INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

ROBERT GORDON UNIVERSITY ABERDEEN

Multimodal Music Information Classification System

Group 18 Final Thesis Report by

Mr. Sathila Samarasinghe

Mr. Thareen Withanage

Mr. Ishan Fernando

Mr. Rikzy Jezuli

Supervised by

Ms. Sachinthani Perera

Submitted in partial fulfillment of the requirements for the BSc (Hons) in Artificial Intelligence and Data Science degree at Robert Gordon University.

April 2023

Declaration

We declare that this is our own research thesis, and this thesis does not incorporate without acknowledgement any material previously published submitted for a degree or Diploma in any other university or institute of higher learning and to the best of our knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Student Name	**Student ID**	**Signature**
Sathila Samarasinghe	20210515	
Thareen Withanage	20211009	
Ishan Fernando	20210549	
Rikzy Jezuli	20210667	

I have read the final thesis and it is in accordance with the approved university final thesis outline.

Signature of the supervisor: Date: 2023.04.08

ABSTRACT

The rapid growth of digital music libraries has underscored the need for efficient and accurate music information classification systems. Existing systems often struggle to keep up with the diverse and evolving nature of the music industry, making it difficult for users to identify the hidden features in a music track that makes them like or dislike it. In response to this challenge, we have developed a multi-model music information classification system named "WhatTrack" that focuses on three critical aspects: music genre classification, chord identification, and musical instrument identification. By leveraging advanced machine learning techniques and carefully selected datasets, this system aims to provide an easy-to-use user-friendly system that enables the identification of key features in a music track.

A music genre classification model using convolutional neural networks (CNN) has been implemented to capture intricate patterns across diverse genres. For chord identification, the K-nearest Neighbors (KNN) algorithm was used to accurately determine a piece's tonal center. Additionally, innovative instrument identification method was done by using machine learning techniques, utilizing KNN for model training, to recognize and differentiate instruments in complex arrangements. The models were trained on datasets from various sources, with thorough pre-processing and feature extraction processes.

The performance of the system was evaluated using standard metrics such as accuracy, F1 score, precision, and recall. 'WhatTrack' outperforms human accuracy and the accuracy state-of-the-art systems when it comes to genre, chord, and instrument identification in a music track. In conclusion, 'WhatTrack' offers a novel and effective solution for organizing and understanding music in digital libraries, with potential applications in music recommendation, music retrieval, and music education. By focusing on the critical aspects of music genre classification, chord identification, and instrument identification, our system provides a comprehensive and user-friendly approach that can adapt to the diverse and evolving landscape of music.

Keywords: Music, Music Recommendation, Music Education, Genre Classification, Chord Identification, Instrument Identification, Convolutional Neural Networks (CNN), K-nearest Neighbors (KNN), Machine Learning, Digital Music Libraries

ACKNOWLEDGEMENT

The successful completion of this project would not have been possible without the generous support and input of numerous individuals. We extend our deepest appreciation to our esteemed supervisor, Ms. Sachinthani Perera, for her guidance, invaluable tips, and unwavering presence throughout this project.

We are also grateful to Mr. Prasan Yapa and the mentoring panel for their valuable advice and guidance in the successful implementation of this project and this report.

Our sincere thanks go out to our lecturers for their continuous help, kindness, and understanding throughout the duration of the project and all the valuable knowledge and skill we have gained from them which helped us immensely in project implementation.

Finally, we would like to express our gratitude to our parents, friends, and everyone who has been involved both directly and indirectly, in supporting us during this project, including domain and technical experts and survey participants who helped us gain valuable insights about the music industry.

Table of Content

List of Tables

List of Figures

List of Abbreviations

Acronym	Description
SVM	Support Vector Machine
CNN	Convolutional Neural Network
CALM	Categorizing And Learning Module
CRNN	Convolutional Recurrent Neural Network
RNN	Recurrent Neural Network
ML	Machine Learning
KNN	K-Nearest Neighbor
Bi-LSTM	Bi-directional Long Short-Term Memory
DNN	Deep Neural Network
HMM	Hidden Markov Model
PLCA	Probabilistic Latent Component Analysis
AI	Artificial Intelligence
UI	User Interface
SSADM	Structured Systems Analysis and Design Method)
DSDM	Dynamic System Development Method
MMIC	Multi-Modal Information Classification System

CHAPTER 01: INTRODUCTION

1.1 Chapter Overview

“WhatTrack”, the proposed project is a Multimodal Music Information Classification System which consists of three key components of how to classify music. They are music genre classification, instrument type prediction and major-minor chord prediction together with music recommendation. Currently, there is no web or mobile application that has the above-mentioned components as a collective. The main objective of this project is to make the users increase their knowledge of music. ML techniques will be used to achieve the above-mentioned components to process real-time audio to get the required output for the user.

1.2 Problem Domain

In the world, Music has a huge impact. From younger generations to the most senior generations music is an important part of life. Most people listen to music to entertain themselves to get into a certain mood, 90% of the world’s population listens to music, according to (**Nielsen Music 360**). Out of the listening music percentage, a low percentage knows about music. (13% of people used to play an instrument and know music).

Music Genres

Music genres are the types of music that have specific distinguishable characteristics. They originated from different backgrounds and parts of the world. Basically pop, hip-hop, rock, jazz, metal, blues & disco are some of them. Modern music is originally based on these types of music. But it is hard to categorize modern music under genres though it has past features.

Musical Instrumental Sounds

Vibration, resonance, and amplification work together to create the sound that musical instruments make. The shape, size, material composition, and playing technique are only a few of the physical characteristics that each instrument possesses and which affect the sound it creates. It can be seen that the ability to recognize musical instrumental sounds (except guitar, violin, piano and drums) is low within the majority of music listeners.

Chords in music (Major/Minor)

In theory of music the key is the group of pitches (frequency), or scale that forms the basis of musical composition of a piece of music, while a chord is a combination of 3 or more notes (pitches) which sounds simultaneously harmonically. There are 2 basic types of chords called major and minor and each having 12 key signatures. Categorizing chords as major and minor manually is harder for those who lack music knowledge.

1.3 Problem Definition

There is no proper system in the world that includes all the features (Music genre classification, audio classification – predicting voice color and instruments used, key/chord prediction, and music recommendation). As mentioned in the problem domain, people listening to music is increasing daily. But people knowing music or understanding music don't grow as music listening or streaming. So, the main problem is to reduce this gap. To reduce this gap the system will collect music, that will be classified into different genres. Instrumental sounds will be identified, voice range will be predicted, key/chord will be generated, and music will be recommended. To achieve this an AI-based application is required. Since it includes machine learning techniques and user inputs of music, the accuracy of this system will be increased, and it will minimize this gap.

1.3.1 Problem Statement

Ordinary music listeners do not have the ability or the training to identify most of the hidden details in a music track such as the genre, the musical instruments used, music chords, and more which severely limits their ability to find similar music tracks they might enjoy.

1.4 Research Motivation

What is music? Music is a song made by different instruments. Music is a song made by voices with unique tone colors. Moreover, music is a set of sounds that harmonize with each other. Do people know these exactly? For this project, the authors' motivation is to educate the basic things of music for normal people with our music knowledge. Developing this app would be a great way to inject good music taste among people around the world.

1.5 Existing Work

Table 1.1 Existing Work

	Limitations	Technology/ Algorithm	Advantages
Existing work on Music Genre Classification Systems			
(Shah et al, 2022)	Limitations in computational power and datasets containing a wider range of genres compelled the research to focus on simpler models instead of complex models for music genre classification. Hence resulted in a relatively less accurate music genre classification system.	CNN	A custom CNN architecture is used for the music recommendation systems which was very efficient in music feature extraction.

(Falola et al, 2022)	An outdated and relatively small dataset is used to train the ML model which could classify only 10 genres. This can cause limitations when predicting genres of modern music which has many new genres that are	Support Vector Machine (SVM)/CALM classifier	It was discovered that using 2 or more classifiers during the genre classification process delivered significantly better results.
Existing work on Music Recommendation Systems			
(Mesghali et al, 2022)	The music recommendation system utilizes past your behavior to recommend music that the user may be interested in but the recommendation model because inaccurate when the user is new and has no past behavior records.	Bi-LSTM	The Bi-LSTM model used in the research required significantly lesser training time compared to the traditional CNN model.
(Lee et al, 2015)	The efficiency of the music recommendation system is measured based on user behaviors of just 5 users which is	MusicRecom (custom)	Research has a novel approach to music

	the very small sample size and the accuracy of the recommendation system measured using the data generated by just 5 users is not very reliable.		recommendation which utilizes both usage history and automatic genre classification to suggest highly personalized music to the user.
Existing work on Music Instrumental Sound Identification Systems			
(Yijie et al 2022)	The model requires further improvement in terms of detecting the instruments when multiple instruments are used in the same audio clip.	XGBoost	The accuracy of the algorithm in this paper on the public dataset medley-solos-DB reached 97. 65%.

(Zhang et al 2021)	The study focuses exclusively on traditional Chinese instruments and the models are not trained to detect sounds from modern musical instruments.	CNN/RNN/DNN	The model used in the research achieved a greater level of accuracy (82 – 99%)

1.6 Research Gap

Due to the Advancement of technology, every past domain has been modernized into different platforms. As it is the same for the music industry. The advancement of technology has increased the crowd's engagement with music. Because of this people tend to use smartphones to listen to music. In recent years the music domain is a well-researched topic. Mainly music genre classification and audio classification are mainly pitched areas. As some researchers have done their relevant work on chord prediction and instrument predictions.

Every researcher has implemented models in predicting the above-mentioned features. The accuracy of the models is different, but the approaches used to classify the above features are the same (Using the same AI Algorithms). No researcher hasn't implemented an app (mobile or web).

So, during this research, the main pitched area was, researchers have just classified music into a genre and recommended music and vice versa. But the research gap proposed a new system that includes every mentioned feature in one user-friendly mobile application (The proposed system is based on mobile because music is listened to mostly using smartphones). The newly added feature which does not exist in any research in the proposed system is to predict the tone color of the user's voice/recorded clip.

1.7 Contribution to the Body of Knowledge

1.7.1 Technological contribution

In this system, implementation is done by processing the spectral coefficients of the received music files. This means the images will be processed by AI algorithms of the system. The most feasible Algorithms are KNN (K-Nearest Neighbor) and CNN (Convolutional Neural Network). This is to classify music genres and predict the major or minor chord and the instrument type.

Music recommendation is recommended using the connected the Spotify API to user preferred genre. To get the most accuracy from the system the training models need to add or shuffle for the existing feature type.

1.7.2 Domain contribution

The project aims to contribute to the music domain in various ways by focusing on 20th and 21st-century music genres. The knowledge of music transformations will lead to an increase in the quality of music and its listeners. The app provides the feature of identifying chords (Major or Minor) for music enthusiasts and predicts the musical instrument of the input wav file are the unique features that could not be found in other existing apps.

1.8 Research Challenge

The team's area of expertise, which included music genre classification, music selection, and chord/key prediction, presented a few difficulties during the project. Looking for datasets separately for each domain and choosing the best one in order to collect the proper datasets took a huge effort.

The team experienced challenges when selecting a model, especially since they had no prior understanding of any model type. Because of this, it took them some time to choose a model and come to an agreement on it.

1.9 Research Questions

R1. How to classify audio clips according to their features?

R2. How to recommend good music tracks according to the features of classified music?

R3. How to increase the accuracy and the quality of our product to give the best user experience?

R4. Which models can be used for the classification process?

1.10 Research Aim

The research aims to classify music information and give any music listener the much-needed basic knowledge about music.

1.11 Research Objectives

Table 1.2: Research Objectives

Research Objectives	Explanation	Learning Outcome
Problem Identification	Identify an efficient and accurate machine learning algorithm to detect and classify audio information.	LO1
Literature Review	RO1 – Identify a suitable programming language and machine learning algorithm to detect audio information. RO2 – Identify datasets to train the Multimodal Music Information Classification System RO3 – To design the system to detect music tracks from plain audio. RO4 – To design the system to further detect the tone and genre and temper of the music track.	LO1

	RO5 – To design the system to suggest music recommendations to the user based on the data gathered in the previous steps. RO6 - Implement the Multimodal Audio Classification System using Mobile/Web Application using an intuitive UI.	
Data Gathering and Analysis	Utilize the datasets of music information with the help of online data libraries.	LO2, LO3
Research Design	Identify the best dataset to train the Multimodal Music Information Classification System from amongst existing publicly available datasets.	L04
Implementation	Implement the Multimodal Music Information Classification System using a mobile/web application.	L04
Testing and Evaluation	Identify the accuracy of the Multimodal Music Information Classification System using user feedback.	L04

1.12 Project Scope

1.12.1 In- Scope

Table 1:3: In-scope

No	Description
1	Identifying selected music genres.
2	Newly trending music will be recommended according to the time.
3	Identifying the Major/Minor Chords of the music.
4	Identifying the musical instruments used to create the music.

1.12.2 Out-scope

Table 1:4: Out-scope

No	Description
1	Classifying Sri Lankan Music into genres.
2	Every audio file cannot be tested.
3	Genres are limited only the most listened genres are tested.
4	Chords will not be generated.

1.12.3 Feature Prototype

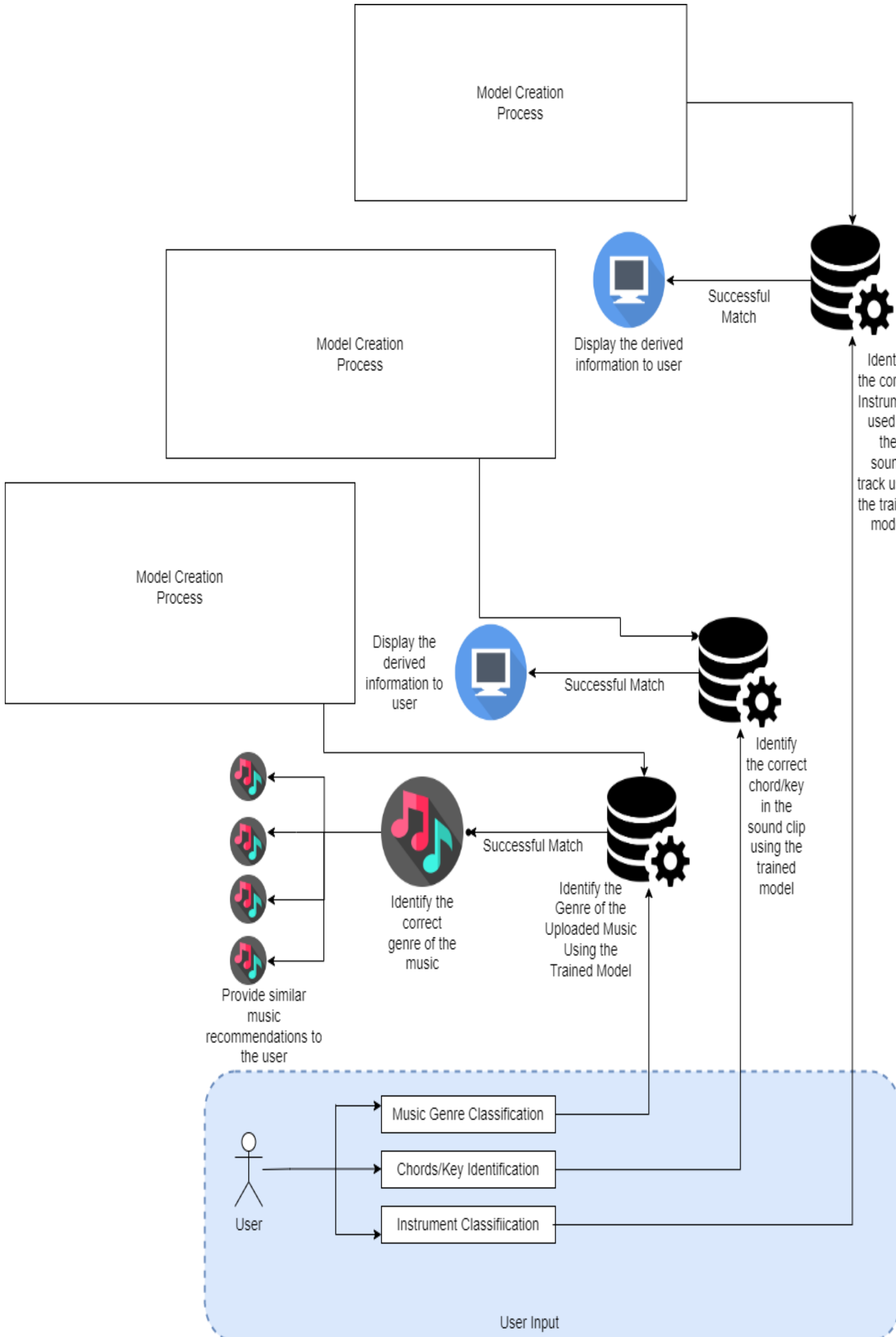


Figure 1.1- Feature Prototype

1.13 Resource Requirements

1.13.1 Hardware Requirements

- Central Processing Unit (CPU) — Intel Core i5 6th Generation processor or higher. An AMD (Advanced Micro Devices) equivalent processor will also be optimal. To train and run the model.
- RAM — 8 GB minimum (7.78 GB usable), 16 GB or higher is recommended. To image processing, create the spectrograms and store these images.
- Graphics Processing Unit (GPU) — NVIDIA GeForce GTX 960 or higher.

1.13.2 Software Requirements

- Python - The main programming language to build is python and it also supports many libraries.
    - TensorFlow – Time series analysis, speech, and image recognition, and training of the model.
    - Numpy – Working with arrays.
    - Scipy - Multidimensional image processing.
    - Pandas - Easy to deal with missing data.
    - Matplotlib - Low memory consumption and better runtime behavior.
    - Magenta, Librosa, and pyAudioAnalysis – working with audio data.
- MS Word – For creating documentation.
- Google Drive – To store huge-size data sets and other valuable documents.
- Windows operating system – To handle huge computational functionalities.
- Git - Version controller and easy to work with a team.
- Flutter - To create UI in the mobile-based application.
- Flask – For the backend

- Anaconda Navigator - Combining all functionalities into one place and using python IDE as Jupiter notebook.
- Draw.io/star UML – To design wireframes for the proposed system.
- Jira – Project management tool.

