

# PRML Major Project

## Speech Emotion Recognition

---

Arvind Kumar Sharma (B21AI006)

Nitish Bhardwaj (B21AI056)

Renu Sankhla (B21AI028)

28 April, 2023

## Contribution

Arvind Kumar Sharma	Nitish Bhardwaj	Renu Sankhla
Dataset Creation	Dataset Creation	Dataset Creation
Mel Spectrogram Dataset	Mel Spectrogram Dataset	Mel Spectrogram Dataset
FFT Dataset	FFT Dataset	FFT Dataset
Decision Tree	Naive Bayes	KNN
Adaboost	Random Forest	SVM
ANN with sigmoid activation	ANN with Tanh activation	ANN with Relu activation

## Problem Statement: **Speech Emotion Recognition**

Speech Emotion Recognition, abbreviated as SER, attempts to recognize human emotion and affective states from speech. This capitalizes on the fact that voice often reflects underlying emotion through tone and pitch. This is also the phenomenon that animals like dogs and horses employ to be able to understand human emotion.

### Pipeline:

1. Data Preprocessing
2. Data Cleaning
3. Feature Engineering
4. Model Selection
5. Prediction
6. Conclusion

#### 1. Data preprocessing :

As data, we were given audio files of different emotions. The librosa library is utilized in order to convert audio files to a numpy array. These numpy arrays were added to the dataset, which also includes the information extracted from their labels. The dataset is saved for future analysis and use.

#### 2. Data cleaning:

Since the audio files were of different intervals, only the first 2.9 seconds of these audio files were used in the creation of the dataset.

### 3. Feature Engineering :

Since the amplitude at the sampled instant is not sufficient to determine the emotion and incorporate the pitch and other acoustic parameters, the dataset was subjected to two additional transformations, and two additional datasets were generated.

The first transformation was the **Fourier transform**. We have only measured its magnitude, not its phase. As a result, training for emotion detection will now be based on the frequency of speech.


The second transformation was the spectrogram of the **mel spectrum**. Now, both time and frequency domains are utilized during model training.

For the sake of complexity, PCA and LDA are used to reduce the dimension to 500 and 7 respectively.

### 4. Model Selection

In terms of model selection, we have chosen a wide variety of models for the problem statement. These are the respective types:

1. **KNN** : We chose KNN because it is a non-parametric algorithm, meaning that it does not make any assumptions about the distribution of the data and can work well with both linear and non-linear data. It works by finding the k closest training data points to a new data point and assigning the new point the label or value that is most common among its k nearest neighbors.

- 
2. **SVM** : The SVM model trained on the Fourier transform and mel spectrum features of the Ravdess dataset performed well in predicting the emotional labels of speech recordings, as evidenced by the high accuracy, precision, and recall scores.
  3. **Decision Tree**: The decision tree was chosen on the assumption that it would classify speech according to its frequency. It is believed that a particular frequency manifests itself in an emotion. Therefore, the decision tree was our first choice.
  4. **Adaboost**: After discovering the poor performance of the decision tree, it was decided to use it as a weak learner and apply Adaboost boosting.
  5. **Random Forest**: Using Random Forest for the RAVDESS dataset offers the potential for capturing complex dependencies in audio data through an ensemble of decision trees, allowing for more accurate emotion classification. It can handle high-dimensional feature spaces, provide feature importance rankings, and generalize well to unseen data.
  6. **Naive Bayes**: It calculates the probability of an emotion using prior probabilities and conditional probabilities of features.
  7. **ANN**: ANN learns the feature on its own. Since none of the aforementioned methods were effective, we have trained various ANNs with distinct activation functions.

## 5. Prediction

## KNN (Number of Neighbors: 8)

---

### KNN Classifier Report on Fourier Transform Dataset

Model Performance	Accuracy: 0.958	Precision: 0.958	Recall: 0.958
-------------------	-----------------	------------------	---------------

### KNN Classifier Report on the Mel Spectrogram Dataset

Model Performance	Accuracy: 0.972	Precision: 0.972	Recall: 0.972
-------------------	-----------------	------------------	---------------

## SVM (Kernel = Linear)

---

### SVM Report on Fourier Transform Dataset

Model Performance	Accuracy: 0.954	Precision: 0.954	Recall: 0.954
-------------------	-----------------	------------------	---------------

