# REPORT

**Task GitHub link:**
https://github.com/MelvinInfant/CognitiveStressClassification/tree/dd7e3857224c30f9832d7fbc218ac4d4b7dc5c7e

Name: Melvin Infant A

University Reg no: 2348533

College: Christ University, Bengaluru

## TASK

The assigned task is to process voice clips to extract patterns that might indicate cognitive impairment or in other words cognitive load, which would affect a person's ability to think, remember and make decision. It is not physical or emotional stress but a cognitive stress.

## DATA

Since the task is to process voice, appropriate data(audio) regarding cognitive impairment was difficult to find and there was no appropriate data available. Hence, I have created data by recording the audio in which I would speak some predefined sentences for the 'no stress' samples and I would randomly speak something for 'stress' samples. As I am speaking randomly for the 'stress' samples automatically cognitive stress is induced on me, using these techniques I created a dataset with 5 samples with names s1, s2…s5.

In which,

s1 and s2 are samples with 'no stress' and,

s3, s4 and s5 are samples with 'stress'.

## FEATURE EXTRACTION

- Pause per sentence
- Speech rate
- Pitch variability
- Jitter
- Shimmer
- Inter word silence

- Sentence Completion
- MFCC
- Hesitation markers

1. **Pause per sentence**
   Measures the number of silent segments ("pauses") in each spoken sentence.
   Usually, one would have more pauses under cognitive stress or hesitation.
2. **Speech rate**
   Measures words spoken per second.
   Cognitive stress can slow down articulation, reducing speech rate.

3. **Pitch variability**
   Measures variability (standard deviation) of the speaker's fundamental frequency (F0).
   Stress often causes unstable or more erratic pitch patterns.
4. **Jitter**
   Cycle-to-cycle variation in the voice's pitch period.
   Increased jitter reflects micro-instability in vocal fold vibration, common under stress.
5. **Shimmer**
   Cycle-to-cycle variation in amplitude.
   Higher shimmer indicates unstable loudness control, often tied to vocal stress.
6. **Inter word silence**
   Total duration (in seconds) of brief silent periods between words.
   Longer or more frequent inter-word silences signal increased hesitation or processing time.
7. **Sentence completion**
   The count of sentence-ending punctuation marks, indicating the completion of the thought process.
   Under stress, speakers may leave more sentences unfinished or trail off.
8. **MFCC (Mel Frequence Cepstral Coefficients)**
   MFCCs are widely used in speech and speaker recognition. In cognitive stress analysis, changes in vocal tract shape and articulation under stress can alter MFCC patterns, making them useful for detecting stress-induced speech variation.

9. **Hesitation markers**
   Hesitation markers are **non-lexical fillers** and **disfluent events** that naturally occur in spontaneous speech, especially under cognitive load, example: uhh, ahh, hmm etc.
   Hesitation markers are a strong behavioural indicator of cognitive stress, uncertainty, or mental effort. They become more frequent and prolonged when individuals experience overload or decision pressure.

A Dataframe was made by extracting these features from the audio and the label for that audio was also added under the column ‘stress’ with values ‘yes’ and ‘no’ indicating the cognitive stress.

