[Gabriel Bermudez, 3965-6965]

CIS4930 Individual Coding Assignment 3

Spring 2023

1. Problem Statement

[The first step in any machine learning project is defining your problem. Here you need to present a high-level problem statement, which may include a description of the problem, why this problem needs to be solved, and how did you solve it.]

The problem consisted of creating a Speech Emotion Recognition (SER) system to identify speech emotion without the need of semantic context. To do this, a subset of the Toronto Emotional Speech Set (TESS) dataset was used, consisting of 400 audio clips for 4 emotions: angry, sad, happy, and fear. Acoustic feature extraction was carried out using the *librosa* library, the outputs of which were used to train three classifier models: SVM, Naive Bayes, and Random Forest.

2. Data Preparation

[Preparing data for training machine learning models is a fundamental step. Describe your data preparation process in detail, which may include how did you clean the data and how did you extract features from the data for the following model training steps.]

Split dataset into training and testing sets

A. Store paths and labels of downloaded audio files.
B. Split data using *train_test_split* with test size 0.3 and random state 42.
C. This results in a *train* dataset size of 280 and *test* dataset size of 120.

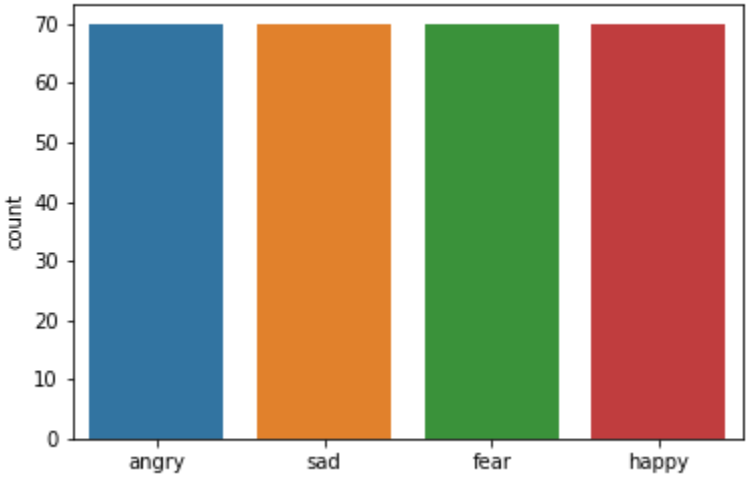
Exploratory data analysis:

A. Create *audio*, *waveplot*, *spectrogram*, and *distribution* functions which display audio files, wave plots (time or frequency), spectrograms, and the data distribution, respectively.
B. Showcase functionality with random examples [see table].

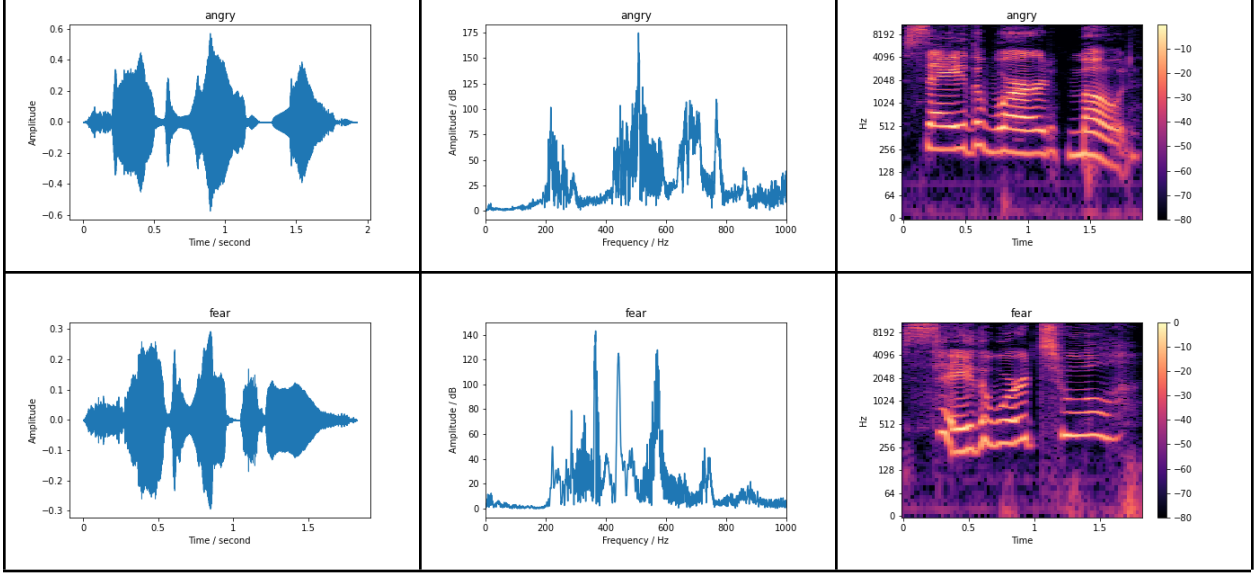
Feature Extraction & Feature Post Analysis:

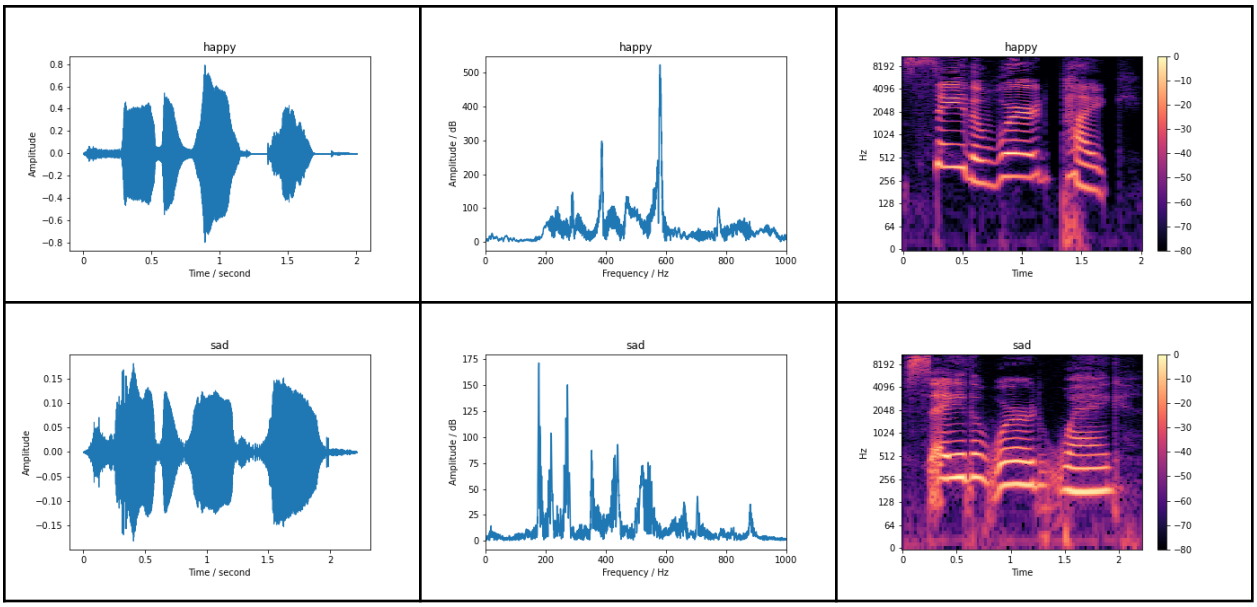
A. Create *scale* function which normalizes a given array to the [-1, 1] range.
B. Extract, scale, and average each feature (loudness, mfcc, zcr, chroma, and melspectrogram).
C. Generate a feature vector for each *.wad* file.
D. Combine all feature vectors to create the train and test datasets.
E. Encode labels (emotions) in decimal values.

Distribution



Exploratory Data Analysis





3. Model Development

- Model Training
  - *[Describe the training phrase, which may include what models did you select, how you split training/validation/test sets, training epochs, and any other parameters.]*

    - The three classification models used were SVM, Naive Bayes, and Random Forest, all imported from the Sklearn library.
    - SVM classifier was initialized with a linear kernel parameter.
    - All other models were run with standard configurations and parameters.
    - Models were fit to train data and labels.

