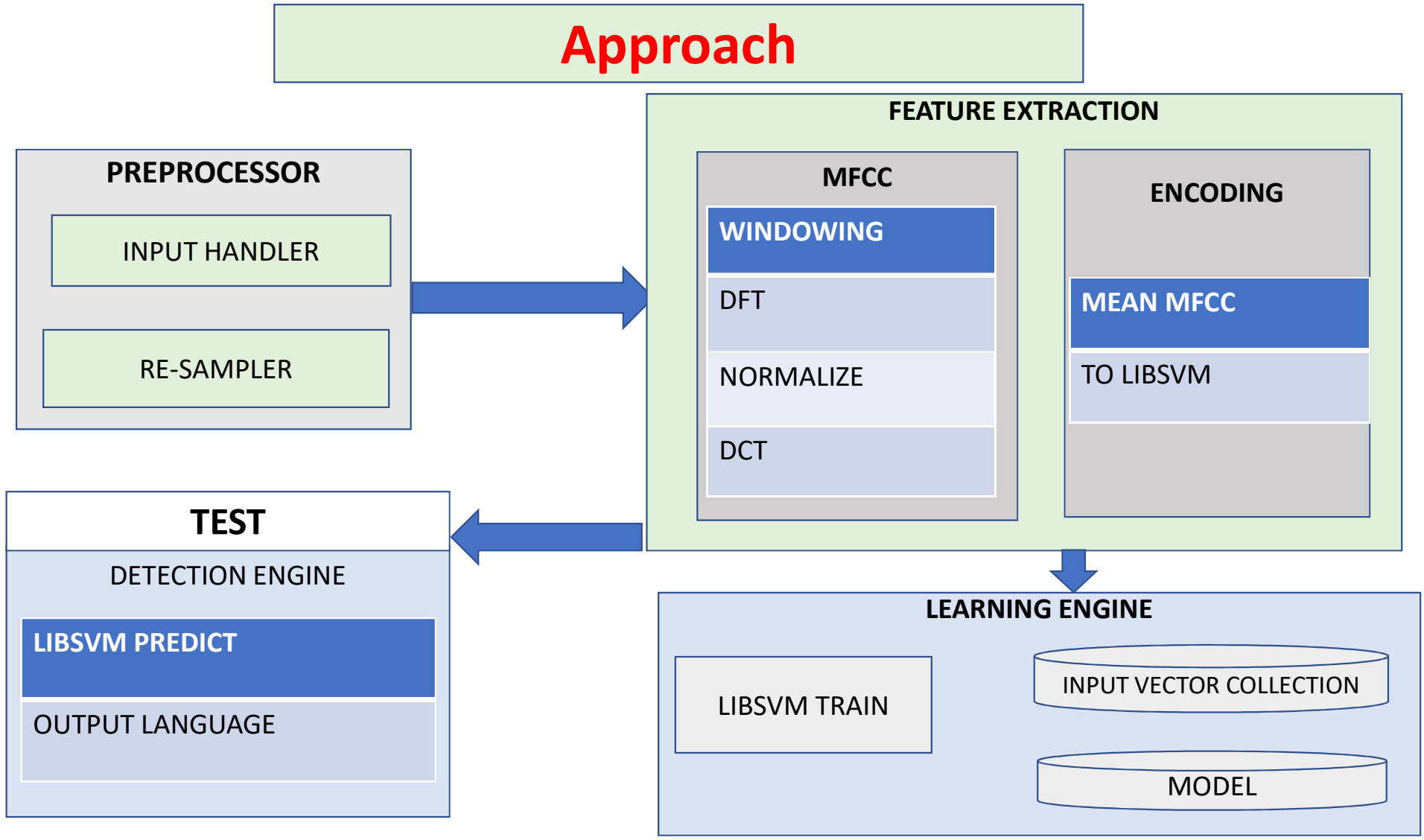
Ministry/ Organization name: Bharat Electronics Limited , Bengaluru
Problem Statement :Open set Language Diarization for Indian Languages audio data.
Team Name :Stars
PS Number:- MU456
Team Leader Name : Rahul kumar
College Code : 1-3511492137

Team member

1. RAHUL KUMAR
2. RAUSHAN KUMAR
3. KAPIL KUMAR
4. KISHOR KUMAR MUNNA
5. DEEPA KUMARI
6. SAPNA KUMARI

Approach


PREPROCESSOR
INPUT HANDLER
RE-SAMPLER
FEATURE EXTRACTION
MFCC
WINDOWING
DFT
NORMALIZE
DCT
ENCODING
MEAN MFCC
TO LIBSVM
TEST
DETECTION ENGINE
LIBSVM PREDICT
OUTPUT LANGUAGE
LEARNING ENGINE
LIBSVM TRAIN
INPUT VECTOR COLLECTION
MODEL

Idea / Approach

The process of language identification is carried out progressively in three stages 1) pre-processing 2) feature extraction 3) machine learning. These three phases together contribute to the language identification.

