Female and Male Voice Recognition

Final Boot Camp Project

Alexej Kasakow

Female and Male Voice Recognition

- Which are the key parameters to recognize a female and male voice ?

- How AI is used to recognize female and male voice?

- What else can we do with the tools for female and male voice recognition ?

Female and Male Voice Recognition

Voice dimorphism refers to the differences in voice characteristics between males and females, typically resulting from differences in larynx size and anatomy. Voice dimorphism can result in differences in vocal pitch, resonance, and other acoustic features, which can impact speech intelligibility, speech production, and voice quality. Voice dimorphism is widely studied in linguistics, speech and hearing sciences, and other related fields, as it plays a role in language development, communication, and social interactions.

Female and Male Voice Recognition with Python

- Voice Recognition: Machine learning problem

- ML: promising result for classification techniques

- In my project:
- K-Nearest Neighbor KNN
- Decision Tree DT
- Ramdon Forest RF
- Support Vector Machine SVM with Poly Kernel and
- Gradient Boosting

Voice signal

- Speech signal - most common means of communication
- Intonation
- Speech Rate
- Duration

Voice signal

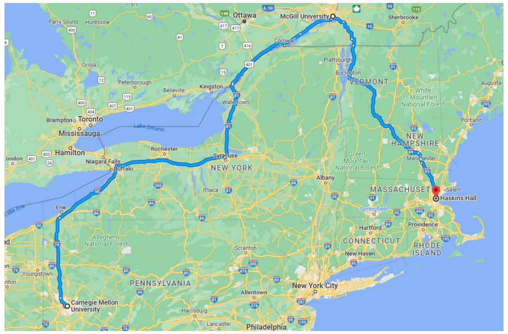
- Human has instincts to identify a difference
- For Computer:
- we need to feed with the biggest possible dataset
- train and test
- provide algorithms

Research

- Kaggle dataset with 3168 recorded voices

- Based upon acoustic properties of the voice and speach

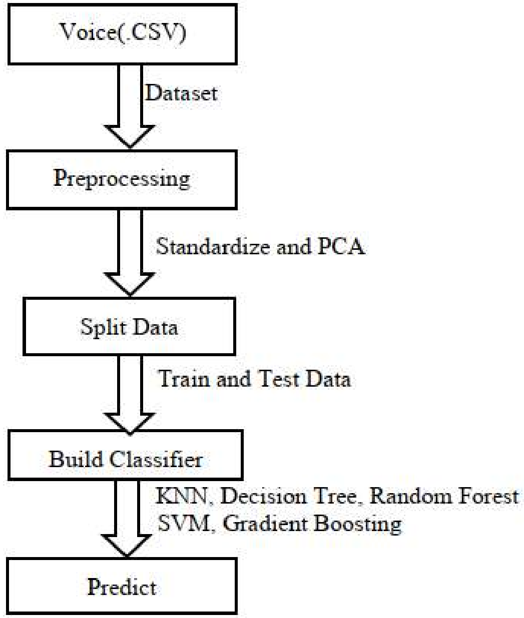
- Research:
- Harvard-Haskins Database Boston
- McGill University,Laval Canada
- Carnegie Mellon University, Pittsburgh



Workflow – Female, Male Voice Recognition

Acoustic properties	
Properties	***Description***
meanfreq	mean frequency (in kHz)
sd	standard deviation of frequency
median	median frequency (in kHz)
Q25	first quantile (in kHz)
Q75	third quantile (in kHz)
IQR	interquantile range (in kHz)
skew	skewness
kurt	kurtosis
sp.ent	spectral entropy
sfm	spectral flatness
mode	mode frequency
centroid	frequency centroid
meanfun	average of fundamental frequency measured across acoustic signal
minfun	minimum fundamental frequency measured across acoustic signal
maxfun	maximum fundamental frequency measured across acoustic signal
meandom	average of dominant frequency measured across acoustic signal
mindom	minimum of dominant frequency measured across acoustic signal
maxdom	maximum of dominant frequency measured across acoustic signal
dfrange	range of dominant frequency measured across acoustic signal
modindx	modulation index. Calculated as the accumulated absolute difference between adjacent measurements of fundamental frequencies divided by the frequency range
label	male or female

Workflow – Female, Male Voice Recognition



Signal - Exploratory Data Analysis

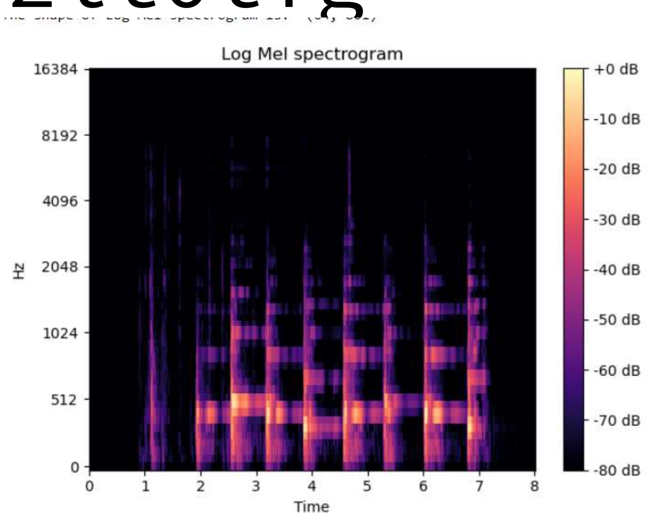
- Pre-Processing techniques
- Missing values
- Standardization