1.13.3 Skills Requirements

Technical skills – Programming, knowledge of Analytical tools, processing large data sets and machine learning and deep learning, and statistical analysis.

Non-technical skills – Strong business acumen, and communication skills.

1.13.4 Data Requirements

This project includes 3 mini-projects with a machine-learning approach. Therefore 4 datasets are required. GTZAN was the most used dataset for music genre classification and will be used here. It includes 1000 audio files with a duration of 30 seconds each. IRMAS: a dataset for instrument recognition in musical audio signals will be another dataset. Music Chords (Major/Minor) dataset with chord sounds of 859 audio clips will be used to identify key chords of the music.

1.14 Chapter Summary

The music industry has undergone many changes in the past decade with the introduction of music streaming platforms. Despite all the advances in music streaming technologies, the public seem to have little to no knowledge of different genres, musical instruments, vocal tones and pitch used in a music track hence limiting their ability to pinpoint what they like or dislike in a given music track. Thus, there is a tremendous interest among music listeners for a centralized, intuitive audio and music classification system. Recent advances in machine learning algorithms have made it feasible to create such a system that is affordable and accessible.

CHAPTER 02: LITERATURE REVIEW

2.1 Chapter Overview

This chapter provides an overview of the current state of the art in the field of audio classification. This chapter will explore various research studies, academic articles, and other relevant sources of information related to audio classification systems. The review will cover the latest advancements in audio classification techniques, including machine learning algorithms, and their application in multi-modal classification systems.

2.2 Problem Domain

In the modern era, people are finding a way to relax by listening to music on their smart devices.

Commonly their approach would be downloading music tracks and categorizing them according to their choice into playlists or refusing that if they don't like it. Here the project mainly aims to help people to classify the music. Moreover, classification of audio track based on musical instruments, classify major/minor chords, classification of the music according to the genre, and recommend more music for the user's preference. Hence, a user could input the list of audio files and get the output including the features of the tracks and genre of music.

2.3 Existing Work

2.3.1 Audio Classification

In the past few years, several approaches have been taken for audio classification using machine learning and deep learning. An audio recognition system for Chinese traditional instruments was made, with Mel spectrum characteristics as input. (Rongfeng Li et al., 2022) has trained an 8-layer convolutional neural network.

Instead of reducing model size using complex methods, (Xubo Liu et al., 2022) has improved a model to eliminate the temporal redundancy in the input audio features (Mel spectrogram) and proposing a family of simple pooling front ends (SimPFs) to reduce redundant information within Mel-spectrogram. Another approach was by (Jianyuan Sun et al., 2022) using deep neural decision forest (DNDF) to classify an audio clip based on the characteristics of the recording environment (ASC). DNDF combines a fixed number of convolutional layers and a decision forest as the final classifier.

There's an approach considering the number of non-target events (Wu Dan, 2019) and problems such as detection strategy, detection time, the decision tree models. (Liang Gao et al., 2022) did 2 multiple representations as inputs to train the networks jointly with a knowledge distillation method.

2.3.2 Music Genre Classification

There have been Several research done to categorize music. Up until now, many techniques have been investigated for music genre classification. A multi-frame approach (Tejas Dalvi et al.,2022) with an average stage to analyze in detail almost the full song. It is used at training time to generate more samples and at test, time to achieve an overview of the whole song. The model to evaluate the performance of the multi-frame approach has been trained with the train partition of hand-made dataset and evaluated using the test partition.

A Convolutional Neural Network (CNN), a deep learning technique was created by (Peace Busola et al., 2022) with a total accuracy of 77.89%. The deep learning approach made by him could play a vital role in classification, audio features such as spectrogram features which are extracted from the signal of the music are one of the best features that gave excellent results.

Another research was conducted by (Hongjuan Zhang et al.,2022) where they tried identifying a novel classification framework incorporating the auditory image feature with traditional acoustic features and spectral features which proposed to improve the classification accuracy. Moreover, the logarithmic frequency spectrogram (Dhevan Lau et al.,2020) rather than linear is adopted to extract the spectral feature to capture the information about the low-frequency part adequately. Other machine learning and deep learning models can still be worked upon for accurate music genre classification. One of the most common datasets used for music genre classification is the GTZAN Database (Hansi Yang et al., 2019), (Lata Gohil et al., 2022) and (Yeshwant Singh et al.,2022)

.

2.4 Technology/Approach/Algorithm Review

There are various technologies, approaches, and algorithms used in existing audio classification systems. These systems typically use machine learning algorithms such as support vector machines (SVM), decision trees, and neural networks to classify audio files. Features extraction is a crucial step in audio classification, and different methods like Mel-frequency cepstral coefficients (MFCC), linear predictive coding (LPC), and wavelet transforms are used to extract relevant features from audio signals. Recent advances in deep learning, such as convolutional neural networks (CNN) and recurrent neural networks (RNN), have also shown significant promise in improving the accuracy of audio classification systems. In addition, multi-modal approaches, which integrate data from multiple sources such as audio and video, have been shown to improve the performance of audio classification systems. Overall, a combination of these technologies, approaches, and algorithms can be used to create an effective audio classification system.

2.5 Chapter Summary

There are some implementation-related issues with earlier studies focused on music genre categorization, audio classification, and music recommendation. When attempting to categorize music genre, the models employed to extract features from audio had several limitations (could not interpret every nuance of scatter spectrograms). There were issues with how to categorize audio features from image spectrograms when training the models. However, the number of participants utilized to assess the accuracy of the models used in these studies was tiny, making it difficult to trust the accuracy of the models used to read, categorize, and recommend music.

Also, the lengthy machine learning models used in these studies for categorizing and recommending music can lead to poor user experience. Multimodal Music Information Classification System is the focus of our project. The audio files that have been used as data are wav files with a maximum duration of 30 seconds and midi files. The audio is categorized and filtered by search algorithms. If the audio files are solely music tracks, then machine learning techniques will be used to classify them into the appropriate genres after audio detection, and later music has been suggested as per feature. The research gap mostly focuses on combining the above three components into a single mobile application.

CHAPTER 03: METHODOLOGY

3.1 Chapter Overview

This chapter outlines the methods used to conduct the research. It provides a detailed description of different methodologies followed to conduct the research, development, design, evaluation, and project management. Furthermore, an expected timeline for the project implementation can be found.

3.2 Research Methodology

Table 3.1: Research Methodology

Research Philosophy	Positivism is chosen as the research philosophy for this project as it involves collection of accurate data, identifying objective patterns in the data using statistics, and classifying them accordingly.
Research Approach	Quantitative approach will be used as it involves gathering larger samples of data and usage of statistical methods to analyze the data.
Research	Experiment Research strategy will be followed in this project as it

Strategy	involves manipulating one or more variables to see how they affect the outcome.
Research Choice	A Multi method research will be chosen as it can help overcome limitations and provide a more comprehensive understanding of the research question.
Time zone	The research uses a pre-defined data set that is collected over a large time.

3.3 Development Methodology

What is the life cycle model and why?

As agile project life cycle model allows to make rapid changes based on outcome and feedback of the client/supervisor. The Agile model allows greater flexibility and improves collaboration among members and greater quality of the final product while enabling the team to stick to implementation timeline.

Design methodology => SSADM or OOAD or Anything else?

Object-Oriented Analysis and Design (OOAD) is chosen over the Structured Systems Analysis and Design Method (SSADM) due to the complex nature of the project and the need for flexibility during the design and implementation process. The OOAD approach enables a more iterative and incremental approach, which allows the development team to make changes and adjustments to the system as it is being built. The OOAD approach also allows individual members to work on various components of the project and integrate them to the main system.

Evaluation methodology => Evaluation metrics and/or benchmarking.

The F1 score is a relevant evaluation metric for this project because the system is anticipated to categorize music tracks into various genres, recognize keys and chords, and identify the musical instruments utilized in each audio track. The F1 score is a single number that represents the system's performance in terms of recall and precision. It can be used to compare the effectiveness of various models.

3.4 Project Management Methodology

3.4.1 Schedule using Gantt Chart after doing WBS.

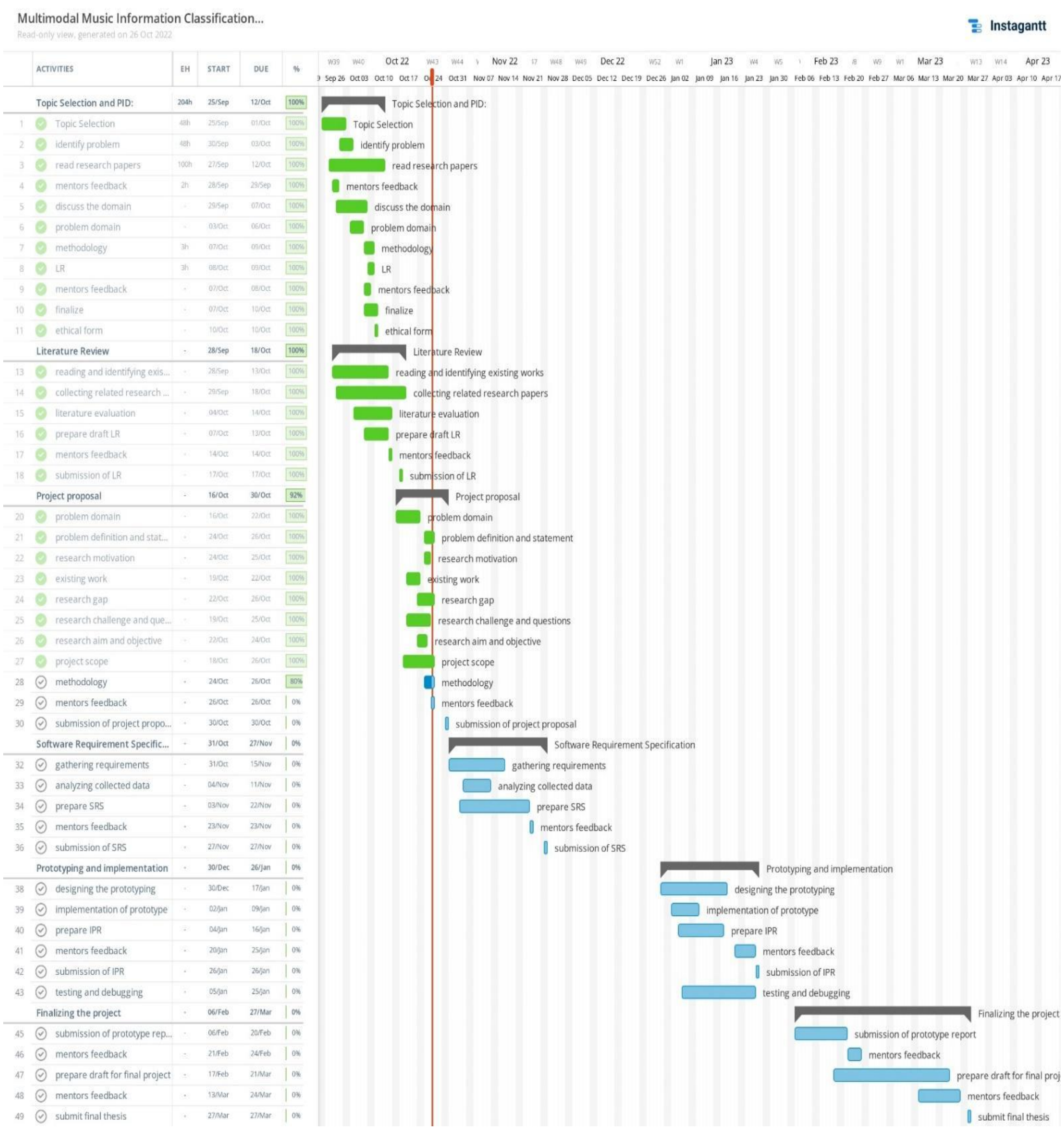


Figure 3.1: Gantt Chart

3.4.2 Deliverables

Table 3.2: Deliverables

Deliverable / Milestone	**Due Date**
Datasets and Components of the project	2nd week
Project Proposal	5th week
SRS	9th week
Learning the tech stack	9th week
Building the components of the project	17th week
Testing	19th week
Integration Process	19th week
Develop CI/CD Pipelines and Integrate to each component	21st week
Final Testing Phase of the applications	23rd week
Evaluation Phase	24th week
Post-Mortem and Research Paper	24th week

3.5 Chapter Summary

Development, design, evaluation, and project management methodologies discussed in this chapter ensures the quality, accuracy, and reliability of the multi modal music classification system. The evaluation metrics selected to measure the accuracy of the system help to determine the success of the implementation.

CHAPTER 04: SOFTWARE REQUIREMENTS SPECIFICATION

4.1 Chapter Overview

The major goal of this document is to identify the important stakeholders. to collect the needs and assess the requirements to determine the information that should be given priority. A detailed picture of the project's internal and external settings as well as an overview of its stakeholders is first drawn. To evaluate the various project techniques, a requirement elicitation was performed. In the last section, functional and non-functional requirements are described along with a use case diagram and its use case description.

4.2 Rich Picture

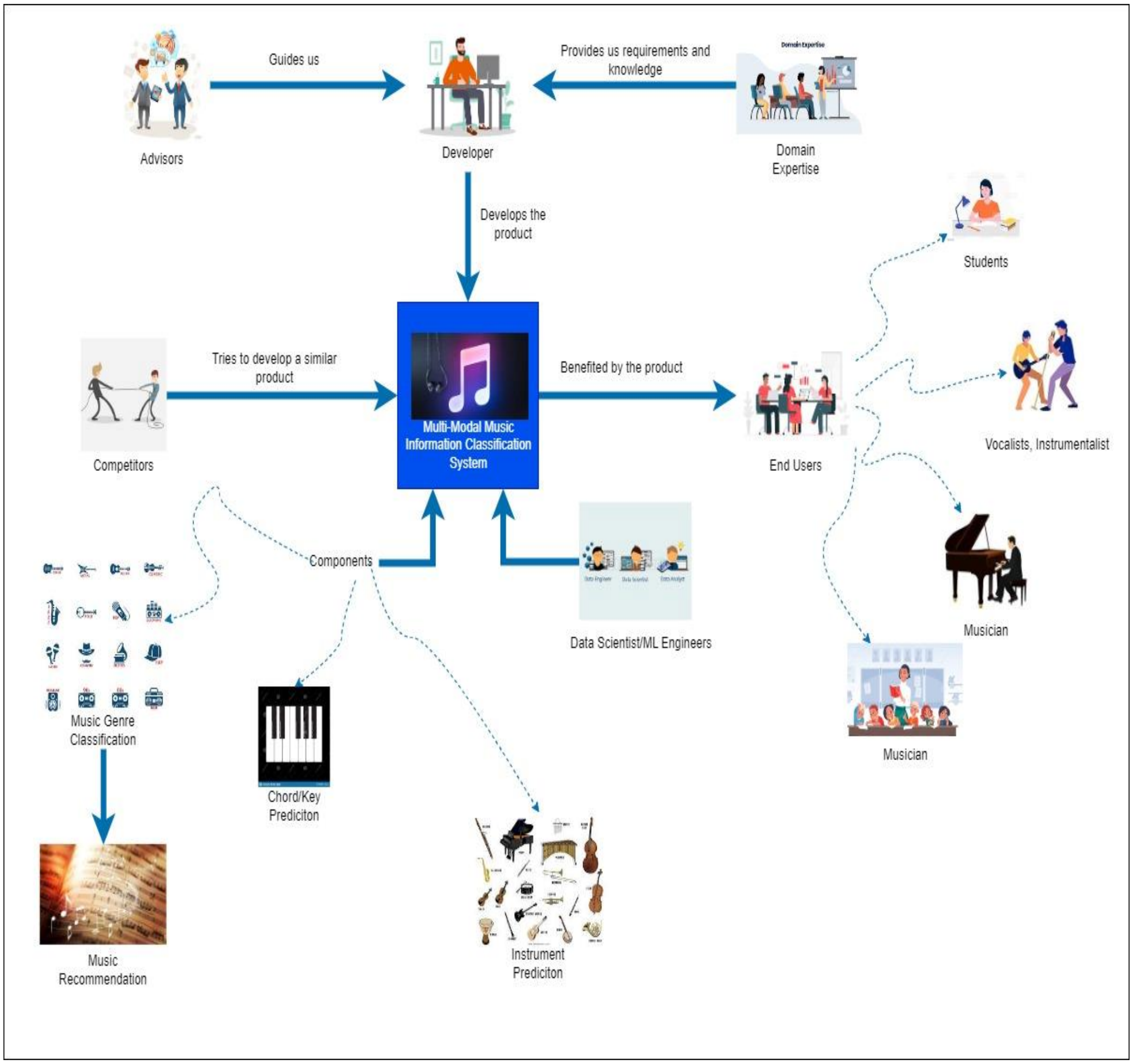


Figure 4.1: Rich Picture

The above rich picture illustrates the different stakeholders engaged with the Multimodal Music Information Classification (MMIC) System for different purposes. Advisors, developers, domain experts, competitors, end users, and data scientists/machine learning engineers are the identified stakeholders.

4.3 Stakeholder Analysis

In the below, the associated stakeholders are shown by the onion model diagram. The role and viewpoint of each stakeholder are identified.

