SPEECH EMOTION RECOGNITION ITSP 2020
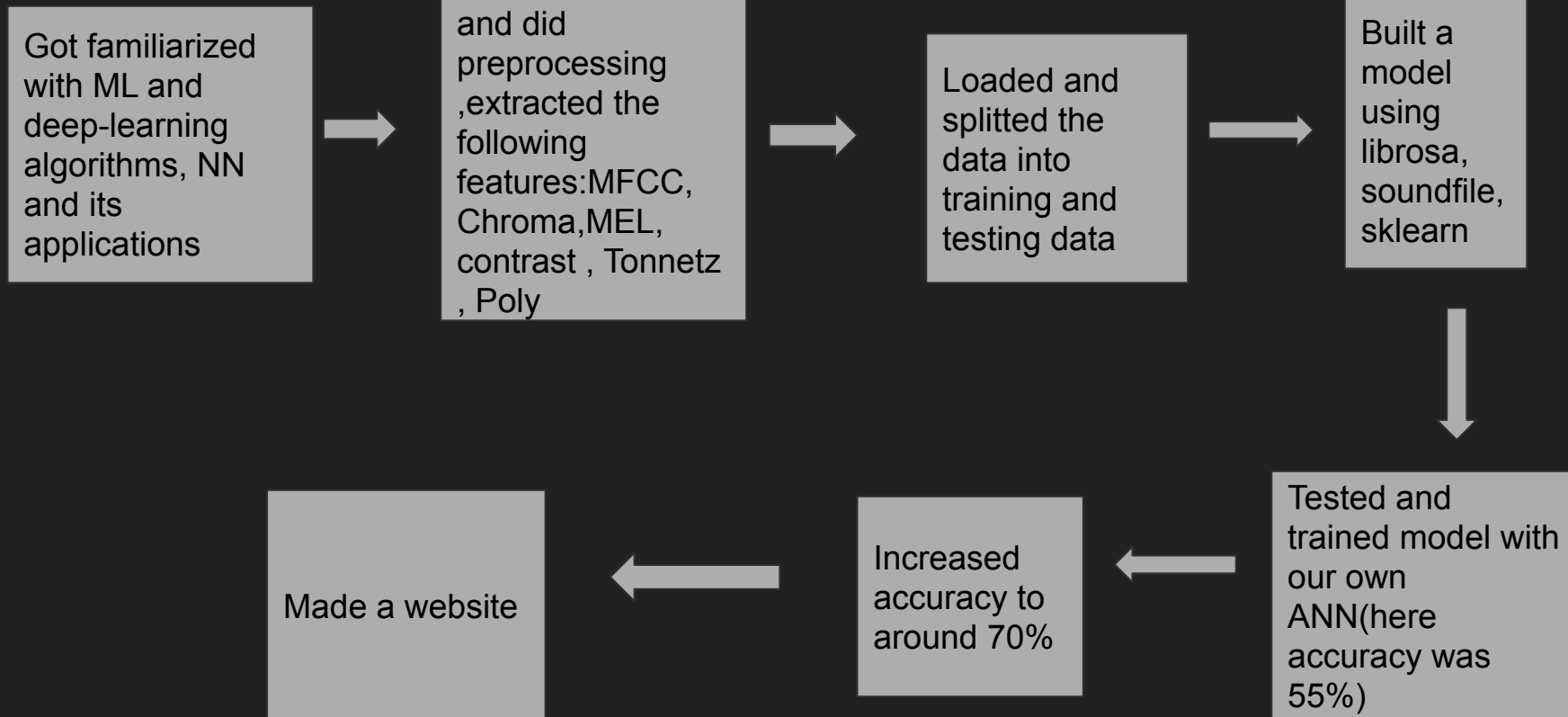
Shivam Modi
Prasada Thombare
Sakshi Kasralikar
Shiv Modi

Idea behind the project

- Speech Emotion Recognition, abbreviated as SER, is the act of attempting to recognize human emotion and affective states from speech.

