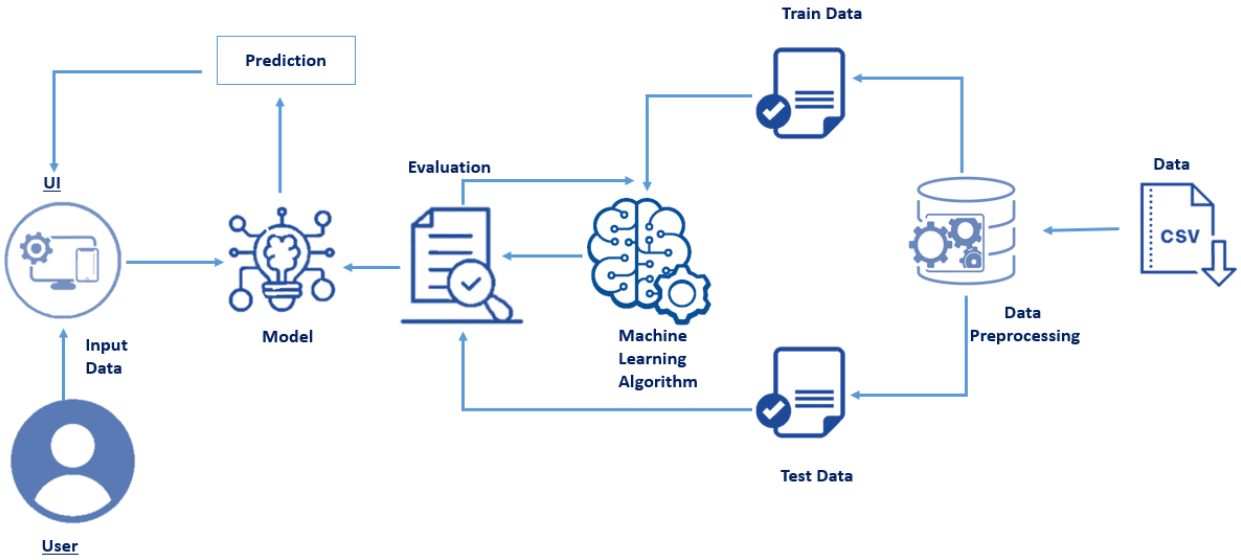
SESHADRI RAO GUDLAVALLERU ENGINEERING COLLEGE

MACHINE LEARNING BASED MUSIC GENRE CLASSIFICATION ON SPOTIFY DATA USING IBM WATSON



SUBMITTED BY:

NAME : K.DEVI SANKAR
PIN NO : 19481A0324

Content

1 INTRODUCTION

1.1 Abstract:

Throughout the course we mainly focused around computer vision tasks and a little bit of NLP. I decided to reach out a new field and use the meachine learning techniques I learned on the field of sound processing. This paper discuss the task of classifying the music genre of a sound sample

1.2 Overview:

Music is like a mirror, and it tells people a lot about who you are and what you care about, whether you like it or not. We love to say “you are what you stream”.

Companies nowadays use music classification, either to be able to place recommendations to their customers (such as Spotify, Soundcloud) or simply as a product (for example Shazam). Determining music genres is the first step in that direction. Machine Learning techniques have proved to be quite successful in extracting trends and patterns from a large pool of data. The same principles are applied in Music Analysis also.

2.LITERATURE SURVEY

2.1 Existing problem

You’ll be able to understand the problem to classify if it is a regression or a classification kind of problem.

You will be able to know how to pre-process/clean the data using different data preprocessing techniques.

You will able to analyse or get insights into data through visualization.

Applying different algorithms according to the dataset and based on visualization.

You will able to know how to find the accuracy of the model.

You will be able to know how to build a web application using the Flask framework.

