1. scotus\_data\_with\_zscore\_of\_audio\_labels.py  
   to obtain concatenation of existing SCOTUS data with M-Turk rater wise Z-Score or petitioner and respondent   
     
   Input :   
   pilot6b\_long\_only\_f\_n\_outcomes.csv  
   remove\_data\_unknown.csv  
   remove\_features\_unknown.csv  
     
   Output :  
   merge\_audio\_data.csv  
   merge\_audio\_features.csv
2. mfcc\_generator\_1955-1998.py  
   Generate MFCC features for audio files between 1955 and 1998  
     
   Input  
   Unzipped Audio Files from 1955 to 1998   
     
   Output  
   mfccFileFeatures\_1955\_1998.csv
3. mfcc\_generator\_1998-2014\_data  
   Generate MFCC features for audio files between 1998 and 2014  
     
   Input  
   Unzipped Audio Files between 1998 and 2014  
     
   Output  
   data/input/mfccFileFeatures\_1998\_2014.csv
4. We merge pilot with mfcc\_generator\_1998-2014\_data (part 3) to get mfcc\_with\_ratings.csv  
     
   Input  
   data/input/mfccFileFeatures\_1998\_2014.csv  
   pilot6b\_long\_only\_f\_n\_outcomes.csv  
     
   Output  
   mfcc\_with\_ratings.csv
5. We so string processing lawyer names for easy pattern matching in the file advocates.csv   
     
   Input  
   advocates.cv  
     
   Output  
   modifiedAdvocates.csv
6. We merge output of mfcc\_generator\_1955-1998 (part 2)to modifiedAdvocates.csv to get advocatesMfccMerged\_1955-1998.csv to link mfcc to lawyer names and petitioner/respondent  
     
   Input  
   modifiedAdvocates.csv  
     
   Output  
   advocatesMfccMerged\_1955-1998.csv
7. We remove duplicates from advocatesMfccMerged\_1955-1998.csv and get into format docket pMFCC and rMFCC as a row  
     
   Input  
   advocatesMfccMerged\_1955-1998.csv  
     
   Output  
   mfcc\_1955\_p\_r.csv
8. Run  
   docket\_predict\_pClass\_label\_rClassLabel\_1955-1998.py  
     
   We train a model from the file obtained from step 4 (mfcc\_with\_ratings.csv) to predict class banalized ratings of the audio traits on mfcc\_1955\_p\_r.csv  
     
   Input  
   mfcc\_with\_ratings.csv ……to train and validate the model  
   mfcc\_1955\_p\_r.csv ………to predict unknown banalized voice trait ratings  
     
   Output  
   docket\_p\_r\_audioCombined.csv
9. We binarize the z score of audio rating by murk workers in pilot6b\_long\_only\_f\_n\_outcomes.csv  
     
   Input  
   pilot6b\_long\_only\_f\_n\_outcomes.csv  
     
   Output  
   docket\_p\_r\_audioSplit.csv
10. Basically output of 8 and 9 are in same format but rows belong to different period in time….so we simply stack them vertically using  
    docket\_pClassLabel\_rClassLabel\_1955-2014.csv  
      
    Input  
    docket\_p\_r\_audioCombined.csv  
    docket\_p\_r\_audioSplit.csv  
      
    Output  
    completeAudioDataClssification.csv
11. Now we combine output of 10 to original scouts data using ScotusData\_docket\_pClassLabel\_rClassLabel.py  
      
    Input  
    completeAudioDataClssification.csv  
    remove\_data\_unknown.csv  
    remove\_features\_unknown.csv  
      
    Output  
    mergeComplete\_audio\_rawData.csv  
    mergeComplete\_audio\_Featuredata.csv