Dataset

- Own research:
- A create a wave file
- B examine wave files
- C create a dataset from wave file to *.csv
- D extract and analyse different features from:
- Voice, speech, breathing, heart beats

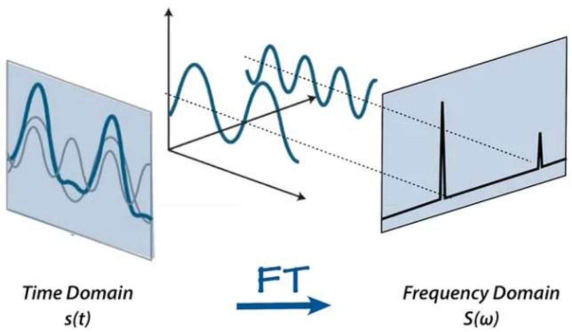
Sound definition

- https://www.youtube.com/watch?v=P9Kozlt0tTg



Sound definition

- The Fourier Transform
- An audio signal is comprised of several single-frequency sound waves. When taking samples of the signal over time, we only capture the resulting amplitudes. The Fourier transform is a mathematical formula that allows us to decompose a signal into it's individual frequencies and the frequency's amplitude. In other words, it converts the signal from the time domain into the frequency domain. The result is called a spectrum.



Exploratory Data Analysis – Mean Frequency

Mean frequency is the key component to distinguish the voice between male and female. The boxplot of mean frequency v/s label i.e. either male or female is given in figure attached

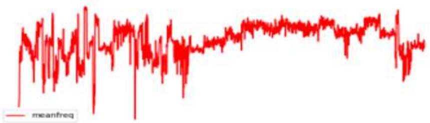


Fig.3. Mean frequency of male

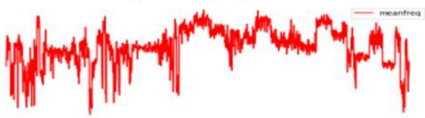
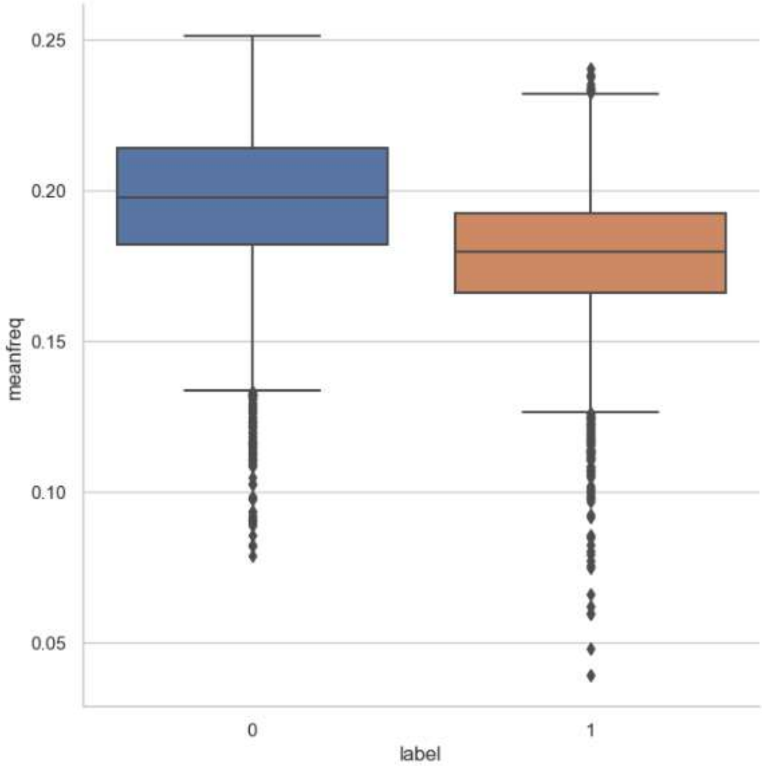
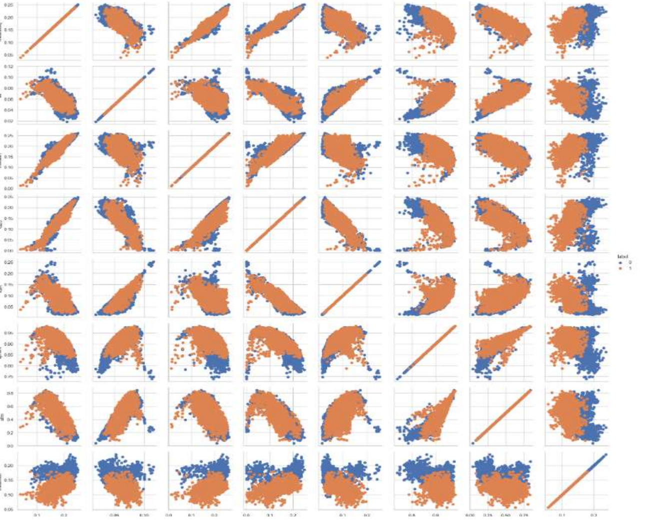