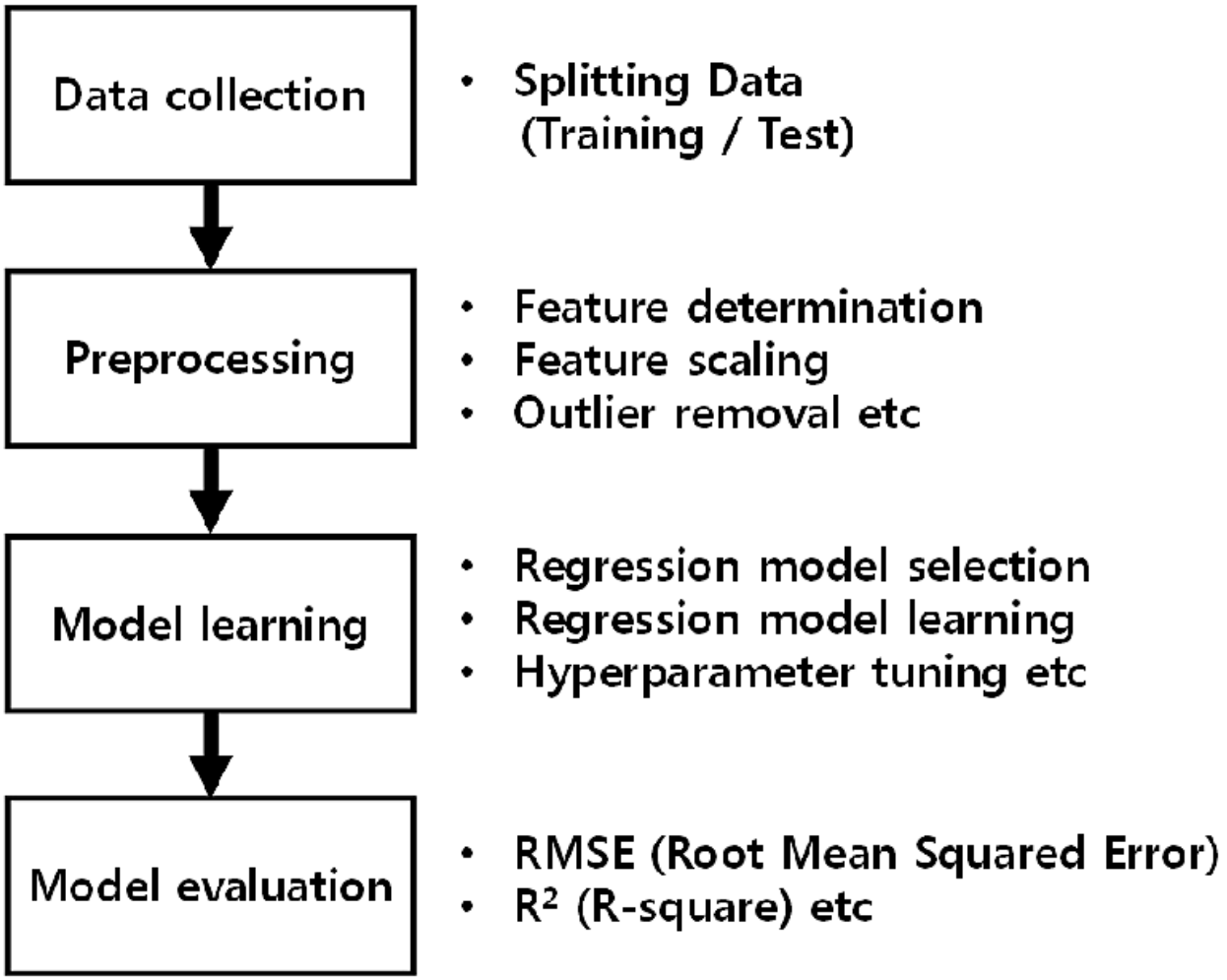
2.2 Proposed solution

By doing this project we classify if it is a regression or a classification kind of problem.And also able to analysis or get insights into data the In this project

we are using the different algorithms according to the dataset and based on visualization

3 THEORITICAL ANALYSIS

3.1 Block diagram



3.2 Hardware / Software Requirements

Recommended System Requirements

Processors:Intel® Core™ i5 processor 4300M at 2.60 GHz or 2.59 GHz (1 socket, 2 cores, 2 threads per core), 8 GB of DRAMIntel® Xeon® processor E5-2698 v3 at 2.30 GHz (2 sockets, 16 cores each, 1 thread per core), 64 GB of DRAMIntel® Xeon Phi™ processor 7210 at 1.30 GHz (1 socket, 64 cores, 4 threads per core), 32 GB of DRAM, 16 GB of MCDRAM (flat mode enabled)

