Abishek shah

Abishek_shah_77466875_Project Specification

Component 1: Project Specification (Initial Project Plan + Risk Register)

Document Details

Submission ID
trn:oid:::26687:126516955

Submission Date
Jan 19, 2026, 9:58 PM GMT+5:45

Download Date
Jan 19, 2026, 10:03 PM GMT+5:45

File Name
Abishek_shah_77466875_Project Specification.docx

File Size
221.5 KB

13 Pages

1,817 Words

9,900 Characters

15% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Exclusions

▸ 3 Excluded Matches

Match Groups

19 Not Cited or Quoted 14%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

1 Missing Citation 1%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

10% Internet sources
8% Publications
15% Submitted works (Student Papers)

Match Groups

- **19** Not Cited or Quoted 14%
  Matches with neither in-text citation nor quotation marks
- **0** Missing Quotations 0%
  Matches that are still very similar to source material
- **1** Missing Citation 1%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

10% Internet sources
8% Publications
15% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Publication	
Zengzhao Chen, Jiawen Li, Hai Liu, Xuyang Wang, Hu Wang, Qiuyu Zheng. "Learni...		2%
2	Student papers	
The British College on 2025-01-06		2%
3	Internet	
iieta.org		1%
4	Student papers	
The British College on 2025-01-06		1%
5	Publication	
Rashmi Gupta, Xiomara Patricia Blanco Valencia, Lalit Mohan Goyal, Jeetendra Ku...		1%
6	Internet	
assets-eu.researchsquare.com		<1%
7	Student papers	
University of Stirling on 2024-08-30		<1%
8	Publication	
Agnieszka Szparaga, Łukasz Szparaga, Ewa Czerwińska, Ireneusz Kapusta, Grzego...		<1%
9	Student papers	
The British College on 2025-01-06		<1%
10	Student papers	
The British College on 2025-01-06		<1%

11 Student papers

Kampala International University on 2023-07-10 <1%

12 Student papers

University of Greenwich on 2025-04-24 <1%

13 Student papers

University of Liverpool on 2022-11-28 <1%

14 Internet

asmp-eurasipjournals.springeropen.com <1%

15 Internet

www.apectraining.com <1%

16 Student papers

University of Essex on 2024-08-27 <1%

17 Student papers

University of Hertfordshire on 2023-05-15 <1%

18 Student papers

University of the West of England, Bristol on 2026-01-09 <1%

BSc (Hons) Computing Course 2025/26

Level 6 Production Project

Name: Abishek shah	**Student I.D.**:77466875
Course: BSc (Hons) Computing	**Supervisor's Name:**

Final Project Individual Aim & Objectives

Title of my Project: Speech Emotion Recognition Using Bilstm and self-Attention

Aim of my Project: The purpose of this project is to develop a speech emotion recognition system that that gets speech audio as input and outputs the predicted emotional category of the speaker using a Bidirectional LSTM with a self-attention mechanism.

- **Objectives of my Project:**

To develop an automated speech emotion recognition system that classifies emotions from speech audio.

To extract relevant acoustic features using MFCC, delta MFCC, and delta-delta MFCC from speech signals.

- To model temporal speech patterns using a Bidirectional LSTM neural network.

• To use a self-attention method to focus on the parts of the speech that are emotionally important.

- To use an emotional speech dataset to train and test the suggested model and find out how well it works.
- To produce a final system that outputs the predicted emotion label for a given speech input.

Specification of my Product:

Functional Requirement	MoSCoW
The website must allow users to upload a speech audio file (e.g: WAV) for emotion detection.	M
The system must preprocess the uploaded audio (cleaning + feature extraction) before running the model.	M
The system must predict the emotion using a BiLSTM + Self-Attention model and display the result clearly.	M
The website must use a backend/API service to send the audio and receive the prediction output.	M
The website should show the confidence/probability score along with the predicted emotion.	S
Users should be able to save prediction results (emotion + score + date/time) in a database.	S
The website would allow users to compare two audio files and see which emotion is stronger in each result.	W

Non-Functional Requirement:

- The website should give emotion results fast and run smoothly.
- The layout should be simple so anyone can upload audio and understand the output.
- It should handle more users and more prediction requests as the project grows.
- It must work well on Chrome Edge and Firefox and fit both mobile and desktop screens.
- It should show stable and show clear messages if an audio file is wrong or something fails.
- User audio and saved results stay private with secure upload and storage.

