Multi-Model Approach to Recommend Personalized Music Playlist

TMP – 2023 – 24 - 065

Final Report

Supervised by – Mr. Thusithanjana Thilakarathne

B.Sc. (Hons) Degree in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

May 2024

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 document 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.

Name	Student ID	Signature
Sumanasekara H. P.	IT20665616	
Gunasekara C. M.	IT20665852	Malshan
Fernando M. P. T. K.	IT20610852	
Dhananjaya W. K. S.	IT20667078	

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

……………………………. 2024.05.09

……………………………… Date

Mr. Thusithanjana Thilakarathne

Supervisor

..

Dr. Darshana Kasthurirathne 2024.05.09

……………………………. Date

Co - Supervisor

Abstract

In today's technologically driven world, the demand for enhanced music experiences drives the exploration of innovative recommendation systems. This research focuses on personalized music suggestions based on users' age, gender, surroundings, and current emotion. Leveraging advancements in machine learning and signal processing, the study delves into emotion recognition through voice frequency, aiming to tailor recommendations to users' emotional states. Additionally, the project emphasizes the profound emotional impact of music, seeking to create a system that comprehensively understands users' emotional responses to music playlists. Moreover, by incorporating real-time weather context, the research proposes a music recommendation system that offers personalized suggestions aligned with prevailing weather conditions, further enhancing the immersive nature of music experiences.

Keywords: CNN, Image Processing, Music Recommendation, Image, Dataset, Current emotions, Current Weather, RNN

Table of Content

Table of Contents

List of figures Page

List of Tables Page

List of Abbreviations	Description
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
IEEE	Institute of Electrical and Electronics Engineers
AWS	Amazon Web Services
DOI	Digital Object Identifier
ML	Machine Learning
DL	Deep Learning
MFCC	Mel-Frequency Cepstral Coefficient
UI	User Interface

1 Introduction

1.1 Background Study

In the fast-paced contemporary world, where everyone is engaged in a race against time, ‘MUSIC’ has become a potent antidote to the stresses and pressures of the modern life. The demand for unique audio experiences grows exponentially as a result. To meet this need, researchers are delving deep into the realms of music recommendation, personalization, and user experience enhancement.

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

Music recommendation systems are essential tools that assist users in discovering new music that aligns with their preferences, thereby enhancing their overall music consumption experiences. These systems have diverse applications in areas such as meditation, medical and psychological purposes, and addressing issues like depression. This research background aims to provide an overview of existing methodologies in music recommendation, highlighting their strengths, limitations, potential areas for improvement, existing gaps, and challenges. The goal is to lay the foundation for the development of a robust and effective music recommender system.

In the realm of traditional recommendation systems, common approaches include collaborative filtering, content-based filtering, and hybrid filtering. Content-based filtering functions by analyzing a user's consumption history from the moment they join the platform, adapting recommendations to their individual preferences and behaviors. Collaborative filtering, on the other hand, identifies user data to combine commonalities and similarities, generating suggestions based on shared preferences. Hybrid filtering increases the strengths of both techniques, leveraging a nuanced understanding of user behavior to provide highly personalized recommendations.

However, these methods can face challenges, such as the cold-start problem when new users or items are introduced, leading to insufficient data for accurate recommendations. To address these challenges, hybrid methods that combine different approaches have been developed and implemented in popular music applications like Spotify, iTunes, and Amazon Music.

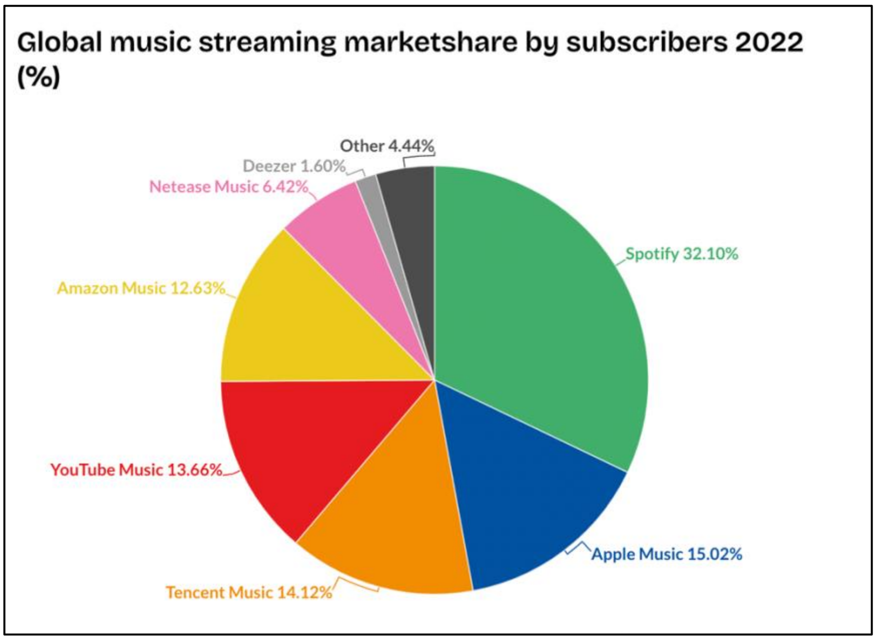


Figure 2 - Utilization of music recommending applications

Researchers have started exploring solutions that incorporate emotional states into music recommendations, recognizing the significant influence of emotions on music preferences and user experiences. By integrating real-time emotional cues into the recommendation process, there is potential to enhance the personalization and relevance of recommended music. Additionally, demographic factors such as gender and age play crucial roles in shaping music preferences, with different demographic groups exhibiting distinct musical tastes. Therefore, considering demographic details alongside emotional states can further refine the precision of music recommendations, catering more effectively to the diverse preferences of users. The diagram below is displaying the effect happens to emotions through music (Figure 3).

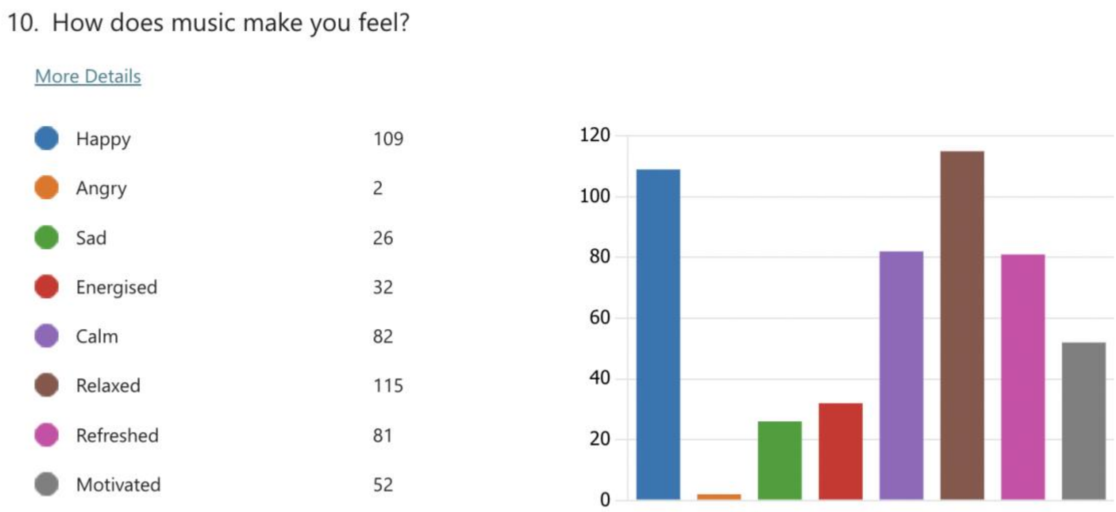


Figure 3 - Effects on emotions caused my music

Moreover, the integration of emotion recognition technology in user interfaces introduces a new dimension to user engagement by enabling computers to understand and respond to users' emotional states in real-time. This trend has led to the development of innovative applications such as voice-based emotion recognition and personalized playlist creation, aimed at tailoring user experiences to their emotions and preferences. Voice-based emotion recognition has made significant advancements, leveraging sophisticated techniques like Mel-Frequency Cepstral Coefficients (MFCCs) and deep learning models to accurately discern emotional states from speech signals. This advancement has paved the way for emotionally intelligent interactions and personalized experiences in music recommendation systems.

The existing research emphasizes the importance of seamlessly integrating real-time emotion detection into personalized recommendation systems. However, this integration poses challenges, such as creating emotionally coherent playlists that align with users' current emotional states while respecting their preferences. Ongoing research is focused on developing intuitive interfaces for emotion input identification and algorithms capable of generating emotionally consistent playlists. Past studies have contributed significantly to this field, exploring emotion-aware recommendation systems and utilizing facial expression detection for tailored music recommendations. These efforts have laid the groundwork for innovative techniques that aim to bridge the gap between emotion identification and playlist generation. Collaboration across disciplines and a user-centric approach are essential for fully realizing the potential of integrating emotions into music recommendation systems to enhance the overall human-computer interaction experience.

This expanded background study delves into the intricate evolution of music recommendation systems and the emerging trend of integrating emotion recognition technology into these systems. By shedding light on the challenges, recent advancements, and future directions in this dynamic field, this study serves as a foundational resource for envisioning and developing sophisticated, emotionally intelligent recommendation systems that are adept at catering to the diverse and evolving needs, preferences, and emotional states of users. The convergence of demographic insights, cutting-edge technologies, and real-time emotion recognition holds the promise of reshaping the landscape of online music discovery and consumption, offering users a more personalized, immersive, and emotionally resonant music experience.

1.2 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 [1][2]. 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 paper "Music recommendation based on embedding model with user preference and context" [6] presents an innovative approach to music recommendation by integrating user preferences and contextual information into an embedding model. The authors emphasize the significance of considering both user preferences and contextual factors to enhance recommendation accuracy. By utilizing embedding techniques, the proposed framework represents music items, users, and contextual features in a unified vector space, facilitating more effective computation of similarity metrics. Key components of the method include capturing user preferences and context information, leveraging collaborative filtering techniques, and incorporating Markov processes for contextual modeling. By combining user preference and contextual information within the embedding model, the proposed approach aims to provide more personalized and relevant music recommendations, thus contributing to the advancement of music recommendation systems.

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 [3], the paper "User Profile-Based Recommendation Engine Mitigating the Cold-Start Problem" presented at the 2022 International Conference on Electrical, Computer, Communications, and Mechatronics Engineering proposes a recommendation engine focused on user profiles to address the cold-start problem. By creating user profiles based on demographic and behavioral data, the system aims to provide personalized recommendations even for new users with limited interaction history. The approach contributes to mitigating the challenges associated with the cold-start problem, ultimately enhancing the effectiveness of recommendation systems.