- Disk space: 2 to 3 GB
- Operating systems: Windows® 10, macOS*, and Linux*

Minimum System Requirements

- Processors: Intel Atom® processor or Intel® Core™ i3 processor
- Disk space: 1 GB
- Operating systems: Windows* 7 or later, macOS, and Linux
- Python* versions: 3.9

Software requirements:

anaconda navigator:

Anaconda is an open-source distribution for python and R. It is used for data science, machine learning, deep learning, etc. With the availability of more than 300 libraries for data science, it becomes fairly optimal for any programmer to work on anaconda for data science.

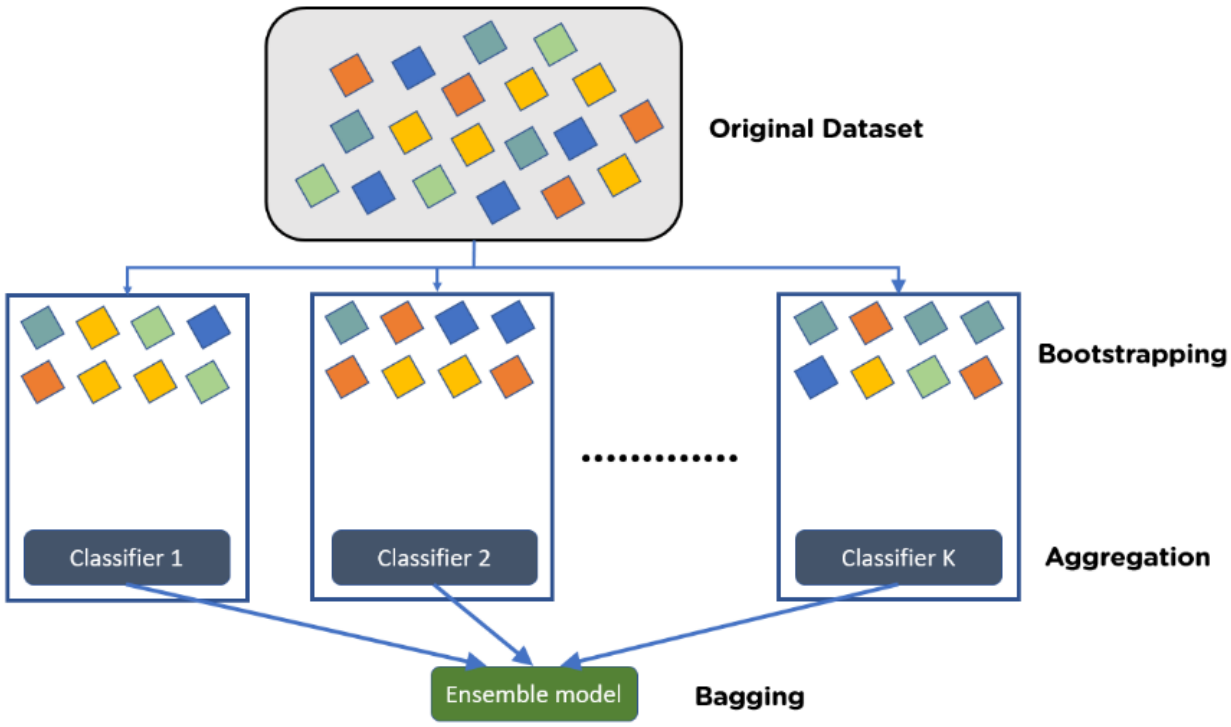
Pycharm:

PyCharm is a dedicated Python Integrated Development Environment (IDE) providing a wide range of essential tools for Python developers, tightly integrated to create a convenient environment for productive Python, web, and data science development.