Research:

Most existing SER papers take one of two paths: they either keep things simple with MFCC-style features and an LSTM, or they add heavier components like CNN backbones, custom gating/optimization, and more complex attention schemes (Singh et al., 2023; Abdelhamid et al., 2022). One RAVDESS-style approach mainly relies on MFCC feature extraction and then trains an LSTM sequential model to learn emotion patterns

over time (Leelavathi et al., 2022).Other work extends this idea by changing how LSTM memory behaves (e.g., modifying the forgetting gate) and applying attention across both the **time** and **feature** dimensions to focus on emotionally important information (Xie et al., 2019).Some approaches then combine CNN+LSTM with attention and train across multiple datasets (or dataset combinations), which typically increases pipeline complexity and training demands (Singh et al., 2023; Abdelhamid et al., 2022). Graph-LSTM goes even further by explicitly constructing a graph based on feature similarity before classification, adding extra computation beyond standard sequence models (Li et al., 2023). In comparison, my stacked BiLSTM with self-attention on MFCC+Δ+ΔΔ keeps the pipeline lightweight while still learning bidirectional context and focusing on the most emotionally informative frames, making it a practical and robust choice for RAVDESS-style emotion dynamics.

Evaluation:

This work involved reviewing the project overall performance, results, and reliability against the goals set at the start. It focused on checking effectiveness tracking progress across development stage and judging the practical impact of the final system.

In this study, the main goal was to evaluate a Speech Emotion Recognition model built using Bidirectional LSTM and self-attention. The evaluation considered whether the model could learn meaningful emotion patterns from MFCC delta and delta delta features and produce consistent predictions. Progress was assessed from preprocessing and feature extraction through training and testing to confirm the system met the project objectives.

