Personalized Music Recommendation Application

TMP – 2023 – 24 - 065

Project Proposal Report

Dhananjaya W.K.S – IT20667078

Supervised by – Mr. Thusithanjana Thilakarathne

B.Sc. (Hons) Degree in Information Technology

Department of Information Technology

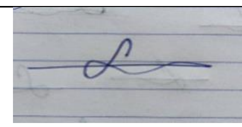
Sri Lanka Institute of Information Technology

Sri Lanka

August 2023

Declaration

I hereby declare that the work presented in this proposal is entirely my own and has been conducted under my initiative and supervision. This proposal does not incorporate, without proper acknowledgement, any material that has been previously submitted for a degree or diploma in 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 acknowledgement 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
Dhananjaya W.K.S	IT20667078	

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

Name of supervisor: Mr. Thusithanjana Thilakarathne
Name of co-supervisor: Mr. Darshana Kasthurirathna

…………………………. ……08/25/2023……….

Signature of the supervisor: ￼ Date

(Mr. Thusithanjana Thilakarathne)

………………………. ……08/25/2023………

Signature of the co-supervisor: ￼ Date

(Mr. Darshana Kasthurirathna)

Abstract

Music, which is sometimes touted as a global language, connects deeply with people by combining emotional resonance, cognitive stimulation, and social cohesiveness. The variety of music genres and their distinctive compositional features, which include pace, rhythm, melody, harmony, and lyrics, have the astonishing power to evoke a range of emotions. In-depth user-specific emotional analysis is the main emphasis of this research, which explores the development of a personalized music recommendation system based on users' emotional states.

The main goal of this project is to create a creative system for making music recommendations that is sensitive to listeners' emotional preferences. The main goal of the system is to understand a user's emotional state prior to involvement, and then analyze that user's emotional response to the playlist that was created by the system after the contact. The precise identification of a user's emotional state and the capturing of subtle facial expressions in video footage provide a key difficulty. This complex system includes not just the analysis of user voice patterns but also continuous emotional journey tracking.

This research enhances our knowledge of how personalized music suggestions may be used to produce immersive and emotionally resonant experiences by addressing the complex interplay between music, emotions, and user involvement. The achievement of this goal has enormous promise, both for the realm of music technology and for improving how people engage with and react to technology on a deeply emotional level.

Keywords: facial emotion recognition, deep convolutional neural network, music recommendation application

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

1. INTRODUCTION

1.1 Background Study

Applications for recommending music today rely on collaborative filtering methods, content filtering, and related algorithms. These techniques help give personalized music suggestions, but it is still difficult to make sure that the recommendations match a person's emotional condition, which calls for the creation of advanced music recommendation systems.

When a person interacts with a music recommendation system for the first time, crucial details like their prior preferences, genre tendencies, and emotional listening patterns are frequently unknown. The need of using cutting-edge technology like computer vision and voice frequency analysis to identify user behaviors and track their emotional responses is highlighted by the lack of historical data. These technologies are crucial to the development of an emotion-based music recommendation system since they make it easier to comprehend customer preferences outside of music alone.

It is interesting that the user's emotional state might change from being neutral to decreasing or moderating after engaging with the machine-generated recommended playlist. The effect of the suggested playlist on the user's emotional state following their listening experience must thus be determined. This necessitates a scientific approach that investigates how the suggested playlist affects the user's mental health, hence assisting in the continuing improvement of the effectiveness of the recommendation system.

In conclusion, a comprehensive emotion-based music recommendation system is built on a foundation of content filtering, collaborative filtering, computer vision, voice frequency analysis, and post-listening emotional evaluation.

The complex interaction between technology improvements and user-centric emotional responses brings up new opportunities for enhancing and individualized music recommendation methods as this discipline continues to develop.

Literature Survey

Recent years have seen a growth in creative research projects focused at improving user experiences across multiple areas as a result of the confluence of technology and human emotion. Emotion-based music recommendation systems have been developed in the area of music consumption in an effort to create emotionally resonant connections between consumers and music. To provide consumers personalized music choices that are in line with their emotional states, these systems take advantage of developments in areas like machine learning, audio analysis, and affective computing. Such systems are justified by the realization that music is more than just aural stimulation; it has the capacity to provoke, reflect, and intensify human emotions.

This section explores the literature that has already been written about emotion-based music recommendation systems. We intend to highlight the present status of the area, major approaches used, obstacles faced, and viable paths for further inquiry by synthesizing the collective knowledge and insights from earlier research. The literature review aims to illuminate the numerous aspects of emotion-based music recommendation, highlighting both the advancements made and the areas that still require more study.

