

# PROJECT REPORT

## On

**SPEECH EMOTION RECOGNITION**

## COMPUTER SCIENCE AND ENGINEERING

**B.E. Batch-2018 in**

**December-2020**



|  |  |
| --- | --- |
| **Under the Guidance of:** | **Submitted By:** |
| Dr. Ashutosh Kumar Dubey | **Parth Ahuja** |
|  | **Id. No. 1811981220** |
|  | **Deepak** |
|  | **Id. No. 1811981097** |
|  | **Atishaya Jain** |
|  | **Id. No. 1811981080** |
|  |  |
|  |  |

**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

**CHITKARA UNIVERSITY**

**HIMACHAL PRADESH**



**CERTIFICATE**

This is to be certified that the project entitled “Speech Emotion Recognition” has been submitted for the Bachelor of Computer Science Engineering at Chitkara University, Himachal Pradesh during the academic semester July 2020-December 2020 is a bonafide piece of project work carried out by “Parth Ahuja(1811981220), Deepak(1811981097), Atishaya Jain(1811981080)” towards the partial fulfillment for the award of the course Professional Practices-Python for Data Science (CS254E) under the guidance of “Dr. Ashutosh Kumar Dubey” and supervision.

### **Signature of Project Guide**:

Dr. Ashutosh Kumar Dubey (Professor, Department of Computer Science and Engineering)



**CANDIDATE’S DECLARATION**

We, **Parth Ahuja(1811981220), Deepak(1811981097), Atishaya Jain(1811981080),** B.E.-2018 of the Chitkara University, Himachal Pradesh hereby declare that the “Name of the Course” Report entitled **“SPEECH EMOTION RECOGNITION”** is an original work and data provided in the study is authentic to the best of our knowledge. This report has not been submitted to any other Institute for the award of any other course.

|  |  |  |
| --- | --- | --- |
| **Sign. of Student 1** | **Sign. of Student 2** | **Sign. of Student 3** |
| Parth Ahuja | Deepak | Atishaya Jain |
| ID No 1811981220 | ID No 1811981097 | ID No 1811981080 |

### **Place:** **Date:**



**ACKNOWLEDGEMENT**

It is our pleasure to be indebted to various people, who directly or indirectly contributed in the development of this work and who influenced our thinking, behavior and acts during the course of study.

We express our sincere gratitude to all for providing an opportunity to undergo “Python for Data Science”- Project as the part of the curriculum.

We are thankful to “Ms. Ravita Chahar” for his/her support, cooperation, and motivation provided to us during the training for constant inspiration, presence and blessings.

We also extend our sincere appreciation to ***“Dr. Ashutosh Kumar Dubey*** who provided his/her valuable suggestions and precious time in accomplishing our project report.

Lastly, we would like to thank the almighty and our parents for their moral support and friends with whom we shared our day-to day experience and received lots of suggestions that improve our quality of work.

|  |  |  |
| --- | --- | --- |
| **Parth Ahuja** | **Deepak** | **Atishaya Jain** |
| **ID No 1811981220** | **ID No 1811981097** | **ID No 1811981080** |



**ABSTRACT**

Speech emotion recognition has been a burning issue since last decade. Researchers have been trying to develop a system more like human, emotion recognizing robots is an example of it. Speech has many parameters which have great weightage in recognizing emotion namely Prosodic and spectral features. Out of prosodic features namely pitch, energy and intensity are popularly used and out of spectral features format Mel frequency cepstral coefficients are commonly used by the researchers worldwide.Further the classifiers are trained by using these features for classifying emotions accurately. This paper is an attempt to give a short review about the work on Emotion recognition from speech.



