

BDM 3035 - Big Data Capstone Project

**MILESTION REPORT 01 FOR
SPEECH EMOTION RECOGNITION PROJECT**

SUBMITTED TO

Instructor Meysam Effati



SUBMITTED BY

ADRIANA M. PENARANDA BARON (C0898944)

ARUNA GURUNG (C0896129)

CARLOS A. REY PINTO (C0868575)

HALDO J. SOMOZA SOLIS (C0904838)

PUJAN SHRESTHA (C0901167)

June 11, 2024

Table of Contents

I. INTRODUCTION 3

II. PROGRESS REPORT 3

III. MODIFIED TIMELINE TABLE 4

IV. NEXT STEPS 7

V. CHALLENGES FACED 8

VI. LESSONS LEARNED 8

VII. CONCLUSSIONS 8

VIII. REFERENCES 9

I. INTRODUCTION

The Speech Emotion Recognition (SER) project focuses on creating a Python-based model that can detect human emotions from speech using the Librosa library and machine learning techniques. This application is particularly useful for call centers, where recognizing customer emotions can improve service quality and boost conversion rates. By leveraging the dataset, which includes emotionally labeled speech and song files, we will extract crucial audio features using Librosa, train a machine learning model, and assess its accuracy.

This project provides practical experience in audio processing and machine learning, showcasing the effective integration of these technologies in understanding human emotions. By carrying out this project, we will gain hands-on experience in audio processing, feature extraction, and machine learning model training, offering a thorough understanding of SER systems.

The GitHub link where the milestones notebooks are is:

https://github.com/adripenaranda/Speech_Emotion_Recognition_Project

II. PROGRESS REPORT

SUMMARY OF TASKS COMPLETED

We have successfully completed several key tasks. Initially, we gathered the dataset, which includes emotionally labeled speech and audio files. We conducted an in-depth exploration to understand its structure and characteristics. Following this, we used the Librosa library to extract essential audio features, such as mel-frequency cepstral coefficients (MFCCs) and chroma features, which are crucial for emotion recognition. We then split the data into training, validation, and test sets to ensure the evaluation process.

With the dataset prepared, we moved on to the modeling phase. We selected the appropriate machine learning model (Librosa) that are well-suited for emotion recognition tasks. The initial model was implemented using the data, and we trained these models on the training dataset. Post-training, the models were tested on the validation set to assess their performance. This evaluation allowed us to identify areas for improvement and refine our models further.

KEY ACHIEVEMENTS AND MILESTONES REACHED

These are the achievements and milestones reached

- Gather the dataset.
- Explore and understand the dataset structure.
- Extract audio features using Librosa.
- Split the data into training, validation, and test sets.
- Select appropriate machine learning model for emotion recognition.
- Implement initial model using data.
- Train model on the training data

- Test model on the validation set.
- Evaluate initial model performance and identify areas for improvement.

DEVIATIONS FROM THE ORIGINAL PLAN AND REASONS FOR THESE CHANGES.

The only change that we have currently contemplated is the visualizations that we are going to include in the project, at the beginning, we had considered using Power BI, however since we have audio, and together with the teacher's advice, we decided that we are going to use Librosa Spectrograms to show the spectrograms of each of the emotions that the model recognizes.

