Machine Learning:

Music Genre Classification

Francesco
Ranieri

UNIVERSITÀ
DEGLI STUDI DI BARI
ALDO MORO

01 Introduction

Study case overview

02 Understanding Audio

Few musical modeling concept explanation

03 Dataset

The data used in this project

04 Experiments

Experiments conducted

05 Final Solution

The final model proposed for Music Genre Classification

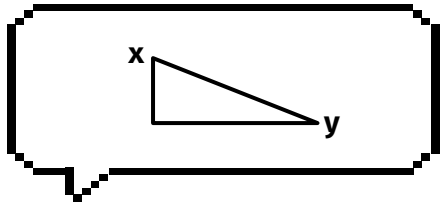
UNIVERSITÀ
DEGLI STUDI DI BARI
ALDO MORO

01 - Introduction

A music genre is a style or a type of music. There are numerous music genres, such as hip-hop, rock, country, pop, etc.

The goal of this project is to build a predictive machine learning model to classify the genre of a given song.

UNIVERSITÀ DEGLI STUDI DI BARI ALDO MORO



02

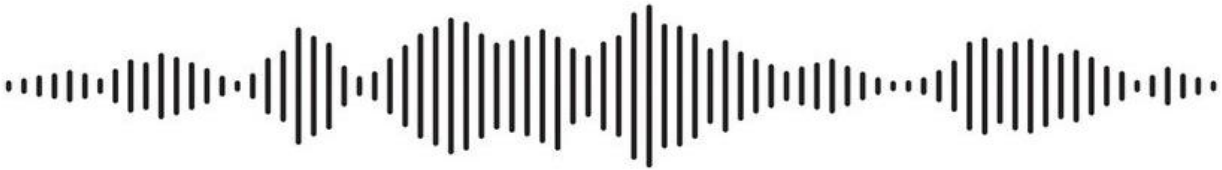
Understanding Audio

Few musical modeling concept explanation

UNIVERSITÀ
DEGLI STUDI DI BARI
ALDO MORO

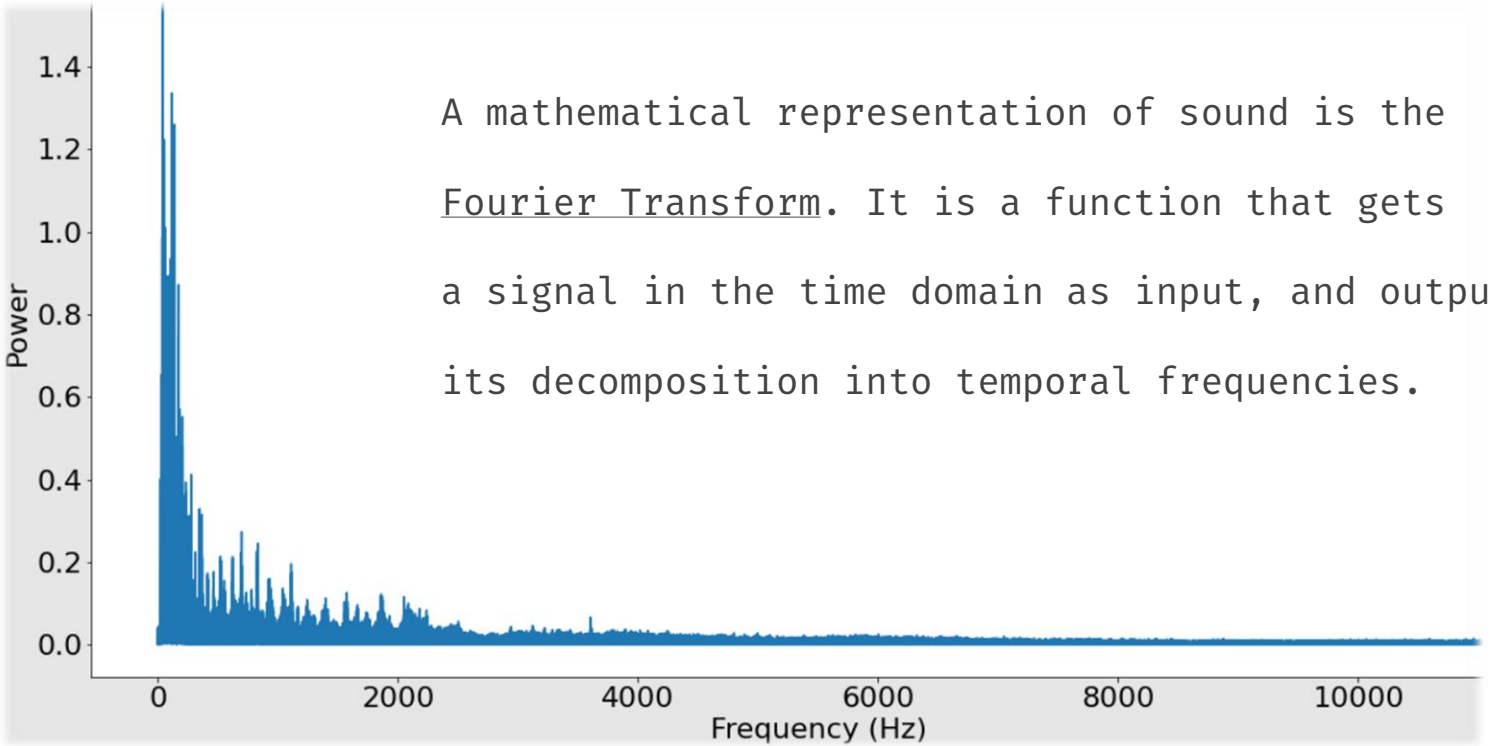
Understanding Audio

Sound is defined as vibrations that travel through the air or another medium as an audible mechanical wave



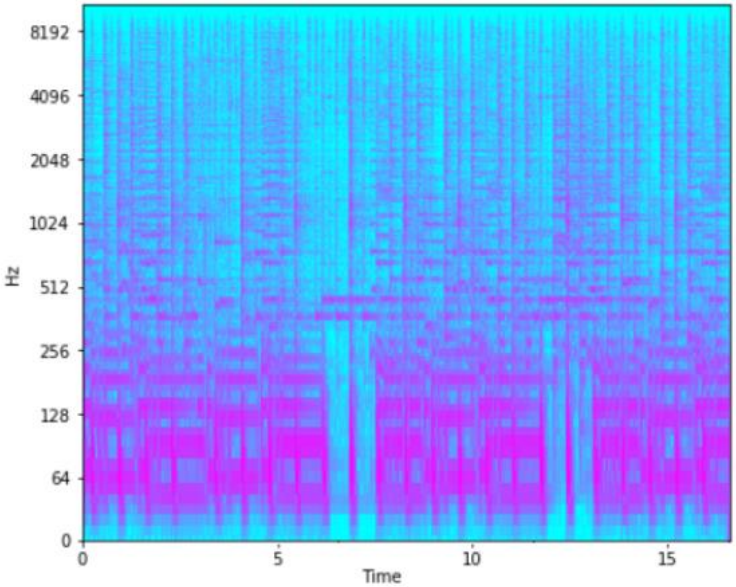
SOUND WAVE

Understanding Audio: Fourier Transform



A mathematical representation of sound is the Fourier Transform. It is a function that gets a signal in the time domain as input, and outputs its decomposition into temporal frequencies.