4 EXPERIMENTAL INVESTIGATIONS

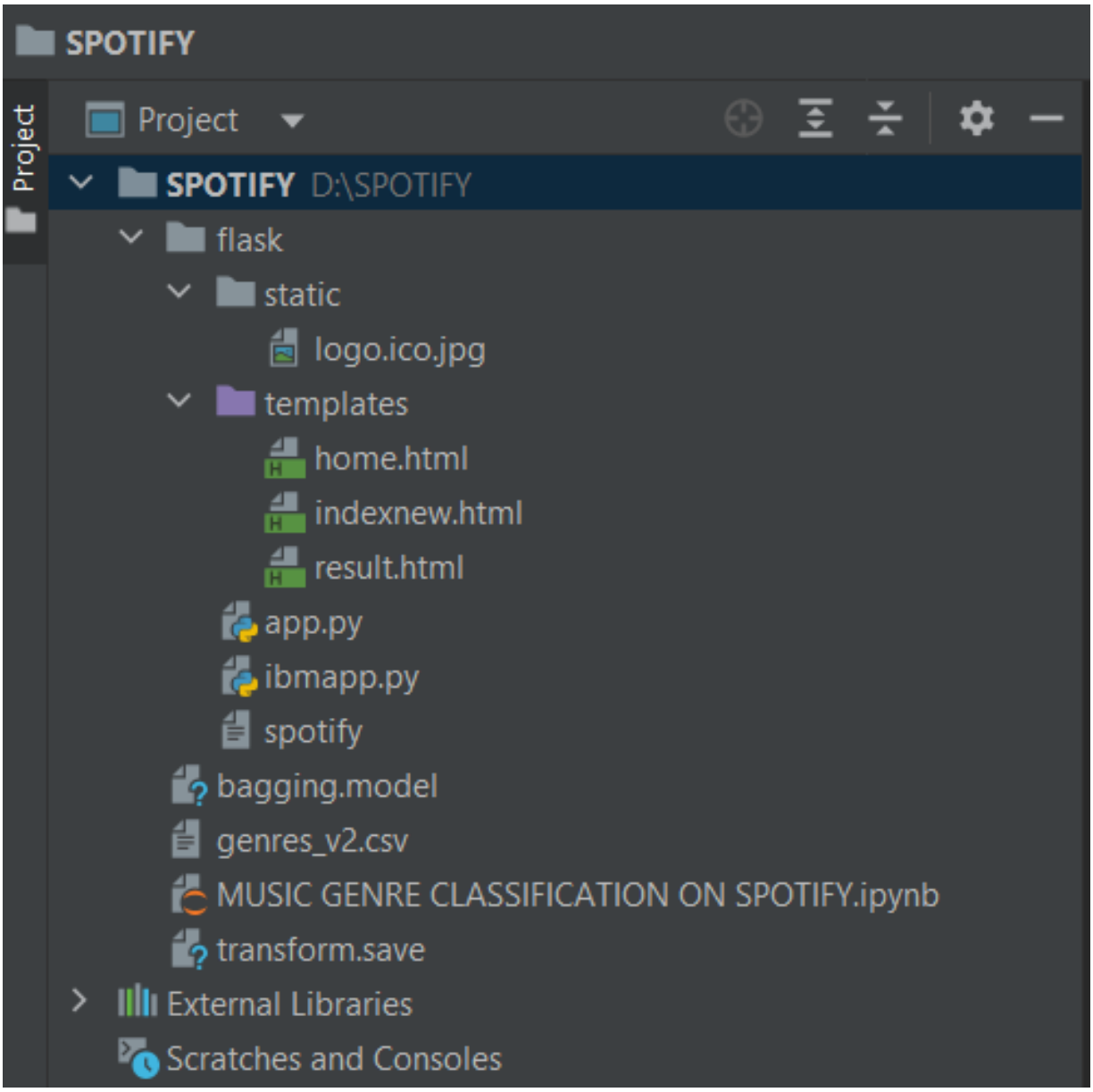
BAGGING CLASSFIER :

Bagging, also known as Bootstrap aggregating, is **an ensemble learning technique that helps to improve the performance and accuracy of machine learning algorithms**. It is used to deal with bias-variance trade-offs and reduces the variance of a prediction model.



