Emotion Detection Sentiment Analysis: Audio Sentiment Analysis

By

Saw Hui Lin

FACULTY OF COMPUTING AND
INFORMATION TECHNOLOGY

TUNKU ABDUL RAHMAN UNIVERSITY OF
MANAGEMENT AND TECHNOLOGY
KUALA LUMPUR

ACADEMIC YEAR
2024/25

Emotion Detection Sentiment Analysis: Audio Sentiment Analysis

By

Saw Hui Lin

Supervisor: Miss Anurehka A/P Magheswaran

A project report submitted to the
Faculty of Computing and Information Technology
in partial fulfillment of the requirement for the
Bachelor of Software Engineering (Honours)

Department of Software Engineering and Technology
Faculty of Computing and Information Technology
Tunku Abdul Rahman University of Management and Technology
Kuala Lumpur

Declaration

The project submitted herewith is a result of my own efforts in totality and in every aspect of the project works. All information that has been obtained from other sources had been fully acknowledged. I understand that any plagiarism, cheating or collusion or any sorts constitutes a breach of TAR University rules and regulations and would be subjected to disciplinary actions.

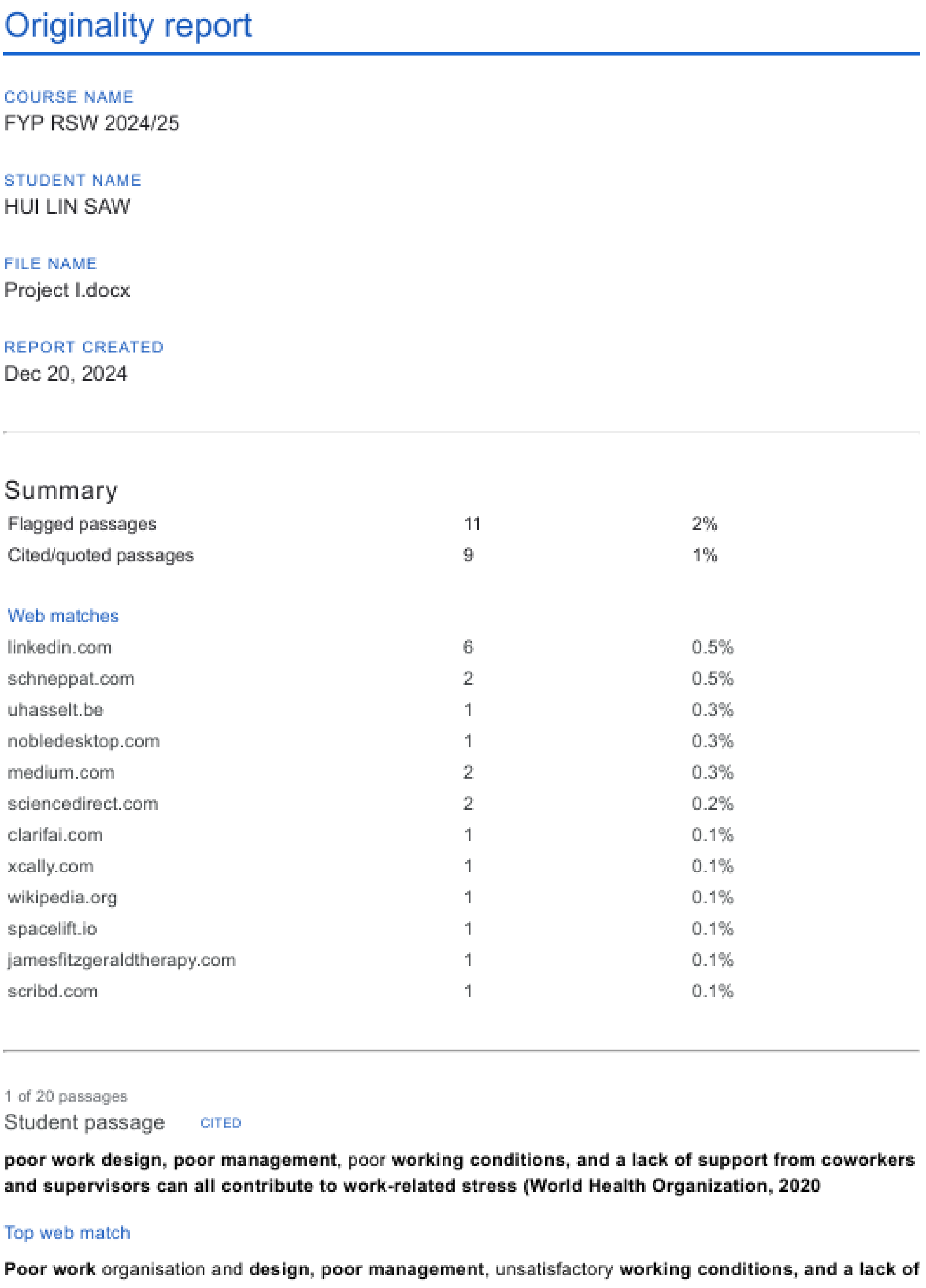
Hui

Saw Hui Lin

Bachelor of Software Engineering (Honours)

ID: 22WMR13626

Originality Report



Originality report

COURSE NAME
FYP RSW 2024/25

STUDENT NAME
HUI LIN SAW

FILE NAME
Project I.docx

REPORT CREATED
Dec 20, 2024

Summary

Flagged passages	11	2%
Cited/quoted passages	9	1%

Web matches

linkedin.com	6	0.5%
schneppat.com	2	0.5%
uhasselt.be	1	0.3%
nobledesktop.com	1	0.3%
medium.com	2	0.3%
sciencedirect.com	2	0.2%
clarifai.com	1	0.1%
xcally.com	1	0.1%
wikipedia.org	1	0.1%
spacelift.io	1	0.1%
jamesfitzgeraldtherapy.com	1	0.1%
scribd.com	1	0.1%

1 of 20 passages
Student passage CITED

poor work design, poor management, poor **working conditions, and a lack of support from coworkers and supervisors can all contribute to work-related stress (World Health Organization, 2020**

Top web match

Poor work organisation and **design, poor management**, unsatisfactory **working conditions, and a lack of**

Abstract

Emotions play a pivotal role in mental health, making effective management critical. Unaddressed emotional distress can escalate into serious mental health issues, a growing concern in both workplace and home environments. This project leverages Natural Language Processing (NLP) and audio processing to develop a comprehensive audio sentiment analysis system that identifies, monitors, and addresses emotional distress. The system includes anonymous feedback channels for employees to safely express concerns, real-time emotion monitoring in virtual meetings, and tools to detect signs of harassment or hostility. By analyzing vocal cues, the system provides real-time insights that allow for timely interventions, reducing emotional distress and promoting a healthier environment. This proactive approach not only supports individuals struggling with emotional challenges but also empowers organizations to foster safer, more supportive spaces.

Table of Contents

Chapter 1

Introduction

1 Introduction

1.1 Objective

a. To study the emotion of the victim

b. To collect the data, process the audio, detect and analyze the emotion

c. To develop an audio sentiment analysis system

1.2 Project Background

The factors that cause an emotional issue are very wide, sometimes it will have a combination of factors (Kandola, 2020). For instance, it could be emotional distress at work (Kandola, 2020), and it will be developed when the employees feel a stressful situation at work (adminhowley, 2022). Some workplaces might be very stressful environments such as the working hours being too long, the salary being lower than expectation, the working conditions being poor and there do not have any overwork control (Kandola, 2020). Employees may feel under strain because of the excessively long work hours and higher demands than they can comfortably handle (Better Health, 2012). Additionally, extended work hours might result in additional health problems like heart disease, stroke, and back discomfort (adminhowley, 2022). A detrimental effect on an employee's physical and emotional well-being as well as the working efficiency, absenteeism, and turnover rates, therefore a suitable working hours is very important (Westover, 2023). Additionally, there are poor working conditions such as bad lighting, outdated technology, and so on, which might result in unmotivated workers, low output, and a higher chance of accidents at work (Malviya, n.d.). When an employee works too hard, too much, or too long, they are exceeding their physical or mental capability, which can have detrimental repercussions on their health, including mental health problems, this is known as overwork (Harris, 2023). Overwork and burnout can result in cynicism, tiredness, boredom, worse work performance, and depression (Harris, 2023). This may set off a downhill spiral in which one becomes increasingly sad and burned out, which lowers one's self-esteem and breeds pessimism, therefore it's difficult to maintain both mental wellness and extended work hours at the same time (Harris, 2023).

Other than that, relationships with colleagues or managers might become one of the factors that affect emotions (Kandola, 2020). The negative relationship will lead to stress issues, we might experience feelings of isolation, loneliness, and overwhelmed if our coworkers don't support us, however well you believe you can compartmentalize, this emotion might nonetheless affect your home life (*MacklinConnection Blog | What Is the Cost of Bad Relationships at Work?*, 2021). Anyone can engage in bullying or harassment at work, and you might not even be aware that you are being victimized, it can be coming from one of the administrative team members, your supervisor, or your coworkers (Watkins, 2020). It can encompass actions like verbal abuse, insults, belittling, humiliation, and exclusion (Gordon, 2022). Increased mental discomfort, sleep disruptions, exhaustion in women and lack of energy in males, anxiety and sadness, adjustment disorders, physical and psychological health problems which include high blood pressure, mood changes, panic attacks, stress and ulcers, and even work-related suicide are examples of the emotional and psychological effects of workplace bullying (Sansone & Sansone, 2015), (Gordon, 2022). At the same time, as there is sexual harassment or a hostile work environment may lead the employees to anxiety and retaliation victims may also be awarded damages for emotional suffering (*Charles Joseph*, 2017). Not only that, but sexual harassment can also cause severe psychological damage, it is common for victims to experience feelings of shame, remorse, anxiety, and melancholy (Legal Specs, 2024). Not only can dealing with a bully be stressful, but it may also be harmful to your mental well-being (Watkins, 2020). A poor work organization, poor work design, poor management, poor working conditions, and a lack of support from coworkers and supervisors can all contribute to work-related stress (World Health Organization, 2020). These factors will lead to a work performance drop, depression or anxiety, and sleep difficulties (Better Health, 2012).

Audio sentiment analysis can provide solutions to various emotional issues by identifying, monitoring, and addressing emotional distress in both workplace and home environments. For example, a feedback channel is provided to establish anonymous audio feedback channels where employees can express their concerns or stressors. This could be a dedicated phone line where employees can record their thoughts and concerns without fear of identification or reprisal. Audio sentiment analysis can help detect common themes and emotional trends, enabling management to address issues more effectively.

Other than that, real-time analysis is applied. Implement audio sentiment analysis in virtual meetings like Zoom, or Google Meet, and calls to monitor employees' emotional states. This can help identify stress, frustration, or dissatisfaction early, allowing management to take proactive measures. It will analyze the audio such as the tone, pitch, and speech rate to identify the emotional states such as happiness, sadness, anger, stress, and many more. These insights can be displayed as real-time feedback for HR or team leaders to observe the employee.

Moreover, monitoring tools which used to detect signs of harassment or hostility in workplace communications. This can help HR teams identify and address such issues more effectively, ensuring a safer work environment.

1.3 Advantages and Contribution

1.3.1 Advantages

The influence of company operations is an advantage of using audio sentiment analysis. Businesses may improve customer happiness and loyalty, maintain the reputation of their brand, and obtain insightful information for future product development by examining consumer feedback since customer pleasure is one of the key performance indicators. Furthermore, audio sentiment analysis helps the HR to know employee satisfaction to improve the workplace condition, and performance management to assist with performance evaluations and offer helpful criticism to create a positive workplace culture. Moreover, it helps in the therapy sessions to enhance patient care in order to deliver more individualized and sympathetic treatment, early detection of conditions like identifying shifts in a patient's emotional state, and help in the early identification of illnesses such as mental health disorders, and also patient satisfaction.

1.3.2 Contributions

The stakeholders will be business and healthcare. For the business aspects, by implementing audio sentiment analysis, HR departments can continuously monitor the emotional states of employees during meetings and communications. This helps in identifying signs of stress, dissatisfaction, or burnout early, allowing for timely interventions. Audio sentiment analysis can provide insights into team dynamics and communication patterns. This information helps HR develop strategies to foster a positive workplace culture, encouraging collaboration and open communication. Moreover, the insights of the analysis can be used to make informed decisions about organizational changes, policy updates, and employee support initiatives. Additionally, audio sentiment analysis can be integrated into therapy sessions to provide therapists with a deeper understanding of patients' emotional states, improving diagnosis and treatment plans. Therefore, by analyzing the data, medical professionals can track emotional trends over time, identifying changes that might indicate progress or areas needing more attention. It also can help in the early detection of mental health issues such as depression or anxiety.

1.4 Project Plan

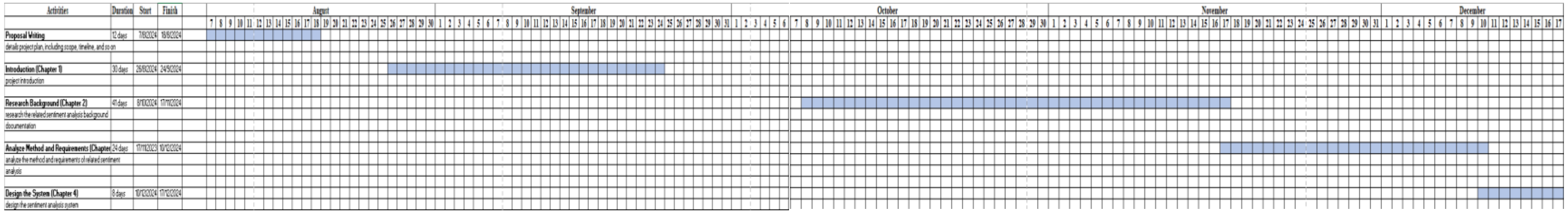
Table 1.1: Project Plan Table

ACTIVITIES	EXPECTED OUTCOME	COMPLETION DATE
Proposal Writing	Detailed project plan, including scope, timeline, resources, and deliverables	18/8/2024
Introduction	Project Introduction (Chapter 1)	24/9/2024
Research Background	Research Background Documentation (Chapter 2)	17/11/2024
Analyze method and requirements	Methodology and Requirements Analysis (Chapter 3)	10/12/2024
Design the System	System Design (Chapter 4)	17/12/2024
System Preview	Source code	24/02/2025
System Testing	Final source code	10/03/2025
Draft FYP Report	Workable Audio Sentiment Analysis System	07/04/2025
Final FYP Report	Completed Audio Sentiment Analysis System	21/04/2025

It is a project plan table that shows the activities, expected outcomes, and also the completion date of each of the activities. In order to ensure that each activity and the project is enough time to complete the task, a planning is required.

Figure 1.1: Project Plan Gantt Chart

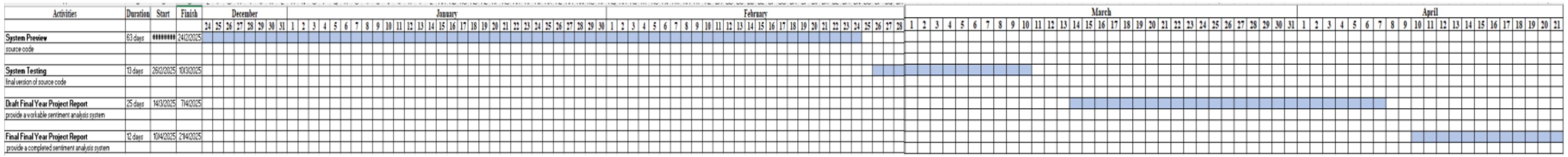
Project I



Activities	Duration	Start	Finish
Proposal Writing	12 days	7/8/2024	18/8/2024
details project plan, including scope, timeline, and so on			
Introduction (Chapter 1)	30 days	26/8/2024	24/9/2024
project introduction			
Research Background (Chapter 2)	41 days	8/10/2024	17/11/2024
research the related sentiment analysis background documentation			
Analyze Method and Requirements (Chapter	24 days	17/11/2023	10/12/2024
analyze the method and requirements of related sentiment analysis			
Design the System (Chapter 4)	8 days	10/12/2024	17/12/2024
design the sentiment analysis system			

The above diagram shows the Project I's planning, it shows the timeline of the project from proposal, Chapter 1 until Chapter 4. It shows the duration, start date and end date of each activity hence it helps to ensure that the project progress is following the planning.

Project II



Activities	Duration	Start	Finish
System Preview	63 days	########	24/2/2025
source code			
System Testing	13 days	26/2/2025	10/3/2025
final version of source code			
Draft Final Year Project Report	25 days	14/3/2025	7/4/2025
provide a workable sentiment analysis system			
Final Final Year Project Report	12 days	10/4/2025	21/4/2025
provide a completed sentiment analysis system			

The above diagram shows the Project II's planning, it show the timeline of the project includes system preview, system testing, draft of final report and also final report of the project. It shows the duration, start date and end date of each activity hence it helps to ensure that the project progress is following the planning.

To see clearer about the Gantt Chart, below is the link:

Gantt Chart.xlsx

1.5 Project Team & Organization

Table 1.2: Project Team & Module

Module	**Sandra Tang Poh Yi**	**Saw Hui Lin**
Data Collection (Procedure for collecting and evaluating data on the relevant variable)	x	x
Text Processing (Text data is analyzed and transformed, using sentiment analysis to find pertinent details, structures, or ideas.)	x	
Audio Processing (Prepare audio data for analysis)		x
Feature Extraction (Extract relevant features from the processed data)	x	x
System Testing (Ensure the system meets the required standards)	X	x

The above table shows the module by each of the team members, it helps to make sure that the module in the system is clearer, each module also follow by a short description to allow others to know about what is the module function.

1.6 Chapter Summary & Evaluation

This project concentrates on building an audio sentiment analysis system that seeks to identify and respond to interpersonal grievances in workplaces and care sectors. The most important goals of the project include the evaluation of emotions in the victims, processing of audio files for emotion recognition, and emotion analysis, and the development of an audio sentiment analysis system. The background depicts the different causes of emotional distress in an organization such as work overload, bad work environment, and conflicts with colleagues. The audio sentiment analysis is to help prevent and recognize emotional states through analysis during virtual meetings by persons who are mentally ill which will improve both workplace conditions and therapy settings. The project mentions the benefits in terms of enhanced customer satisfaction, better employees' state, and an improved patient care, and what it brings to business and healthcare.

Chapter 2

Literature Review

2 Literature Review

2.1 Background & Related Work

A natural language processing (NLP) technique called sentiment analysis is used to determine if the information is optimistic, negative, or objective (*Why Is Sentiment Analysis Important? - Voxco*, n.d.). The goal of audio sentiment analysis is to analyze and accurately deduce the speaker's sentiment from speech signals, and it has attracted major research attention (Audio Sentiment Analysis by Heterogeneous Signal Features Learned from Utterance-Based Parallel Neural Network, n.d.). Humans communicate and interact through a range of emotions, including fear, rage, happiness, sadness and more. As expressed verbally or in writing, these feelings often become hazy. As a result, the use of audio to recognize and evaluate complex emotion in human contact and speech serves as an extra feedback mechanism without changing the original meaning or linguistic content (Naman Dhariwal et al., 2023). However, since it can be challenging to precisely identify the robust feature set required to recognize feelings represented within the audio signal, sentiment analysis utilizing audio signals is a substantial problem (Luitel & Anwar, 2022). There are many different things that might lead to an emotional problem. Occasionally, multiple elements will combine to generate it, for instance, emotional distress at work like long working hours, low salary, poor work conditions, relationships with colleagues or managers (Kandola, 2020). Other than that, sexual harassment or a hostile work environment, poor work organization, poor work design, poor management will also be the factors that affect the emotional problem (*Can You Sue for Emotional Distress? | Working Now and Then*, 2024), (World Health Organization, 2020).

Moreover, the technologies of audio sentiment analysis nowadays are automatic speech recognition (ASR), WaveNet, Mel-Frequency cepstral coefficients (MFCCs) and more. Automatic speech recognition (ASR) uses voice recognition technology to convert spoken utterances into text (*Top 7 Methods for Audio Sentiment Analysis in 2024*, 2024). Next, sentiment analysis is performed on the transcription using natural language processing (NLP) methods (Foster, 2021). Over the past ten years, the area has experienced exponential growth with the ASR systems becoming more and more prevalent in everyday applications such as Zoom for meeting transcription, Spotify for podcast transcription, Instagram and TikTok real-time captioning and more (Foster, 2021). An explosion of applications utilizing ASR technology will occur when it rapidly reaches human accuracy levels, and the accuracy, affordability, and accessibility of ASR technology are increasing because of Speech-toText APIs like AssemblyAI (Foster, 2021). At the same time, WaveNet produces speech that sounds more natural, it mimics human speech by emphasizing and distorting words, phonomenes, and syllables more closely

(*Types of Voices*, 2024). The text-to-speech technology powers Google Translate, Google Translate, and Google Search voice output. WaveNet technology offers a novel approach to producing synthetic speech, not merely a collection of artificial voices (*Types of Voices*, 2024). WaveNet uses deep neural networks to directly analyze raw audio waveforms and extract audio properties, it may extract fine features from the audio stream without the need for audio transcription and probabilistic approach provides state-of-the-art outcomes using a multimodal dataset that combines text and audio (*Top 7 Methods for Audio Sentiment Analysis in 2024*, 2024). Moreover, the short-term power spectrum of sound is represented by Mel-Frequency cepstral coefficients (MFCCs), they are taken out of audio files and added to sentiment analysis models as features (*Top 7 Methods for Audio Sentiment Analysis in 2024*, 2024). Not only this, MFCCs are frequently utilized to describe speech signals due to their ability to convey acoustic information that is understood by the human auditory system (Dwivedi et al., 2023). This technology is less vulnerable to aberrations in the channel and background noise and it is effective to convey the spectral qualities of sound by highlighting the sense of hearing in humans (Upadhyay et al., 2024). It analyzes by using a mathematical model by converting the Mel frequency to real acoustic frequency (Upadhyay et al., 2024).

