# WOC7010 AGILE SOFTWARE DEVELOPMENT

## ANIMOJI PROJECT – WEEK 3 - DTW

Sundaram Aishwarya Sundaram
S2029083/1
19 January 2022

# CONTENTS

# SPEECH RECOGNITION

Speech recognition is the process of "identifying sounds based on words spoken by converting an acoustic signal, which is captured by an audio device". It recognizes the words spoken by the human and digitalize it for processing by the computer. This allows a device to recognize and understand the words spoken by digitizing words and matching the digital signals with a certain pattern stored in a device. It has the following basic processes (Washani & Sharma, 2015),

- signal pre-processing
  - noise reduction
  - end point detection
- feature extraction
- speech classification/recognition with models

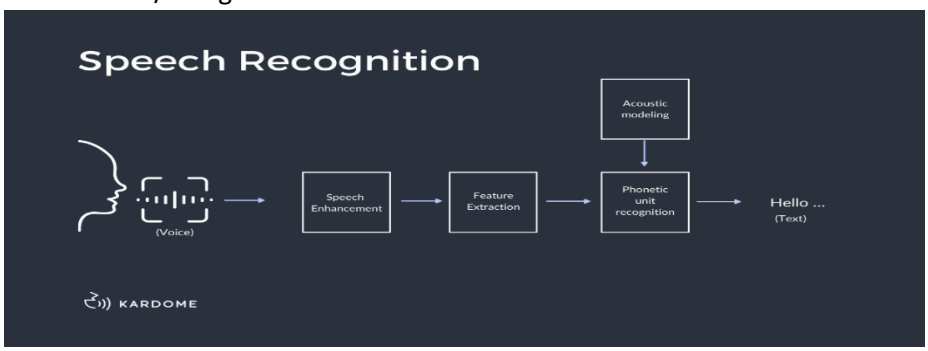
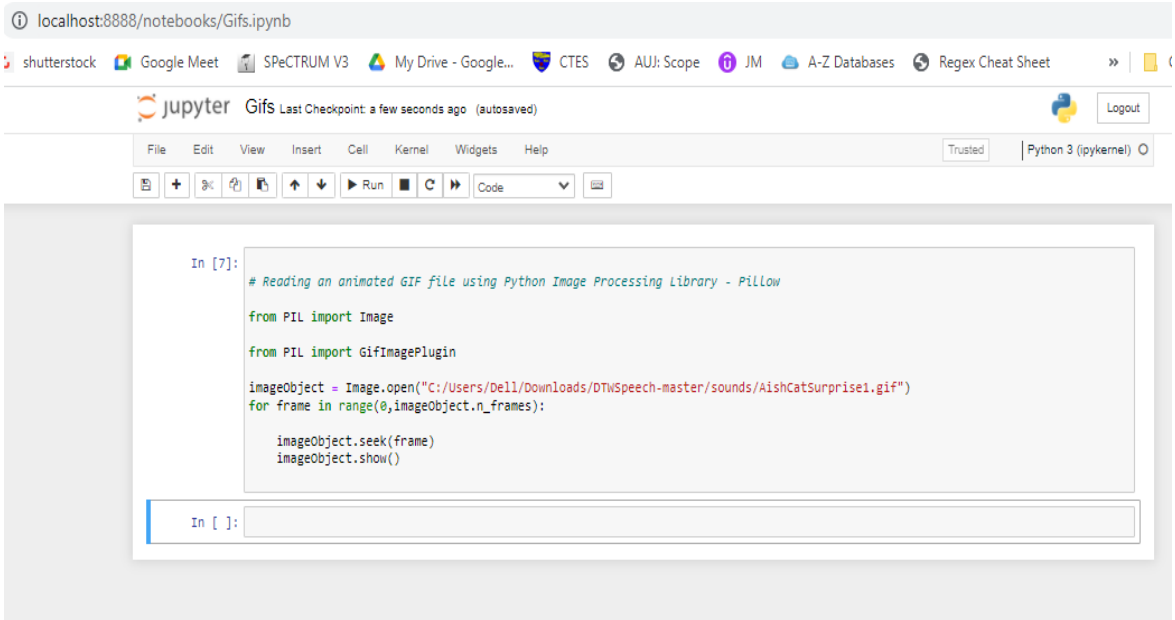