Another approach is discussed in Darshna's paper, presented at the 2018 International Conference on Inventive Systems and Control in Coimbatore, India, introduces a music recommendation system that combines content-based and collaborative approaches while addressing the cold-start problem. The system aims to overcome the challenge of providing accurate recommendations for new users with limited interaction history. By integrating both content-based analysis, which evaluates music attributes, and collaborative filtering, which considers user interactions, the system strives to offer personalized recommendations to users of varying familiarity with the platform.

On the other hand, one of the major challenges is to accurately extract and identify the accurate data of the user. Up to date, there has been research conducted emphasizing this matter. One of the studies I found has [9] 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 to enhance accuracy and personalization.

Train human-computer interaction (HCI) is also used to recognize facial emotions in some research [4]. These apps are trained to store sensitive data like ambient conditions, indoor/outdoor temperature, time, status, etc. To this end, the new system works to provide a personalized experience by using scene analysis [7] to find the most appropriate music recommendation based on users’ current surroundings.

Moreover, In the realm of human-computer interaction, there's a growing interest in integrating emotion recognition with personalized recommendations. Being able to accurately detect users' emotional states from voice inputs and subsequently provide playlist suggestions tailored to their context holds immense promise in revolutionizing human interactions with technology. This section commences with an introduction to the background and context of voice-based emotion recognition and playlist construction. It then proceeds to delve into a comprehensive literature review, shedding light on significant research contributions, challenges, and existing gaps in the field.

The human voice can express emotions through changes in how we speak, like tone and rhythm. Researchers have made progress in recognizing these emotions using methods like Mel-Frequency Cepstral Coefficients (MFCCs), which analyze sound, and deep learning models [9][10]. These studies show that we can tell how someone feels from their speech in real-time.

In summary, the literature review reveals a diverse range of approaches and techniques employed in music recommendation systems, emphasizing the significance of considering demographic details, user behavior, content features, and emotional cues to enhance recommendation accuracy and personalization.

1.3 Research Problem

The research problem at hand is to explore the development of music recommendation systems that transcend conventional approaches by seamlessly integrating real-time environmental factors, user preferences, and post-listening emotional states. This entails leveraging cutting-edge technologies such as deep learning, computer vision, and audio classification to create personalized recommendations tailored to individual moods, surroundings, and emotional responses. By addressing challenges such as user fatigue, privacy concerns, and the cold start problem, the aim is to revolutionize the overall listening experience. This research seeks to bridge the gap between existing methods and user-centric needs, paving the way for a new generation of music recommendation systems that not only understand users' music tastes but also adapt to their unique contexts in real-time, thereby enriching user engagement and satisfaction with music streaming applications.

1.4 Research Gap

A notable research gap in the domain of music recommendation and personalization pertains to the insufficient integration of voice-based emotion recognition with the generation of personalized playlists. Despite substantial research advancements in both domains over the past decade, there is a noticeable absence of cohesive platforms that effectively amalgamate these elements to provide customized music suggestions based on users' real-time emotional states. While individual studies have made significant progress in enhancing emotion recognition algorithms and playlist creation techniques, the lack of synergy between these components impedes the development of comprehensive systems capable of delivering tailored recommendations that align with users' current emotional contexts. Moreover, existing research often overlooks practical implementation and user-centered design considerations, thereby disregarding the intricacies of contemporary interaction dynamics and contextual recommendation processes. Consequently, there is a critical need for the establishment of integrated systems that not only leverage state-of-the-art technologies but also prioritize user experience by incorporating intuitive interfaces and adaptive engagement strategies. Addressing this research gap is vital for advancing the field of human-computer interaction and adopting better user satisfaction in the increasingly digitalized landscape.

Table 1 – Research Gap

Features	**Proposed System**	**Existing Systems / Research**									
		[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
Music Recommendation based on age	YES	No	No	No	No	Yes	No	No	No	No	No
Music Recommendation based on gender	YES	No	No	No	No	Yes	No	No	No	No	No
Get user details using a selfie	YES	No	No	Yes	No	Yes	No	No	No	No	No
Predict a user profile	YES	No	No	No	Yes	No	No	No	No	No	No

Reduce cold start problem	YES	No	Yes	No	Yes	No	No	No	No	No	No
Predict current weather	YES	No	No	No	No	No	Yes	No	Yes	No	No
Predict weather indoor or outdoor	YES	No	No	No	No	No	Yes	No	Yes	No	No
Predict user's current emotion	YES	No	No	No	Yes	No	No	No	No	No	No
Predict user emotion by a voice clip	YES	No	No	No	No	No	No	No	No	No	Yes
Assess effect on the user's emotion after listening to the generated playlist	YES	No	No	No	No	No	No	No	No	No	No
Generate personalized Playlist	YES	Yes	Yes	Yes	No	No	Yes	No	No	No	No

2 Objectives

2.1 Main Objective

The primary objective of this recommender system is to enhance user experience by delivering personalized music playlists tailored to individual preferences and emotions. This entails improving user-friendliness and recommendation accuracy through a series of challenging steps, including capturing high-quality selfies, analyzing facial details, assessing environmental cues, and utilizing voice commands to identify user emotions. By integrating these inputs, the system aims to generate bespoke playlists that resonate with each user's unique tastes and mood. Additionally, a sub-objective involves leveraging selfies to collect personal data and predict user profiles, thereby further refining the recommendation process.

2.2 Specific Objectives

1. Predicting a user profile using a selfie

- **Collecting a high-quality image to process facial data:**

  This objective involves designing a mechanism within the mobile application to capture high-quality images of users' faces. These images serve as input for the facial data processing pipeline.

- **Processing facial data into image classification model:**

  Upon image capture, the facial data undergo preprocessing to extract relevant features. Subsequently, this processed data is fed into an image classification model, such as a convolutional neural network (CNN), to identify key attributes such as gender, age, and potentially nationality or geographical region.

- **Training the image classification model:**

Before deployment, the image classification model must undergo rigorous training using labeled datasets. This involves iteratively adjusting model parameters to optimize performance in accurately classifying facial attributes.

- **Predicting the user profile based on extracted details:**

  Once the image classification model is trained, it can predict user profiles based on the extracted facial details. These profiles encompass demographic information such as gender, age, and potentially nationality or geographical region.

- **Sending the user profile as input to the recommender system:**

  The predicted user profile is then transmitted as input to the music recommender system. This enables the recommender system to tailor music recommendations based on the user's demographic characteristics and preferences.

2. Predicting indoor/outdoor and the current weather by using a photo.

- **Obtaining Data Sets:**

  Obtaining suitable datasets is crucial for training and testing image analysis and feature recognition models. These datasets should include images of various weather conditions, such as sun, clouds, rain, snow, and indoor-outdoor conditions. Data sources can be publicly available, weather databases. Careful selection and processing of high-quality images ensure model strongly and accuracy.

- **Image Analysis and Context Detection:**

  The study uses image processing techniques and deep learning algorithms to analyze datasets and develop models capable of recognizing indoor-outdoor features, particularly weather conditions. Convolutional neural networks (CNNs) are used to learn hierarchical features from visual data. The trained models can identify patterns and features indicating different weather conditions, allowing them to accurately classify prevailing weather conditions in real time.

- **Validating the Performance of the Developed Models:**
  Validation is crucial for evaluating the performance and efficiency of image analysis and feature recognition models. It involves accuracy, retesting, and evaluation using separate validation data sets. Techniques like validation and comparison with existing methods ensure strongness and generalizability across different scenarios. Continuous optimization based on validation results improves model performance and reliability.

3. Predicting user emotions by a voice emotion.

- **Collect speech data inputs from the Kaggle dataset for an emotion classification model that can recognize various emotions.**

  This step involves sourcing speech data from the Kaggle dataset, specifically curated for emotion classification tasks. The dataset should encompass a diverse range of emotional expressions, ensuring comprehensive coverage across various emotional states. Selecting a representative dataset with ample samples for each emotion category is crucial for training a robust emotion classification model.

- **Convert raw voice signals from the Kaggle dataset into a set of acoustic features using feature extraction.**

  Once the speech data is collected, the next stage is to preprocess the raw voice signals and extract relevant acoustic features. Techniques like Mel-frequency cepstral coefficients (MFCCs), pitch, intensity, and formants are employed to capture essential acoustic properties from the audio data. These extracted features serve as input for the subsequent emotion classification model.

- **Train the emotion classification model on speech samples from the Kaggle dataset.**

  With the preprocessed audio features in hand, the emotion classification model is trained using machine learning or deep learning algorithms. Selection of an appropriate model architecture, such as LSTM-based recurrent neural networks, is crucial for effective emotion recognition. Training involves optimizing model parameters and evaluating performance through cross-validation techniques to ensure accurate classification of emotional states.

- **Train the music recommendation model:**

  Concurrently with the emotion classification model, the music recommendation model is trained using merged datasets that incorporate emotion-labeled speech samples. Leveraging the emotional insights derived from speech data, the recommender model learns to suggest personalized music playlists tailored to inferred emotional states. While lacking explicit user listening data, these datasets provide valuable emotional cues for influencing music selection.

4. Assessing user's post emotions.

- **Emotional Expression Examination:**

  Observe users' emotional reactions post-listening to music via live video feeds to enhance emotional experiences.

- **Emotional Pattern Identification:**

  Analyze visual data to detect users with distinctive emotional patterns, refining music suggestions accordingly.

- **Engagement Level Analysis:**

  Assess visual cues to determine user engagement levels, aiding in refining music recommendations for sustained interest.

- **Genre Preference Inference:**

  Analyze visual cues to infer users' genre preferences, enabling personalized music suggestions aligned with emotional responses.

- **Emotional Insight Storage:**

  Store emotional insights from users' experiences for personalized music recommendations tailored to individual preferences and emotional states.

5. Integration.

- **Training the music recommender model:**

  Concurrently, the music recommender model undergoes training using historical user data and music attributes. This process involves learning patterns and preferences to generate accurate music recommendations.

- **Developing the user interfaces of the mobile application accordingly:**

  With the backend components in place, the mobile application is developed to provide a seamless user interface for capturing images, processing facial data, and accessing music recommendations. This entails designing intuitive user interfaces and integrating backend functionalities.