4.2.1 Onion Model

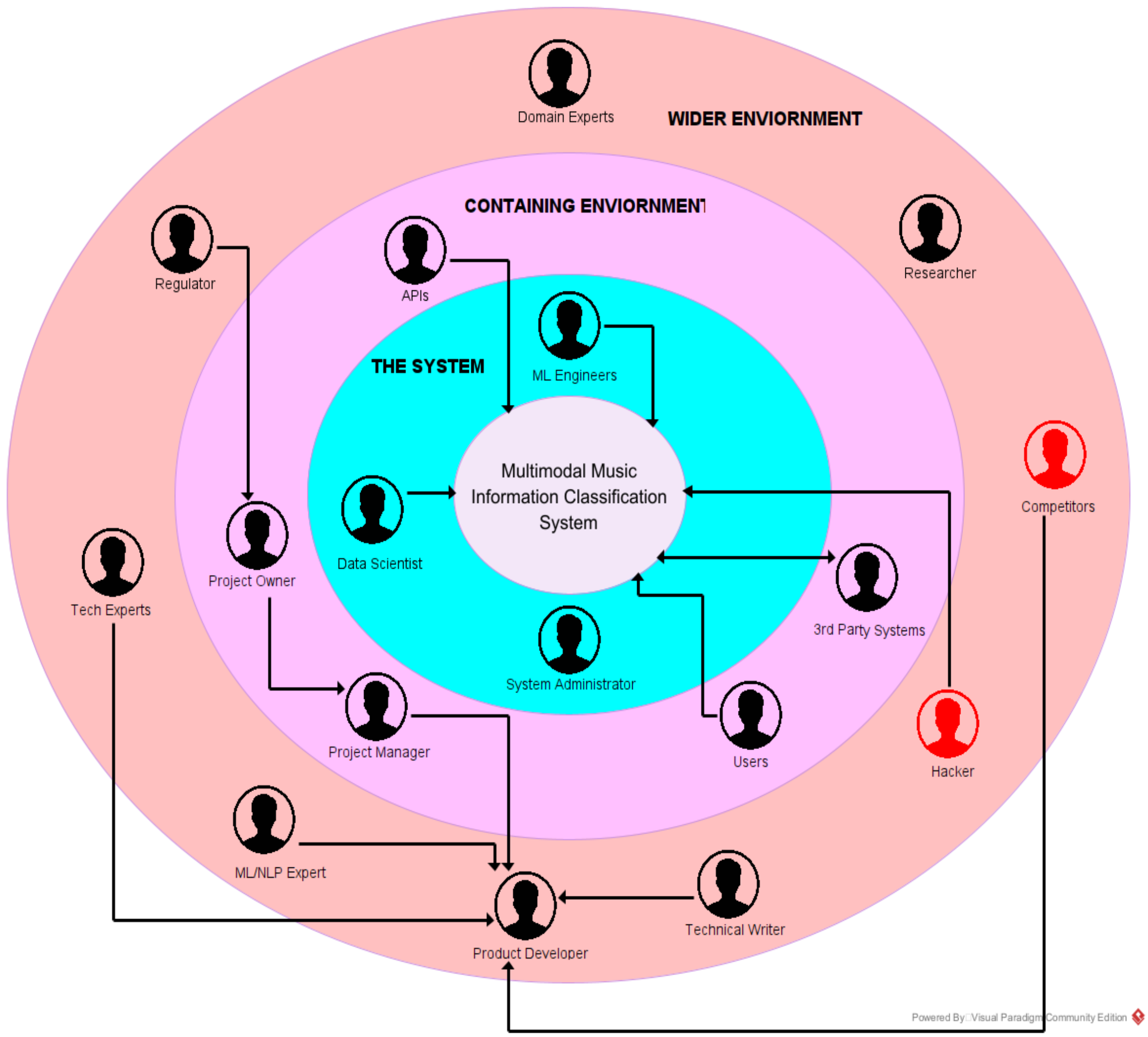


Figure 4.2: Onion Modal

4.3.2 Stakeholders Viewpoints

Table 4.1: Stakeholder's Viewpoints

Stakeholder	Role	Benefits
Data scientists, ML Engineers	Operational Maintenance	Design and development of the MMIC System process and models.
System Admin	Operational Administration	Deploying and configuring the application for various environments.

Product Owner	Functional Beneficiary	Owner of the MMIC system.
Product Manager	Managerial Support	Managing the application and processes to ensure that the project works properly.
Users, 3rd Party Systems, APIs	Functional Beneficiary	Using the developed MMIC application through various channels (Android) or integrating it with other systems.
Regulator	Quality Regulator	Monitor the application to make sure that data usage and processing fall within the established privacy policies.
Tech Experts	Expert	Determines if a collection of specifications is supported by the platform.
Domain Experts	Expert	Provides a domain view of the project's technology and methodologies.
Product Developer	Developer, Operational Maintenance	Implements and maintains the system.
Technical Writer	Operational Support	Provides support on the creation of the documentation for the system.
Researcher	Educational Beneficiary	Analyses the current systems and approaches to increase the efficiency of the current process and techniques.
Competitor	Negative Stakeholder	Creates an application that has similar features to the proposed system.

Hacker	Negative Stakeholder	Poses a threat to the proposed system due to the possibility of unauthorized access to the application and its data.

4.4 Selection of Requirement Elicitation Techniques

The various methods of gathering requirements are referred to as requirement elicitation. Requirement Elicitation Methods are employed with the assistance of clients, users, and other stakeholders to ascertain the system's requirements. The best strategies taken into consideration are described in the sections that follow, along with some of their benefits and drawbacks.

4.4.1 Observing Existing Systems and Literature Review

Reviewing current systems that have similar features and approaches may be the initial step for requirement elicitation. The current work in the domain is classified and analyzed to identify features that could be improved.

Table 4.2: Pros and Cons of existing systems and literature review

Advantages	**Disadvantages**
·The main components for MMIC system, approaches for implementation can be identified. ·Useful in identifying the feature gaps which can be further improved.	·Even though identifying the features of the available commercial products are not complex, reviewing the existing research and developed systems are complex, since the objective of each research is different.

4.4.2 Surveys & Questionnaires

The MMIC system can be used by potentially anyone who is interested in music, or regular music listener who has access to internet, a computer or even a smartphone. Since the MMIC system has a very broad target audience, using surveys and questionnaires is ideal as a wide audience can be covered and large amounts of reliable data can be gathered.

Table 4.3: Pros and cons of survey and questionaries

Advantages	**Disadvantages**
• Ability to cover a very large, diversified audience. • Less time-consuming and can be completed with ease by anyone. • Easy to classify and analyze survey data.	• Multiple choice questions can be tricky as it is quite hard to know exactly what a person feels about a question. • The question can be misunderstood, or the answers may not be honest.

4.4.3 Interviews

Interviews can provide an insight into current gaps and opportunities for improvement in the domain of the MMIC system. Experts in the music industry, regular music listeners and students who study music could be interviewed for this purpose.

Table 4.4: Pros and cons of interviews

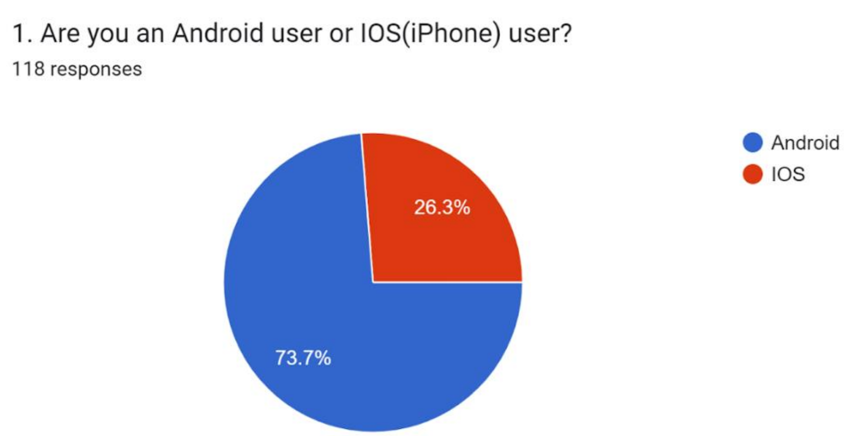
Advantages	**Disadvantages**

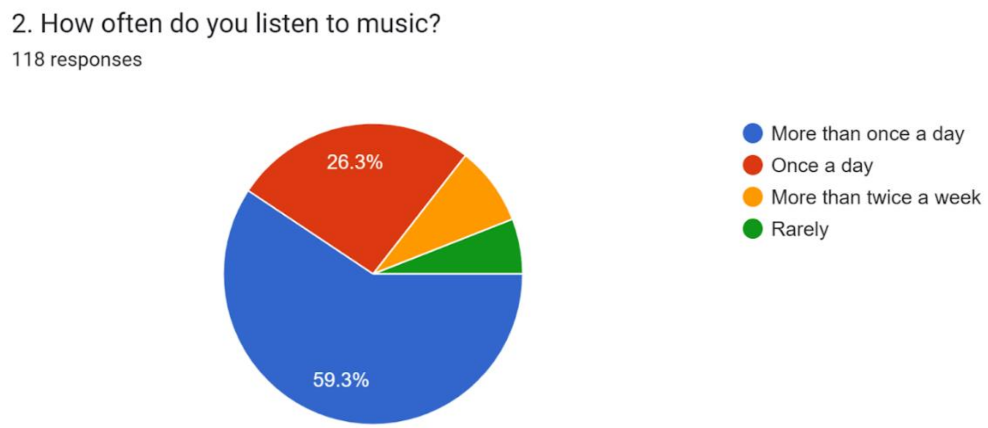
• Direct face to face interactions can provide a unique point of view of the problem. • The face-to-face interview reduce the chances of misinterpretations compared to surveys and questionnaires.	• Covering a wide diversified audience is challenging and time consuming. • Everyone may not have to the time to allocate for the interviews and may not have enough knowledge to understand the tech used in the MMIC system.

4.5 Discussion of Results

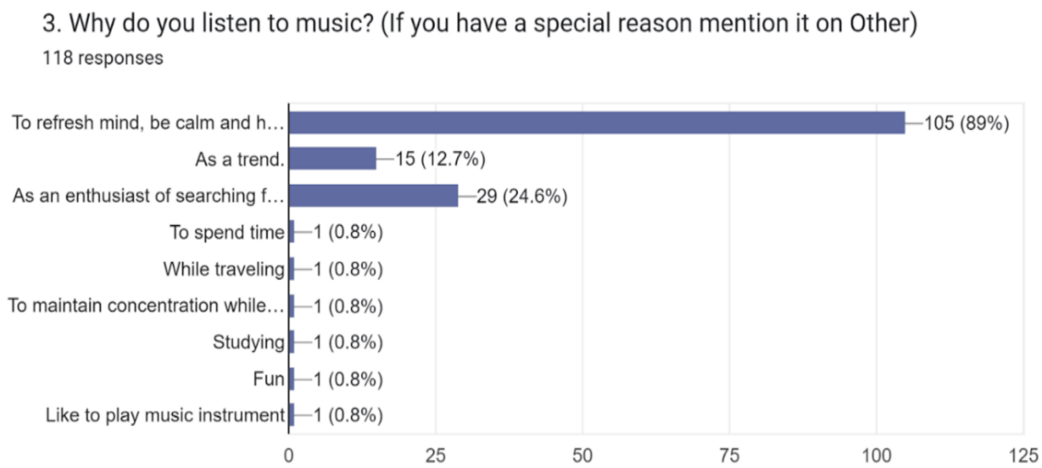
Table 4.5: Survey Findings

Question	Are you an Android user or IOS (iPhone) user?
Aim of the Question	To see the number of users who use IOs and Android operating systems.
Observations	



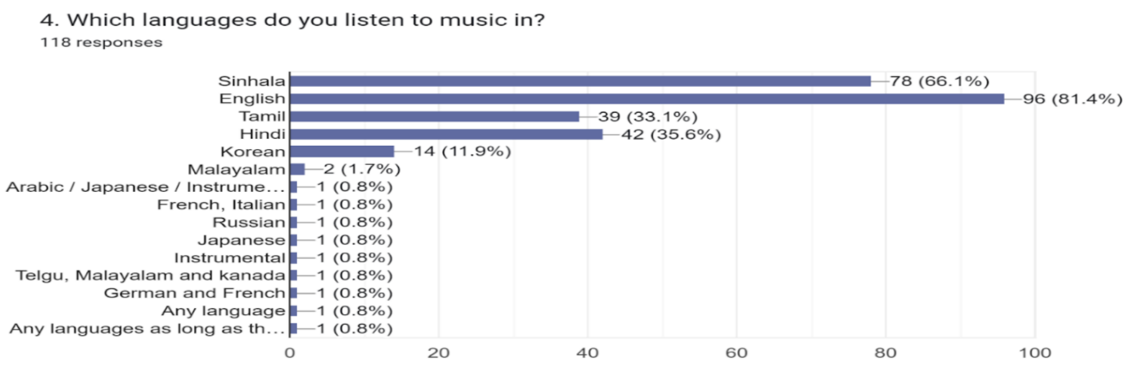
The observation is analyzed from 73.7% of Android users and 26.3% of IOs users.	
Conclusions	Most of the participants are Android users. So, the product is an Android OS-based mobile application.
Question	How often do you listen to music?
Aim of the Question	To see how users listen to music.
Observation	
 It is observed that 59.3% of participants listen to music more than once a day, 59.3 % once a day and 15.4% listen to music twice a week and rarely.	
Conclusion	It is observed that most of the participants listen to music frequently and a least number of participants neglect music.
Question	Why do you listen to music?
Aim of the Question	To analyze why the users, listen to music.

Observation



It is observed that 89% of the participants listen to music to freshen and calm themselves to release their stress of work. 24.6% listen to music for their passion for learning music and 12.7% listen to music as a trend in this society.

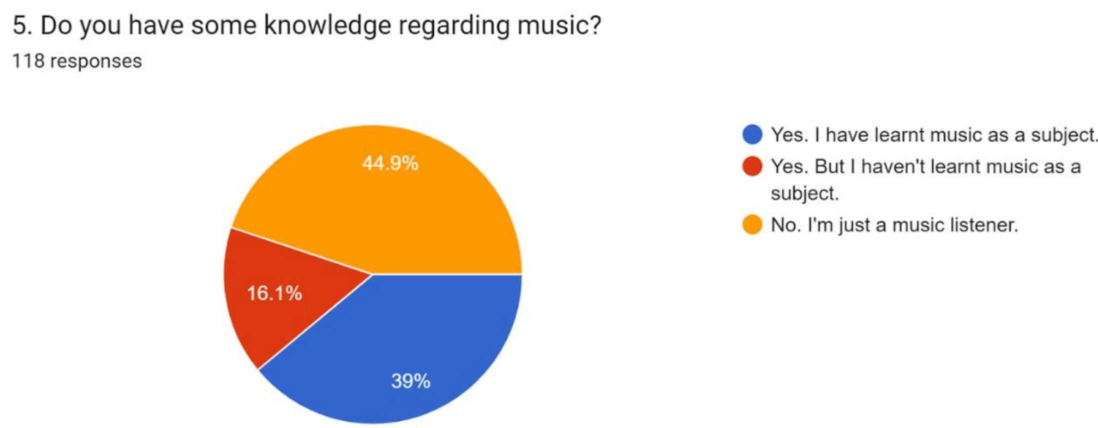
Conclusion	Today, due to the high workload and survival of life people are stressed. So, the only remedy to get rid of it is to listen to music and most people have a lesser knowledge of music.
Question	Which language do you listen to music in?
Aim of the Question	To identify the taste in music they listen to.
Observation	



The top 3 languages that users listen to are English, Sinhala and Hindi. Out of them English has the highest percentage of 81.4%.

Conclusion	English is the most listened to type of music. So that the product is an English-only classifying system.
Question	Do you have some knowledge regarding music?
Aim of the Question	To see the participants' knowledge of music.

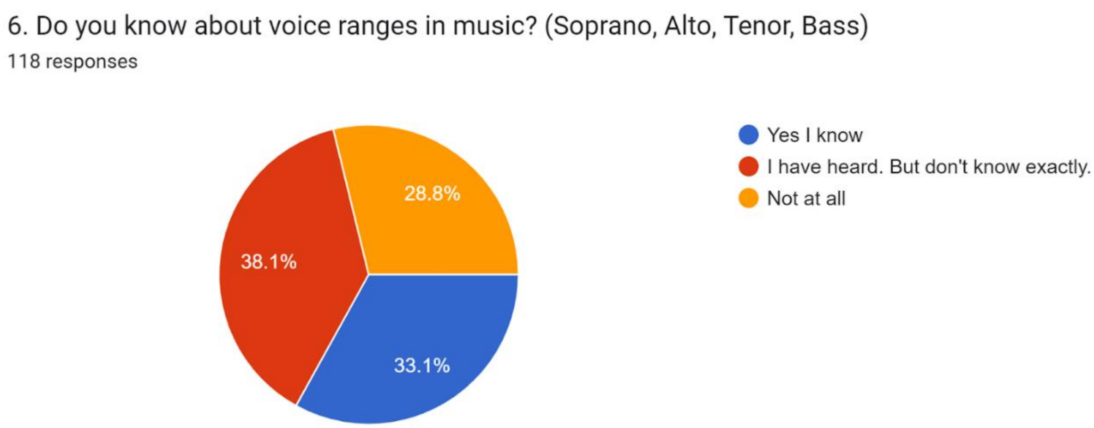
Observation

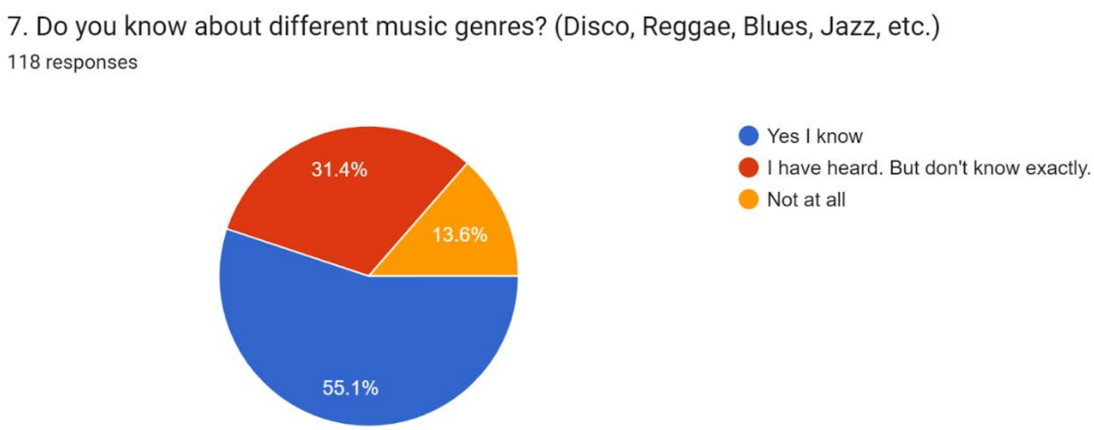


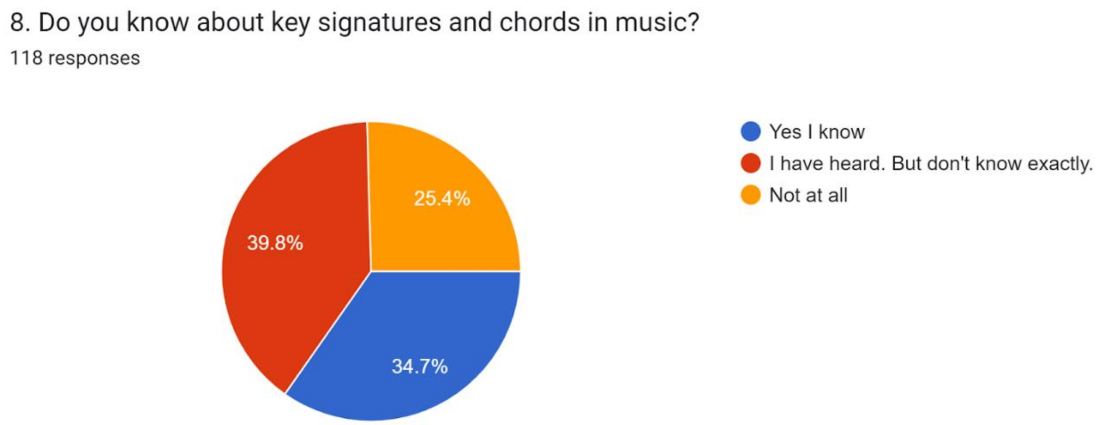
44.9% of participants are just music listeners, 39% have learnt music as a subject and 16.1% haven’t learnt music as a subject.

