Multi-Model Approach to Recommend Personalized Music Playlist

TMP – 2023 – 24 - 065

Project Proposal Report

Sumanasekara H. P. – IT20665616

Supervised by – Mr. Thusithanjana Thilakarathne

B.Sc. (Hons) Degree in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

August 2023

1 Declaration

I hereby declare that the work presented in this proposal is entirely my own and has been conducted under my own initiative and supervision. This proposal does not incorporate, without proper acknowledgment, any material that has been previously submitted for a degree or diploma at any other university or institute of higher learning. To the best of my knowledge and belief, this proposal does not contain any material that has been previously published or written by another person, except where explicit acknowledgment is made within the text. I take full responsibility for the originality and authenticity of the content presented in this proposal. Any sources, ideas, or contributions from external individuals or works have been appropriately cited and referenced.

Name	Student ID	Signature
Sumanasekara H. P.	IT20665616	

The above candidates are carrying out research for the undergraduate Dissertation under my supervision.

…….

25.08.2023

…….

Date

Mr. Thusithanjana Thilakarathne

Supervisor

..

25.08.2023

…….

Date

Dr. Darshana Kasthurirathne

Co - Supervisor

2 Abstract

This proposed research topic focuses on developing a multi-model approach to provide a personalized music experience to users based on many facts, including real-time emotions. The objective is to ensure that the user experiences a personalized playlist that enhances the current emotion based on the provided playlist. The initial challenge is to accurately analyze the users' demographic data using a selfie, including age, gender, and nationality. The system must be using major technical mechanisms like image processing, CNN (convolutional neural networks), object detection, feature extraction, etc. to extract accurate user details to feed to the recommender system. Then we hope to use existing recommender algorithms like content filtering, collaborative filtering, and hybrid filtering. This research endeavors to establish a comprehensive recommender system with the potential to significantly enhance user experiences in the domain of music consumption.

Keywords: Content Filtering, Collaborative Filtering, CNN, Image Processing, Music Recommendation, Image, Dataset, Feature Extraction, Object Detection, Cold-start Problem.

Table of Content

List of figures Page

List of Tables Page

List of Abbreviations Description

CNN	Convolutional Neural Network
IEEE	Institute of Electrical and Electronics Engineers
AWS	Amazon Web Services
DOI	Digital Object Identifier

3 Introduction

3.1 Background Study

Music recommendation in today's world has significantly grown in recent years due to technological advancement and increase of the demand for personalized user experiences.

Figure 1 - Music consumption against other stress releasing activities in USA in 2020

Music recommender systems play a crucial role in assisting users to discover new music that aligns with their preferences, enhancing their overall music consumption experiences. Moreover, in today's world, there are so many practical applications for music recommendation, such as meditation, medical reasons, psychological reasons, depression, etc. (Figure 1). This background study aims to provide an overall idea of existing approaches in music recommendation, highlighting their strengths, limitations, potential areas for improvement, gaps, and problems, in order to lay the foundation for the development of a comprehensive music recommender system. Considering traditional recommender systems, this approach often relies only on some already existing machine learning algorithms like collaborative filtering, content-based filtering, and hybrid filtering. Basically, collaborative filtering involves analyzing user behaviors and preferences according to user-item interaction to generate recommendations. Conversely, content-based filtering is mainly focused on item-to-item interactions. This approach has been effective for a time period until it discovers its limitations, like the cold-start problem, which occurs when a new user or item is introduced to the system and limited data is available for accurate recommendations.

Then the hybrid methods came into play in order to get the best out of those already existing algorithms [2]. Initially, the researchers combined collaborative filtering and content-based approaches. There are so many mobile applications out there on the market that have these recommender systems, and some of them are Spotify, iTunes, Amazon Music, etc. Most people are using these applications rather than searching for a song or piece of music one by one just because of the song recommendation.

