BDM 3035 - Big Data Capstone Project

MILESTIONE REPORT 02 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)

July 01, 2024

Table of Contents

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 were is uploaded the milestones notebooks is:

https://github.com/adripenaranda/Speech_Emotion_Recognition_Project

II. PROGRESS REPORT

SUMMARY OF TASKS COMPLETED

In this second installment we have successfully achieved several key tasks, in our progress we initially did cross validation to confirm that our model generalized well and was not overfitting. Likewise, we applied some tuning to the model using grid search, managing to improve the accuracy of the model. We also developed three new codes that we will use in the following stages, these are: 1. Downloading youtube audios, 2. Converting audios to mp3 files and 3. Splitting audios into small parts.

KEY ACHIEVEMENTS AND MILESTONES REACHED

These are the achievements and milestones reached

Milestone 1:

- 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.

Milestone 2:

- Fine-tune model hyperparameters.
- Implement additional feature extraction techniques if needed.
- Retrain and test models with refined parameters.
- Perform cross-validation to ensure model robustness.
- Compare performance metrics and select the best model.

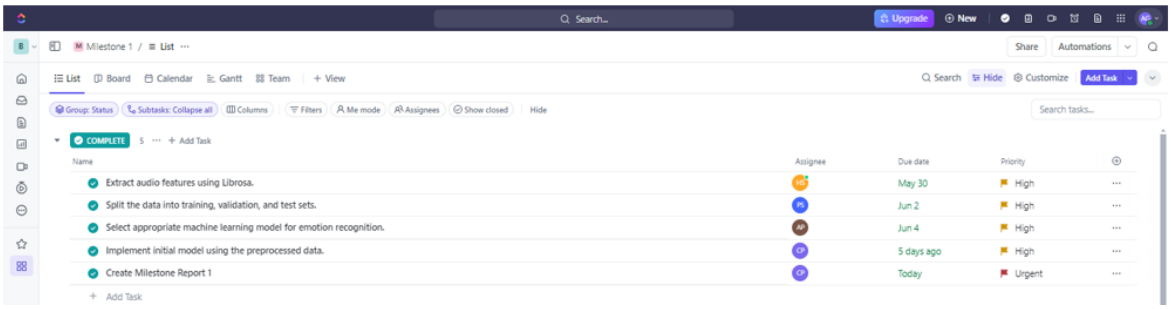
DEVIATIONS FROM THE ORIGINAL PLAN AND REASONS FOR THESE CHANGES.

The advances that were implemented in this milestone were all achieved as established. Among the tasks that are pending for the next milestone, which were not on the agenda, is checking whether the dataset is balanced or not to apply a final improvement to the model, improving its accuracy. But in general, the objectives established for this report were all made.