- Consider there are n observations and m features in the training set. You need to select a random sample from the training dataset without replacement
- A subset of m features is chosen randomly to create a model using sample observations
- The feature offering the best split out of the lot is used to split the nodes
- The tree is grown, so you have the best root nodes

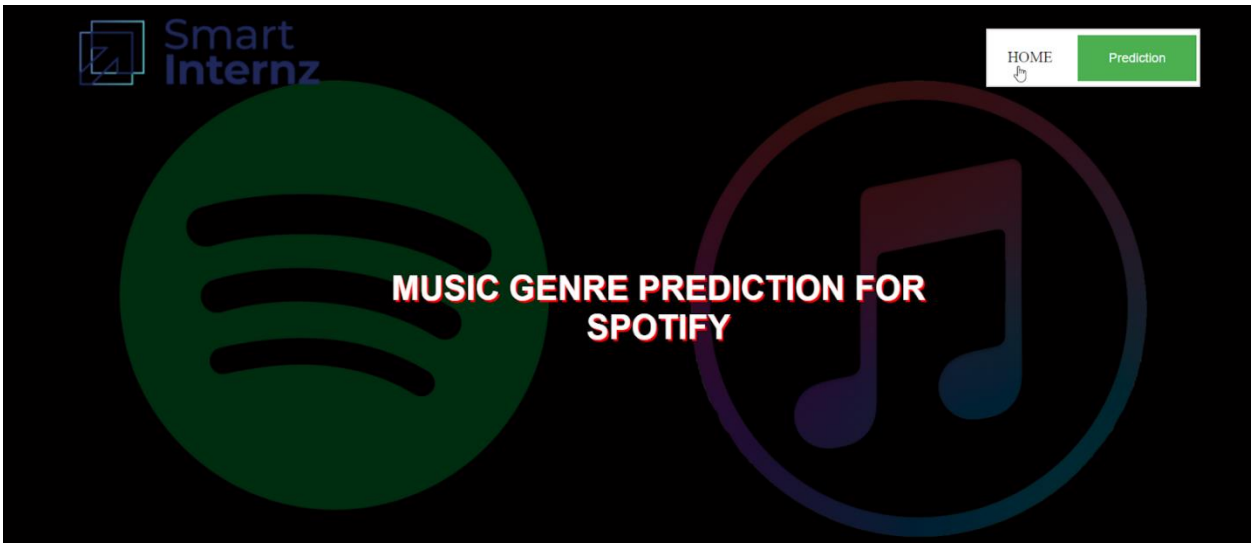
5 FLOW CHART


SPOTIFY
Project
Project
SPOTIFY D:\SPOTIFY
flask
static
logo.ico.jpg
templates
home.html
indexnew.html
result.html
app.py
ibmapp.py
spotify
bagging.model
genres_v2.csv
MUSIC GENRE CLASSIFICATION ON SPOTIFY.ipynb
transform.save
External Libraries
Scratches and Consoles