Project Planning & Methodology

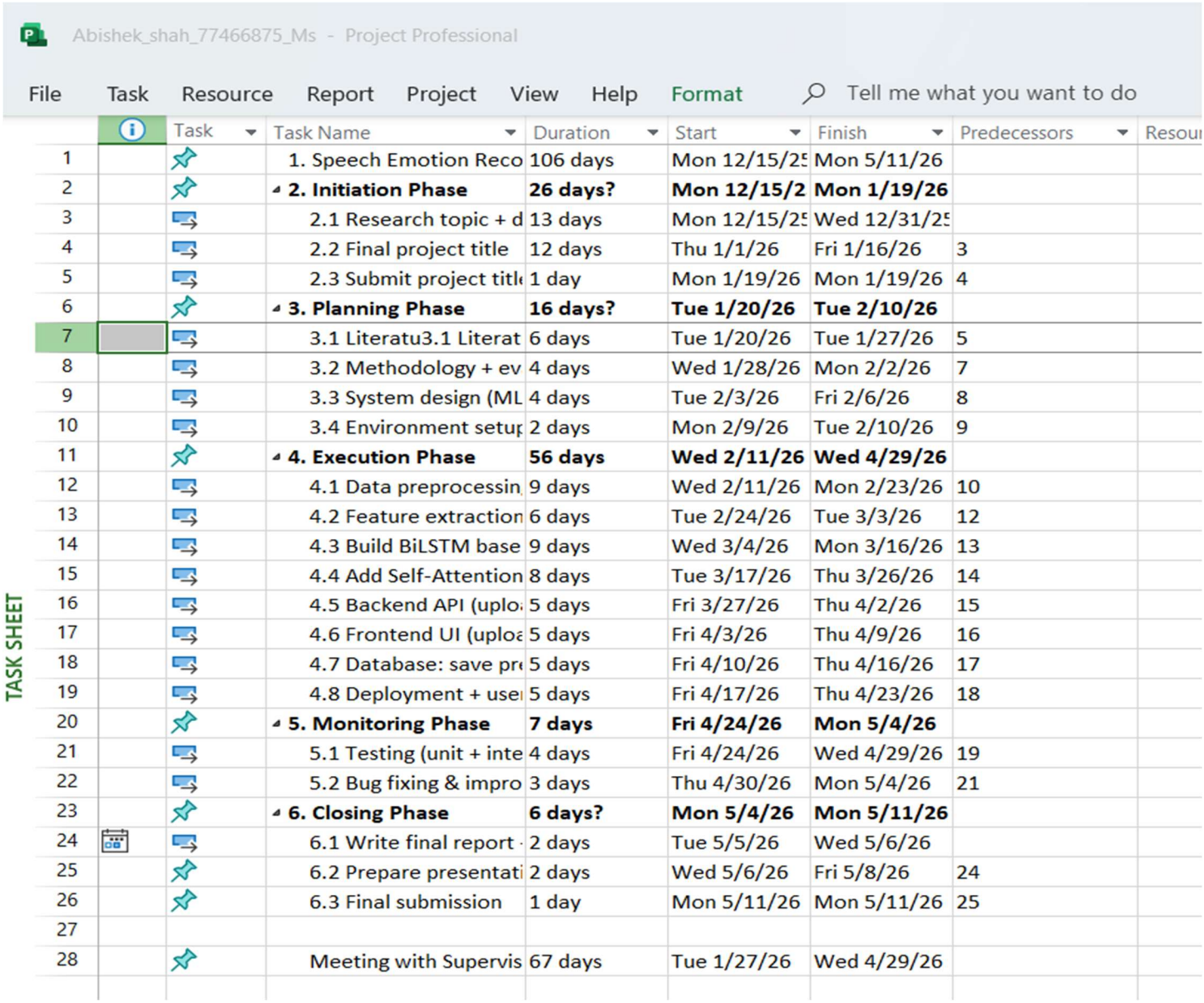
Project Planning:

Project planning was used to split the work into weekly step such as research dataset preparation feature extraction model building testing and final reporting. This kept task organized and reduced lastminute issues.

Figure 1: Gantt chart showing the project schedule and task timelines.

Gantt chart:

A Gantt chart was made to show task start and end dates and dependencies. It helped track progress manage time, and ensure key steps were done in order.