Besides, audio sentiment analysis can be found in market research, healthcare industry, social media monitoring and so on. In market research, sentiment analysis is a useful technique for gaining insight into consumer attitudes and views about goods, services, and brands, the uses are numerous (xtn, 2024). Therefore, it helps to improve the products and services quality (giosmin, 2023). With voice recording and sentiment analysis, it is more instantaneous to discover customers' views toward a brand or product and identity measures to be made to boost satisfaction (giosmin, 2023). By analyzing the sentiment, it helps to compare against the rivals and quickly identify unfavorable news and mitigate public relations problems to safeguard the brand reputation (Anastasov, 2023). In addition, sentiment analysis serves as a revolutionary force in healthcare that has the potential to greatly enhance the provision of healthcare (Tech, 2024). It enhanced the patient care like healthcare professionals that are aware of the patients' emotions is able to give more compassionate and individualized treatment and comprehend the thoughts, worries and experiences that patients have with their services, early detection of conditions, for example, identifying shifts in a patient's emotional state can help in the early identification such as mental health disorders and so on (*Top 7 Methods for Audio Sentiment Analysis in 2024*, 2024), (optisolnew1, 2023). Additionally, audio recordings from podcasts can also be subjected to audio sentiment analysis; it helps in analyzing public opinion by evaluating speech sentiments to determine the general public's viewpoint on a range of subjects, understanding the emotional responses of the audience to various forms of material can have an impact on content production techniques, and trend analysis by recognizing new themes and opinions in social media discussions, which helps businesses remain ahead of the curve in terms of marketing (*Top 7 Methods for Audio Sentiment Analysis in 2024*, 2024).

The similar idea that is similar to audio sentiment analysis is text sentiment analysis. It evaluates digital text to see if the message's emotional tone is neutral, positive, or negative (AWS, 2023). Text sentiment analysis is highly scalable, it is able to analyze large volumes of text data quickly, making it highly scalable for applications such as social media monitoring, customer reviews, and more, and it is cost-effective, it eliminates the need for manual analysis of large datasets like instead of hiring a team to manually shift through thousands of reviews. However, contextual understanding and ambiguity in language might be the disadvantages such as struggle with the understanding of context, the models often misinterpret sarcastic comments or indirect expressions of emotion, for example, the sentence "Wow, another Monday!" might be flagged as positive as it is actually negative and text is often ambiguous, and the same words can convey different sentiments depending on the context, tone, and structure, the models may misclassify the sentiment in ambiguous sentences. For instance, "The movie was sick!" could mean positive in slang but it might be interpreted as a negative sentence by a model that lacks contextual understanding. The simple algorithm might be it gets input from a tweet, reviews, emails, and more firstly. Then, it preprocesses the text like tokenization, removes stop words, and convert text into vectors. After that, it feeds the text vectors into the machine-learning model to predict the sentiment and provide the output.

2.2 Literature Review

The tool or software that needs to be used is an audio processing library such as Librosa. It is a Python library for audio and music analysis. It offers a number of tools to help you rapidly extract important metrics and audio attributes from the audio files, the audio file formats like MP3, OGG, WAV, and more may all be analyzed and worked with the Librosa library (Technocrat, 2023). Moreover, it provides a broad range of algorithms to extract several types of audio characteristics, including zero-crossing rate, spectral contrast, chroma features, and Mel-frequency cepstral coefficients (MFCCs). Waveforms, spectrograms, and feature plots are just a few of the visual representations of audio data that Librosa can provide to aid in analyzing the qualities of the audio. Not only that, it is useful for tasks like music transcription and instrument recognition since it can estimate the pitch and tonal richness of audio (Jeevitha M, 2023). Furthermore, another library that might involved is the Praat which allows to examination several facets of speech, such as voice quality, pitch, formant, and intensity. Spectrograms which show how sound changes over time, and cochleagrams which are particular spectrograms that more precisely mimic how sound enters the inner ear are allowed to access. With Praat, it may help to produce speech either by using the pitch curve and filters like acoustic synthesis or by using the muscles like articulatory synthesis (*Praat/Praat*, 2021).

The hardware involved in audio sentiment analysis might be the processor which is the CPU in order to audio processing and model training are computationally intensive. A high-performance processor will ensure the execution becomes smoother. Additionally, memory like RAM is required because it is used to store the audio and manipulate the datasets during analysis. A cooling system or some external hardware such as an external SSD, and headphones might required too as the cooling system helps to ensure the laptop stays at optimal performance without thermal throttling since a high-performance laptop tends to get hot during intensive tasks including model training and many more, external SSD may act as a backup storage to storing audio files and datasets, it has a fast read and write speeds for transferring large datasets while headphones may be used to listen for the audio more clearly as audio is involved in the sentiment analysis.

Support Vector Machines (SVM) are powerful supervised learning models used for both classification and regression tasks, though they are primarily used for classification. The core principle of SVM is to find a hyperplane that best divides a dataset into different classes by maximizing the margin between the closest data points, called support vectors. This hyperplane serves as the decision boundary that separates different classes. SVMs can handle both linearly separable and non-linearly separable data by utilizing various kernel functions like the linear kernel for simple problems and the Radial Basis Function (RBF) or polynomial kernels for more complex data (Sasidharan, 2021; IBM, 2023). In audio sentiment analysis, the process begins with preprocessing the audio input by applying noise reduction and voice activity detection. Important features like Mel-frequency cepstral coefficients (MFCCs), pitch, energy, and duration are extracted, which capture essential voice characteristics. After normalization and dimensionality reduction to scale the features, the SVM is trained on labeled emotion data, aiming to classify emotions such as happiness, sadness, or anger. During the prediction phase, the model extracts features from new audio inputs and classify the emotions based on those features (Scikit-learn, 2018; Javatpoint, n.d.). Support Vector Machines (SVM) have several advantages, particularly in high-dimensional spaces where the number of features can exceed the number of data points. SVM is effective for both linearly and non-linearly separable data, thanks to its ability to apply kernel functions like the Radial Basis Function (RBF) and polynomial kernels, making it a versatile choice for tasks like text or audio sentiment classification. Additionally, SVM models tend to be robust to overfitting when there is a clear margin of separation between classes, which contributes to their ability to generalize well even in complex datasets (Kanade, 2022; Gandhi, 2018). However, one of its significant downsides is that it can be computationally expensive, especially when working with large datasets since the training time scales poorly with the number of samples. This makes SVM less suitable for very large datasets. Moreover, SVM is sensitive to noise in the data, such as mislabeled or overlapping classes, which can adversely affect the model's accuracy. Finally, selecting the appropriate kernel function and tuning its parameters for example C and gamma can be challenging, as the model's performance depends heavily on these choices (Sasidharan, 2021; Javatpoint, n.d.).

At the same time, Convolutional Neural Networks (CNNs) have proven to be highly effective in audio classification tasks by converting audio signals into visual representations called spectrograms. A spectrogram is a visual representation of the frequencies of a signal as they vary with time, and it allows CNNs to exploit their strength in image recognition for analyzing audio data (IBM, 2024; Nandi, 2021). Spectrograms are generated by transforming the raw audio data using techniques like Short-Time Fourier Transform (STFT), which decomposes the audio signal into time-frequency components. CNNs can then learn patterns from these spectrograms that are linked to different audio classes or sentiments, such as speech, music, or environmental sounds (emanuelbuttaci, 2023; *lassification of Sound Using Convolutional Neural Networks | IEEE Conference Publication | IEEE Xplore*, n.d.). The basic algorithm works as follows: first, the raw audio data is converted into spectrograms. These spectrograms are then fed into the CNN, which processes them using convolutional layers to detect relevant features, such as frequency bands or harmonics. These features are passed through pooling layers, which reduce dimensionality, followed by fully connected layers that perform the classification. In training, the CNN learns to distinguish between different audio classes by adjusting the weights of the convolutional filters through backpropagation (Badshah et al., 2017; Nanni et al., 2021). Besides, CNNs are particularly effective at capturing intricate patterns in the time-frequency domain, which makes them well-suited for tasks like speech and music recognition. Their ability to leverage local and global features through convolutional layers allows them to excel in complex pattern recognition tasks, leading to high accuracy (Nanni et al., 2021). CNNs are also highly scalable, meaning they can handle large amounts of data, and transfer learning enables them to use pre-trained models, significantly reducing training time and enhancing performance for different audio tasks (IBM, 2024). Additionally, CNNs' ability to automatically extract meaningful features from spectrograms, without the need for manual feature engineering, streamlines the process and boosts their robustness (emanuelbuttaci, 2023). In contrast, it require significant computational power, particularly during training, which can be costly in terms of both time and resources (Nandi, 2021). Another challenge is their reliance on large labeled datasets, which may be difficult to obtain, as accurate labeling of audio data can be time-consuming and expensive (jeffprosise, 2019). Furthermore, CNNs are susceptible to overfitting, especially when used with small datasets, and need careful tuning with techniques such as regularization or dropout to mitigate this risk (*lassification of Sound Using Convolutional Neural Networks | IEEE Conference Publication | IEEE Xplore*, n.d.).

Not only that, Long Short-Term Memory (LSTM) is a type of Recurrent Neural Network (RNN) designed to address the limitations of traditional RNNs, particularly their inability to learn long-term dependencies. LSTM networks are highly effective in processing sequential data, such as time series, natural language, and speech, by utilizing a specialized memory cell that retains information over long periods. Unlike standard RNNs, which suffer from vanishing and exploding gradient problems, LSTMs incorporate gates—namely the input gate, forget gate, and output gate—to control the flow of information, ensuring that important information is preserved and irrelevant data is discarded as needed (Chugh, 2019; Brownlee, 2017). In terms of the algorithm, an LSTM cell operates by first deciding what information to forget or keep through the forget gate. Then, the input gate updates the cell state with new information, while the output gate determines the final output for the current timestep. This ability to selectively remember or forget enables LSTMs to capture both short-term and long-term dependencies in data, which makes them particularly well-suited for tasks like sentiment analysis, where the emotional context of words over time is crucial (*Sentiment Analysis with LSTM*, 2022; *Long Short-Term Memory Network - an Overview | ScienceDirect Topics*, n.d.). Next, LSTMs excel in handling sequential data and are highly effective in tasks where both long-term and short-term dependencies matter, such as natural language processing (NLP), speech recognition, and sentiment analysis (Hamad, 2023). Their internal memory structure makes them more robust for managing complex time dependencies compared to simpler RNNs (Yadav et al., 2023). Additionally, LSTMs are capable of learning context over long input sequences without the gradient vanishing problem, which is common in standard RNNs (Chugh, 2019). Despite their advantages, LSTMs are computationally expensive and require more training time due to their complex architecture. Their intricate gating mechanism also makes them prone to overfitting, particularly when applied to smaller datasets. Furthermore, tuning LSTM models can be challenging, as they require careful adjustment of hyperparameters like learning rate, batch size, and the number of layers (Brownlee, 2017; Luay, 2023). LSTMs are widely used in sentiment analysis tasks, where they help to classify sentiments by analyzing word sequences in text. For example, LSTM networks can capture the sentiment embedded in longer texts, such as product reviews or social media comments, making them a powerful tool for understanding consumer behavior in e-commerce (Yadav et al., 2023; *Sentiment Analysis with LSTM*, 2022).

For the prototype for the sentiment analysis, we plan to use the website as an interface to communicate with the user, and there are many suitable tools used for webpage development. Firstly, Visual Studio Code (VSCode) is a powerful and versatile code editor that supports various programming languages, including HTML and CSS. It offers an array of features such as syntax highlighting, intelligent code completion, and integrated version control, making it an excellent choice for web development (Microsoft, 2016; V*isual Studio vs Visual Studio Code - What's Best in 2022?*, n.d.). The editor's built-in Emmet support enhances productivity by allowing developers to write HTML and CSS code quickly using abbreviations (*Visual Studio Code*, 2016). Furthermore, VSCode is lightweight, has a vast library of extensions, and is highly customizable, which enables users to tailor the environment to their preferences. It also includes integrated debugging tools, terminal support, and collaboration features (*Visual Studio Code: Read This before You Get Started*, n.d.). On the downside, some users may find that the sheer number of extensions can lead to confusion, and the initial setup may require some time to configure optimally (V*isual Studio vs Visual Studio Code - What's Best in 2022?*, n.d.). Additionally, while it is generally fast, performance can lag when handling very large files or projects.

The second tool is NetBeans, it is a robust Integrated Development Environment (IDE) designed for Java development but also supports web technologies such as HTML and CSS. It provides features like syntax highlighting, code templates, and built-in validation tools that streamline the web development process (*Welcome to Apache NetBeans*, n.d.; *Easy Web Site Creation in the NetBeans IDE*, 2019). NetBeans also allows developers to create web applications quickly, offering a straightforward approach to building modern web interfaces (*Easy Web Site Creation in the NetBeans IDE*, 2019). One of the main advantages of NetBeans is its all-in-one environment, which includes powerful tools for debugging and testing, making it easier for developers to manage projects (Apache NetBeans, 2017). It also supports various programming languages, allowing for versatility in development (*Welcome to Apache NetBeans*, n.d.). However, NetBeans can be resource-intensive, potentially leading to slower performance on less powerful machines. Additionally, while it offers a wealth of features, some users may find the interface less intuitive compared to other lightweight editors (*Easy Web Site Creation in the NetBeans IDE*, 2019).

Table 2.1: Audio Processing Tools

Aspect	**Librosa**	**Praat**
Purpose	General-purpose audio processing and analysis	Specialized in speech analysis and synthesis
Feature Extraction	Extracts features like MFCCs, spectral contrast, and chroma	Focuses on speech-specific features like pitch, formants, and voice quality
Visualization	Waveforms, spectrograms, and feature plots	Spectrograms, cochleagrams
Supported Formats	Works with MP3, WAV, OGG and more	Primarily works with WAV and specific formats
Ease of Use	Python-based, well-documented, and widely used in machine learning	Offers GUI and scripting support but has a steeper learning curve
Applications	Music analysis, general audio feature extraction	Speech analysis, pitch tracking, and voice quality studies
Limitations	Not tailored for speech-specific metrics like formants	Limited general audio processing capabilities compared to Librosa

The above table shows the audio processing tools which are Librosa and Praat, which summarize from the above Literature Review. It compares the library in different aspect include purpose, feature extraction and so on. A table is able to increase the efficiency to know about the difference of the library and make a decision on using which of the library in the system.

Table 2.2: Machine Learning Models for Audio Sentiment Analysis

Aspect	**Support Vector Machine (SVM)**	**Convolutional Neural Network (CNN)**	**Long Short-Term Memory (LSTM)**
Type	Supervised learning model for classification and regression	Deep learning model specialized for spatial data (images, spectrograms)	Recurrent Neural Network (RNN) optimized for sequential data
Data Type	Structured data or feature vectors (MFCCs, pitch)	Time-frequency representation (spectrograms)	Sequential data (audio waveforms, text, or sequences of features)
Feature Extraction	Requires manual feature engineering (MFCCs, pitch)	Automates feature extraction from spectrograms	Automates features extraction from sequential data
Handling Nonlinear Data	Excellent with kernel function as like RBF and polynomial	Handles complex patterns in spectrograms	Capture long-term and short+term dependencies in sequences
Training Efficiency	Computationally efficient for small to medium datasets	Computationally expensive, especially with large datasets	Computational expensive due to complex architecture and gating mechanisms
Performance	Effective for small datasets with clear class separation	High accuracy for large datasets and complex patterns	Excels in tasks with sequential dependencies, such as speech or text sentiment analysis
Overfitting Risk	Low risk if margin is well defined	High risk, especially with small datasets, mitigated with regularization techniques	High risk, especially with small datasets, mitigated with techniques like dropout and proper regularization
Scalability	Less scalability for very large datasets	Highly scalable, especially with GPUs and pre-trained models	Moderately scalable , training time increases with sequence length and model complexity

Interpretability	Relatively interpretable (support vectors, decision boundary)	Less interpretable due to deep network layers	Less interpretable, but gating mechanisms provide some insights into memory handling
Applications	Emotion classification from extracted features	Audio sentiment classification using spectrograms, speech or music or environment recognition	Speech based sentiment analysis, sequential modelling, emotion recognition in long utterances

The above table shows the machine learning models of audio sentiment analysis which are Support Vector Machine (SVM), Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM), which also summarize from the above Literature Review. It compares the models in different aspect include type, data type, feature extraction, handling nonlinear data, training efficiency, and many more. By using a table to show the comparison of the models, it helps to reduce the time to know about each model and easier to make a selection to use which models to develop the system.

Table 2.3: Web Development Tools

Aspect	Visual Studio Code (VSCode)	NetBeans
Primary Use	Lightweight code editor supporting HTML, CSS, JavaScript, and extensions for other languages	Integrated Development Environment (IDE) primarily for Java, with web development support
Customization	Highly customizable with extensions	Limited customization options compared to VSCode
Debugging Tools	Integrated debugging for multiple languages	Powerful debugging tools, especially for Java
Ease of Use	Beginner-friendly with intuitive UI and strong community support	Steeper learning curve, especially for non-Java developers
Performance	Lightweight; suitable for small to medium projects	Resource-intensive; better suited for large-scale development
Collaboration	Built-in tools for real-time collaboration	Lacks native collaboration features
Applications	Web development, frontend coding	Comprehensive Java projects, backend systems, and web applications
Limitations	Performance may lag with very large files	Less intuitive for beginners and slower on less powerful machines

The above table is about the web development tools of audio sentiment analysis system which are Visual Studio Code (VSCode), and NetBeans. It compares the development tools in different aspects such as primary use, customization, debugging tools, performance and many more. It shows the differences of the models, hence it increases the understanding of each of the web development tools and make the selection easier.

Table 2.4: Overall Studies of the Chapter

Category	Tool/Method	Pros	Cons
Audio Processing Tools	Librosa	- Rich feature extraction (MFCCs, chroma) - Support multiple audio formats - Visualizations like spectrograms	- Limited to Python - Not tailored specifically for speech features like Praat
	Praat	- Advanced speech analysis (voice, quality, pitch, formants) - Suitable for speech synthesis	- Limited general audio processing capabilities compared to Librosa
Machine Learning Models	Support Vector Machine (SVM)	- Handles high-dimensional data - Effective for small to medium datasets - Versatile kernel functions	- Computationally expensive for large datasets - Sensitive to noise and parameter tuning
	Convolutional Neural Network (CNN)	- Automatically extracts features - Effective for time-frequency data (spectrograms) - Scalable with transfer learning	- Requires large labeled datasets - High computational and training cost
	Long Short-Term Memory (LSTM)	- Captures long and short-term dependencies - Avoids vanishing gradient issues - Effective for sequential data	- Computationally expensive - Required careful tuning and prone to overfitting
Web Development Tools	Visual Studio Code (VSCode)	- Lightweight and highly customizable - Supports extensions for HTML, CSS, and debugging - Collaborative tools	- Can be overwhelming due to a vast library of extensions - Performance issues with large files
	NetBeans	- Comprehensive for Java and web technologies - Built-in debugging and testing tools	- Resource-intensive - Less intuitive interface for beginners

The table above is an overall study of the chapter, it shows the advantages and disadvantages of each tool or methods. Pros and cons also a condition to determine whether which tools or methods will be chosen to use in develop the system.

2.3 Feasibility Study

2.3.1 Economic Feasibility

Firstly, economic feasibility entails calculating the price of purchasing the required hardware, software, and cloud infrastructure in addition to continuing server fees for audio data processing. Important factors include recruiting qualified experts, training employees, and licensing costs for proprietary tools or APIs. Using open-source tools and making the most of cloud resources are two cost-cutting techniques that can assist with cost management. Long-term advantages including higher employee happiness, better customer service, and greater brand reputation are what generate return on investment (ROI), which should exceed startup and operating expenses.

2.3.2 Operational Feasibility

Furthermore, operational feasibility looks at the system's ability to satisfy user demands and integrate into the organization. For non-technical stakeholders, the solution should include an easy-to-use interface that allows for flexibility in adapting to different communication channels or linguistic support. Maintaining relevance requires regular changes, such as system scalability and model retraining. User input and pilot projects may guarantee acceptability and alignment with corporate objectives. Operational viability is further increased with a strong maintenance plan that includes routine monitoring and prompt problem-solving.

2.3.3 Technical Feasibility

Besides, the availability of appropriate infrastructure and tools for audio sentiment analysis is guaranteed by technical viability. Development may be accelerated by using pre-trained models for sentiment identification and speech-to-text conversion, such Wav2Vec2.0 or BERT, while data processing is efficiently managed by robust cloud platforms or on-premise solutions. Large audio files must be supported by reliable storage solutions, and success depends on features like encryption, noise reduction, and smooth system integration. The technical dependability of the system is further improved by scalable architecture and effective resource utilisation.

2.3.4 Schedule Feasibility

Additionally, the timescale for creating and implementing the system is evaluated by schedule feasibility. This covers the time needed for user training, testing, model building, and data gathering. To keep the project on schedule, it is crucial to set clear milestones, and considering potential delays from external dependencies is key to keeping the project on track. To guarantee that the system is adopted and used effectively, sufficient time must also be set out for staff training.

2.3.5 Legal Feasibility

In addition, legal feasibility guarantees adherence to data privacy laws such as the CCPA and GDPR, especially when it comes to the gathering and examination of audio data. Audio recording and processing require user consent, and data anonymisation and retention guidelines should be set up. To prevent legal issues, it is also necessary to handle intellectual property rights and adherence to industry-specific legislation. Legal and ethical compliance is further improved by putting policies in place to stop sentiment model bias, keeping thorough records, and carrying out frequent audits.

2.4 Chapter Summary and Evaluation

Audio sentiment analysis is introduced as a technique to interpret emotions from speech using technologies like Automatic Speech Recognition (ASR), WaveNet, and Mel-Frequency Cepstral Coefficients (MFCCs). It has applications in areas such as market research, healthcare, and social media. The covers tools like Librosa and Praat for feature extraction and machine learning models like Support Vector Machines (SVM), Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM) for classifying emotions. It also highlights the hardware needed for processing, including processors, RAM, and cooling systems. Other than that, the feasibility study assesses costs, technical needs, and legal compliance. Economically, it considers software, hardware, and cloud expenses, while operationally, it focuses on ease of use and scalability. Technical feasibility ensures proper tools are available, and the schedule looks at development timelines, while legal feasibility ensures compliance with data privacy laws.