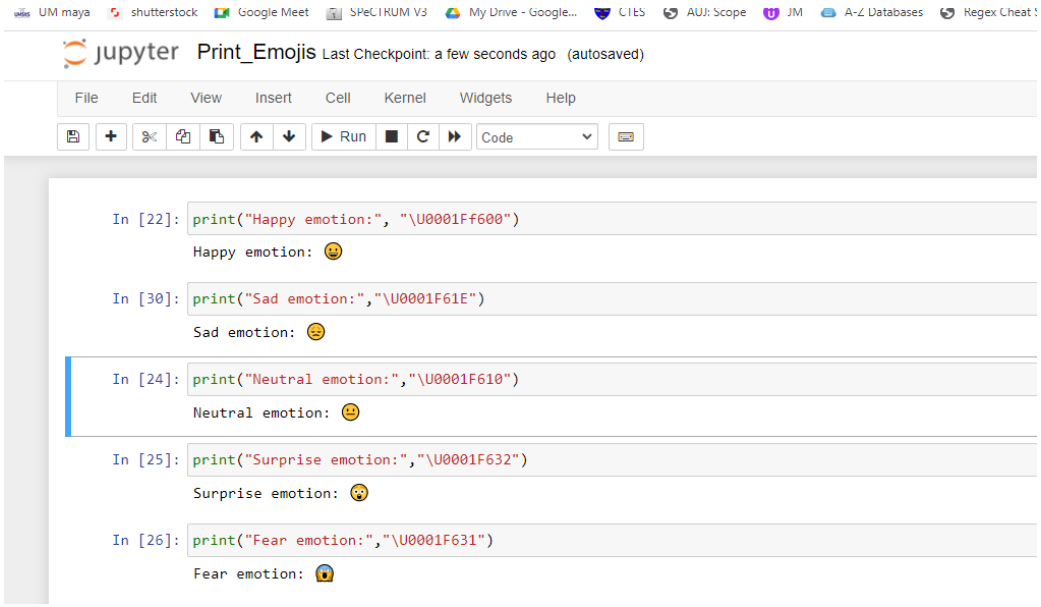

Figure 1. Speech recognition (Source: www.kardome.com)