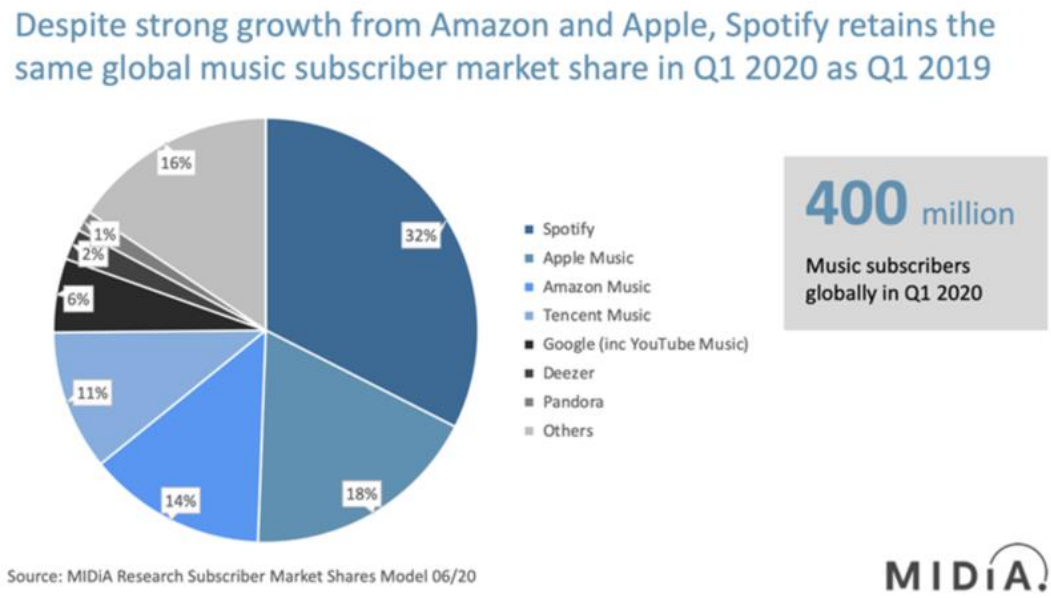

Figure 2 - Utilization of music recommender mobile applications.

With the rapid growth of technology, this approach no longer met the desires of users to the expected extent. Therefore, the researchers started to explore solutions to consider the emotional state when recommending music. As we all know, emotions play a crucial role in music preferences and can significantly impact the user's music consumption experience. Integrating real-time emotions into the recommendation process can enhance the personalization and relevance of suggested music.

Furthermore, when it comes to demographic details like, gender, age, nationality etc. the music preferences differ.

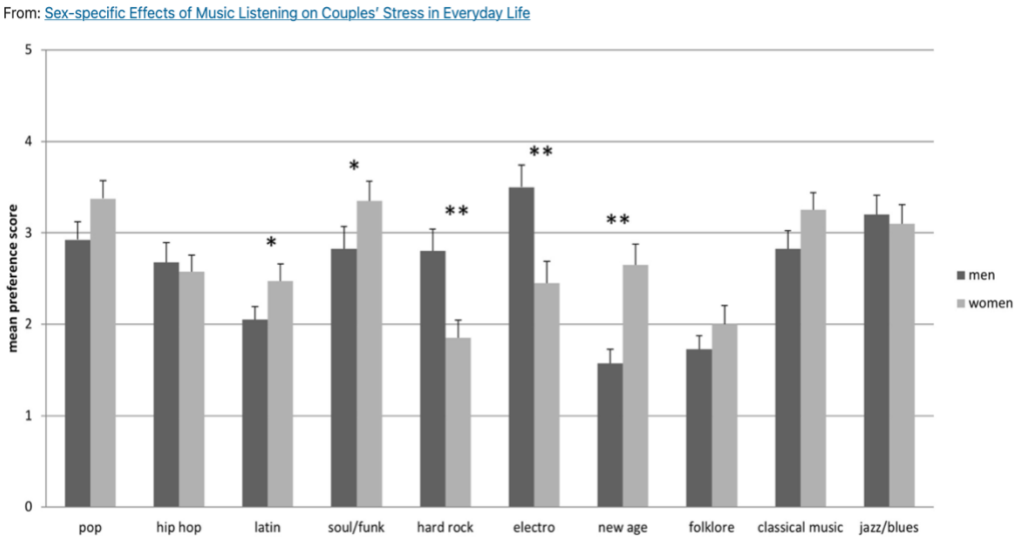

Figure 3 - Fluctuation of music preferences against gender

The above graph proves that demographic details are a valid key point in recommending songs to individuals. As I conclude, the domain of music recommendation systems has evolved from traditional collaborative and content-based filtering to leveraging advancements in deep learning and image analysis, which offer new possibilities for capturing user preferences and enhancing the accuracy of recommendations. This proposed music-recommended system aims to integrate this advancement to provide a personalized experience to the user on a new level.

3.2 Literature Review

Music recommendation has been a major aspect of life these days. There are many platforms that we can use to listen to music. Some of them are our traditional and oldest: YouTube, Spotify, iTunes, and many others. There are so many mechanisms to give filtered recommendations, like content-based filtering, collaborative filtering, and hybrid filtering [2]. [1] Recommender systems were created to bridge that gap between information gathering and analysis by filtering all available data to offer only what is most important to the user. Some research has found that content-based filtering similarity results reach up to 80% similarity for the song and 50% similarity for the artist, which means this type of filtering works well for our recommendation system [1]. Therefore, there are already tested and proven machine learning algorithms in use in recommender systems [8]. This proposed system will take this approval another step ahead and use user emotions, surroundings, and many other inputs to enhance the accuracy of the recommender system along with the content, collaborative, and hybrid filtering algorithms.