**INTRODUCTION**

Speech emotion detection refers to analyzing vocal behavior as a marker of affect, with focus on the nonverbal aspects of speech. Its basic assumption is that there is a set of objectively measurable parameters in voice that reflect the affective state a person is currently expressing. Speech is one of the basic and natural way of communication among human beings. Emotion makes speech more expressive and effective. Different ways like laughing, yelling, teasing, crying, etc. are used by humans to express their emotions. Emotion detection can be an easy task for humans but a difficult one for machines. So, there is a need of such emotion recognition systems that can make human computer interaction quite easy. The speech signal contains other frequency related characteristics that are spectral features. Mel Frequency Cepstral Coefficients (MFCCs) are generally used in speech recognition with great accuracy in emotion detection. The fields of psychology and psycholinguistics provided interesting results about how prosodic cues, fundamental frequencies and the intensity of the voice can show variability levels across different speakers. Speech emotion recognition thus can be defined as the extraction of the emotional state of the speaker from his or her speech signal to make human machine interface more convenient. The core part of this project exploits a balanced mixture of theory and implementation to create small, compact, highly optimized pairwise emotion classifiers. The machine learning framework shows several classifiers used in several tasks related to emotion recognition. Each classifier has advantages and disadvantages in order to deal with the speech emotion recognition problem.



**Background**

As human beings speech is amongst the most natural way to express ourselves. We depend so much on it that we recognize its importance when resorting to other communication forms like emails and text messages where we often use emojis to express the emotions associated with the messages. As emotions play a vital role in communication, the detection and analysis of the same is of vital importance in today’s digital world of remote communication. Emotion detection is a challenging task, because emotions are subjective. Choosing to follow the lexical features would require a transcript of the speech which would further require an additional step of text extraction from speech if one wants to predict emotions from real-time audio. There is no common consensus on how to measure or categorize them. We define a SER system as a collection of methodologies that process and classify speech signals to detect emotions embedded in them. Such a system can find use in a wide variety of application areas like interactive voice based-assistant or caller-agent conversation analysis. In this project we attempt to detect underlying emotions in recorded speech by analyzing the acoustic features of the audio data of recordings.

**Problem Statement**

To Develop a Machine Learning Project that can Recognize human emotions through speech.



**SOFTWARE AND HARDWARE REQUIREMENT SPECIFICATION**

For Speech Emotion Recognition project we will be using the Anaconda compiler for the compilation of the code and we will use many libraries of python. A set of libraires are used, for machine learning, digital signal processing and GUI (Graphical User Interface) development. A Dataset is required for the input in the form of different type of voice to recognize the emotion with the help of speech.

**Methods**

1. Dataset: The Emotional characteristics are extracted from the Ryerson Audio-visual Database. The database is based on six universal emotions like anger, disgust, fear, joy, sadness and surprise. For this project the Multimedia File Reader object from the DSP System Toolbox Library of MATLAB was used to read the group

of audio frames from each multimedia file.

1. Features: The data are acquired directly from the group of waveform Audio files and they were transformed in 264 vectors of features. A wide range of possibilities exist for parametrically representing a speech signal and its content in a vector, with the

intention of the extraction of relevant information from it. A variety of choices for this



task can be applied to represent the speaker’s speech in a large number of parameters,

in which the changes in these parameters will result in corresponding change in emotions.

1. Machine Learning Techniques: The binary classification algorithm Support Vector Machine (SVM) which originated in statistical learning theory, offers robust classification to a very large number of variables and small samples. SVM is capable of learning complex data from classification models applying mathematical principles to avoid overfitting. The more used kernels in SVM are polynomial and linear.

**Programming/Working Environment**

The Program is implemented in python. This is done to allow interfacing with the openSMILE

feature extraction library and the highest performing Machine learning libraries. Python provides us many libraries which are helpful for using the dataset in our project. Librosa library will be used for analyzing audio. For using dataset in a proper way, Multi-Layer Perceptron (MLP) classifier is used for the importing the dataset. Confusion matrix is used for evaluating the performance of the dataset. We used an MLPClassifier for this and made use of the sound file library to read the sound file, and the librosa library to extract features from it.

**Requirements to Run the Application**

* Dataset for the input in the project.
* Librosa library for analyzing audio.

