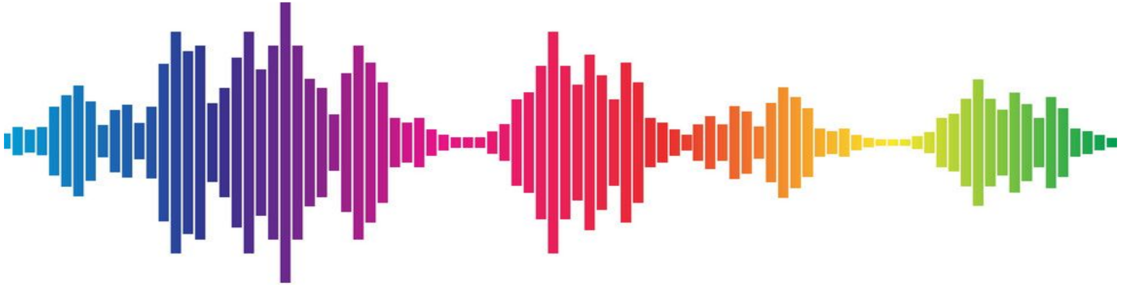
Speech Emotion Recognition (SER) through Machine Learning

Submitted By:
Avani Pal
Apurva Panchal

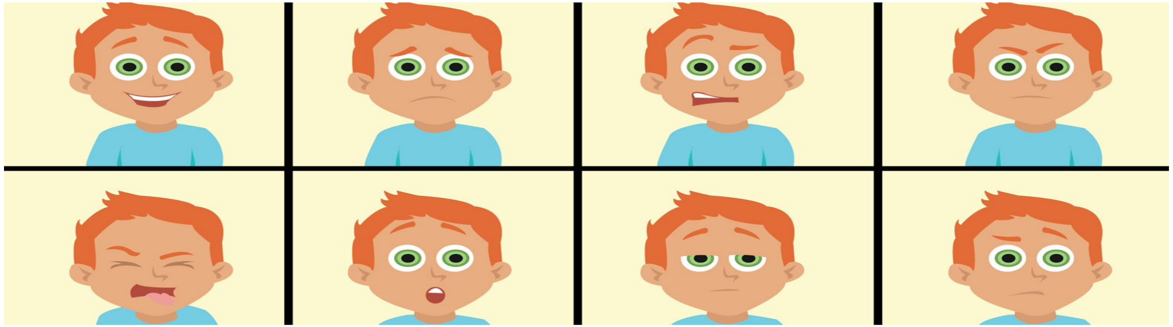
DATABASE USED:

The portion of the RAVDESS contains 2880 files: 60 trials per actor x 24 actors = 2880. The RAVDESS contains 24 professional actors (12 female, 12 male), vocalizing two lexically-matched statements in a neutral North American accent, fearful, surprise, and disgust expressions. Each expression is produced at tro levels of emotional intensity (normal, strong), with an additional neutral expression.

ABSTRACT



Speech is the fundamental strategy for us to communicate with each other. However, emotions play a vital role in communication. It is a medium of expression of one's perspective to others. It includes several emotions such as anger, happiness, sadness, fear, passion, disgust, etc. It has been a topic of research for a long time but it is a challenging task to perform as human emotions are very subjective and sometimes it is even harder for humans to notate them.The main aim of this project is to classify the emotional state of the speaker from the speech signal. In this, we will be making SER (Speech Emotion Recognition) which identifies speech signals to detect the emotions underlying them. In particular, we will be creating a classification model elicited by speeches based on Deep Learning and Machine Learning algorithms ie. (MLP classification, CNN classifiers, SVM) by analysing the acoustic features.

