

CLASSICALLY PUNK

Music Classification by Genres using
Deep Learning

Presentation by **Nikita Gaidamachenko**



TABLE OF CONTENTS



01

Task

Goals of the project

02

Model

Why use CNN?

03

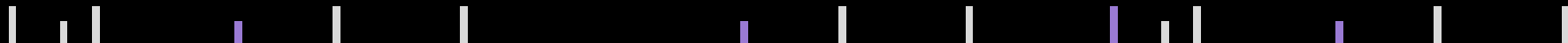
Data

About the dataset and
feature extraction

04

Results

How accurate is the
model?

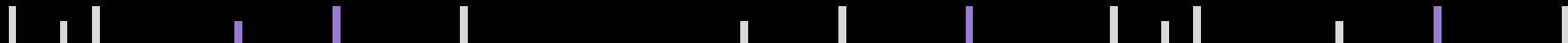




Task



Classify different musical genres from an audio snippet





Task

Build an application in Python that automatically classifies different musical genres from an audio snippet.

Application should:

- Handle large datasets;
- Analyze media files to generate & identify patterns;
- Classify audio files by genres.

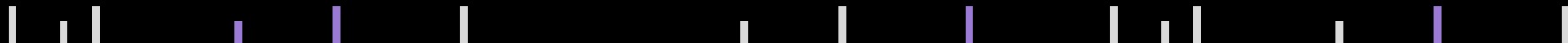


02

Dataset



Dataset and audio features



GENRES



1000 audio files:

- 10 genres presented in *.wav format
- 100 tracks per genre

BLUES

CLASSICAL

COUNTRY

DISCO

HIPHOP



JAZZ

METAL

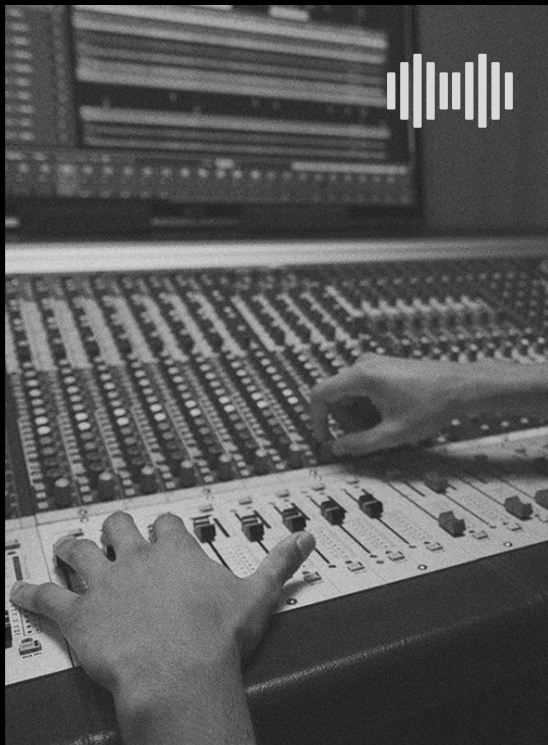
POP

REGGAE

ROCK

ENGINEER PORTFOLIO





AUDIO FEATURES

1. Frequency Domain Features:

- 13 Mel-frequency Cepstral Coefficients
 - 13 is industry standard
- Chroma Short-Time Fourier Transform
- Spectral Centroid
- Spectral Bandwidth
- Spectral Rolloff

In short, these features are responsible for “shape” of music, notes, keys, pitch, range of frequencies, sharpness of sounds.



AUDIO FEATURES

2. Temporal Domain Features:

- Zero Crossing Rate
- RMS Energy

In short, these features are responsible for how often sound waves change their direction, and average volume of sound

3. Tempo – beats per minute



AUDIO FEATURES

4. Timbral Features:

- Spectral Contrast
- Tonnetz

In short, these features help recognize different instruments and voices, and whether melody is sad or happy

In total with mean and variance, over 50 features will be extracted for each track.

03

Model

Convolutional Neural Network



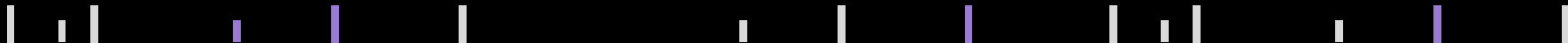


Convolutional Neural Network

The CNN model is made up of layers that process the input data, adding complexity and patterns to understand different features, before making a final prediction about the genre of the music.