	Task	Task Name	Duration	Start	Finish	Predecessors	Resou
1		1. Speech Emotion Reco	106 days	Mon 12/15/25	Mon 5/11/26		
2		**2. Initiation Phase**	**26 days?**	**Mon 12/15/2**	**Mon 1/19/26**		
3		2.1 Research topic + d	13 days	Mon 12/15/25	Wed 12/31/25		
4		2.2 Final project title	12 days	Thu 1/1/26	Fri 1/16/26	3	
5		2.3 Submit project titl	1 day	Mon 1/19/26	Mon 1/19/26	4	
6		**3. Planning Phase**	**16 days?**	**Tue 1/20/26**	**Tue 2/10/26**		
7		3.1 Literatu3.1 Literat	6 days	Tue 1/20/26	Tue 1/27/26	5	
8		3.2 Methodology + ev	4 days	Wed 1/28/26	Mon 2/2/26	7	
9		3.3 System design (ML	4 days	Tue 2/3/26	Fri 2/6/26	8	
10		3.4 Environment setup	2 days	Mon 2/9/26	Tue 2/10/26	9	
11		**4. Execution Phase**	**56 days**	**Wed 2/11/26**	**Wed 4/29/26**		
12		4.1 Data preprocessin	9 days	Wed 2/11/26	Mon 2/23/26	10	
13		4.2 Feature extraction	6 days	Tue 2/24/26	Tue 3/3/26	12	
14		4.3 Build BiLSTM base	9 days	Wed 3/4/26	Mon 3/16/26	13	
15		4.4 Add Self-Attention	8 days	Tue 3/17/26	Thu 3/26/26	14	
16		4.5 Backend API (uplo	5 days	Fri 3/27/26	Thu 4/2/26	15	
17		4.6 Frontend UI (uploa	5 days	Fri 4/3/26	Thu 4/9/26	16	
18		4.7 Database: save pre	5 days	Fri 4/10/26	Thu 4/16/26	17	
19		4.8 Deployment + user	5 days	Fri 4/17/26	Thu 4/23/26	18	
20		**5. Monitoring Phase**	**7 days**	**Fri 4/24/26**	**Mon 5/4/26**		
21		5.1 Testing (unit + inte	4 days	Fri 4/24/26	Wed 4/29/26	19	
22		5.2 Bug fixing & impro	3 days	Thu 4/30/26	Mon 5/4/26	21	
23		**6. Closing Phase**	**6 days?**	**Mon 5/4/26**	**Mon 5/11/26**		
24		6.1 Write final report	2 days	Tue 5/5/26	Wed 5/6/26		
25		6.2 Prepare presentati	2 days	Wed 5/6/26	Fri 5/8/26	24	
26		6.3 Final submission	1 day	Mon 5/11/26	Mon 5/11/26	25	
27							
28		Meeting with Supervis	67 days	Tue 1/27/26	Wed 4/29/26		

Figure 2: Task sheet showing weekly tasks and key milestones.