We make use of the acoustic model for Language Identification. The acoustic model means that only those features which are independent of prosodic or phonotactic information are used to model languages. One such feature is Mel Frequency Cepstral Coefficients (MFCC). They are derived from a type of cepstral representation of the audio sample which is a nonlinear "spectrum-of-aspectrum". The MFC frequency bands are equally spaced on the mel scale, which approximates the human auditory system's response more closely than the linearly-spaced frequency bands. The main task of the feature extraction block is to extract MFCC for the audio sample. It then computes the mean MFCC. We makes use of Support Vector Machine (SVM) as the classifier. The process has two phases: training the SVM with cepstral data (mean MFCC) and testing it with speech samples. The SVM in the training phase creates a model based on the input feature vectors for different languages. This model file is used by the classifier in the testing phase to predict language. The final result is the language identified for the given test samples.

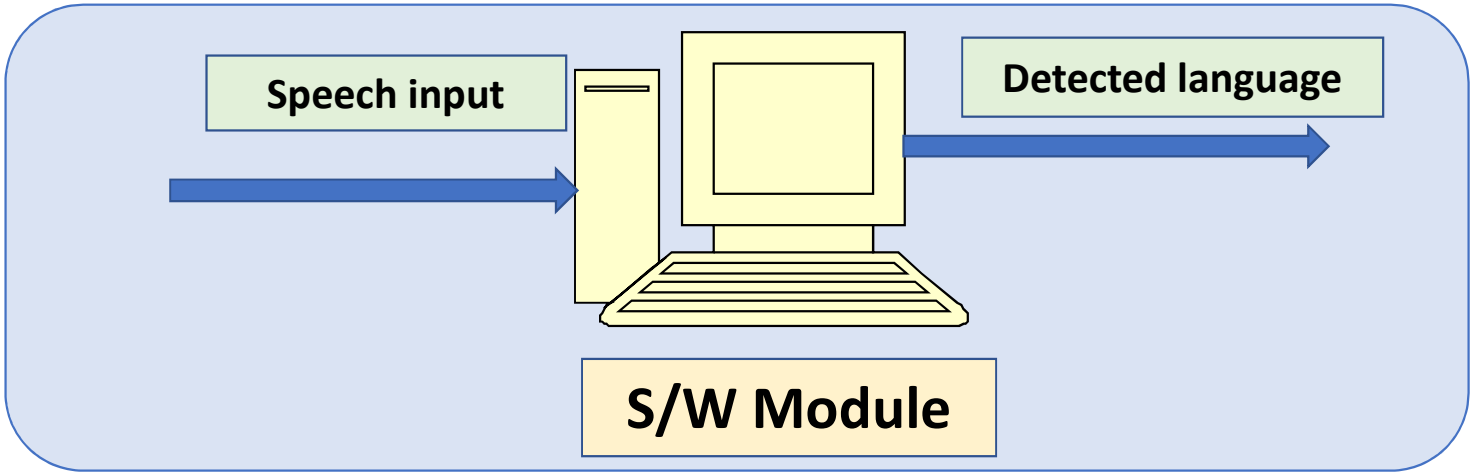
Programming language

- Python

AI

- Automatic speech Recognition(ASR) (in=audio data, out=text)
- Natural language understanding(NLU) (in=text, out=intent)
- Feature Extraction (in=intent, out=structured data)
- Deep neural network (DNN)
- Support vector machine (SVM)

Use case/ Show stopper



Dependencies

- The machine should have the support of vector machine libraries.
- Machine should have python support.

Show stopper

- If input speech sample contain abrupt utterances, noise or background-noise, this is especially if background-noise is also speech, then module will be unable to predict the language correctly.
- In India many languages come from the same origin. Many words are common in Indian languages. In that case, module will predict different languages as one (if in audio data, both languages have same words).