# PYTHON SIMPLE VOICE TO EMOTION RECOGNITION

## 1. PRINT GIFS

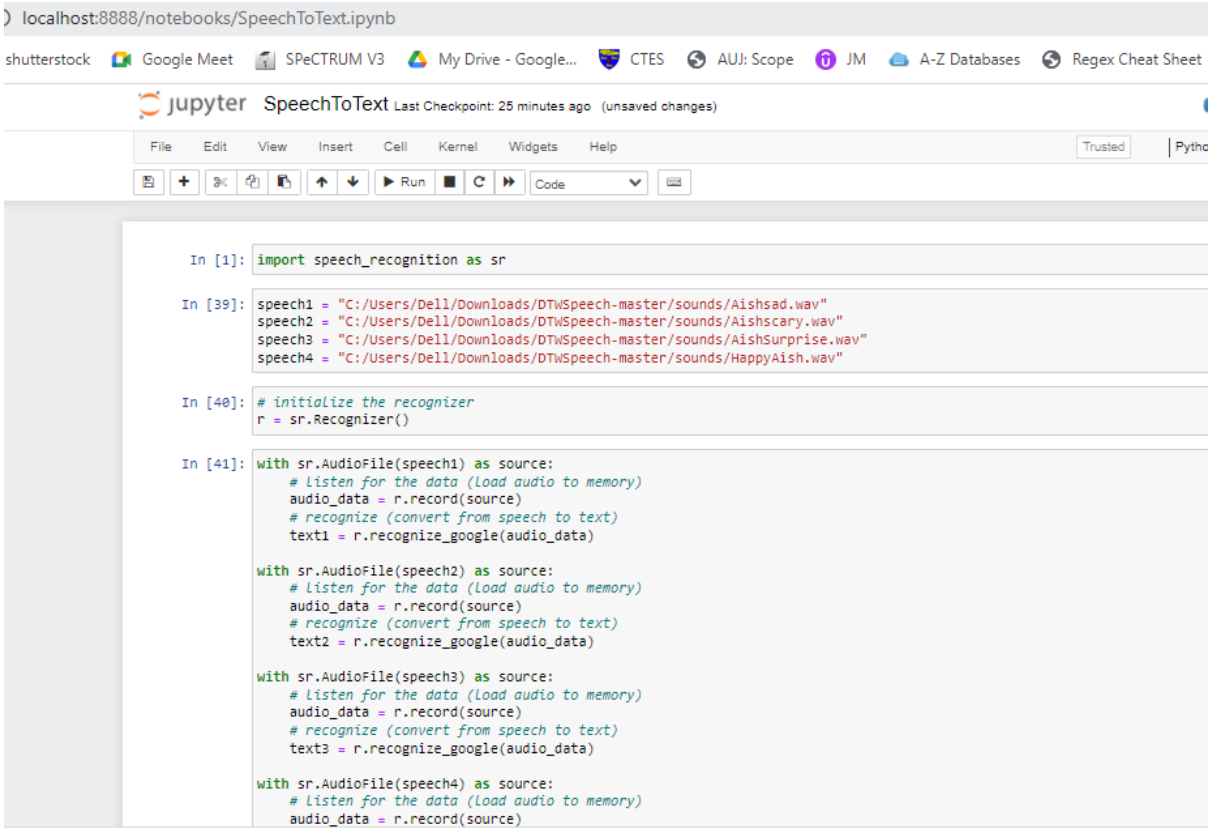


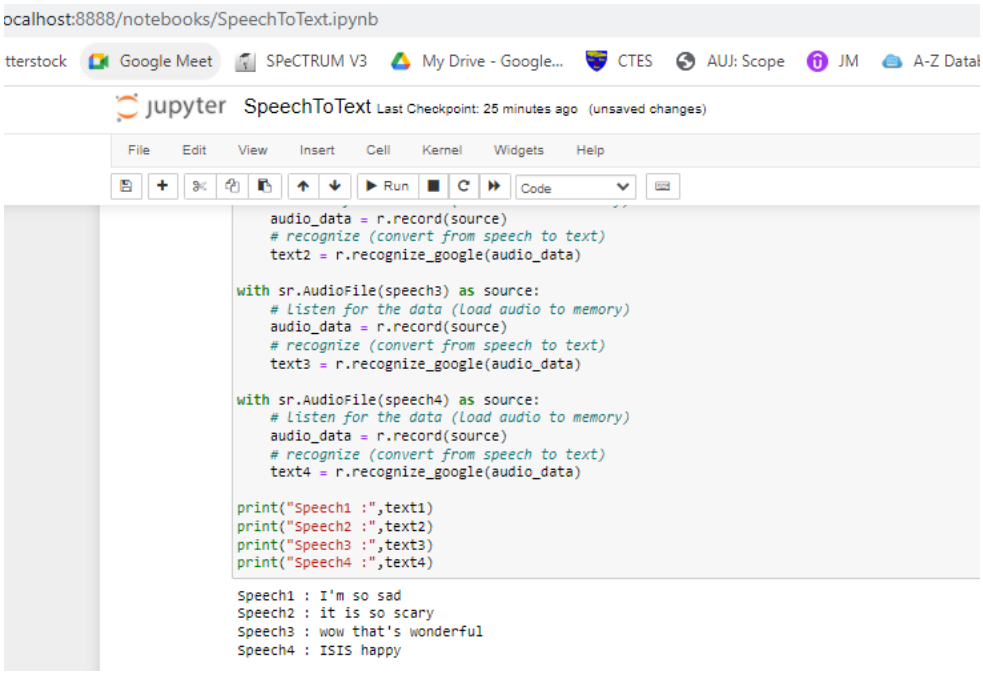
## 2. PRINT EMOJIS




Created file name: Print_Emojis.ipynb

## 3. VOICE TO TEXT

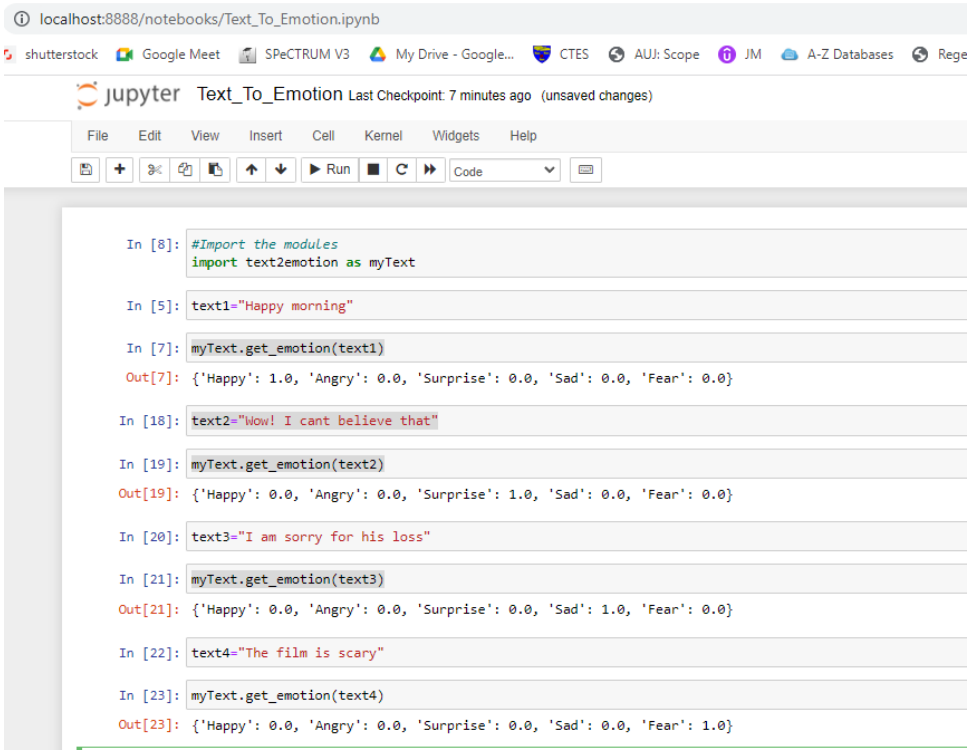





Created file name: SpeechToText.ipynb

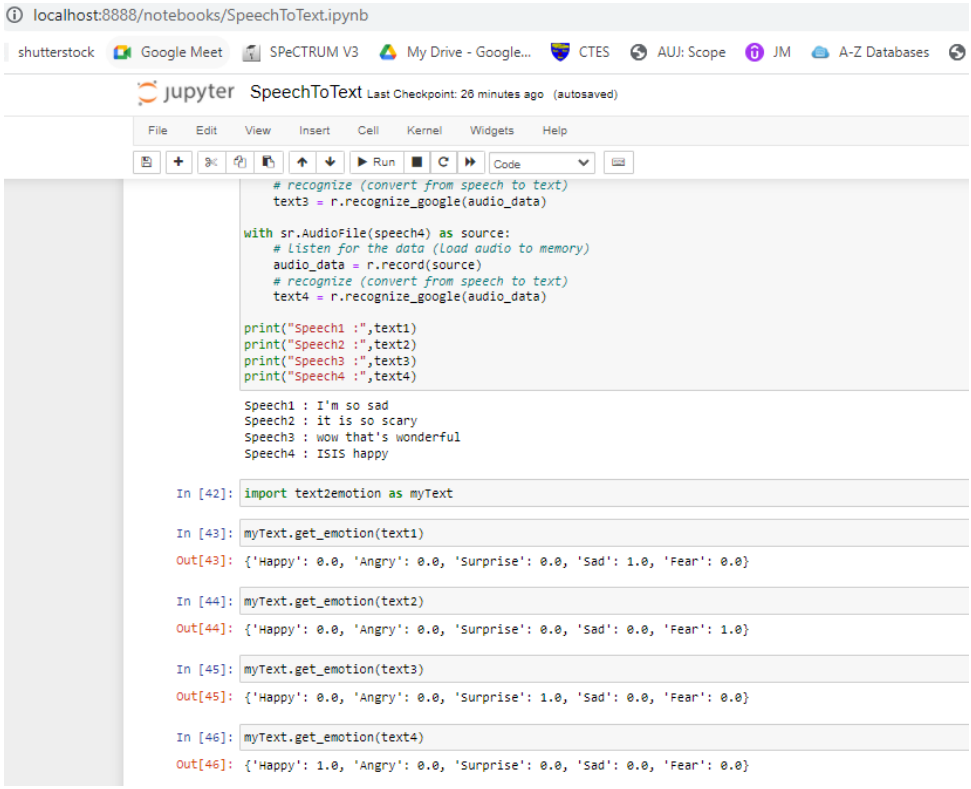
## 4. TEXT TO EMOTION

### SAMPLE




Created file name: Text_To_Emotion.ipynb

## FROM VOICE INPUT




Created file name: SpeechToText.ipynb

# DTW ALGORITHM IN PYTHON

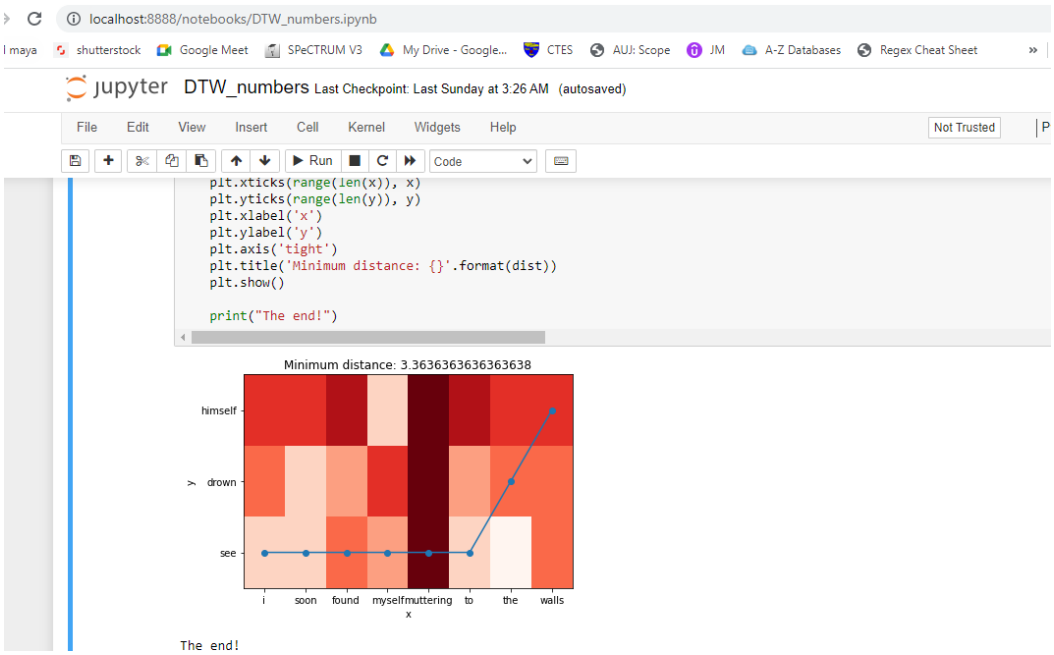
## DOCUMENTATION

Please see the help for the dtw.dtw() function which is the package's main entry point.