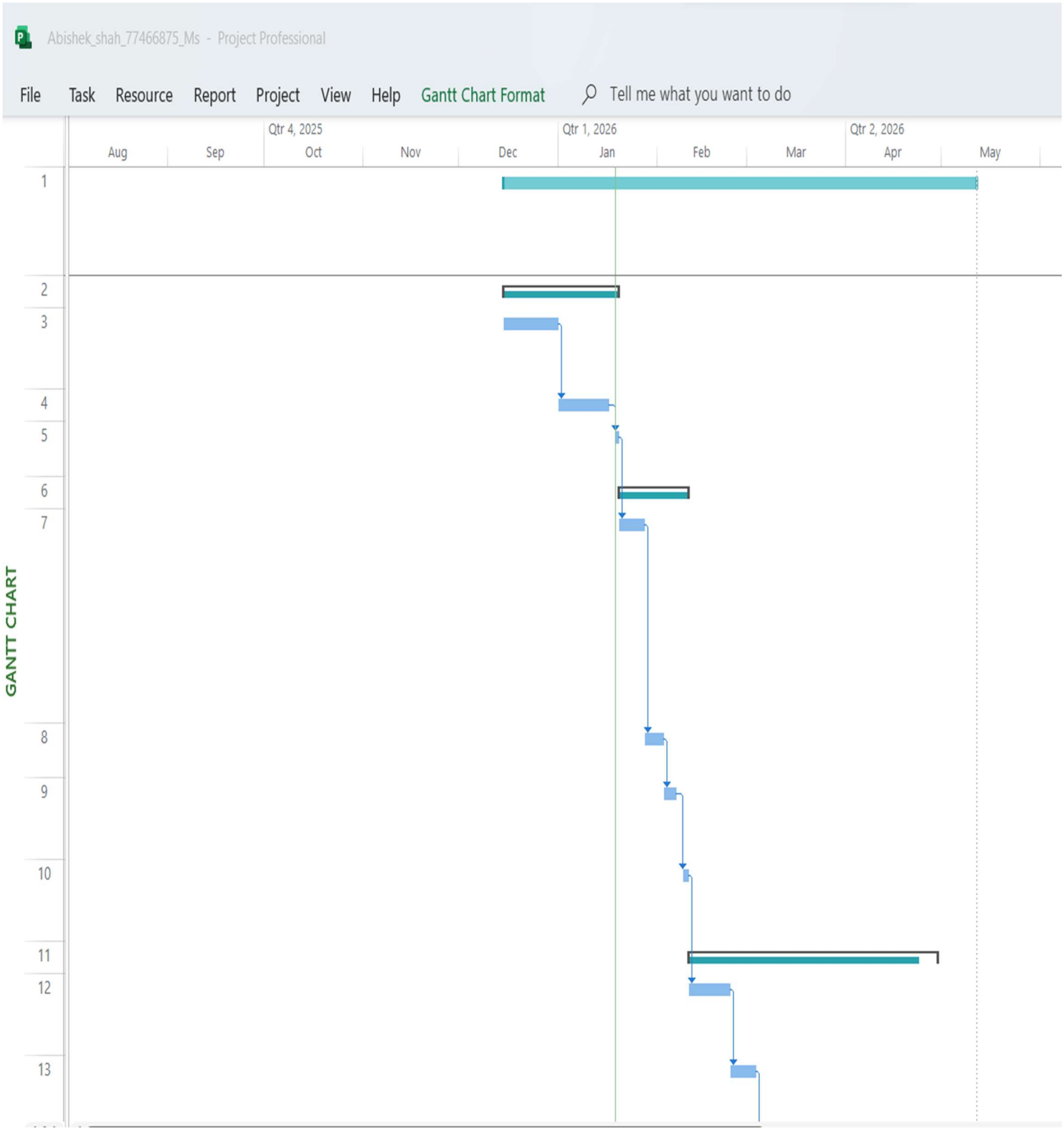


Figure 1(a): Gantt chart (Part 1).

