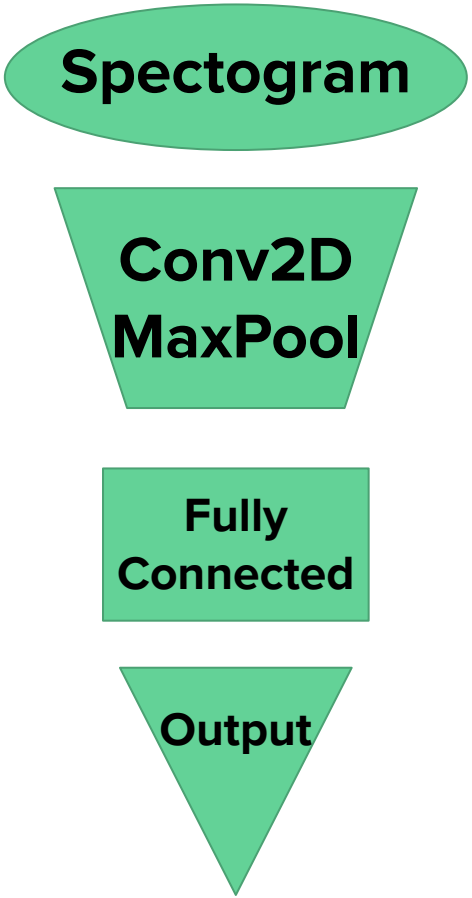


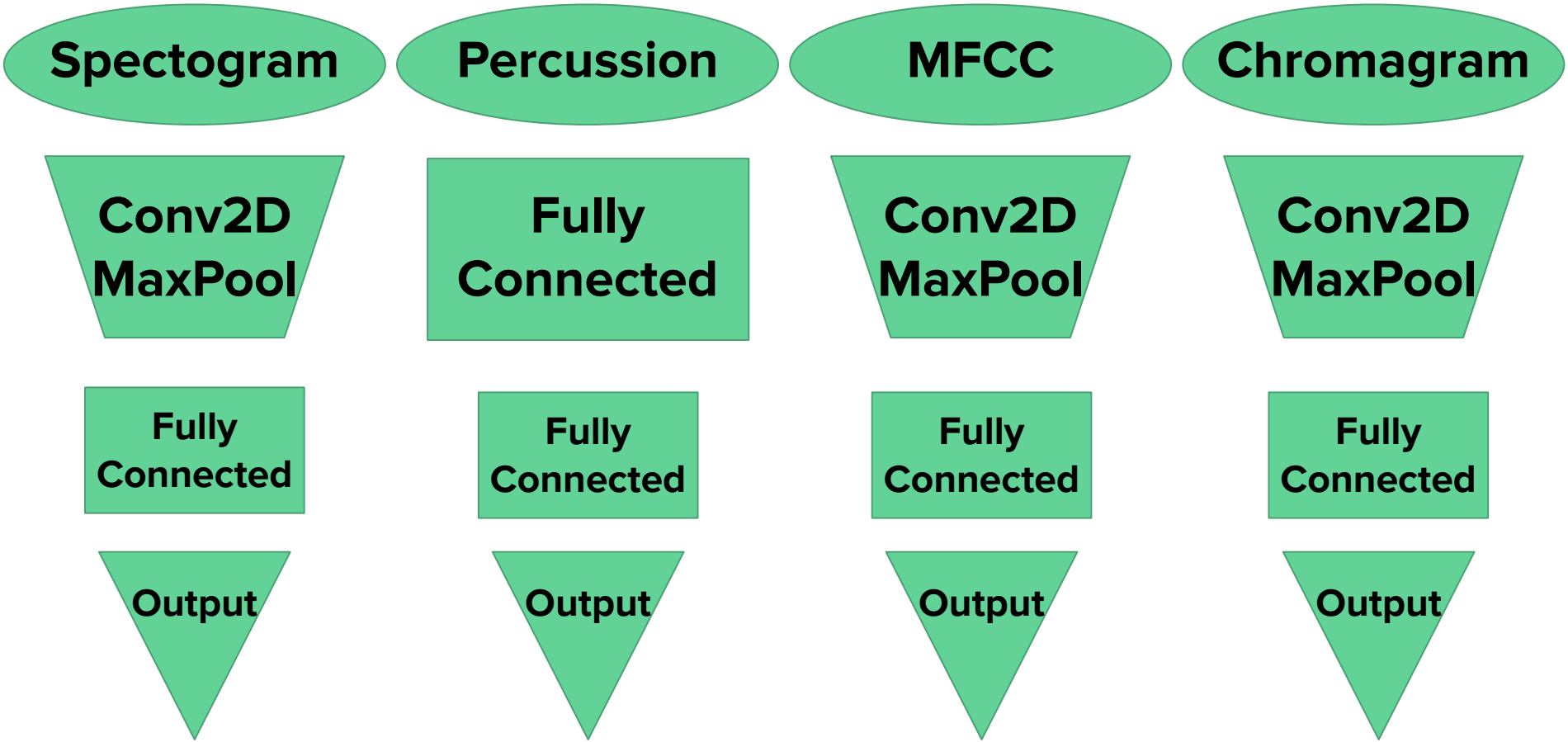
Chromagram

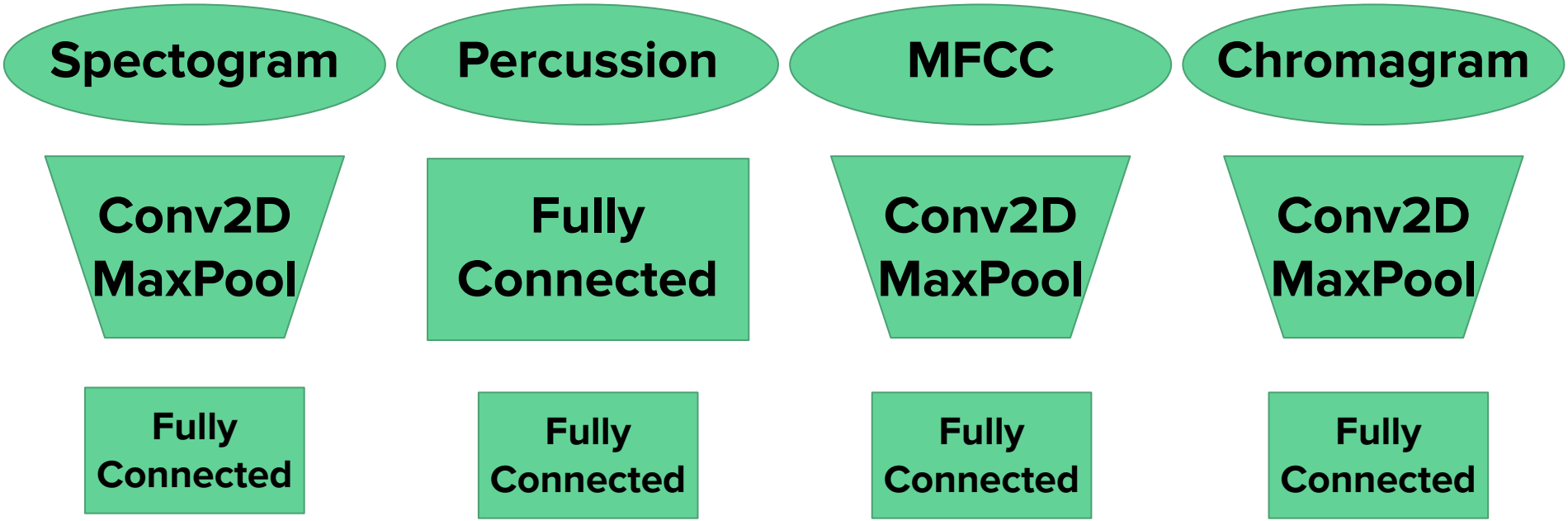


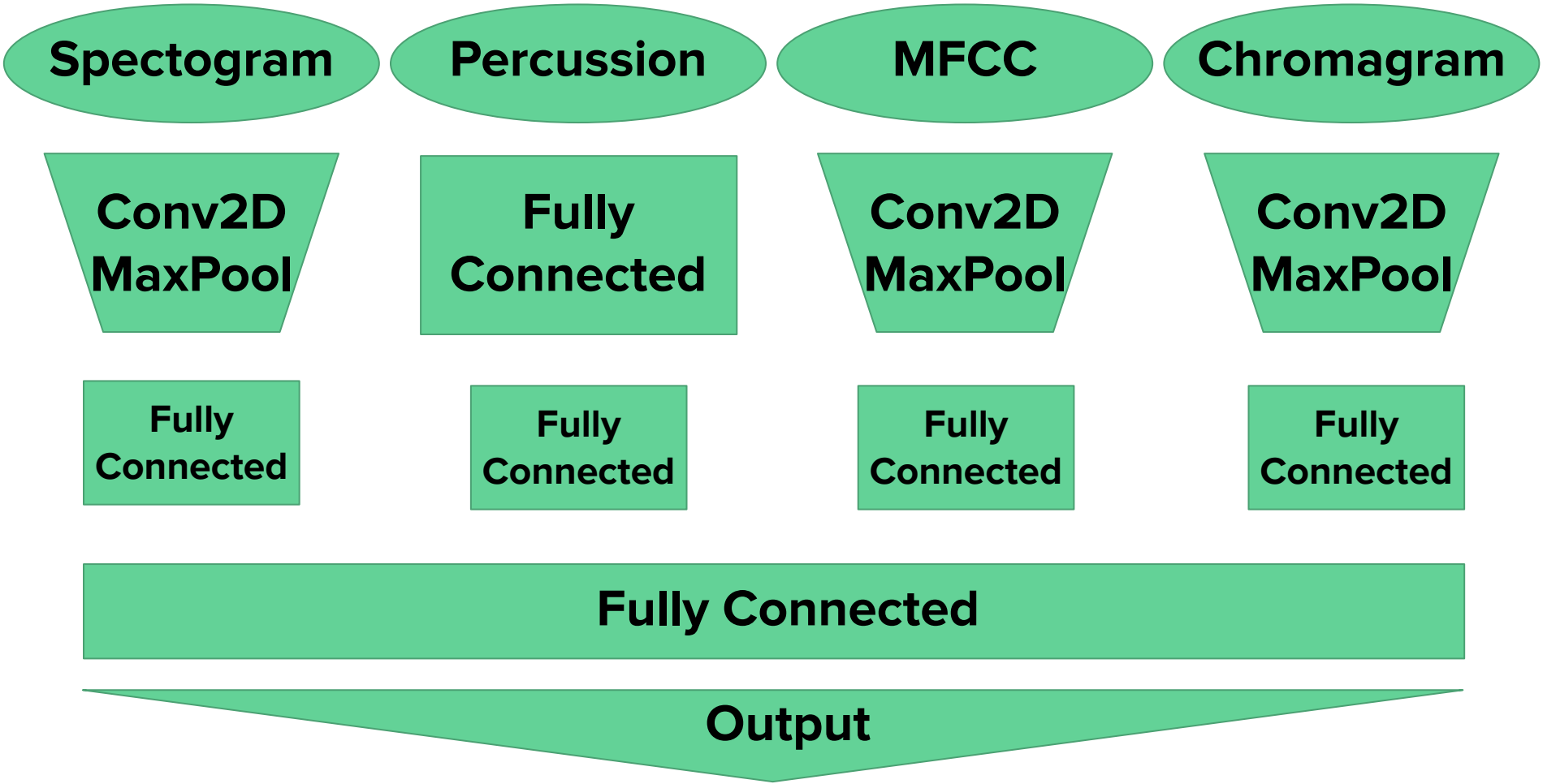
Data

	NN Type	Test Accuracy
Spectogram	2D CNN	**.38**
Percussion	Fully connected	.23
MFCC	2D CNN	.30
Chromagram	2D CNN	.35


Spectogram
Conv2D
MaxPool
Fully
Connected
Output


Spectogram
Percussion
MFCC
Chromagram
Conv2D
MaxPool
Fully
Connected
Conv2D
MaxPool
Conv2D
MaxPool
Fully
Connected
Fully
Connected
Fully
Connected
Fully
Connected
Output
Output
Output
Output