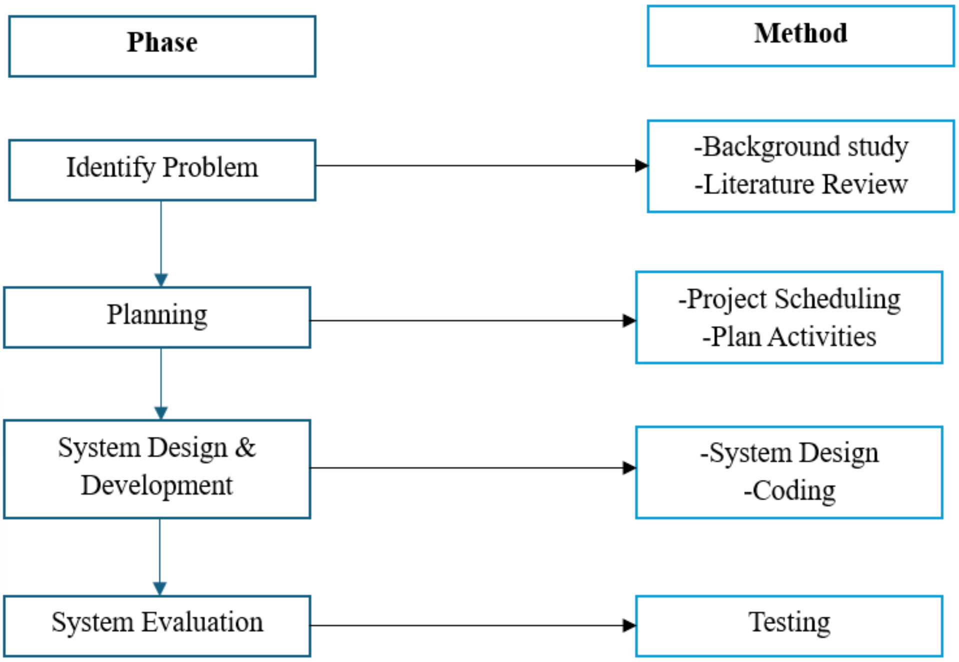
Chapter 3

Methodology and Requirements Analysis

3 Methodology and Requirements Analysis

3.1 Overall Methodology

Figure 3.1: Overall Methodology



The diagram illustrates the overall methodology for project development, breaking it into four distinct phases which are Identify Problem, Planning, System Design & Development, and System Evaluation, with associated methods.

The first step in the process is problem identification, which is backed up by background research and literature review to comprehend current difficulties. Since we completed the literature review and background research, we are aware of the problems, resources, and methods used in audio sentiment analysis, including the audio processing libraries Praat and Librosa. Sentiment analysis may be done in a variety of ways using techniques like Support Vector Machines (SVM), Convolutional Neural Networks (CNNs), and many more. Hardware is a crucial prerequisite for audio sentiment analysis, in addition to tools and algorithms.

The project moves on to the following stage, planning, once the issue has been recognised. Project scheduling and activity planning are two techniques that involve project planning in order to provide sufficient time and resources for every stage. To ensure that the project is proceeding according to schedule, for example, a Gantt chart and a table that details the activity, expected result, and completion date have been created.

Furthermore, the system must be designed and developed with careful design and coding to guarantee that it is efficient and easy to use. To guarantee the system's development and future enhancement, a system design should be established earlier to system development. As a result, several diagrams, including use cases, state diagrams, activity diagrams, and others, must be produced. During the development phase, audio processing techniques include noise reduction, voice activity detection, and spectrogram conversion. Additionally, Python may be used to construct audio sentiment analysis. Additionally, the user interface may employ a basic website to facilitate communication between the user and the system, which consists of HTML and CSS created with NetBeans or Visual Studio Code (VSCode).

The evaluation step, which comes last, will evaluate the system to ensure that its performance and functionality are satisfactory. To make sure the system can perform as expected and meet both functional and non-functional criteria, it will be trained and tested using the actual dataset.

3.1.1 Selected Models

Convolutional Neural Networks (CNNs) are the chosen algorithm. First, data collection is required; collect labelled audio data that represents various emotions, such as neutral, negative, and positive. To train and test the system, real-world audio datasets must be obtained or discovered online. These can be either custom data collection by recording audio samples from volunteers or pre-existing datasets such as the Ryerson Audio-Visual Database of Emotion Speech and Song (RAVDESS) from online platforms. Data preprocessing and cleansing, the second necessary step after data collection, involves consistently structuring and eliminating unwanted noise from the raw audio data to prepare it for analysis. For instance, managing missing or damaged audio files, reducing background noise using the Librosa library, normalizing audio to achieve a consistent loudness, and trimming the audio to maintain consistent length throughout datasets.

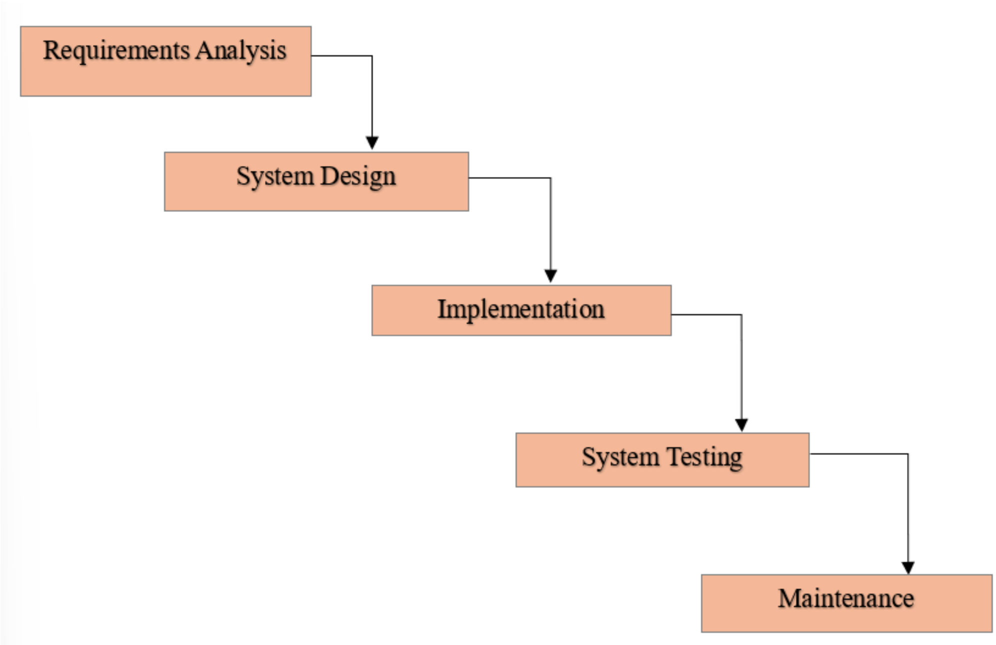
The feature extraction is another. It converts the unprocessed audio into representations that the CNN can understand, such as spectrograms, which show the frequency spectrum over time graphically. It creates spectrograms that will capture the qualities of the audio by using the Mel-Frequency Cepstral Coefficients (MFCCs). By changing the pitch to produce variations in audio samples, shifting time to shift the audio signal forward or backward, and adding background noise to introduce controlled noise to stimulate different recording environments, data augmentation improves the training dataset and enhances the model's robustness and generalisation capabilities. Additionally, by making sure the audios are on a same size, feature scaling and transformation are the following steps that are required to get the extracted features ready for input into the CNN. It prepares the data to meet the CNN architecture's input criteria and standardises and reshapes the data, adjusting the features to have a mean of zero and a normal deviation of one.

Additionally, the datasets should be divided into training, validation, and test sets. It may be trained, validated, and tested using percentages such as 80, 10, 10, or 70, 15, 15. A number of layers are included in the CNN model architecture, including the input, convolutional, pooling, fully connected, and output layers. The input layer receives the input spectrograms following preprocessing and feature extraction. Next, the convolutional layers determine the number of filters, filter size, and stride for feature extraction from the input, while the pooling layer determines the maximum or average pooling to minimize the size of the feature maps. Moreover, the dense layer is connected to the fully connected layer by flattened gesture maps from the preceding layer and the output layer employs activation functions such as softmax for classification.

In order to reduce error and increase accuracy, the CNN model is then fitted into the training data while its performance is tracked on the validation set. This helps to guarantee that the CNN recognises patterns in the spectrograms that correlate to certain emotions. Upon training, the weights of the model are adjusted to reliably categorise audio spectrograms into specific sentiment categories. Model testing and validation come next. The validation dataset is used to evaluate the model's performance during training; metrics like precision, recall, and f1-score are tracked. The test dataset is used to evaluate the model's final performance after it has been trained; metrics like accuracy and confusion metrics will be generated.

3.2 Proposed Software

Figure 3.2: Software Process Model: Waterfall Model



The figure illustrates the phases involved in creating an audio sentiment analysis system which are requirements collection, analysis, design, coding, testing, and maintenance. The waterfall model is the selected software process model. It breaks down project activities into sequential, linear phases, each of which is based on the deliverables of the previous one and represents a speciality of tasks. The technique is widely used in many areas of engineering design.

First, all of the functional and non-functional requirements for the audio sentiment analysis system must be gathered and recorded during the requirement analysis stage. It aids in determining the system's goal, such as feedback processing, voice analysis, or emotion recognition. Furthermore, the types of audio inputs such as MP3, WAV, and so on, processing techniques, for example, pitch analysis, MFCC extraction, and the desired result like positive, negative, or neutral must all be specified. These include machine learning models such as SVM, CNN, and LSTM, as well as libraries like Librosa and Praat.

Moreover, system design is the second state. To understand how the system would be constructed in this pertinent stage, a blueprint is required. It creates the architecture for feature extraction, audio processing, and emotion analysis. In order to construct the system, it is also necessary to plan the modules such as preprocessing, which includes noise reduction, feature extraction, and classification. Additionally, the programming language, libraries, and tools must be chosen for the entire system. Create system processes, including data flow diagrams and many more after that.

In the third step, known as the implementation stage, the system components are developed and integrated according to the design. Written code is used to process audio recordings, extract features, and create machine learning models. Once the code is finished, the CNN model that was chosen must be trained using the sentiment analysis datasets. Additionally, VSCode will be used to create a basic web-based user interface for adding audio samples and showing analysis results.

The next step is system testing, which verifies that the system satisfies the specifications and functions as planned. To make sure the algorithm can accurately categorise the emotions, it also conducts functional testing. Precision, recall, f1-score, and accuracy are among the measures used to assess the system's performance.

The final step is maintenance, which guarantees that the system will continue to function and be effective following deployment. The public may provide input on the system's functionality or performance after it has been made available to the general public; in response, the system may improve its functioning. Aside from that, update the system to accommodate new specifications or enhance accuracy by keeping the models with more data. For system maintenance, it is also necessary to keep an eye out for any mistakes, bugs, or performance problems with the system.

The waterfall model may be used since the project's requirements are set, its timeframe is known, its technology is well understood, and so on. Early changes to the specifications and design of the system are easy to implement. The waterfall approach also facilitates understanding, adhering to, and planning tasks. However, the waterfall technique is not the greatest option for complex and high-risk projects because no working product is available until later in the project lifetime.

3.3 Chapter Summary and Evaluation

3.3.1 Functional Requirements

Table 3.1: Functional Requirements

No.	Modules	Actors/Users	Requirements
1.	User Management	1.1 User	1.1.1 The system should allow the user to register for an account in the sentiment analysis platform. 1.1.2 The system should allow the user to log in and access their account securely.
2.	Audio Sentiment Analysis	2.1 User	2.1.1 The system should allow the user to upload the audio. 2.1.2 The system should have the extracted features, such as spectrograms from the audio signals. 2.1.3 The systems should be able to be preprocessing and cleansing the missing or corrupted audio files. 2.1.4 The system should be able to classify the audio data into sentiment categories such as sad, fear, happy, and so on.
3.	User Interface	3.1 User	3.1.1 The system must be able to display the sentiment analysis result in a user-friendly format. 3.1.2 The system should provide a user interface for users to upload the audio and view results.

3.3.2 Non-Functional Requirements

Performance Requirements

No.	Requirements
1.	The system response to the user action such as uploading the audio should take no more than 20 seconds.
2.	The system should be able to handle at least 1500 users simultaneously.
3.	The display of the result of the sentiment analysis should not be more than 10 seconds.
4.	The result downloading process must be completed within 10 seconds.

Safety and Security Requirements

No.	Requirements
1.	The system must protect the sensitive information that is the result of the sentiment analysis in a safe condition.

Software Quality Attributes

No.	Attributes	Requirements
1.	Usability	The average time taken by users to complete an analysis should not be more than 5 minutes.
2.	Reliability	The system should achieve a 99.99% uptime, with no more than 1 minute downtime per month.

3.4 Other Requirements

3.4.1 Hardware Requirements

Table 3.2: Hardware Requirements

No.	Category	Requirements
1.	Devices	1.1 The CPU of the laptop or desktop such as Intel i5, i7, i9, AMD Ryzen, or more for the basic processing.
		1.2 At least 8GB RAM to handle the data smoothly and for model training.
		1.3 A high-quality audio interface card for accurate input and output processing.
		1.4 500GB of SSD required for storing audio files, preprocessed data, and training the model.
2.	Networking	1.1 The signal strength number needs to be closer to 0, which is a strong signal and does not lower than -85dBm to avoid slow or no connection.
		1.2 A firewall and instruction detection system (IDS) need to be set up.
		1.3 The latency should be under 00ms to provide a smooth user experience.
		1.4 The minimum bandwidth is 11178 GB per month and the maximum bandwidth is 16767 GB per month.

3.4.2 Software Requirements

Table 3.3: Software Requirements

No.	Category	Requirements
1.	Operating System	1.1 A Windows 10 and above, or MacOS 13 and above operating system is required.
2.	Programming Tools	2.1 Python 3.8 and above programming language is required to develop the system.
		2.2 HTML4 and above is required to develop the simple user interface.
		2.3 Lirosa or Praat library is applied to the system.

3.5 Chapter Summary and Evaluation

This chapter provides an in-depth overview of the methodology and system requirements for developing an audio sentiment analysis system. The algorithm that was selected for the audio sentiment analysis is the Convolutional Neural Network (CNNs), it involved data collection, data preprocessing and cleansing, data augmentation, feature scaling, model training, model validation and model testing. The proposed methodology is structured into distinct phases, adhering to the waterfall model include problem identification, planning, system design, coding, testing, and evaluation. Each phase includes specific techniques and deliverables, ensuring systematic progress and a robust final product.

The functional requirements emphasize the system's capabilities, including enabling users to upload audio files, extracting features such as spectrograms, classifying sentiments such as positive, negative, neutral, and displaying results in a user-friendly format. Non-functional requirements focus on performance, usability, and reliability, such as swift responses to user actions, supporting multiple users simultaneously, and maintaining system uptime of 99.99%.

Hardware requirements highlight the need for a high-performance computing environment, including a powerful CPU like Intel i5/i7/i9, AMD Ryzen, at least 8GB RAM, SSD storage, and quality audio interfaces for precise processing. Software requirements include modern operating systems, for instance, Windows 10 and above or macOS 13 and above, programming tools like Python, and libraries such as Librosa and Praat.

This chapter concludes with a comprehensive outline of the structured approach to ensure that the system is both functional and efficient while meeting user needs and technical specifications.

Chapter 4

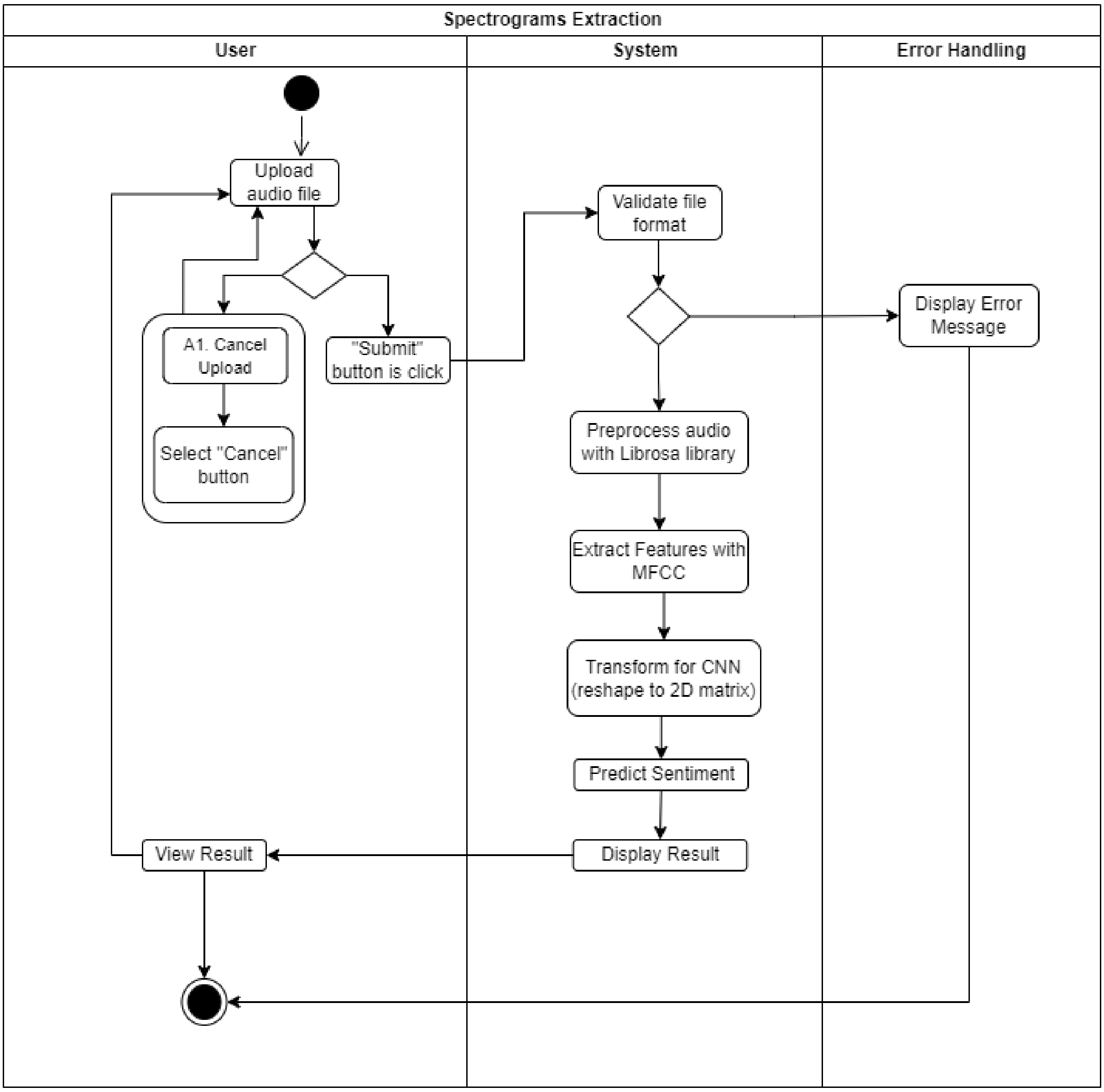
System Design

4 System Design

4.1 Algorithm / Process Design

4.1.1 Spectrograms Extraction

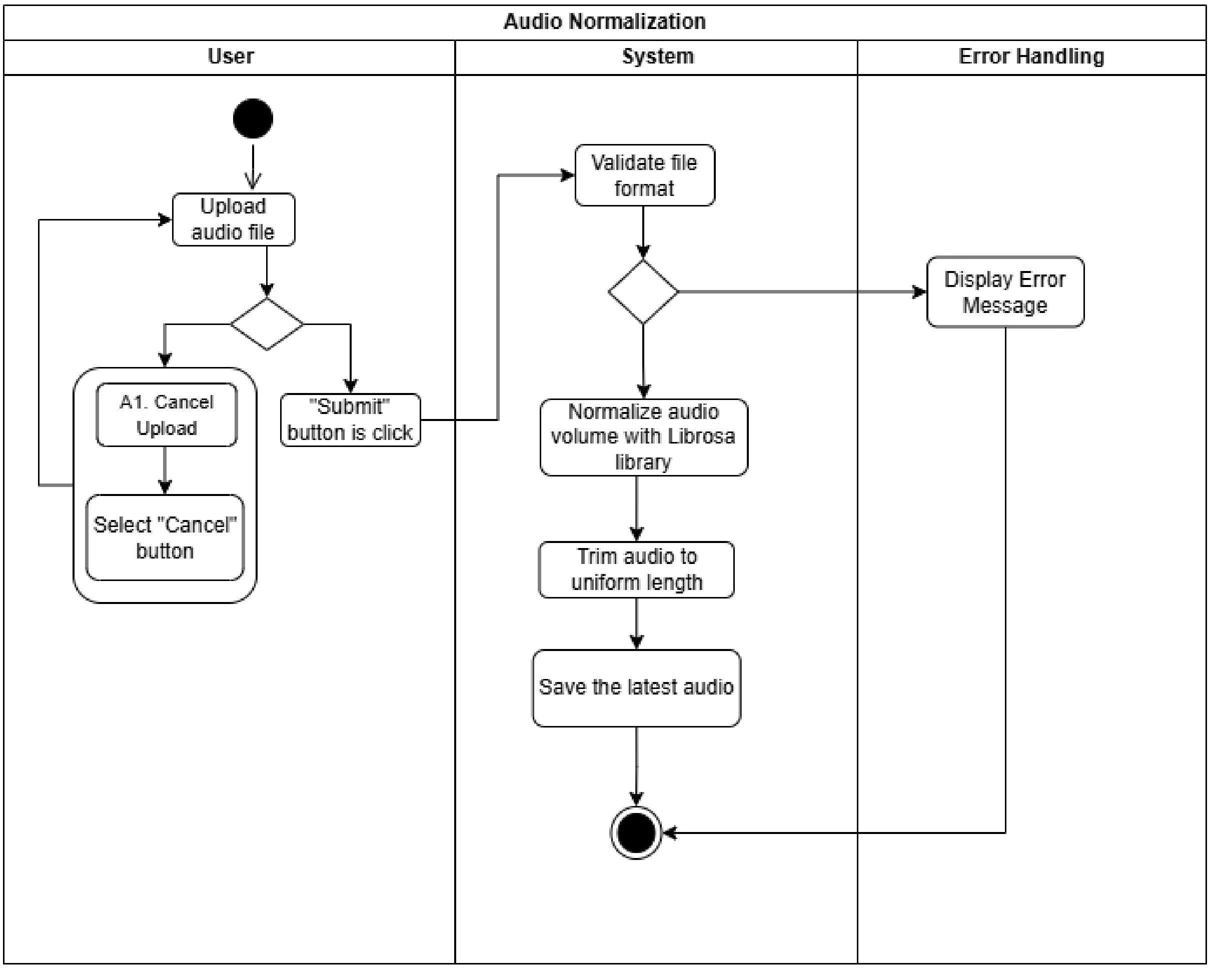
Figure 4.1: Spectrograms Extraction Activity Diagram



The activity diagram above shows the data flow of spectrogram extractions. A spectrogram, which is a frequency used in audio analysis, is a graphic depiction of a signal's frequency spectrum across time. The user first needs to upload the audio in a correct format, then the system will preprocess with the Librosa library, extract features with MFCC, and transform for CNN. After that, it will predict the sentiment of the audio and display the result of the audio. It also has an error handling to handle the audio which is uploaded in an incorrect format, it will end the process directly.