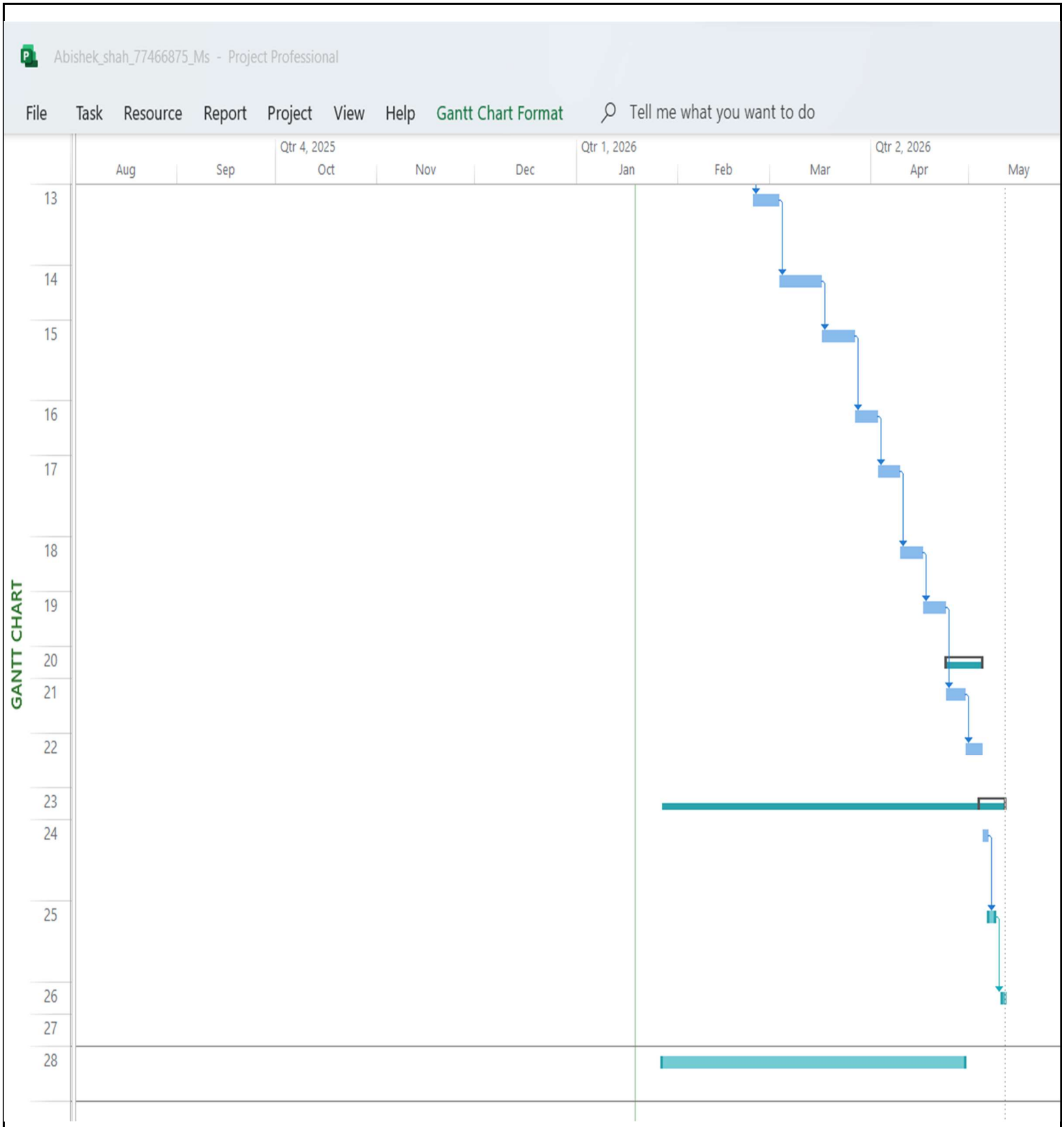


Figure 1(b): Gantt chart (Part 2).

Methodology:

In this Speech Emotion Recognition project, a BiLSTM with self-attention was built using MFCC, delta, and delta-delta features. Developing it required lots of short experiments trying different feature settings, padding length, model depth, and learning rate then quickly checking the validation results and improving the next version. This is why Agile/Scrum is the most suitable approach: it is designed for complex work and follows an iterative, incremental cycle that supports frequent improvement and risk control. In contrast **Waterfall** is more sequential and was even warned against when feedback and iteration are missing, which does not fit ML experimentation well. RAD helps for fast prototypes, but it is more software-prototype focused than model evaluation. JAD is useful early on to agree labels, requirements, and evaluation criteria through workshops.

Resources

Dataset

- RAVDESS

Software:

- Windows 10/11
- Python 3
- Google Colab / Jupyter Notebook
- TensorFlow + Keras
- Librosa to load MFCC, delta, delta–delta features
- NumPy, Pandas
- Scikit-learn
- Pickle
- (Flask / FastAPI)

Frontend Software

- HTML, CSS, JavaScript

Other Supporting Software

- Google Chrome
- Microsoft Word
- Microsoft Excel
- Microsoft PowerPoint
- Microsoft Access

Hardware

- Laptop/PC with CPU, 8GB+ RAM
- GPU or Google Colab GPU

Human Resource

I am working on my Project with the following people

Name: Abishek shah	**Role: not needed**
	Module Leader: Mr. Rohit raj, Pandey
	Supervisor

Initial Bibliography

1. **Xie, Y., Liang, R., Liang, Z., Huang, C., Zou, C. and Schuller, B.** (2019) 'Speech emotion classification using attention-based LSTM', *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 27(11), pp. 1675–1685. Available at: https://ieeexplore.ieee.org/document/8794694 (Accessed: 16 January 2026).

2. **Abdelhamid, A.A., El-Kenawy, E.S.M., Alotaibi, B., Amer, G.M., Abdelkader, M.Y., Ibrahim, A. and Eid, M.M.** (2022) 'Robust speech emotion recognition using CNN-LSTM based on stochastic fractal search optimization algorithm', *IEEE Access*, 10, pp. 49265–49284. Available at: https://doi.org/10.1109/ACCESS.2022.3172954 (Accessed: 16 January 2026).

