Voice Emotion Detection:

Introduction:

Voice Tone Emotion Detection is the process of detecting and predicting the human emotions from audio signals using deep learning techniques. In this project, a voice emotion detection system has been designed which will detect the emotion in a voice signal by analysing its tone.

Background:

This project has been implemented using Python Programming Language. And the various libraries used are Librosa (for audio analysis), Tensorflow (for backend), Numpy (to perform numeric operations) and Keras (to build the classification model). The input voice tone is played and our model detects the emotion of the voice by analyzing its tone.

Learning Objectives:

The main objective of this Voice Emotion Detection system is to produce an automatic model which can detect different emotions in people based on their voice tone and can identify the people who are sad, happy, neutral, surprised, disgusted, angry, fear etc. This has many applications like lie detection, in emergency call centers, security issues etc.

Activities and Tasks:

The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) dataset is used to train the model. The input (a voice clip file) is uploaded and then it is passed through the emotion detection model and then the model detects the respective emotion of the voice clip and predicts the emotion accordingly.

Skills and Competencies:

Installation of Python 3.8 version, install all the necessary libraries in it like numpy, matplotlib, tensorflow, jupyter notebook, streamlit, librosa etc. Installation of Anaconda prompt for checking the output. Python coding is done in Jupyter notebook and in Visual Studio Code. Convolutional Neural Network layers has to be defined. Programs are run on Kaggle also for a better optimized output.

Feedback and Evidence:

There is a lot of things that we can learn from this project. It is a complete use of Deep Learning of CNN in Python. References are taken from:

1. Text Book “Python Programming by Reema Thareja”.
2. Few research papers.
3. Few previous works on which people have already worked.

Challenges and Solutions:

The main challenge of this emotion detection system is that sometimes it cannot detect the voice or sometimes it cannot give very accurate results. And the solution to this challenge is that we have to train the model with more number of epochs to get better and accurate results or we should train the model with more number of voice clips.

Outcomes and Impact:

Our neural network which was trained with voice emotion expression dataset is able to predict the emotion of each voice tone in the audio file and categorize them into happy, sad, anger, fear, surprised, disgusted or neutral. We have used streamlit to display our outputs.

Conclusion:

A classification method is implemented in which the voice of different tones are used to train a classifier predictor which predicts the seven basic human emotions. Our predictor model successfully predicts the human emotions based on their voice tone and classify them as sad, angry, happy, neutral, disgust, surprise or fear.