### SVM Report on Mel Spectrogram Dataset

Model Performance	Accuracy: 0.958	Precision: 0.958	Recall: 0.958
-------------------	-----------------	------------------	---------------

## SVM (Kernel = rbf)

---

### SVM Report on Fourier Transform Dataset

Model Performance	Accuracy: 0.972	Precision: 0.972	Recall: 0.972
-------------------	-----------------	------------------	---------------

## SVM Report on Mel Spectrogram Dataset

Model Performance    Accuracy: 0.958    Precision: 0.958    Recall: 0.958

## Decision tree

---

### Decision Tree Report on Fourier Transform Dataset with PCA

Weighted Accuracy with Gini : 0.29    Entropy: 0.19 log-loss : 0.17

### Decision Tree Report on Fourier Transform Dataset with LDA

Weighted Accuracy with gini : 0.29    entropy : 0.16    log-loss : 0.16

### Decision Tree Report on Mel Spectrogram Dataset with PCA

Weighted Accuracy with gini : 0.23    entropy : 0.21    log-loss : 0.25

### Decision Tree Report on Mel Spectrogram Dataset with LDA

Weighted Accuracy with gini : 0.23    entropy : 0.24    log-loss : 0.21

## Adaboost

---

### Ada boost Report on Fourier Transform Dataset with PCA and max depth = 7

Weighted Accuracy with gini : 0.24    entropy : 0.22    log-loss : 0.17

Ada boost Report on Fourier Transform Dataset with LDA and max depth = 7

Weighted Accuracy with gini : 0.21    entropy : 0.25    log-loss : 0.26

Ada boost Report on Fourier Transform Dataset with PCA and hyperparameter tuning of decision tree

Weighted Accuracy with gini : 0.24    entropy : 0.22    log-loss : 0.22

Ada boost Report on Fourier Transform Dataset with LDA and hyperparameter tuning of decision tree

Weighted Accuracy with gini : 0.24    entropy : 0.22    log-loss : 0.22

Ada boost Report on Mel Spectrogram Dataset with PCA and max depth = 7

Weighted Accuracy with gini : 0.30    entropy : 0.27    log-loss : 0.24

Ada boost Report on Mel Spectrogram Dataset with LDA and max depth = 7

Weighted Accuracy with gini : 0.31    entropy : 0.31    log-loss : 0.28



Ada boost Report on Mel Spectrogram Dataset with PCA and hyperparameter tuning of decision tree

Weighted Accuracy with gini : 0.24    entropy : 0.20    log-loss : 0.20

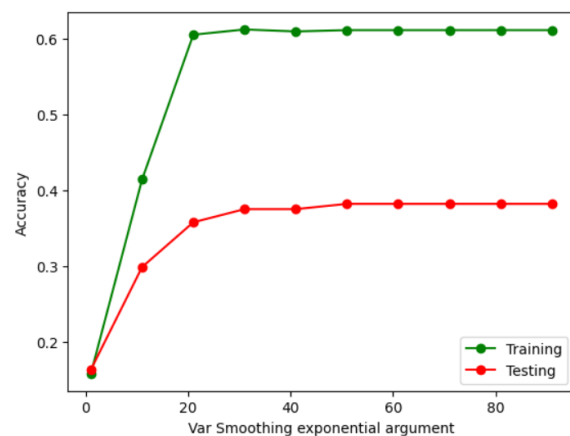
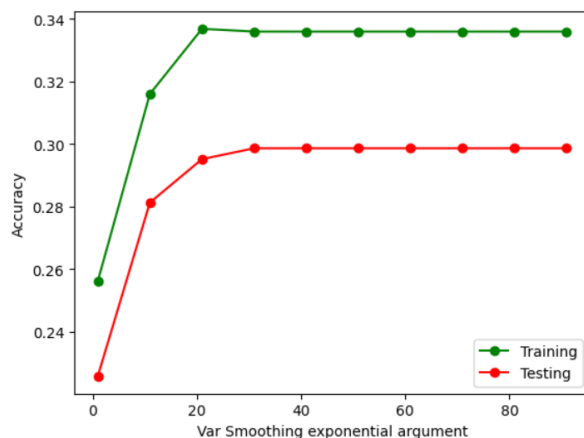
Ada boost Report on Mel Spectrogram Dataset with LDA and hyperparameter tuning of decision tree

Weighted Accuracy with gini : 0.24    entropy : 0.20    log-loss : 0.20

## Naive Bayes

- In Naive Bayes, we have only one hyper-parameter “var smoothing” to vary. “Var smoothing” is used to handle zero probability problems.