Fig.4. Mean frequency of female

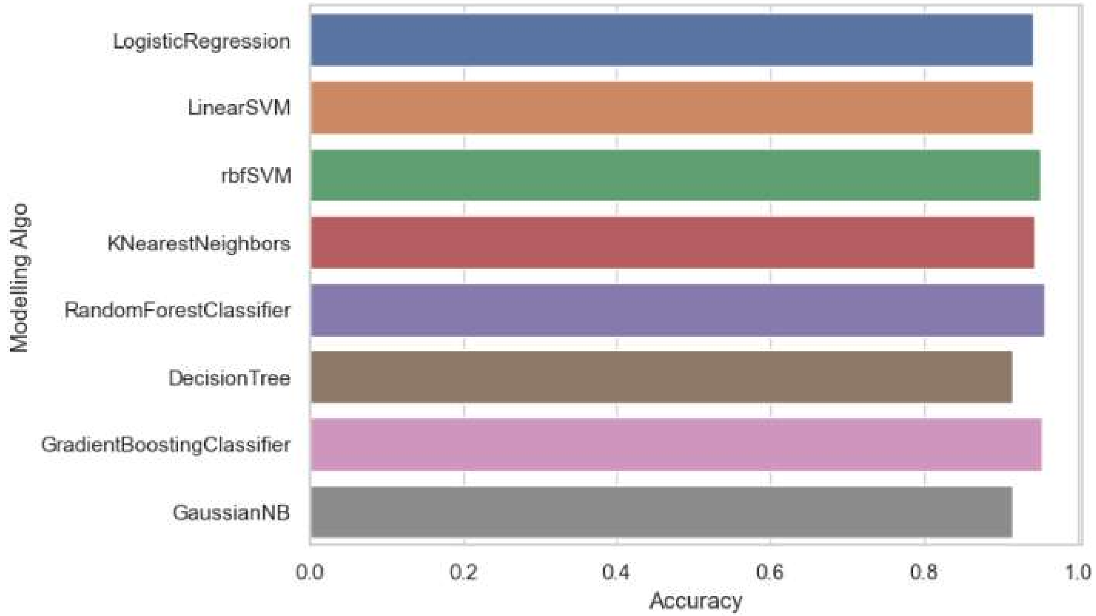


Exploratory Data Analysis – Pairs of features

Again high difference in females and males mean fundamental frequency. This is evident from the heat map which clearly shows the high corelation between meanfun and the 'label'.
Now we move onto analyzing different features pairwise. Since all the features are continuous the most reasonable way to do this is plotting the scatter plots for each feature pair. I have also distinguished males and feamles on the same plot which makes it a bit easier to compare the variation of features within the two classes.¶



Exploratory Data Analysis – Mean Frequency



Signal - Exploratory Data Analysis

- Pre-Processing techniques
- Missing values
- Standardization

Today’s Presentation

From Fig.5, it can be deduced that as number of components increases accuracy also increases. PCA is performed to increase the accuracy. For almost all the classifiers, accuracy remains constant after number of components reaches 16. But there is no use of performing PCA and reducing the dimensions to 16 from 20 as it will not make much difference. The reason that PCA does not perform better is that for each different classifier, different number of features are showing importance. Performance depends on the feature which the classifier takes while building the model. This can be seen in Fig.6 which shows feature extraction of Gradient Boosting classifier and Fig.7 which shows feature extraction of Random Forest.

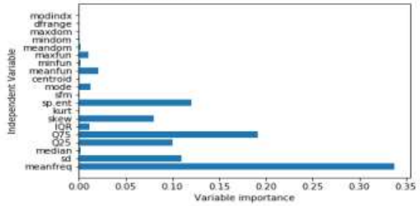


Fig.6. Feature extraction of Gradient Boosting

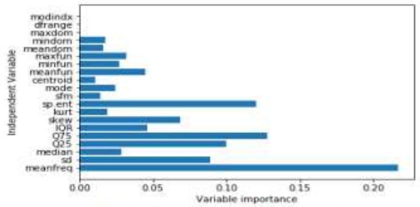


Fig.7. Feature extraction of Random Forest

Thanks for your attention !