The survey starts off by going through the fundamental studies on music emotion detection methods. These methods, which include sentiment analysis, physiological signal interpretation, and audio analysis, form the basis for deducing consumers' emotional states from their listening to music. The study then digs into research on studies that combine music emotion detection with recommendation algorithms, emphasizing the interaction between technological improvements and user-centric emotional involvement.

The survey also aims to discuss the subtleties of personalization in emotion-based recommendation systems. It investigates how to customize recommendations to a user's unique emotional tendencies and how to use contextual aspects to make recommendations more applicable in a variety of situations.

Ethical considerations and user-centered evaluations are becoming more important as the landscape of emotion-based music selection continues to change. As a result, the study looks at literature that discusses the moral ramifications of using user feedback to improve recommendation systems and highlights the value of using personal data for emotional analysis.

In essence, this literature survey serves as a compass, guiding us through the diverse terrain of emotion-based music recommendation systems. By comprehensively assessing prior research, we endeavor to pinpoint the gaps that beckon exploration and the avenues that hold promise for advancing the field's efficacy and user impact.

Using deep learning is one method of detecting user emotions when they use the recommendation system. Using audio signals, they do future extraction methods to find user emotions were proposed by V. Mounika, G. B. V. K. UDAY, Y. Charitha and A. Pravin **[1]**. They extracted emotions in audio signals by using CNN (Convolutional Neural Networks), MFCC and other tools.

[2] Wu, Y., Gu, R., Yang, & Luo investigated the relationship between heart rate and emotion. 32 pupils were employed in the experiment. The affective video snippets are linked to the individuals' emotional experiences. The category-matching ratios of amused and neutral video clips were greater than those of furious and terrified video clips, according to Bonferroni's correction.

Studying and understanding facial expressions has been a long-standing problem. Wisal Hashim Abdulsalam and Mohammed Najm Abdullah proposed video-based emotion detection. **[2]** They trained it using the ADFES-BIV dataset after which they employed it. For the two phases of the proposed system—training and testing—frames were taken from the input video at a rate of 13 frames per second. The dataset was unsuitable for DCNNs since there were not many samples in it. We examined two techniques, one for the training phase and the other for the testing phase, to address this issue. We employed DCNN, which needs a lot of training data because more data increases the accuracy of the inferences. The Warsaw Set of Emotional Facial Expression Pictures (WSEFEP), which contains 210 images of facial expressions, was then downloaded during the testing phase. To test whether their technique works properly, they were all combined with the frames taken from the ADFES-BIV dataset's "Practice" subdirectory. (WSEFEP), which contains 210 images of face expressions, was used together with other operations after extracting the frames from the input video.

C. -C. Hsiao, W. -D. Zheng, R. -G. Lee and R. Lin **[3]**. They suggested playing four different types of games to elicit the feelings of pleasure, happiness, fear, and anger. and after that, the heart rate signals associated with the identified emotions are extracted, and these are used to compute the normalized features. The Artificial Neural Network (ANN) for machine learning is then employed as the input and output for seven normalized features and determined emotions. The trained ANN model can then be used to categorize the subjects' gaming-related emotions.

1.2 Research Gap

A fascinating and less-traveled route in the field of music recommendations involves utilizing technology to monitor listeners' reactions to computer-generated playlists. This field, which is still in its infancy, provides a novel viewpoint on how both sound and sight can influence our emotions.

As technology develops, pairing what we hear with what we see could significantly enhance our ability to recommend music that is in tune with how people are feeling. The idea of utilizing computer vision to identify how people's emotions change after listening to music is new, even though much study has gone into suggesting music based on its content or what others like.

How exactly a playlist created by a machine affects and elevates a person's mood is a key question in this situation. We refer to how music might change or improve someone's mood as "improvement." What exact elements of the suggested music, such as the songs, lyrics, or style, are in charge of these emotional shifts is another intriguing subject. We also want to know if the impact occurs right away or if it takes time.

In essence, this research void is like a brand-new experience where emotions, music, and technology all collide. It's about figuring out how technology may offer suggestions that appeal to our emotions and aid us in understanding. Investigating this area might change the way we recommend music, giving music lovers a more comprehensive and emotionally rewarding experience.