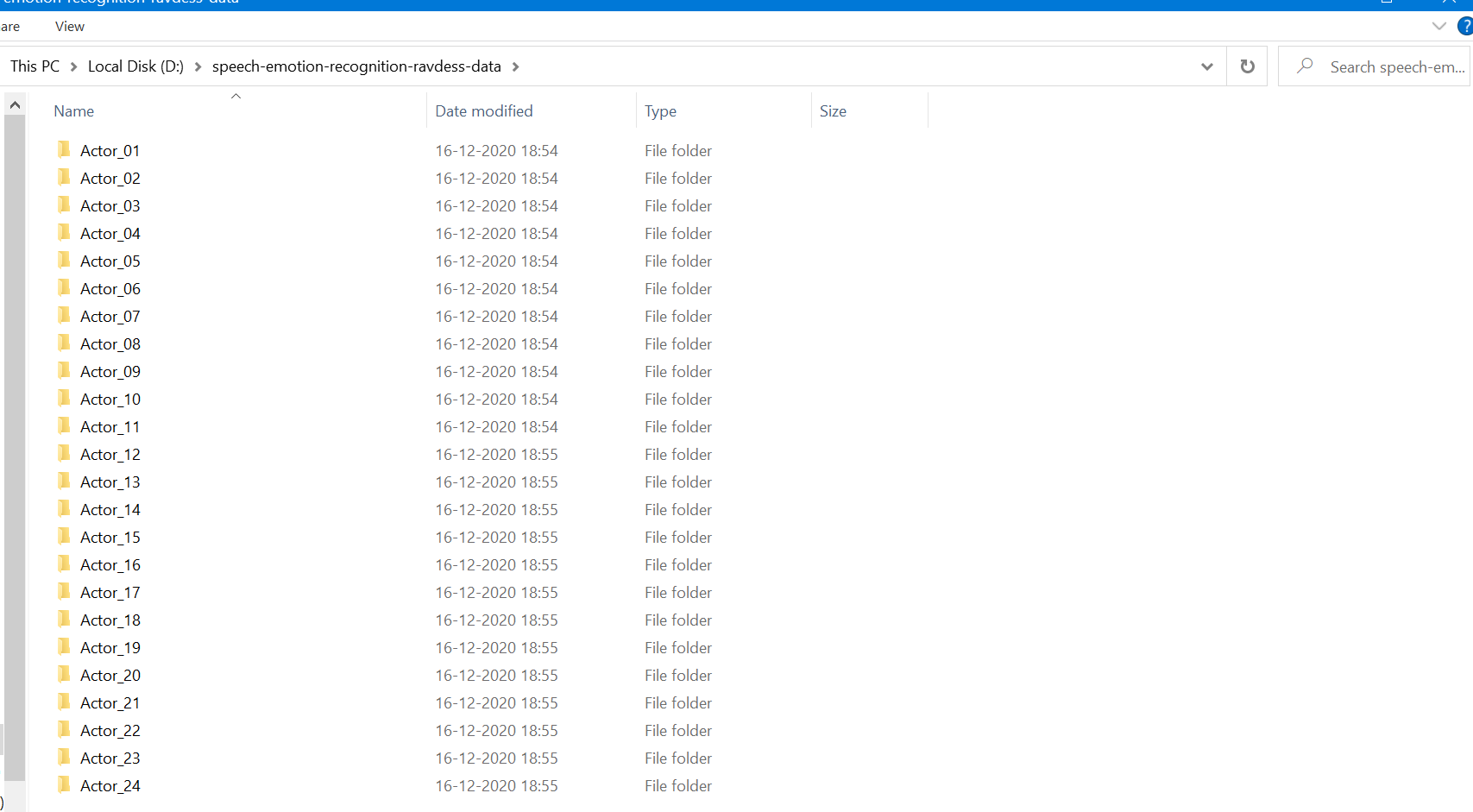


* MATLAB for signal processing toolbox.
* Online compiler for running the program.
* Corpora for emotion detection and speech quality analysis.

