

Module 9: Supervised Learning -II

Case Study – 1

edureka!

edureka!

© Brain4ce Education Solutions Pvt. Ltd.

Case Study –1

Domain – Media

focus – optimize the selection process

Business challenge/requirement

Motion Studios is the largest Radio production house in Europe. Their total revenue is \$ 1B+. The 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 the first round, participants have to upload their voice clip online and the clip will be evaluated by experts for selection in the next round. There is a separate team in the first round for the evaluation of male and female voices.

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

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

Key issues

Voice samples are across accents

Considerations

The output from the pre-processed WAV files was saved into the CSV file

Data volume

- Approx 3000 records – file voice-classification.csv

Fields in Data

- meanfreq: mean frequency (in kHz)
- sd: standard deviation of the 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 the highest energy)
- meanfun: average of fundamental frequency measured across the acoustic signal
- minfun: minimum fundamental frequency measured across the acoustic signal
- maxfun: maximum fundamental frequency measured across the acoustic signal
- meandom: average of dominant frequency measured across the acoustic signal
- mindom: minimum of dominant frequency measured across the acoustic signal
- maxdom: maximum of dominant frequency measured across the acoustic signal
- dfrange: range of dominant frequency measured across the 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, the time to select candidates is very short. The whole success of the show and hence the profits depends upon quick and smooth execution