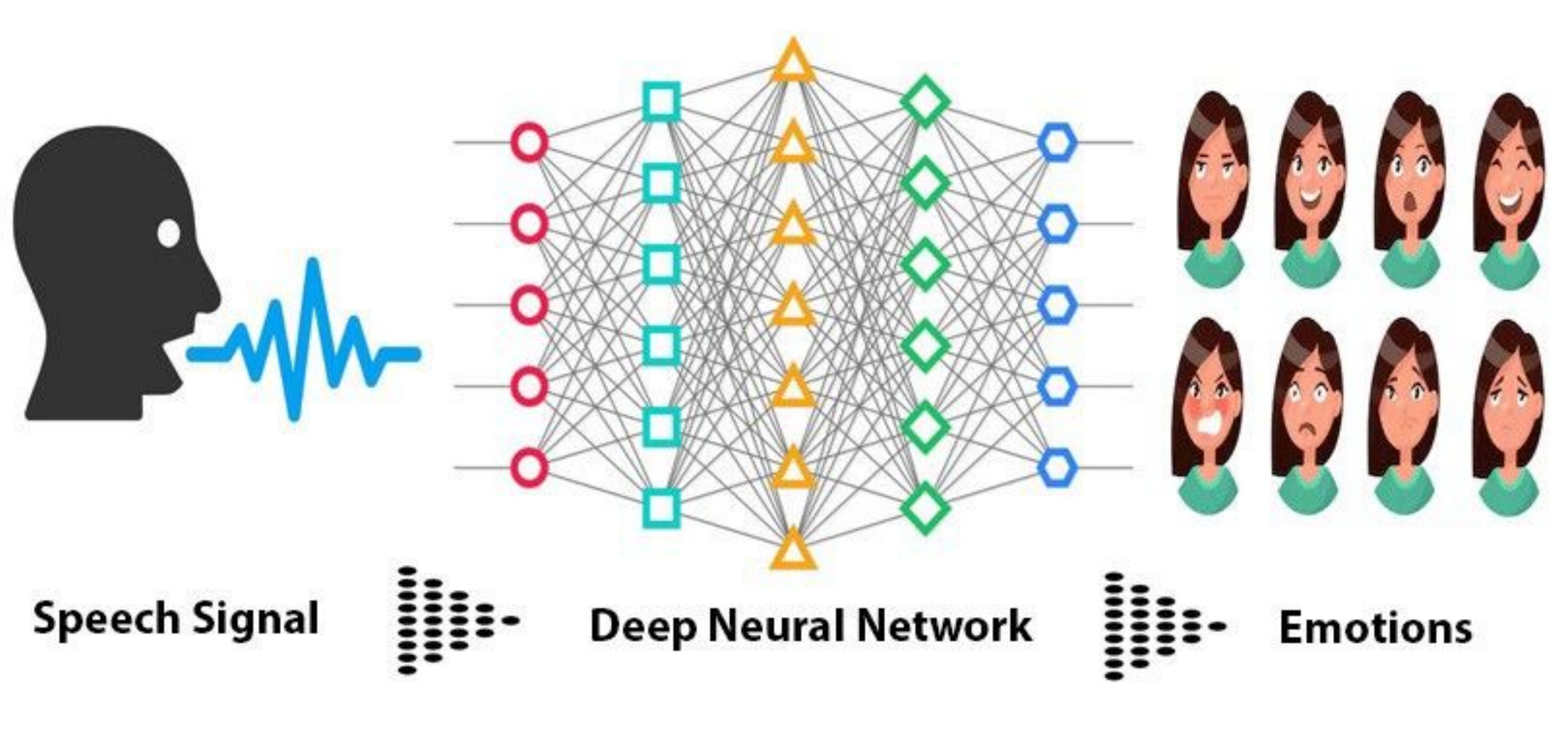


Emotion refers to the wide range of affective processes such as moods, feelings, affects and well-being.

Several machine learning paradigms were used for the emotion classification task. A recurrent neural network (RNN) classifier is used first to classify seven emotions. Their performances are compares later to multivariate linear regression (MLR) and support vector machine (SVM) techniques, which are widely used in the field of emotion recognition for spoken audio signal.

Voices are an important modality for emotional expression. Speech is a relevant communicational channel enriched with emotions: the voice in speech not only conveys semantic message but also the information about the emotional state of the speaker.