| | |
|---|---|
| countPaths (d[, debug]) | Count the number of warping paths consistent with the constraints. |
| dtw (x[, y, dist_method, step_pattern, ...]) | Compute Dynamic Time Warp and find optimal alignment between two time series. |
| dtwPlot (x, type, **kwargs) | Plotting of dynamic time warp results |
| dtwPlotAlignment (d[, xlab, ylab]) | |
| dtwPlotDensity (d[, normalize, xlab, ylab]) | Display the cumulative cost density with the warping path overimposed |
| dtwPlotThreeWay (d[, xts, yts, ...]) | Plotting of dynamic time warp results: annotated warping function |
| dtwPlotTwoWay (d[, xts, yts, offset, ...]) | Plotting of dynamic time warp results: pointwise comparison |
| itakuraWindow (iw, jw, query_size, reference_size) | |
| mvmStepPattern ([elasticity]) | Minimum Variance Matching algorithm |
| noWindow (iw, jw, query_size, reference_size) | |
| rabinerJuangStepPattern (ptype[, ...]) | Construct a pattern classified according to the Rabiner-Juang scheme (Rabiner1993) |
| sakoeChibaWindow (iw, jw, query_size, ...) | |
| slantedBandWindow (iw, jw, query_size, ...) | |
| warp (d[, index_reference]) | Apply a warping to a given timeseries |
| warpArea (d) | Compute Warping Path Area |
| DTW (obj) | The results of an alignment operation. |
| StepPattern (mx[, hint]) | Step patterns for DTW |