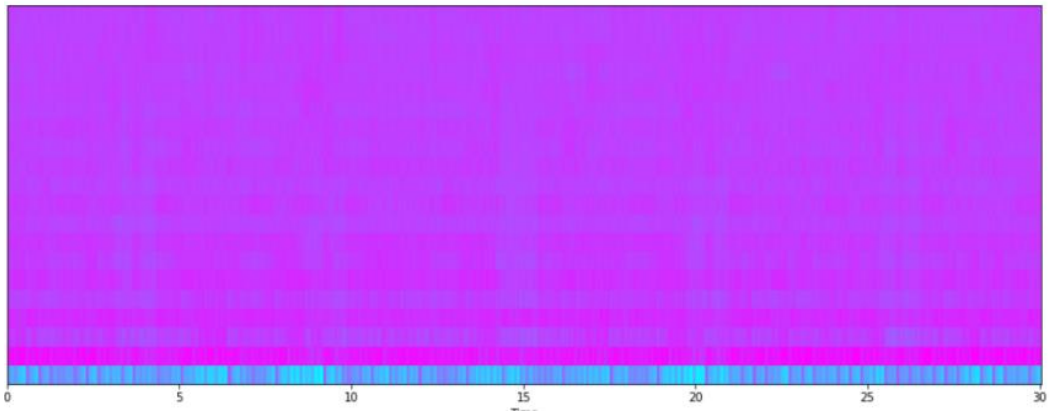
Understanding Audio: Spectrogram



Then we have the Spectrogram which is a visual representation of the spectrum of frequencies of a signal as it varies with time.

Understanding Audio: MFCCs

The Mel frequency cepstral coefficients (MFCCs) of a signal are a small set of features (usually about 10–20) which concisely describe the overall shape of a spectrogram envelope.



03

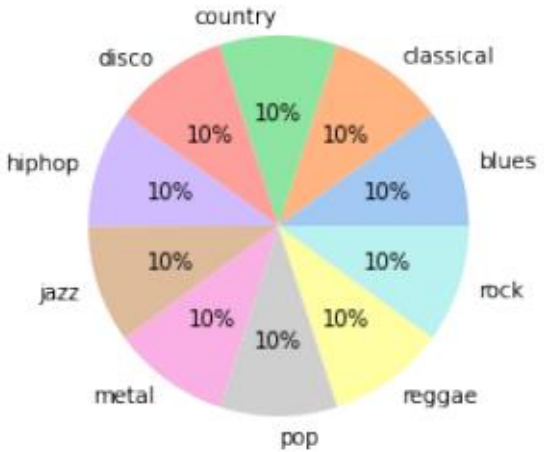
DATASET

The data used in this project

Dataset: GTZAN

The original GTZAN dataset consists of:

- 1000 audio files
- 10 genre classes (disco, classical, …)
- Each song is 30 seconds in duration
- Each song is described by 60 features



In order to have more entries in the dataset, each song is split in 10 songs of 3 seconds each. In this way, the dataset has been expanded to 10,000 songs.

UNIVERSITÀ DEGLI STUDI DI BARI ALDO MORO

Dataset: MFCCs

The MFCCs dataset is inspired by the GTZAN dataset and it tries to represent a song by using only information present in Mel Frequency Cepstral Coefficients (MFCCs) in order to isolate a representative time series information.

It is created by analysing the 3-seconds GTZAN songs and thanks to librosa python package it was possible to extract 13 set of features (MFCCs) per song.
The overall size of the two datasets are the same: 10000 entries.

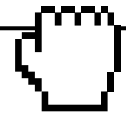
04

Experiments

Experiments conducted

Enviroment

- Google Colab Pro
- GPU Enviroment
- Scikit-Learn
- Keras
- Librosa

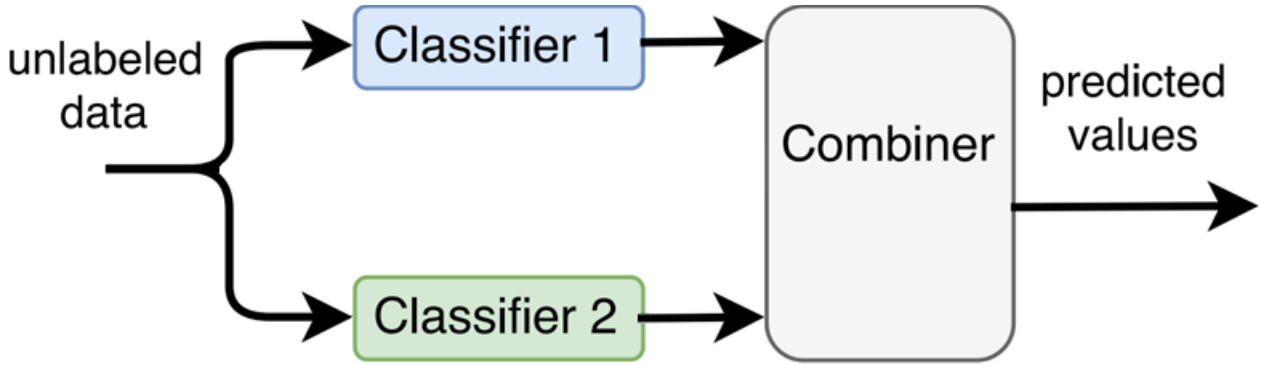


UNIVERSITÀ DEGLI STUDI DI BARI ALDO MORO

Main Idea

The main idea is to train two different models, one per dataset and combine the results of this two classifiers as the final model.

