classification model to predict the gender

**Project Description:**

**Problem statement:** Create a classification model to predict the gender (male or

female) based on different acoustic parameters

**Context:** This database was created to identify a voice as male or female, based upon

acoustic properties of the voice and speech. The dataset consists of 3,168 recorded

voice samples, collected from male and female speakers. The voice samples are preprocessed by acoustic analysis in R using the seewave and tuneR packages, with an

analyzed frequency range of 0hz-280hz (human vocal range).

**Column 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 (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 differencebetween adjacent measurements of fundamental frequencies divided by the

frequency range

• label: male or female

**Steps to consider:**

1)Remove/handle null values (if any)

2)Depict percentage distribution of label on a pie chart

3)Considering all the features as independent feature and label as dependent feature,

split the dataset training and testing data with test size=20%

4)Apply the following classifier models on training dataset and generate predictions for

the test dataset

a. Decision Tree Classifier

b. Random Forest Classifier

c. KNN Classifier

d. Logistic Regression

e. SVM Classifier

5)Also generate confusion\_matrix and classification report for each model generated inQ4.

6)Report the model with the best accuracy