III. MODIFIED TIMELINE TABLE

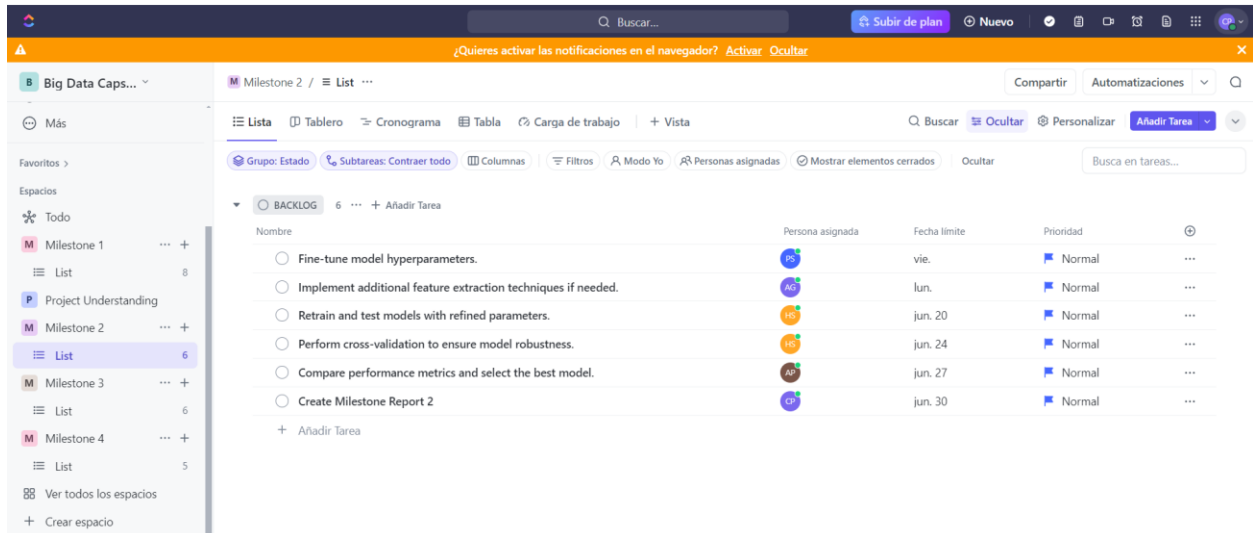
We made some small modifications to our tasks, below you will find the screenshots of the updated subtasks updated in click up:

MILESTONE 1

The screenshot shows the ClickUp interface for 'Milestone 1'. It features a sidebar with navigation icons and a main task list. The tasks are organized into three groups: 'COMPLETE' (5 tasks), 'ON HOLD' (1 task), and 'UPDATE REQUIRED' (2 tasks). Each task entry includes a name, an assignee (indicated by a colored circle), a due date, a priority (High or Urgent), and a status icon. The 'UPDATE REQUIRED' group is highlighted with a yellow background.

Group	Name	Assignee	Due date	Priority
COMPLETE	Extract audio features using Librosa.	AS	May 30	High
	Split the data into training, validation, and test sets.	AS	Jun 2	High
	Select appropriate machine learning model for emotion recognition.	AS	Jun 4	High
	Implement initial model using the preprocessed data.	AS	5 days ago	High
	Create Milestone Report 1	AS	Today	Urgent
ON HOLD	Evaluate initial model performance and identify areas for improvement.	AS	2 days ago	High
UPDATE REQUIRED	Train model on the training data	AS	4 days ago	High
	Test model on the validation set.	AS	3 days ago	High

