cogito

Intelligence Augmentation (IA) for AI

Our Approach →

Team: ML Enthusiast

Overview

Detecting emotions is one of the most important marketing strategies in today’s world. We could personalize different things for an individual specifically to suit their interest. Getting the emotion of the customers will improve the outcome of the products. Normally, emotion recognition is done through the text data, but we had to do it with audio data for this challenge.

Goals

These were the goals that we decided to work on to improve our pipeline.

Efficiency

Our pipeline should audio files efficiently and generate useful features for training.

Performance

Our developed model should perform well on unknown data as well as noisy data.

Latency

Our model should have less latency and should require fewer resources to evaluate audio files.

Preprocessing

We used the Librosa library in Python to process and extract features from the audio files. Librosa is a python package for music and audio analysis. It provides the building blocks necessary to create music information retrieval systems.

Using the librosa library, we extracted features, i.e. MFCC(Mel Frequency Cepstral Coefficient). MFCCs are a feature widely used in automatic speech and speaker recognition.

We then compressed the 2D MFCC features to 1D by extracting mean, std, skew, high, low, etc., along the y axis. This preserves maximum features from the 2D MFCC and enables us to use ML models.

We now repeat this process for various other augmentations, which further helps the model to generalize better on the Test Dataset

Modeling

After trying many models, We finally selected 3 best performing models for our final pipeline. The predictions are generated by weighted(8:2:1) soft voting of these 3 models:

LGBM Classifier

OOF AUC: 83.00

OOF Accuracy: 59.00

ExtraTrees Classifier

OOF AUC: 82.80

OOF Accuracy: 59.26

RandomForest Classifier

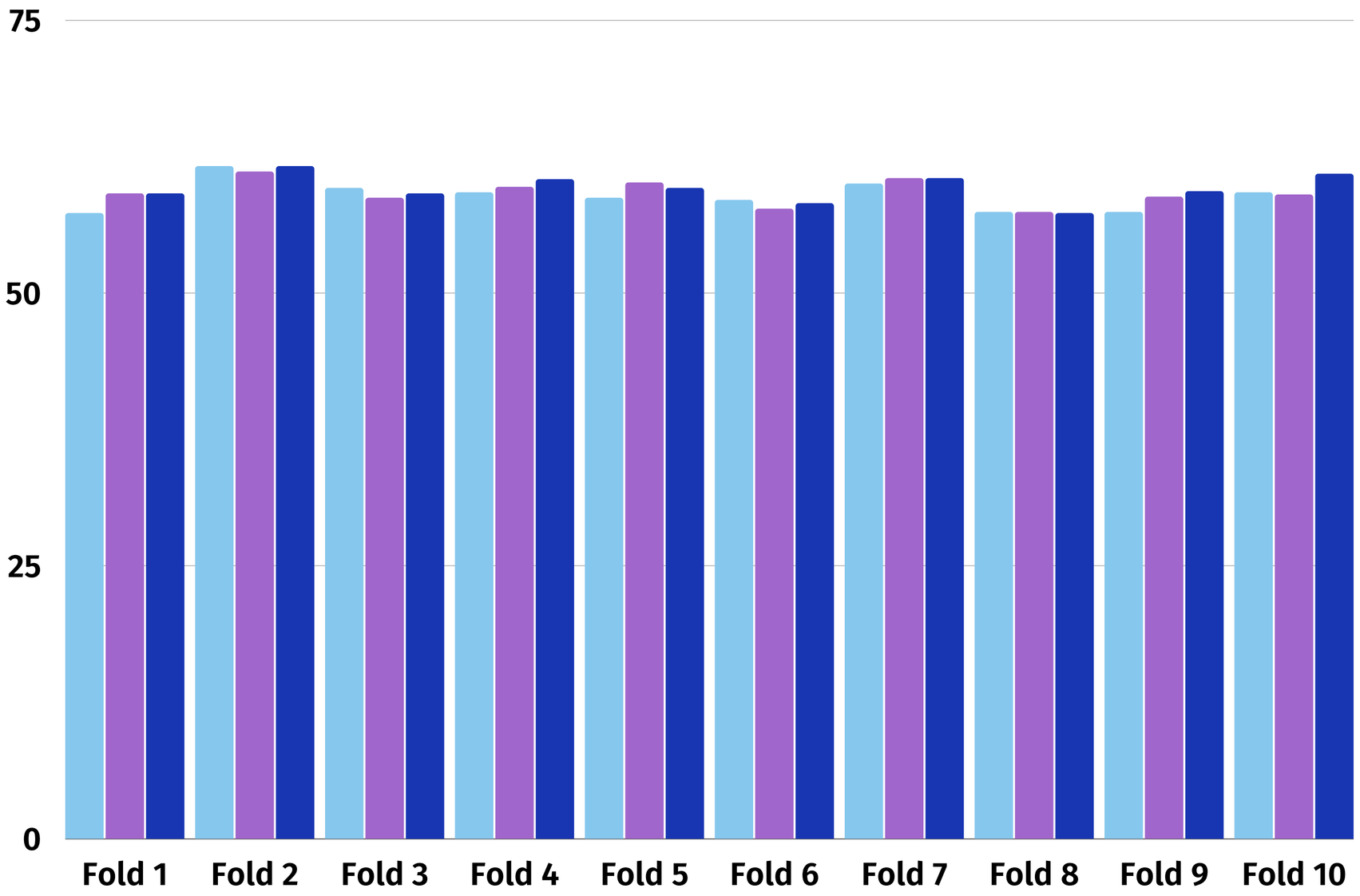
OOF AUC: 83.76

OOF Accuracy: 59.64

K-Fold Cross Validation Scores

To train the model on complete data and help the model generalize better on unseen data, We used Stratified K-Fold Cross-Validation.