- **Integrating above features into the developed mobile application:**

  Finally, the developed components, including the image capture module, facial data processing pipeline, user profiling, and music recommendation systems, are seamlessly integrated into the mobile application. This integration ensures that users can easily access personalized music recommendations based on their demographic characteristics and preferences, thereby enhancing user experience and engagement.

3 Feasibility studies

3.1 Technical Feasibility:

The technical feasibility of developing a Music Recommendation System using Python and machine learning algorithms is the primary objective of this project. The necessary hardware, including devices for capturing images and audio, should be widely accessible and cost-effective, ensuring practical implementation.

Before initiating development, it is essential to gain proficiency in key technologies such as image and voice recognition, as well as relevant programming languages and libraries like Python, TensorFlow, and PyTorch. This may involve a learning curve, but it is necessary for effectively implementing the system.

Additionally, Python and machine learning libraries like TensorFlow are open source, minimizing development expenses and providing access to extensive resources and community support. The availability of skilled developers and experts in machine learning and Python further strengthens the project's technical feasibility. These factors collectively ensure the successful development and deployment of the Music Recommendation System within various user environments.

3.2 Economic Feasibility:

From an economic standpoint, the Music Recommendation System exhibits significant potential. Leveraging open-source technologies like Python and machine learning algorithms reduces development costs substantially. Operational expenses remain minimal, mainly limited to hardware maintenance. However, the system's economic viability hinges on its ability to generate a return on investment (ROI) for music platforms or streaming services. By enhancing user engagement and satisfaction through personalized music recommendations, the system can potentially drive increased user retention and revenue generation. This improvement in user experience may lead to a positive ROI by attracting more subscribers and increasing overall platform usage.

3.3 Ethical Feasibility:

Ethical considerations are pivotal in determining the viability of our Music Recommendation System. It's imperative to prioritize data privacy and adhere to regulatory standards throughout the system's development. To address these ethical considerations, stringent protocols for data anonymization will be implemented, and legal experts will be consulted to establish transparent consent procedures for users. Upholding ethical principles not only ensures compliance with regulations but also fosters trust among users, promoting responsible and ethical usage of the system.

4 Requirements Gathering

4.1 Requirement Gathering and Analysis

The process of gathering requirements for our Music Recommendation System was thorough and involved several essential steps to ensure effective and ethical achievement of our research goals. Initially, we conducted a detailed review of existing research and analyzed available systems, using various online resources to expand our understanding.

A critical part of our study was identifying and examining systems similar to our proposed system, which provided deeper insights into common techniques and technologies used in their development. Our research methodology evolved through an extensive requirement gathering process and information sources.

We began by establishing a clear understanding of our research objectives through consultations with music industry professionals, technology experts, and researchers in relevant fields. Their valuable insights guided us in defining specific parameters for evaluation, focusing primarily on enhancing user experience through multi model approach to generate a personalized music playlist.

Next, we identified the necessary technology and tools for data acquisition and processing, collaborating closely with experts in machine learning, emotion recognition, and data analytics. Their expertise was instrumental in selecting appropriate algorithms and libraries for tasks such as emotion detection, sentiment analysis, and user profiling.

Ethical considerations were crucial throughout the requirement gathering phase. Ensuring user privacy protection and adhering to ethical guidelines were important as well. Collecting the necessary data for testing and validation involved capturing user interactions within the application. This required transparent communication with users, development of consent forms, and careful selection of recording equipment. By gathering requirements our research methodology was tailored precisely to address our specific objectives. It ensured a comprehensive evaluation of our Music Recommendation System while upholding the highest standards of ethical conduct and research integrity.

4.2 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.

- **Capturing the image:**

The system interface must incorporate a camera feature enabling users to capture a selfie image in real-time. The captured image should be of sufficient quality to facilitate accurate facial recognition and subsequent analysis. Additionally, the interface should provide feedback to ensure users are aware of the image quality requirements.

- **Analyzing the captured image:**

Following image capture, the system must employ image processing techniques to isolate and extract the user's face from the captured image. This step is crucial for subsequent facial attribute analysis.

- **Classifying the age, gender, current weather, current emotion of the user:**

Once the face is extracted, the system should utilize convolutional neural network (CNN), RNN and MFCC to classify personal details like age, gender, current emotion, current weather, and surroundings. This classification process enables the system to generate a tailored music playlist.

- **System should generate a personalized playlist accordingly:**

Leveraging the user profile and other contextual information, such as indoor-outdoor status, emotion, the system should generate a personalized music playlist tailored to the individual user's preferences and mood. This playlist may include recommendations based on music genre, artist preferences, and past listening history.

- **User should be able to listen to playlist from the software application:**

The system interface should provide users with the ability to access and listen to the generated personalized playlist directly within the software application. This functionality enhances user experience by providing seamless access to recommended music content.

- **System should be able to track and train the models for future recommendations:**

To continuously improve recommendation accuracy, the system should incorporate mechanisms for tracking user interactions and feedback. This data can be used to retrain and refine the image classification and recommendation models, ensuring that future recommendations remain relevant and personalized to users' evolving preferences and behavior.

4.3 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:**

  - Response Time: The system should respond promptly to user interactions, with minimal delay between input and output.
  - Loading Time: Application and content loading times should be optimized to provide a seamless user experience.
  - Resource Utilization: The system should efficiently utilize resources such as CPU, memory, and network bandwidth to minimize resource wastage and maximize performance.

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

  - Intuitive Design: The user interface should be intuitive and easy to navigate, allowing users to perform tasks efficiently without unnecessary cognitive load.
  - Visual Appeal: The interface should adhere to modern design standards, with visually appealing layouts, colors, and typography.
  - Accessibility: The application should be accessible to users with disabilities, following accessibility guidelines such as WCAG (Web Content Accessibility Guidelines) to ensure inclusivity.

- **Security and Privacy:**

  - Data Encryption: Sensitive user data should be encrypted during transmission and storage to protect against unauthorized access and data breaches.
  - Authentication and Authorization: The system should implement robust authentication and authorization mechanisms to ensure that only authorized users can access and modify data.
  - Compliance: Developers should adhere to data privacy regulations such as GDPR (General Data Protection Regulation) or CCPA (California Consumer Privacy Act) to safeguard user privacy and comply with legal requirements.

- **Scalability:**

  - Database Scalability: The system architecture should support horizontal scaling of databases to accommodate growing data volumes and user traffic.
  - Server Scalability: Server infrastructure should be designed to scale dynamically based on demand, ensuring consistent performance under varying loads.

- **Reliability and Availability:**

  - System Uptime: The system should strive for high availability, minimizing downtime for maintenance and updates.
  - Fault Tolerance: The system should be resilient to failures and errors, with mechanisms in place to handle exceptions gracefully and prevent service disruptions.
  - Disaster Recovery: Procedures and mechanisms should be in place to recover data and restore service in the event of a catastrophic failure or disaster.

4.4 System Requirements

The system requirements for implementing the proposed solution include:

- **Laptop/Desktop:**

A laptop or desktop computer serves as the primary hardware platform for developing and deploying the system. It should meet the minimum hardware specifications required to run the necessary software tools and libraries.

- **Cameras and Audio Recorders:**

High-quality cameras and audio recorders are essential for capturing user selfies, surrounding images, and voice inputs. These devices should be capable of producing clear and detailed images and audio recordings, ensuring accurate analysis and processing.

- **Internet Connection:**

A stable internet connection is necessary for accessing online resources, such as cloud-based development platforms, datasets, and APIs. Additionally, internet connectivity enables seamless communication between components of the system and facilitates real-time updates and data synchronization.

- **Mobile Phone:**

A mobile phone serves as the endpoint for delivering personalized music recommendations to the user. It should be compatible with the mobile application developed as part of the system. The mobile phone provides users with access to their personalized playlists and allows them to interact with the recommendation system on the go.

5 Methodology

5.1 Tools and technologies.

- **Programming language - Python**

We selected Python and Google Colaboratory as our primary tools for developing machine learning models due to several compelling reasons. Python is widely recognized for its simplicity, versatility, and extensive ecosystem of libraries, making it an ideal choice for machine learning tasks. By leveraging libraries like NumPy, pandas, TensorFlow, and PyTorch, we could efficiently build, train, and evaluate our models with ease.

- **Google Colaboratory**

Google Colaboratory offered several advantages that aligned perfectly with our project requirements. Its provision of free access to GPU and CPU resources allowed us to accelerate model training without incurring additional costs. Additionally, Colab's seamless integration with Python and built-in support for data preprocessing, analysis, and visualization facilitated collaborative development and streamlined our workflow. The platform's ability to visualize metrics such as loss and accuracy enabled us to evaluate and optimize our models effectively.

- **Kaggle**

We went for Kaggle as our data source platform due to its extensive collection of datasets and educational resources tailored to data science and machine learning projects. Leveraging Kaggle's datasets allowed us to access diverse data sources relevant to our machine learning tasks, ranging from weather prediction to age, gender, and voice emotion recognition. The mentioned datasets were taken from the Kaggle for the use of ML model.

Weather-prediction

- https://www.kaggle.com/datasets/vijaygiitk/multiclass- weather-dataset/

Age prediction

- https://susanqq.github.io/UTKFace/

Gender prediction

- https://susanqq.github.io/UTKFace/

Voice emotion recognition

- https://www.kaggle.com/datasets/ejlok1/toronto- emotional-speech-set-tess
- https://www.kaggle.com/datasets/ejlok1/cremad/data
- https://www.kaggle.com/datasets/uwrfkaggler/ravdess-emotional-speech-audio
- https://www.kaggle.com/datasets/ejlok1/surrey-
- audiovisual-expressed-emotion-savee

- **TensorFlow**

TensorFlow emerged as our preferred machine learning framework due to its robustness, scalability, and comprehensive toolset for building deep learning models. With TensorFlow, we could harness state-of-the-art algorithms and techniques to develop powerful machine learning models tailored to our project requirements.

- **Flutter, Visual Studio Code, and Firebase**

Flutter, Visual Studio Code, and Firebase were chosen for their respective roles in developing our mobile application. Flutter's cross-platform capabilities enabled us to build a single codebase for multiple platforms, ensuring compatibility and consistency across devices. Visual Studio Code provided a feature-rich development environment with seamless integration for writing, debugging, and deploying code. Meanwhile, Firebase offered backend services such as authentication, database management, and cloud storage, facilitating the seamless integration of our machine learning models with our mobile application while ensuring scalability, security, and real-time data synchronization.