3. **Singh, J., Saheer, L.B. and Faust, O.** (2023) 'Speech emotion recognition using attention model', *International Journal of Environmental Research and Public Health*, 20(6), 5140. Available at: https://doi.org/10.3390/ijerph20065140 (Accessed: 16 January 2026).

4. **Sreelavathi, R., Deepthi, S.A. and Aruna, V.** (2022) 'Speech emotion recognition using LSTM', *International Research Journal of Engineering and Technology (IRJET)*, 9(1), pp. 586–590. Available at: https://www.irjet.net/archives/V9/i1/IRJET-V9I1101.pdf (Accessed: 16 January 2026).

5. **Li, Y., Wang, Y., Yang, X. and Im, S.K.** (2023) 'Speech emotion recognition based on graph-LSTM neural network', *EURASIP Journal on Audio, Speech, and Music Processing*, 2023(40). Available at: https://doi.org/10.1186/s13636-023-00303-9 (Accessed: 16 January 2026).

9

Risk Register:

ID	Risk	Risk Description	Likelihood	Impact	Severity	Owner	Mitigation	Status
1	Dataset path missing	So, the code won't be able to find any wav files because the dataset folder might not be located where it should be.	3	5	15 (High)	Student	Check the folder and count the audio files before you train. Stop and show a clear message if none are found.	Open
2	Broken audio files	Some audio files may be damaged or not readable	3	3	9 Medium	Student	Keep a list of failed files and show how many failed. Replace or re-download files.	Open
3	Audio normalization error	If a sound file is almost empty, dividing it by the highest number leads to NaN or Inf, which can stop train.	2	5	10(Medium)	Student	Add a small safety check don't divide that file or skip it if the max value is too small.	Open
4	Padding uses too much memory	Some audio repetitions may be long. So, padding everything to the longest can cause huge arrays and crash RAM/GPU.	3	5	15 (High)	Student	Set a fixed max length and cut longer sequences. Also reduce batch size if memory is low.	Open
5	Wrong label from filename	The label that was extracted out by split ('-')[2] might be wrong if the	3	5	15(High)	Student	To start, print some filename-to-label samples. Before	Open

		filename is incorrect					training make sure that the labels fit the expected emotion classes.	
6	Class imbalance	Some emotions have more samples than others. so, the model needs to learn the common emotions.	3	4	12(Medium)	Student	If you need to use class weights and check the class numbers after the stratified split. Not only accuracy scores but also points for each class.	Open
7	Overfitting	The model may learn the training data to well and perform poorly on new voice sample.	4	5	20(High)	Student	Don't train for too many times and use early ending and dropout. frequently compare the results of training and evaluation.	Open
8	Unstable training	Training may become slow or accuracy may not improve due to learning rate, model complexity.	3	3	9(Medium)	Student	You can already use ReduceLROnPlateau, try a slower learning rate or batch size, and save the best model while it being trained.	Open
9	Upload/API not working	The site might not be able to send	3	4	12(Medium)	Student	Use Postman or your	Open

		the audio file to the service because of the size of the file the wrong format, or a problem with CORS.					browser to test the upload early. Set rules for the size and kind of files and provide error messages that are easy to understand.	
10	Confidence score misleading	The model might show high confidence even when it is wrong can confuse users.	3	3	9(Medium)	Student	Display it as model confidence. Also show top-2 predictions so users can see alternatives.	Open
11	Privacy and security	User audio files and prediction history could be exposed if storage and access are not handled safely.	3	5	15(High)	Student	Do not store audio longer than needed. Delete temporary files, restrict access, and store only necessary prediction details.	Open

Notes:
ID values may be useful to refer back to in your final documentation. Number these in order. This register should be included in the appendix
Risk description provides an outline of the issue
Please use Low, Medium and High to identify the risk level and colour code.
Typically the owner will be you, but it maybe the case in team work or other projects that have external clients, other activities may impact on the project
Mitigation implies on how you will manage the risk and to reduce the likelihood of it occurring
Status – has the risk event now passed. It should indicate an Open and Closed status.