First graph is of Fourier Transform Dataset and other of Mel spectrogram Dataset.



Naive Bayes Report upon best hyper-parameters on Fourier Transform Dataset

Model Performance Accuracy: 0.32 Precision: 0.33 Recall: 0.32 F1-score: 0.30

## Naive Bayes Report upon best hyper-parameters on Mel Spectrogram Dataset

Model Performance Accuracy: 0.36 Precision: 0.34 Recall: 0.45 F1-score: 0.39

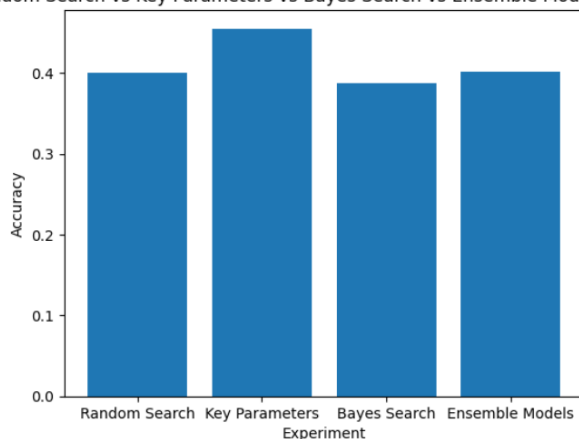
## Random Forest

---

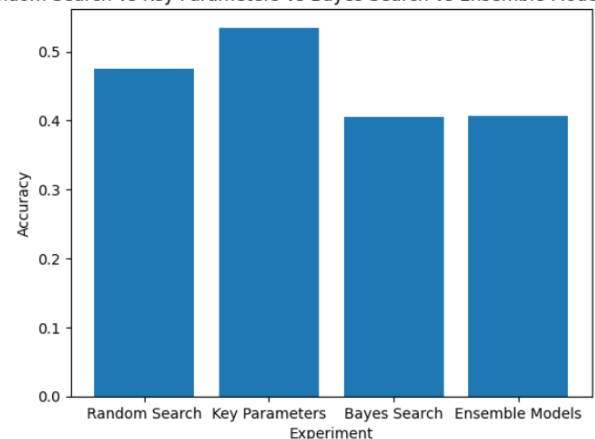
- To do effective hyper-parameter tuning, we employed methods to check performance upon
  - Default model
  - Randomized Search CV & Key Parameters
  - Bayes Search CV
  - Ensemble Models

First graph is of Fourier Transform Dataset and other of Mel spectrogram Dataset.

Random Search vs Key Parameters vs Bayes Search vs Ensemble Model Accuracy



Random Search vs Key Parameters vs Bayes Search vs Ensemble Model Accuracy



## Random Forest Report upon best hyper-parameters on Fourier Transform Dataset

Model Performance Accuracy: 0.410 Precision: 0.443 Recall: 0.403 F1-score: 0.379

## Random Forest Report upon best hyper-parameters on Mel Spectrogram Dataset

Model Performance Accuracy: 0.458 Precision: 0.436 Recall: 0.438 F1-score: 0.432

### ANN

#### FFT

	Relu	Tanh	Sigmoid
LDA Accuracy	0.604	0.941	0.948
PCA Accuracy	0.385	0.319	0.342

#### Mel Spectrum

	Relu	Tanh	Sigmoid
LDA Accuracy	0.649	0.990	0.993
PCA Accuracy	0.355	0.233	0.295



## 6.Conclusion :

- Decision and Adaboost did not perform properly as a fixed frequency cannot be used for classification of emotion. ANN with Sigmoid activation function performs very well.
- Naive Bayes and Random Forest did not work well on the RAVDESS dataset due to complex relationships between features and emotions and the limited expressiveness of the models in capturing intricate patterns in audio data. Therefore, we worked upon ANN with Tanh activation function which performed well.
- SVM and KNN performed well but ANN with ReLU as activation function didn't perform well due to its simple processing and was not able to effectively learn the intricate patterns and nuances present in the audio data.