- LGBM Classifier
- Extra Trees Classifier
- Random Forest

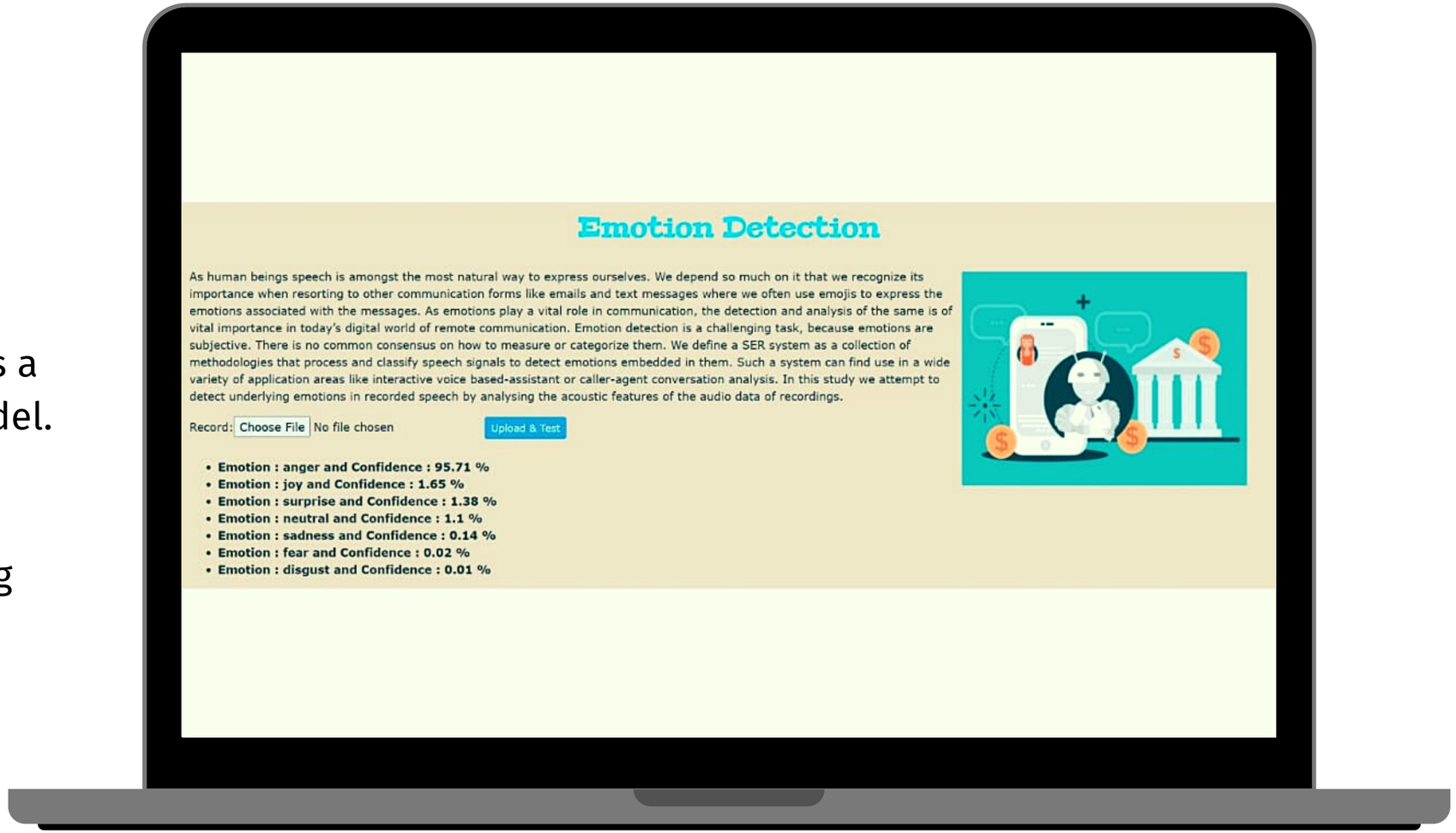


Post Processing

After analysing the test data, We found that the data was very imbalanced. Neutral class dominated the data. Neutral audio files contained much noise and were not easily identifiable, But as it was in the majority, most of the models had a high bias towards it. As we did not have any idea about the distribution of the testing set, We decided to use post-processing to remove bias. For this, We extracted those cases in which our model was about 50-51% confident of the audio being neutral. Then we gave the next most confidently predicted class for these cases.

Final Product

We built a Web-based Application using Django as a frontend to deploy our model. Now, the user just needs to upload the audio files and within 10-20 sec (depending upon the size) the model predicts the emotion.



Meet our Team

Kshitij Mohan

Undergraduate IIIT-D

Madhav Mathur

Undergraduate NSUT