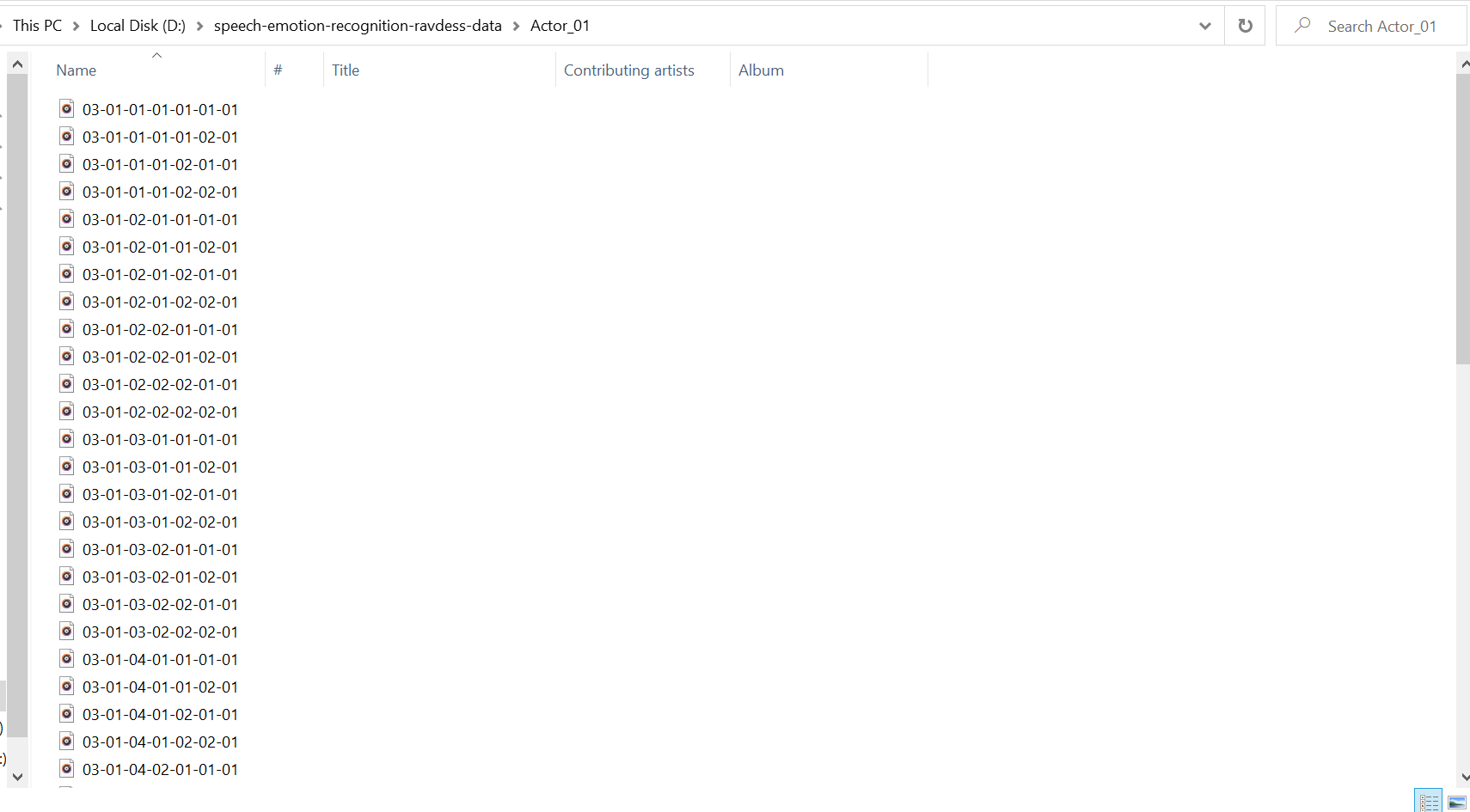


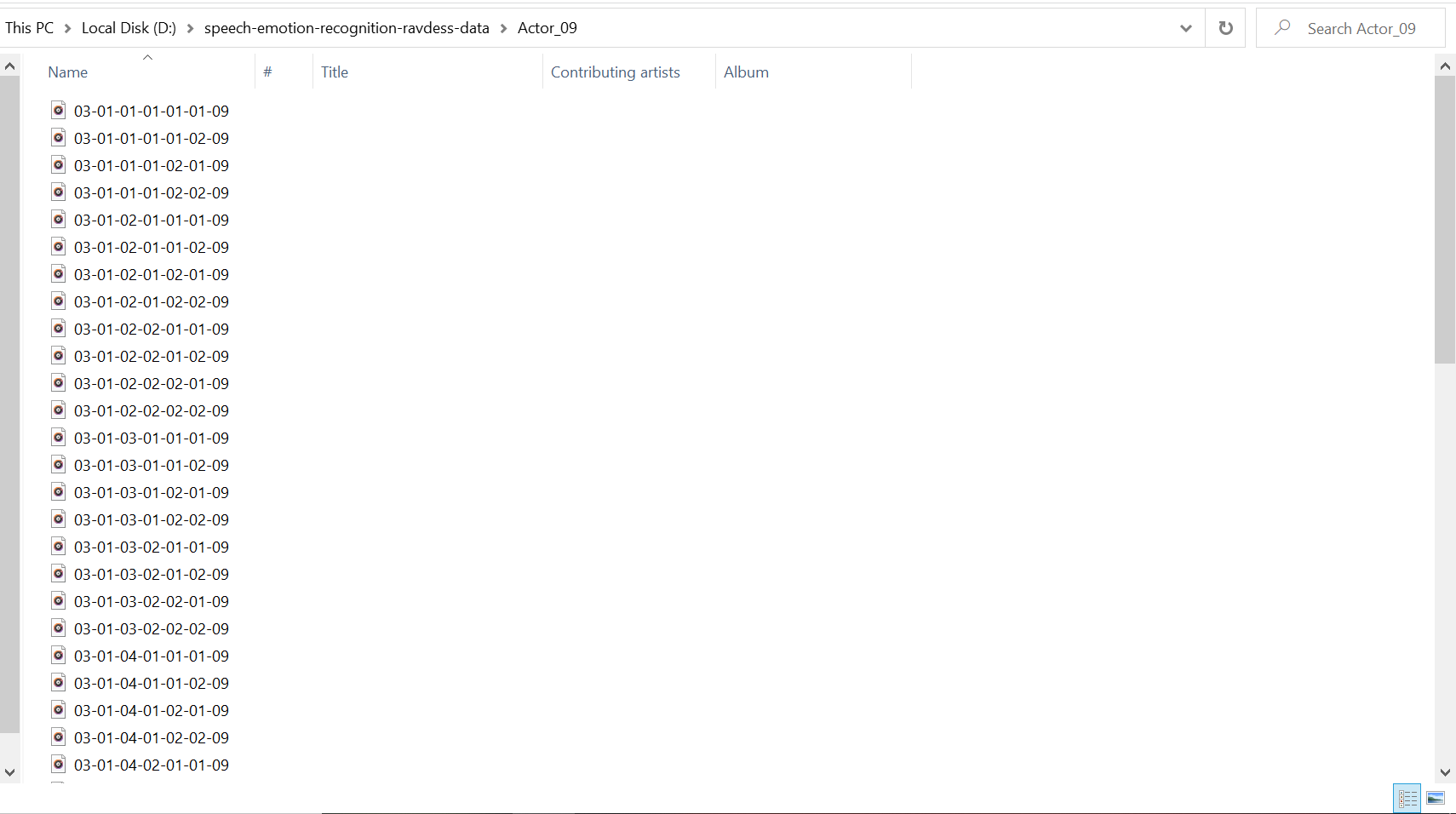
**PROGRAM’S STRUCTURE AND ANALYZING AND GUI CONSTRUCTING**

Snapshots of Dataset (Voice of Different speakers):