Conclusion	This product will be implemented to reduce the gap of music listeners to convert them to music knowledgeable listeners.
Question	Do you know about voice ranges in music?Do you know about different ranges of music? Do you know about key signatures?
Aim of the Question	In this question the main objective is to identify the different knowledge about music.

Observation







Out of the three questions most of the survey participants have heard about music genres (31%), voice ranges (38%) and key/chord (40%).

Conclusion	The participants have a minimum knowledge about the above questioned categories.

4.5.1 Interviews

Interviews were conducted with domain experts to discuss the scope of this project. Mainly we met and interviewed parties who engage in the music industry. We mainly discussed the gap in knowledge of music and how this product can be utilized to decrease this gap. Mainly discussed topics are,

Table 4.6: Discussion of Interview Results

Question 1	Present society's knowledge and passion about music?
Conclusion	People in this society have less of an impression and a passion to learn music compared to the last decades.
Question 2	What reason for this drop in passion?

Conclusion	Due to the increase in living costs people are struggling to work on their aesthetic taste and the increase in technology. The younger generation tends to play video games without learning an art subject.
Question 3	Will the proposed system help this gap to be reduced?
Conclusion	As this system is a technological-based simple mobile application it is a well-adaptable system. So, there is a positive aspect that this product will reduce this gap.

4.6 Summary of Findings

Table 4.7: Summary of findings

Findings	**Literature Review**	**Questionnaire**	**Existing Systems**	**Interviews**
A system with all mentioned components is not there.	**X**	**X**		**X**
Most people have preferred a mobile application with Android as its OS.		**X**		
Using ML and DL techniques is necessary for implementing the components	**X**			
Knowledge of music is considerably less		**X**		**X**
A music classification system will help to increase the knowledge of music				**X**

No chord key prediction systems which use audio files to predict.			X	
Apps that classify music into genres do not use the input to recommend music.	X		X	
Music Recommendation Systems recommend music according to user preference only.	X		X	

4.7 Context Diagram

Figure 4.3: Context Diagram

4.8 Use Case Diagram

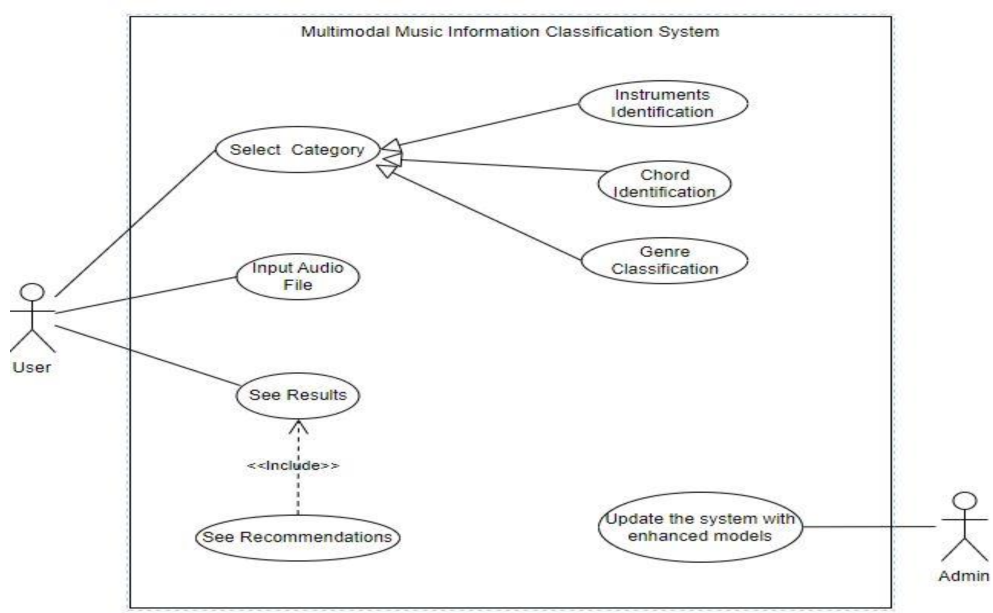


Figure 4.4: Use case Diagram

4.8.1 Use Case Description

Table 4.8: Use case description

Use Case ID	UC001
Use Case Name	Select Category
Actor	User
Purpose	Select a category from 3 categories in the main interface.
Overview	The user needs to select a category. The user selects a category from the main interface. The user directs to the selected category interface.
Pre-Conditions	The user should open the app.
Post Conditions	The user should be in the selected category interface.
The Typical Course of Events.	**Actor action:** The user selects a category from the main interface. **System response:** The system directs the user to the selected category interface.
Alternative Courses	-

Use Case ID	UC002
Use Case Name	Input Audio File
Actor	User
Purpose	Input the audio clip to get the relevant output.
Overview	The user needs to classify music genre/voice type/instruments or identify key chords on a music track. The user selects the music track from the device storage. The user inputs the audio file into the system.
Pre-Conditions	The user should be in the selected category interface.
Post Conditions	The user should successfully input the audio file into the system.

The Typical Course of Events.	Actor action	System response
	The user gives permits to access device files. The user selects a music track from the device storage.	Access device files. Copy the selected audio clip into the system.
Alternative Courses	Unable to copy larger files or incorrect file formats into the system.	

Use Case ID	UC003	
Use Case Name	View Results	
Actor	User	
Purpose	View the results.	
Overview	The user needs to view the result. The user asks to show the result. The user views the result.	
Pre-Conditions	The user should have input the audio file successfully.	
Post Conditions	The user should see the result.	
The Typical Course of Events.	Actor action	System response
	The user clicks on 'view results	The system runs the audio track and shows the result.
Alternative Courses	Unidentified audio clip.	

4.9 Functional Requirements

The functional requirements of the system are listed in the table below, along with their priority level.

Table 4.9: Functional Requirements

Priority Level	Description

Critical	Main features and functionalities must be mandatorily included in the system.
Important	A proposed feature or functionality that will further add value to the system but is not mandatory.
Desirable	An out-of-scope requirement

	Requirement and Description	**Priority**
FR01	Accepting the audio track	Critical
	The system should be able to accept the audio track the user wants to analyze.	

FR02	Save the results of the analysis to further improve the models	Critical
	Despite not having any use for the end user, metadata generated by each operation should be saved to further improve the ML models.	
FR03	Detect the genre of the audio track	Critical
	The trained ML model for detecting music genre is used to detect the genre of the music track	
FR04	GUI and other Interface support	Desirable
	A user-friendly intuitive UI and API interfaces for the 3rd party apps to connect with the system.	
FR05	Detect the instruments used in the music track	Critical
	The trained ML model for detecting instruments is used to detect the music instruments used in the track.	

FR06	Identify chords used in the music track *The trained ML model for detecting chords is used to detect the music instruments used in the track.*	Critical
FR07	Recommendation of similar music *Based on the genre detected by the model, the system should be able to recommend similar music to the user.*	Important

4.10 Non- Functional Requirements

Performance constraints – Reliability, security, response time, performance

Interface constraints - Usability

Economic constraints – Marketing analysis, resources, development cost

Lifecycle requirements – Quality of the design, solution evaluation, portability.

Table 4.10: Non-functional requirements

	Requirement and Description	**Specification**	**Priority**
NFR01	The accuracy of the system in detecting various features in an audio track should be high	Accuracy	Important
NFR02	The system should be secured to avoid unauthorized access and data breaches.	Security	Important
NFR03	Model training should be efficient and not take a long time	Performance	Important
NFR04	The process should be done with minimum resource requirements to support as many devices as possible.	Performance	Non-important
NFR05	The system will have an intuitive GUI	Usability	Desirable

4.11 Chapter Summary

An onion model and a thorough system analysis were presented in the chapter's opening paragraphs. A questionnaire result and specified requirement elicitation methods were also provided. A context diagram and a use case diagram were also used for the demonstration. Moreover, requirements that were both functional and non-functional were listed with priorities.

CHAPTER 05: SOCIAL, LEGAL, ETHICAL AND PROFESSIONAL ISSUES

5.1 Chapter Overview

This chapter gives a comprehensive assessment of all the social, legal, ethical, and professional challenges that might affect the final result. These external influences can be examined and handled using the SLEP analysis. As a result, SLEP analysis is utilized to pinpoint potential problems that could occur; methods taken to address these problems are then covered in more detail in this chapter.

5.2 SLEP Issues and Mitigation

5.2.1 Social Issues

- Diversity and Inclusion: The system is designed to accommodate a wide range of musical styles and cultural perspectives eliminating any bias towards certain musical styles.
- Bias: The models are trained to detect a diverse range of musical genres and instruments eliminating the potential bias towards certain genres or artists.
- Language: English is used as the primary language of the product due to its universality, but this does not hinder the ability to classify or analyze audio tracks that may belong to languages other than English.
- Cultural appropriation: The system is trained in a diverse dataset using automated

classification techniques to overcome misclassification and to ensure that music is classified accurately and respectfully.

5.2.2 Legal Issues

- Copyright infringement: The system is designed with a complete database of music metadata that includes information about copyright ownership and licensing. The system is built in such a way that illegal download or streaming of music is prevented.
- Data protection: User data is not sold or distributed to 3rd parties. The system is transparent about how the user data is collected, stored, and used.
- Dataset: All the datasets used in the training of the models are available in the public domain and are obtained with relevant permissions.

5.2.3 Ethical Issues

- Responsible data use: The system is built with responsible data use practices, with a focus on data privacy, security, and user control.
- Anonymity: The information gathered during surveys and interviews are stored anonymously to protect the identity of the participants.
- Bias and Discrimination: The system uses a diverse range of data in training the models to prevent misclassification and bias towards certain music types.
- Datasets: Datasets used in the training of models are chosen after ensuring that the data is gathered by ethical means and does not infringe upon the copyright of any 3rd party.

5.2.4 Professional Issues

- Performance and reliability: The system is designed by adhering to wide accepted design principles making updating and integration seamless provide a positive user experience.
- Scalability and maintenance: Given the huge target audience the system is designed

with scalability and ease of maintenance in mind. A complete documentation and testing process is in place to ensure that new features and updates can be integrated seamlessly.

- Collaboration and communication: The system is designed and updated by the team using version control such as git and GitHub. In addition to that task flow is managed using Trello and Email and WhatsApp is used primarily for communication.

5.3 Chapter Summary

This chapter provides an overview of the possible social, legal, ethical, and professional issues and how they might affect the system and steps taken for mitigation of these issues that may arise.

CHAPTER 06: SYSTEM ARCHITECTURE & DESIGN

6.1 Chapter Overview

The main objective of this chapter is to provide the system and architecture design of the product. In this chapter system architecture design, design goals, system design displayed using design paradigm, component diagram, class diagram, sequence diagram and UI design.

6.2 Design Goals

Table 6.1: Design Goals

Design Goal	**Description**
Mobile	The mobile app is the only platform where the user can interact with the product. Mainly an android application as per survey results.

Performance	Aim for higher accuracy and higher F1 score when training the AI modals. Try to build a higher accuracy-based modal.
Security	User's data should be protected and only be accessed by the current user only.
Scalability	The mobile application should be able to easily cooperate more functionalities/ features and support when increase of multiple users. The application must have the ability to balance the load of user's logged at the same time.
UI/UX	Create a user-friendly interface for the mobile application where user can have real time user experience while using the application

6.3 System Design Architecture

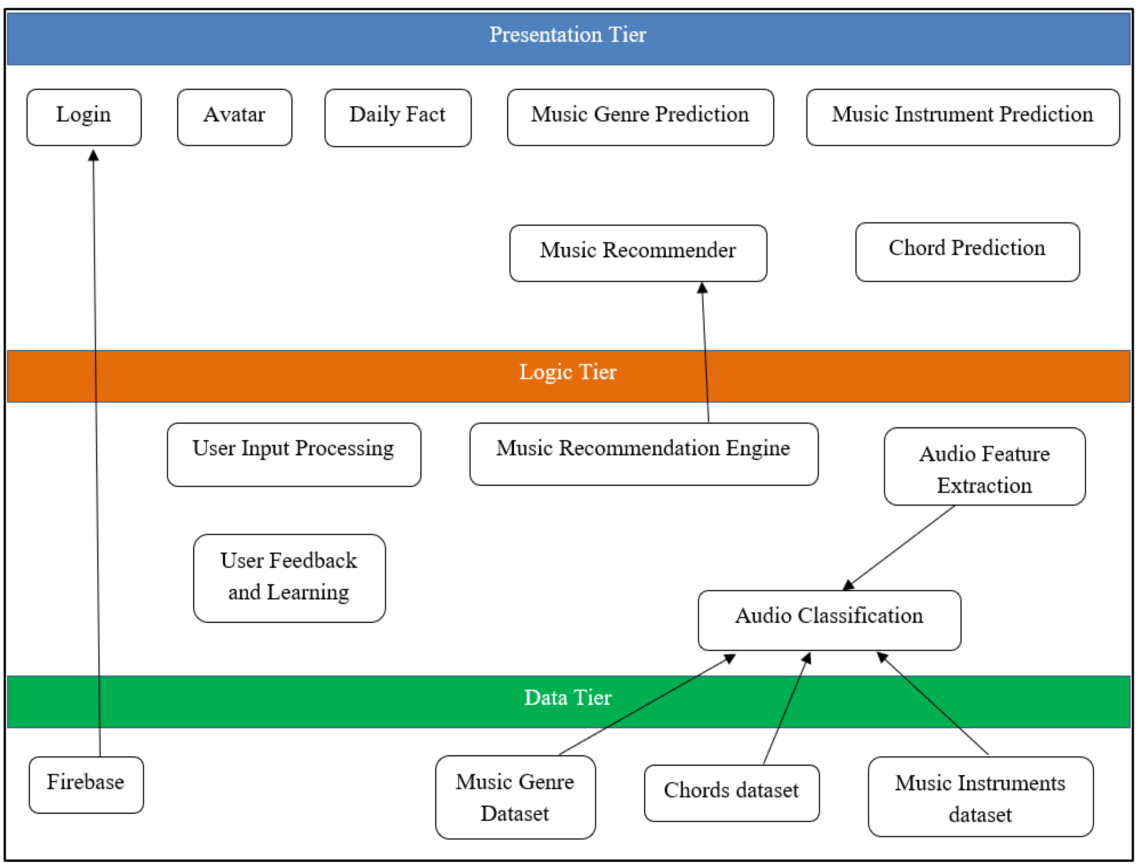


Figure 6.1:System Design Architecture

The above figure shows the system architecture design with three layers, presentation tier, logic tier and data tier.

1. Presentation Tier
    a. Login – This gives the user to access the app.
    b. Avatar – Avatar is generated randomly according to the gender of the user..
    c. Daily fact – This is to enhance user's music knowledge by providing daily facts in music.
    d. Music genre prediction – This UI is to input audio files of unknown music genre.
    e. Music instrument prediction – This UI is to input audio file of unknown music instrumental sound.
    f. Chords prediction – This UI is to input audio file of unknown chord.
    g. Music recommender – This UI is to display music recommendation by accessing. Spotify music playlist.

2. Logic Tier

Logic tier combines algorithms and techniques from audio signal processing, machine learning and deep learning concepts.

a. Audio feature extraction - Features are mainly extracted as MFCCs and stored into json files.

b. Audio classification - ML/DL algorithms are used and tested to classify audio with the best accuracies. Training models by the annotated music samples are used.

c. User input processing - This involves converting user input to a standardized format and performing preprocessing steps.

d. Music recommendation engine: Based on the music genre prediction a separate music recommendation system is provided by accessing Spotify music.

e. User feedback and learning: This is to collect user feedback on recommended music and accuracy of the predicted results. This will help to improve the performance of the application by techniques such as reinforcement learning or active learning.

3. Data Tier

   a. Firebase – All the user details will be stored in the firebase to access when needed.

   b. Music genre dataset – Dataset used to train the music genre classification system.

   c. Music instruments dataset – Dataset used to train the music instrumental sound classification system.

   d. Chords dataset - Dataset used to train the chords identification system.

6.4 System Design

System design refers to the methodology that is used to achieve the specified requirements for architecture, components, modules, interfaces, and data for the system. In other words, system design provides a roadmap for creating the software system that meets the requirements of its users and other stakeholders.

SSADM (Structured Systems Analysis and Design Method) and OOAD (Object-Oriented Analysis and Design) are two different system design methodologies used in software.

6.4.1 Choice of Design Paradigm

OOAD system design methodology is chosen for this system largely because it enables consistency and reusability. As this system includes diverse features and components, OOAD provides an ideal framework for this system be built on improving its reliability and resilience. In addition to that SSADM is based on Waterfall methodology whereas this system uses the Agile methodology, and hence this is another reason why OOAD is chosen over SSADM.

