Module 9: Supervised Learning -II

Case Study

Case Study

Domain – Media

focus – optimize selection process

<u>Business challenge/requirement</u>
Motion Studios is the largest Radio production house in Europe. Their total revenue $ 1B+. Company has launched a new reality show – "The Star RJ". The show is about finding a new Radio Jockey who will be the star presenter on upcoming shows.

In first round participants have to upload their voice clip online and the clip will be evaluated by experts for selection into the next round. There is a separate team in the first round for evaluation of male and female voice.

Response to the show is unprecedented and company is flooded with voice clips.

You as a ML expert have to classify the voice as either male/female so that first level of filtration is quicker.

<u>Key issues</u>
Voice sample are across accents

<u>Considerations</u>
The output from the pre-processed WAV files were saved into the CSV file

<u>Data volume</u>
- Approx 3000 records – file voice-classification.csv
 Fields in Data

- 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 (see note in specprop description)
- kurt: kurtosis (see note in specprop description)
- sp.ent: spectral entropy
- sfm: spectral flatness
- mode: mode frequency

- centroid: frequency centroid (see specprop)
- peakf: peak frequency (frequency with highest energy)
- 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

Additional information
- NA

Business benefits
Since "The Star RJ" is a reality show, time to select candidates is very short. The whole success of the show and hence the profits depends upon quick and smooth execution