# **AI Translator-Tech**

Proposed technical libraries for the project

* **Librosa** (audio management)
* **Pykaldi** (audio management)
* **Spacy** (text management)
* **RNN/LSTM** or **TensorFlow** (data processing and model training)

## **Project Resource Links**

1. <https://www.kaggle.com/mayarmohsen/cnnw7/notebook>
2. <https://www.kaggle.com/harunshimanto/speech-classification-using-cnn> (speech classification with CNN)
3. <https://www.kaggle.com/davids1992/speech-representation-and-data-exploration> (Ben this is your own resource here containing data cleaning aspect)
4. <https://www.kaggle.com/shivamburnwal/speech-emotion-recognition>
5. <https://medium.com/wavey-ai/end-to-end-speech-recognition-f13f0d0197c7>
6. <https://medium.com/algoanalytics/automatic-labelling-of-text-for-nlp-5270e70a2f5f>
7. <https://towardsdatascience.com/audio-deep-learning-made-simple-sound-classification-step-by-step-cebc936bbe5>
8. <https://medium.com/mlearning-ai/audio-classification-using-wavelet-transform-and-deep-learning-f9f0978fa246>
9. <http://www.ijamtes.org/gallery/123-feb2019.pdf> (Feature extraction research documentation.)

The main phases that will be used in the development of this AI project are as follows;

## **TEXT TREATMENT**

...........

## **AUDIO FILE TREATMENT**

* **Investment and hypothesis:** This involves the different hypotheses involved during the data creation of the project which factors have been taken into consideration in order to properly monitor the quality and validity of the acquired data. The high biased issue.

1. We have a dictionary of words and phrases.
2. Each voice collected is represented into two categories of people is Young(18 - 34) and Old (35 - 80) both male and female recorded in a compensated manner such that bias is avoided to the maximum.

* **Raw Data:** The voices are regarded regarding the category of the language spoken in this case we are focusing on two of them which are **Yemba** and **Ngeumba**. This data is supposed to be recorded into either in the .wav or .mp3 format numerical processing.
* **Data Quality Control:** This phase mostly takes into consideration the quality of the data been generated by the translators, making sure that it contains the least reverberation and background noise.

1. This phase also keeps track of the semanticity of the language used, taking note of the form, structure and appearance of each voice recorded related to a particular word or phrase into the projects dictionary. <https://en.wikipedia.org/wiki/Semantics> .

Seamantics

Phonology and form in linguistics

* **Data Cleaning:** This phase mostly entails making sure that the data is uniform.  
  This entails the following procedures to be adapted on the diverse dataset of audio files.

1. Silence removal (This entails the removal of background effects if exists
2. Dimensional reduction( reducing the frequency length in voice using matrix reduction technique while conserving the audio data quality )

* **Feature Extraction steps:**

1. Resampling
2. VAD (Voice activity detection: also known as speech activity detection or speech detection, is the detection of the presence or absence of human speech, used in speech processing)
3. Maybe padding with 0 to make signals be equal length
4. Log spectrogram (or MFCC, or PLP or LPCC(Linear Prediction Cepstral Coefficients))
5. Features normalization with mean and std
6. Stacking of a given number of frames to get temporal information

* **Deep learning modelling simulations:** This entails the model selection which has several proposals to which from diverse articles. Our input here is a text and output a voice matrix. This will be later on be converted into a voice
* **Model deployment:** This entails just the extraction of the produced model to be incorporated in the application of the model