6.4.2 Component Diagram

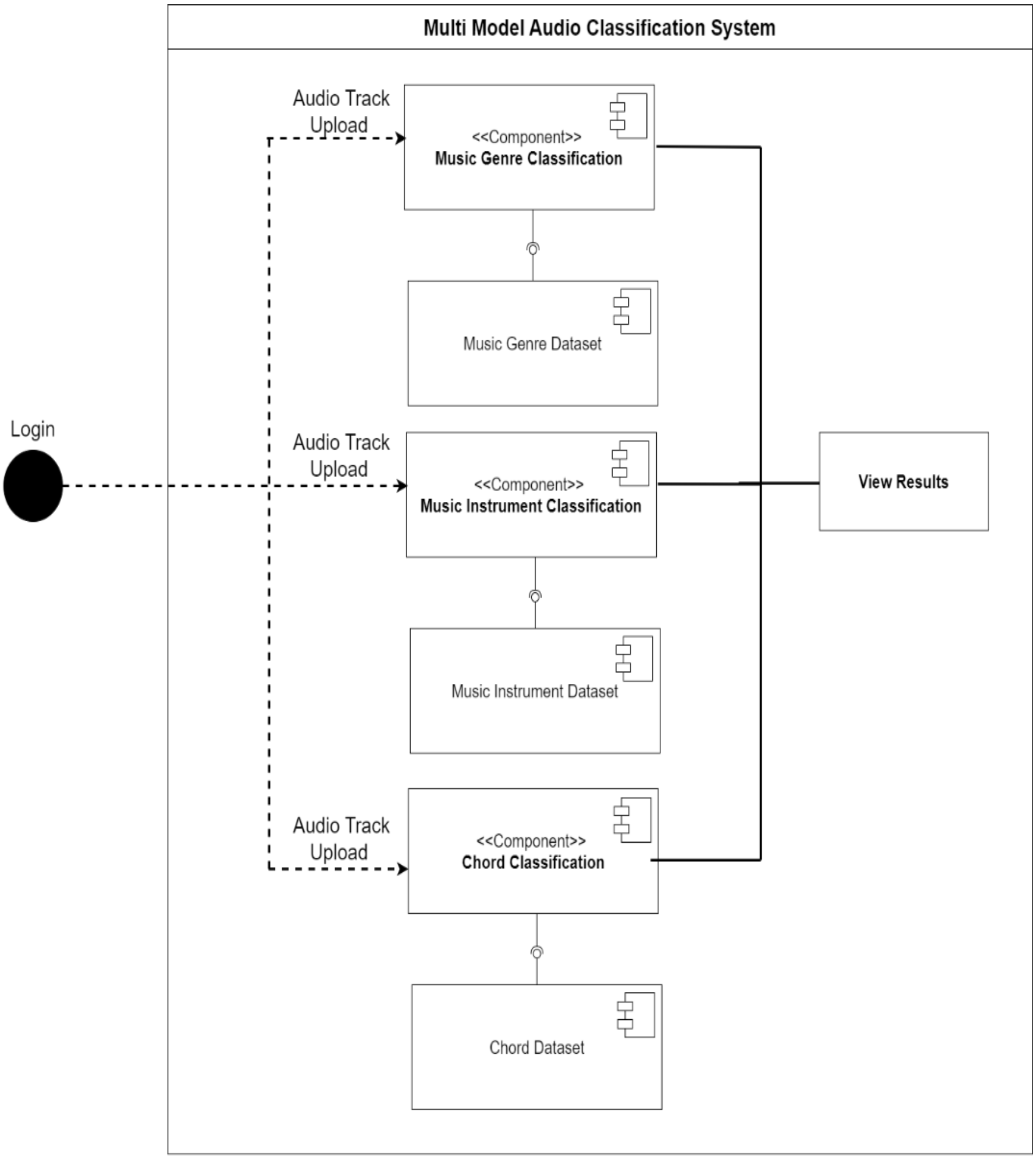


Figure 6.2:Component Diagram

6.4.3 Class Diagram

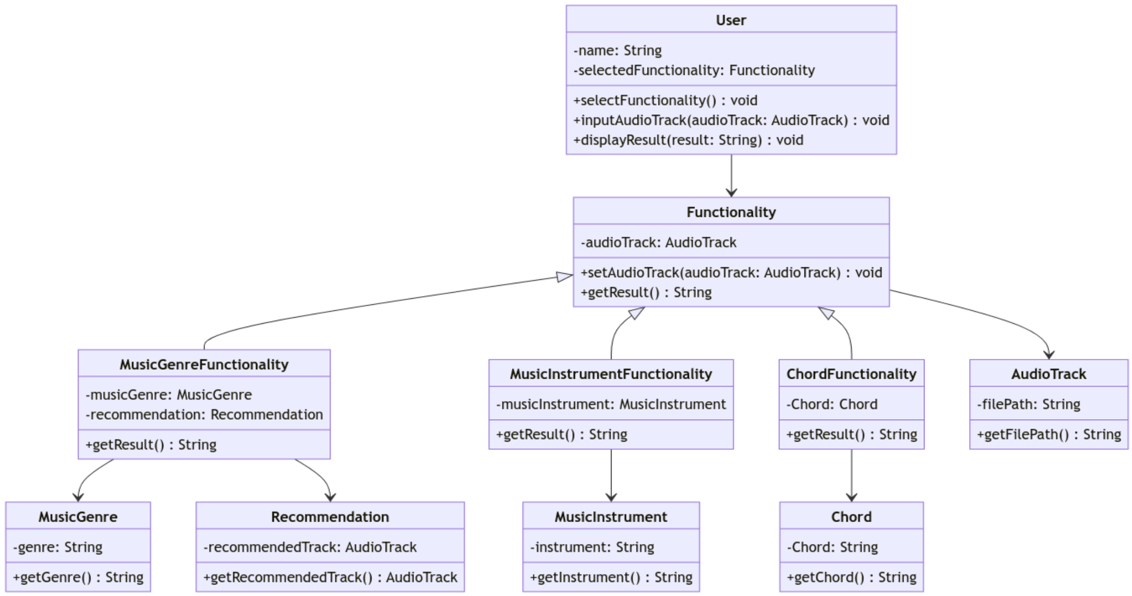


Figure 6.3:Class Diagram

6.4.4 Sequence Diagram

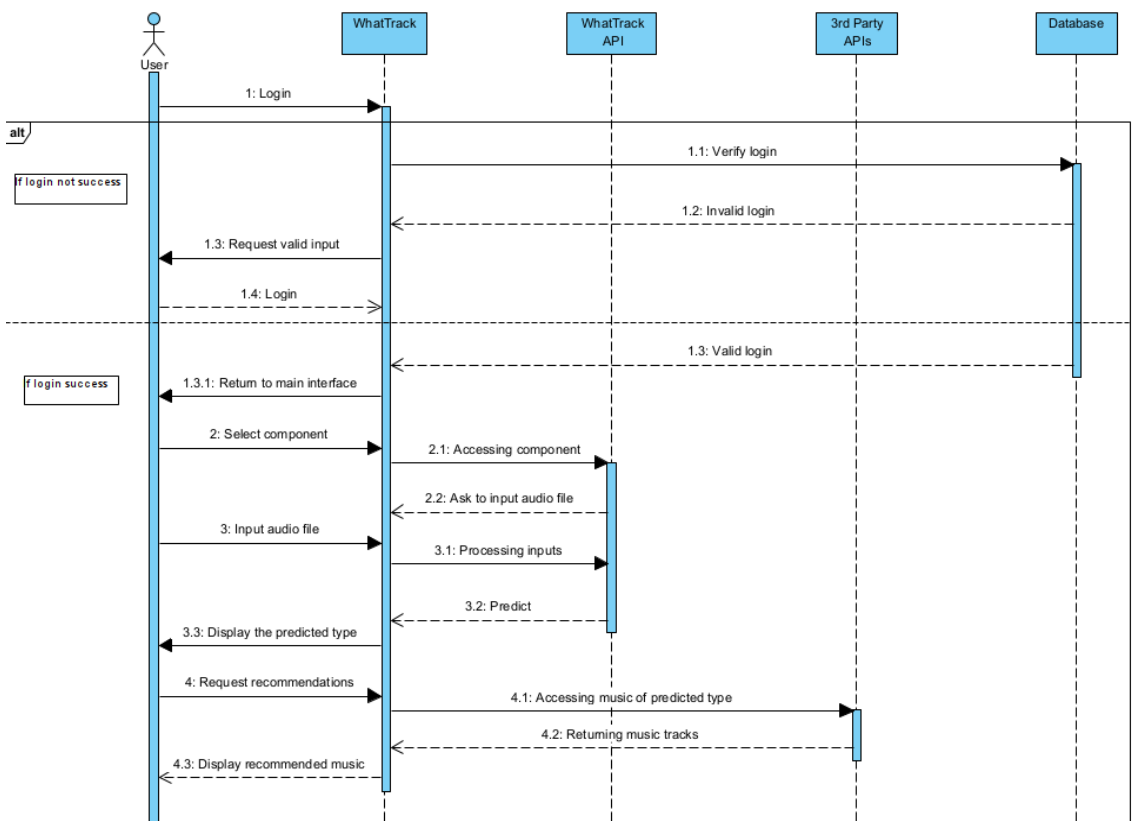


Figure 6.4:Sequence Diagram

6.4.5 UI Design

Below the UI design of the mobile application “WhatTrack” is presented. User’s input and output is displayed. This UI design will cover the whole application.

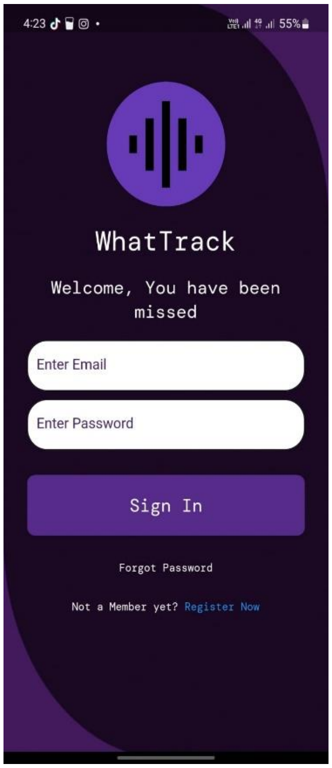


Figure 6.5:Login Page

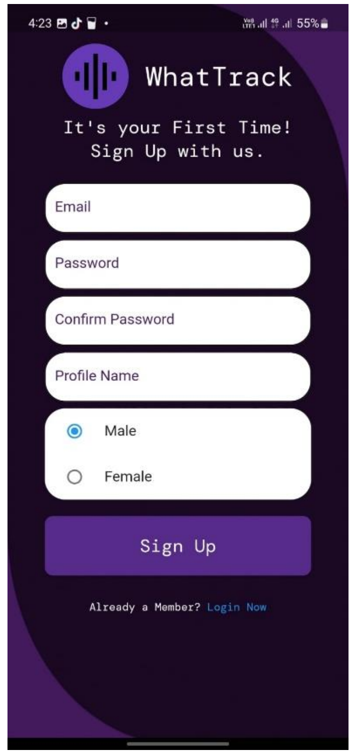


Figure 6.6: Signup Page

Figure 6.7: Home

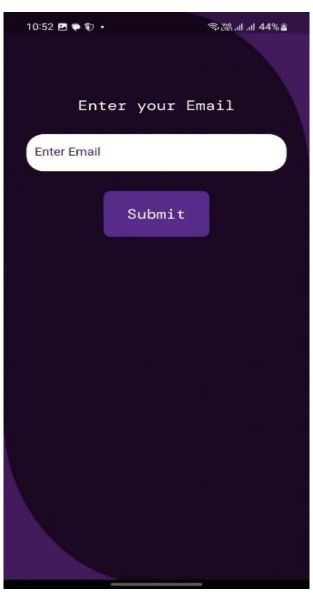


Figure 6.8: Forget password.

Figure 6.9: About us

Figure 6.10: Major/Minor

Figure 6.11: Profile

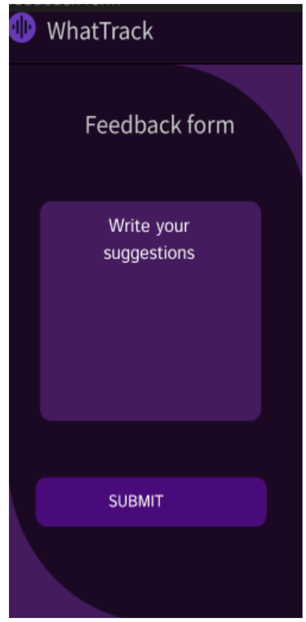


Figure 6.12: Feedback Form

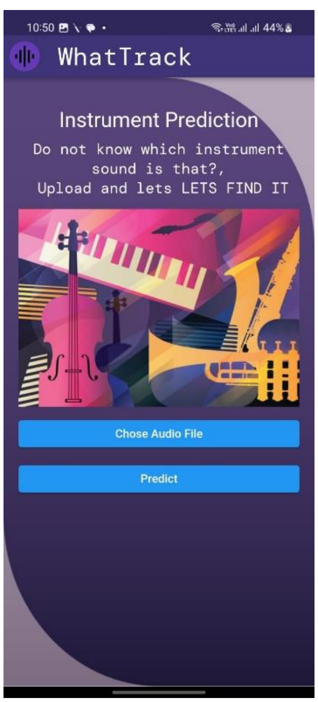


Figure 6.13: Instrument Prediction

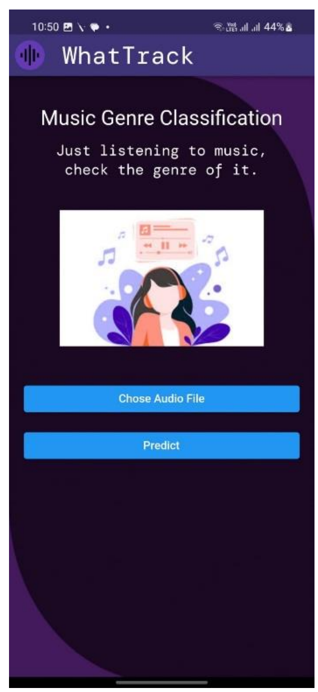


Figure 6.14: Music Genre

6.4.6 Process Flowchart

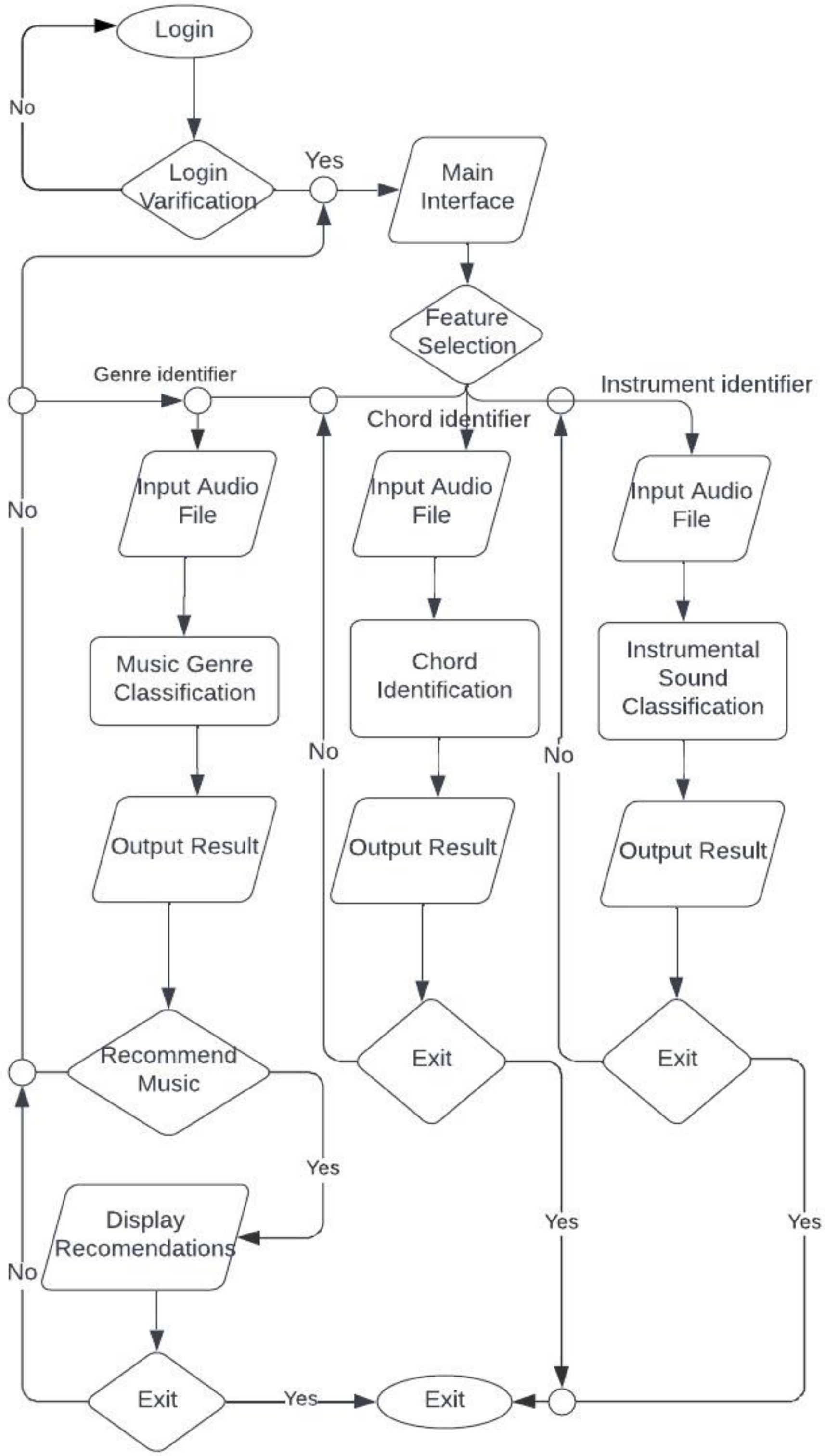


Figure 6.15: Process Flowchart

6.5 Chapter Summary

This chapter provided a brief overview of the design and architectural elements described above. Component diagrams, class diagrams, sequence diagrams, and UI/UX designs in the previously supplied design provide a clear explanation of the desired design goals for the system.