4.1.2 Audio Normalization

Figure 4.2: Audio Normalization Activity Diagram

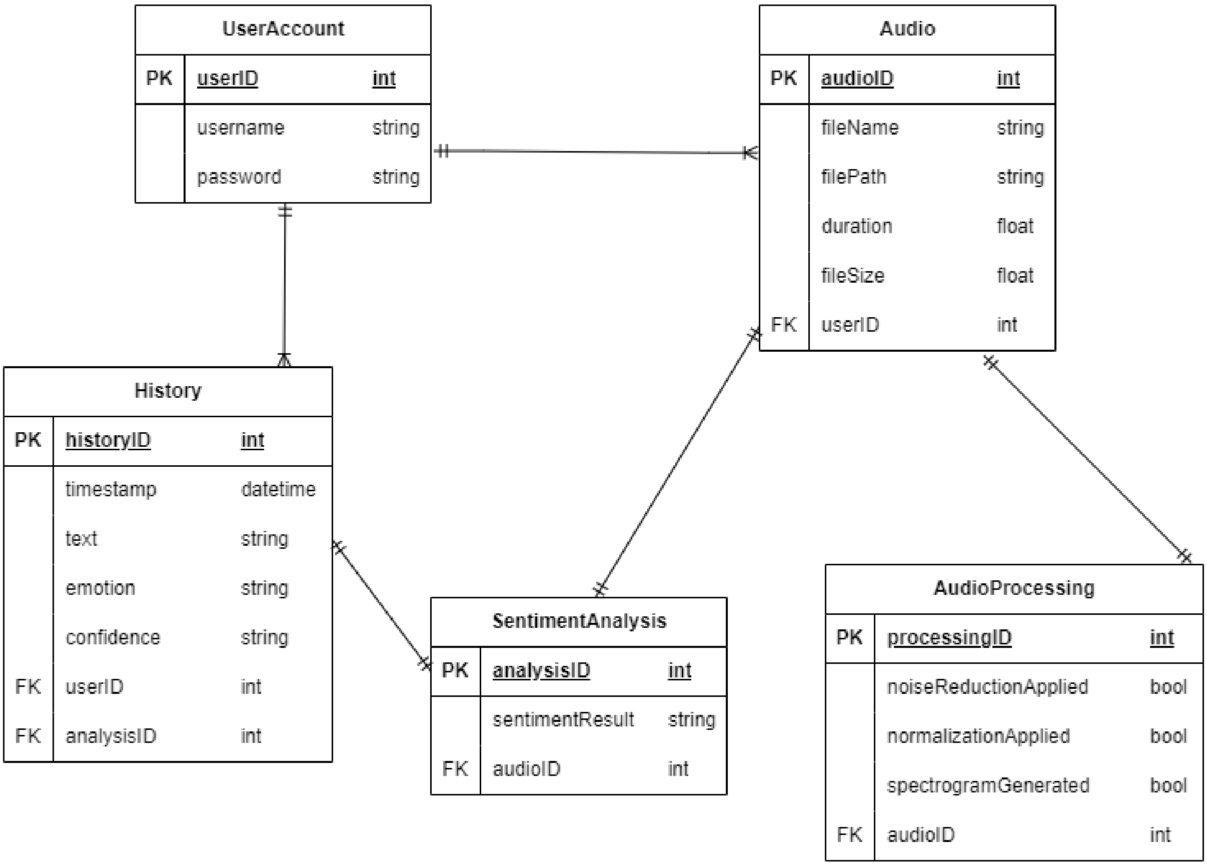


The above activity diagram is related to audio normalization, which normalize the volume of the audio and trim the audio into a uniform length. Same with the spectrogram normalization, the user should upload the audio at first in order to go into the audio normalization process. After the audio's volume has been normalize and trim to a uniform length, it will save the latest audio. Furthermore, it also has error handling to handle the audio which uploaded in an incorrect format, it will display an error message to notify that the audio format is wrong.

4.2 Database Design & Structural Models

4.2.1 ERD

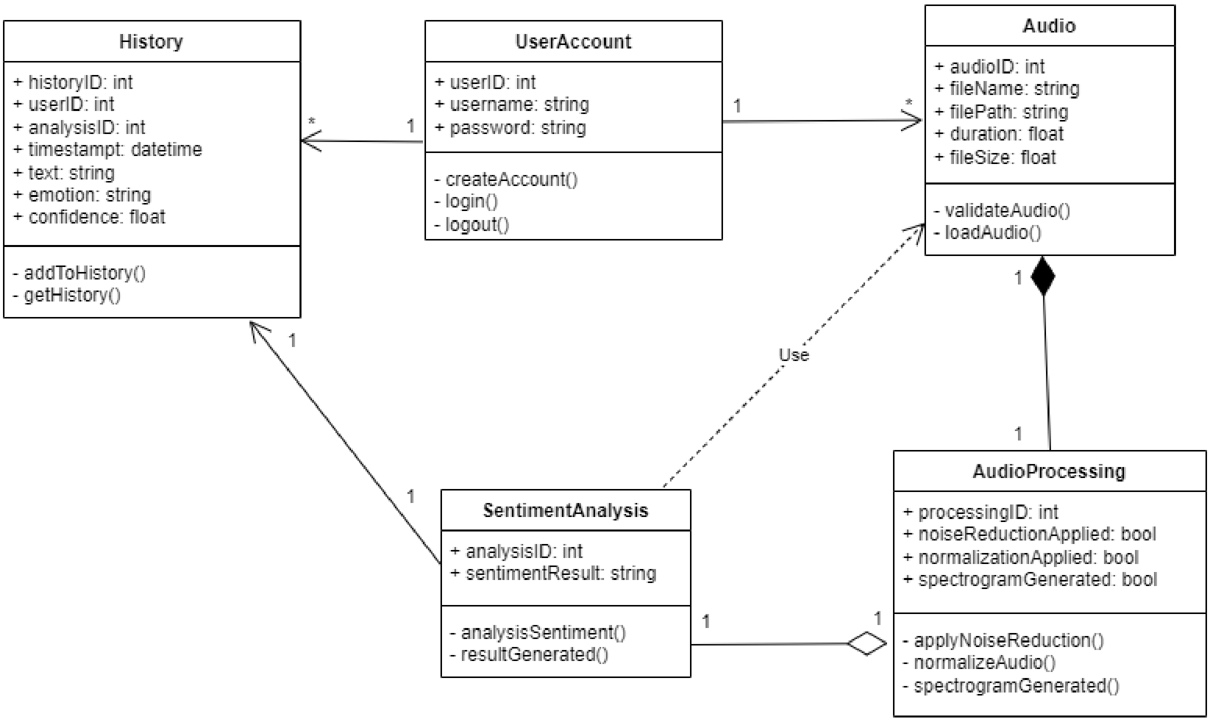
Figure 4.3: ERD of Audio Sentiment Analysis



The ERD diagram shows the data structure of the system. It has 5 class involves UserAccount, AudioProcessing, Audio, SentimentAnalysis, and History class. The ERD shows the relationship between each of the class. A UserAccount is able to upload one or more than one audio, while one audio only belongs to a UserAccount. Furthermore, Audio only has a AudioProcessing and a SentimentAnalysis while a AudioProcessing and a SentimentAnalysis only belongs to an Audio. Audio, AudioProcessing and SentimentAnalysis is a one-to-one relationship. Additionally, the History may have one or more than one history from UserAccount, and a one-to-one relationship with the SentimentAnalysis since one history only own by one sentiment analysis and the audio is unique, therefore each audio’s analysis has a unique history.

4.2.2 Class Diagram

Figure 4.4: Class Diagram of Audio Sentiment Analysis

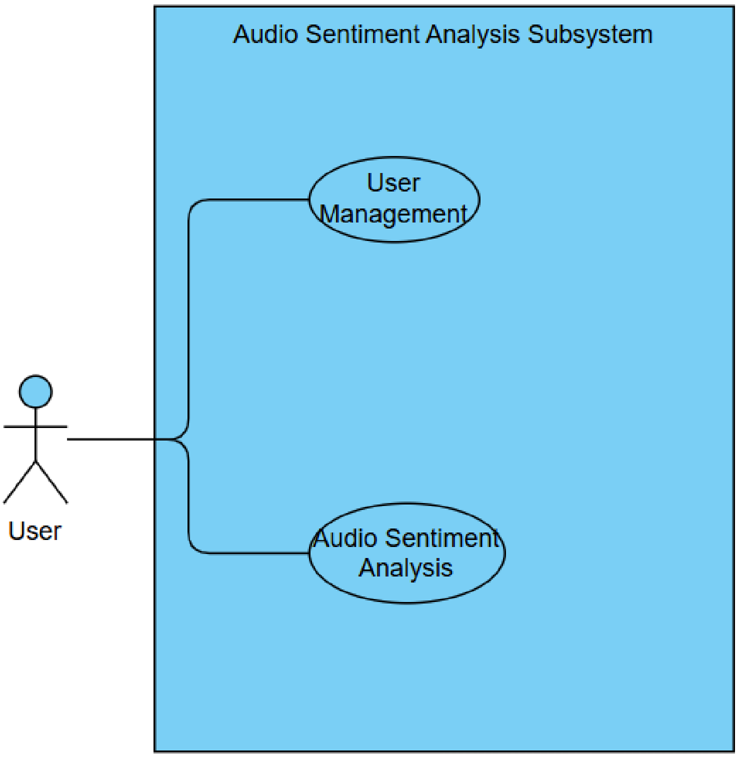


By displaying a system's classes, attributes, methods, and relationships between them, the class diagram illustrates its structure. A UserAccount is able to upload multiple Audio files, it is an association, and one-to-many relationship between them. Moreover, the relationship between the Audio and AudioProcessing is composition, and it is a one-to-one relationship. An Audio file owns an AudioProcessing instance, means that the AudioProcessing cannot exist without Audio. At the same time, the relationship between AudioProcessing and SentimentAnalysis is aggregation, and also a one-to-one relationship. A processed AudioProcessing instance is used to generate a SentimentAnalysis result. The SentimentAnalysis depends on the AudioProcessing but it does not own the AudioProcessing strictly. Other than that, the relationship between Audio and SentimentAnalysis is indirect association. The Audio indirectly connects to SentimentAnalysis through AudioProcessing, the SentimentAnalysis uses audioID from Audio via the AudioProcessing class. Last but not least, the History has a one-to-one relationship with SentimentAnalysis and also a one-to-many relastionship with UserAccount.

4.3 Interaction Models

4.3.1 Main Use Case Diagram

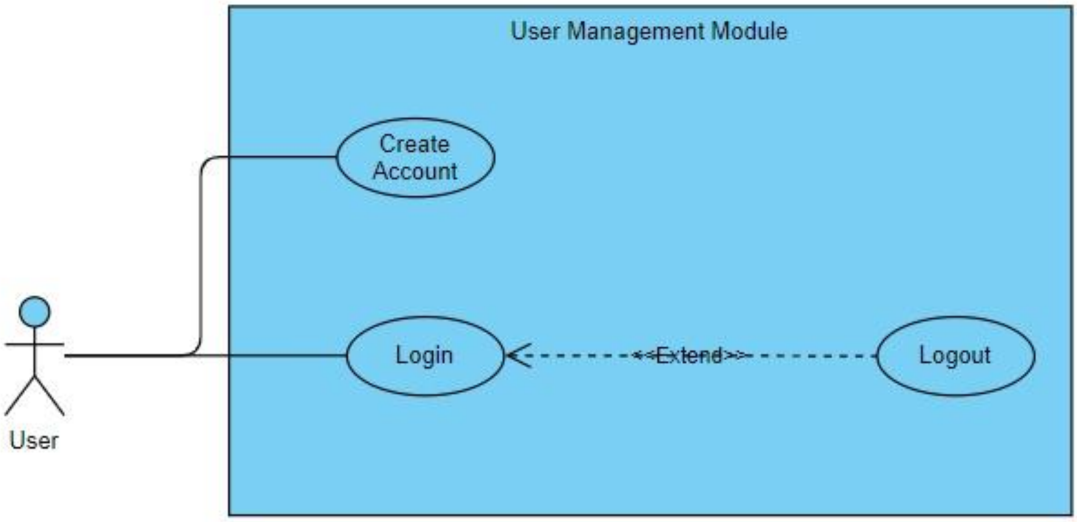
Figure 4.5: Audio Sentiment Analysis Subsystem Main Use Case Diagram



It is an overall use case diagram that show the module in the subsystem, in the audio sentiment analysis subsystem, it has 2 modules which are User Management Module and Audio Sentiment Analysis Module.

4.3.2 Use Case #1 User Management Module

Figure 4.6: Audio Sentiment Analysis Subsystem User Management Module Use Case Diagram



It is the use case diagram for the User Management Module. The user is able to create account, and login. A new user needs to create account before they login to the system to do the sentiment analysis. Furthermore, after the user login the system and finish the analysis, they might need to choose whether they want to logout their account or no.

Use Case Description Table

Table 1

<table>
<tr><td colspan="2"><b>Use Case Name:</b> Create User Account</td></tr>
<tr><td colspan="2"><b>Actor:</b> User</td></tr>
<tr><td colspan="2"><b>Brief Description:</b> This use case allows the user to create an account.</td></tr>
<tr><td colspan="2"><b>Pre-condition: -</b></td></tr>
<tr><td colspan="2"><b>Main Flow of Events:</b></td></tr>
<tr><td><b>Actor Action</b></td><td><b>System Response</b></td></tr>
<tr><td>1) The user navigates to the registration page.</td><td>2) The system displays the registration form.</td></tr>
<tr><td>3) The user fills in the required details such as username and password.
4) Select the “Submit” button to submit the form.</td><td>5) The system validates the details whether they meet the requirements
6) The account creation is successful if the details are valid and in the correct format.</td></tr>
<tr><td colspan="2"><b>Alternative Flow of Events:</b>
<u>A2. Step 5</u>
If the details are not valid and not in the correct format, the system will display an error message and ask the user to fill in again the details.</td></tr>
</table>

Table 4.1: Description Table of Create User Account

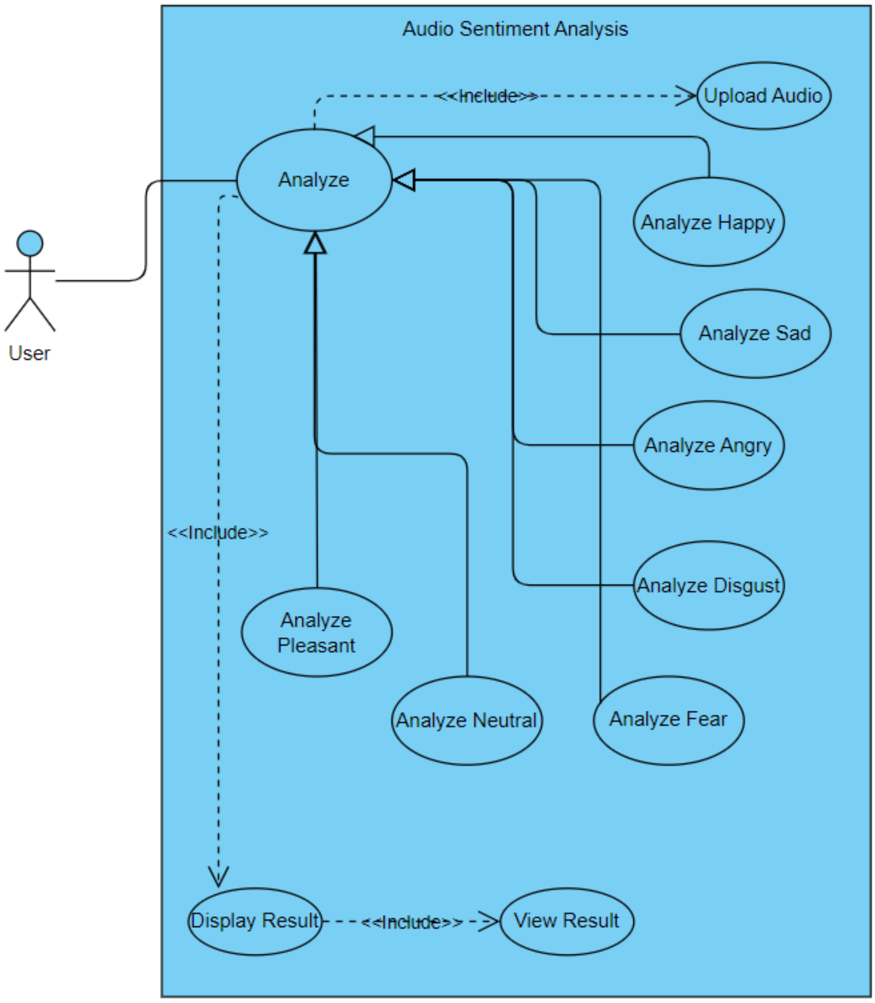
Table 2

Use Case Name: Login Account	
Actor: User	
Brief Description: This use case allows the user to login to their account.	
Pre-condition: The user needs to have an active account.	
Main Flow of Events:	
Actor Action	**System Response**
1) The user navigates to the login page.	2) The system displays the login form.
3) The user fills in the required details such as username, and password. 4) Select the "Login" button to submit the form.	5) The system compares the details with the database. 6) If the details match with the database, the user is able to login successfully.
Alternative Flow of Events: A2. Step 5 If the details are not valid and not in the correct format, the system will display an error message and ask the user to fill in again the details.	

Table 4.2: Description Table of Create User Account

4.3.3 Use Case #2 Audio Sentiment Analysis Module

Figure 4.7: Audio Sentiment Analysis Subsystem Audio Sentiment Analysis Module Use Case Diagram



It is the use case diagram for the Audio Sentiment Analysis Module. The user is able to analyze the audio. The user needs to upload the audio in a correct format to start the analysis. After the audio processing and analysis, the system will display the result, hence the user is able to view the report. The system will analyze the audio whether is angry, sad, fear, neutral, pleasant, happy, or disgust.

<u>Use Case Description Table</u>

<u>Table 1</u>

Use Case Name: Analyze Sentiment of Audio	
Actor: User	
Brief Description: This use case allows the user to analyze the sentiment of the audio.	
Pre-condition: An audio must be uploaded	
Main Flow of Events:	
Actor Action	**System Response**
1) The user navigates to the sentiment analysis page.	2) The system displays the page.
3) The user uploads the audio file in the correct format 4) Select "Analyze" button	5) The system validates the audio format 6) If the format is valid, it starts to upload the audio 7) The system starts to analyze the audio 8) The system will display the analysis result
9) View the result	
Alternative Flow of Events: <u>A1. Step 2</u> If the user selects the "X" button, the user needs to upload the audio again and return to step 1. <u>A2. Step 3</u> If the audio format is invalid, the user must upload the audio again and return to step 1.	

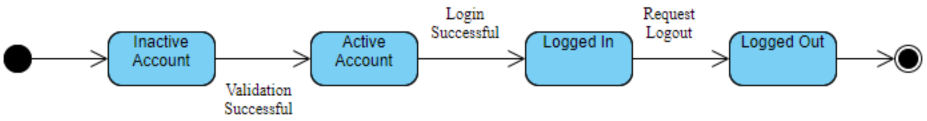
Table 4.4: Description Table of Analyze Sentiment of Audio

4.4 Behavioral Models

4.4.1 User Management Module

UserAccountClass

Figure 4.8: State Diagram of UserAccount Class of User Management Module

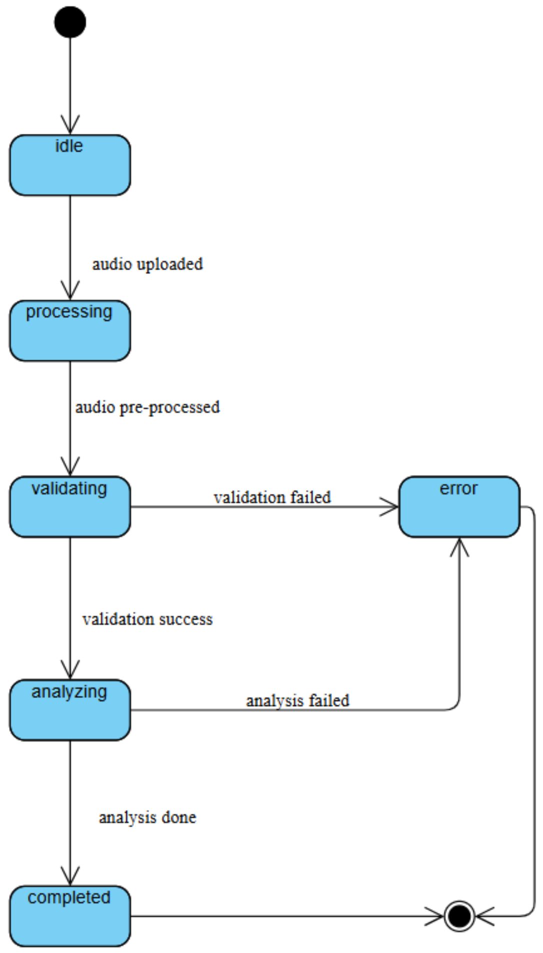


The state diagram of UserAccount Class is shown as above. The user must have an active account to login to the system. The user account will log out from the system if the user request to logout, then the process will end.

