VIETNAM NATIONAL UNIVERSITY HO CHI MINH CITY
UNIVERSITY OF INFORMATION TECHNOLOGY

UIT
UNIVERSITY OF INFORMATION TECHNOLOGY

PROJECT REPORT

SUBJECT: COMPUTATIONAL THINKING

TOPIC

MUSIC GENRE CLASSIFICATION

Lecturer: Ngo Duc Thanh

Team's member: Tran Vi Hao - 19521482
To Thanh Hien - 19521490
Nguyen Duong Hai - 19521464

Class: CS117.L22.KHCL

Ho Chi Minh City, July 16, 2021

Content

Part 1: Introduction and description of the problem

Categorizing music by genres is a necessity as online (and offline) music listening platforms grow and become more popular. Currently, the music listening platforms available on the market mainly focus on classifying the information based on the label attached to a certain song, not on the classification by rhythm and knowledge. the sound structure contained in that song, so that the user's inclusion of an arbitrary song is obtained from another source, not provided by the music listening platform itself, and wishes to be classified as a music genre. very difficult. For that reason, the above problem was born to solve the problem of classifying music genres based on the rhythm of a song.

1. Input and Output

In this project, my group will limit the problem as follows:

- Input:
  + There will be **only one** input
  + It's a piece of music **without lyrics**
  + To ensure that the input is definitely a piece of music, the input must be an audio clip with **rhythm** with a length of at least 30 seconds and not more than 5 minutes
  + Can follow any audio format such as mp3, wav,…
  + The input piece must be in **one of the following ten** musical genres:
    - Blues
    - Classical
    - Country
    - Disco
    - Hiphop
    - Jazz
    - Metal
    - Pop
    - Reggea
    - Rock
- Output: displayed as text, output the name of one of the ten music genres above corresponding to the input music.
- Abstraction: classifying the music genre (limited to the ten musical genres mentioned above) of any piece of music, that piece of music must satisfy all the binding requirements as stated in the Input section. The final result of the problem we get is the name of the music genre corresponding to the piece of music we are looking for.

2. Dataset

In this project, the dataset used is the GTZAN dataset available on Kaggle. This dataset was used in a famous article on genre classification in 2002, it consists of 1000 30-second audio files in 16-bit WAV 22050Hz Mono format divided into ten folders. named after the ten musical genres mentioned above. Each folder contains 100 corresponding files.

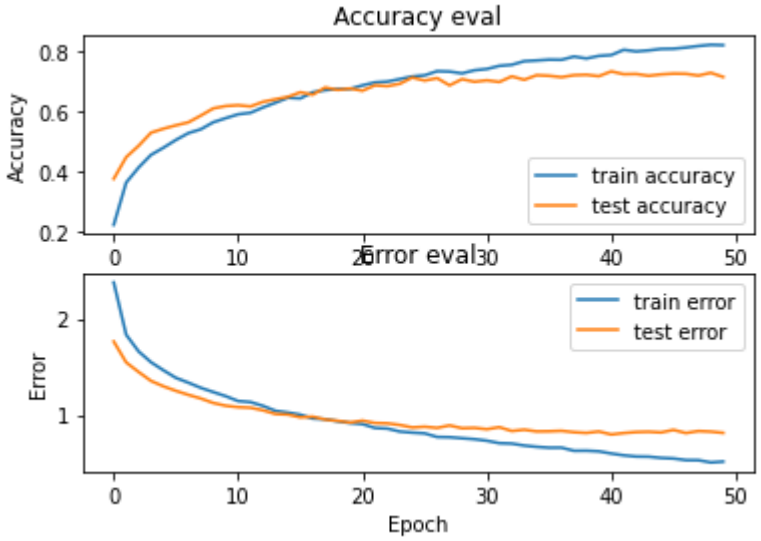
Category	Quantity
Blues	100
Classical	100

Country	100
Disco	100
Hiphop	100
Jazz	100
Metal	100
Pop	100
Reggea	100
Rock	100
Total:	***1000***