The cold start problem comes into play with these recommendation algorithms, and it was a notable problem found in the above algorithms. The "cold start" problem means that when a user first logs in to the system, there is no user history or input for these algorithms to run. This problem has also been addressed by various research projects so far and has many kinds of solutions. For instance [5], this research mitigates the cold-start problem by using matrix factorization and spatial information for users with few restaurant visits in the past. In this proposed system, we are hoping to address this problem by predicting a user profile with the use of image processing, object detection, feature extraction, etc.

Then the main challenge is to accurately extract and identify the sensitive details of the user by using a selfie. So far, there has been research conducted emphasizing this matter [3] [6]. One of the studies I found has [6] three neural network-based models to detect age, gender, and emotion, respectively, and depending on this combination, a personalized playlist has been suggested. In this case, only those combinations of inputs are sent to the recommender system, and in my research component, we are predicting a user profile at the very beginning, which will be combined with many other inputs like surroundings and voice-based emotion detection in order to enhance accuracy and personalization.

When it comes to user demographics, I have found that many researchers have proven that music preferences may differ based on the age, gender, and nationality of the user. If I could integrate these details into the recommendation, I found that it would enhance the accuracy of the recommendation. I propose the approach of integrating image analysis techniques like facial recognition to extract demographic data of the user from a selfie, which presents an innovative way to gather user information while potentially mitigating the cold-start problem.

3.3 Research Problem

At present, as I have mentioned above, music and relevant music recommendations play a crucial role in every aspect of life. There are many successful mobile applications and IoT devices at play in the world, such as Spotify, iTunes, Alexa, Amazon Music, etc. These applications have been tremendously popular and actively in use around the world up until today (Figure 2).

In contrast to existing research, I'm concerned about an underexplored area in music streaming app development. Music listening habits are highly related to personal data like age, gender, and nationality. However, providing lengthy forms for users to complete to access personalized song recommendations is not a good solution in this era. If the user gets tired of filling out forms and giving sensitive personal details to the system, the customer base might decrease. According to the literature survey, the main challenge I see here is to collect accurate user data without tiring out the user. also to predict a user profile in the very beginning so that this data will help to reduce the cold start problem.

In our proposed system, the main goal is to overcome this challenge by introducing an innovative approach. I hope to implement a convolutional neural network (CNN)-based facial image classification model. This model will harness deep learning technologies to accurately discern users' age, gender, and nationality through their facial features. Integrating this facial image classification model with our music recommendation system will create a unified mobile application.

3.4 Research Gap

According to my literature review, music recommendation and personalization have been very popular among researchers these days. Our end goal is to make a multi-model approach to recommend personalized music playlists, including demographic data and surrounding classification and emotion to enhance accuracy and the personalized experience. This has great potential to dominate the marketplace among all other existing music players.

My subcomponent is to classify the demographic details of the person, extracted through a selfie, and predict an initial user profile to send to the recommender system as an input. The below table shows the comparison, contrast, and novelty of this component against existing research.

Table 1 – Research Gap

Features	**Proposed System**	**Existing Systems / Research**				
		[1]	[3]	[4]	[5]	[6]
Music Recommendation based on age	Yes	No	No	No	No	Yes
Music Recommendation based on gender	Yes	No	No	No	No	Yes
Music Recommendation based on nationality / geographical region	Yes	No	No	No	No	No
Get user details using a selfie	Yes	No	No	Yes	No	Yes
Predict a user profile	Yes	No	No	No	Yes	No
Reduce cold start problem	Yes	No	Yes	No	Yes	No
Generate personalized Playlist	Yes	Yes	Yes	Yes	Yes	Yes

4 Objectives

4.1 Main Objective

The main objective of this proposed system is to provide a personalized music playlist to the user. The primary goal is to increase the user friendliness and accuracy of the music recommendations. To achieve this goal, there are challenging steps to take. They are capturing a high-quality selfie, analyzing the facial details, analyzing the surroundings, taking a voice command to identify the emotion of the user, and by using these inputs, generating a personalized playlist. For the above system, I identified a sub-objective to collect personal data through the selfie and to predict a user profile.

4.2 Specific Objectives

I have distributed my component into few specific objectives, and I have mentioned them below.