4.4.2 Audio Sentiment Analysis Module

SentimentAnalysis Class

Figure 4.9: State Diagram of SentimentAnalysis Class of Audio Sentiment Analysis Module

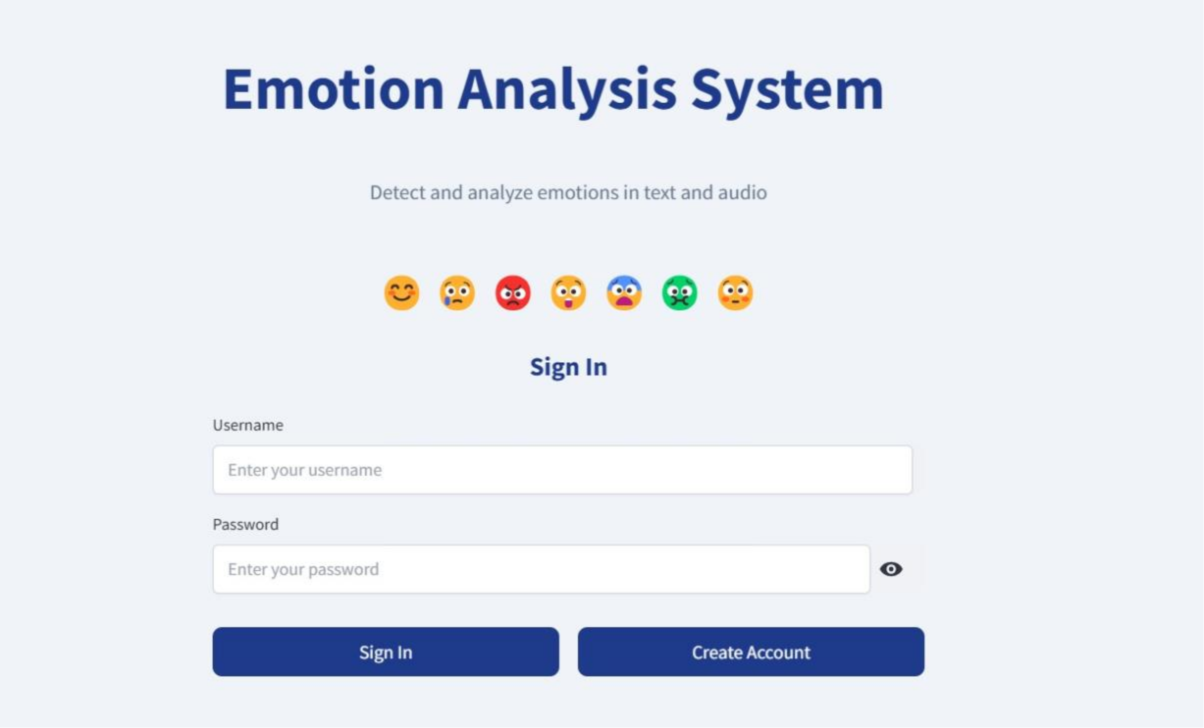


The above diagram is a state diagram of SentimentAnalysis Class, it shows the sate of the sentiment analysis. The audio needs to be uploaded in a correct format then the system only will analyze the audio. After the analysing is done, the state will become completed and end the process, else if the validation failed means that the audio is not upload in a correct format, it will become error and end the process. Not only that, it the analysis is failed, it will also become error and end the process.

4.5 Other Designs – UI Design

4.5.1 User Management Module

1. System Main Page



Emotion Analysis System

Detect and analyze emotions in text and audio

Sign In

Username

Enter your username

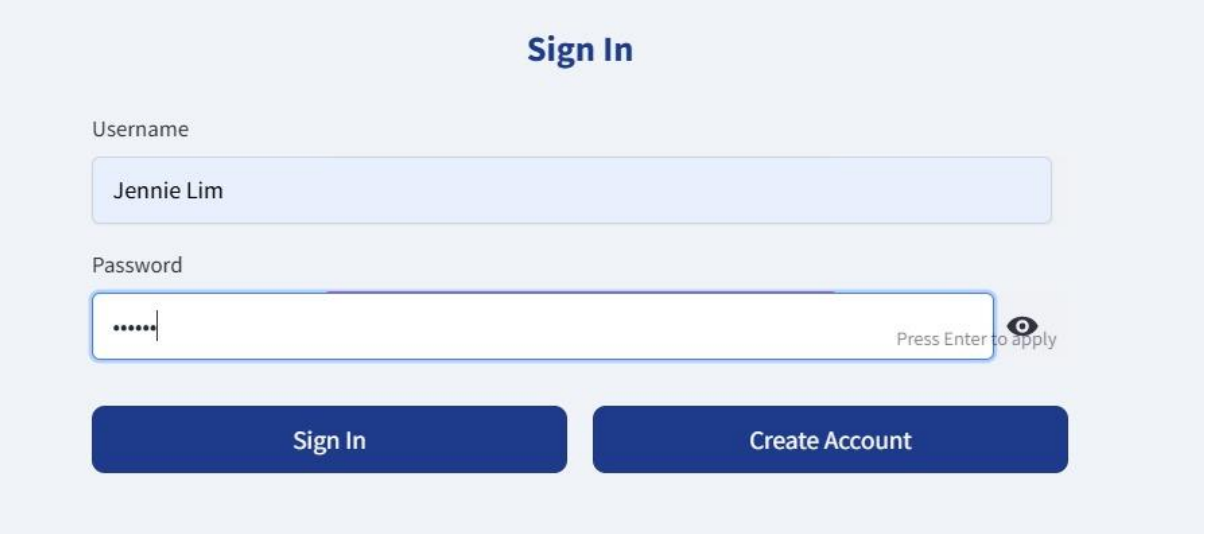
Password

Enter your password

Figure 4.10: System Main Page – A main page that allows user to login or sign up

The system's main page is shown as above, it will display the default sign-in page and 2 buttons which are “Sign In” and “Create Account” that allow the user to choose whether they want to login or sign up a new account.

2. Login Page



Sign In

Username

Jennie Lim

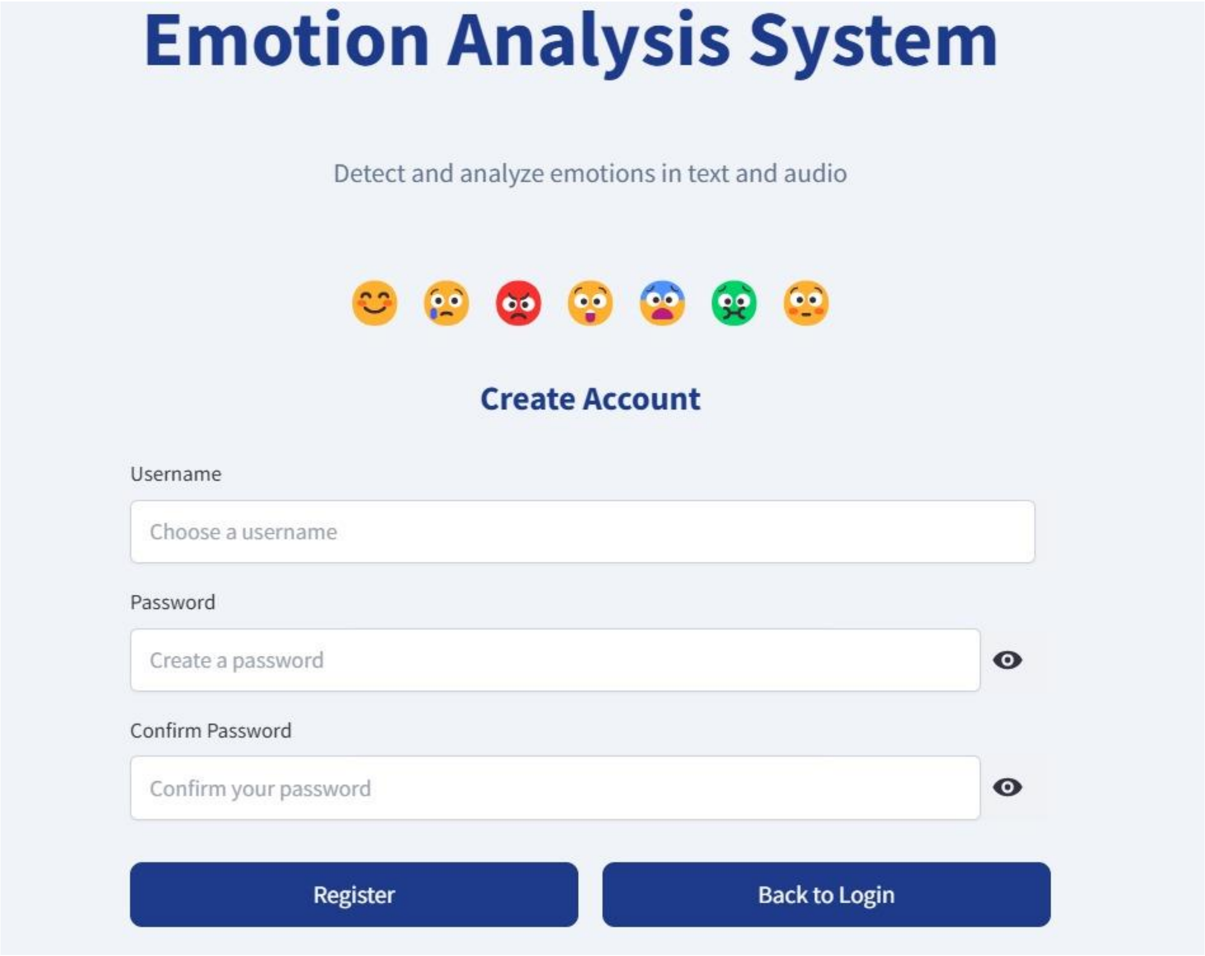
Password

••••••

Figure 4.11: Login Page – A page that allows user to login

The UI design of the Login page is shown above. The user needs to enter the correct username and password that match the database to log in to their account. If the user enters the correct username and password, the system will bring the user to the sentiment analysis page.

3. Sign Up Page



Emotion Analysis System

Detect and analyze emotions in text and audio

😊 😢 😡 😲 😨 🤢 😳

Create Account

Username

Choose a username

Password

Create a password

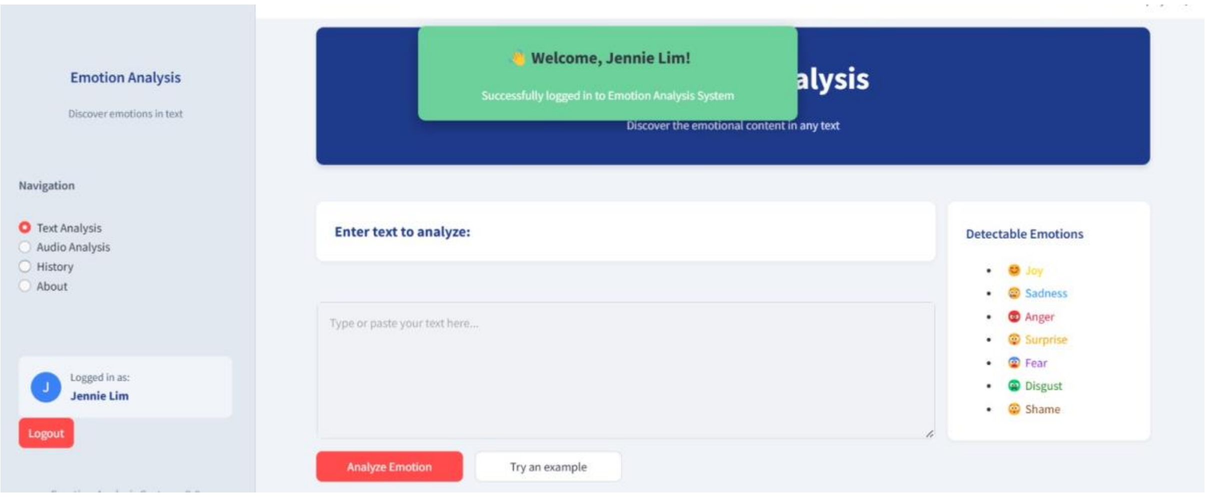
Confirm Password

Confirm your password

Figure 4.12: Sign Up Page – A page that allows user to sign up

The sign up page is designed as above, the user is required to fill in their details, including username, password, and confirm password. The username must be a unique username that does not have the same username as the database, else the user is required to enter another username. For the password and confirm password, these 2 fields must be the same, else it will display the error message to notify the user. The "Register" and "Back to Login" button also placed at the right bottom corner as a consistent.

4. Main Dashboard


👋 Welcome, Jennie Lim!

Successfully logged in to Emotion Analysis System

Discover the emotional content in any text

Enter text to analyze:

Type or paste your text here...

Logged in as: Jennie Lim

Figure 4.13: Main Dashboard Page – A page that allows user to choose the navigation

The main dashboard is shown as above, a welcome message will show as the user is login successfully. Furthermore, the user is allowed to choose the function from the navigation whether they want to use the text analysis or audio analysis to do the sentiment analysis.

4.5.2 Audio Sentiment Analysis Module

1. Subsystem Main Page

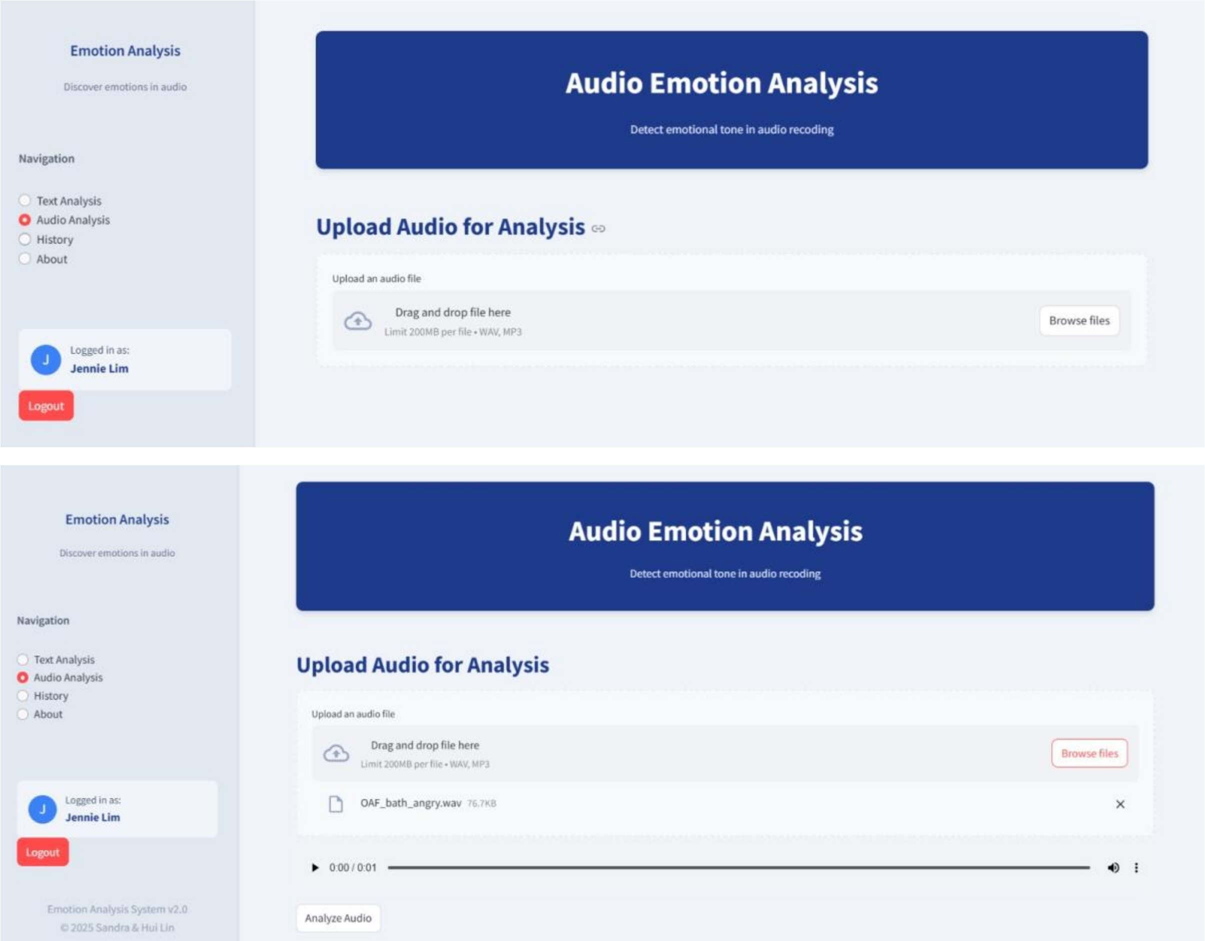
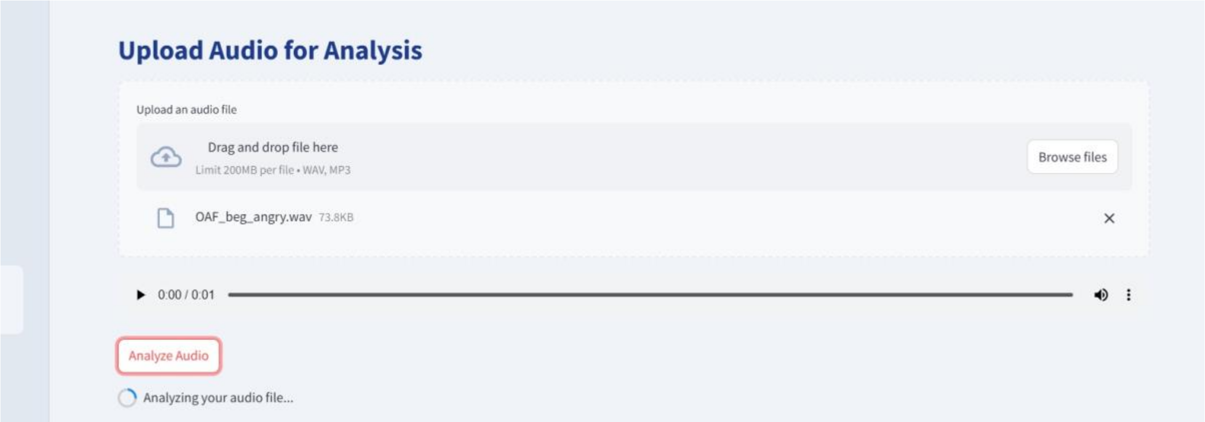


Figure 4.13: Subsystem Main Page – Allows user to upload the audio

It is a subsystem main page that allows users to upload audio. The user allows to drag and drop the audio files or browse the files to upload the audio. After the audio is uploaded, the user is required to click the "Analyze Audio" button to the next stage of progress, while the "X" button allows the user to remove the audio if they want to change another audio or cancel uploading.

2. Processing Page



Upload Audio for Analysis

Upload an audio file

Drag and drop file here

Limit 200MB per file • WAV, MP3

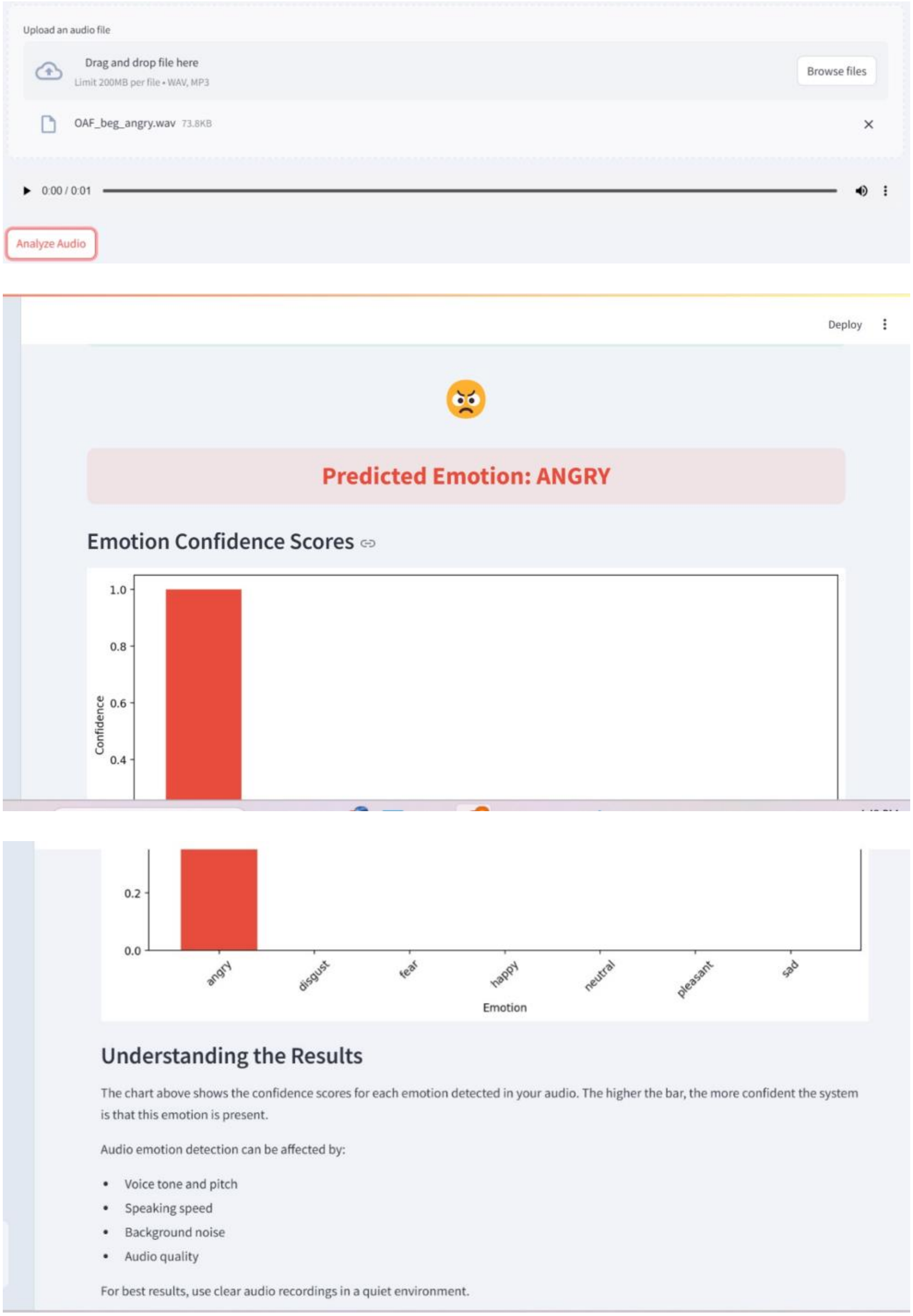
OAF_beg_angry.wav 73.8KB

Analyzing your audio file...

