Kaggle Challenge - 2 (30 marks)

Unsupervised Learning

Instructions.

- This is a kaggle competition so you have to follow these instructions carefully.
- Create a single notebook in the competition. Multiple notebooks of same submission will lead to **0 marks**
- You are free to use any available classical machine learning techniques and methods for data pre-processing.
- You have to submit code and your code will be checked for plagiarism, IIITD plagiarism policy will be followed strictly.

Dataset Description

65,000 one-second long utterances of 30 short words, by thousands of different people. **Download Link**

Task Description

You have to make a Gaussian Mixture Model (GMM) classifier for the given dataset which can classify between 30 words, you are allowed to use any available source, resource to you for the pre-processing, algorithm implementation, testing. Make sure you do not use any kind of deep learning which reward you 0 marks.

- Only *test_train_split* from SkLearn must be used to split your datasets, or else your submission will be invalid.
- Your have to submit your code and a python script which takes input as a text file containing path to testing audios (will be provided at time of demos.) and predicts the class of the audio file. Test audios are given in *SpeechCommandTest* folder.
- **Evaluation criteria**: Your final marks will be a combination of 60%(Accuracy at time of demo) + 40%(Leader-board position).
  Your marks are divided in two parts 60% of 30 i.e. (18) and 40% of 30 i.e. (12). Say you got 45% accuracy at the time of demo and your leader-board position is 20, your marks will be (45% of 18) + (leader-board ranking based score out of 12)
- This is a simple task to perform, you will be awarded marks only if you get accuracy above **65%**
- Class mapping = ['right': 0, 'eight': 1, 'cat': 2, 'tree': 3, 'bed': 4, 'happy': 5, 'go': 6, 'dog': 7, 'no': 8, 'wow': 9, 'nine': 10, 'left': 11, 'stop': 12, 'three': 13, 'sheila': 14, 'one': 15, 'bird': 16, 'zero': 17, 'seven': 18, 'up': 19, 'marvin': 20, 'two': 21, 'house': 22, 'down': 23, 'six': 24, 'yes': 25, 'on': 26, 'five': 27, 'off': 28, 'four': 29]