1.3 Research Problem

Using computer vision to more accurately identify how individuals feel after listening to music is a significant problem in the field of music recommendation systems. The research issue we're examining is how to employ computer vision images to improve the precision and efficacy of music recommendations. This entails utilizing visual cues to better align the music with audience emotions.

Many articles have been written about the use of technology in music recommendation systems. Systems for recommending music use a variety of techniques, including collaborative filtering, content-based filtering, and hybrid filtering. They employ spoken audio recognition, image processing, and video processing techniques to identify the user's sentiment before making a recommendation. The finest uses of computer vision technology for what occurs to a user after listening to the suggested playlist or songs, however, have not been well studied.

Existing music recommendation algorithms, however, are not focused on identifying the user's post-emotional states. How can we connect what we see in computer-generated images with how individuals feel after listening to music, if we take a closer look at this issue first? Here, we should work to improve the users' ability to post-emotionally detect themselves.

Additionally, the system should be able to identify which songs on computer-generated music playlists can amplify listeners' emotions. How can we quantify and enhance these components to better propose music that reflects emotions?

2. OBJECTIVE

2.1 Main Objective

After the user has listened to the machine-generated playlist, the system's primary goal is to employ computer vision technology to determine his or her emotional state. This section aims to investigate methods to alter users' moods depending on playlists created automatically.

Video feeds from cameras can be analyzed with the aid of computer vision technology, which can swiftly and precisely determine viewers' moods. They can determine what happens to a user's emotional state after they recommend something by analyzing post-user emotional state.

The overall objective of this section is to develop and put into practice an integrated framework that makes use of computer vision technology to precisely capture and interpret users' facial expressions. This will allow for the identification of their post-listening emotional states and the investigation of the possibility that music recommendations may have a positive effect on users' emotional health and wellbeing. Create methods for selecting playlists that correspond to consumers' emotional inclinations, possibly promoting their long-term emotional wellbeing. analyses the precise musical qualities and traits included in playlists created automatically that help to improve consumers' emotional moods. Examine these characteristics to find trends that may be used to enhance recommendation systems.

2.2 Specific Objective

The following specific sub objectives will be achieved in advance to achieve the above main objective.

- **Examine Users' Post-Listening Emotional Expressions:** Assessing users' post-listening emotional reactions is the first sub-goal. It is necessary to record and inspect live video feeds from the source to look for any unusual or distinctive emotional reactions. This could involve emotions like joy, grief, zeal, or relaxation. By monitoring users' emotional reactions in real time, the system can provide timely insights and enable adjustments to enhance their emotional experiences.
- **Analyze the Visual Data and Identify Emotional Patterns:** By examining the visual data, the second sub-goal is to identify people who exhibit remarkable emotional patterns. The system should be able to identify users who have unexpected emotional reactions, including heightened enthusiasm or significant mood fluctuations. Computer vision methods like pattern recognition and emotion detection can be applied in this situation. By recognizing these emotional patterns, the algorithm may better tailor suggestions to users' emotional preferences.

- **Analyze the Visual Data and Determine Engagement Levels:**
  The third sub-goal comprises assessing the visual data to assess user engagement and interaction levels. This entails identifying situations in which users demonstrate higher degrees of interest, either through gestures, attitudes, or bodily reactions. By identifying active users, The system may learn more about the effectiveness of suggested music in evoking emotional reactions and maintaining user attention.

- **Analyze the Visual Data and Identify Genre Preferences:**
  The fourth sub-goal is to analyze the visual data to determine user preferences for musical genres. By watching users' facial expressions and body language while they are listening to various genres, the algorithm may be able to infer which genres are more likely to make them feel a specific way. This data may be used to provide tailored suggestions based on the users' chosen genres and emotional cues.

- **Store Emotional Insights for Personalized Recommendations:** Keeping track of the emotional insights learned from users' post-listening experiences is the last sub-goal. This involves gathering data on consumers' emotional propensities, levels of engagement, and favored genres. The system may eventually produce more specialized and emotionally impactful personalized music suggestions by keeping track of this data.