CHAPTER 07: IMPLEMENTATION

7.1 Chapter Overview

In this chapter mainly focuses on the technologies used to implement the mobile application “WhatTrack”. Here analyzation of technology stack, methodologies used for data selection, selection of framework, programming language, libraries and IDEA used are described.

7.2 Technological Selection

7.2.1 Technology Stack

The system uses a wide array of technologies to achieve specific requirements. TensorFlow and Scikit learn is used to build the Machine Learning models. Flutter is primarily used to build mobile applications and Flask is used to build the backend of the mobile application.

7.2.2 Data Selection

After carefully analyzing the literature and other similar projects that achieved significant accuracies and the datasets that were used, we decided to select the following datasets to build the machine learning models. The table below gives an overview of the datasets used for training the Music Genre, Chord, and Musical Instrument Identification models.

Table 7.1 Data Selection

Domain	Dataset	Description
Music Genre Identification	GTZAN Dataset - Music Genre Classification	The dataset consists of 1000 audio tracks each 30 seconds long. It contains 10 genres, each represented by 100 tracks. The tracks are all 22050Hz Mono 16-bit audio files in .wav format.
Chord Identification	Shords Dataset - Major VS Minor guitar chords	This dataset consists of 500+ audio files in wav format. Each audio represents 1 chord played in a major or minor key.
Musical Instrument Identification	IRAMAS Dataset	Dataset consists of sounds of 6 different musical instruments.

7.2.3 Selection of Development Framework

7.2.3.1 Frontend

To build the mobile application Flutter and DART was used. Due to flutter is cross-platform we could write for both Ios and Android. Learning flutter took less time due to that it was easier. Flutter consists of many widgets which ease the work when developing.

7.2.3.2 Backend

Mainly for the backend firebase authentication was used. Flask was used to write the APIs to connect the AI modals.

7.2.4 Programming Language

The primary programming language utilized to create our mobile application WhatTrack is DART. It is mostly utilized because it is an easier-to-learn cross-platform programming language. DART is a Google product, thus there is little need for error management, which increases the effectiveness of designing the app.

Python served as the primary programming language for the creation of machine learning modals. While training our modals, TensorFlow and Keras are the two frameworks that are most frequently employed. When building modals, using these frameworks made the
process much easier and took much less time.

7.2.5 Libraries

Table 7.2 Libraries

Mobile Application	
Flutter Version	1.0.0+1
GoogleFonts	-
cupertino_icons:	^1.0.2
firebase_auth	^4.3.0
firebase_core	^2.8.0
cloud_firestore	^4.4.5
provider	^6.0.5
get	^4.6.5
shared_preferences	^2.0.20
loading_indicator	^3.1.0
email_validator	^2.1.17
firebase_storage	^11.0.16
file_picker	^5.2.6
http	^0.13.5
Machine Learning Modal	
TensorFlow	2.11.0
Numpy	1.21.4

Librosa	0.9.2
Scikit learn	0.24.2
Matplotlib	3.6.3
Seaborn	0.12.2

7.2.6 IDE

Android Studio – Electric Eel and VS code is the mainly used IDEAs to build the mobile app. To build the machine learning models PyCharm, Jupyter Notebook via anaconda was frequently used.

7.2.7 Summary of Technology Selection

Table 7.3 Summary of Technology Selection

Component	Tool/ Technology	Version
Programming Language	PyCharm	2022.3.2
	Flutter	3.7.1
	Dart	2.19.1
UI	Figma	93.4.0
IDE	Android Studio – Electric EEl	3.2-7.4
	Visual Studio Code	1.77.0
	Jupyter Notebook	6.1.4
	PyCharm	2022.3.2

7.3 Implementation of Core Functionalities

7.3.1 Music Instrument Classification

The main aim of this component is to predict the music instruments used to create the audio file.

Start
Import all necessary libraries
Load audio files and calculate features
While load audio files
For:
extracts features using the get_features() function
Scale features using Standard Scaler, mean of 0 and standard deviation of 1
End for
k-Nearest Neighbors (kNN) classifier is initialized with k=1
Fit kNN model: The kNN model is fit to the training set.
Predict using the Test Set
Evaluate performance
Find wrong predicted audio files
End

7.3.2 Music Genre Classification

Main aim of this modal is to identify the music genres of an audio file. Mainly 10 genres are classified. They are blues, classical, music, disco, hip-hop, jazz, metal, pop, reggae, and rock.

Start
Import the necessary libraries
Define the path to the JSON file containing the dataset
Define a function load_data that loads the dataset from the JSON file and returns two NumPy arrays: X, containing the input data
DO:
Define a function load_data that loads the dataset from the JSON file and returns two NumPy arrays: X, containing the input data. Define a function build_model that builds

Keras Sequential model
For:
Define a function predict that takes a trained Keras model, an input data sample X,
and its corresponding target label y, predicts the label for the sample using the model, and prints the expected and predicted labels.
End For
End Do
In the __main__ block, call the prepare_datasets function with the desired test and validation set sizes, and build the Keras model using the build_model function and the input shape of the training data
Compile the Keras model with the categorical cross-entropy loss function, the Adam optimizer with a learning rate of 0.0001
Train the model on the training set using a batch size of 32, a maximum of 100 epochs, and the validation data for validation
Evaluate the trained model
Call the predict function with a sample input and its target label to test the trained model.
End

7.3.3 Major Minor Chord Classification

In this component music chords are classified into major or minor.

Start Load audio files and calculate features using librosa library. Normalize the audio signals. Scale the features using Standard Scaler. While load audio files For: extracts features using the get_features() function

```
    Scale features using Standard Scaler, mean of 0 and standard deviation of 1
    End for
Create a KNN Classifier with the desired number of neighbors classifier is initialized with k=1
Fit KNN model: The KNN model is fit to the training set.
Predict using the Test Set
Evaluate performance.
Find the misclassified samples and their corresponding audio files.

End
```

7.4 Chapter Summary

The tech stack that was used to develop WhatTrack was described in this chapter. A high-level grasp of how the system functioned was provided by the pseudocode implementation of the key functionalities.

CHAPTER 08: TESTING

8.1 Chapter Overview

This chapter focuses on verifying that the system fulfills all functional requirements as expected. All the models of this multi modal music information system is tested individually to ensure optimal functionality. Agile software design methodology requires that testing is completed immediately after implementation of functional requirements. This chapter explores the objectives and goals of the testing process and the results obtained from the tests.

8.2 Objectives and Goals of Testing

One of the important reasons for conducting tests is to ensure that the functional requirements of the system are functioning at an optimal level. The following tasks are performed inorder to ensure proper testing:

- Ensure that the code follows the best quality standards and coding practices.
- Detect bugs that may have been overlooked during the initial stages of implementation.
- Make sure that all the functional requirements perform at an ideal level.
- Check whether the non-functional requirements are satisfied.

8.3 Testing Criteria

When it comes to testing, the quality of the final code is of utmost importance. Testing for quality of the code can be done by black box testing, white box testing, active tests, inactive test etc. The quality of the code can be identified by:

- Maintainability – ensures that the code is readable, can be easily updated and modified.
- Performance – concerns with how efficiently the system utilizes its resources and how it influences response times, scalability, and customer satisfaction.
- Security – Ensures that all the necessary security measures are in place in order to prevent unauthorized access and protect customer data.
- Reliability – this refers to the ability of the system to work under heavy traffic without any failures and bugs.

8.4 Module Evaluation

8.4.1 Confusion Matrix

Accuracy:

$$\boldsymbol{Accuracy = \frac{True\ Positives + True\ Negatives}{True\ Positives + False\ Positives + True\ Negatives + False\ Negatives}}$$

Recall:

$$\boldsymbol{Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}}$$

Precision:

$$\boldsymbol{Precision = \frac{True\ Positives}{True\ Positives\ +\ False\ Positives}}$$

F1-Score:

$$\boldsymbol{F1\ Score = 2 * \frac{Precision * \ Recall}{Precision + \ Recall}}$$

- True Positives - instances that are positive and have been correctly classified as positive by the classification model.
- True Negatives - instances that are negative and have been correctly classified as negative by the classification model.
- False Positives - instances that are negative but have been incorrectly classified as positive by the classification model.
- False Negatives - instances that are positive but have been incorrectly classified as negative by the classification model.

8.4.1.1 Music Genre Classification

```
Accuracy: 0.7627627627627628
Recall: 0.7627627627627628
Precision: 0.7646079911329398
F1 Score: 0.762566138992498

Classification Report:
              precision    recall  f1-score   support

           0       0.74      0.76      0.75       105
           1       0.91      0.95      0.93       103
           2       0.66      0.70      0.68        86
           3       0.67      0.70      0.69       115
           4       0.73      0.77      0.75        88
           5       0.90      0.80      0.85        95
           6       0.88      0.90      0.89        98
           7       0.75      0.79      0.77        98
           8       0.79      0.67      0.72       106
           9       0.62      0.60      0.61       105

    accuracy                           0.76       999
   macro avg       0.77      0.76      0.76       999
weighted avg       0.76      0.76      0.76       999
```

Figure 8.1: Classification report of music genre

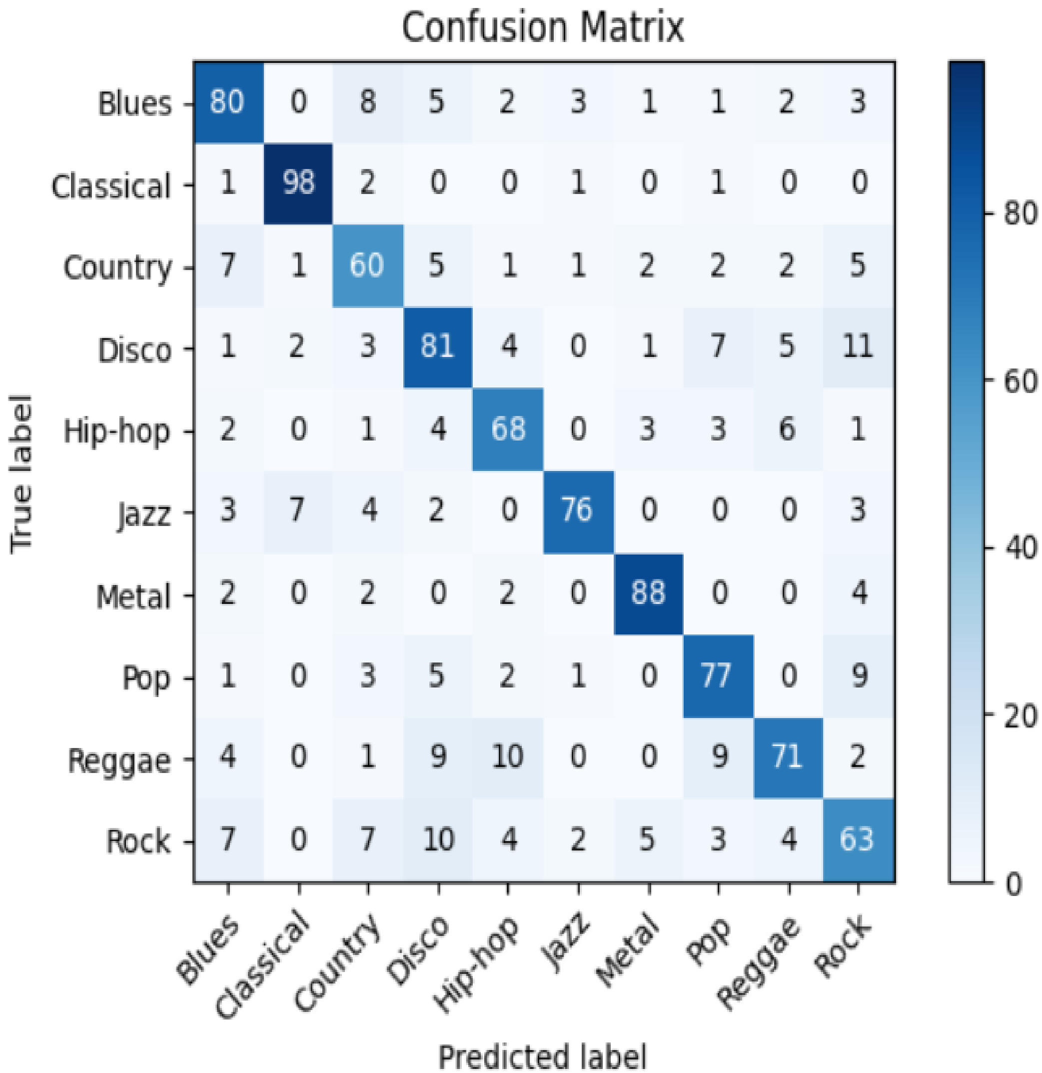


Figure 8.2: Confusion Matrix Genre

8.4.1.2 Music Instruments Classification

Classification Report:

```
Recall                :  [0.67 0.8 0.46 0.77 0.76 0.82]
Precision             :  [0.68 0.74 0.6 0.72 0.81 0.8]
F1-Score              :  [0.67 0.77 0.52 0.74 0.79 0.81]
Accuracy              : 0.75, 641
Number of samples     : 856


Train set shape       : (2567, 13)
Test set shape        : (856, 13)
Train classes shape   : (2567,)
Test classes shape    : (856,)
```

Confusion Matrix:

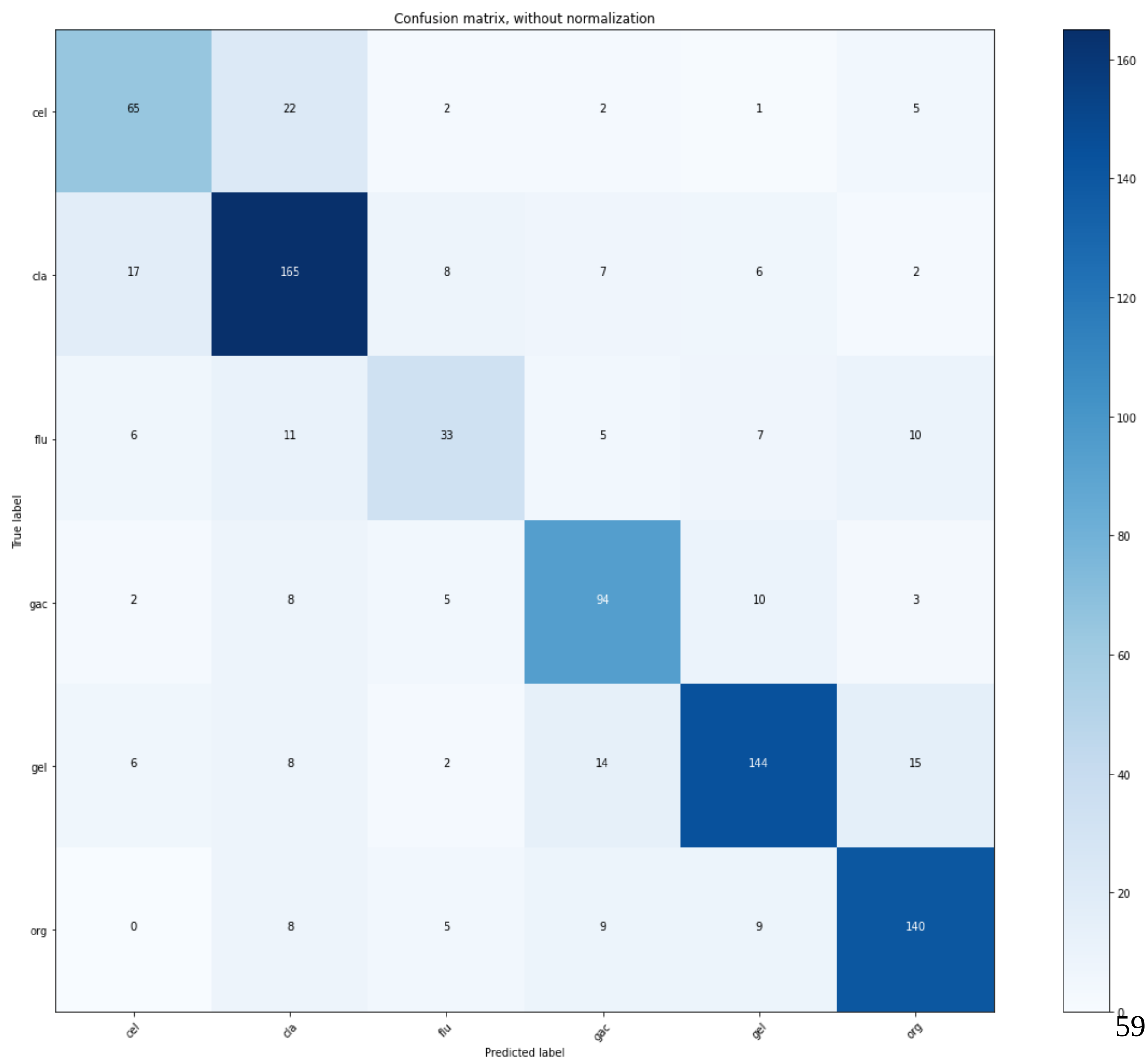


Figure 8.3: Confusion Matrix of Instrument

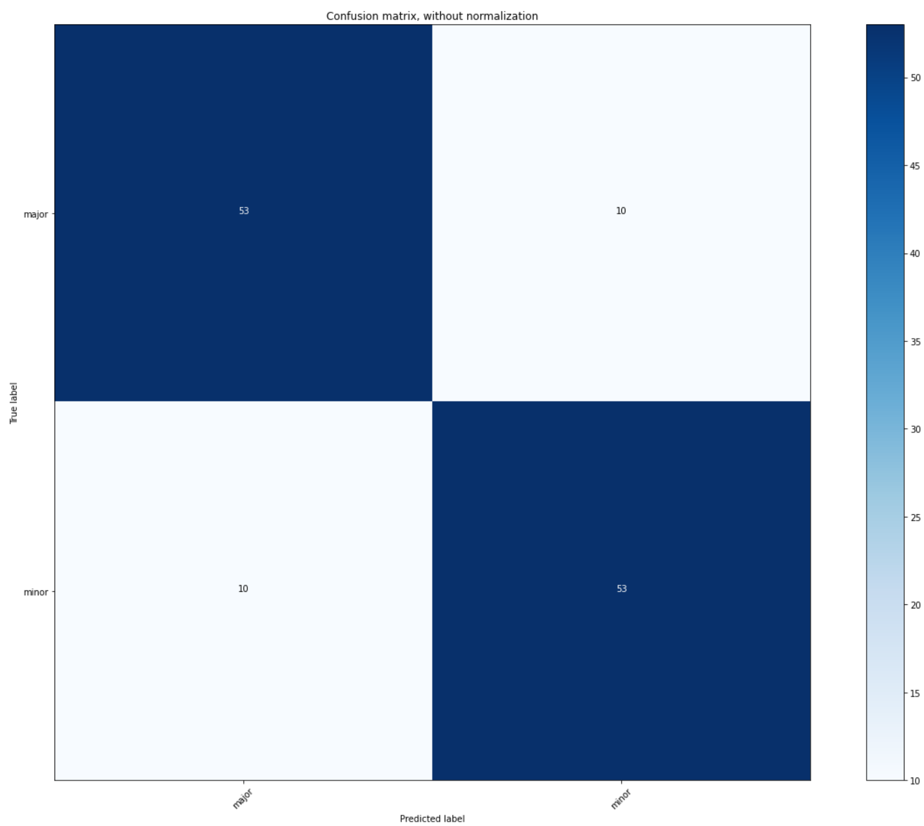
8.4.1.3 Major/Minor Chords Classification

Classification Report:

```
Recall             :  [0.84126984 0.84126984]
Precision          :  [0.84126984 0.84126984]
F1-Score           :  [0.84126984 0.84126984]
Accuracy           :  0.84  , 106
Number of samples :  126

Train set shape        : (376, 13)
Test set shape         : (126, 13)
Train classes shape    : (376,)
Test classes shape     : (126,)
```

Confusion Matrix:



8.5

Figure 8.4: Major/minor confusion matrix

8.5 Benchmarking

Benchmarking is the process of comparing one product against another leading product already in the market for the same category. This main goal of benchmarking is to find areas that are lacking in the systems and identify areas for improvement. There are 2 types of benchmarking.