Part 2: Solution/Algorithm

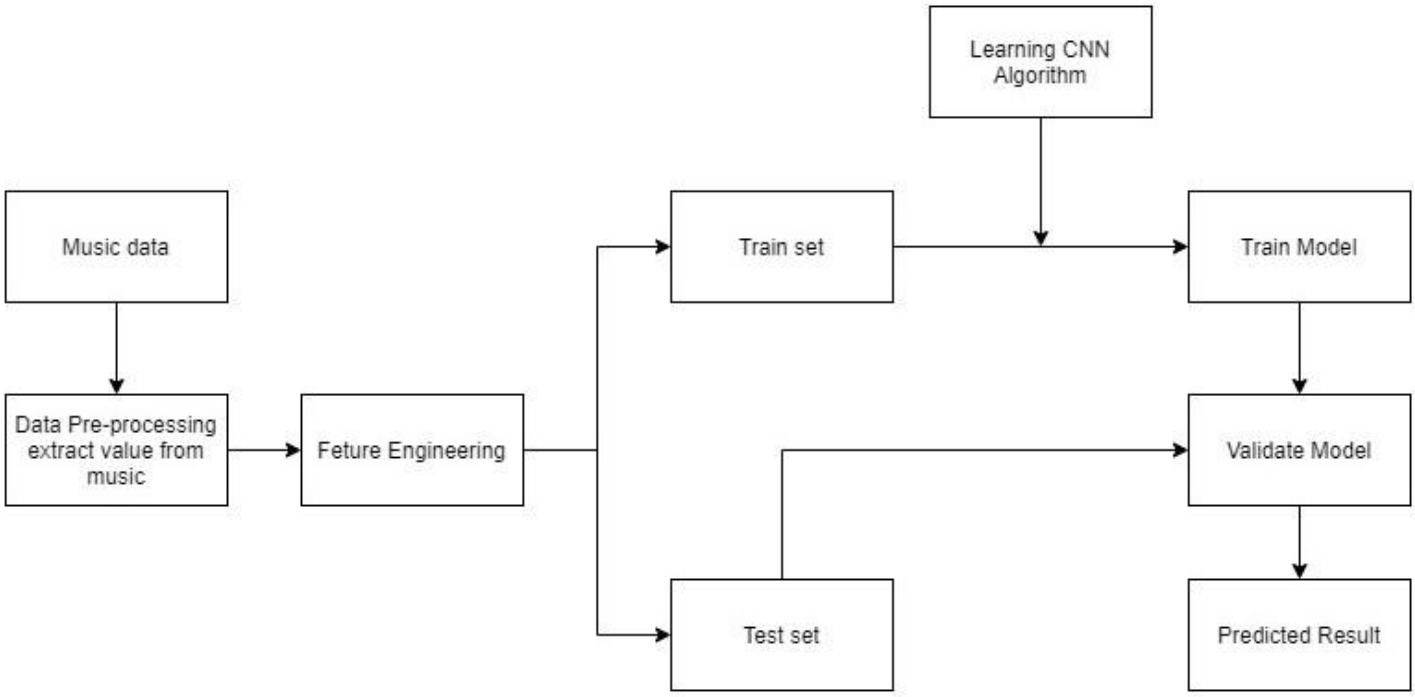
My team will use ***Convolutional Neural Network (CNN)*** to solve this problem, the accuracy of the model is about 70%.

```
[ ]  # evaluate model on Test Set
     test_loss, test_acc = model_cnn.evaluate(X_test, y_test, verbose=2)
     print('\nTest accuracy:', test_acc)

     79/79 - 1s - loss: 0.8364 - accuracy: 0.7036

     Test accuracy: 0.7036443948745728
```

