**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| Team Member : 1. Kartik Pandey  2. Aniket Nichat  3. Rohit Thawali  4. Sagar Khekale  Emali: 1. [kartikpande12@gmail.com](mailto:kartikpande12@gmail.com)  2. [vrushabhnichat@gmail.com](mailto:vrushabhnichat@gmail.com)  3. [rohitthawali25@gmail.com](mailto:rohitthawali25@gmail.com)  4. sagarkhekale2@gmail.com  Contribution :  Kartik Pandey:   1. Worked on EDA 2. Count of Emotion 3. Compare the MFCC feature for male and female angry audio clips 4. Deploy Model CNN   Aniket Nichat:   1. Worked on Data Augmentation 2. Noise added in Audio 3. Stretched Audio 4. Shift Audio   Rohit Thawali:   1. Worked on Feature Extraction 2. Data Preprocessing 3. Collect Dataset From Kaggle 4. Worked on KNN   Sagar Khekale :   1. Worked on Deploy model 2. Worked on MLP Classifier 3. Work on Confusion Matrix |
| **Please paste the GitHub Repo link.** |
| <https://github.com/Vrushabh9975/Speech-Emotion-Recognition> |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** |
| **Summary:**  **Speech Emotion Recognition, abbreviated as SER is the act of attempting to recognize human emotion and affective stages from speech. This is capitalizing on the fact that voice often reflects underlying emotions through tone and pitch. This is also the phenomenon that animals like dogs and horses employ to be able to understand human emotion. SER is tough because emotions are subjective and annotating audio is challenging**  **In this project, we divide this project into four-part First EDA, Data Augmentation, Feature Extraction, Model.**  **First EDA in this Technique we get an insight of data information, using EDA The key feature was**  **We use MFCC(Mel Frequency Cepstral Coefficients), Mel Spectrogram.**  **MFCC –**  **MFCC is taken on the Mel scale which is a scale that relates the perceived frequency of a tone to the actual measured frequency. It scales the frequency To match more closely what the human ear can hear. The envelope of the temporal power spectrum of the speech signal is representative of the vocal tract and MFCC accurately represents this envelope.**  **Mel Spectrogram :**  **A Fast Fourier Transform is computed on overlapping windowed segments of the signal and we get what is called the spectrogram. This is just a spectrogram that depicts amplitude mapped on a Mel scale.**  **Now, The Second Part is Data Augmentation .**  **Data augmentation is the process by which we create new synthetic data samples by adding small perturbations to our initial training set.**  **To generate syntactic data for audio, we can apply noise injection, shifting time, changing pitch, and speed.**  **Now, Feature Extraction**  **In Feature Extraction, we extract features and processing the data. Broadly feature extraction techniques are classified as temporal analysis and spectral analysis techniques. In temporal analysis, the speech waveform itself is used for analysis. In spectral analysis spectral representation of the speech, a signal is used for analysis.**  **Now, Model**  **In Model after we ran the CNN model, we get model3.h5 file and we store it and we used into the testing purpose for web application**  **Problem Statement :  Verbal Communication is valuable and sought after in workplace and classroom environments alike. There is no denying the notion that Indians lack verbal communication and consequently lag behind in the workplace or classroom environments. This happens despite them having strong technical competencies. Clear and comprehensive speech is the vital backbone of strong communication and presentation skill.**  **Approach :**  **My approach towards the Project First I understand the data, data is in a good shape or not. And my Next Aim is to convert the audio waveform into a spectrogram. I used Three way to Augment data Time Warping, Frequency Masking, Time Masking. When I used Time Warping did not improve model performance a lot, if a resource is limited, this approach is discarded. To generate syntactic data for audio, I have applied noise injection, shifted time, changed pitch and speed.**  **Conclusion:**  **Through this project, we showed how we can leverage Machine learning to obtain the underlying emotion from speech audio data and some insights on the human expression of emotion through voice. This system can be employed in a variety of setups like Call Centre for complaints or marketing, in voice-based virtual assistants or chatbots, in linguistic research, etc.** **In this project, we use some beautiful techniques like MLP classifier, CNN, KNN every model gave us good accuracy but we take only one model that names model3.h5 file. After that we also use Data augmentation is a technique used to increase the amount of data by adding slightly modified copies of already existing data or newly created synthetic data from existing data. It acts as a regularizer and helps reduce overfitting when training a machine learning model. So this project helps to predict emotion using speech.**  **Drive Link:**  **https://drive.google.com/drive/folders/16By\_ijvHFI14PU3IP7Db1sc4WwXHndQm?usp=sharing**  . |
|  |