Report On Machine Learning Minor Project

**Gender Prediction And Analysis Based On Classification Models**

Name : Akash L M

Email Id: [akashlmnmamit@gmail.com](mailto:akashlmnmamit@gmail.com)

Batch: February

Project Name: Machine Learning February Minor Project

Topic : Machine Learning

Submitted on : 27 March 2022

**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 difference between adjacent measurements of fundamental frequencies divided by the frequency range

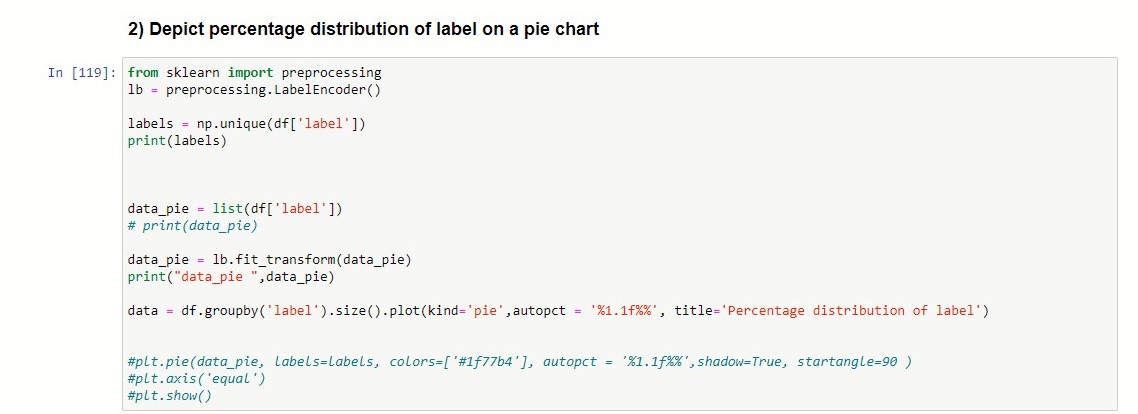
• label: male or female

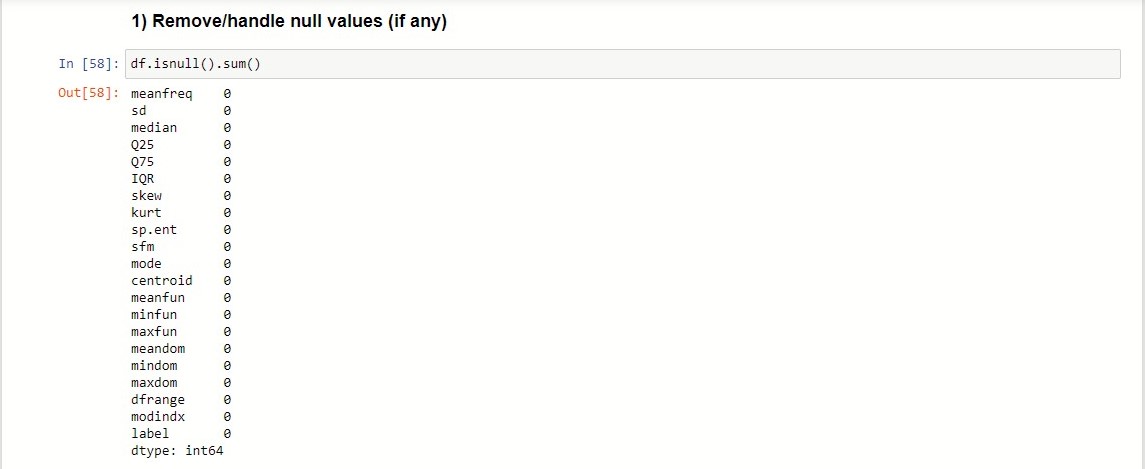
**Dataset:** <https://drive.google.com/file/d/10b2NKUtKzNm1by28AiKh5Dd_NFfARRHy/view?usp=sharing>