- Collect a high-quality image to process facial data into image classification model to identify the gender, age, and nationality / geographical region of the user.
- Train the image classification model.
- Predict the user profile based on extracted details.
- Send the user profile as an input to recommender system.
- Train the music recommender model.
- Develop the mobile application accordingly.
- Integrate above features into the developed mobile application.

5 Requirements

5.1 Functional Requirements

Out of all the functional requirements, the most important thing for the developer to do is build up the solution. Functional requirements are the key components of a system, and they describe the end goal and the user's expectations of the system. During the research, I found the following key functional requirements:

- Capture an image

First and foremost, the interface should provide a camera to capture a selfie image of the user and it should be in enough high quality for further processes. Since we consider the surroundings along the face, the image should be taken and uploaded in real time.

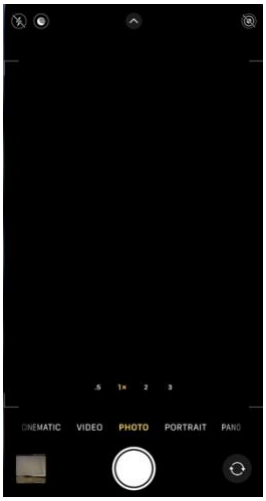

Figure 4 - Camera Interface

- System should be able to extract the face from the image.
- System should classify personal details like age, gender, nationality etc. out from the image.
- System should predict a user profile.
- System should generate a personalized playlist accordingly.
- User should be able to listen to playlist from the software application.
- System should be able to track and train the models for future recommendations.

5.2 Non-Functional Requirements

Non-functional requirements are a critical aspect of software development because they define how a software application operates and performs, rather than just its functional features. These requirements focus on aspects such as performance, security, usability, scalability, and other qualities that contribute to the overall user experience and system effectiveness. The success of the product is dependent on these non-functional requirements.

- **Performance**

This is very important. It mainly focuses on the speed, responsiveness, and overall efficiency of the system. As developers, we should emphasize response time, loading time, and minimum and efficient resource utilization in order to give a smooth user experience.

- **User Interface and User Experience (UI / UX)**

When we create a mobile application for commercialization, the app should have an intuitive and visually appealing interface that aligns with modern design standards. and the navigation should be clear to the user. We also need to think about accessibility for people with special needs.

- **Security and Privacy**

This is the most important point in this era. We should use the latest technologies to encrypt data for authentication and authorization. Developers should also be aware of data privacy rules and regulations.

- **Scalability**

Should think about future utilization and ensure whether the databases and servers are scalable.

- **Reliability and Availability**

Users need to have trust in the system, and it needs to be available and dependable whenever they need it. The system should have minimum downtime for maintenance and updates.

6 Methodology

The methodology for the proposed multi-model music recommender system that enhances the personalized experience of the user is mentioned below. There are tools and technologies that are going to be used for the implementation of the system.

Algorithms

- CNN
- RCNN
- ANN

(Selection of the algorithms can be differ during the implementing based on the best approach)

Integrated development environment (IDE)

- PyCharm or Anaconda

Databases

- MongoDB
- Firebase

Backend

- Image Processing
- Image Analysis
- Feature Extraction
- Python - handle algorithms OpenCV framework

Datasets

- Kaggle

Overall, the above process includes data collection, data preprocessing, implementing a machine learning model, testing and training the model, deploying the system, maintenance, and updates for the improvement of the system's performance. The system's success depends on the quality and quantity of data used to train the model, the accuracy and reliability of the machine learning algorithms, and the effectiveness of the system in classifying user details.

6.1 Overall System Diagram

The below figure demonstrates the over system in a high-level view.