- **Swagger**:

Swagger is an open-source software framework that allows developers to design, build, document, and consume RESTful web services. We integrated Swagger into our development process to streamline API documentation and facilitate communication between frontend and backend teams. With Swagger, we could efficiently document and test our APIs, ensuring consistency and reliability across our application ecosystem.

Overall, the combination of Python, Google Colaboratory, Kaggle, TensorFlow, Flutter, Visual Studio Code, and Firebase provided us with a robust and versatile toolkit to effectively develop, deploy, and integrate machine learning models into our mobile application, thereby fulfilling our project requirements with efficiency and efficacy.

5.2 Introduction

"Melowave" represents a pioneering mobile application designed to revolutionize the music listening experience through personalized recommendations driven by machine learning and artificial intelligence. At its core, Melowave aims to generate music playlists tailored to each user's unique characteristics, including age, gender, contextual factors, and the user's current emotional state.

In the initial stages of development, extensive data gathering was important to inform the system's algorithms and ensure its efficacy. This involved a comprehensive evaluation process to identify key factors that influence music preferences and mood. Specifically, the research focused on collecting data to delineate user demographics, age groups, as well as contextual elements such as weather conditions. Therefore, as the initial step, a survey has been conducted to identify details regarding our components. *(Figure 4).*

From these results, the proportion of gender is different *(Figure 4)*. Therefore, it is proven that it's essential to recognize that preferences can vary widely among individuals, and gender is just one of many factors influencing music consumption behavior.

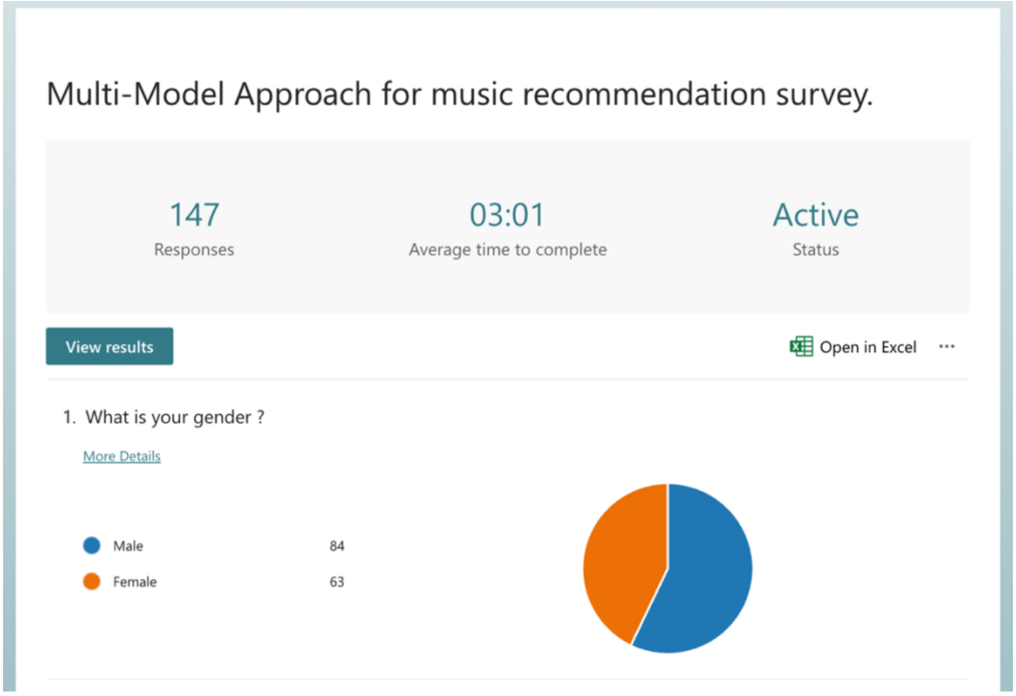


Figure 4 - Survey results of the gender

The diagram above, *(Figure 5).*, reveals how music consumption varies with age groups. Our analysis of the data shows a clear trend: individuals aged 18-24 listen to music more frequently than other age groups. Following closely behind, those aged 25 to 34 also show a significant interest in music. These findings were helpful in developing our machine learning model. By understanding these age-related preferences, we can better categorize users and provide more accurate music recommendations tailored to their tastes.

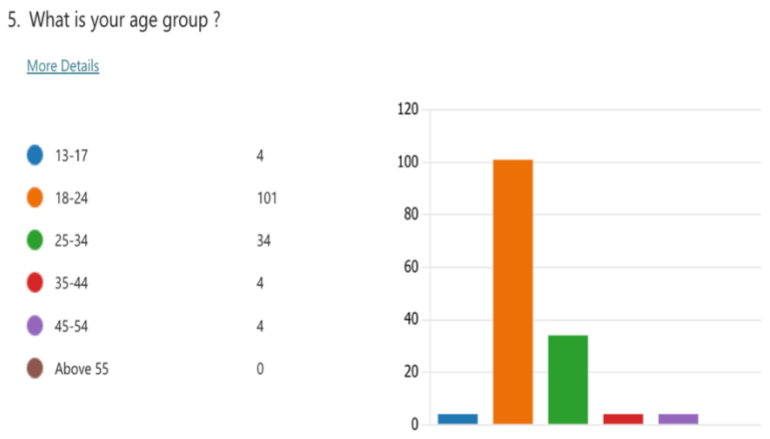


Figure 5 - survey results of age groups

Furthermore, the survey was conducted to investigate the correlation between environmental factors and music consumption patterns among users as well. The results revealed distinct preferences, notably indicating a higher inclination towards music listening during nighttime under cloudy weather conditions. Additionally, the study elucidated a prevalent preference for outdoor settings over indoor environments. These empirical findings played a pivotal role in the development of a sophisticated machine learning model, enabling improved user categorization, and subsequently enhancing the precision of music recommendations.

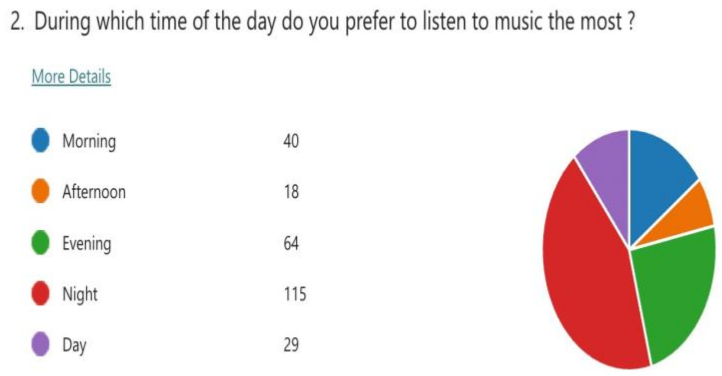


Figure 6 - Proof that the time and surrounding matters to the music consumption.

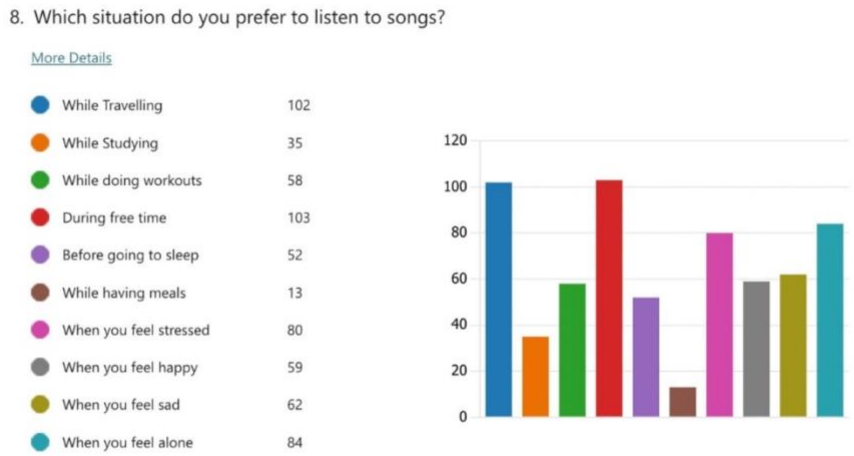


Figure 7 - survey results of impact of Indoor/outdoor

Moreover, it was proven that music has an impact on user's emptions, and it has been proven through the above survey that has been conducted by us. By listening to different types of songs impact on user's emotion and according to a psychological base, music can be used as a remedy to reduce the stress, depression and tiredness of a person.

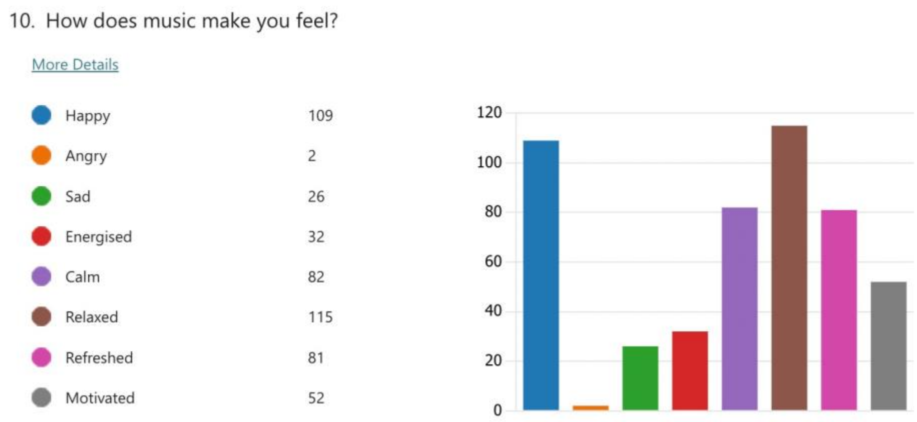


Figure 8 - How music consumption impacts on users' emotions

5.3 Overall System Diagram

In the proposed solution for a multi-model music recommendation system, we have identified four main sub- components. The System overall Diagram is illustrated below *(Figure 9)*.