Figure 4.14: Processing Page – A page that shows the audio is being analyzed

The processing page of the audio sentiment analysis is shown as above. The brief description of the audio sentiment analysis and the audio file field will still remain on the page, while the process information or the process status will display at the below of the audio file in order to let the user know about the process.

3. Result Page



Predicted Emotion: ANGRY

Emotion Confidence Scores

Understanding the Results

The chart above shows the confidence scores for each emotion detected in your audio. The higher the bar, the more confident the system is that this emotion is present.

Audio emotion detection can be affected by:

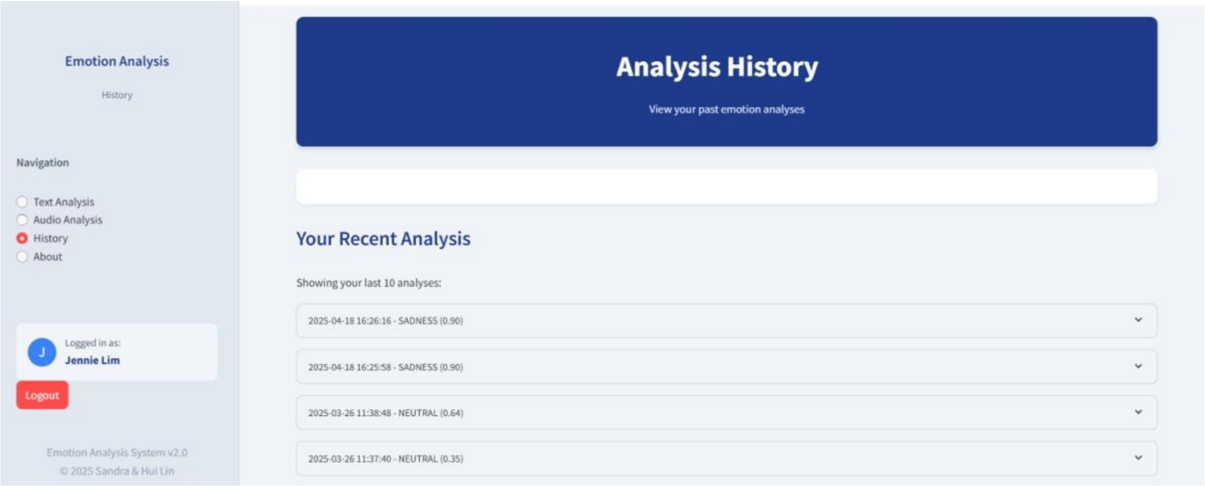
- Voice tone and pitch
- Speaking speed
- Background noise
- Audio quality

For best results, use clear audio recordings in a quiet environment.

Figure 4.15: Result Page – A page that shows the analyse result of the audio

The result page of the audio sentiment analysis is shown as above. A picture will be displayed to show the result of the analysis whether is angry, sad, fear or many more. In addition, an emotion confidence graph that show the result will also be displayed. As the user clicks the “X” button, the user is able to upload another audio to do the analysis again.

4.5.3 History Page



Emotion Analysis

History

Navigation

○ Text Analysis
○ Audio Analysis
● History
○ About

Logged in as:
Jennie Lim

Logout

Emotion Analysis System v2.0
© 2025 Sandra & Hui Lin

Analysis History

View your past emotion analyses

Your Recent Analysis

Showing your last 10 analyses:

2025-04-18 16:26:16 - SADNESS (0.90)

2025-04-18 16:25:58 - SADNESS (0.90)

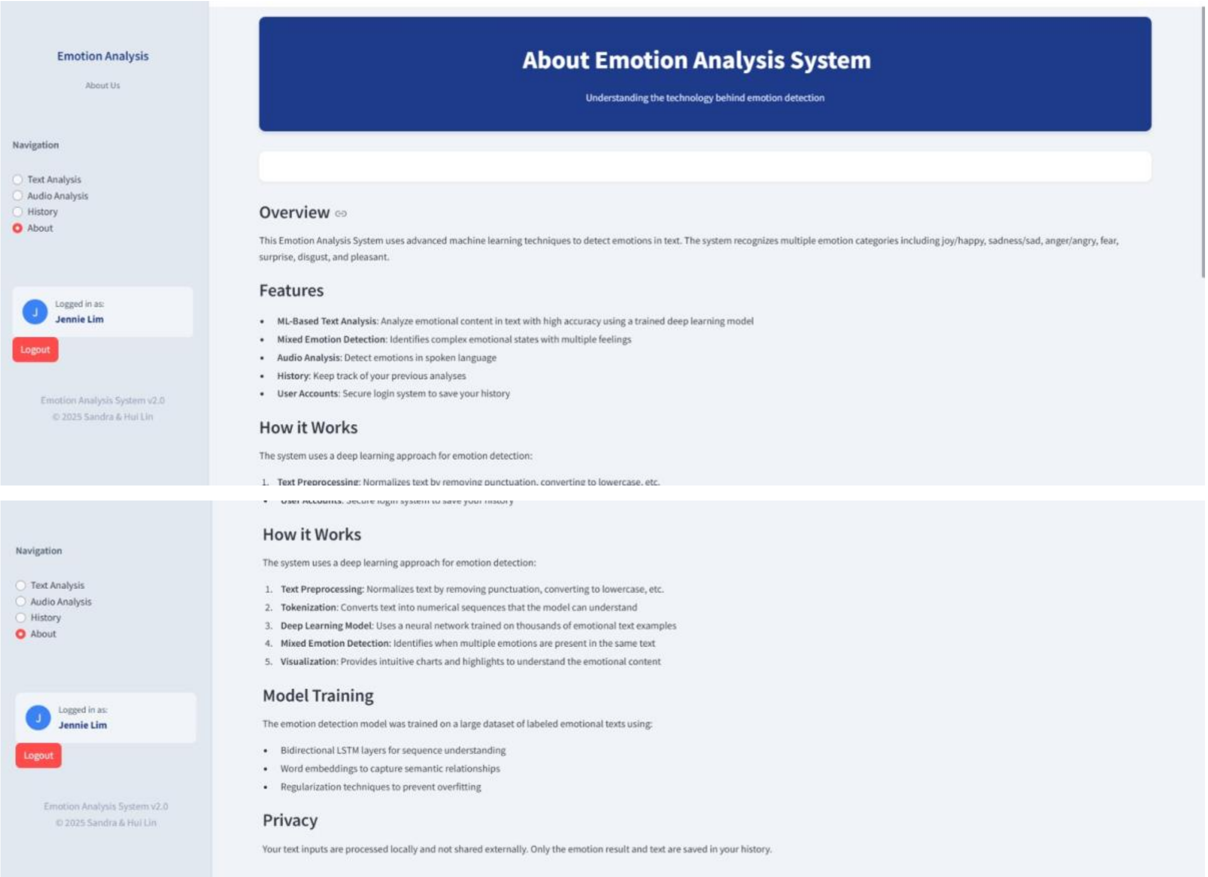
2025-03-26 11:38:48 - NEUTRAL (0.64)

2025-03-26 11:37:40 - NEUTRAL (0.35)

Figure 4.16: History Page – A page that shows the analysis history result of the audio

The history page allows the user to check the history of the analysis of the text or audio sentiment analysis. The history will show the date and time, type of analysis, like text or audio, sentiment analysis result. The history page will store and display the latest 10 results of the analysis.

4.5.4 About Us Page



Emotion Analysis

About Us

Navigation

○ Text Analysis
○ Audio Analysis
○ History
● About

Logged in as:
Jennie Lim

Logout

Emotion Analysis System v2.0
© 2025 Sandra & Hui Lin

About Emotion Analysis System

Understanding the technology behind emotion detection

Overview

This Emotion Analysis System uses advanced machine learning techniques to detect emotions in text. The system recognizes multiple emotion categories including joy/happy, sadness/sad, anger/angry, fear, surprise, disgust, and pleasant.

Features

- **ML-Based Text Analysis**: Analyze emotional content in text with high accuracy using a trained deep learning model
- **Mixed Emotion Detection**: Identifies complex emotional states with multiple feelings
- **Audio Analysis**: Detect emotions in spoken language
- **History**: Keep track of your previous analyses
- **User Accounts**: Secure login system to save your history

How it Works

The system uses a deep learning approach for emotion detection:

1. **Text Preprocessing**: Normalizes text by removing punctuation, converting to lowercase, etc.

Navigation

○ Text Analysis
○ Audio Analysis
○ History
● About

Logged in as:
Jennie Lim

Logout

Emotion Analysis System v2.0
© 2025 Sandra & Hui Lin

- **User Accounts**: Secure login system to save your history

How it Works

The system uses a deep learning approach for emotion detection:

1. **Text Preprocessing**: Normalizes text by removing punctuation, converting to lowercase, etc.

2. **Tokenization**: Converts text into numerical sequences that the model can understand
3. **Deep Learning Model**: Uses a neural network trained on thousands of emotional text examples
4. **Mixed Emotion Detection**: Identifies when multiple emotions are present in the same text
5. **Visualization**: Provides intuitive charts and highlights to understand the emotional content

Model Training

The emotion detection model was trained on a large dataset of labeled emotional texts using:

- Bidirectional LSTM layers for sequence understanding
- Word embeddings to capture semantic relationships
- Regularization techniques to prevent overfitting

Privacy

Your text inputs are processed locally and not shared externally. Only the emotion result and text are saved in your history.

Figure 4.17: About Page – A page that shows the information of the system

The about page shows some of the information of the system such as overview, features, how it works, model training, and also privacy.

4.6 Chapter Summary and Evaluation

The process design, database structures, interaction models, and behavioral models are all covered in this chapter's overview of the audio sentiment analysis system's architecture. Spectrogram extraction and audio normalization are important procedures that get audio data ready for examination. System elements like UserAccount, Audio, AudioProcessing, and SentimentAnalysis are defined in the Class Diagram and ERD, along with their connections. Use cases for audio sentiment analysis such as uploading audio, verifying files, and analyzing sentiment and user management includes creating and maintaining user accounts are described in depth in the Interaction Models. System states for user activities and sentiment analysis procedures are depicted by state diagrams. System structure and user-friendly interfaces are the main topics of other designs, such as architectural and user interface design.

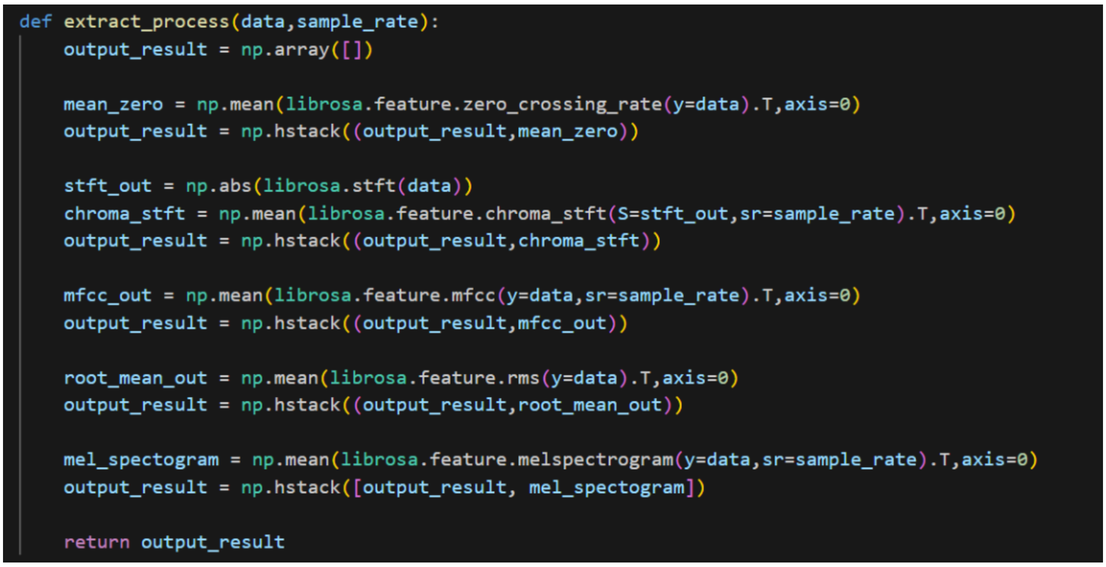
Chapter 5

Implementation and Testing

5 Implementation and Testing

5.1 Implementation

5.1.1 Feature Extraction Process

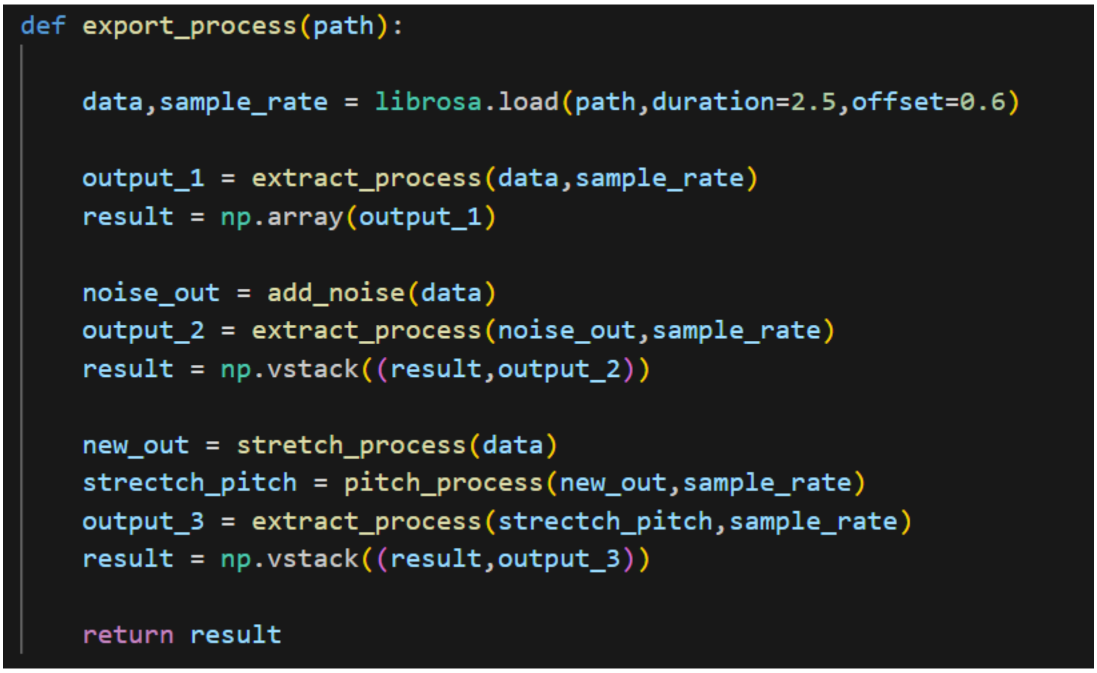


```
def extract_process(data,sample_rate):
    output_result = np.array([])

    mean_zero = np.mean(librosa.feature.zero_crossing_rate(y=data).T,axis=0)
    output_result = np.hstack((output_result,mean_zero))

    stft_out = np.abs(librosa.stft(data))
    chroma_stft = np.mean(librosa.feature.chroma_stft(S=stft_out,sr=sample_rate).T,axis=0)
    output_result = np.hstack((output_result,chroma_stft))

    mfcc_out = np.mean(librosa.feature.mfcc(y=data,sr=sample_rate).T,axis=0)
    output_result = np.hstack((output_result,mfcc_out))

    root_mean_out = np.mean(librosa.feature.rms(y=data).T,axis=0)
    output_result = np.hstack((output_result,root_mean_out))

    mel_spectogram = np.mean(librosa.feature.melspectrogram(y=data,sr=sample_rate).T,axis=0)
    output_result = np.hstack([output_result, mel_spectogram])

    return output_result
```

The function extracts certain audio features that are important for classifying the emotions. A feature vector is created by the frequency at which the signal changes from positive to negative, known as the zero-crossing rate. Furthermore, extracting the chromogram features from magnitude Short-Time Fourier Transform, calculating the short-term sound power spectrums known as Mel-Frequency Cepstral Coefficients (MFCCs). Other than that, it also enhances the loudness by appending root mean square energy and also adds the characteristics to the Mel spectrogram that show the energy at different frequencies.

In order to provide a complete representation of the qualities of the audio signal that may be utilized for emotional content recognition, the role combines these several acoustic characteristics into a single array.

5.1.2 Data Augmentation Pipeline



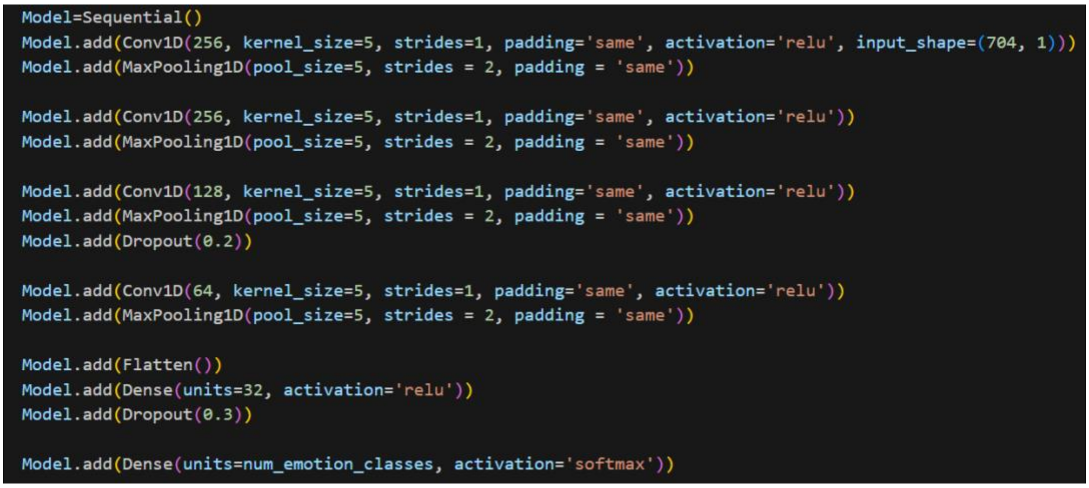
```
def export_process(path):

    data,sample_rate = librosa.load(path,duration=2.5,offset=0.6)

    output_1 = extract_process(data,sample_rate)
    result = np.array(output_1)

    noise_out = add_noise(data)
    output_2 = extract_process(noise_out,sample_rate)
    result = np.vstack((result,output_2))

    new_out = stretch_process(data)
    strectch_pitch = pitch_process(new_out,sample_rate)
    output_3 = extract_process(strectch_pitch,sample_rate)
    result = np.vstack((result,output_3))

    return result
```

This function implements into practice an advanced process for data augmentation that loads a 2.5s audio sample starting 0.6s into the file, extracts the features from the original audio, creates a noisy version of the audio, and extracts its features. Besides, it also creates a time-stretched version with modified pitch and extracts the features, and stacks all the feature sets into a unified result.

By subjecting the model to variants of the same emotional content with distinct acoustic features, the function boosts model resilience and triples the effective dataset size. It is essential for enhancing generalisation across various speakers and recording scenarios.

The initial implementation had a more basic feature extraction process without this comprehensive augmentation pipeline. Three variants of each audio sample are produced by this improved version, which is the original, noise-augmented, and time-stretched with pitch change. Next, it also produces more varied training data by sequentially combining many augmentation approaches like pitch shifting after stretching. The function essentially triples the training data from the same audio recordings by returning a stacked array of all three feature sets. Furthermore, it tackles the widespread issue of insufficient training data for emotion identification. It also increases the model's resilience to various acoustic circumstances.

5.1.3 CNN Model Architecture for Audio Analysis

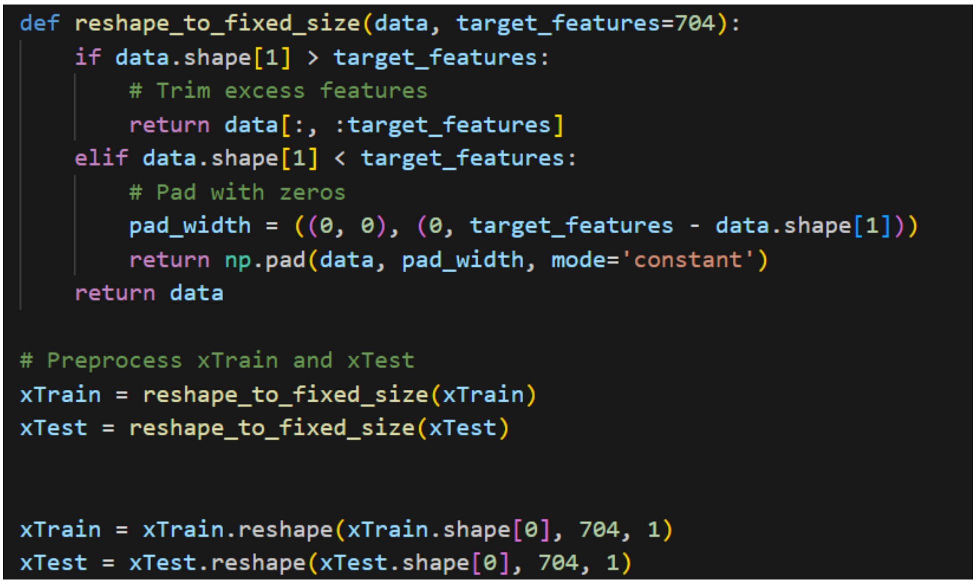


```
Model=Sequential()
Model.add(Conv1D(256, kernel_size=5, strides=1, padding='same', activation='relu', input_shape=(704, 1)))
Model.add(MaxPooling1D(pool_size=5, strides = 2, padding = 'same'))

Model.add(Conv1D(256, kernel_size=5, strides=1, padding='same', activation='relu'))
Model.add(MaxPooling1D(pool_size=5, strides = 2, padding = 'same'))

Model.add(Conv1D(128, kernel_size=5, strides=1, padding='same', activation='relu'))
Model.add(MaxPooling1D(pool_size=5, strides = 2, padding = 'same'))
Model.add(Dropout(0.2))

Model.add(Conv1D(64, kernel_size=5, strides=1, padding='same', activation='relu'))
Model.add(MaxPooling1D(pool_size=5, strides = 2, padding = 'same'))

Model.add(Flatten())
Model.add(Dense(units=32, activation='relu'))
Model.add(Dropout(0.3))

Model.add(Dense(units=num_emotion_classes, activation='softmax'))
```

This segment defines a 1D Convolutional Neural Network architecture specifically designed for audio emotion classifications. It reduces the dimensions from 256 to 64 filters using a progressive filtering technique with four Convolutional 1D layers. In order to maintain the temporal information, each layer has a kernel size of 5 and the same amount of padding. For the MaxPooling layers, shorten sequences without compromising the crucial characteristics. At the same time, the 0.2 and 0.3 dropout layers guard against overfitting. The emotion probabilities are generated via a final dense layer with SoftMax activation.