1. **Competitive benchmarking** - checks how well the product being tested performs compared to the leading product.
2. **Technical Benchmarking** – identifies the capabilities of the products by comparing it to a product already in the market.

Table 8.1: Benchmarking

ML Model Type	Human Level & State of-the-Art Performance	Comparison of Model used in WhatTrack
Music Genre Classification	Human Accuracy: 60-70% (source) Leading system accuracy: 74% (source)	Dataset Used: GTZAN (source) Accuracy: 76% WhatTrack model for genre identification has surpassed the accuracy levels of both human and state of the art machine learning system.
Chord Identification	Human Accuracy: 50-60% (source) Leading system accuracy: 65% (source)	Dataset Used: Shords dataset (source) Accuracy: 84% Even though there were some serious limitations in the dataset the WhatTrack system for Chord identification performed quite well compared to human and state of the art system accuracy.

Instrument Identification	Human Accuracy: 60-70% (source) Leading system accuracy: 62% (source)	Dataset Used: IRMAS (source) Accuracy: 75% Despite the lack of data for training, our models were quite accurate in Identifying music instruments compared to human and state of the art systems.

8.6 Functional Testing

Table 8.2: Functional Testing

Test Case	Description	Input	Expected Result	Actual Result	Status
1	Checking whether the user can add details and sign up	Email, Password, Confirm Password, Profile Name, Gender	User creates the account successfully.	Account created successfully.	Passed
2	Checking whether the user can login to the system using Email and password	Email, Password	User logs in successfully	User logs in successfully	Passed
3	After successful login check whether user can access the Home Screen	No input, Redirects from the login page	User can access the components	User can access the components	Passed
4	User selects the component to predict major and minor chord prediction	Inputs an audio file	Predicts whether there is a major or minor key	Predicts major or minor according to the audio file.	Passed
5	User selects the component to predict instrument.	Inputs an audio file.	Predicts out of following instruments	Predicts one of the following	Passed

			(violin,piano,organ,trumpet, Voice and flute)	instruments	
6	User selects the component to predict music genres	Inputs and audio file.	Predicts the genre of the audio file.	Predicts the genre of the audio file.	Passed
7	After genre is predicted user selects to recommend music.	No input	Recommend a list of music names for the predicted genre of music	Recommends a list of music names for the predicted genre of music	Passed

8.7 Module and Integration Testing

Module testing, commonly referred to as unit testing, is a category of testing that focuses on examining the operation of distinct parts or modules of a software program. Developers typically carry out this task, which entails testing a module separately from the rest of the system. Module testing is used to find bugs in the code at an early stage of development before they spread to other components of the system.

Contrarily, integration testing is a sort of testing that focuses on confirming the interactions between various software application modules or components. The purpose of integration testing is to identify defects in the interactions between modules, as well as defects that arise when different modules are combined. So, integration testing has helped to ensure that the system works correctly as a whole.

8.8 Non-Functional Testing

8.8.1 Accuracy and Performance Testing

8.8.1.1 Music Genre Classification

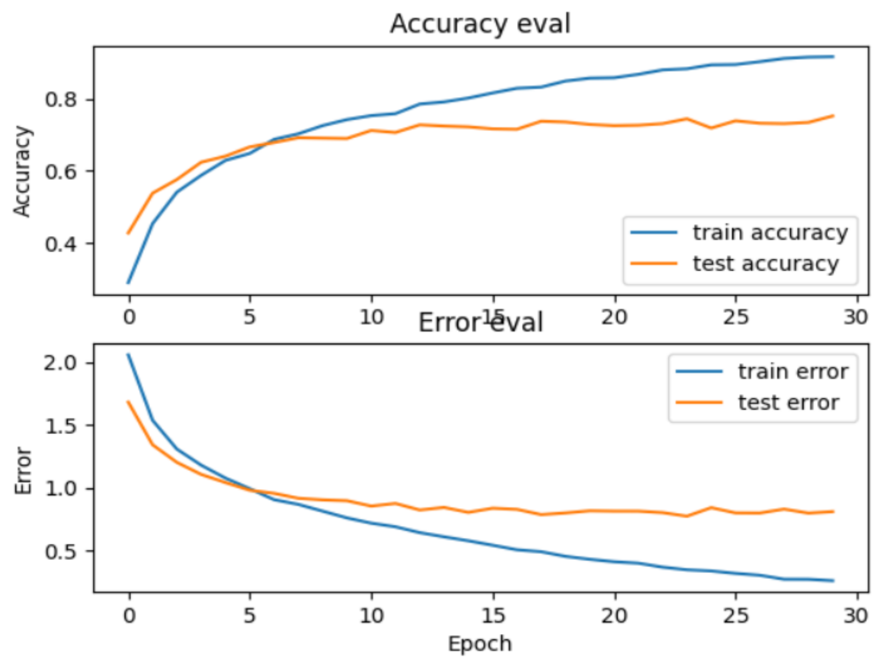


Figure 8.5: Accuracy and error graph of Music Genre

8.8.1.2 Music Instruments Classification

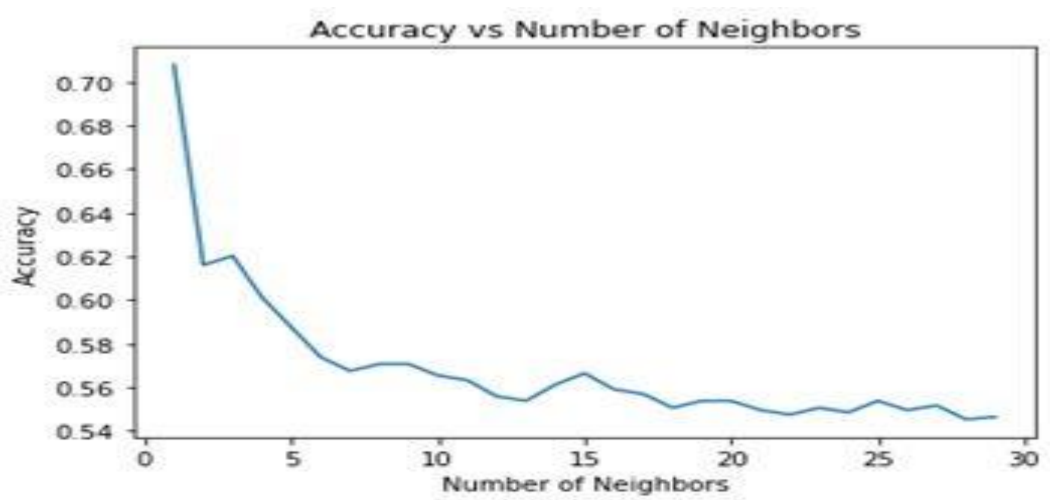


Figure 8.6: Accuracy for k=1 graph of Music Instrument

8.8.1.3 Major/Minor Chords Classification

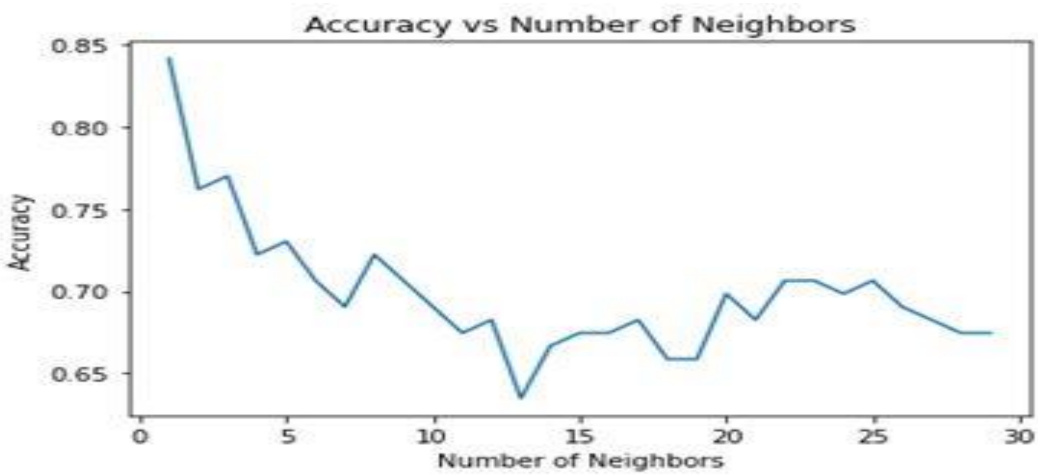


Figure 8.7: Accuracy graph of Major/minor for K=1 value

8.8.2 Security Testing

When the user uses the app for the first time, the user is required to sign up using an email and password of their choice. Users will remain signed in until they are singed out manually or for security reasons. In this case, the user is required to enter the email address and password that was used during the registration, in order to access internal pages of the application. Entering email or password that is invalid or unavailable in our database will prompt the user an error message suggesting to reverify the information that was entered.

The image of the login page is shown below:

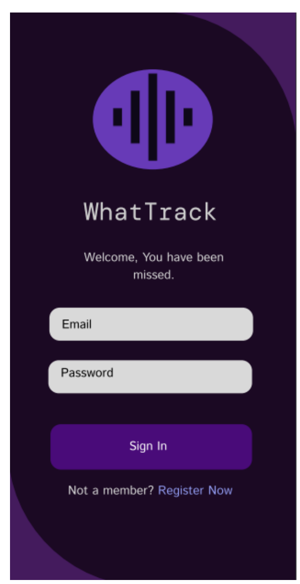


Figure 8.8: Security Testing

8.8.3 Load Balancing

The process of distributing the testing environment's workload across several resources such as servers, containers, or virtual machines for the purpose of increase the total efficiency and reliability of the process of testing. Effective load balancing has improved the performance and reliability of the mobile application, ensuring that the users can access the app quickly and efficiently, regardless of the load on the system.

8.9 Limitations

Some of the system's limitations that were identified during the testing process are listed below:

1. Processing Power: A relatively large amount of processing power is required to run the ML models.
2. Datasets: Unavailability large datasets the covers a broader set of music genres and music instruments in the public domain affects the systems accuracy.
3. Scalability: The system may struggle to handle large datasets or a high volume of simultaneous users, resulting in slow response times or even crashes.
4. Data Privacy: Handling sensitive user information such as private playlists and user history can pose a challenge.
5. Copyright issues: Legal concerns surrounding copyrighted material may limit the system's ability to provide certain information or access specific music resources.
6. Slow Response Times: Handling large number of users simultaneously can become a challenge due lack of server resources.
7. Foreign Languages: As the datasets are not very broader there is a risk of misclassification when it comes to foreign languages.

8.10 Chapter Summary

The chapters provide a brief overview of the different types of tests conducted to ensure that all the functional requirements perform at an optimal level. As we have discussed in this chapter, all the functions are tested to proper standards. The objectives of the tests were properly identified, and the maintainability, security and reliability of the system were properly tested and analyzed.

CHAPTER 09: EVALUATION

9.1 Chapter Overview

In this section, evaluations of various aspects of the system provided by domain and technical experts are provided. In addition to that self-evaluation of the system by the authors is also discussed. Functional and non-function requirements that are implemented are also discussed in this section.

9.2 Evaluation Methodology and Approach

The system uses ML models to detect key information in a audio track such as, the genre, musical instruments used, the key etc., hence its evaluation can involve consulting musicians and technical specialists. They can provide insights on these aspects:

- AI Model Effectiveness
- System's Ease of Use
- User-friendliness of the system

9.3 Evaluation Criteria

The evaluators assessed the following aspects:

1. The project's overall concept
2. Project's scope
3. Design and Architecture of the system
4. Proposed solution and prototype
5. User interface and application experience

9.4 Self-Evaluation

Table 9.1: Self-Evaluation

Criteria	Author's Evaluation
Project overall concept	A vast majority of music listeners do not have the required expertise to identify key features of a song or music track. Our systems provide an easy and straightforward way for users to detect key information in the song or music track such as genre, instrument types and key.
Scope of the project	The ML models used to music information have relatively high accuracy, which will improve as we train models regularly with new datasets.
System design, architecture, and implementation	The design and architecture have been created with usability or ease of use in mind and have met expectations. The high-level architecture of the system allows for further improvements and modifications when necessary.
Solution and Prototype	The prototype is easily accessible due to its simple and efficient design and implementation. Both the web and mobile applications contain the expected components.

9.5 Selection of Evaluators

Evaluators were chosen from two groups:

1. Domain experts with expertise in the medical field
2. Technical experts skilled in artificial intelligence applications.

The evaluators were selected from these categories as follows:

Table 9.2: Selected Evaluators

Group	Affiliation	Reason
Domain Expertise	Mr. Kanchana Jayasinghe (Composer and Pianist)	He currently works as a senior composer and pianist. Works in the music field for more than 15 years.
Technical Expertise	Mr. Rohan Fernando (Associate Director)	He works at Acuity Knowledge Partners and currently works as an associate director in specialized solutions. Has an experience of more than 10 years in the IT field.
Technical Expertise	Ms. Niwarthana Kariyabadhuge (Lecturer)	She works as a lecturer at IIT and has mentored the DSGP in tutorial sessions.

9.6 Evaluation Results

9.6.1 Overall Concept

Table 9.3: Overall Concept

Question	
What do you think about the overall project?	
Person	**Feedback**
Mr. Rohan Fernando	"Great and a unique project which can create an app"
Mr. Kanchana Jayasinghe	"A good way to enhance knowledge of music"
Ms. Niwarthana Kariyabaduge	"Good approach and a unique project, Dark theme of UI goes with the project."

Summary
The Evaluators has positive feedback about the project and happy about the approach that have taken.

9.6.2 Scope of the System

Table 9.4: Scope of the system

Question	
What do you think about the scope of the project?	
Person	**Feedback**
Mr. Rohan Fernando	"Focus should be on algorithm and I'm more curious on how to test the accuracy of this"
Mr. Kanchana Jayasinghe	"An initiative idea. Would be more beneficial for Students who are beginners to learning music"
Ms. Niwarthana Kariyabaduge	"Creating such an app to produce the listeners educational and more info about music is good."
Summary The Evaluators has positive feedback about the methods used to implement the app and happy about the system.	

9.6.3 Design, Architecture and Implementation

Table 9.5: Design, architecture

Question	
What do you think about the design, architecture, and implementation of the project?	
Person	**Feedback**
Mr. Rohan Fernando	"Great, any idea on managing

	compression techniques. Nice architecture”
Mr. Kanchana Jayasinghe	“Attractive Interface”
Ms. Niwarthana Kariyabaduge	“Dark theme goes with the project, instructs to change the menu icons to be visible more, happy about the security.”
Summary The Evaluators has positive feedback about the design and points some screens to be colorful.	

9.6.4 Solution and Prototype

Table 9.6: Solution

Question	
What do you think about the Solution and prototype of the project?	
Person	**Feedback**
Mr. Rohan Fernando	“Great UI design. Dark theme goes well”
Mr. Kanchana Jayasinghe	“Focus on predicting correct results”.
Ms. Niwarthana Kariyabaduge	“Great design some small adjustments overall introduced model is a success.”
Summary The Evaluators has positive feedback about the design and happy about how easy the users can utilize the application.	

9.7 Limitations

Table 9.7: Limitations

Person	**Feedback/ Suggestion**
Mr. Rohan Fernando	“Focus on algorithms may be better to try out some of the transformer models”
Mr. Kanchana Jayasinghe	“Try to apply the system to Sri Lankan genres and instruments”.

Ms. Niwarthana Kariyabaduge	"Increase the datasets and trained the model to get predictions for more larger time audio files, trained not only for one file format.