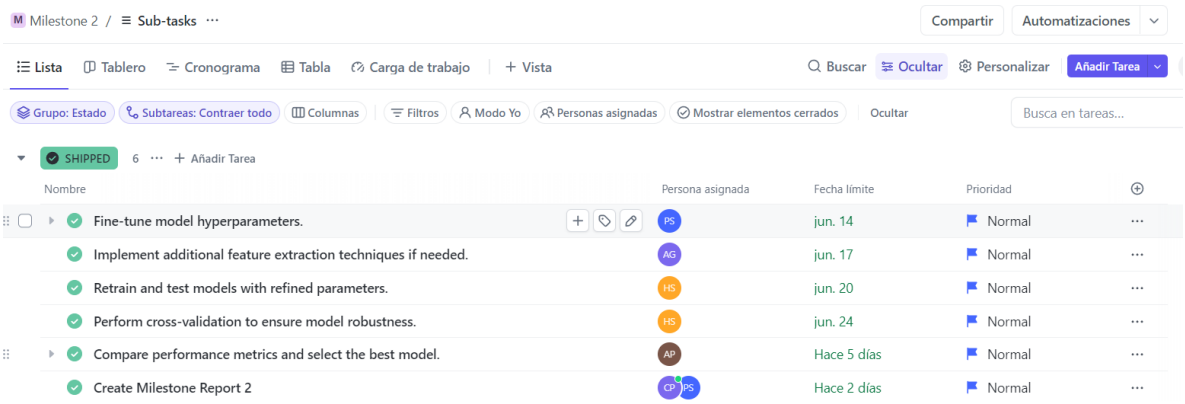
III. MODIFIED TIMELINE TABLE

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

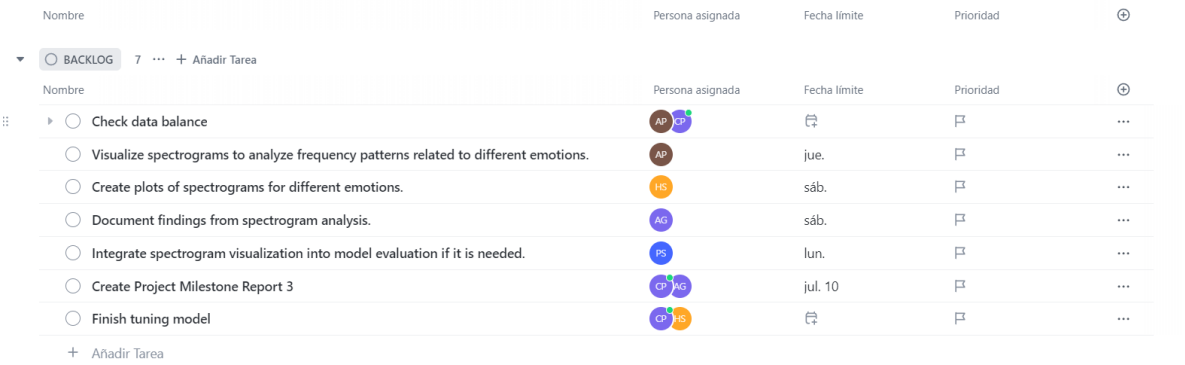
MILESTONE 1



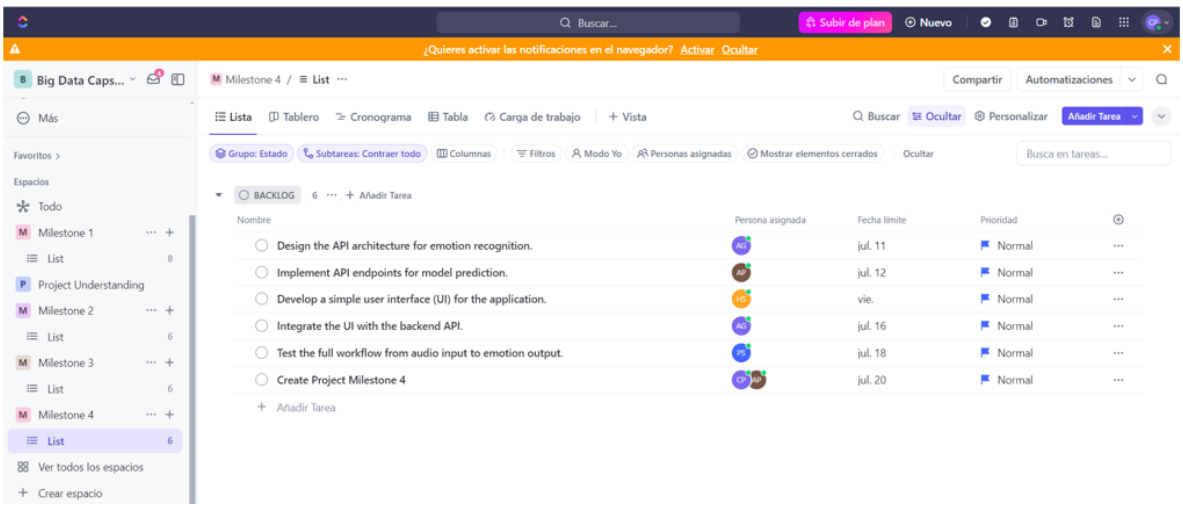
MILESTONE 2



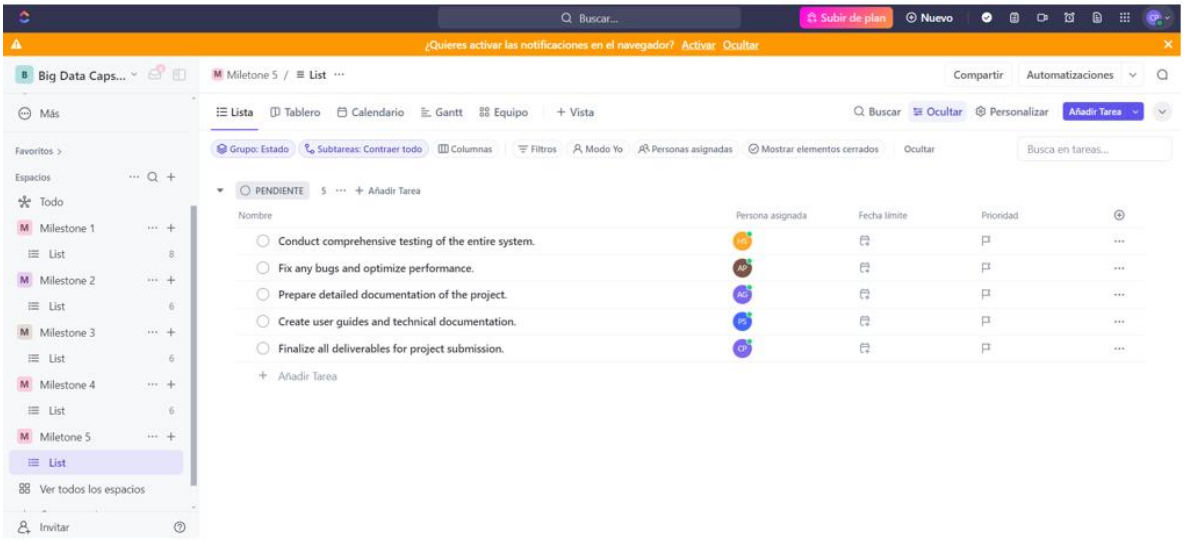
MILESTONE 3



MILESTONE 4



MILESTONE 5



IV. NEXT STEPS

UPCOMING STEPS AND ACTIVITIES PLANNED FOR THE PROJECT.

Brief Description of Tasks to be Undertaken

Milestone 3: June 30, 2024 - Model Evaluation and Refinement

1. **Finish tuning model:**

Description: Finalize the process of optimizing the hyperparameters of the machine learning model. This involves using techniques like Grid Search or Random Search to find the best combination of hyperparameters that yield the highest accuracy and performance on the validation dataset.

Purpose: To ensure that the model is performing at its best and is capable of accurately predicting emotions from speech data.

2. **Visualize spectrograms to analyze frequency patterns related to different emotions:**

Description: Generate and study spectrograms, which are visual representations of the spectrum of frequencies in a sound signal as it varies with time. These visualizations can help in understanding how different emotions affect the frequency patterns of speech.

Purpose: To gain insights into the distinctive frequency patterns associated with various emotions, which can be used to improve feature extraction and model accuracy.

3. **Create plots of spectrograms for different emotions:**

   Description: Create and plot spectrograms for each emotion category in the dataset. These plots will visually demonstrate the differences in frequency patterns for each emotion.

   Purpose: To provide a clear visual representation of how emotions influence speech patterns, facilitating better understanding and communication of these differences.

4. **Document findings from spectrogram analysis:**

   Description: Record and detail the observations and insights gained from analyzing the spectrograms. This includes noting any consistent patterns or anomalies that are associated with specific emotions.