## FEATURE SELECTION

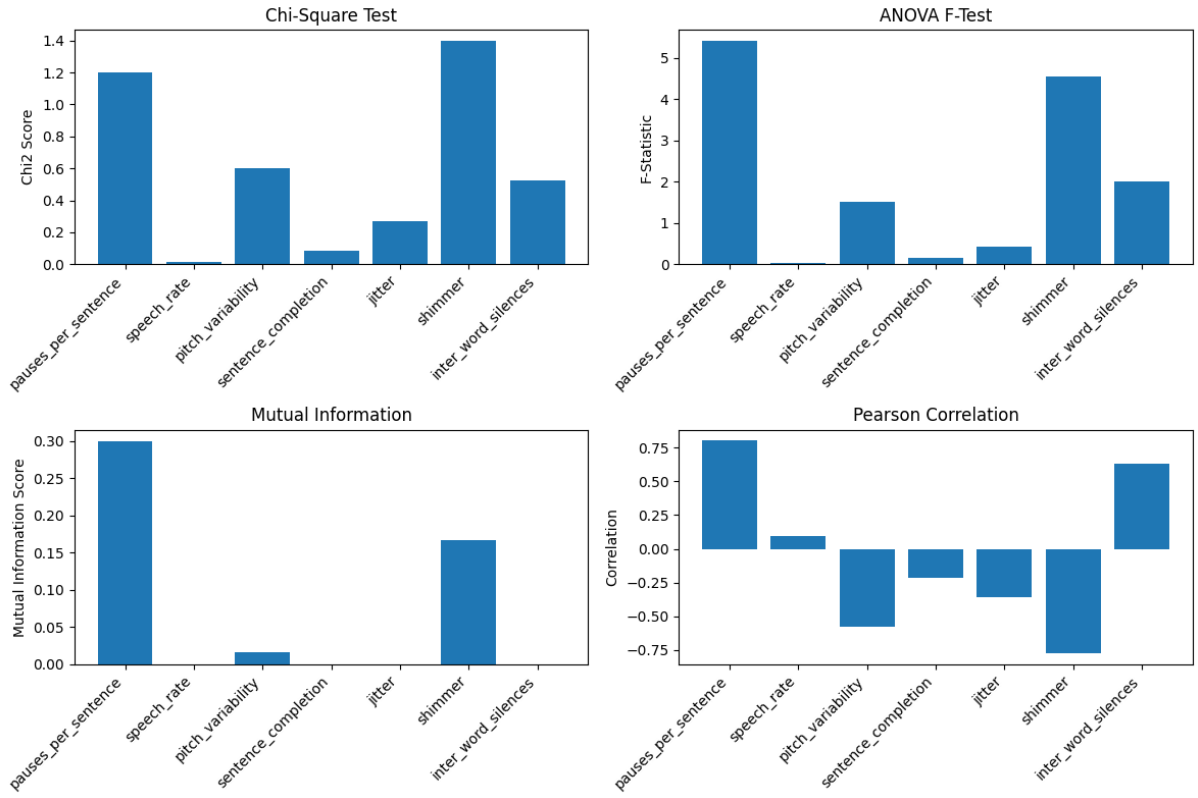
However, in our case we didn’t use MFCC as out task is to extract patterns which might indicate the cognitive stress and,

Capturing hesitation markers were not successful as speech recognition models usually clean up the voice before converting to text. I also tried with speech model (Whisper - Large) by OpenAI for converting the audio to text and it too didn’t work. Even though hesitation markers might be an important factor in determining cognitive stress, in this case it had been not possible, with access to more models and required hardware for processing using the model might result in successful execution of the task.