Basic Analogy of Project



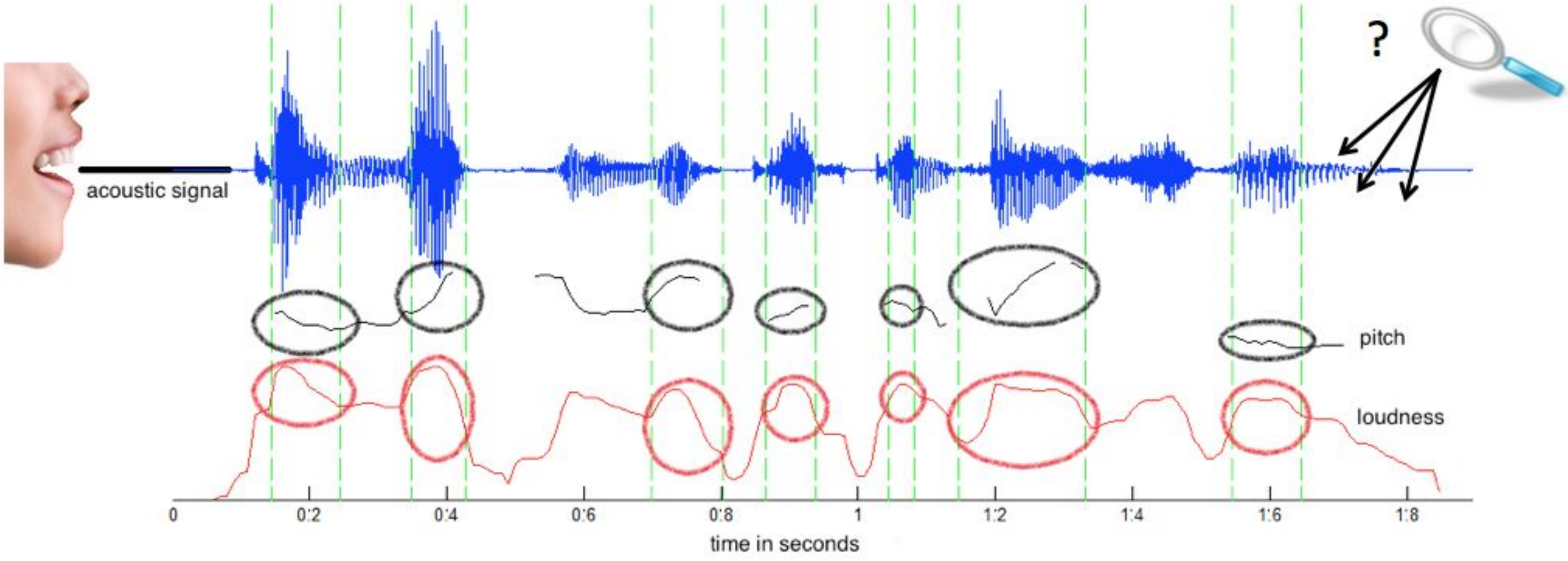
IDEOLOGY BEHIND THE PROJECT

The idea behind creating this project was to build a machine learning that could detect emotions from the speech we have with each other all the time.

So, why not have an emotion detector that will gauge your emotions and in the future recommend different things based on your mood. This can be used by multiple industries to offer different services like marketing companies suggesting you buy products based on your emotions, the automotive industry can detect the emotions and adjust the speed of autonomous cars as required to avoid any collisions etc.

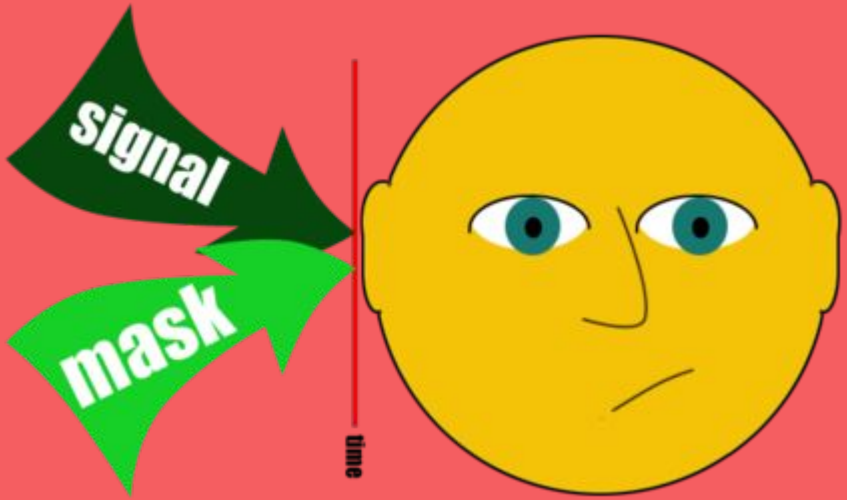
Analyzing AUDIO Signals:

The first step is to test out the audio files by plotting out the waveform and spectrogram to see the sample audio files.



MASKING AND CLEANING OF AUDIO FILES:

Next step is to clean the audio files by lowering down the sample rate removing the unwanted noise around the raw audio via MASKING.

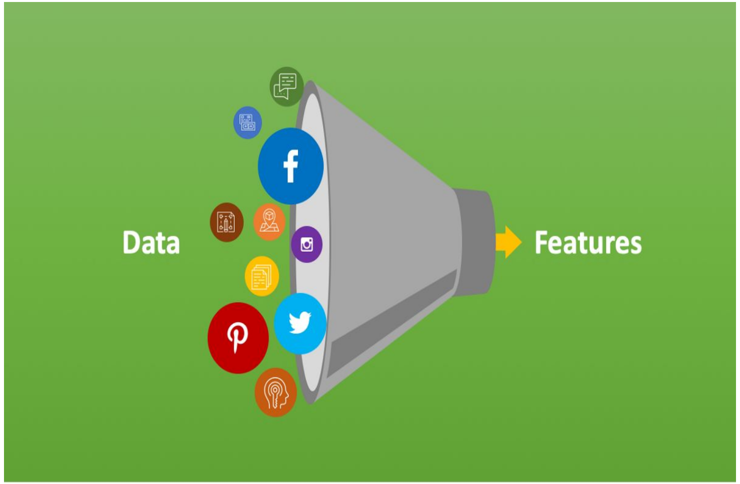


FEATURE EXTRACTION:

Feature Extraction aims to reduce the number of features in a dataset by creating new features from the existing ones.

The next step involves extracting the features from audio files which will help our model learn between these audio files. For feature extraction we make use of the LIBROSA library python which is one of the libraries used for audio analysis.

Also there are labels of EMOTIONS defined, When the Clean Dataset is loaded with the calling of Feature Extraction process, every audio is classified into the labels defined.



