**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  Email: 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. Worked on GRU 5. Deploy Model CNN   Aniket Nichat:   1. Worked on Data Augmentation 2. Noise added in Audio 3. Stretched Audio 4. Shift Audio 5. Worked on LSTM   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 4. Worked on Decision Tree |
| **Please paste the GitHub Repo link.** |
| https://github.com/kartik567/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:**  **What is Speech Emotion Recognition?**  **In Today’s era the speech has more important because using speech we detect what the feeling of person in various communication field the speech has more important, So What is speech, Speech is nothing but,humans express the feeling and thoughts using coherently sounds, speech emotion recognition easily understand the emotion, Speech is the foremost common way of communicating ourselves as people. It is as it were common at that point to expand this communication medium to computer applications. We characterize speech feeling acknowledgment (SER) frameworks as a collection of techniques that prepare and classify speech signals to identify the inserted feelings.**  **In this project, we divide this project into four-part First EDA, Data Augmentation, Feature Extraction, Model.**  **To begin with EDA amid this Procedure we get an understanding of information data, utilizing EDA**  **The Key Feature we use MFCC and Mel Spectrogram.**  **MFCC –**  **MFCC is taken on the Mel scale which may be a scale that relates the seen recurrence of a tone to the specific measured recurrence. It scales the recurrence to coordinate more closely what the human ear can listen. The envelope of the worldly control range of the speech flag is agent of the vocal tract and precisely speaks to the MFCC envelope.**  **Mel Spectrogram :**  **A Fast Fourier Transform is computed on overlapping windowed segments of the signal and that we get what's called the spectrogramA spectrogram may be a visual way of speaking to the flag quality, or “loudness”, of a flag over time at different frequencies display in a specific waveform.**  **Now, The Second Part is Data Augmentation:**  **Data augmentation is that the method by which we make unused engineered information tests by including little annoyances to our introductory preparing set. To create syntactic data for sound, we are going apply commotion infusion, moving time, changing pitch, and speed.**  **Now, Feature Extraction**  **In Feature Extraction, we extract features and processing the info. Broadly highlight extraction procedures are classified as worldly examination and unearthly examination techniques. In temporal analysis, the speech waveform itself is employed for analysis. In spectral analysis spectral representation of the speech, a sign is employed for analysis.**  **Now, Model**  **In Model, we use six models but we take only one model for deployment and take this model for web application**  **Problem Statement :**  **Verbal Communication is effective and wanted in workplace and classroom environments alike. there's no denying the notion that Indians lack verbal communication and consequently lag behind within the workplace or classroom environments. This happens despite them having strong technical competencies. Clear and comprehensive speech is that the vital backbone of strong communication and presentation skills.**  **Approach:**  **My approach towards the Project First thing first use EDA it gets intuition about dataset after intuition use MFCC and Mel spectrogram to extract the feature from speech dataset and after that when if working on sequence dataset like speech GRU comes in mind, so create model of a GRU then moving ahead next use CNN model This model is also use in deployment.**  **Conclusion:**  **With this explanation, we found out a spread of notary or marketing call centers, voice based virtual assistants or chatbots, linguistic translators to realize some intuition of some human expressions of emotions and voluntary emotions inherent within the data we speak. The model gives us better accuracy, but we only take one model which is known as model3. h5 file. We then use data augmentation which is employed to extend the quantity of knowledge by adding slightly modified copies of existing data or newly created synthetic data from existing data. It acts as a regulation and helps to scale back excess fitting while training the machine learning model. So this project helps to predict emotions using speech.**  . |
| **Drive Link:**  **https://drive.google.com/drive/folders/1pbOV1qCzoIcCpWwEvnMm5F4uK0FPm\_tx** |