Because of the computational efficiency and the ability to capture both local and global patterns in the sequential audio features, this architecture works especially well for audio.

5.1.4 Data Processing and Reshaping Logic



```python
def reshape_to_fixed_size(data, target_features=704):
    if data.shape[1] > target_features:
        # Trim excess features
        return data[:, :target_features]
    elif data.shape[1] < target_features:
        # Pad with zeros
        pad_width = ((0, 0), (0, target_features - data.shape[1]))
        return np.pad(data, pad_width, mode='constant')
    return data

# Preprocess xTrain and xTest
xTrain = reshape_to_fixed_size(xTrain)
xTest = reshape_to_fixed_size(xTest)


xTrain = xTrain.reshape(xTrain.shape[0], 704, 1)
xTest = xTest.reshape(xTest.shape[0], 704, 1)
```

Standardizing the feature dimensions is a fundamental data processing activity that is covered in this part. The purpose of this is to verify whether the feature vector is larger than the desired size, which is 704, and to truncate it if necessary. It also maintains the model input data dimensions constant by padding shorter vectors with zeros to meet the target length.

Upon the standardization of the feature lengths, the algorithm transforms data into the samples, timesteps, and features that are needed by the Convolutional 1D layers. In the real world of audio processing, where the recordings may vary in length or feature extraction may provide varied dimensions, this preprocessing guarantees that the model can effectively handle variable-length audio inputs.

For the reshaping logic, the initial version assumed the consistent feature dimensions; if the feature extraction yielded variable-length vectors, it would fail. With the improvement, the feature vectors of any size may be handled by an intelligent reshaping algorithm that either removes the extra elements that go beyond the desired size, which is 704, else, if they are less than the necessary length, it will be padding with zeros. It allows a greater variety of audio inputs, including ones with varying durations or quality, to be used by the model, and the runtime mistakes that arise when feature dimensions do not match the model expectations are eliminated.

5.2 Testing

5.2.1 Test Plan

Overview

The system that will be approached is the emotion detection sentiment analysis system, which involves text sentiment analysis and audio sentiment analysis. Using extracted acoustic characteristics and data augmentation techniques, the strategy focuses on verifying that the machine learning model can correctly categorize audio samples into emotional categories such as angry, disgusted, fear, happy, neutral, sad, and pleasant. Testing will guarantee that the model satisfies accuracy standards, operates dependably in the intended setting, and works consistently across various audio sources.

Context of Testing

The testing shall be conducted within a controlled development environment prior to deployment with attention to both system-level and component-level validation. The feature extraction pipeline, data augmentation processes, model training, and ultimate classification accuracy shall be tested.

Project

The Audio Sentiment Analysis System is a machine learning program that uses 1D Convolutional Neural Networks to analyse audio recordings and identify emotional states. The TESS Toronto emotional speech dataset is used to classify emotions after the algorithm extracts acoustic characteristics from audio recordings and enhances the training data.

Test Item

Component	Description
Feature Extraction (extract_process)	Functions that extract acoustic features from audio, including zero-crossing rate, chromogram, MFCCs, RMS energy, and mel spectrograms
Data Augmentation Pipeline (add_noice, stretch_process, shift_process, pitch_process, export_process)	A collection of routines that use pitch alteration, time stretching, shifting, and noise injection to produce different audio samples
CNN Model Architecture	An arrangement of many Conv1D layers, MaxPooling, Dropout, and Dense layers in a 1D Convolutional Neural Network
Data processing and Reshaping Logic (reshape_to_fixed_size)	Functions that guarantee model compatibility by standardising feature dimensions by padding or clipping
Model Training Process	Model fitting with early halting, optimiser setup, and loss function implementation
Model Evaluation Metrics	Confusion matrix, classification report creation, and accuracy evaluation
Saved Model Artifacts (audio_sentiment_model.keras, encoder_label.pkl, scaler_data,pkl)	Label encoders, data scalers, and the trained model are stored indefinitely

Test Scope

1. Functional Testing

Functional testing is to verify that the system components perform their intended function. It will test whether the software functions as it is expected and ensure that the system behaves in a correct way based on the requirements. The main objectives of the tester are to make sure that the software functions as intended while adhering to the specifications listed in the document instead of study and understand the code. The testers will be independent during functional testing as they won't be swayed by the inner workings of the program. Since the testing doesn't require coding expertise, it is also simpler for testers to complete. For example, if the user wants to check the history of the analysis, the system should display the history to allow the user to check on the history. The testing should make sure that the process of the system is smooth, and it is essential to guarantee that every function is able to operate without error.

2. Performance Testing

Performance testing allows to assess how well the functions in various scenarios. It considers factors such as responsiveness, stability, growth-handling capabilities, dependability, and system resource consumption. Performance testing determines whether the system operates effectively and smoothly, while functional testing determines is the system accomplishes what it should. Performance testing is crucial in the system because it determines if a system meets the performance requirements before deployment such as the time taken to complete a request, and identifies the system architecture's weak areas and performance limitations.

3. Accuracy Testing

Accuracy testing is a quality check that makes sure that the system is precisely doing its function. It looks at whether the system is handling data and the result correctly as the expectation. In simple terms, it is about checking if the system is doing the calculation properly, showing this correct information, or making the right decisions based on the input. It might help to catch the mistakes or errors early, hence the final product can be trusted to work accurately. For instance, the precision, recall, F1 score, accuracy, and confusion matrix which are used to determine the accuracy. The methodology that was used to test the accuracy is defined accuracy requirements, prepared test data, designed test cases, execute test, and so on.

Assumptions and Constraints

1. Having sufficient computational resources for model training and evaluation
2. Testing will be conducted on the pre-recorded audio files instead of live audio
3. Maximum audio length constraints of 2.5s per sample
4. The testing environment has the necessary libraries, which is the librosa, tensorflow, sklearn, and many more.

Risk Register

Product Risk

1. **Insufficient Model Accuracy**

   Risk

   For accurate emotion recognition, the 80% accuracy criterion is essential. The users cannot rely on the system's classification of the emotion such as angry, happy, fear, and so on if it falls under of this standard.

   Management

   To manage the system to better accuracy, it may make iterative changes to the model architecture such as LSTM and many more, hyperparameter modifications like learning rate, layer depth and so on, and feature selection which include spectral, linguistic and others. Other than that, it can also implement validation metrics beyond the simple accuracy, like F1 Score and confusion matrix.

2. **Overfitting**

   Risk

   The model performs well on training data but poorly on the new audio. Furthermore, instead of learning the generalisable aspects of emotional speech, the model memorizes the training data pattern.

   Management

   The method to manage the overfitting is to adjust dropout rates to prevent neurons from becoming too specialized, use cross-validation to ensure the model performance is consistent across different data subsets. Moreover, it improves data augmentation, like adding noise, pitch shifting, and time stretching to improve the generalization.

3. **Feature Extraction Inconsistency**

   Risk

   The audio feature extraction produces inconsistent results across different samples. Even when the emotional content of audio signals stays the same, the complex information they convey can be significantly altered by a variety of variables.

   Management

   Normalise features to make sure all the features are on a comparable scale. Other than that, standardise audio preprocessing such as filtering, framing are applied consistently, and the robust feature selection methods to identify features that remain stable across conditions.

Project Risk

1. **Computational Resource Limitations**

   Risk

   Deep learning models and audio processing require a lot of computing power. If there is not enough memory or processing capacity, it can slow things down and make the training take much longer. In some cases, it might mean having to use simpler models or not being able to handle large amounts of data at all.

   Management

   Optimise the feature extraction workflow to reduce the dimensions before the model training and use batch processing, which allows handling larger datasets with limited memory. Next, the potential use of cloud resources for intensive training phases is also be the one management solution for the computational resource limitation.

2. **Library Compatibility Issues**

   Risk

   Conflicts between different versions of required libraries for audio processing and machine learning can cause inconsistent behavior, integration errors or prevent deployment in production environments.

   Management

   Use containerization such as Docker to create isolated environments with specific library versions. Document all dependencies clearly, and use version pinning to ensure the environment remains consistent and reproducible across development and deployment.

3. **Data Quality Issue**

   Risk

   Emotion classification depends heavily on high-quality, balanced audio data. If the training data is unbalanced or contains low-quality samples, the model may become biased, like it performing well on common emotions but poorly on rare ones.

   Management

   Apply class balancing techniques to ensure all emotions are fairly represented. Use the data augmentation to generate more samples for underrepresented emotions and implement the quality filtering to remove noisy or unstable audio samples from the dataset.

Test Strategy

Test Sub-Processes

The testing methodology follows an incremental approach where each component is validated separately before being included in the entire system as part of our progressive testing process. This approach creates a strong basis for the machine learning pipeline and permits early problem discovery. Using carefully chosen audio samples, feature extraction algorithms are first tested to ensure that acoustic properties, including MFCCs, zero-crossing rates, and spectral components, are accurately captured. Next, we examine how well data augmentation methods produce realistic audio sample changes while maintaining emotional content in order to verify them. The model stability depends on the consistent input dimension and appropriate normalization, which are the main goals of the data preparation pipeline assessment. Moreover, evaluation of model training focusses on convergence patterns, the optimisation process, and how early interrupting and appropriate regularisation prevent overfitting. Lastly, we evaluate the model's overall performance using a different test dataset to gauge its accuracy and generalization potential in the real world.

Test Deliverables

The comprehensive test cases that outline the inputs, processes, anticipated results, and pass or fail criteria for every component will be created. The correctness and consistency of the extracted audio characteristics across various sample types will be recorded in the Feature Extraction Validation Report. Using quantitative measurements and visualizations, the Data Augmentation Effectiveness Report will examine how augmentation approaches increase model robustness. Other than that, throughout the training process, the Model Training and Validation Report will monitor validation metrics, convergence trends, and the outcomes of hyperparameter optimization. Comprehensive analysis, including confusion matrices, class-specific metrics, and error analysis on difficult samples, will be included in the Final Model Performance Report. A Test Summary Report that summarises all of the findings, highlights the positives and negatives, and offers suggestions for future development will be created.

Test Design Techniques

The black box testing techniques for the audio sentiment analysis does not look at the core system activities, it only focusses on the relationship between the input and the output. Using the audio with exceptionally high or low quality, unusually short or long durations, and acoustic characteristics at the edges of normal human speech like whispers, shouts, or extremely fast or slow speech, the testers purposefully choose the audio samples that push the boundaries od the acceptable parameters when using boundary value analysis. The systematic testing of boundary conditions helps to identify where the system performance might deteriorate. In addition, equivalence partitioning separates the enormous array of the potential audio inputs into logical group according to the speaker attributes, acoustic surroundings, and emotional content. Hence, the testers can confirm the system behaviour effectively without thoroughly testing every potential input by evaluating representative samples from each division, such as speakers expressing the same emotion at different intensities. This well-rounded strategy manages the infinitely varied range of human emotional expression in audio form whole guaranteeing comprehensive coverage of the system’s expression.

Test Completion Criteria

1. All the test cases are executed at least once, and the results are documented.
2. Minimum 95% of test cases are passed
3. Maximum 5% of test cases are failed
4. Minimum 80% validation accuracy threshold
5. At least one test cycle is completed

Metrics to be Collected

1. Test management metrics include counts of planned, executed, passed, and failed test cases to track the testing progress and coverage
2. Accuracy on the validation set, which represents overall classification correctness, will be the primary metrics used to access performance
3. For each emotion category, per-class precision, recall, and F1-scores will enable fine-grained analysis.
4. Confusion matrix distributions will be used to visualize misclassification trends between emotion pairings to assist with discovering comparable emotions that the algorithm finds difficult to distinguish.
5. To track convergence and identify any possible overfitting problems, training and validation loss curves will be displayed.
6. The model convergence rate, which maximizes the training time and hyperparameters has been monitored.
7. Feature extraction timing metrics to identify the preprocessing pipeline's performance.

Test Data Requirements

Module	Function	Data	Data Type	Description	Sample
Account	User Sign Up	Username	string	Valid: The username contains the alphabet and number.	cookie11
				Valid: The username contains alphabet only.	cookie
				Invalid: The username contains number only.	123
				Invalid: The username contains special characters only.	###
				Invalid: The username field is empty.	Null
		Password	string	Valid: The password contains the alphabet and number.	cookie123
				Valid: The password contains alphabet only.	cookie
				Valid: The password contains number only.	123456
				Invalid: The password is less than 6 characters.	1234
				Invalid: The password field is empty.	Null
		Confirm Password	string	Valid: The re-type password is the same as the password.	cookie123
				Invalid: The re-type password is not the same as the password.	cookie12
				Invalid: The confirm password field is empty.	Null
	User Login	Username	string	Valid: The username matches the database.	cookie11
				Invalid: The username does not match the database.	cookie1
		Password	string	Valid: The password matches the database.	cookie123
				Invalid: The password does not match the database.	cookie12

Audio Sentiment Analysis	Audio Upload	Audio File	file	Valid: The audio file was uploaded in the correct format.	audio.mp3
				Invalid: The audio file was not uploaded in the correct format.	video.mp4

Test Environment Requirement

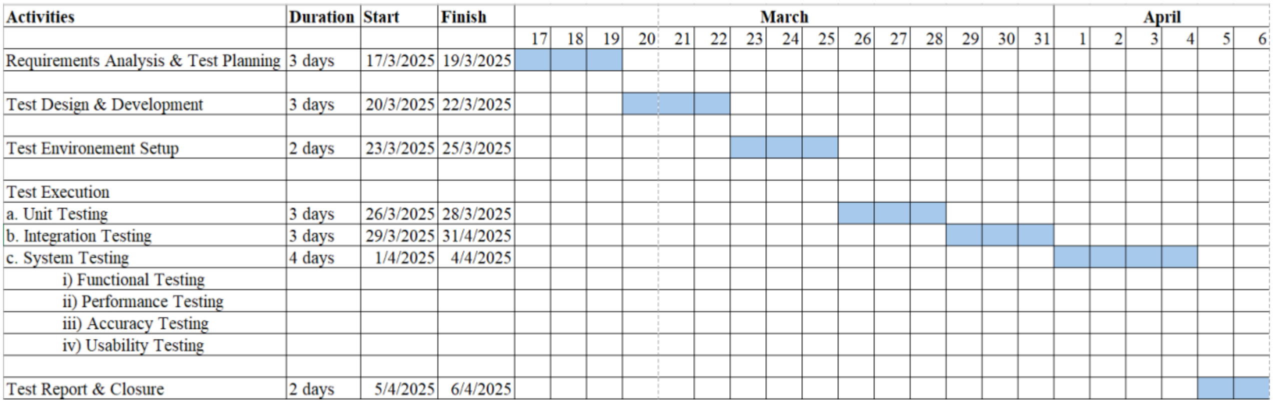
Hardware

1. 2 laptops with a processor of Intel i5 or above, 8GB of RAM or above, and 256GB SSD storage
2. PC with processor of Intel i5 or above, 8GB of RAM or above, and 500GB SSD storage

Software

1. Python version 3.10 and above
2. TensorFlow
3. Librosa library
4. Jupyter Notebook
5. Visual Studio Code
6. Scikit-learn
7. Numpy
8. Pandas

Test Activities and Estimates



Activities	Duration	Start	Finish	March															April					
				17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	1	2	3	4	5	6
Requirements Analysis & Test Planning	3 days	17/3/2025	19/3/2025	🟦	🟦	🟦																		
Test Design & Development	3 days	20/3/2025	22/3/2025				🟦	🟦	🟦															
Test Environement Setup	2 days	23/3/2025	25/3/2025							🟦	🟦	🟦												
Test Execution																								
a. Unit Testing	3 days	26/3/2025	28/3/2025										🟦	🟦	🟦									
b. Integration Testing	3 days	29/3/2025	31/4/2025													🟦	🟦	🟦						
c. System Testing	4 days	1/4/2025	4/4/2025																🟦	🟦	🟦	🟦		
i) Functional Testing																								
ii) Performance Testing																								
iii) Accuracy Testing																								
iv) Usability Testing																								
Test Report & Closure	2 days	5/4/2025	6/4/2025																				🟦	🟦

5.2.2 Test Cases

Overview

The test cases template is provided, and it has been required to follow the template to create the test case in order to test the functionality of the system to make sure that the system is run as expected. Each of the test cases includes the test case ID, name, system, and module being tested. Moreover, it also includes details such as who designed the test case, the date of the test case designed, who executed it, and the execution date. Pre-conditions are also involved in the test case, which indicate the requirements that must be met before the test can be executed. Next will be the step-by-step procedures, like the action to be performed, expected response by the system, pass or fail status, and comments. Lastly, the post-conditions used to describe the expected state or action performed by the system after the test is completed. In addition to providing a record of test results that can be consulted during the development and quality assurance processes, this organised method guarantees consistency in testing and explicit documentation of test procedures.

Test Cases

Test Case Template	
Test Case #: TC_AccountModule_UserSignUp_001	**Test Case Name: User Sign Up**
System: Emotion Detection Sentiment Analysis	**Module: Account Module**
Design By: Sandra Tang Poh Yi	**Design Date: 20/3/2025**
Executed By: Saw Hui Lin	**Execution Date: 22/3/2025**
Short Description: Test the user sign up with a valid username, password, and confirm password.	

Pre-conditions: The user does not have an account.

Step	Action	Expected System Response	Pass/Fail	Comments
1	Open the emotion detection sentiment analysis system	System shows the sign in page.	Pass	
2	Click “Create Account” button	System navigates to the sign up page.	Pass	
3	Enter username: june11			
4	Enter password: june123			
5	Enter confirm password: june123			
6	Click “Register” button	System shows the message that the account was created successfully.	Pass	

Post-conditions: The system navigates to the sign in page.

Test Case Template

Test Case #: TC_AccountModule_UserSignUp_002	**Test Case Name: User sign up**
System: Emotion Detection Sentiment Analysis	**Module: Account Module**
Design By: Sandra Tang Poh Yi	**Design Date: 20/3/2025**
Executed By: Saw Hui Lin	**Execution Date: 22/3/2025**
Short Description: Test the user sign up with an invalid username, valid password, and confirmed password.	

Pre-conditions: The user does not have an account.

Step	Action	Expected System Response	Pass/Fail	Comments
1	Open the emotion detection sentiment analysis system	System shows the sign in page.	Pass	
2	Click “Create Account” button	System navigates to the sign up page.	Pass	
3	Enter username:			
4	Enter password: june123			
5	Enter confirm password: june123			
6	Click “Register” button	System prompts an error message: “Please fill in all fields”.	Pass	

Post-conditions: The system shows an error message, and the user is unable to sign up.

Test Case Template	
Test Case #: TC_AccountModule_UserSignUp_003	**Test Case Name: User sign up**
System: Emotion Detection Sentiment Analysis	**Module: Account Module**
Design By: Sandra Tang Poh Yi	**Design Date: 20/3/2025**
Executed By: Saw Hui Lin	**Execution Date: 22/3/2025**
Short Description: Test the user sign up with an invalid username, valid password, and confirmed password.	

Pre-conditions: The user does not have an account.

Step	Action	Expected System Response	Pass/Fail	Comments
1	Open the emotion detection sentiment analysis system	System shows the sign in page.	Pass	
2	Click “Create Account” button	System navigates to the sign up page.	Pass	
3	Enter username: 123			
4	Enter password: june123			
5	Enter confirm password: june123			
6	Click “Register” button	System prompts an error message: “Username must include letters”.	Pass	

Post-conditions: The system shows an error message, and the user is unable to sign up.

Test Case Template	
Test Case #: TC_AccountModule_UserSignUp_004	**Test Case Name: User sign up**
System: Emotion Detection Sentiment Analysis	**Module: Account Module**
Design By: Sandra Tang Poh Yi	**Design Date: 20/3/2025**
Executed By: Saw Hui Lin	**Execution Date: 22/3/2025**
Short Description: Test the user sign up with an invalid username, valid password, and confirmed password.	

Pre-conditions: The user does not have an account.

Step	Action	Expected System Response	Pass/Fail	Comments
1	Open the emotion detection sentiment analysis system	System shows the sign in page.	Pass	
2	Click “Create Account” button	System navigates to the sign-up page.	Pass	
3	Enter username: ###			
4	Enter password: june123			
5	Enter confirm password: june123			
6	Click “Register” button	System prompts an error message: “Username must contain only letters & numbers”.	Pass	

Post-conditions: The system shows an error message, and the user is unable to sign up.

Test Case Template	
Test Case #: TC_AccountModule_UserSignUp_005	**Test Case Name: User sign up**
System: Emotion Detection Sentiment Analysis	**Module: Account Module**
Design By: Sandra Tang Poh Yi	**Design Date: 20/3/2025**
Executed By: Saw Hui Lin	**Execution Date: 22/3/2035**
Short Description: Test the user sign up with valid username, and confirmed password and invalid password.	

Pre-conditions: The user does not have an account.

Step	Action	Expected System Response	Pass/Fail	Comments
1	Open the emotion detection sentiment analysis system	System shows the sign in page.	Pass	
2	Click “Create Account” button	System navigates to the sign up page.	Pass	
3	Enter username: june11			
4	Enter password:			
5	Enter confirm password: june123			
6	Click “Register” button	System prompts an error message: “Please fill in all fields”.	Pass	

Post-conditions: The system shows an error message, and the user is unable to sign up.

Test Case Template	
Test Case #: TC_AccountModule_UserSignUp_006	**Test Case Name: User sign up**
System: Emotion Detection Sentiment Analysis	**Module: Account Module**
Design By: Sandra Tang Poh Yi	**Design Date: 20/3/2025**
Executed By: Saw Hui Lin	**Execution Date: 22/3/2025**
Short Description: Test the user sign up with valid username, valid password and invalid confirmed password.	