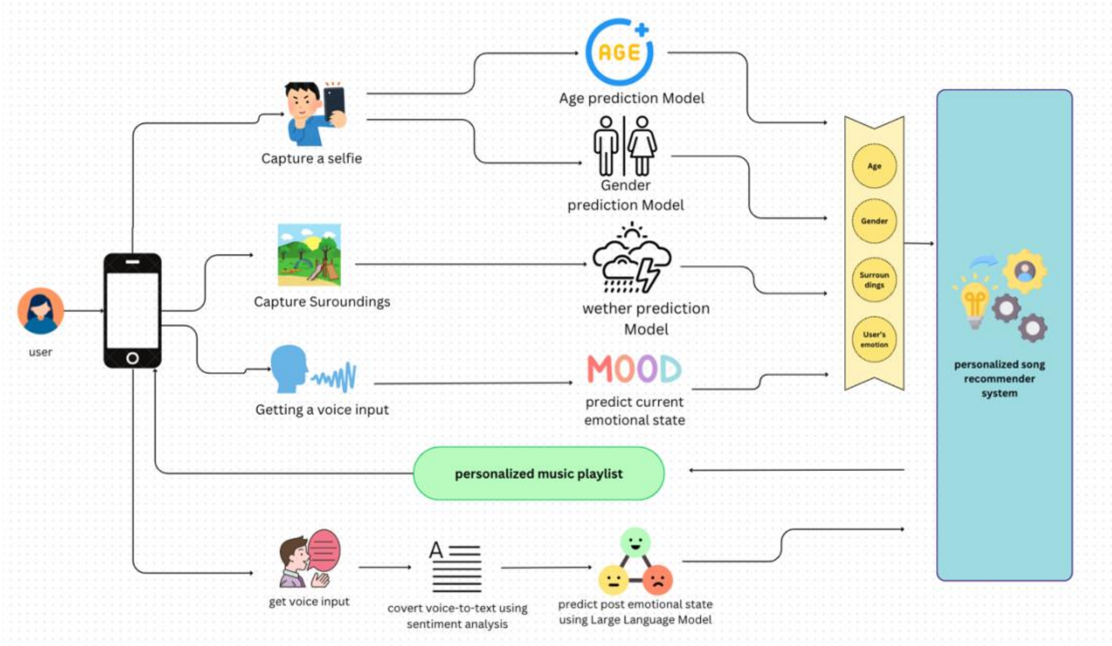


Figure 9 - System Overview Diagram

As depicted in Figure 9, our proposed system comprises four main sub-components designed to achieve optimal functionality. The first component involves retrieving an image of the user for analysis, with a focus on extracting essential user details such as age and gender. This information serves as crucial input for the recommender system, enabling personalized music recommendations tailored to the user's demographic profile.

The second component of our system entails capturing an image of the user's surroundings to identify current weather conditions. This data is then extracted and forwarded to the recommender system as additional input. By incorporating weather information, our system enhances the relevance of music recommendations, aligning them with the prevailing atmospheric conditions and user preferences.

Moving on to the third component, we gather voice input from the user to extract their current emotional state. This emotional data provides valuable insight into the user's mood, allowing the recommender system to curate music selections that resonate with their feelings and emotions at that moment.

Once all the relevant inputs are collected, our music recommendation models spring into action, generating a tailored playlist designed to enhance the user's entertainment experience. By leveraging advanced machine learning algorithms, our system ensures that each recommendation is finely tuned to the user's preferences, demographics, weather conditions, and emotional state.

Furthermore, our system includes an evaluation mechanism to assess the post-listening emotions of the user. This feedback loop enables continuous refinement and improvement of the music recommendation process, ensuring that future recommendations are even more aligned with the user's evolving tastes and emotional responses. Overall, our comprehensive approach aims to deliver a truly personalized and immersive music listening experience for every user.

5.4 User interfaces of the Mobile application

From the user's perspective, the "Melowave" mobile application begins with a login window prompting the user to sign in with their Google account *(Figure 10).*

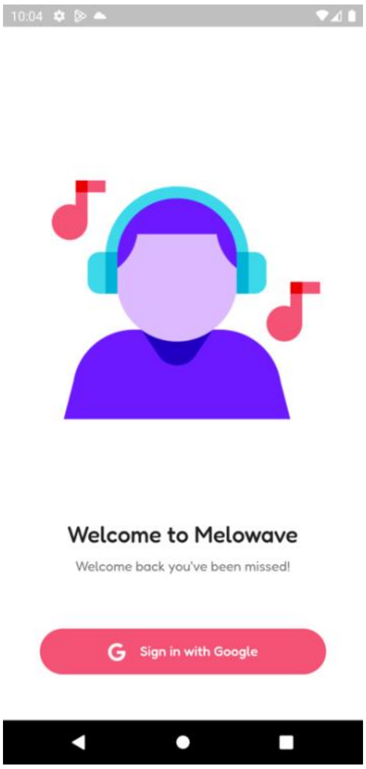


Figure 10 - Sign in UI

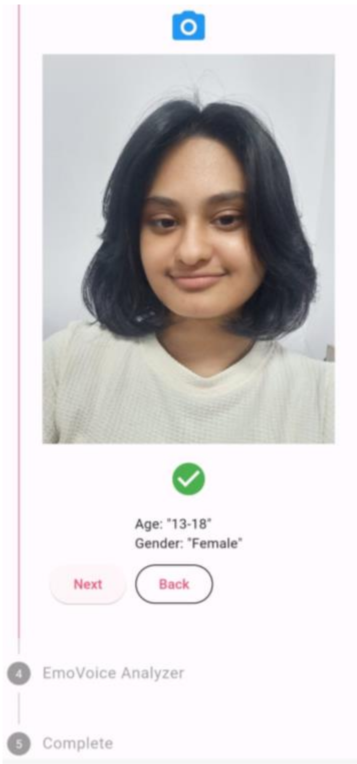


Figure 11 - Image Predictor step in the UI

Once logged in, the application guides the user through a series of steps using a stepper interface. As the first step, a consent will be prompt for the permissions to use the camera, gallery, and the microphone of the user's mobile phone. After accepting the consent, the next step involves taking a selfie using the device's camera, capturing the user's facial features for analysis *(Figure 11).*

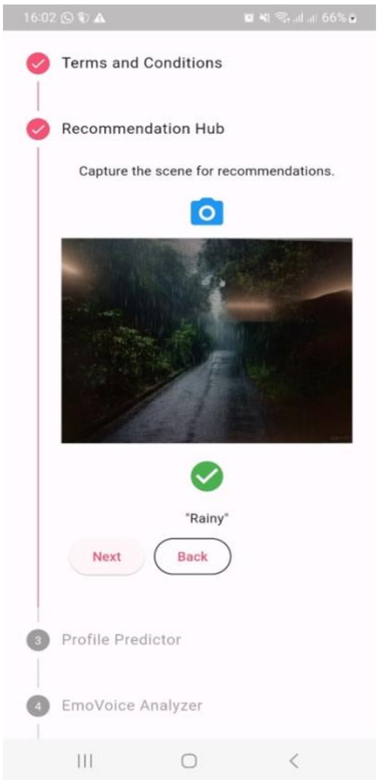


Figure 13 - Surroundings capturing UI

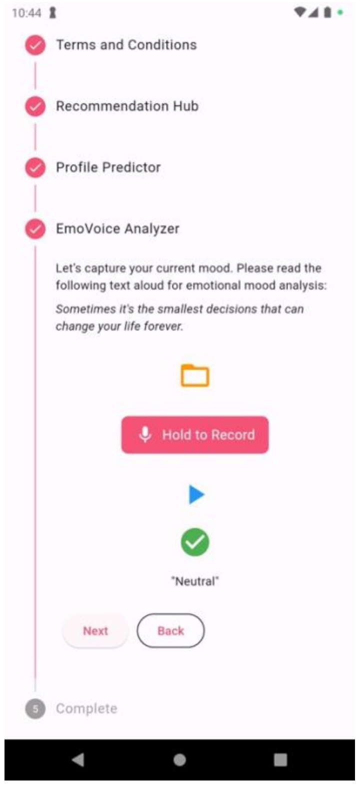


Figure 12 - Getting a voice clip from the user

Moving on to the next step, the application again utilizes the camera to capture the user's surroundings, gathering data on the current environment *(Figure 13)*. Finally, in the third step, the app prompts the user to record a voice clip using the microphone, allowing the system to analyze the user's current emotional state *(Figure 12)*.

After clicking the "next" button, the trained machine learning models are deployed. These models include the age prediction model, gender prediction model, weather prediction model, and emotion prediction model. They analyze the collected data and generate outputs such as the user's age, gender, current weather conditions, and emotional state *(Figure 16)*.

Subsequently, the music recommender model generates a personalized music playlist based on the user's age, gender, current weather, and emotional state *(Figure 15)*. However, if the user is not satisfied with the playlist or if their mood does not improve, the post-emotion classification model intervenes *(Figure 14)*.

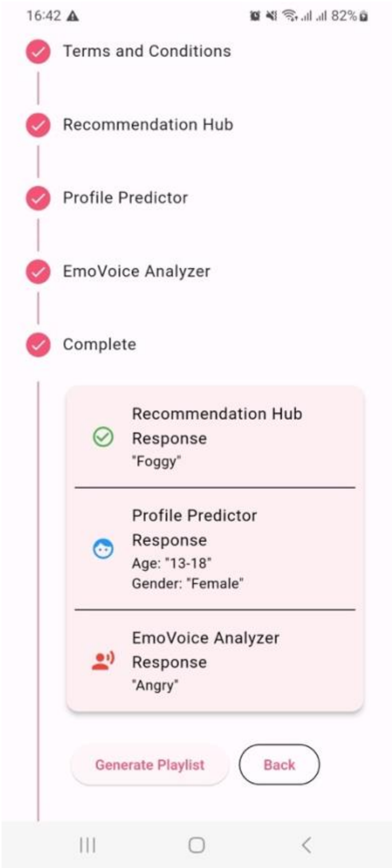


Figure 16 - Summary of the predictions UI

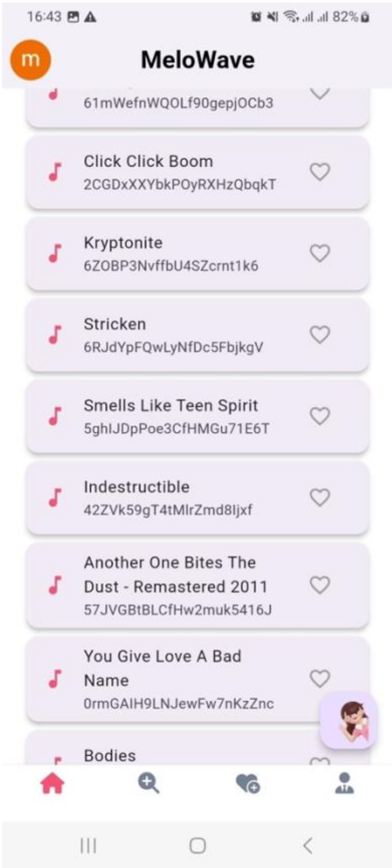


Figure 15 - Songs list UI

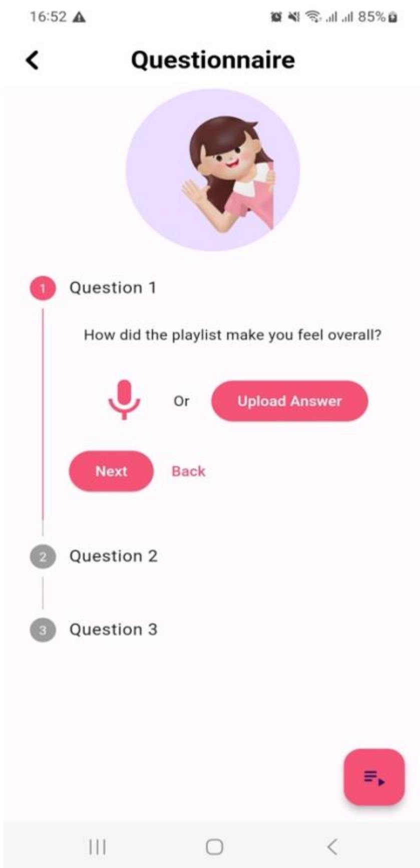


Figure 14 - UI of the questionnaire

5.5 Machine leaning models and accuracy levels.

1) Gender identification through image processing:

As the first component extracting user gender, we have used the CNN algorithm, image processing, deep learning methods, and the pre-trained model – "EfficientNetB3" from the Kaggle to develop the gender prediction machine learning model. After training the model with modifications, I could achieve accuracy level of 0.91. The following images convey the accuracy levels of the Gender Classification model after evaluation *(Figure 17)*.

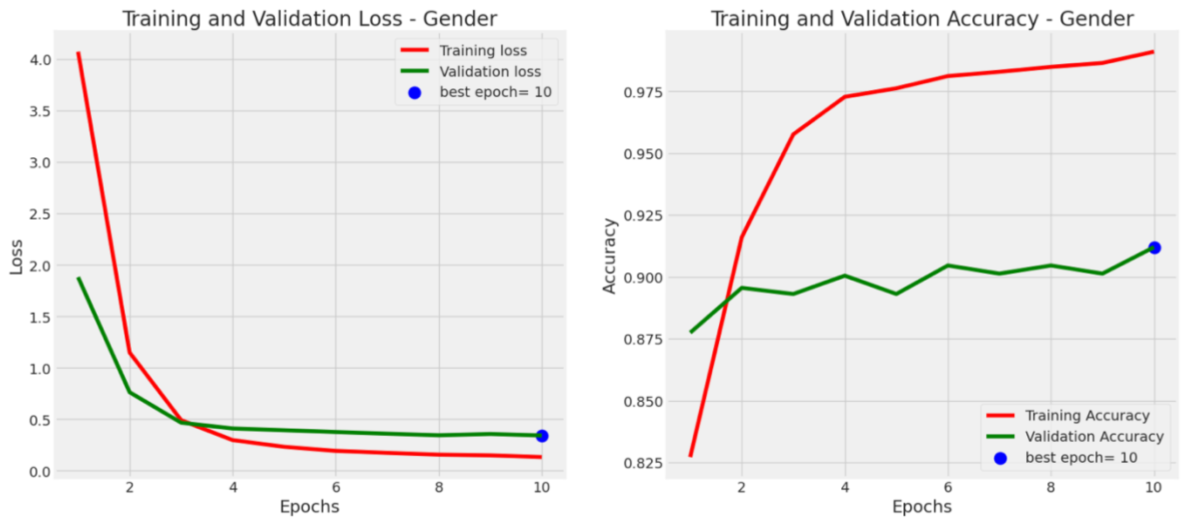


Figure 17 - Accuracy and the loss of the gender prediction model.

2) Age identification through image processing:

Same as the Gender classification model, here we have used the CNN algorithm, image processing, deep learning methods, and the pre-trained model – "EfficientNetB3" from Kaggle. The Confusion matrix of the Age prediction model is as follows *(Figure 18)*.

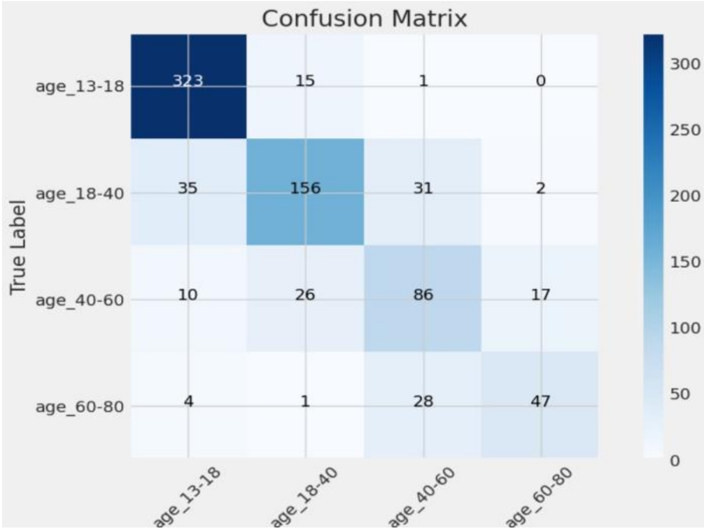


Figure 18 - Confusion Metrix of the age prediction model.

3) Surrounding identification through image processing:

we have used the CNN algorithm, and weather conditions and image classification using VGG19 to identify features has added value to the system image processing, deep learning methods, and the pre-trained model – “VGG19” from the Kaggle to develop the weather prediction machine learning model. After training the model with modifications, achieved accuracy level of 0.88 (Figure 19).

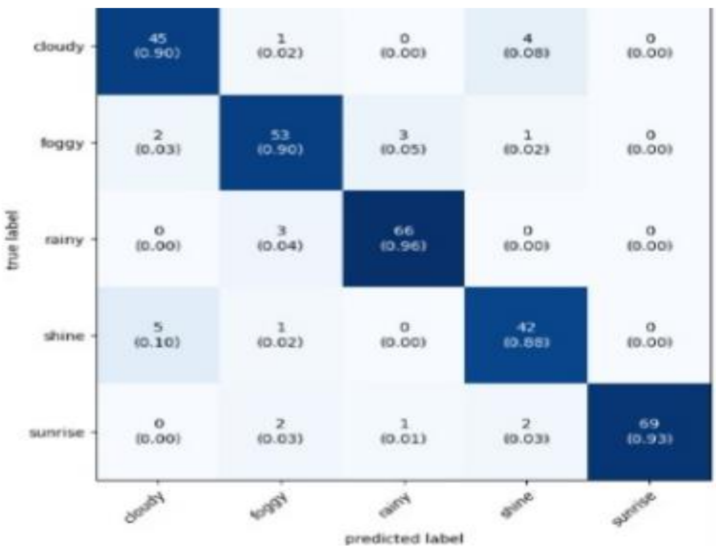


Figure 19 - Confusion Metrix of weather prediction model

4) Emotion Prediction model.

In this part, the focus is predicting the user's present emotion. The technique was tested with 50 people, and 45 of them made commendably accurate predictions. Furthermore, the model's prediction based on frequency produced good results. Thus, despite the dataset being in English, it was also excellent at distinguishing emotions in Sinhala. The below diagram shows training and validation accuracies (Figure 20).

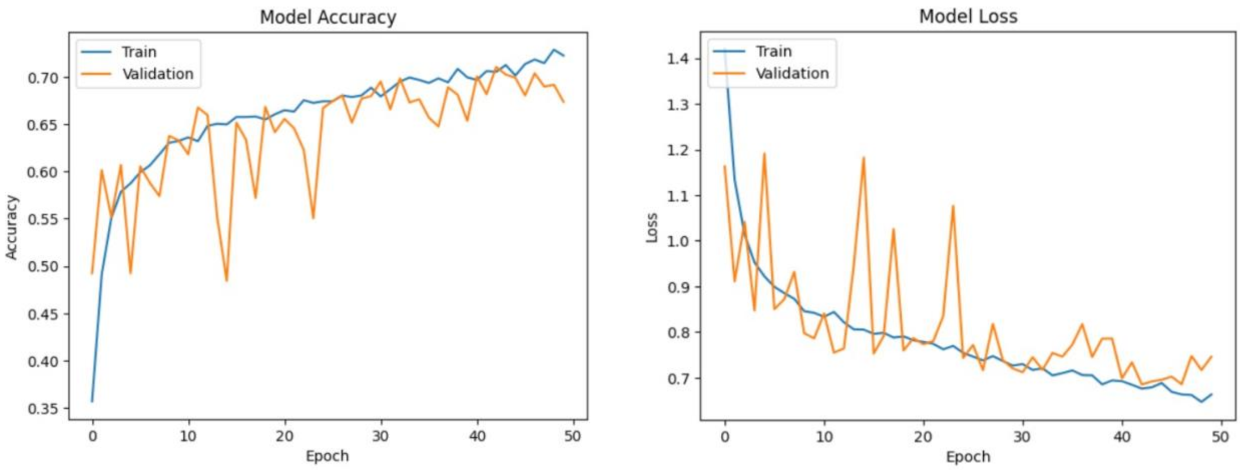


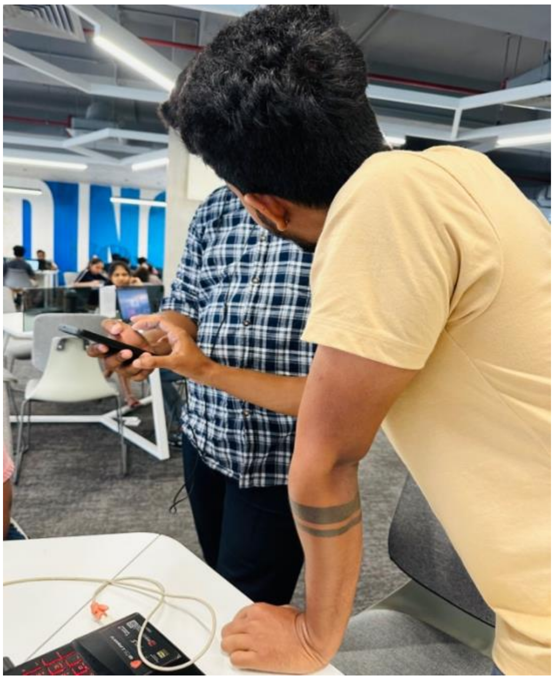
Figure 20 - Emotion prediction model accuracy and the loss

5.6 Field visits and Feedbacks

Field visits

After implementing the mobile application and after integrating, as always user feedbacks were needed to test and to make changes. Considering field visits and feedbacks, it plays a crucial role in ensuring the success of our mobile application. By conducting field visits, we can directly engage with our target users and observe their behaviors, preferences, and pain points in real-world settings. This firsthand experience allows us to gain valuable insights into how our application is being used and how it can be improved to better meet user needs. Additionally, collecting feedback from users enables us to gather direct input on their experiences, likes, dislikes, and suggestions for enhancements. Incorporating this feedback into the development process enables us to iteratively refine and optimize our mobile application, ensuring that it remains relevant, useful, and user-friendly. Ultimately, field visits and feedback serve as essential tools for understanding user requirements, validating design decisions, and ultimately delivering a mobile application that truly resonates with its intended audience.