# IMPLEMENTATION

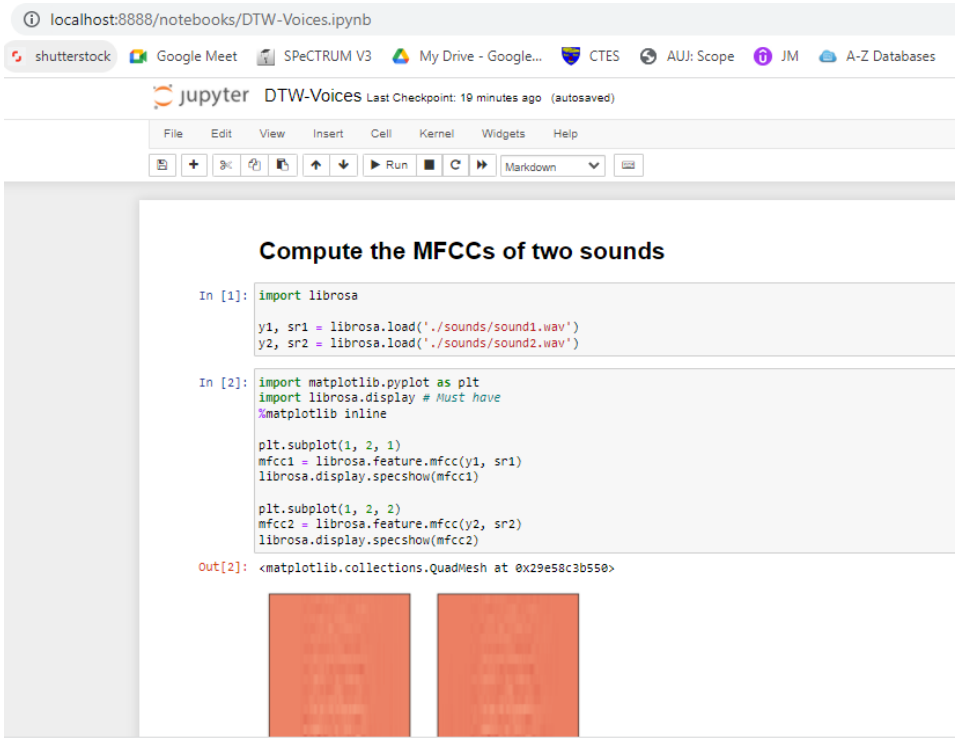
## 1. FOR NUMBERS



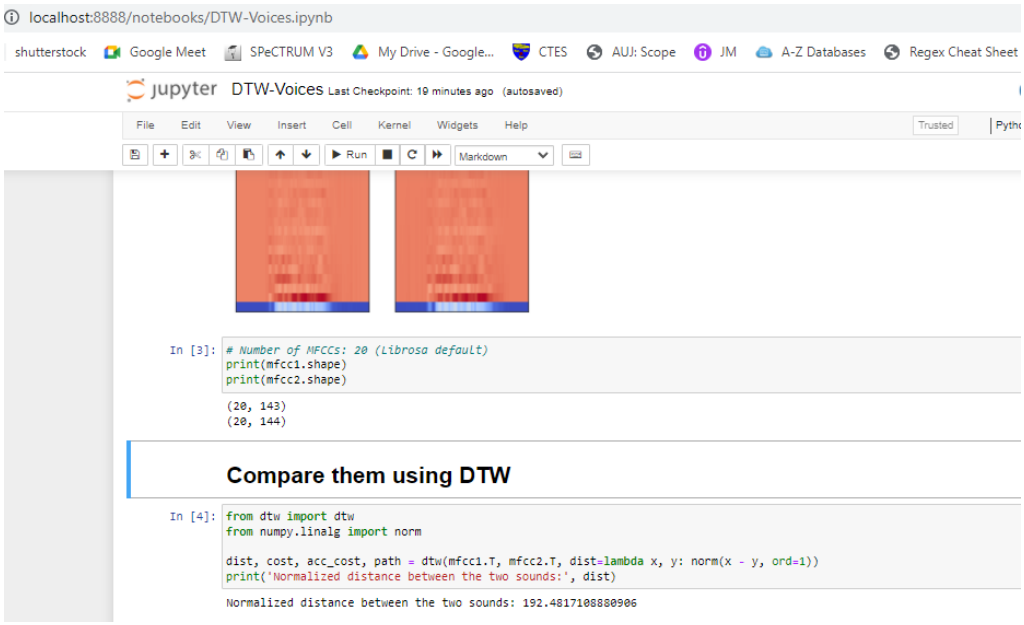

Created file name: DTW_numbers.ipynb

## 2. FOR VOICES

1) Compute Mel-frequency cepstral coefficients (MFCCs)