3. METHODOLOGY

Methodology:

1. **Specify the Purpose and Needs:** The precise emotional signals to be captured as well as the necessary gear and software for visual-emotion analysis are defined in this stage.
2. **Purchase and install the required hardware and software:** Select the appropriate cameras and place them in a strategic location to record consumers' emotional responses. Establish the technical foundation required for processing and analyzing the visual input in an effective manner.
3. **Gather and annotate emotional data:** Compile videos of the emotional responses—both positive and negative—that listeners experience. To make machine learning model training easier, classify the video data with emotions.
4. **Train and verify models for emotion recognition:** To train models using tagged emotional data, use relevant machine learning techniques like CNN, RCNN, or ANN. Validate the trained models using a specific set of labelled data to ensure their accuracy in detecting users' emotional reactions.
5. **Models should be integrated into the recommendation system:** As part of the music recommendation system, employ trained emotion recognition models to analyze user visual emotional signals in real-time. This connection will allow the algorithm to recognize emotional changes after listening to suggested playlists.
6. **Test and refine the system:** To ensure that the system can detect odd behavior, it must be put to the test in the actual world. Any issues that surface during testing will be resolved, and the system will be adjusted, as necessary.
7. **Deploy and Maintain the System:** Implement the improved system into the framework for music recommendations. To maintain the system's operation over time, be sure to do routine maintenance, upgrades, and monitoring. To ensure reliable and efficient functioning, conduct regular reviews.

Additionally, I intended to investigate the relationship between heart rate and user post-emotions. When the optimal technique is used, the methodology choice will be made with certainty.
The following technique may be used to explain the proposed personalized music recommendation system that uses video analysis and machine learning algorithms to identify user emotions.

(Technologies used only for cut-shape identification are mentioned here)

Integrated development environment (IDE)

PyCharm or Anaconda

Back end:

Image processing

Python - handle algorithms

OpenCV framework

Algorithms

1. CNN
2. RCNN
3. ANN

(Selection of the algorithms will be finalized when implementing based on the best approach)

Database:

MongoDB and Firebase

The process comprises collecting, preparing, and training data, as well as testing and evaluating the model, deploying the system, taking the appropriate action, and constantly improving the system's performance. The system's capacity to recognize human emotions, the accuracy and reliability of the machine learning algorithms, and the quantity and quality of the data used to train the model are all crucial factors in its success.