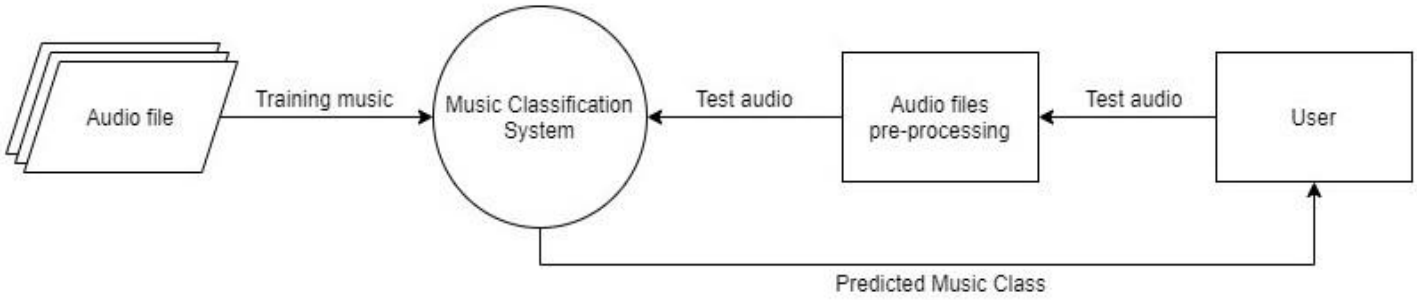


1. Flowchart

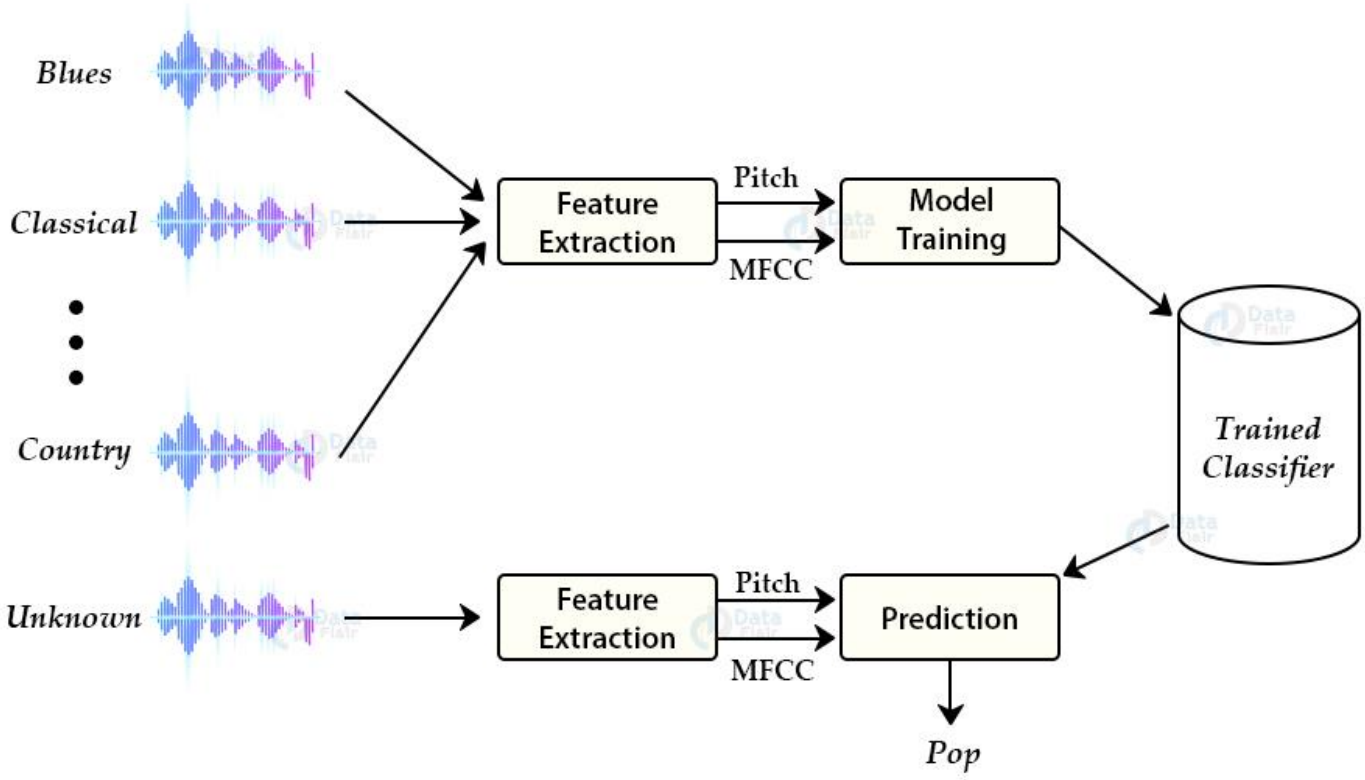
Below is a flowchart showing the algorithm to solve the problem, my team will divide the problem into 2 phases (phases) to make it easier to follow.



Pic 1: Training Phase



Pic 2: Test Phase



Pic 3: Full Phase

Code (Google Colab) :
https://colab.research.google.com/drive/1TaCJqqmauSlu6Nr0HoSOhNDrak5WEgMb?usp=sharing

2. Evaluate

First, this problem encounters two rather big problems:

1. The defined music genres are not clear, there are many inadequacies and contradictions in the way of determination. So much so that people often argue about the genre of a song.

2. Extracting features separate from audio data to put into a model is not a small task.

When the model learns from the GTZAN dataset, it is easy to see that this model almost never predicts Blues music, but mostly misclassifies it into Jazz and Rock music. This can be explained as follows, Jazz and Blues are very similar musical styles, Rock music is also heavily influenced and actually born from Blues music.