We could identify many mistakes and points to improve during our field visit *(Figure 15).*



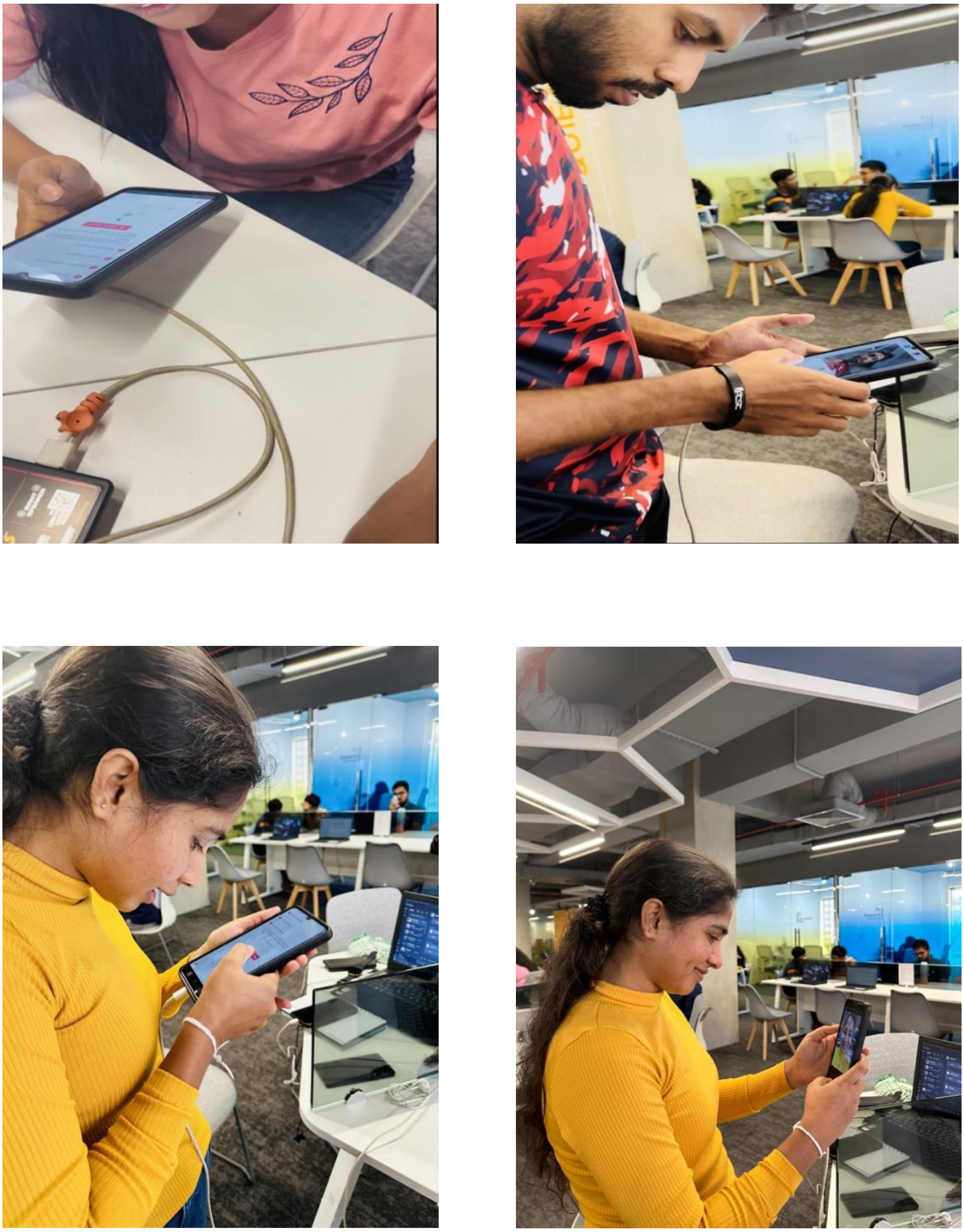
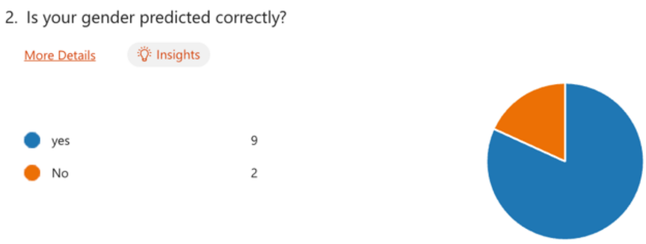


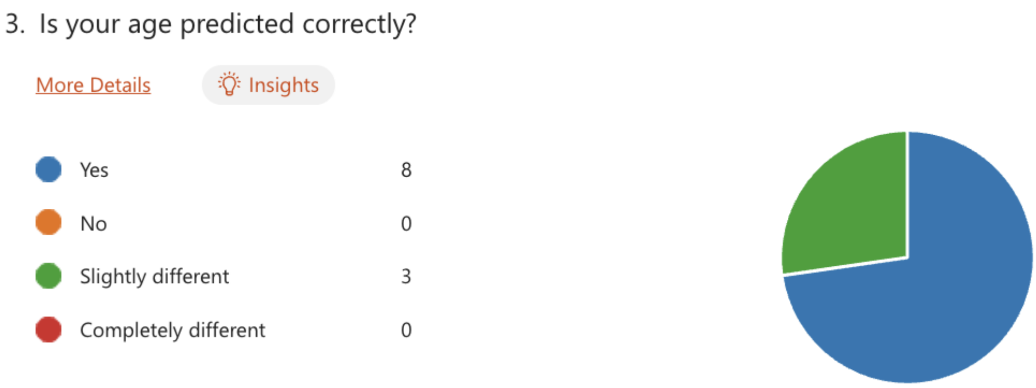
Figure 21 - Field visits and testing

Feedbacks

Feedback is indispensable for the success of our mobile application as it provides essential insights into user experiences, preferences, and areas for improvement. By actively soliciting and responding to user feedback, we can enhance the overall user experience, identify, and address bugs and technical issues promptly, validate features and functionality, tailor music recommendations to user preferences, and foster user engagement and loyalty. Utilizing feedback as a guiding force enables us to continuously refine and improve our application, driving user satisfaction and contributing to its long-term success in the market.

Here are some feedbacks we receive from the users *(Figure 16)*.