3.1 System Overview Diagram.

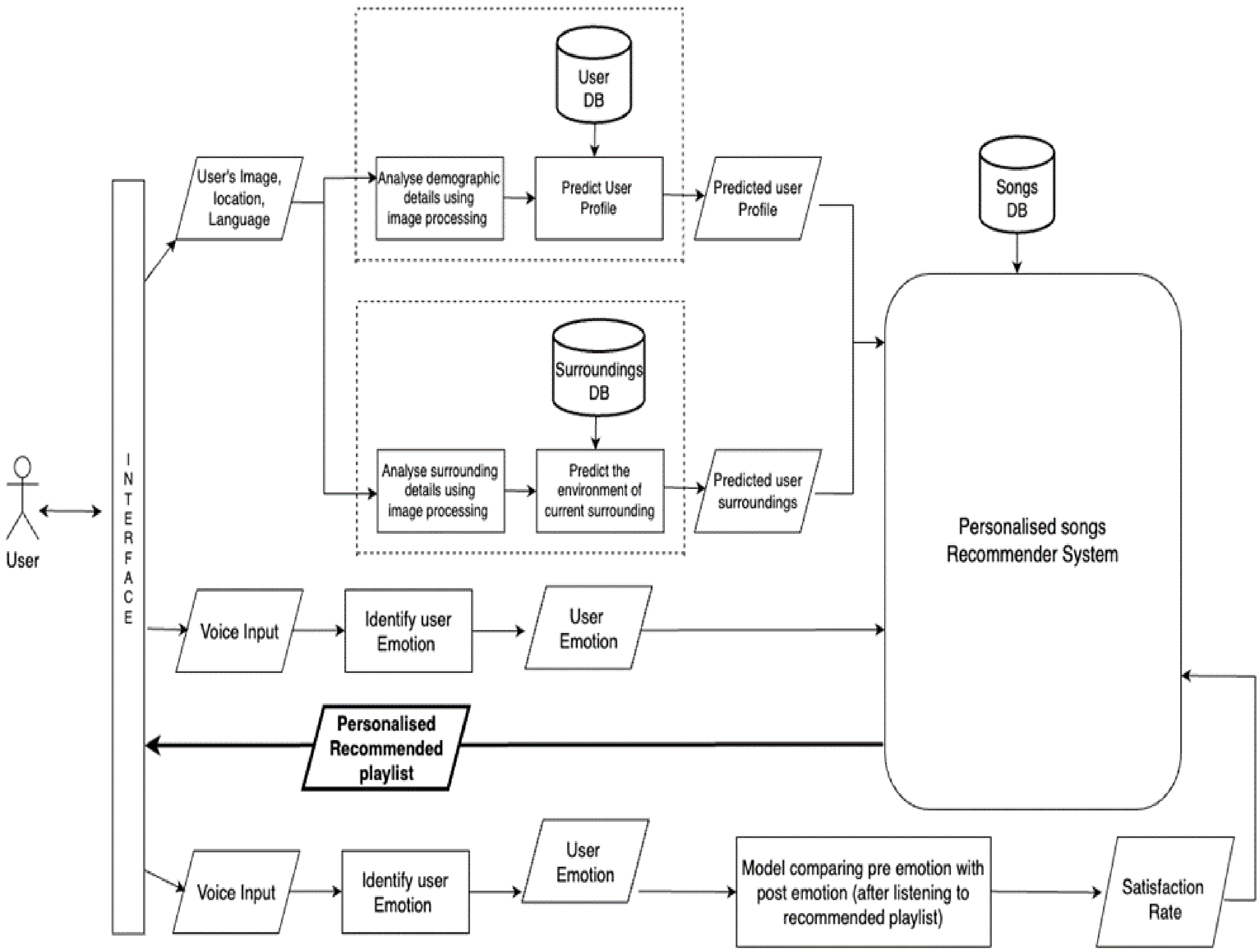


Figure 3.1 System overview diagram

3.2 Component Overview Diagram.

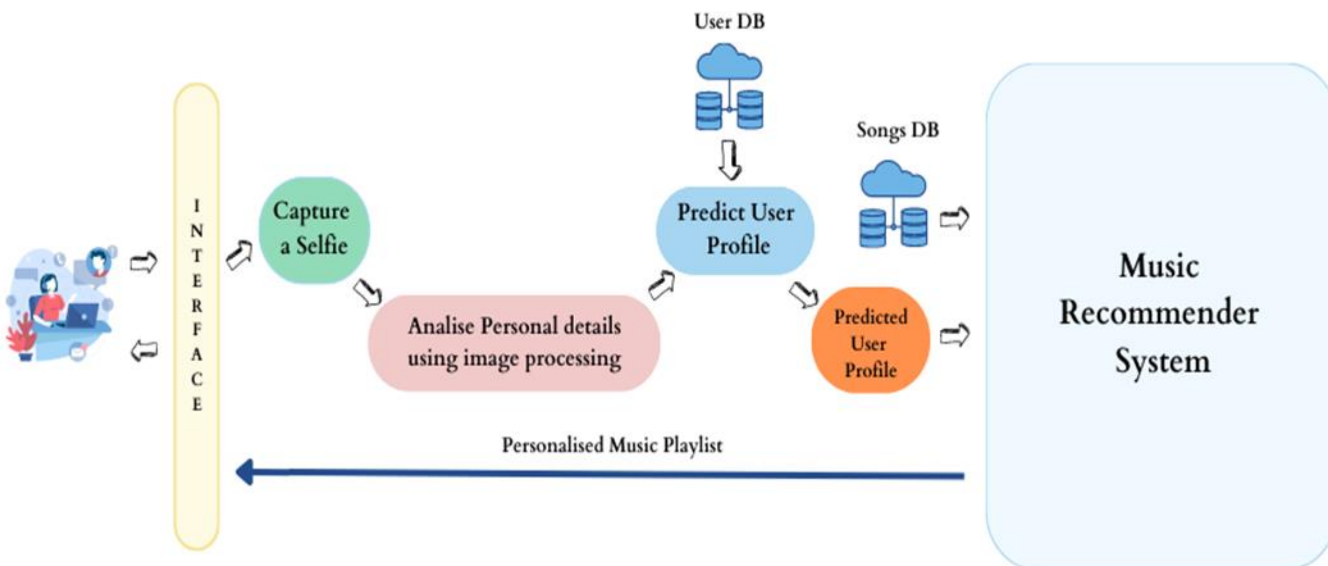


Figure 3.2 Component Overview Diagram

3.3 Gantt Chart.

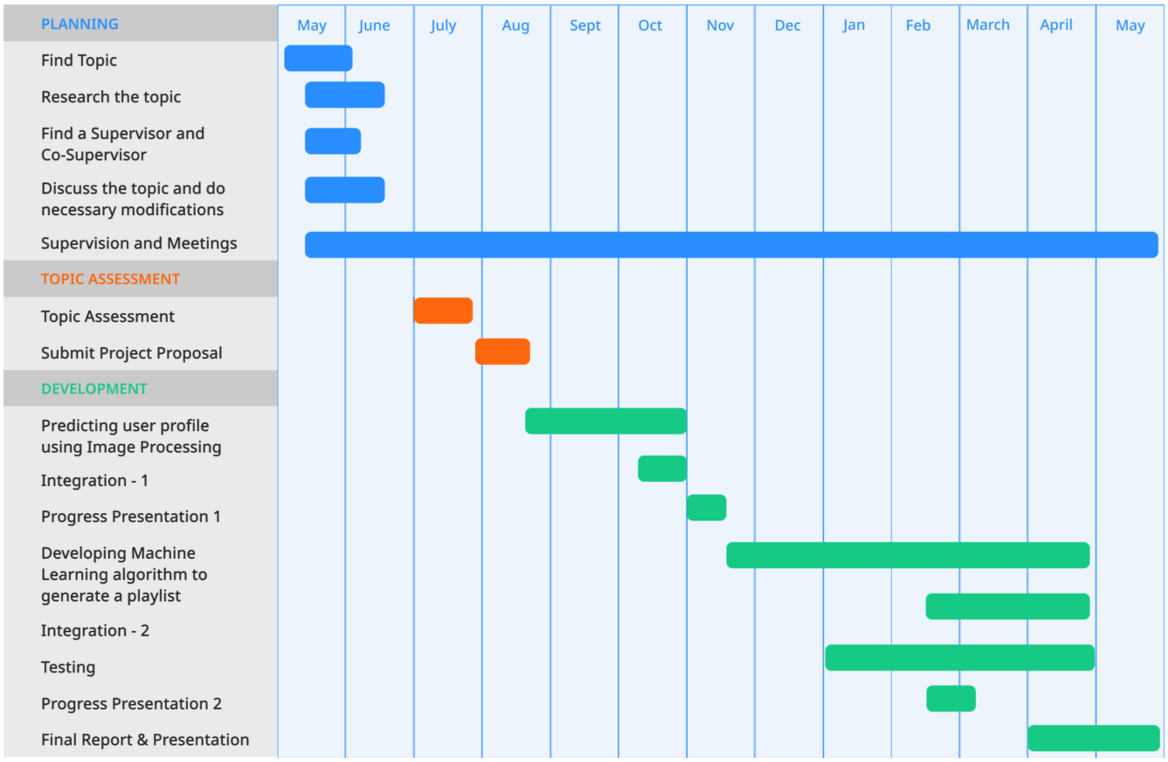


Figure 3.3 Gantt chart

3.4 Work Breakdown Chart

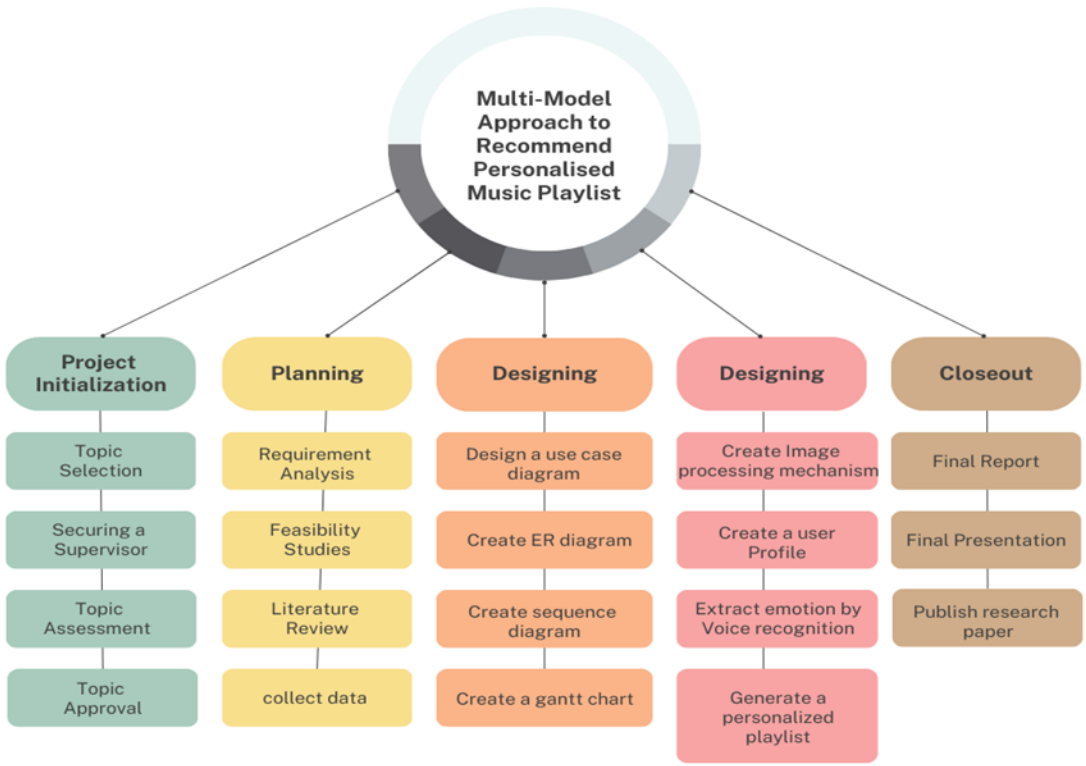


Figure 3.1 Work breakdown chart.

4. PROJECT REQUIREMENTS

The following requirements will be focused only related to the component of post emotion detection.

4.1 Functional Requirements

- Capturing and analyzing video using cameras.
- Ability to detect and identify user's emotional states.
- Analyze the voice recognition model.
- Feature extraction and classification recognition.
- Predict the user's mood after listening to a playlist.
- Find their connection between heart rate and emotions using smartwatch data.

4.2 Non-Functional Requirements

- Because they specify how a software program functions and performs rather than merely its functional capabilities, non-functional requirements are a crucial component of software development. These requirements concentrate on features that improve the entire user experience and system effectiveness, including performance, security, usability, scalability, and other attributes. These non-functional needs are essential to the product's success.
- **Performance**

This has major significance. The system's overall effectiveness, reactivity, and speed are its key points of emphasis. To provide a seamless user experience, developers should prioritize response speed, loading time, and minimal and efficient resource utilization.

- **User Interface and User Experience (UI / UX)**