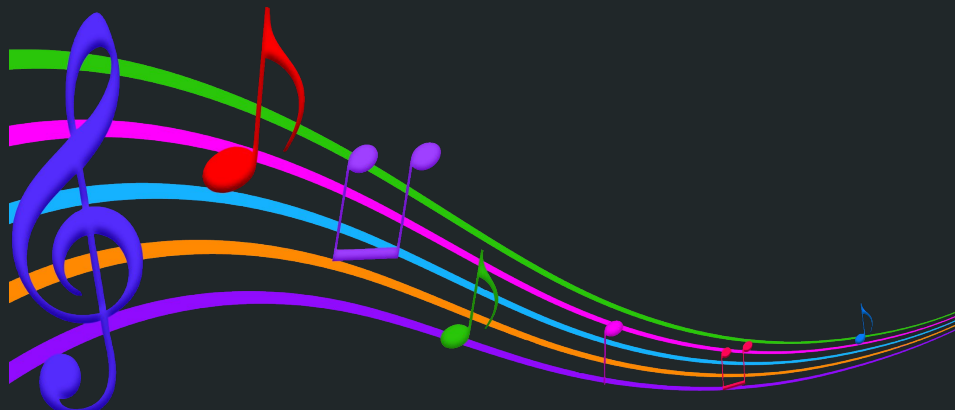

Spectogram
Percussion
MFCC
Chromagram
Conv2D
MaxPool
Fully
Connected
Conv2D
MaxPool
Conv2D
MaxPool
Fully
Connected
Fully
Connected
Fully
Connected
Fully
Connected


Spectogram
Percussion
MFCC
Chromagram
Conv2D
MaxPool
Fully
Connected
Conv2D
MaxPool
Conv2D
MaxPool
Fully
Connected
Fully
Connected
Fully
Connected
Fully
Connected
Fully Connected
Output

42%

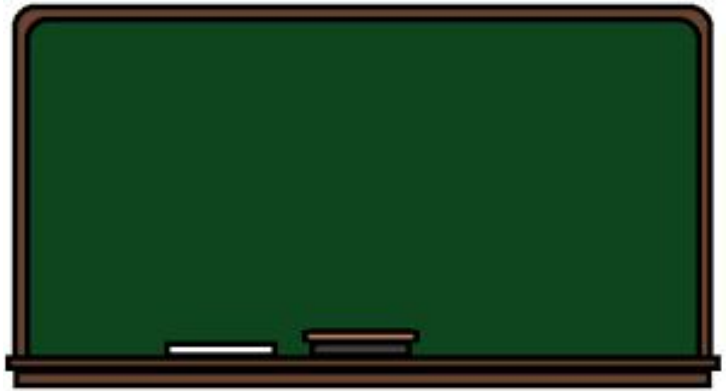
Model prediction accuracy

What’s next?



What’s next?

- Back to the drawing board

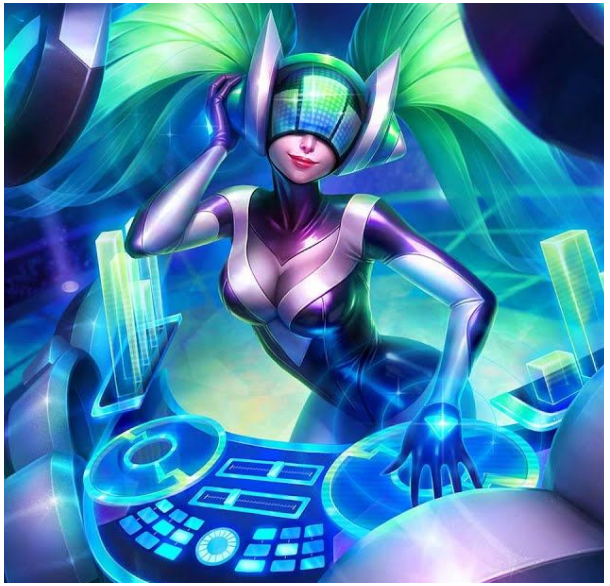


What’s next?

- Back to the drawing board
- Look at what other people have done
  - Other folks claim to be 90%+ with this dataset

What’s next?

- Back to the drawing board
- Look at what other people have done
  - Other folks claim to be 90%+ with this dataset
- Include other genres!



End.