Methods used to select features are,

- Chi-Square test
- Anova f-test
- Mutual Information
- Pearson Correlation

The results are as follows,



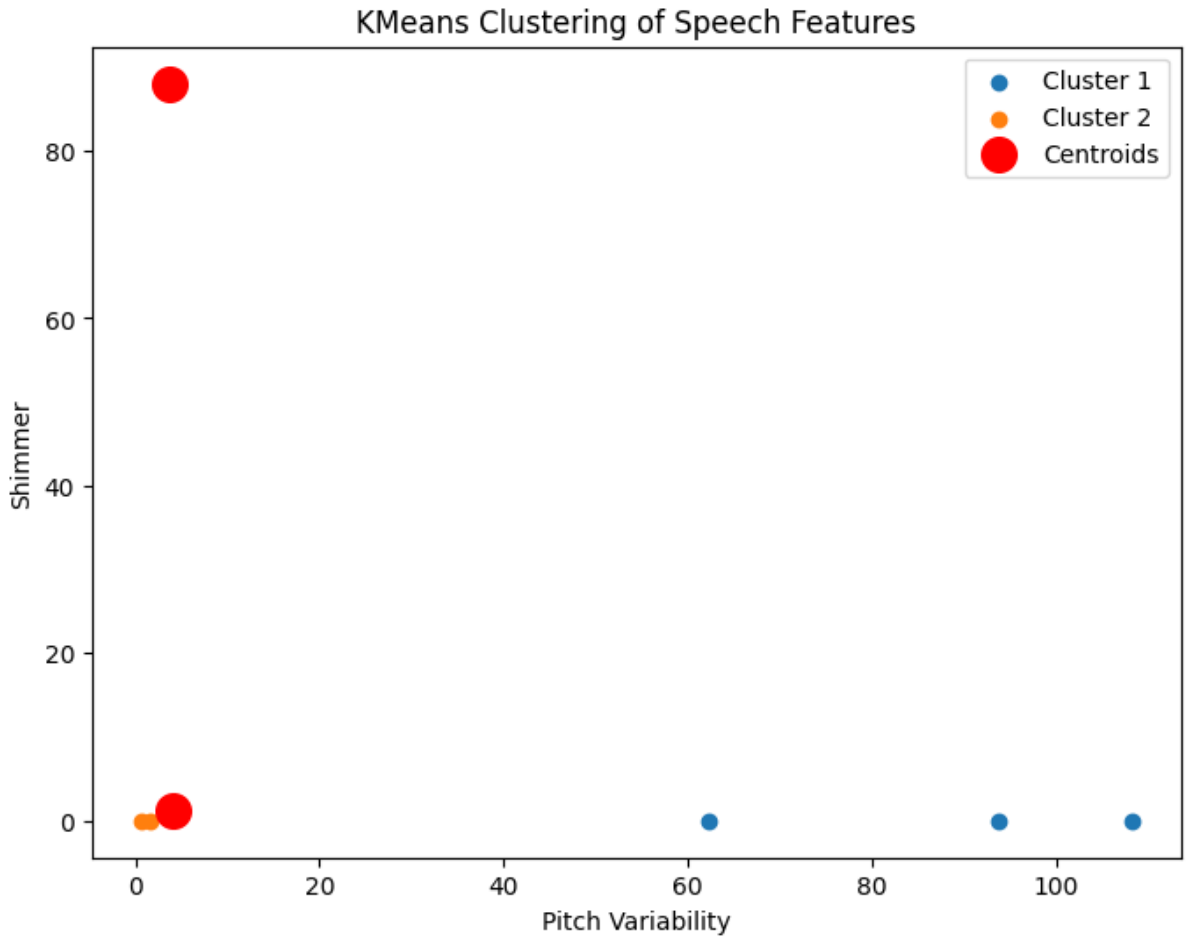

Upon assessing the results,

- ‘pauses_per_sentence’ and ‘shimmer’ appear in the top-2 for all three filter methods (Chi-Square, ANOVA, Mutual Information).
- ‘inter_word_silences’ also shows a strong linear correlation ($|r| > 0.6$), making it a good secondary candidate.
- ‘pitch_variablity’ also showed better result.

Hence, these features are selected for implementation.

## IMPLEMENTATION

Clustering algorithm (K-means) was implemented using the data. And the result is shown below,




```
Sample s1: Cluster 1
Sample s2: Cluster 1
Sample s3: Cluster 1
Sample s4: Cluster 2
Sample s5: Cluster 2
```

Original values,

| | samples | stress |
|---|---|---|
| 0 | s1 | no |
| 1 | s2 | no |
| 2 | s3 | yes |
| 3 | s4 | yes |
| 4 | s5 | yes |

Upon comparing with the original values, we can see 80% of the data (4 points) are classified into the right cluster.

## FURTHER IMPROVEMENTS

The size of the data I used was very small which would yield no generalized pattern upon building a model for classification. In order to improve the performance and to build a robust model,

- Dataset size or sample size should be increased.
- Missed featured should be added (Hesitation markers) and Cross validation methods should be incorporated for getting the best result.
- The data plays the major role, hence the data used should be of good quality.