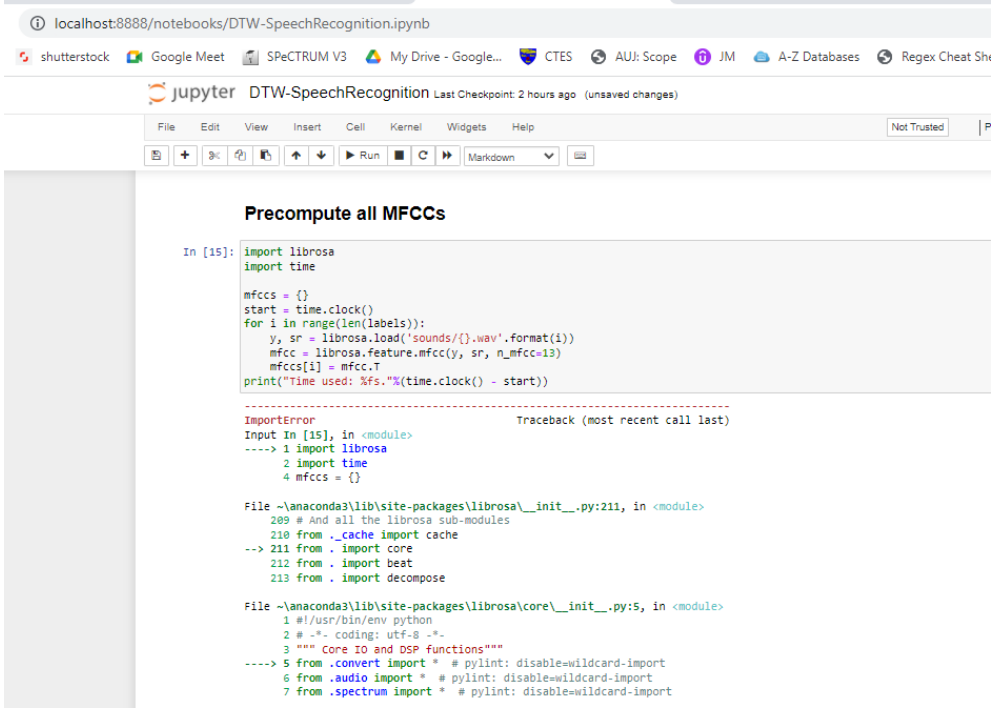


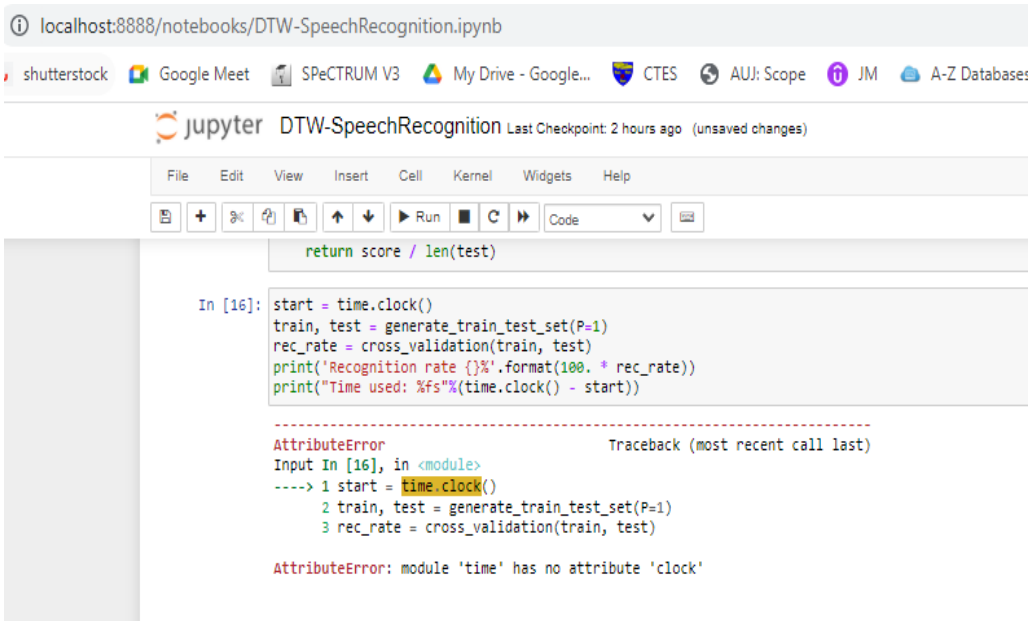

2) Input to DTW for comparison




Created file name: DTW-Voices.ipynb

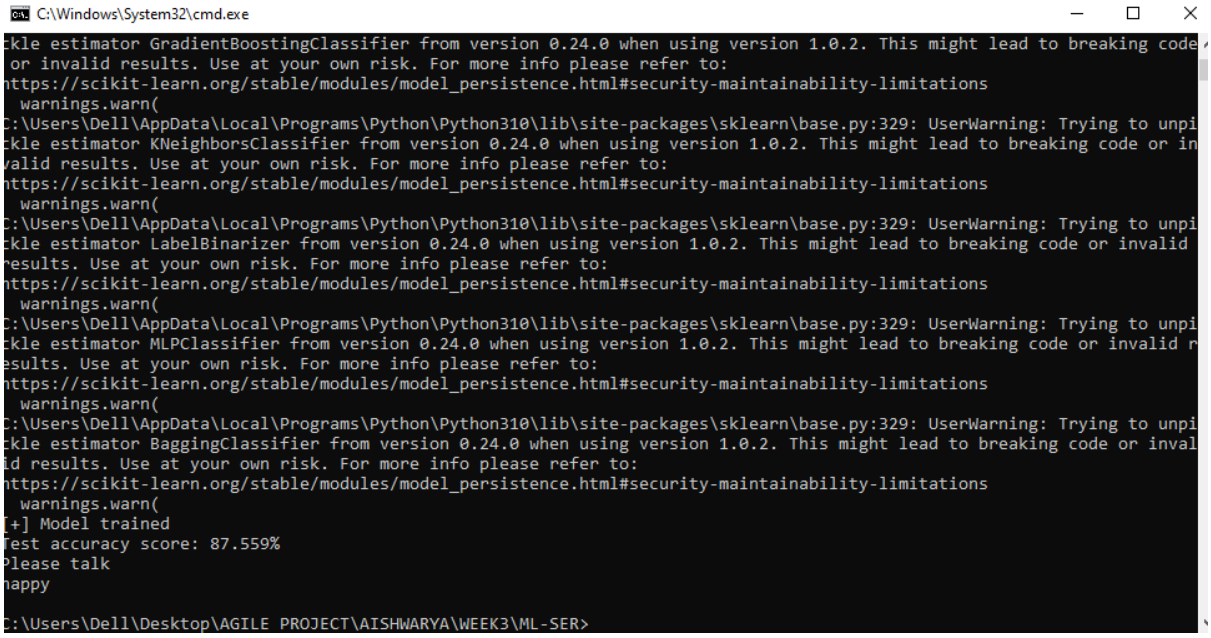
## 3. DTW-SPEECH RECOGNITION






Created file: DTW-SpeechRecognition.ipynb

# SCIKIT-LEARN MACHINE LEARNING ALGORITHM FOR SPEECH EMOTION RECOGNITON IN PYTHON



# APPENDIX

INSTALLATION AND ERROR FIXING

- pip3 uninstall PyCrypto
- pip3 install -U PyCryptodome
- time.clock() to time.time()
- pip install dtw-pythonconda
- install -c conda-forge unidecode
- conda install -c conda-forge emoji
- conda install jupyterlab
- conda install -c conda-forge librosa
- pip3 install SpeechRecognition pydub
- pip install pipwin
- pipwin install pyaudio
- pyaudio comp pbm
- conda install numba
- pip install text2emotion
- pip install pydub
- PyAudio-0.2.11-cp38-cp38-win_amd64.whl is not a supported wheel on this platform
- https://www.lfd.uci.edu/~gohlke/pythonlibs/#pyaudio
- time.process_time()
- \ and /
- pip install --upgrade ipython
- pip3 uninstall PyCrypto
- pip3 install -U PyCryptodome
- pip install --upgrade flask_sqlalchemy
- https://visualstudio.microsoft.com/visual-cpp-build-tools/
- pip3 install librosa==0.6.3 numpy soundfile==0.9.0 sklearn pyaudio==0.2.11