   Purpose: To compile a comprehensive report on the frequency characteristics of emotions in speech, which can be referenced for further research and model improvement.

5. **Integrate spectrogram visualization into model evaluation if it is needed:**

   Description: Incorporate the spectrogram visualizations into the model evaluation process. This might involve using the insights gained from spectrogram analysis to refine the model or to visually validate the model's predictions.

   Purpose: To ensure that the spectrogram analysis contributes to the model's evaluation and improvement, making the model more robust and accurate.

EXPECTED OUTCOMES AND GOALS FOR THE NEXT PHASE

Optimized Model Performance: Achieving the highest possible accuracy and reliability of the speech emotion recognition model through meticulous tuning.

Insightful Spectrogram Analysis: Gaining a deeper understanding of the frequency patterns associated with different emotions, which can be used to refine the model further.

Comprehensive Documentation: Creating detailed documentation of the spectrogram analysis findings, which will serve as a valuable resource for future work and model enhancements.

V. CHALLENGES FACED

The biggest obstacle we had for this delivery was improving the accuracy of the model, we are still trying to improve it, but we have had to document ourselves and support ourselves with our teacher to continue finding better alternatives to optimize the parameters of the model and achieve a high, superior accuracy. of 77.

VI. LESSONS LEARNED

1. **Importance of Hyperparameter Tuning**:
    - **Lesson**: Fine-tuning hyperparameters is critical for optimizing model performance. Even small adjustments can significantly impact the accuracy and robustness of the model.
    - **Example**: Through Grid Search and Random Search, we discovered optimal parameter configurations that enhanced our model's accuracy.
2. **Value of Feature Engineering**:
    - **Lesson**: Effective feature extraction is essential for capturing the nuances of speech emotions. Exploring various features, such as MFCCs, chroma, and mel spectrograms, provided valuable insights into how emotions manifest in speech patterns.
3. **Balanced Datasets are Crucial**:
    - **Lesson**: Ensuring a balanced dataset is important for training a robust and unbiased model. Imbalanced datasets can lead to skewed predictions and lower overall accuracy.
    - **Example**: Identifying the need to check and potentially balance the dataset was a key insight for future improvements.
4. **Iterative Model Improvement**:
    - **Lesson**: Continuous iteration and testing are necessary to refine the model. Cross-validation helps ensure that the model generalizes well to unseen data.
    - **Example**: Performing cross-validation and comparing performance metrics across different models allowed us to select the best-performing model.
5. **Collaboration and Guidance**:
    - **Lesson**: Seeking guidance from mentors and collaborating with peers can provide new perspectives and solutions to challenges.
    - **Example**: Collaborating with our teacher helped us explore alternative parameter optimization strategies to achieve higher accuracy.

IV. CONCLUSSION

The milestone focused on fine-tuning the model's hyperparameters and enhancing feature extraction techniques to improve the accuracy and robustness of our speech emotion recognition model. Despite facing challenges in achieving the desired accuracy, we made significant progress in optimizing the model and understanding the intricacies of speech emotion features.

1. **Successful Hyperparameter Tuning**:
    - We successfully fine-tuned the model's hyperparameters, leading to an improved accuracy of 74%. This was achieved through extensive experimentation with different parameter configurations.
2. **Enhanced Feature Extraction**:
    - By incorporating additional feature extraction technique, we gained deeper insights. This helped in refining the model's feature set and improving its predictive capabilities.
3. **Model Robustness and Validation**:
    - Cross-validation played a crucial role in ensuring the model's robustness. It provided a reliable measure of the model's performance across different data splits, helping us select the best model configuration.
4. **Challenges and Future Directions**:
    - The primary challenge was achieving a high accuracy, which required continuous documentation, research, and collaboration with our teacher. We identified the need to check the dataset balance as a crucial step for future improvement of the model tuning.