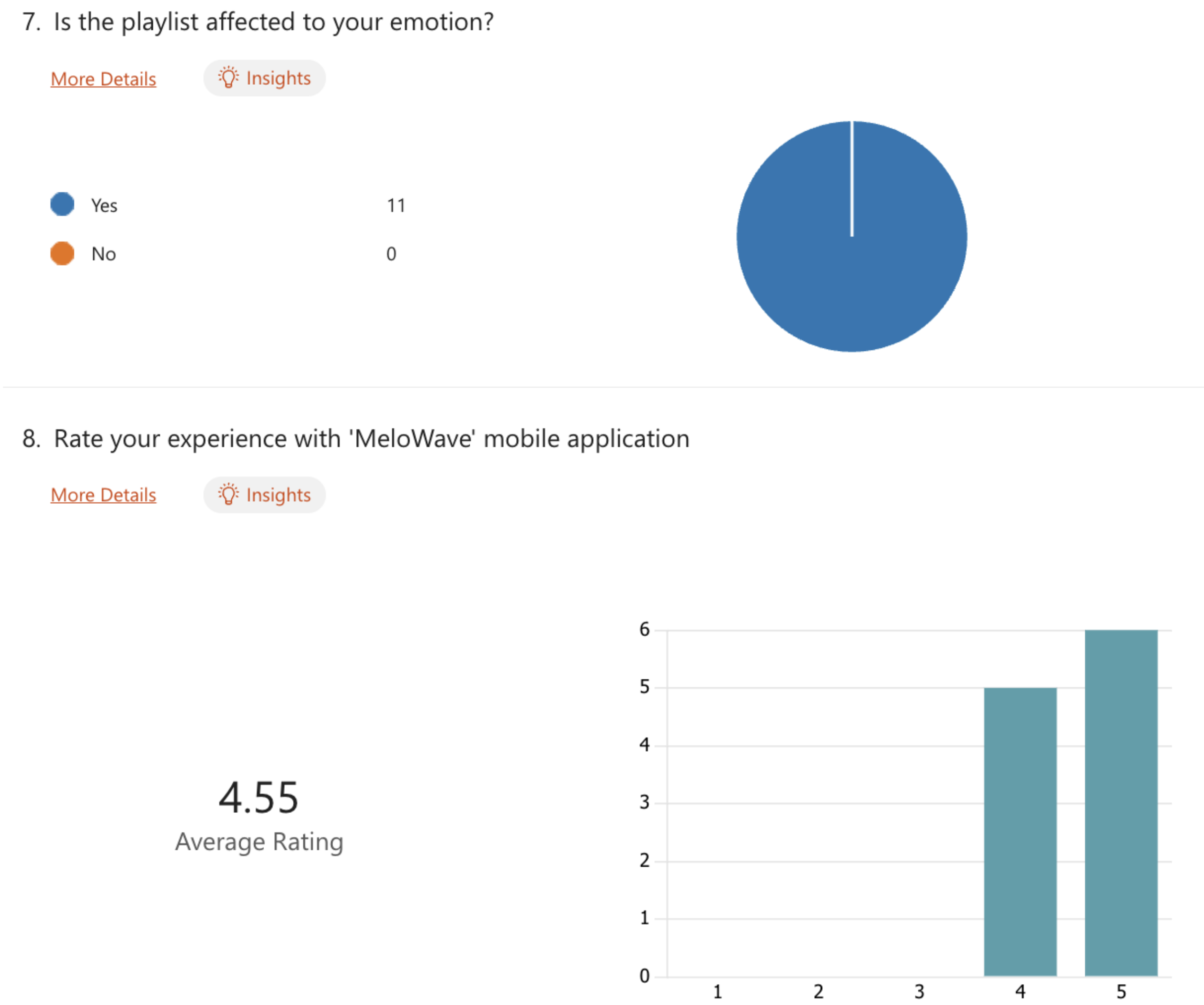


Figure 22 - Feedbacks from the users

5.7 Gantt Chart

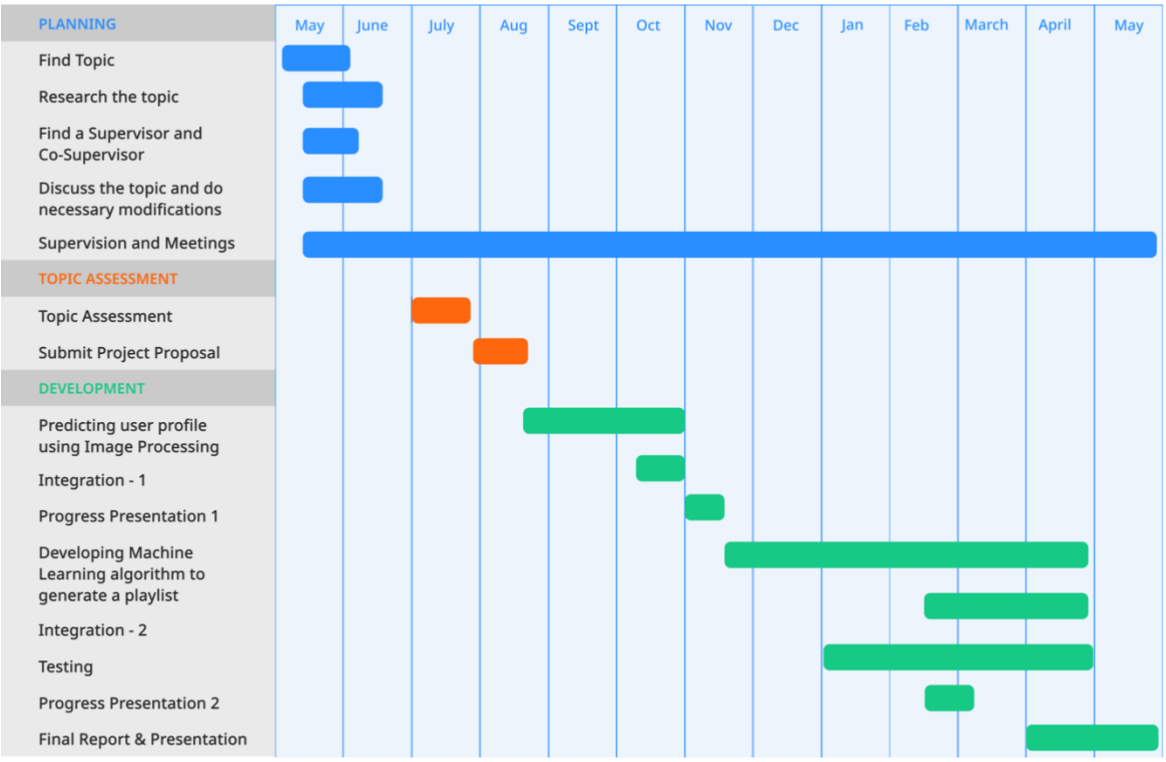
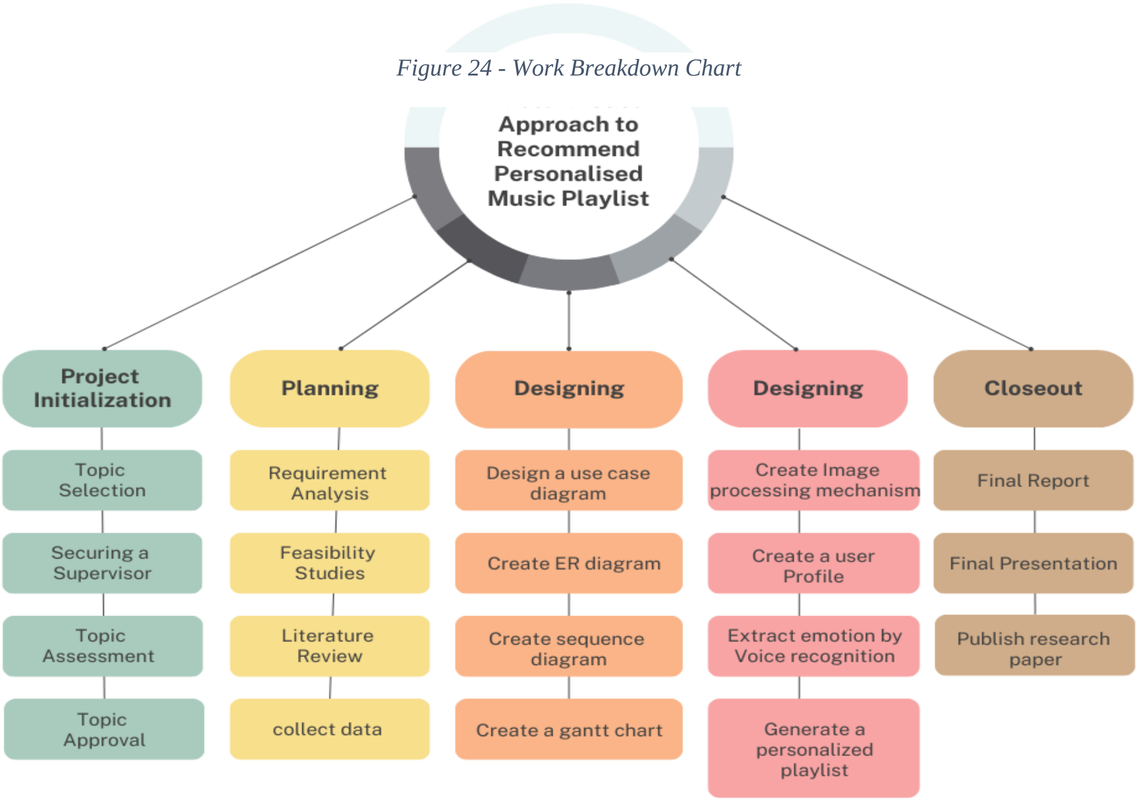


Figure 23 - Gantt Chart

5.8 Work Breakdown Chart

Figure 24 - Work Breakdown Chart



6 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.

7 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.

7.1 Test cases

Age and gender prediction component.

Table 2 - Test case 01 - Selfie capture and analysis

Test Case ID	01
Test Case	Selfie Capture and Analysis
Test scenario	Ensure the camera functionality properly captures a selfie image of the user.
Input	A selfie
Expected output	Capturing an image through given camera functionality
Actual output	Successfully captured an image
Status (Pass / Fail)	PASS

Table 3 - Test case 02 - Selfie capture and analysis

Test Case ID	02
Test Case	Selfie Capture and Analysis
Test scenario	Validate that the captured image is of sufficient quality for facial analysis.
Input	A selfie from a low-quality phone as well as from a high-quality phone
Expected output	Predictions are accurate in both pictures
Actual output	Low quality photo was not predicted properly. Prediction of the photo from the quality camera is more accurate.
Status (Pass / Fail)	FAIL

Table 4 - Test case 03 - Selfie capture and analysis

Test Case ID	03
Test Case	Selfie Capture and Analysis
Test scenario	Verify that the system accurately analyzes facial details such as age and gender.
Input	A selfie
Expected output	Predicting correct age and gender
Actual output	Predicted correct age and gender
Status (Pass / Fail)	PASS

Emotion detection through voice analysis component

Table 1- Test case 01 – verify the expected response time

Test Case ID	01
Test Case	Verify the response time for given voice clip
Test scenario	Ensure the output's response time
Input	Voice clip
Expected output	Within 30 seconds, the output will be given
Actual output	Within 30 seconds, the output will be given
Status (Pass / Fail)	PASS

Table 5 - Test case 02 – analyze the voice clip's frequency.

Test Case ID	02
Test Case	Give voice clip and Analysis
Test scenario	Ensure the given voice clip's frequency matches with the emotion.
Input	A voice clip
Expected output	Successfully identified current emotion
Actual output	Successfully identified current emotion
Status (Pass / Fail)	PASS

Table 4 - Test case 04 – give sad based emotion voice clip.

Test Case ID	03
Test Case	Give voice clip based on sad state
Test scenario	Validate that given voice clip for feature extraction
Input	Sad state-based voice clip
Expected output	Predict voice clip as a sad state
Actual output	Predict voice clip as a neutral state
Status (Pass / Fail)	FAIL

8 Budget and Commercialization

Considering the widespread use of music players in people's daily lives, this project has significant potential for commercial success. Individuals are willing to invest in an improved music player experience, indicating substantial commercial value. However, with established market leaders like Spotify, iTunes, and Deezer already in place, it's crucial to devise a competitive and fair pricing strategy for the music player. While popular subscription models like those of Spotify, Apple Music, and Deezer typically charge around $10 per month, some users find this pricing too high for the perceived value. Hence, a different subscription model is proposed to drive the commercialization of this mobile app.

Table 6 - 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 7 - 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

9 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]. 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.

[5]. 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.

[6] L. Jin, D. Yuan and H. Zhang, "Music recommendation based on embedding model with user preference and context," 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA), Beijing, China, 2017, pp. 688-692, doi: 10.1109/ICBDA.2017.8078723.

[7] Johnson, M., White, E. (2019). "Analyzing Image Texture for Scene Classification." Proceedings of the IEEE International Conference on Computer Vision, 789-802. DOI:

[8] .) J. G. B. Vitório and C. N. Silla, "Music Recommendation System for Shared Environments," 2023 30th International Conference on Systems, Signals and Image Processing

(IWSSIP), Ohrid, North Macedonia, 2023, pp. 1-5, doi: 10.1109/IWSSIP58668.2023.10180270.

[9]. 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.

[10] K. R. Nambiar and S. Palaniswamy, "Speech Emotion Based Music Recommendation," 2022 3rd International Conference for Emerging Technology (INCET), Belgaum, India, 2022, pp. 1-6, doi: 10.1109/INCET54531.2022.9824457.

10 Appendix

10.1 Plagiarism Report