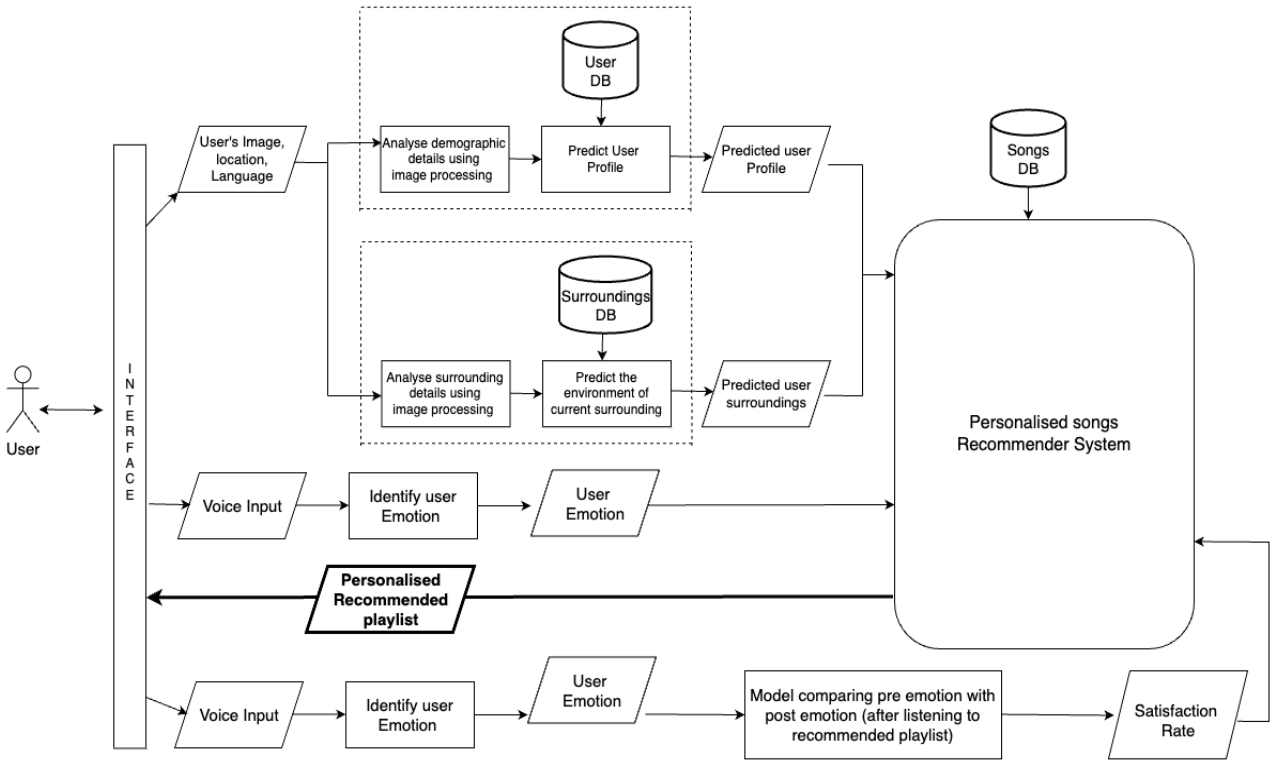

Figure 5 - System Overview Diagram

As the very first step, the user must capture an image of the face to collect facial specifications to predict the age, gender, geographical location (county or island), region, and nationality to give the very first high-level music recommendation to reduce the cold start problem. Then the application typically collects user data through various means, including speech recognition, image processing, and the surroundings. Then there is a new aspect to the post-emotions after listening to the recommended songs to enhance the accuracy of the recommendation in the future. After integrating all these models, algorithms, and subcomponents, there will ultimately be personalized play lists for different moods and occasions for the user (Figure 5).