- Emotion recognition is the part of speech recognition which is gaining more popularity and need for it increases enormously.

- The main motive of doing this project is to gain experience in the field of Machine Learning using Python.

- As **SER** is one of the applications of Advanced Machine Learning, by working on this project we intend to experience the essence of Machine Learning.

Workflow

Got familiarized with ML and deep-learning algorithms, NN and its applications

→

Found dataset and did preprocessing ,extracted the following features:MFCC, Chroma,MEL, contrast , Tonnetz , Poly

→

Loaded and splitted the data into training and testing data

→

Built a model using librosa, soundfile, sklearn

↓

Made a website

←

Increased accuracy to around 70%

←

Tested and trained model with our own ANN(here accuracy was 55%)

Project Details

- Humans are emotional beings and their emotions play a crucial role in understanding their thoughts which further control their actions.
- We attempted to create a website where we can take Speech as an input and output the emotion associated with me.
- We found two libraries i.e: librosa and soundfile as an important part of our project as we used them to extract 6 important features of human speech: [1) MFCC 2) Chroma 3) MEL 4) Contrast 5) Tonnetz 6)Poly] which are essential for understanding emotions by making a function extract_feature.
- For this project, we used CREMA-D dataset.

- Currently our goal is to find 4 emotions only i.e: Happy, Sad, Neutral and Angry because these are basically general emotions and adding more emotions can affect our accuracy as well.
- We got 195 features after applying feature scaling and we also to_categorical on the emotions for our ANN.
- Our ANN looks like this:-

```
] # create the model
new_model = Sequential()
new_model.add(Dense(780, input_shape=(195,)))
new_model.add(Dense(780, activation='relu'))
new_model.add(Dropout(0.3))
new_model.add(Dense(195, activation='relu'))
new_model.add(Dropout(0.2))
new_model.add(Dense(4, activation='softmax'))
new_model.compile(loss='categorical_crossentropy',optimizer='adagrad' , metrics=['accuracy'])
print(new_model.summary())
```

- Our test_size was 0.25 and we got the accuracy of nearly 70%.

Results (Images/Videos)

https://sakshikasralikar.github.io/speech_emotion/

Skills learnt

The biggest motivation for doing this project was to learn Machine Learning and this project was proven to be appropriate for us as we didn't lack motivation till the end for our project and after this, we also aim to further learn and work more on various other ML projects.

As Machine Learning is also a Python project, we learnt how we can work and implement Python to get our desired Project done.

We learnt about various Python and ML libraries like numpy, librosa, soundfile and sklearn which helped us a lot while working on our model and are expected to be useful for future projects also.

Challenges faced

- The challenge that we faced was about finding a proper dataset. Earlier we had decided that we will use the RAVDESS dataset but due to its size we need to change it.
- So we wanted to change our dataset and finally after 1.5 days of research, we found our new dataset <u>(CREMA-D)</u>. Most of our time gone in vectorization and embedding of audio data because that was a difficult task for us.
- Early we thought that we will detect emotion from both text and speech for better accuracy but we can't input text and speech at same time.
- Next challenge that we faced was about increasing the accuracy for our model. We tried ensembling but due to coding error that we couldn't debug we had to stop that idea. We also try using feature engineering but that didn't work either.
- Last challenge we faced was related to Python script, we debugged it but it showed the same error when we use python IDE, so for better transparency, we will redirect it to google colab where successfully emotion is getting predicted.

Conclusion

Emotions play an important role in human mental life. The importance of SER has increased in recent days to improve both the naturalness and efficiency of human – machine interactions. And we have tried to create as such one project that can be used to detect human emotions using Machine Learning.

Future plans

If the project will be appreciated by the panel, We can add CNN into our model and add corresponding feature too in our website and take input according to that because both of the parts together will be more favorable in case of emotion recognition.