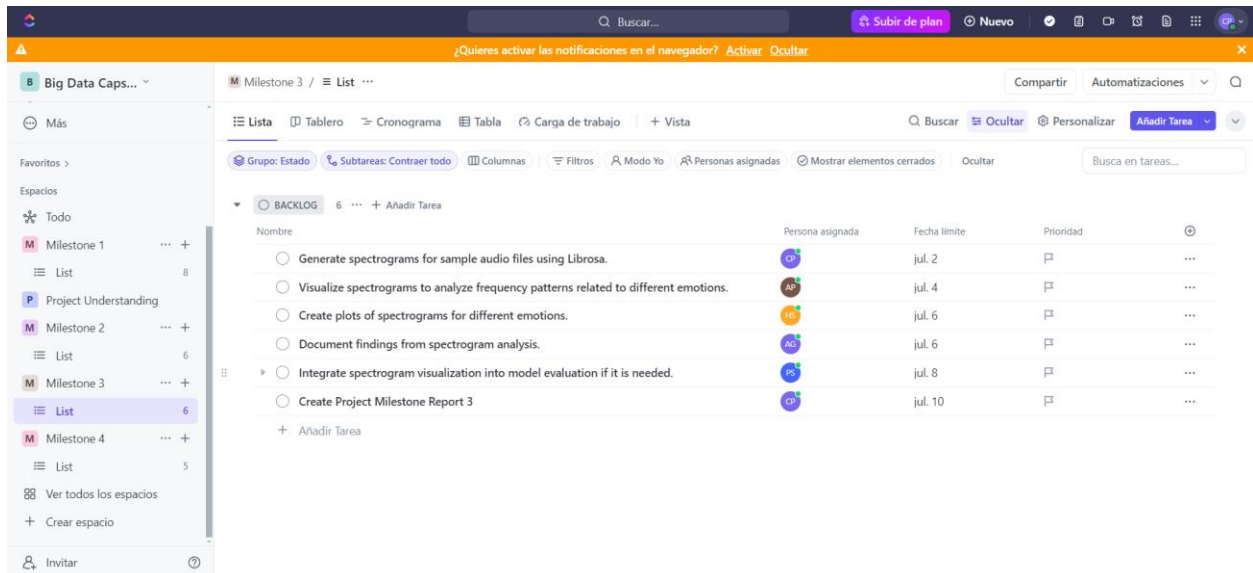
MILESTONE 2



The screenshot shows the Jira interface for Milestone 2. The left sidebar displays a project hierarchy with Milestone 2 selected. The main area shows a backlog of tasks with columns for Name, Assigned Person, Due Date, and Priority.

Nombre	Persona asignada	Fecha límite	Prioridad
<input type="radio"/> Fine-tune model hyperparameters.	PS	vie.	Normal
<input type="radio"/> Implement additional feature extraction techniques if needed.	AP	lun.	Normal
<input type="radio"/> Retrain and test models with refined parameters.	IS	jun. 20	Normal
<input type="radio"/> Perform cross-validation to ensure model robustness.	IS	jun. 24	Normal
<input type="radio"/> Compare performance metrics and select the best model.	AP	jun. 27	Normal
<input type="radio"/> Create Milestone Report 2	CP	jun. 30	Normal

MILESTONE 3



The screenshot shows the Jira interface for Milestone 3. The left sidebar displays a project hierarchy with Milestone 3 selected. The main area shows a backlog of tasks with columns for Name, Assigned Person, Due Date, and Priority.

Nombre	Persona asignada	Fecha límite	Prioridad
<input type="radio"/> Generate spectrograms for sample audio files using Librosa.	CP	jul. 2	Normal
<input type="radio"/> Visualize spectrograms to analyze frequency patterns related to different emotions.	AP	jul. 4	Normal
<input type="radio"/> Create plots of spectrograms for different emotions.	IS	jul. 6	Normal
<input type="radio"/> Document findings from spectrogram analysis.	AP	jul. 6	Normal
<input type="radio"/> Integrate spectrogram visualization into model evaluation if it is needed.	PS	jul. 8	Normal
<input type="radio"/> Create Project Milestone Report 3	CP	jul. 10	Normal

MILESTONE 4

The screenshot shows a project management interface for 'Big Data Caps...'. The left sidebar lists milestones 1 through 4, with Milestone 4 selected. The main area displays the 'BACKLOG' for Milestone 4, showing 6 tasks. The tasks are listed in a table with columns for Name, Assigned Person, Due Date, and Priority.

Nombre	Persona asignada	Fecha limite	Prioridad
Design the API architecture for emotion recognition.	AG	jul. 11	Normal
Implement API endpoints for model prediction.	AG	jul. 12	Normal
Develop a simple user interface (UI) for the application.	IS	vie.	Normal
Integrate the UI with the backend API.	AG	jul. 16	Normal
Test the full workflow from audio input to emotion output.	PS	jul. 18	Normal
Create Project Milestone 4	AG, IS	jul. 20	Normal

MILESTONE 5

The screenshot shows the same project management interface, but now displaying the 'PENDIENTE' (Pending) view for Milestone 5. The left sidebar shows milestones 1 through 5, with Milestone 5 selected. The main area displays 5 pending tasks in a table.

Nombre	Persona asignada	Fecha limite	Prioridad
Conduct comprehensive testing of the entire system.	IS		
Fix any bugs and optimize performance.	AG		
Prepare detailed documentation of the project.	AG		
Create user guides and technical documentation.	PS		
Finalize all deliverables for project submission.	CP		

IV. NEXT STEPS

UPCOMING STEPS AND ACTIVITIES PLANNED FOR THE PROJECT.