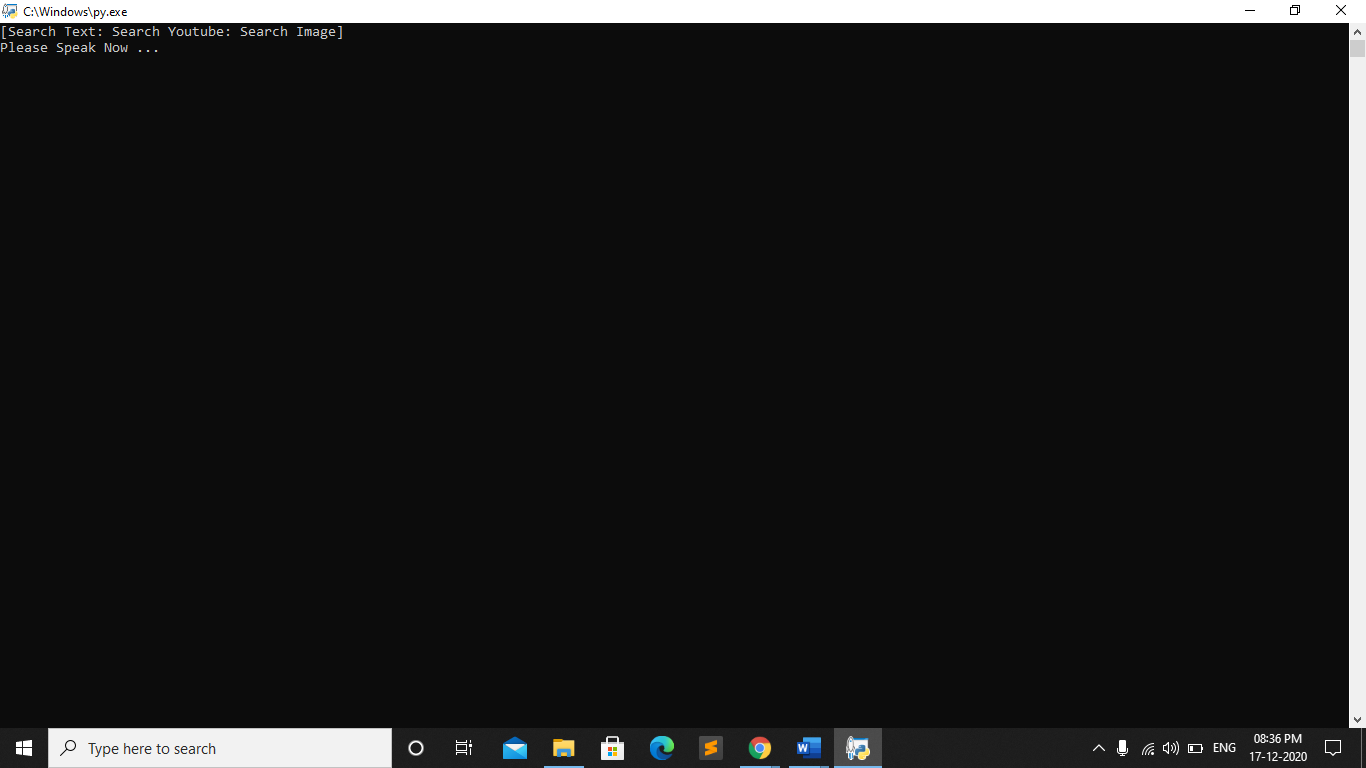




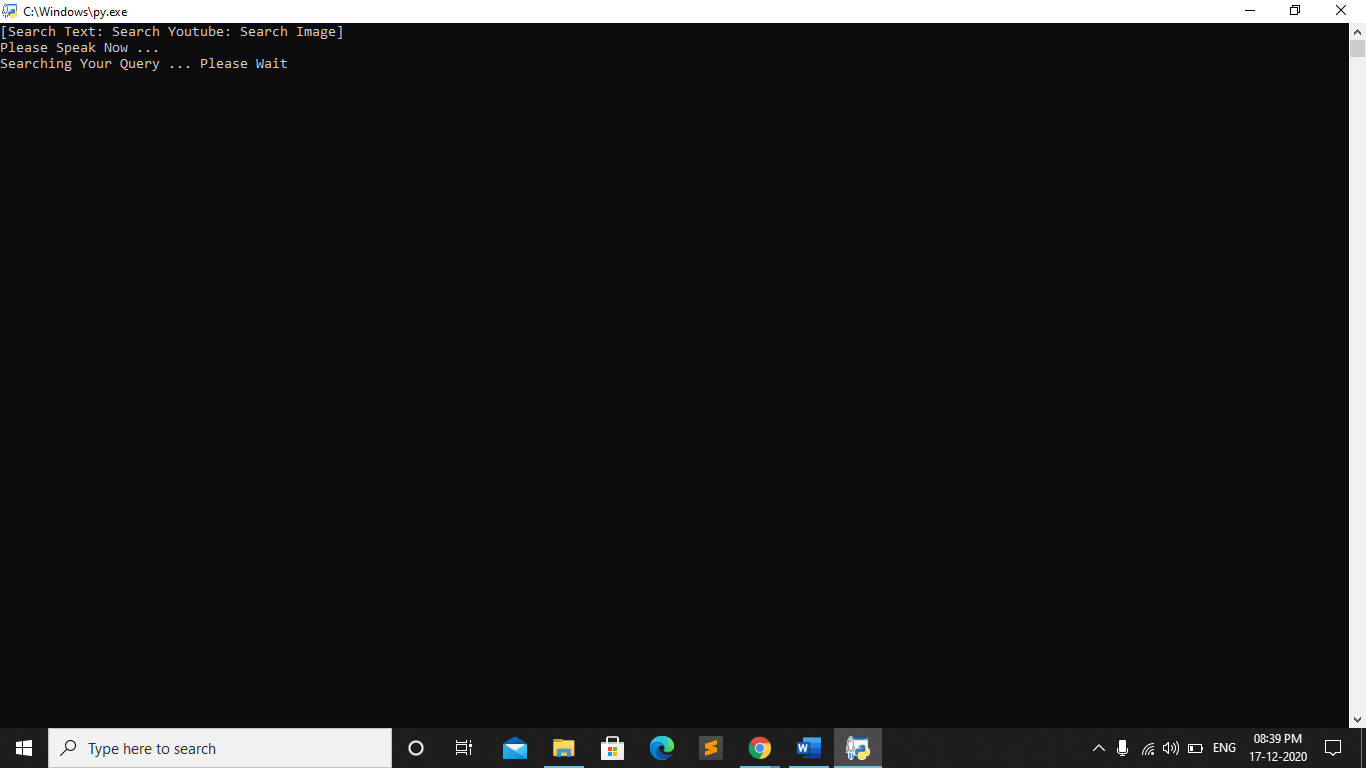




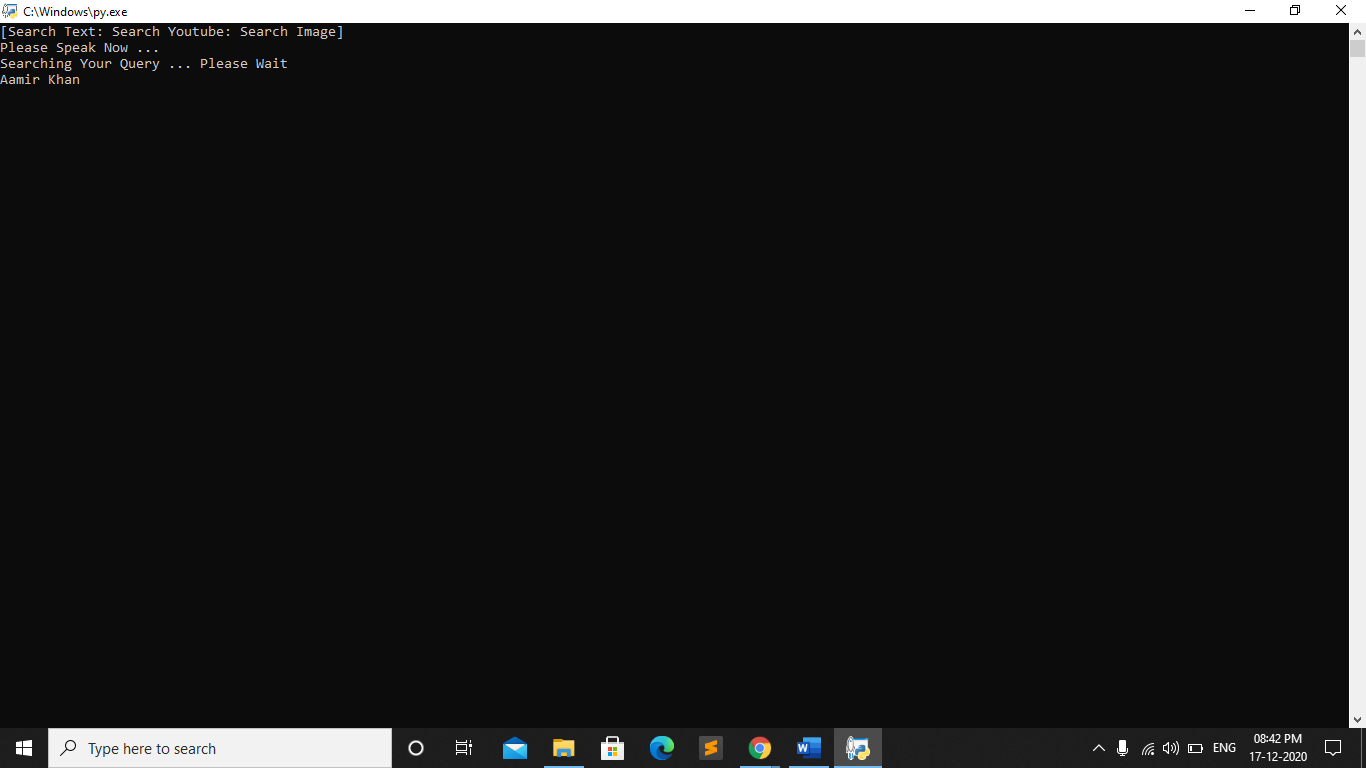
Output Of Our Project (Snapshot)--













**LIMITATIONS**

* It does not work on different Languages other than English.
* It is not like Google which can tell every single information about a person.



**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. A collection of audio data from several videos related to human emotional expressions were gathered and turned into a data set. As a real application, it could be considered a real-time system that can serve like a motor of emotional knowledge in order to understand the autistic children, to describe accurately their internal state and show the real content of their emotions. This system can be employed in a variety of setups like Call centers for complaints or marketing, in voice based virtual assistants or chatbots, in linguistic research, etc.