The two chosen classifiers are the best performing models, based on the accuracy on the test set.



Loss Functions

Sparse Categorical Cross-entropy

It is used when class labels are mutually exclusive for each data, meaning each data entry can only belong to one class

$$J(w) = \sum_{i=1}^{size_output} y_i * \log \hat{y}_i$$

- y_i is the true label
- $\hat{y}_i$ is the predicted label

GTZAN Experiments Overview

UNIVERSITÀ DEGLI STUDI DI BARI ALDO MORO

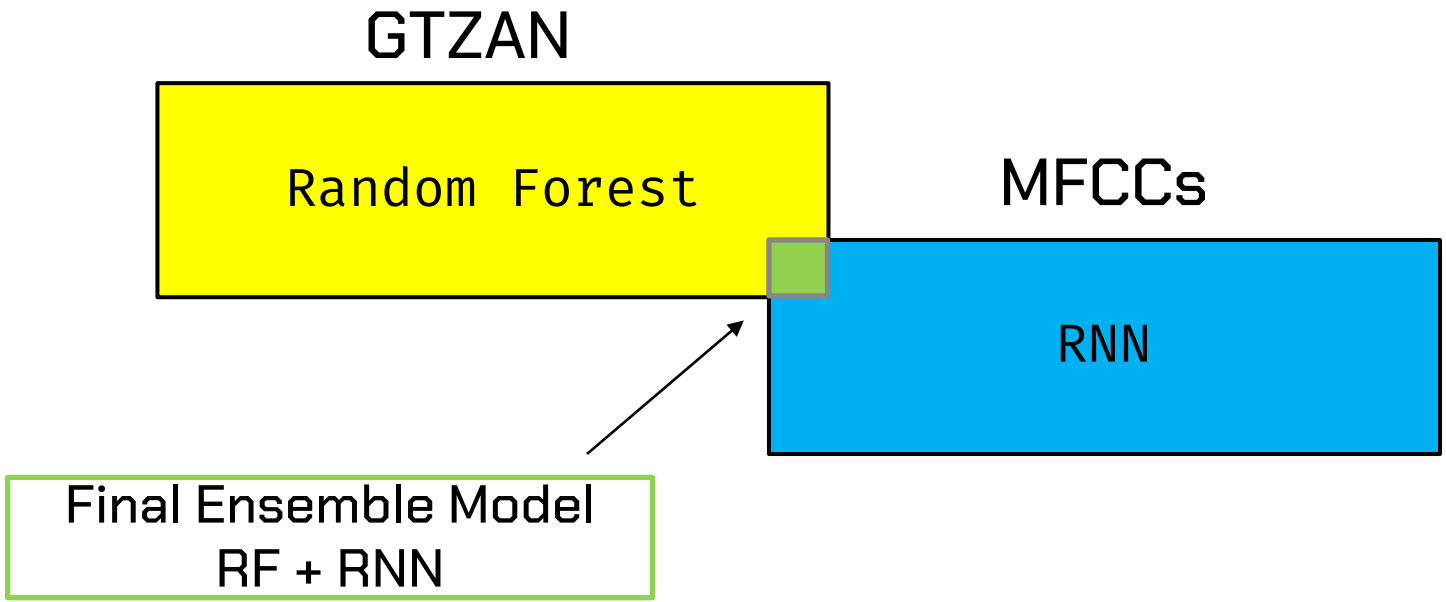
Naive Bayes		42%
Nearest Neighbors		28%
Logistic Regression	max_iter = 10000	45%
Random Forest	n_estimatorsint, default=100	86%
Ada Boost	n_estimators = 100	40%
Gradient Boosting	n_estimatorsint, default=100	82%
Decision Tree		60%
Support Vector Classifier	max_iter = 10000	23%
MLP	hidden_layer_sizes=(128, 64, 8), max_iter=250, learning_rate_init=0.000001	11%
FNN	Dense (32, 16, 10), epoch = 100 relu on dense layer, the last one with SoftMax	81%