# REFERENCES


Garg, K., & Jain, G. (2016). A comparative study of noise reduction techniques for automatic speech recognition systems. *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. https://doi.org/10.1109/icacci.2016.7732361

Kardome Technologies. (2021). *The Difference Between Speech and Voice Recognition*. Www.Kardome.Com. https://www.kardome.com/blog-posts/difference-speech-and-voice-recognition

Librosa Development Team. (2021). *librosa.feature.mfcc — librosa 0.8.1 documentation*. Librosa.Org. https://librosa.org/doc/main/generated/librosa.feature.mfcc.html

Malek, A. (2020, January 25). *Signal framing — Ayoub Malek's blog*. Signal Framing. https://superkogito.github.io/blog/SignalFraming.html

Permanasari, Y., Harahap, E. H., & Ali, E. P. (2019). Speech recognition using Dynamic Time Warping (DTW). *Journal of Physics: Conference Series*, *1366*(1), 012091. https://doi.org/10.1088/1742-6596/1366/1/012091

Rockikz, A. (2019, July 29). *How to Make a Speech Emotion Recognizer Using Python And Scikit-learn - Python Code*. ThePythonCode. https://www.thepythoncode.com/article/building-a-speech-emotion-recognizer-using-sklearn

Toni, G. (2019). *Welcome to the dtw-python package — The dtw-python package 1.1.6 documentation*. DTW. https://dynamictimewarping.github.io/py-api/html/

Washani, N., & Sharma, S. (2015). Speech Recognition System: A Review. *International Journal of Computer Applications*, *115*(18), 7–10. https://doi.org/10.5120/20249-2617