6.2 Component Overview Diagram

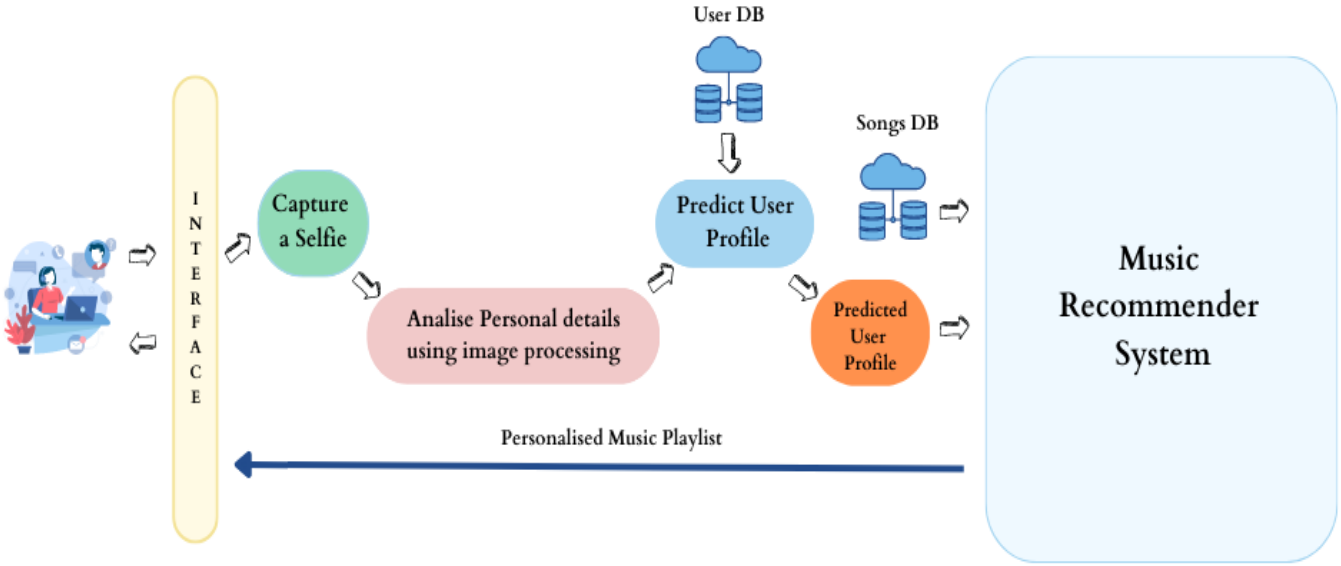

Figure 6 - Component Overview Diagram

6.3 Gantt Chart

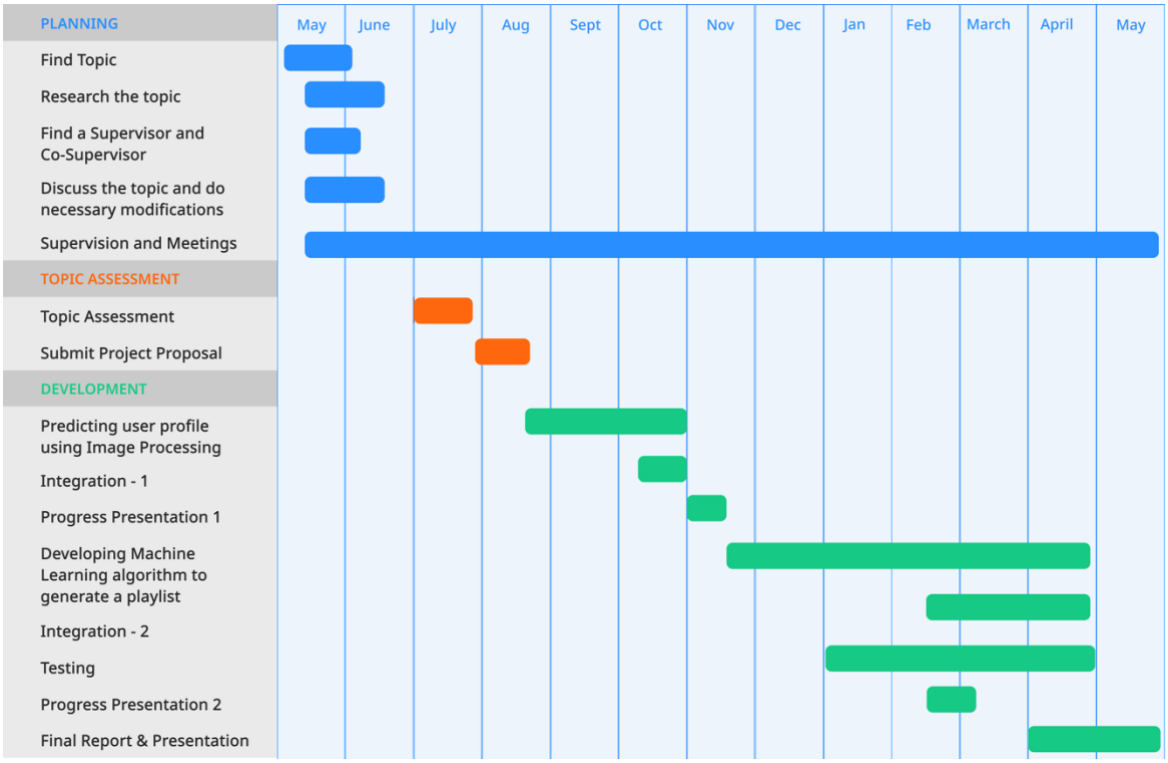

Figure 7 - Gantt Chart

6.4 Work Breakdown Chart

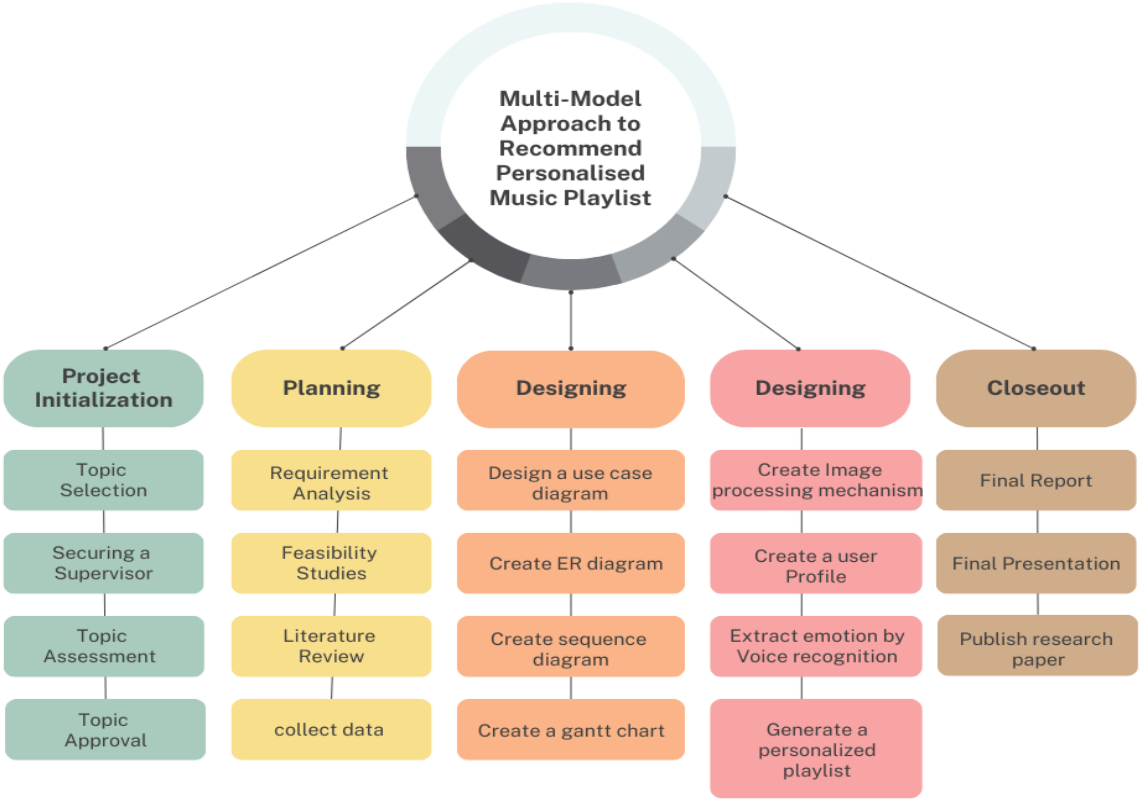

Figure 8 - Work Breakdown Chart

7 Limitations and Challenges

My research scenario presents an interesting and innovative approach to tackling the challenges of personalizing music recommendations without burdening users with lengthy forms. However, as always there are several limitations and challenges that I face when implementing above solution.

- **Privacy Concerns**

The first and foremost and one of the most important challenge is to gather images of users because it comes with significant privacy concerns. Users might be uncomfortable and unwilling to share their images.

- **Biases in Data**

The accuracy of the recommender model heavily corelated on the diversity of the training data. If the training dataset is biased in a particular group, the predictions might be inaccurate for certain demographics.

- **User Acceptance**

Users might resist the idea of using their facial images for demographic prediction. Specially in third world countries like Sri Lanka, the new technology acceptance is very low. The level of user acceptance could impact the success of the proposed system.

- **Technical Complexity**

Implementing a convolutional neural network (CNN) for facial image classification is complex. Ensuring the model's efficiency and accuracy could raise technical challenges.

- **User Experience**

While my main goal is to reduce the old-fashioned forms, some users might still find sharing facial images more intrusive or inconvenient than filling out forms. Ensuring a smooth and comfortable user experience is crucial.

- **Accuracy and Reliability**

The accuracy of facial image classification models can vary based on the quality of images, lighting conditions, facial expressions, and more. It's important to thoroughly evaluate the model's reliability in predicting age, gender, and nationality.

8 Test Plan

Testing for the suggested system will occur at different project stages. This aids in identifying bugs within each component, facilitating their independent resolution instead of addressing the entire project. Consequently, the testing approach will encompass multiple phases and protocols.

1. **Unit Testing**

   Individual unit testing will take place for every element, including both the facial image classification model and the music recommendation model. This approach allows for the isolation and rectification of bugs within each element. In this context, the researchers will concentrate on two primary dimensions,

   a) Performance testing of the component.
   b) Accuracy testing of the component.

2. **Integration Testing**

   Integration of the components will be a major task of this research project. Components will be integrated one by one and tested simultaneously because integration can cause major bugs in the system.

3. **Final Testing**

   Final testing will be done to make sure the system is performing well without any issues. The finished product will be tested using different test cases and sample data. In the second phase of the final testing, the mobile application will be given to some selected users, and their feedback will be taken. The user experience of the mobile application will also be measured by the users, and we will fine-tune the user interface of the mobile app to provide a better user experience to the end-user.