Model also had difficulty distinguishing between Reggae and Hiphop. Half of the misclassifications for Reggae are Hiphop and vice versa. Again, this means that Reggae music is heavily influenced by hip-hop music and has similar characteristics.

The model misclassified several genres as Rock, specifically Blues and Country. Not surprisingly, there are many subgenres of rock music that branch out into other genres.

This shows us that the children's model is facing the same difficulties as a human. It is clearly learning some of the distinguishing elements of musical genres, but it is struggling with those that share characteristics with other genres. Again, this goes back to the first issue, and that is the nature of musical genres. They are very difficult to distinguish.

In addition, when we put the model into the actual classification, we continued to encounter some rather painful objective problems. It is almost impossible for the input music to belong to only a

certain genre of music, the piece of music can be mixed from many different music genres, making it difficult to determine the music genre.

However, with an accurate prediction rate of 70%, this model can also be considered as good, there will be a lot of room for improvement in the future if you want to get a higher accuracy rate. In general, being able to produce a predictive model with the highest possible accuracy rate on music genre classification will bring a lot of commercial value, and in the future can open many doors. new development directions such as music classification according to people's moods. This will be a “fertile land” and full of fun for us to continue to explore.

Part 3: Graphic Organizer and Project Justification

In this part, my group would like to present in the Final project form of the course Problem Solving Using Computational Thinking by University of Michigan on Coursera.

Project Justification

After you complete the graphic organizer below, use this project justification document to explain how you used computational thinking in your project.

Problem Identification.

For each iteration of your problem, please explain how you arrived at your identified problem.

In Interation 1, my group raised a question with a rather large implication, through which issues need to be clearly defined when solving this problem: What is the input input? Does the input have any special requirements? What is Output? Based on what to choose the most suitable method to solve the problem?

From Interation 2 onwards, the problems are expanded when learning methods deeply about the problem and also the to solve it, thus breaking down the problem into smaller solvable problems. These connections are made possible by the common pattern recognition process defined in Interation 1.

Decomposition.

For each iteration where you decomposed an identified problem, please explain how this decomposition helped you solve your identified problem.

Decomposition stages are very important in all 6 Interations. In Interation 1 the group decomposes into 4 smaller problems, in Interation 2 it's 2, in Interation 3 it's 2, in Interation 4 it's 4, in Interation 5 it's 2 and Interation 6 is 2. Together, all Those sub-problems have made the main problems clearer and clearer, helping the group to have a more specific view of the nature of the big problem, ask questions and come up with a specific solution to help solve the problem. larger is more complete.

Pattern Recognition.

For each iteration where you recognized patterns in data, please explain how these patterns helped you solve your identified problem.

For this project, my team used the pattern recognition phase in Interation 1. This is an important stage to help us make clear the problem, identify the main problems to be solved, and then decompose. broken down into smaller subproblems, then just by solving these subproblems, the big problem is also solved.

Abstraction.

For each iteration where you abstracted information, please explain how abstraction allowed you to solve your identified problem.

Abstraction phases are used in all Interations. For example, the abstraction in Interation 4 helped my team focus on only two methods of Librosa and MFCCs to process data, and both methods are extremely effective. These Abstraction stages helped the group focus only on the most important and relevant information to solve the problem, not too much on less important information, avoiding wasting time.

Iteration 1

Problem Identification

I have a song that I want to categorize music genres (note that this song is not supported by the music platform I am using), the online and offline music platforms available on the market today. Nowadays, almost everyone has difficulty with this problem, if not more specifically, there is no foundation for this classification function to be born completely. So, to solve this problem, what do I need to determine?

To set up your identified problem

Decomposition (How would you break down your problem into sub-problems?

To solve this problem, I need to clarify:

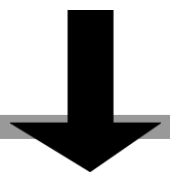
+ What is the input of this problem?
+ What special requirements does the input input have?
+ What is output?
+ Based on what to choose the most suitable

Pattern Recognition (Are there related solutions to draw on?)

Based on the previous solutions, it can be seen that the common point of this problem is to classify a piece of melodic audio into the available music genres in the world, regardless of whether the audio segment has any How's the tune? But it will certainly fall into one of those musical genres.

Abstraction (How would you abstract this problem?)