Pre-conditions: The user does not have an account.

Step	Action	Expected System Response	Pass/Fail	Comments
1	Open the emotion detection sentiment analysis system	System shows the sign in page.	Pass	
2	Click “Create Account” button	System navigates to the sign up page.	Pass	
3	Enter username: june11			
4	Enter password: june123			
5	Enter confirm password:			
6	Click “Register” button	System prompts an error message: “Password does not match”.	Pass	

Post-conditions: The system displays an error message, and the user is unable to sign up.

Test Case Template	
Test Case #: TC_AccountModule_UserSignUp_007	**Test Case Name: User sign up**
System: Emotion Detection Sentiment Analysis	**Module: Account Module**
Design By: Sandra Tang Poh Yi	**Design Date: 20/3/2025**
Executed By: Saw Hui Lin	**Execution Date: 22/3/2025**
Short Description: Test the user sign up with a valid username, invalid password, and invalid confirm password.	

Pre-conditions: The user does not have an account.

Step	Action	Expected System Response	Pass/Fail	Comments
1	Open the emotion detection sentiment analysis system	System shows the sign in page.	Pass	
2	Click “Create Account” button	System navigates to the sign up page.	Pass	
3	Enter username: june11			
4	Enter password: 12345			
5	Enter confirm password: 12345			
6	Click “Register” button	System prompts an error message: “Password must be at least 6 characters long”.	Pass	

Post-conditions: The system displays an error message, and the user is unable to sign up.

Test Case Template	
Test Case #: TC_AccountModule_UserLogin_001	**Test Case Name: User Login**
System: Emotion Detection Sentiment Analysis	**Module: Account Module**
Design By: Sandra Tang Poh Yi	**Design Date: 20/3/2025**
Executed By: Saw Hui Lin	**Execution Date: 22/3/2025**
Short Description: Test the user login with a valid username and valid password.	

Pre-conditions: The user must register as a user of the system.

Step	Action	Expected System Response	Pass/Fail	Comments
1	Open the emotion detection sentiment analysis system	System shows the sign in page.	Pass	
2	Enter username: june11			
3	Enter password: june123			
4	Click “Sign In” button	System allows the user to log in.	Pass	

Post-conditions: The system directs the user to the text sentiment analysis page.

Test Case Template	
Test Case #: TC_AccountModule_UserLogin_002	**Test Case Name: User Login**
System: Emotion Detection Sentiment Analysis	**Module: Account Module**
Design By: Sandra Tang Poh Yi	**Design Date: 20/3/2025**
Executed By: Saw Hui Lin	**Execution Date: 22/3/2025**
Short Description: Test the user with an invalid username and valid password.	

Pre-conditions: The user must register as a user of the system.

Step	Action	Expected System Response	Pass/Fail	Comments
1	Open the emotion detection sentiment analysis system	System shows the sign in page.	Pass	
2	Enter username: may			
3	Enter password: june123			
4	Click “Sign In” button	System prompts an error message: “Invalid username or password”	Pass	

Post-conditions: The system prompts an error message and the user is unable to log in.

Test Case Template	
Test Case #: TC_AccountModule_UserLogin_003	**Test Case Name: User Login**
System: Emotion Detection Sentiment Analysis	**Module: Account Module**
Design By: Sandra Tang Poh Yi	**Design Date: 20/3/2025**
Executed By: Saw Hui Lin	**Execution Date: 22/3/2025**
Short Description: Test the user with a valid username and invalid password.	

Pre-conditions: The user must register as a user of the system.

Step	Action	Expected System Response	Pass/Fail	Comments
1	Open the emotion detection sentiment analysis system	System shows the sign in page.	Pass	
2	Enter username: june11			
3	Enter password: june			
4	Click “Sign In” button	System prompts an error message: “Invalid username or password”	Pass	

Post-conditions: The system prompts an error message and the user is unable to log in.

Test Case Template	
Test Case #: TC_AudioSentiment_FileValidation_001	**Test Case Name: Audio File Validation**
System: Emotion Detection Sentiment Analysis	**Module: Audio Sentiment Analysis Module**
Design By: Saw Hui Lin	**Design Date: 21/3/2025**
Executed By: Sandra Tang Poh Yi	**Execution Date: 22/3/2025**
Short Description: Test the user to upload the audio in the correct format.	

Pre-conditions: The user must log in to the system.

Step	Action	Expected System Response	Pass/Fail	Comments
1	Open the emotion detection sentiment analysis system	System shows the sign in page.	Pass	
2	Upload the audio in correct format(.wav/.mp3)	System receives the audio.	Pass	
3	Click “Analyze Audio” button	System starts to analyze the audio and display the result.	Pass	

Post-conditions: The system displays the result of the emotion and save it to the history.

Test Case Template	
Test Case #: TC_AudioSentiment_FileValidation_002	**Test Case Name: Audio File Validation**
System: Emotion Detection Sentiment Analysis	**Module: Audio Sentiment Analysis Module**
Design By: Saw Hui Lin	**Design Date: 21/3/2025**
Executed By: Sandra Tang Poh Yi	**Execution Date: 22/3/2025**
Short Description: Test the user to upload the audio in the wrong format.	

Pre-conditions: The user must log in to the system.

Step	Action	Expected System Response	Pass/Fail	Comments
1	Open the emotion detection sentiment analysis system	System shows the sign in page.	Pass	
2	Upload the audio in wrong format(.mp4)	System prompts the error message: “video/mp4 files is not allowed”	Pass	

Post-conditions: The system displays the error message, and the user is unable to see the result.

Chapter 6

Discussions and Conclusion

6 Discussions and Conclusion

6.1 Summary

This project developed an audio sentiment analysis system that could detect and classify human emotions using speech. The system applies deep learning based on the architecture of a Convolutional Neural Network (CNN) employing 1D convolutions to learn meaningful audio features and classify them into discrete emotional labels. The deployment is an end-to-end pipeline from audio preprocessing to model training and a minimalistic GUI where the user is able to upload and process the audio files, display the outputs, and maintain a history of their processes.

The project addressed the issue of speech emotion recognition by utilizing strong audio feature extraction using Librosa, data augmentation using noise injection, time stretching, and pitch shifting. Additionally, designing and training a CNN model specialized in audio classification, constructing an end-to-end graphical user interface with user administration, and merging the system elements to allow for real-time emotional analysis.

The final system is able to detect seven emotions, including sad, happy, angry, fear, disgust, neutral, and pleasant, at high precision, demonstrating the potential of deep learning techniques in emotion recognition for audio data.

6.2 Achievements

The project was successfully accomplished with accurate emotion classification, robust feature extraction, data augmentation, and many more. For the accurate emotion classification, the CNN model used attained over 85% accuracy on the test set, being effective in classifying emotional states from the audio input. In addition, it implements a robust feature extraction pipeline that extracts vital audio features such as zero-crossing, chroma features, Mel-Frequency Cepstral Coefficients (MFCCs), Mel spectrograms, and so on. Thirdly, the data augmentation enhanced the model's robustness through careful audio augmentation strategies that include applying random noise to simulate realistic environments, using time stretching to offset speech rate differences, and employing pitch shifting to handle voice quality variations.

Moreover, it also includes a user-friendly GUI using Streamlit, and the GUI supports functions such as user authentication and account management, audio file upload and analysis, emotion detection output visualization, and history tracking and analysis review. It also has a model deployment, like bundling the trained model together with the necessary preprocessing elements for seamless integration into the application. Last but not least, the real-time processing refactored the processing pipeline to enable near-real-time analysis of audio files uploaded by the users.

6.3 Contributions

The several significant contributions to the field of audio sentiment analysis. The implementation of a specialist 1D CNN architecture designed for audio feature processing demonstrates an effective approach for emotion recognition of speech signals. The structure of consecutive convolutional and pooling layers in the network performed extremely well to detect temporal patterns in audio features. Furthermore, the project designed a systematic way of extracting and combining multiple audio features into one feature, which gives a comprehensive description of emotional speech features. The system is generalizable to other audio classification tasks beyond emotion recognition.

By the use of advanced machine learning and ease of interface, the project showcases an end-to-end process from research to utilization. This end-to-end solution bridges the gap between practice and theory in models and deployment. The system's architecture allows deployment across different platforms, making emotion analysis more accessible to users with any level of technical proficiency. Additionally, the codebase and implementation are a valuable asset for students and researchers who are interested in speech analysis, deep learning use cases, emotion detection, and audio processing.

6.4 Limitations and Future Improvements

Despite its success, the system has several limitations that provide opportunities for future enhancement. The limitations of the audio sentiment are that the model was primarily trained on the TESS Toronto emotional speech dataset, which may not reflect the full spectrum of emotional expressions across different cultures and languages. Future work should incorporate more diverse datasets to improve cross-cultural robustness. In addition, the current system only analyzes individual short audio segments and not the general conversational or situational context. Incorporating contextual information may improve accuracy when handling ambiguous utterances.

The current system separates emotions into discrete categories, whereas in human beings, emotions are usually compound mixtures. Future versions might take a dimensional emotion model (valence-arousal) or even facilitate the detection of mixtures of emotions. The other limitation is that the current version requires whole audio files to analyze. Extension of the system to accommodate real-time audio streams would make available applications like monitoring emotions in real-time during a conversation or call.

Not only that, the CNN model, as much as it is efficient, is still possible to optimize for implementation on low-resource devices. The methods of quantization of models, pruning, or knowledge distillation can be applied to decrease the model footprint without significantly impacting accuracy. Another approach would involve the use of user-specific calibration to account for individual differences in emotional expression that would increase system accuracy for ordinary users. Finally, enhancing emotion detection outcome visualization would more adequately inform users about emotional trends within their speech.

6.5 Issues and Solutions

Throughout the development process, there were several technical problems that were encountered and overcome. One of the key problems was feature dimensionality inconsistency, where the features extracted had different dimensions for different audio samples, leading to errors in model training. This was rectified by using a standardization function (reshape_to_fixed_size) to ensure all feature vectors had a consistent dimension, which is 704 features, by either dropping surplus features or padding with zeros. Additionally, the model convergence was likewise problematic, as early training attempts demonstrated overfitting and unstable convergence. Dropout layers (with 0.2 and 0.3 rates) were used at key points of the network architecture to address this, and early stopping with 3 3-epoch patience was added to prevent overfitting.

The variability in audio quality was also challenging, whereby irregular feature extraction due to different audio file qualities affected consistency. The solution to this involved normalization of the audio processing pipeline using consistent offset (0.6 seconds) and duration (2.5 seconds) parameters and adding data augmentation for facilitating the model to generalize for different audio qualities. The GUI integration was also problematic, whereby integration of the trained model into Streamlit's GUI was marred by path and dependency problems. This was resolved by using robust path handling and error checking to ensure the model and its dependencies load properly in any deployment environment.

The user interface needed to be addressed when early interface designs were found to be non-intuitive for non-technical users. The UI was then redesigned with intuitive controls, visual indicators of emotion, and explicit instructions to enhance usability. Model serialization was the last significant challenge, as saving and loading the model along with its preprocessing components introduced compatibility problems. This was achieved by separating the model, scaler, and encoder serialization with different methods, including Keras save and load for the model, pickle for the scaler and encoder, and adding strong error checking when loading.

References

adminhowley. (2022, December 24). *Can I Sue My Employer For Emotional Distress? - Howley Law*. Howley Law Firm. https://howleylawfirm.com/can-i-sue-my-employer-for-emotional-distress#:~:text=Workplace%20emotional%20distress%20develops%20when

Anastasov, R. (2023, November 3). *Sentiment Analysis: Understanding Perception for Better Marketing*. Mention. https://mention.com/en/blog/power-of-sentiment-analysis/

Apache NetBeans. (2017). Introduction to Developing Web Applications. Apache.org. https://netbeans.apache.org/tutorial/main/kb/docs/web/quickstart-webapps/

AWS. (2023). *What is Sentiment Analysis? - Sentiment Analysis Explained - AWS*. Amazon Web Services, Inc. https://aws.amazon.com/what-is/sentiment-analysis/

Badshah, A. M., Ahmad, J., Rahim, N., & Baik, S. W. (2017). Speech Emotion Recognition from Spectrograms with Deep Convolutional Neural Network. 2017 International Conference on Platform Technology and Service (PlatCon). https://doi.org/10.1109/platcon.2017.7883728

Better Health. (2012). *Work-related stress*. Better Health Channel. https://www.betterhealth.vic.gov.au/health/healthyliving/work-related-stress

Brownlee, J. (2017, July 19). A Gentle Introduction to Long Short-Term Memory Networks by the Experts. Machine Learning Mastery. https://machinelearningmastery.com/gentle-introduction-long-short-term-memory-networks-experts/

Can You Sue for Emotional Distress? | Working Now and Then. (2024, August 15). Working Now and Then. https://www.workingnowandthen.com/can-you-sue-for-emotional-distress/

Charles Joseph. (2017). Working Now and Then. https://www.workingnowandthen.com/can-you-sue-for-emotional-distress/#:~:text=Emotional%20distress%20damages%20compensate%20an

Chugh, A. (2019, January 16). Deep Learning | Introduction to Long Short Term Memory. GeeksforGeeks. https://www.geeksforgeeks.org/deep-learning-introduction-to-long-short-term-memory/

Classification of Sound using Convolutional Neural Networks | IEEE Conference Publication | IEEE Xplore. (n.d.). Ieeexplore.ieee.org. https://ieeexplore.ieee.org/document/10072823

Dwivedi, D., Ganguly, A., & V.V. Haragopal. (2023). Contrast between simple and complex classification algorithms. *Elsevier EBooks*, 93–110. https://doi.org/10.1016/b978-0-323-91776-6.00016-6

Easy Web Site Creation in the NetBeans IDE. (2019). Oracle.com. https://www.oracle.com/java/technologies/website-creation-netbeans-ide.html

emanuelbuttaci. (2023, March 12). Speech recognition with spectrograms and CNN. Kaggle.com; Kaggle. https://www.kaggle.com/code/emanuelbuttaci/speech-recognition-with-spectrograms-and-cnn

Foster, K. (2021, November 9). *What is ASR? An Overview of Automatic Speech Recognition*. AssemblyAI Blog. https://www.assemblyai.com/blog/what-is-asr/

Gandhi, R. (2018, June 7). Support Vector Machine — Introduction to Machine Learning Algorithms. Towards Data Science. https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47

giosmin, E. (2023, July 28). *Sentiment analysis by voice recording: how to understand what your customers are thinking - XCALLY Motion*. XCALLY Motion. https://www.xcally.com/news/sentiment-analysis-by-voice-recording-how-to-understand-what-your-customers-are-thinking/

Gordon, S. (2022, February 21). *How Workplace Bullying Negatively Affects the Victim and Profits*. Verywell Mind. https://www.verywellmind.com/what-are-the-effects-of-workplace-bullying-460628

Hamad, R. (2023, December 11). What is LSTM? Introduction to Long Short-Term Memory. Medium. https://medium.com/@rebeen.jaff/what-is-lstm-introduction-to-long-short-term-memory-66bd3855b9ce

Harris, H. (2023, March 3). *Overworking: The Effects on Mental Health | Nivati*. Www.nivati.com. https://www.nivati.com/blog/overworking-the-effects-on-mental-health

IBM. (2023, December 27). *What is support vector machine? | IBM*. Www.ibm.com. https://www.ibm.com/topics/support-vector-machine

IBM. (2024). What are Convolutional Neural Networks? | IBM. Www.ibm.com; IBM. https://www.ibm.com/topics/convolutional-neural-networks

JavaTPoint. (n.d.). Support Vector Machine (SVM) Algorithm - Javatpoint. Www.javatpoint.com. https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm

Jeevitha M. (2023, September 4). *Librosa is a popular Python library for audio and music analysis. It provides tools for various audio-related tasks, including feature extraction, visualization, and more.* Linkedin.com. https://www.linkedin.com/pulse/exploring-librosa-comprehensive-guide-audio-feature-extraction-m

jeffprosise. (2019). Deep-Learning/Audio Classification (CNN).ipynb at master · jeffprosise/Deep-Learning. GitHub. https://github.com/jeffprosise/Deep-Learning/blob/master/Audio%20Classification%20(CNN).ipynb

Kanade, V. (2022, September 29). What Is a Support Vector Machine? Working, Types, and Examples. Spiceworks. https://www.spiceworks.com/tech/big-data/articles/what-is-support-vector-machine/

Kandola, A. (2020, November 27). Emotional distress: What are the causes and symptoms? Www.medicalnewstoday.com. https://www.medicalnewstoday.com/articles/emotional-distress#causes

Long Short-Term Memory Network - an overview | ScienceDirect Topics. (n.d.). Www.sciencedirect.com. https://www.sciencedirect.com/topics/computer-science/long-short-term-memory-network

Luitel, S., & Anwar, M. (2022, August 1). *Audio Sentiment Analysis using Spectrogram and Bag-of- Visual-Words*. IEEE Xplore. https://doi.org/10.1109/IRI54793.2022.00052

MacklinConnection Blog | What is the Cost of Bad Relationships at Work? (2021, August 24). Macklinconnection.com. https://www.macklinconnection.com/blog/what-is-the-cost-of-bad-relationships-at-work

Malviya, R. (n.d.). *7 Examples of Poor Working Conditions and How to Improve Them*. Www.pulpstream.com. https://www.pulpstream.com/resources/blog/working-conditions#:~:text=In%20contrast%2C%20poor%20workplace%20conditions

Microsoft. (2016, April 14). Visual Studio Code. Visual Studio Code. https://code.visualstudio.com/docs/editor/whyvscode

Naman Dhariwal, Sri Chander Akunuri, None Shivama, & K Sharmila Banu. (2023). Audio and Text Sentiment Analysis of Radio Broadcasts. *IEEE Access*, *11*, 126900–126916. https://doi.org/10.1109/access.2023.3331226

Nandi, P. (2021, July 30). CNNs for Audio Classification. Medium. https://towardsdatascience.com/cnns-for-audio-classification-6244954665ab

Nanni, L., Maguolo, G., Brahnam, S., & Paci, M. (2021). An Ensemble of Convolutional Neural Networks for Audio Classification. Applied Sciences, 11(13), 5796. https://doi.org/10.3390/app11135796

optisolnew1. (2023, March 31). *5 Benefits of Using Sentiment Analysis for Understanding Patients Feedback*. OptiSol. https://www.optisolbusiness.com/insight/5-benefits-of-using-sentiment-analysis-for-understanding-patients-feedback

praat/praat. (2021, January 3). GitHub. https://github.com/praat/praat

Sansone, R. A., & Sansone, L. A. (2015). Workplace bullying: a tale of adverse consequences. *Innovations in Clinical Neuroscience*, *12*(1-2), 32–37. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4382139/#:~:text=Emotional%2Fpsychological%20consequences%20of%20workplace

Sasidharan, A. (2021, January 20). *Support Vector Machine Algorithm*. GeeksforGeeks. https://www.geeksforgeeks.org/support-vector-machine-algorithm/

scikit learn. (2018). 1.4. Support Vector Machines — scikit-learn 0.20.3 documentation. Scikit-Learn.org. https://scikit-learn.org/stable/modules/svm.html

Sentiment Analysis with LSTM. (2022, January 17). Analytics Vidhya. https://www.analyticsvidhya.com/blog/2022/01/sentiment-analysis-with-lstm/

Tech, C. (2024, July 17). *Sentiment Analysis: Accelerating Innovation in Healthcare & Life Science*. Cogitotech. https://www.cogitotech.com/blog/sentiment-analysis-accelerating-innovation-in-healthcare-and-life-science/

Technocrat. (2023, September 14). *Introduction to LibROSA*. CoderHack.com. https://medium.com/coderhack-com/introduction-to-librosa-912c2c109f41

Top 7 Methods for Audio Sentiment Analysis in 2024. (2024). AIMultiple: High Tech Use Cases & Tools to Grow Your Business. https://research.aimultiple.com/audio-sentiment-analysis/

Types of voices. (2024). Google Cloud. https://cloud.google.com/text-to-speech/docs/voice-types

Upadhyay, A., Singh, A., Gupta, A., Sharma, A., Mall, A., & Pooja Tomar. (2024). Sentiment Analysis on Speech Data using MFCC. *IJNRD.ORG IJNRD2407252 International Journal of Novel Research and Development*, *9*(7), 2456–4184. https://www.ijnrd.org/papers/IJNRD2407252.pdf

Visual Studio Code. (2016, April 14). Visualstudio.com. https://code.visualstudio.com/Docs/languages/html

Visual Studio Code: Read this before you get started. (n.d.). Daily.dev. https://daily.dev/blog/visual-studio-code-read-this-before-you-get-started

Visual Studio vs Visual Studio Code - What's Best In 2022? (n.d.). Www.turing.com. https://www.turing.com/kb/ultimate-guide-visual-studio-vs-visual-studio-code

Watkins, H. (2020, April 13). *7 Causes of Stressful Work Environments and How to Fix Them*. Solvo Global. https://solvoglobal.com/how-to-fix-7-causes-of-stressful-work-environments/

Welcome to Apache NetBeans. (n.d.). Netbeans.apache.org. https://netbeans.apache.org/front/main/index.html

Westover, J. H. (2023, October 3). *What Research Says about the Dangers of Long Working Hours*. HCI Consulting; HCI Consulting. https://www.innovativehumancapital.com/article/what-research-says-about-the-dangers-of-long-working-hours#:~:text=The%20research%20is%20clear%20that

Why Is Sentiment Analysis Important? - Voxco. (n.d.). https://www.voxco.com/blog/why-is-sentiment-analysis-important/

World Health Organization. (2020, October 19). Occupational health: Stress at the workplace. Www.who.int. https://www.who.int/news-room/questions-and-answers/item/ccupational-health-stress-at-the-workplace#:~:text=Work%2Drelated%20stress%20can%20be

xtn. (2024, February 4). *Sentiment Analysis: Decoding Customer Emotions in Market Research*. Medium. https://medium.com/@xtn13/sentiment-analysis-decoding-customer-emotions-in-market-research-fbdda26a91ab

Yadav, V., Verma, P., & Katiyar, V. (2023). Long short term memory (LSTM) model for sentiment analysis in social data for e-commerce products reviews in Hindi languages. International Journal of Information Technology (Singapore), 15(2), 759–772. https://doi.org/10.1007/s41870-022-01010-y