9 Budget and Commercialization

Given the daily usage of music players by individuals, this project holds great potential for commercial success. People are willing to invest reasonably in an enhanced music player experience, showing a substantial commercial value. With established market leaders like Spotify, iTunes, and Deezer already present, ensuring a competitive and equitable pricing strategy for the music player is essential. Notably, prevailing subscription models for Spotify, Apple Music, and Deezer typically demand around $10 per month. However, this pricing is regarded as steep by some, prompting them to question its value. Consequently, the following subscription model is proposed to drive the commercialization of this mobile app.

Table 2 - Subscription Types

	Free version	**Paid version (<$10/month)**
Advertisements	Yes	No
Monthly charges for the users.	No Revenue will be generated from the advertisements showed to the user while the user is using the mobile application.	Yes Revenue will be generated from the monthly charges paid by the user.
Features	All features	All features

The final mobile application will be focused on different user groups; therefore, it will be marketed to each user group using different methods.

1. Young People – social media, gaming advertisements
2. Adults – worldwide news
3. Tech People – in-depth technical advertisements, new technologies, new trending applications

Below is the budget that has been planned for the project. Charges will be changed from time to time, and final charges will be based on the consumption of the resources used in the cloud environment.

Table 3 - Budget Plan3

Description	**Amount (USD)**
1. AWS Cloud database (S3) for facial images • To store user images collected through mobile app.	0.023 per GB / Month
2. AWS Cloud database (EFS) for user demographic data. • To store demographic data of the users.	0.30 per GB / Month
3. AWS glacier to store User logging from the mobile application.	Storing = $0.004 per GB / Month Retrieving = $0.01 per GB
4. Paper Publications and documentation.	50 - 100

10 References

[1]. J. T. Anthony, G. E. Christian, V. Evanlim, H. Lucky and D. Suhartono, "The Utilization of Content Based Filtering for Spotify Music Recommendation," 2022 International Conference on Informatics Electrical and Electronics (ICIEE), Yogyakarta, Indonesia, 2022, pp. 1-4, doi: 10.1109/ICIEE55596.2022.10010097.

[2]. J. Singh, "Collaborative Filtering based Hybrid Music Recommendation System," 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), Thoothukudi, India, 2020, pp. 186-190, doi: 10.1109/ICISS49785.2020.9315913.

[3]. P. Darshna, "Music recommendation based on content and collaborative approach & reducing cold start problem," 2018 2nd International Conference on Inventive Systems and Control (ICISC), Coimbatore, India, 2018, pp. 1033-1037, doi: 10.1109/ICISC.2018.8398959.

[4]. O. Ghosh, R. Sonkusare, S. Kulkarni and S. Laddha, "Music Recommendation System based on Emotion Detection using Image Processing and Deep Networks," 2022 2nd International Conference on Intelligent Technologies (CONIT), Hubli, India, 2022, pp. 1-5, doi: 10.1109/CONIT55038.2022.9847888.

[5]. E. Mayrhuber and O. Krauss, "User Profile-Based Recommendation Engine Mitigating the Cold-Start Problem," 2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), Maldives, Maldives, 2022, pp. 1-6, doi: 10.1109/ICECCME55909.2022.9988037.

[6]. J. Jayakumar and P. Supriya, "CNN based Music Recommendation system based on Age, Gender and Emotion," 2022 6th International Conference on Electronics, Communication and Aerospace Technology, Coimbatore, India, 2022, pp. 1356-1359, doi: 10.1109/ICECA55336.2022.10009391.

[7]. S. Gilda, H. Zafar, C. Soni and K. Waghurdekar, "Smart music player integrating facial emotion recognition and music mood recommendation," 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai, India, 2017, pp. 154-158, doi: 10.1109/WiSPNET.2017.8299738.

[8]. Y. -C. Wang, P. -L. Yang, S. -I. Sou and H. -P. Hsieh, "The MuTube Dataset for Music Listening History Retrieval and Recommendation System," 2020 International Computer Symposium (ICS), Tainan, Taiwan, 2020, pp. 55-60, doi: 10.1109/ICS51289.2020.00021.

[9]. V. P. Sharma, A. S. Gaded, D. Chaudhary, S. Kumar and S. Sharma, "Emotion-Based Music Recommendation System," 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 2021, pp. 1-5, doi: 10.1109/ICRITO51393.2021.9596276.

[10]. M. Schedl, D. Hauger, K. Farrahi and M. Tkalčič, "On the Influence of User Characteristics on Music Recommendation Algorithms", Lecture Notes in Computer Science, pp. 339-345, 2015. Available: http://link.springer.com/10.1007/978-3-319-16354-3_37.

11 Appendix

11.1 Plagiarism Report