6 RESULT

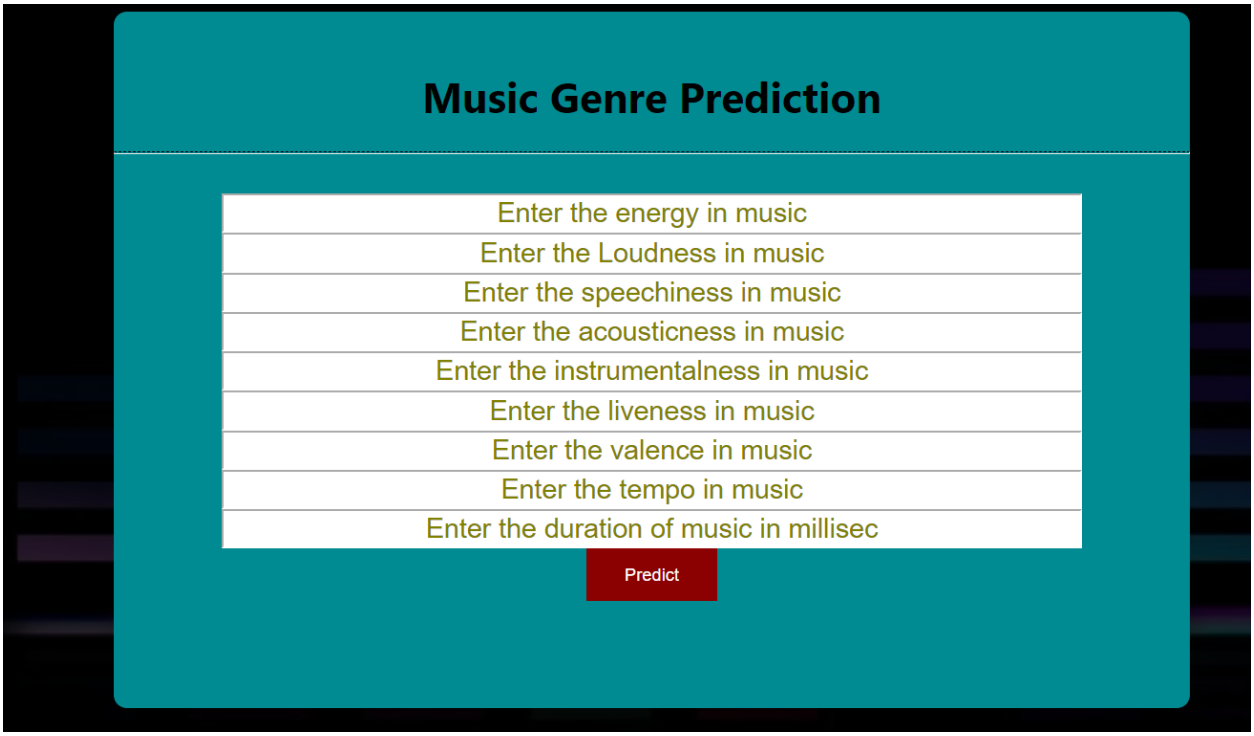
For Getting Music Genre prediction for Spotify. We will Enter Our Results

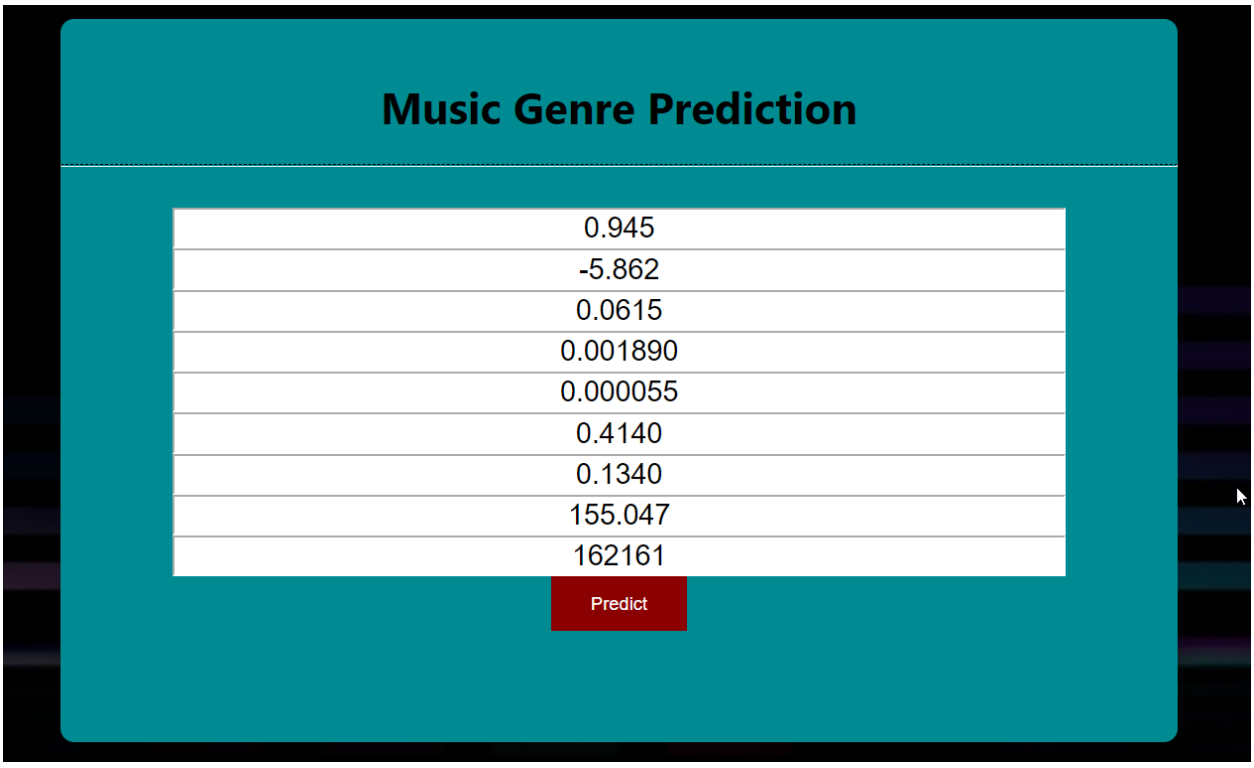
Of All The Fields Enter Here And Check The Probability :