Brief Description of Tasks to be Undertaken

Milestone 2: June 30, 2024 - Model Evaluation and Refinement

1. Fine-tune Model Hyperparameters

Description: Adjust the hyperparameters of the selected machine learning models to optimize their performance. This involves experimenting with different values and configurations to find the most effective settings. Also improve the accuracy of the model, implementing seeds.

2. Implement Additional Feature Extraction Techniques if Needed

Description: Introduce new feature extraction methods to enhance the dataset's representational power. This may include exploring advanced techniques such as pitch tracking, spectral contrast, or other relevant audio features.

3. Retrain and Test Models with Refined Parameters

Description: Retrain the models using the fine-tuned hyperparameters and any new features extracted. Subsequently, test the models to evaluate improvements in performance and accuracy.

4. Perform Cross-Validation to Ensure Model Robustness

Description: Conduct cross-validation to validate the model's performance across different subsets of the dataset. This process helps ensure the model is robust and generalizes well to unseen data.

5. Compare Performance Metrics and Select the Best Model

Description: Compare the performance metrics of all trained models, such as accuracy. Select the model that demonstrates the best overall performance for the final implementation.

EXPECTED OUTCOMES AND GOALS FOR THE NEXT PHASE.

In the upcoming phase of the project, our expected outcomes and goals are focused on optimizing and significantly improving the performance of the emotion recognition models through several key actions. Firstly, we will fine-tune the hyperparameters of the selected models to maximize their accuracy and robustness, implementing seeds to ensure result reproducibility. Additionally, we will explore additional feature extraction techniques, such as pitch tracking and spectral contrast, to enrich the dataset's representation. Subsequently, we will retrain the models with refined parameters and new features, evaluating their enhanced performance through thorough testing. Through cross-validation, we will ensure the models are robust and generalize well to unseen data. Finally, we will meticulously compare performance metrics to select the most suitable model for implementation in the project's final phase.

V. CHALLENGES FACED

OBSTACLES OR DIFFICULTIES ENCOUNTERED DURING THE PROJECT AND HOW WE ADDRESSED.

During the project, several challenges were encountered and addressed effectively in various tasks. For example, during the data collection phase for the dataset, issues with incomplete downloads were mitigated by finding alternative sources and validating data integrity. In the development and initial testing of the Milestone 1 model, the challenges in principle were knowing how to implement books properly, and also knowing what type of data preprocessing we could implement for this type of models, later we found that our accuracy varied each time. that we executed the code, so we had to consult with our teacher to implement future improvements in the next delivery, in this case we will work with SEEDS to guarantee the same accuracy.

VI. LESSONS LEARNED

As we reflect on the project's progress so far, we have gained valuable insights and insights in several critical areas. First, navigating the complexities of data collection, particularly with the dataset, highlighted the importance of strong data management strategies and proactive problem-solving. Extracting audio features with Librosa helped us better understand how these models work and be able to improve it for the next advance.

VII. CONCLUSIONS

Currently, the project stands at the initial model development and testing phase, scheduled for completion by June 11, 2024. We have successfully implemented preprocessing pipelines, split data into training/validation/test sets, and commenced training initial models. Moving forward, for the next

milestone will be focus is on refining model hyperparameters, exploring additional feature extraction techniques if needed, and conducting rigorous cross-validation to ensure robustness.

VIII. REFERENCES

Madanian, S., Chen, T., Adeleye, O., Templeton, J. M., Poellabauer, C., Parry, D., & Schneider, S. L. (2023). Speech emotion recognition using machine learning — A systematic review. *Intelligent Systems with Applications*, 20(200266), 200266. <https://doi.org/10.1016/j.iswa.2023.200266>

Speech emotion recognition. (2020, May 28). Kaggle.com; Kaggle.
<https://www.kaggle.com/code/shivamburnwal/speech-emotion-recognition>

Susile, Y. S., & Herawam, J. (n.d.). Speech Emotion Recognition Using Librosa. *Aijmr.com*. Retrieved May 21, 2024, from <https://www.aijmr.com/papers/2023/1/1003.pdf>

(N.d.). Data-Flair.Training. Retrieved May 21, 2024, from https://data-flair.training/blogs/python-mini-project-speech-emotion-recognition/#google_vignette