Keras – Sequential model

Sequential model allows to build neural networks layer by layer in a sequential manner. Its straightforwardness allows to design and tailor different architectures for tasks such as image recognition or sequence prediction, making it perfect for our program.



MODEL SCOPE



2 SETS OF DATA

- 1,000 30-sec tracks
- 10,000 3-sec slices from those tracks (10 per track)

120 EPOCHS

- Set is passed 120 times
- 120 was chosen as optimal number between loss, resource allocation and results

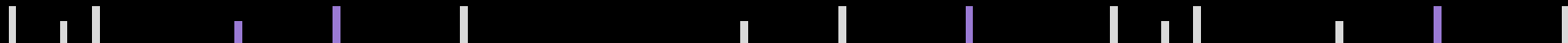


5 LAYERS

- 512 neurons are divided by 2 until they reach 64 at the lowest layer.

20% DROPOUT

- 20% of neurons are randomly dropped after each layer to prevent overfitting





04

Results

Convolutional Neural Network

ENGINEER PORTFOLIO AUDIO ENGINEER PORTFOLIO

11,100

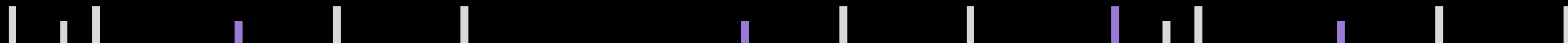
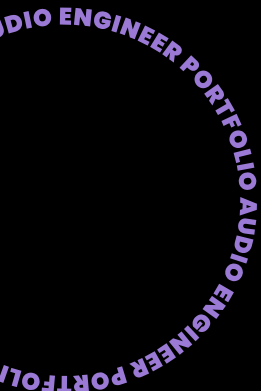
Samples

550,000

Features collected

93%

Accuracy





Accuracy

93.18%
Accuracy

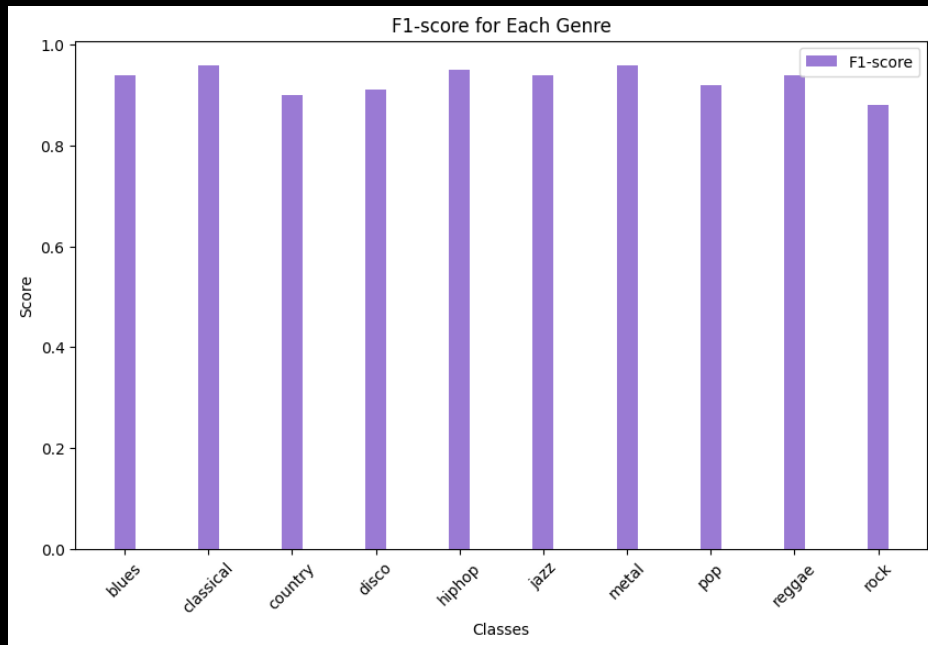
The model correctly identified genre in 93% of cases

96%
Highest

Both Classical music and Metal scored a total of 96%

88%
Lowest

Rock music scored the lowest in both precision and recall – 88% in total





FUTURE IMPROVEMENTS

- Create a CNN model based on spectrographs
- Combine it with pre-trained VGG model for better results
- Minor refactoring



THANKS!

Does anyone have any questions?

Contact [Nikita Gaidamachenko](#) on Discord

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**

Please keep this slide for attribution