If u click on the Home button it will refresh the home page
If u click on the prediction button it will redirect to new web page

Enter input Of All The Fields Enter Here And Check The Probability


Music Genre Prediction
Enter the energy in music
Enter the Loudness in music
Enter the speechiness in music
Enter the acousticness in music
Enter the instrumentalness in music
Enter the liveness in music
Enter the valence in music
Enter the tempo in music
Enter the duration of music in millisec
Predict


Music Genre Prediction
0.945
-5.862
0.0615
0.001890
0.000055
0.4140
0.1340
155.047
162161
Predict

7 ADVANTAGES OF THE PROJECT

ADVANTAGES :

1. The main thing to identify and divide the audio into different features is amplitude and frequency that changes within a short span of time.
2. We can visualize the audio frequency wave of amplitude and frequency with respect to time in form of a wave plot that can be easily plotted using librosa.
3. MFCCs total provides 39 features related to frequency and amplitude. In that 12 parameters are related to the amplitude of frequencies. It means it provides us with enough frequency channels to analyze audio and this is the reason MFCCs are used everywhere for feature extraction in audios.
4. The key working of MFCC is to remove vocal excitation (pitch information) by dividing audio into frames, make extracted features independent, adjust the loudness, and frequency of sound according to humans, and capture the context.

8 CONCLUSION

Automatic genre classification is a difficult and problematic task that none the less has important value in terms of both pure research and commercial application. Continuing research in automatic genre classification has much to of-

fer, as does parallel research involving other aspects of musical similarity. Automatic genre classification performance appears to have fallen into a local maximum recently, and serious modifications to the approaches used are needed in order to realize further improvements.

10 BIBILOGRAPHY

Installation of Anaconda Navigator:

https://www.youtube.com/embed/5mDYijMfSzs

Installation Of Pycharm Professionals:

https://www.youtube.com/embed/z73PyNDgVyQ

Installation Of Python Packages:

https://www.youtube.com/embed/akj3_wTploU

Data Collection:

https://www.kaggle.com/datasets/lucascantu/spotify-dataset-albums/download?datasetVersionNumber=1

Data Pre-processing:

https://thesmartbridge.com/documents/spsaimldocs/Datapreprocessing.pdf

Handling Null Values:

https://towardsdatascience.com/7-ways-to-handle-missing-values-in-machine-learning-1a6326adf79e

Data Visualization:

https://www.youtube.com/embed/TLdXM0A7SR8

Splitting Dependent And Independent Columns:

https://www.youtube.com/embed/A_V6daPQZIU

Splitting The Data Into Train And Test:

https://www.youtube.com/embed/xgDs0scjuuQ

Training And Testing The Model:

https://www.youtube.com/embed/yIYKR4sgzI8

Model Evaluation:

https://towardsdatascience.com/the-5-classification-evaluation-metrics-you-must-know-aa97784ff226

Flask Frame Work Reference:

https://www.youtube.com/embed/lj4I_CvBnt0

Flask Refarance To Run:

https://www.youtube.com/embed/UbCWoMf80PY

Train The Model On IBM:

Account Creation:

https://www.youtube.com/embed/4y_zD-0Q3F8

Train Model On IBM Watson:

https://www.youtube.com/embed/TysuP3KgSzc

Integrate Flask With Scoring Endpoint:

https://www.youtube.com/embed/ST1ZYLmYw2U