MFCCs Experiments Overview

FNN	Dense128, Dropout 0.4, Dense 64, Dense 32, Dense 16, Dense 10	42%
CNN	Conv 2D 8, Maxpooling (2,2), Conv 2D 32, Maxpooling (2,2), Flatten, Dense 128, Dense 64, Dense 10	63%
RNN	LSTM 64, LSTM 64, Dense 64, Dropout 0.3, Dense 10	89%

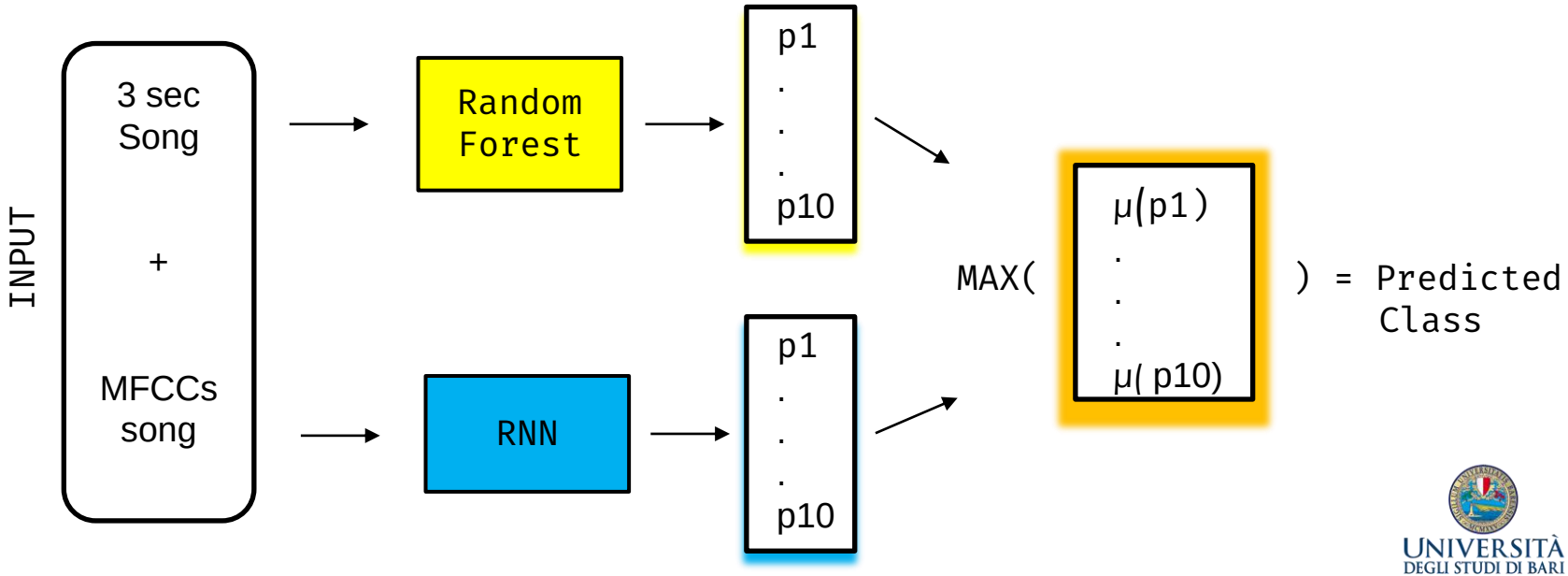
UNIVERSITÀ DEGLI STUDI DI BARI ALDO MORO

Final Solution



The result is an ensemble learning approach by using the chosen models. The prediction of the ensemble model is the genre class with higher mean probability of the two models.

The ensemble model scores on test set an accuracy of 87%.



UNIVERSITÀ DEGLI STUDI DI BARI ALDO MORO

Thanks!

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