9.8 Evaluation on Functional Requirements

Table 9.8: Evaluation on functional requirements

Requirement ID	Requirement and Description	Evaluation	Priority
FR1	The system should be able to accept the audio track the user wants to analyze.	Implemented	Critical
FR2	Despite not having any use for the end user, metadata generated by each operation should be saved to further improve the ML models.	Implemented	Critical
FR3	The trained ML model for detecting music genre is used to detect the genre of the music track	Implemented	Critical
FR4	A user-friendly intuitive UI and API interfaces for the 3rd party apps to connect with the system.	Implemented	Desirable
FR5	The trained ML model for detecting instruments is used to detect the music instruments used in the track.	Implemented	Critical
FR6	The trained ML model for detecting chords is used to detect the music instruments used in the track.	Implemented	Critical
FR7	Based on the genre detected by the model, the system should be able to recommend similar music to the user.	Implemented	Important

9.9 Evaluation on Non-Functional Requirements

Table 9.9: Evaluation on Non-functional Requirements

	Requirement and Description	Evaluation	Priority
NFR 01	The accuracy of the system in detecting various features in an audio track should be high	Implemented	Important
NFR 02	The system should be secured to avoid unauthorized access and data breaches.	Implemented	Important
NFR 03	Model training should be efficient and not take a long time	Implemented	Important
NFR 04	The process should be done with minimum resource requirements to support as many devicesas possible.	Implemented	Non- important
NFR 05	The system will have an intuitive GUI	Implemented	Desirable

9.10 Chapter Summary

This chapter outlined the assessment of various factors by industry experts. It began by identifying the categories discussed with experts and providing a brief explanation of the evaluated criteria. The author's self-assessment was also included. Next, the documentation of expert evaluations across different categories was presented. Lastly, the evaluation of functional and non-functional requirements was discussed.

CHAPTER 10: CONCLUSION

10.1 Chapter Overview

This chapter provides an overview of all the aspects of the project that have been discussed so far. The chapter starts with the discussions of the aims and objectives of the project. The challenges that were met in implementing the functional and non-functional requirements, the skills that had to be learned to overcome the limitations are also discussed. In addition to that, the future enhancements and the limitations of the current system, ethical consideration and individual contribution to the project are discussed in detail.

10.2 Achievements of Research Aims & Objectives

10.2.1 Project Aim

The system that is developed is a multi-model music information classification system which allows users to input a music track and detect the key features in the music track like the genre, chords, and instruments used. The proposed system is implemented as a mobile application and can be used as a handy tool for both music listeners and musicians to quickly identify key information embedded in any music track.

10.2.2 Completion of objectives of the project

Table 10.1: Completion of objectives of the project

Description	Status
Literature Review	
Evaluation of existing work and systems	Completed
Software Requirements Specification	
Detailed overview of the product, requirements, and Stakeholder analysis	Completed
Design	
Design a user-friendly mobile application	Completed
Development	
Developed a working Mobile app according to the design and requirements	Completed
Testing	
Tested the ML models, Mobile Application, and the Backend.	Completed

10.3 Utilization of Knowledge from the Course

Table 10.2: Utilization of knowledge from the course

Module	Description
CM2603 - Data Science Group Project	Module laid a good foundation on how to plan, design and implement a project, how to conduct a literature review, write an SRS and provided a good understanding of market research, marketing, and business thinking.
CM2604 - Machine Learning	The module provided a good understanding on working with data, machine learning algorithms and model building, testing and evaluation of models.
CM2607 - Advanced Mathematics for Data Science	Mathematics principles learnt in this module enabled understanding of machine learning models and specially about sound waves as a Fourier Series.
CM2601 - Object Orientated Development	Learnt about version controlling, designing the software.

10.4 Use of Existing Skills

Existing skills provided the foundation when developing the project and was crucial for implementing many functional and non-function requirements of the project.

10.4.1 Machine learning

The foundations of Machine Learning were learnt using the Machine Learning Module. The knowledge helped to develop good understanding about Machine Learning models such as KNN and how they worked.

10.4.2 Mobile App Development

Mobile Development with Flutter was learnt from mainly online video tutorials available in platforms such as YouTube and other online resources such as programming websites and forums such as stack overflow.

10.5 Use of New Skills

During project implementation, the team was able to build new skills and enhance many existing skills. The project provided a good understanding on software project and the team was able gain valuable experience in field.

10.5.1 Deep learning

The foundations of Deep Learning were learnt using online sources such as YouTube and other programming related websites. This knowledge helped develop good understanding about Deep Learning models such as CNN and how they worked.

10.5.2 Practical Deep Learning and Machine Learning

Through a thorough analysis of existing literature and learning new domain-specific knowledge, the team was able to acquire extensive knowledge about the ML models that were implemented. The team allocated a considerable time to gain a deep understanding of the deep learning models that were implemented. The team made careful choices about optimizers and other heuristics in order to achieve the best possible performance from the data. The team used a data-centric approach to build the models, and they also gained extensive knowledge about feature engineering and pre-processing techniques.

10.5.3 Integrating AI Models with the Mobile Interface

Incorporating ML models into Mobile Interface presented a new challenge. Flask backends were developed to serve the ML models, with a focus on ensuring optimal inference speed and model performance.

10.5.4 Version Control

Leveraging version control with GitHub for a large-scale project of this nature proved to be advantageous in expediting development and were able to easily organize the coding tasks among the team. In addition to that, version control with GitHub provided a safe environment for app development and rigorous testing and modifications.

10.6 Achievement of Learning Outcomes

10.6.1 Develop skills necessary to collaborate within a team on a software development project.

- Version control with Git and GitHub was employed for collaborative development, promoting organization and effective collaboration.
- Trello was utilized as a team workspace for task management, improving time management skills.
- Weekly meetings conducted via Google Meet, coupled with daily discussions via WhatsApp, facilitated clear communication and synchronization among team members.

10.6.2 Present an analysis of the user centered design process, cognitive aspects, research methods, modelling and prototyping used to produce applications related to Data Science with a reflection on legal, ethical, professional and social issues.

- A Literature Review and SRS document were created before prototype development, after analysis of existing work and research.
- The product design and features were developed with a business-centric approach, aimed at generating revenue after proper market research and consultation of domain experts.
- A cross-platform mobile application was made available for most platforms.
- The entire system features cutting-edge and robust ML models optimized for accuracy.
- All datasets/data used were obtained with permission from authors and relevant parties, anonymized, and ethical boundaries were respected.

10.7 Problems and Challenges Faced

Table 10.3: Problems and challenges faced.

Problems/Challenges	Solutions/Workarounds
Hardware Requirements	Due to the complexity of the ML models and, in some cases, the large amount of data, a computer with a large RAM and GPU was required to train the models effectively.
Power Outages	Due to power outages, it was quite difficult to work at specific times. The team had to adjust their schedules and work during the nighttime when the outages lasted less, to make progress in implementations.
Time management and distance	As residents of team members were located in many cities, meeting offline and discussing issues face to face was a challenge. And managing the time between other projects and responsibilities posed a significant challenge. These challenges were mitigated with effective communication and clever delegation of the workload among members.
Dataset Issues	Despite the availability of datasets, finding large enough dataset with proper parameter was an issue. This was overcome by further dividing the audio files to increase the number of audio files available for training.
Lack of Domain Knowledge	Lack of knowledge in the music industry and music in general was a challenge. Hence, many domain experts were consulted prior development of the app.

10.8 Deviations

Initially, there were four models to be implemented. But later one model was removed ("Voice tone detection") since it wasn't practical to real life.

10.9 Limitations of the Development

Table 10.4: Limitations of the development

Model	Limitations
Music Genre Classification	The model can detect only 10 music genres. The model works with the best accuracy only for audio files that are 30 seconds long. Allows only wav files.
Music Instruments Prediction	The model can detect only 6 instruments. Works with the best accuracy only for audio files less than 5 seconds. Audio input should be a single instrumental sound. Allows only wav files.
Major Minor Chords Prediction	The model works for audio files which are 3 seconds long. Audio input should be a single music chord (played from a guitar or piano) Allows only wav files.

10.10 Future Enhancements

- **Enable users to input lengthy audio files.**
  The system only works properly for shorter audio files because the trained datasets contain shorter audio files. By training more audio data with longer time durations app can be made to work for long audio files.
- **An Inbuilt Music Player**
  An inbuilt music player that enables the user to stay in the app and play music without the need to leave the app can be a game changer for this app. This will boost the engagement rate of the users making it popular among many music lovers.

- **Utilizing User History to develop an efficient music recommendation engine.**
  Currently the app does not utilize user history to recommend personalized music to the user. A machine learning model can be used to learn from user history and recommend music that user is likely to enjoy more. Developing this feature will provide the user with a personalized experience and will make the user to stay on the platform for longer periods of time boosting the engagement rate.
- **Improvements in the range of genres and instruments.**
  Current training data is very limited when it comes to classifying a wide range of genres and musical instruments. Creating a larger dataset with a wide range of labeled music data can enhance or improve the accuracy of current models.
- **Improving prediction in almost all the instruments used to create the track.**
  In the model which has been implemented, the model can only predict the most used instrument. It cannot predict all the instruments used. So, increasing the datasets and training the model to achieve this.
- **Predicting the voice tone color**
  There are four main voice ranges. They are soprano, alto, tenor, and bass. This model will give an additional benefit to the mobile application.

10.11 Achievement of the contribution to the body of knowledge

The development of the project made several contributions to the body of knowledge. **WhatTrack** is the first multi-model music information detection system. The pre-processing techniques, feature engineering, and different ML model architectures used in the project can pave the way for further research in this area.

10.12 Individual Contribution

In implementing the proposed model “**WhatTrack”** and writing the thesis each and everybody on the team has a key role and everybody worked as a team.

Table 10.5: Individual Contribution

Team Member	Contribution
Sathila Samarasinghe	• KNN model for major and minor prediction • Audio feature extraction and preprocessing for music genre classification. • User input preprocessing and prediction of music genre classification • Daily music fact generator.
Thareen Renuja	• Preprocessing data for instruments classification • KNN model for instrument classification. • Accuracy testing for chord prediction. • Accuracy testing for instrument prediction
Ishan Fernando	• Preprocessing data for major minor chord prediction. • UI designing • Login page, authentication, and security • Integration of frontend and backend using Flask. • Integration of Firebase to collect user information. • Random avatar image to the profile based on gender.
Rikzy Jesuli	• CNN model for music genre classification. • Music Recommendation Engine from Spotify API • User feedback page to improve music recommender and predicted results.

10.13 Chapter Summary

This chapter provided an overview of the multi-model music information system prototype developed. The project afforded the team a wealth of experience and exposure in the areas of Data Science, Machine Learning, Deep Learning and Software Development. The prototype can be readily expanded into a fully-fledged business-ready product for commercialization.

APPENDIX PART 1 – REFERENCES

I. Shah, M., Pujara, N., Mangaroliya, K., Gohil, L., Vyas, T. and Degadwala, S. (2022). Music Genre Classification using Deep Learning. 2022 6th International Conference on Computing Methodologies and Communication (ICCMC). doi:10.1109/iccmc53470.2022.9753953.

II. Falola, P.B., Alabi, E.O., Ogunajo, F.T. and Fasae, O.D. (2022). MUSIC GENRE CLASSIFICATION USING MACHINE AND DEEP LEARNING TECHNIQUES: A REVIEW.ResearchJet Journal of Analysis and Inventions, [online] 3(03), pp.35–50. doi:10.17605/OSF.IO/FZQXW.

III. Dalvi, T., Kamble, S. and Yadav, S. (2022). MUSIC GENRE CLASSIFICATION. International Journal of Research Publication and Reviews Journal homepage: www.ijrpr.com, [online] 3, pp.2029–2034. Available at: https://ijrpr.com/uploads/V3ISSUE4/IJRPR3596.pdf [Accessed 15 Oct. 2022].

IV. Cai, X. and Zhang, H. (2022). Music genre classification based on auditory image, spectral and acoustic features. Multimedia Systems. doi:10.1007/s00530-021-00886-3.

V. 1Mesghali, S. and Askari, J. (2022). Designing Music Recommendation System based on Music Genre by using Bi-LSTM. [online] IEEE Xplore. doi:10.1109/ICEE55646.2022.9827016

VI. Lee, J., Shin, S., Jang, D., Jang, S.-J. and Yoon, K. (2015). Music recommendation system based on usage history and automatic genre classification. [online] IEEE Xplore. doi:10.1109/ICCE.2015.7066352

VII. Liu, Y., Yin, Y., Zhu, Q. and Cui, W. (2022). Musical Instrument Recognition by XGBoost Combining Feature Fusion. arXiv:2206.00901 [cs, eess]. [online] Available at: https://arxiv.org/abs/2206.00901 [Accessed 25 Oct. 2022].

VIII. Li, R. and Zhang, Q. (2022). Audio recognition of Chinese traditional instruments based on machine learning. Cognitive Computation and Systems, 4(2), pp.108–115. doi:10.1049/ccs2.12047.

IX. Miller, J., O’Hanlon, K. and Sandler, M.B. (2022). Improving Balance in Automatic Chord Recognition with Random Forests. [online] IEEE Xplore.

Available at: https://ieeexplore.ieee.org/abstract/document/9909558 [Accessed 25 Oct. 2022].

X. Schramm, R., McLeod, A., Steedman, M., Benetos, E. and Conference (ISMIR 2017), 18th I.S. for M.I.R. (2017). Multi-pitch detection and voice assignment for a cappella recordings of multiple singers. qmro.qmul.ac.uk. [online] Available at: https://qmro.qmul.ac.uk/xmlui/handle/123456789/25292 [Accessed 25 Oct. 2022].

XI. Musical Pursuits. (2022). Music Streaming Statistics in 2022 (US & Global Data). [online] Available at: https://musicalpursuits.com/music-streaming/#:~:text=Music%20streaming%20makes%20up%2084 [Accessed 27 Oct. 2022].

XII. HeadphonesAddict (2022). 30+ People Listening to Music Statistics & Trends (2022). [online] HeadphonesAddict. Available at: https://headphonesaddict.com/listening-to-music-statistics/.

XIII. www.youtube.com. (n.d.). 16- How to Implement a CNN for Music Genre Classification. [online] Available at: https://www.youtube.com/watch?v=dOG-HxpbMSw [Accessed 2 Mar. 2023].

XIV. www.youtube.com. (n.d.). 15- Convolutional Neural Networks Explained Easily. [online] Available at: https://www.youtube.com/watch?v=t3qWfUYJEYU&t=26s [Accessed 2 Mar. 2023].

XV. www.youtube.com. (n.d.). 14- SOLVING OVERFITTING in neural networks. [online] Available at: https://www.youtube.com/watch?v=Gf5DO6br0ts [Accessed 2 Mar. 2023].

XVI. www.youtube.com. (n.d.). 13- Implementing a neural network for music genre classification. [online] Available at: https://www.youtube.com/watch?v=_xcFAiufwd0&t=574s [Accessed 24 Feb. 2023].

XVII. www.youtube.com. (n.d.). 12- Music genre classification: Preparing the dataset. [online] Available at: https://www.youtube.com/watch?v=szyGiObZymo [Accessed 28 Jan. 2023].

XVIII. www.youtube.com. (n.d.). 11- Preprocessing audio data for Deep Learning. [online] Available at: https://www.youtube.com/watch?v=Oa_d-zaUti8&t=2s [Accessed 1 Jan. 2023].

XIX. www.youtube.com. (n.d.). 10 - Understanding audio data for deep learning. [online] Available at: https://www.youtube.com/watch?v=m3XbqfIij_Y [Accessed 16 Dec. 2022].

XX. www.youtube.com. (n.d.). 5- Computation in neural networks. [online] Available at: https://www.youtube.com/watch?v=QUCzvlgvk6I [Accessed 16 Dec. 2023].

XXI. www.youtube.com. (n.d.). 1- Deep Learning (for Audio) with Python: Course Overview. [online] Available at: https://www.youtube.com/watch?v=fMqL5vckiU0 [Accessed 15 Dec. 2022].

APPENDIX PART 2 – OTHER RELEVANT DETAILS

A survey was conducted to get information from the general comunity.

Find the survey here.

Section 1 of 6

Multimodal Music Information Classification System

Dear respondents, Thank You for taking the time to fill the form!

We are a group of second year students of BSc (Hons) in Artificial Intelligence and Data Science from the Informatics Institute of Technology (IIT), affiliated with Robert Gordon University, UK.

This is a form to collect data for our Data Science Group Project which is a Multimodal Music Information Classification System to enable the user to classify music, based on hidden details of audio clips using Machine Learning.

This survey has only 10 questions and wouldn't take more than 2 minutes. Please spare your valuable time to complete this survey which would immensely assist us in identifying the requirements of the proposed solutions and their challenges.

Responses to this survey will be treated as highly confidential and will not be disclosed under any circumstances and strictly be used for academic / research purposes only.

We want to thank you once again for helping us!

1. Are you an Android user or IOS(iPhone) user? *

- Android
- IOS

This survey was conducted to get feedback from technical and domain expertise.

WhatTrack

Section 1 of 3

Multi-Modal Music Information Classification Mobile Application

In the world, music is the language which connects people of different ethnics, religions or from younger to elder age gap. Music is the best way to relax and enjoy lives. Mainly people has the trend just to listen to music without the knowledge of it. They listen to music just as a trend or as a stress releaser for their life problems. Since this has created a gap between people who knows music and doesn't.

Hence in order to minimize this gap, Group 18 of 2nd Year Data Science Group Project in Informatics Institute of Technology (IIT) has come up with a solution ***in the form of a mobile application by utilizing techniques of Data Science and Machine Learning concepts. "WhatTrack"*** *a mobile application is introduced as the solution. The below survey is conducted to get the feedback from* ***Domain Expertise, Technical Expertise and Users.***

Mainly this mobile application consists of three main components helping to minimize this gap. Our mobile application has the ability to classify the music genres of music tracks, ability to differentiate major or minor keys and has the ability to classify the instruments used to build the music track. Users have the opportunity to add an audio track (wav file) to test the components and gets the desired prediction form the mobile application. The three components are explained below.

Music Genre Classification

Mainly ten genres of music can be classified by this component. They are ***blues, classical, music, disco, hip-hop, jazz, metal, pop, reggae, and rock.*** *After the prediction text is displayed the user can use the* ***Music Recommendation*** *to get 10 trending music lists.*

1. Upload an audio file
2. Then the API (flask API) will load the model classifier and predict the outcome. (Outcomes will be one of the above mentioned genres)
3. Sends the message to the frontend.
4. Displays the predicted genre.
5. User can use the music recommendation (using Spotify API)

The link is [here](#).