**FUTURE SCOPE**

* An accurate implementation of the pace of the speaking can be explored to check if it can resolve some of the deficiencies of the model.
* Exploring other acoustic features of sound data to check their applicability in the domain of speech emotion recognition.
* Following lexical features-based approach towards SER and using an ensemble of the lexical and acoustic models. This will improve the accuracy of the system because in some cases the expression of emotion is contextual rather than vocal.
* Adding more data volume either by other augmentation techniques like time-shifting or speeding up/slowing down the audio or simply finding more annotated audio clips.



**BIBLIOGRAPHY/REFERENCES**

* Librosa: - <https://librosa.org/librosa/tutorial.html>
* RAVDESS: - <https://zenodo.org/record/1188976#.XvbvZudS_IU>
* Kernel Reference: -https://data-flair.training/blogs/python-mini-project-speech-emotion-recognition/
* TESS: - <https://tspace.library.utoronto.ca/handle/1807/24487>
* Sets and Booleans: - <https://data-flair.training/blogs/python-set-and-booleans-with-examples/>
* MLPClassifier: - <https://www.analyticsinsight.net/speech-emotion-recognition-ser-through-machine-learning/>
* Mel Frequency Cepstral Coefficients (MFCC): - <http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/>
* Chroma: - <https://www.audiocontentanalysis.org/code/audio-features/pitch-chroma-2/>
* Mel Spectrogram: - <https://www.mathworks.com/help/audio/ref/melspectrogram.html>
* Pipeline: - <http://scikit-learn.org/stable/modules/generated/sklearn.pipeline.Pipeline.html>
* Sklearn: - <https://scikit-learn.org/>
* Other Libraries: - <https://docs.python.org/3/library/>



* Literature Reference: - <https://doi.org/10.1186/1687-4722-2012-16>
* Kernel Reference: - https://github.com/marcogdepinto/emotion-classification-from-audio-files?fbclid=IwAR2T4hhtWWfKdU4FwLS8LOAnF5sBwnmfc6PQHTGidzLaLl1uUVOvicx7TVw