BUILDING THE MODEL

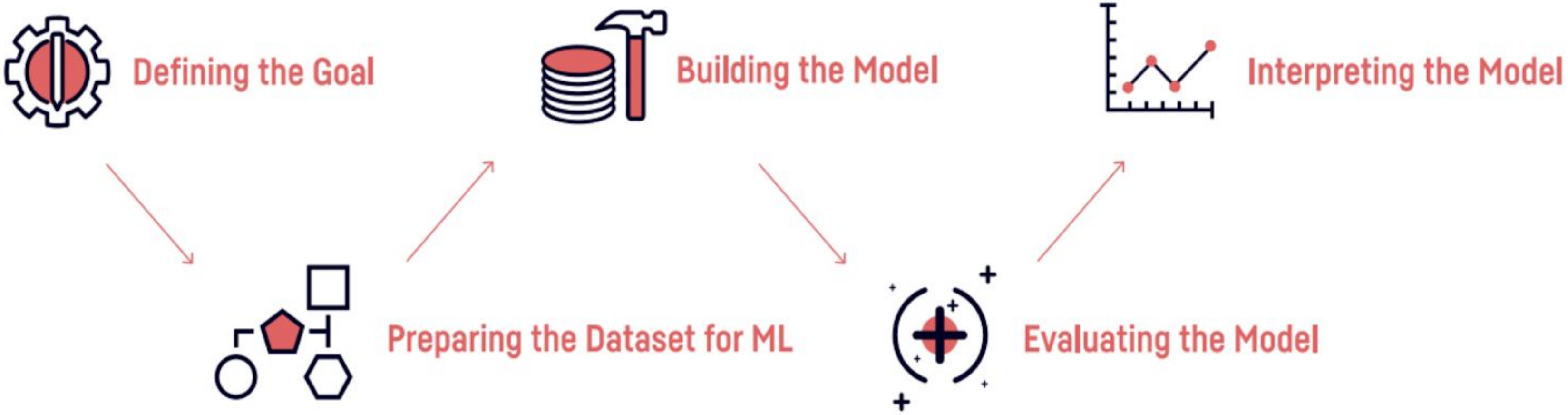


Figure I-5: Key stages of building a machine learning model.

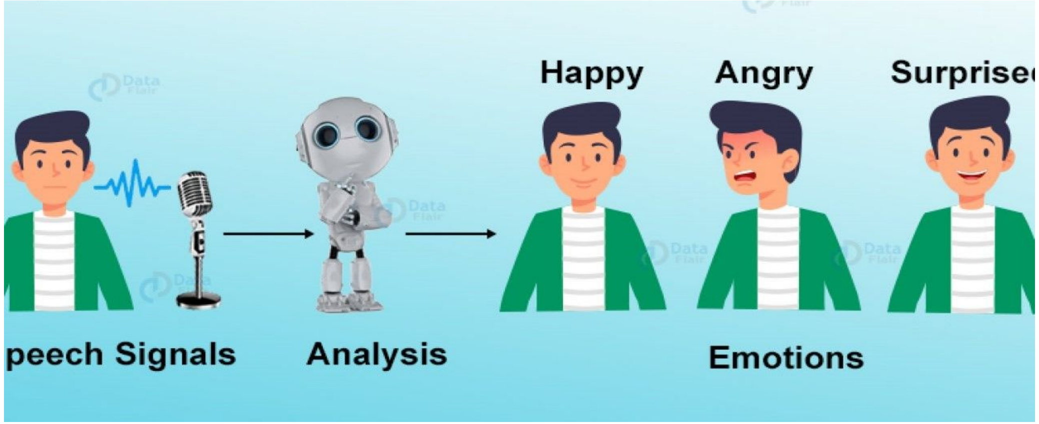
BUILDING THE MODEL

Since the project is a classification problem, MultiLayer Perceptron seems the obvious choice. We chose this model to predict the right emotions.

This classifier connects to the neural network. Unlike other classification algorithms such as Support Vector Machine or Naive Bayes Classifier, MLPClassifier relies on an underlying Neural Network to perform the task of classification.

PREDICTIONS:

After tuning the model, tested it out by predicting the emotions for the test data. Following the splitting of training and testing data to saving the model. Model is loaded again to predict the test data and store its result in .csv file along with its labels for mapping individual result to its wav file name



LIVE DEMO FOR PREDICTION:

Conclusion and Challenges:

- The literature in speech emotion recognition is not very rich and researchers are still debating what features influence the recognition of emotion in speech. There is also considerable uncertainty as to the best algorithm for classifying emotion, and which emotions to class together.
- In the real problems, different individuals reveal their emotions in a diverse degree and manner. There are also many differences between acted and spontaneous speech.

This model can be used by various apps, online shopping website and so on to know about the user’s emotions. Further improvements can be made to the model so that it can perform well in real time. For improving the accuracy of the model, we can increase the size of the dataset. The classifier can be embedded in a software or an app so that it can work in real time . Moreover, we look forward to come up with more industrial application of this model.

Due to less availability of datasets our accuracy was not as expected. So, in further work we can increase the number of datasets so as to get higher accuracy.

Future Scope:

- You can try different other classifiers to predict the emotions the emotion behind the audio like SVM, CNN, etc.
- Predicting the live audio takes a lot of processes and its sometimes difficult to process as it is unlike the binary data with some csv file associated with it.
- In future, we can predict the random recorded audio as well.
- We can also embedded our UI interface with ML model. And we can build web based app with ML using flask in future, this point can be a great scope.
- More advancements can be more variety of voices can be trained and dataset can be increased to deploy more realistic model.

Thank You!