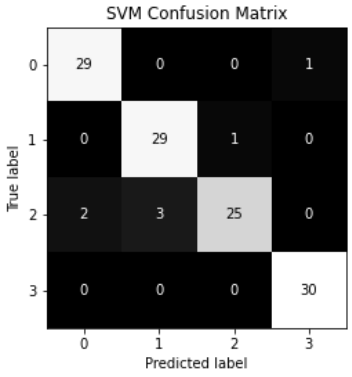
- Model Evaluation
  - *[Present the results of your models, basic evaluation looks at metrics such as accuracy, precision, or F1 score to determine which model is the best fit to solve the problem. You may find the model performance not good enough, in which case you can experiment further to improve the model performance with different features or more complex deep learning models.]*

    - All models performed exceptionally well, with accuracy scores greater than or equal to 0.94. Model performance appears to be proportional to the number of features. The higher the number of mfcc and melspectrogram features, the better the performance. With n_mfcc and n_mels set to 12, the models performed with roughly 0.80 accuracy, as opposed to the near-perfect accuracy of those trained

with n_mfcc and n_mels set to 60 (134 total features). The specific, final results are as follows:

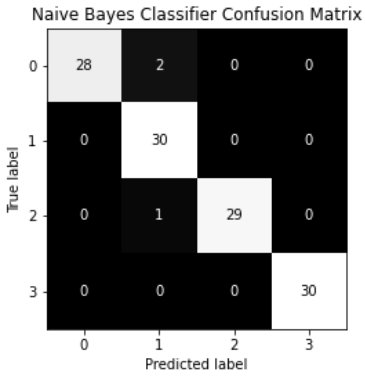
SVM

	precision	recall	f1-score	support
0	0.94	0.97	0.95	30
1	0.91	0.97	0.94	30
2	0.96	0.83	0.89	30
3	0.97	1.00	0.98	30
accuracy			0.94	120
macro avg	0.94	0.94	0.94	120
weighted avg	0.94	0.94	0.94	120



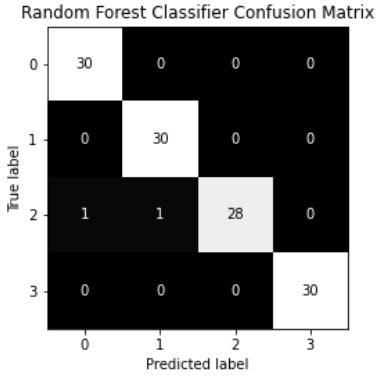
Naive Bayes Classifier

	precision	recall	f1-score	support
0	1.00	0.93	0.97	30
1	0.91	1.00	0.95	30
2	1.00	0.97	0.98	30
3	1.00	1.00	1.00	30
accuracy			0.97	120
macro avg	0.98	0.97	0.98	120
weighted avg	0.98	0.97	0.98	120



Random Forest Classifier

	precision	recall	f1-score	support
0	0.97	1.00	0.98	30
1	0.97	1.00	0.98	30
2	1.00	0.93	0.97	30
3	1.00	1.00	1.00	30
accuracy			0.98	120
macro avg	0.98	0.98	0.98	120
weighted avg	0.98	0.98	0.98	120



4. Discussion

- *[Does your model perform well enough? Discuss the potential reasons that your model fixes the problem well or not]*

  The models performed well beyond my expectations. This may have been due to one of two causes: either feature extraction and post-processing was performed exceptionally well, or a certain oversight artificially skewed the results. I hope, and suspect, it was the former.

- *[What are the challenges you met during both the data preparation and model development processes? How did you solve them?]*

  Surprisingly, I did not encounter too many challenges this time around. Of course I faced several obstacles along the way, but these were not too time-consuming and mostly due to poor conceptual understanding and a lack of technical know-how.

- *[Any reflections or thoughts on this assignment?]*

  This was a great assignment. It was very interesting, and just the right amount of challenging.

5. Appendix

- https://github.com/skinnii/multimodal-ml