- When we develop a mobile application for commercialization, the app should have a user-friendly interface that adheres to contemporary design standards and is visually appealing. The user should be able to easily navigate. We also need to consider how accessible things are for those with special needs.
- **Security and Privacy**
- In this time period, this is the most crucial aspect. The most recent technology should be used to encrypt data and for authentication and authorization. Rules and laws pertaining to data privacy should also be known by developers.
- **Scalability**

We 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 a user needs it. The system should have minimum downtime for maintenance and updates.

4.3 System Requirements

- Laptop/desktop
- Cameras and audio recorders
- Internet connection
- Smartwatch

5. TEST PLAN

1. Optimum Music's mobile application will be tested at several points throughout the project. Finding the faults in each component makes it simple to address them individually as opposed to fixing the entire product. Consequently, the entire testing process will include a number of phases and procedures.

2. Unit Testing

   any component, such as the model for categorizing face images and the model for music recommendations, will undergo separate unit testing. In this way, any bugs in any component will be found and addressed as necessary. Here, the researchers' attention will be divided between two primary areas:

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

   The above testing methods will be important for the overall usability of the final product.

3. 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.

4. Final Testing

Final testing will be carried out to ensure that the system is operating faultlessly. Test cases and sample data will be used to evaluate the final output. The mobile application will be delivered to a few chosen beta users as the second stage of the final testing, and their input will be recorded.

The beta users will also evaluate the mobile application's user experience, and the researchers will adjust the user interface to improve the experience for the final user.

6. BUDGET AND BUDGET JUSTIFICATION

People use music players on a regular basis; thus, this product might be quite profitable. For a better music player, people are prepared to spend a reasonable sum of money. The cost of the music player must be reasonable and competitive given the presence of some industry leaders like Spotify, Apple Music, and Deezer. Nearly all the players (Spotify, Apple Music, and Deezer) cost around $10 per month. The following subscription strategy is suggested for the commercialization of this mobile app since some individuals believe the above approach to be pricey and not worth it.

The primary locations where this software will be sold are the Google Play Store, Apple AppStore, and Huawei software Gallery. For this smartphone, there will be two variants,

	Free version	**Paid version (<$10/month)**

Advertisements	Yes Advertisement networks such as Google AdSense /Ad mob will run on this version of the mobile application.	No No advertisements will be displayed in the mobile app
Monthly charges for the users.	No Revenue will be generated from the advertisements shown 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. The younger generation – Social media advertisements (Facebook/Instagram)
2. Mature generation – Newspapers etc.
3. Musical experts – Face-to-face demonstrations with musical associations/groups.