Graphic Organizer



Iteration 2

Problem Identification

In this problem, I have reduced the problem to just solving the classification of an audio clip with a length of at least 30 seconds and no more than 5 minutes with any audio format, and I also to summarize the classification in just 10 music genres: Blues, Classical, Country, Disco, Hiphop, Jazz, Metal, Pop, Reggea, Rock. But how to ensure that the input audio that I put into it is definitely a piece of music or a song, most likely it's just pure sounds, or maybe noise. What do I need to do to solve this problem?

To set up your identified problem

Decomposition (How would you break down your problem into sub-problems?)

In order for Input to be a piece of music that meets the desired requirements, I need:

+Make sure the input is an audio clip with a rhythmic melody to make sure it's music

+Make sure the piece of music belongs to one of the ten genres that I can classify, the length must be greater than 30 seconds but less than 5 minutes

Pattern Recognition (Are there related solutions to draw on?)

Abstraction (How would you abstract this problem?)

Specifies that the input input is exactly a piece of music or a song.

Graphic Organizer

Iteration 3

Problem Identification

In this problem, I will use Machine Learning to solve the problem. So the question here is, among so many algorithms out there, which ML algorithm will I choose to optimize the problem? Should I choose an algorithm that specializes in audio processing, or test as many algorithms as I can and choose the one that gives me the highest accuracy rate?

To set up your identified problem

Decomposition (How would you break down your problem into sub-problems?)

To solve this problem, I need to know:

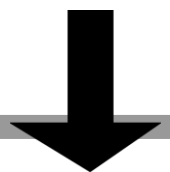
+ What algorithm did the predecessors use?
+ How will those algorithms if run on the same dataset?

Pattern Recognition (Are there related solutions to draw on?)

Abstraction (How would you abstract this problem?)

Determine the accuracy of each algorithm on the same dataset (here is the GTZAN dataset), thereby choosing the most optimal algorithm for the problem, and in this problem I finally chose CNN.

Graphic Organizer



Iteration 4

Problem Identification

The GTZAN dataset includes 1000 wav files, so how do I process the data to prepare for the Training Phase?

To set up your identified problem

Decomposition (How would you break down your problem into sub-problems?)

At this time, I need to learn more about a Python module that is Librosa specialized in audio processing to:

+Load data

+Calculation Spectrum

Alternatively, use MFCCs to extract the features of each audio file and save the information into a single json file.

Pattern Recognition (Are there related solutions to draw on?)

Abstraction (How would you abstract this problem?)

Using the Librosa module and MFCCs for the data preprocessing step, is an important step that can affect the accuracy rate of the algorithm.

Graphic Organizer

Iteration 5

Problem Identification

In addition, after the data learning algorithm is finished, we come to the Test Phase (the phase of testing the model on a specific song), whether the input is given to satisfy the requirements or not is very important, in This step has a question, what if the music exceeds the time allowed? And will a piece of music that includes the singer's voice in the model give good results?

To set up your identified problem

Decomposition (How would you break down your problem into sub-problems?)

I need:

+Make sure the input is a piece of music without lyrics

+ Process input before putting it into the model

Pattern Recognition (Are there related solutions to draw on?)

Abstraction (How would you abstract this problem?)

Process input to ensure that the input meets the requirements that I have set, for example: input must be instrumental music, longer than 30 seconds but less than 5 minutes, etc…

Graphic Organizer

Iteration 6

Problem Identification

Input processing (Audio files pre-processing)

To set up your identified problem

Decomposition (How would you break down your problem into sub-problems?)

The input processing is broken down into 2 main jobs:

+Calculate start and finish sample for current segment

+Extract MFCC

Pattern Recognition (Are there related solutions to draw on?)

Abstraction (How would you abstract this problem?)

Input processing involves truncation and feature extraction of an audio clip to obtain an input in an “ideal” state when fed to the model.

Graphic Organizer

Part 4: Table of assignment of tasks and Completion level

Student ID	Name	Tasks	Completion level
19521482	Tran Vi Hao	Break down the big problem into subproblems and write a report	100%
19521490	To Thanh Hien	Trainning and Test model	100%
19521464	Nguyen Duong Hai	Translate the report into English and support the analysis of large problems into subproblems	100%

Part 5: Documents

1. https://www.youtube.com/watch?v=OpAdgx9wpXc
2. https://www.youtube.com/watch?v=a7rR9lOXUcc&t=2092s
3. https://www.youtube.com/watch?v=Tjl-Qn6ausw
4. https://data-flair.training/blogs/python-project-music-genre-classification/
5. In addition, you can read the Vietnamese version of the report here.