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

Description	Amount (USD)
1. AWS Cloud database (S3) for facial images • To store user images collected through the mobile app.	0.023 per GB / Month
2. AWS Cloud database (EFS) for user demographic data. • To store the 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.	

Table 6 the budget allocation for the research including the cloud charges and AWS cost.

REFERENCES

[1] M. {. R. a. M. D. Levine, "A Deep Learning Method Based on Two-Stage CNN Framework for Recognition of Chinese Reservoirs with Sentinel-2 Images," vol. 117, pp. 1436-1452, 2013.

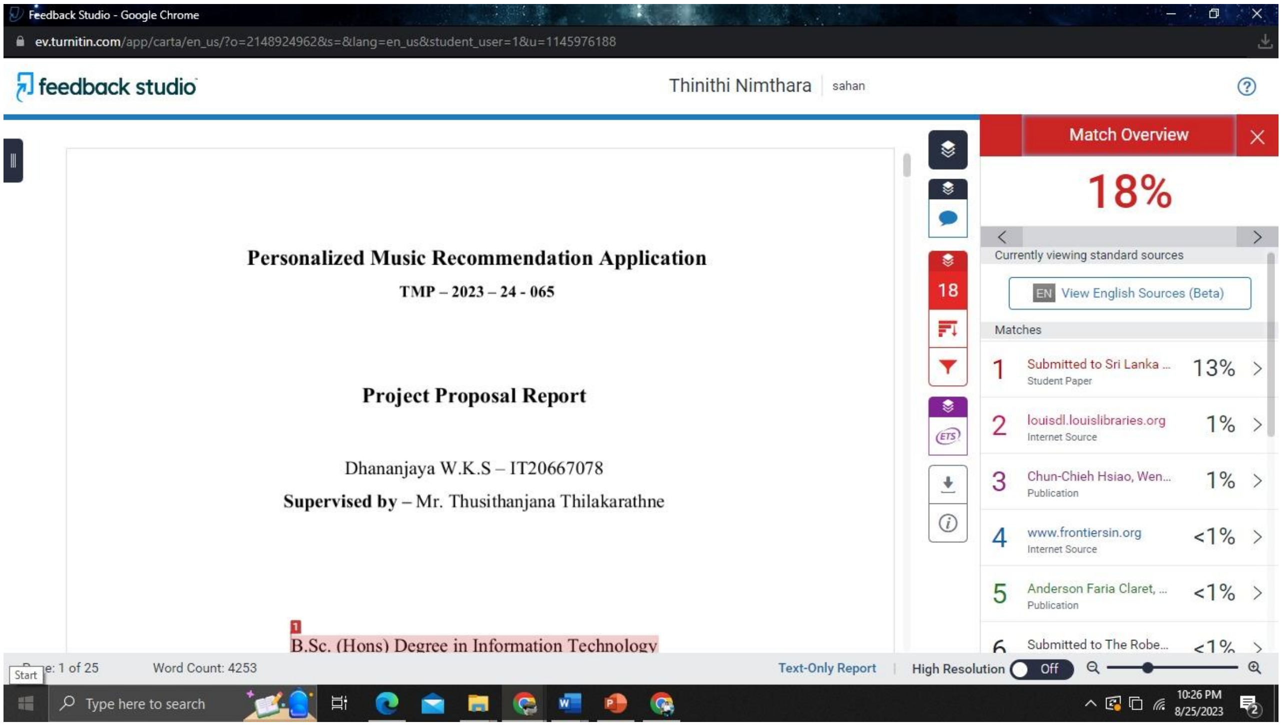
[2] D.-G. a. S. H.-I. a. P. S.-K. a. L. S.-W. Lee, "Motion Influence Map for Unusual Human Activity Detection and Localization in Crowded Scenes," ***IEEE Transactions on Circuits and Systems for Video Technology,*** **vol. 25, pp. 1612-1623, 2015.**

[3] V. a. L. W. a. B. V. a. V. N. Mahadevan, "Anomaly detection in crowded scenes," in ***2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*****, 2010, pp. 1975-1981.**

[4] X. Z. &. J. C. Shaoci Xie, "Video crowd detection and abnormal behavior model detection based on machine learning method," vol. 31, no. 1, pp. 175-184, 01 January 2019.

[5] D.-Y. C. a. P.-C. Huang, "Motion-based unusual event detection in human crowds," ***Journal of Visual Communication and Image Representation,*** **vol. 22, pp. 178-186, 2011.**

APPENDIX



Appendix - A Turnitin report