**Images :**



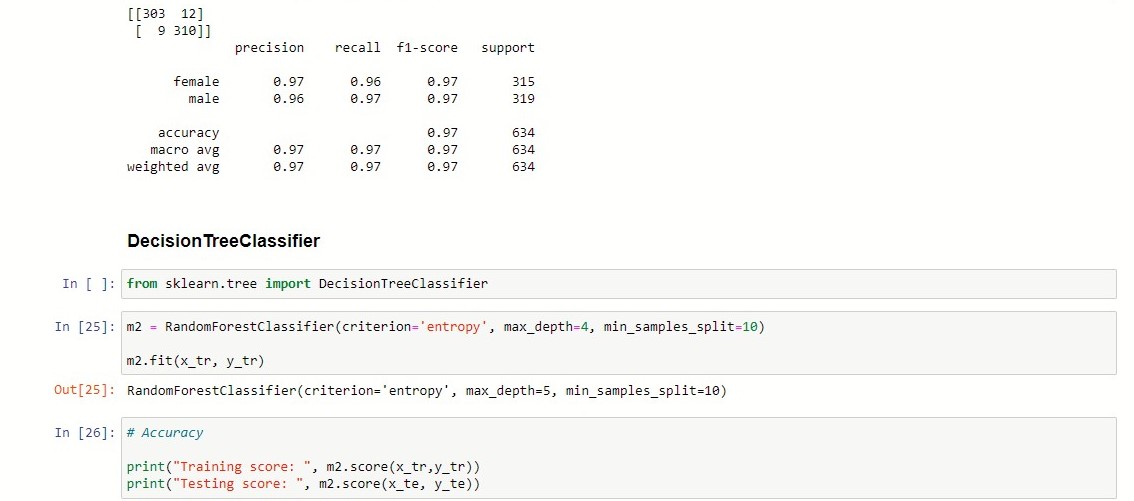
**Steps 1 :**

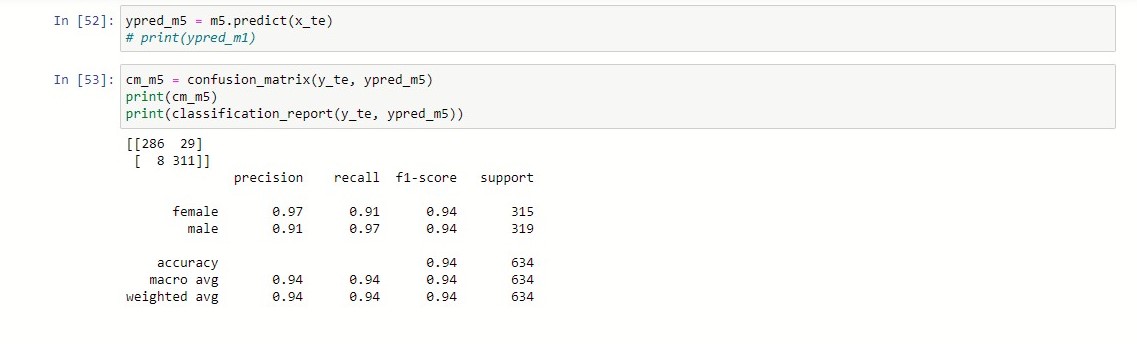
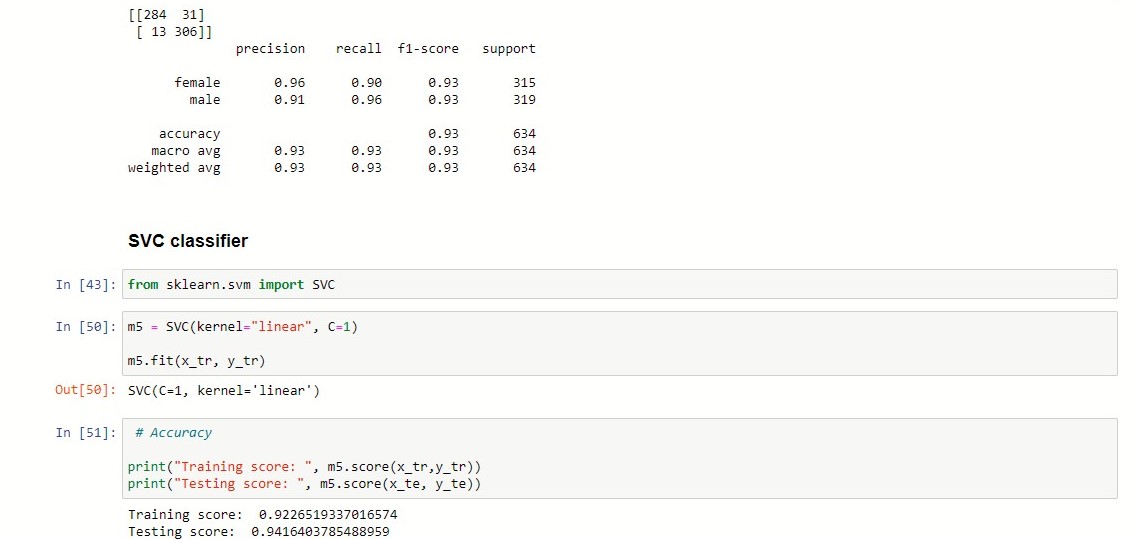
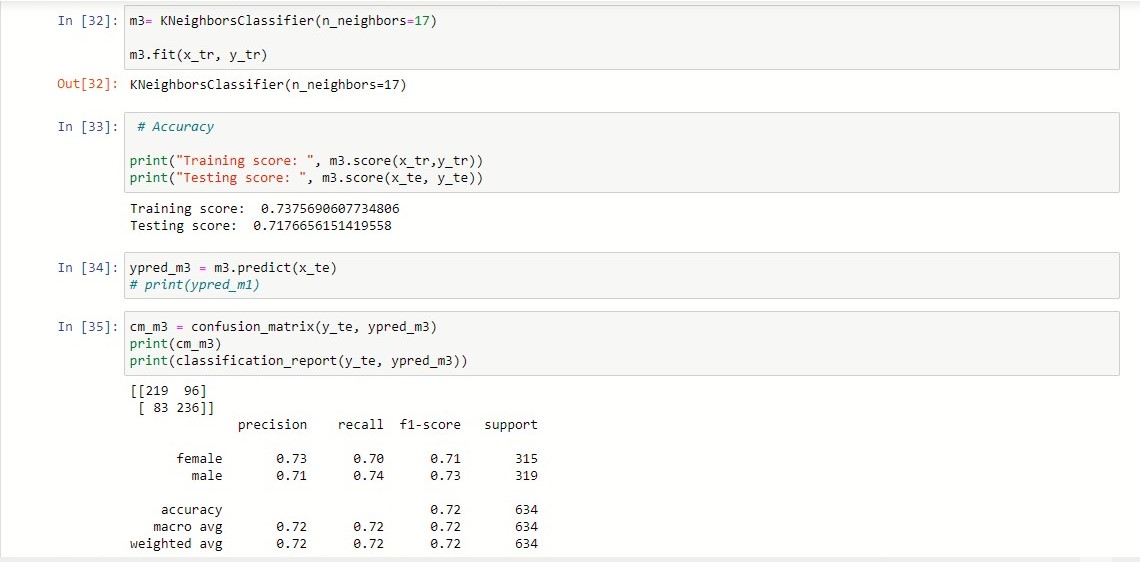
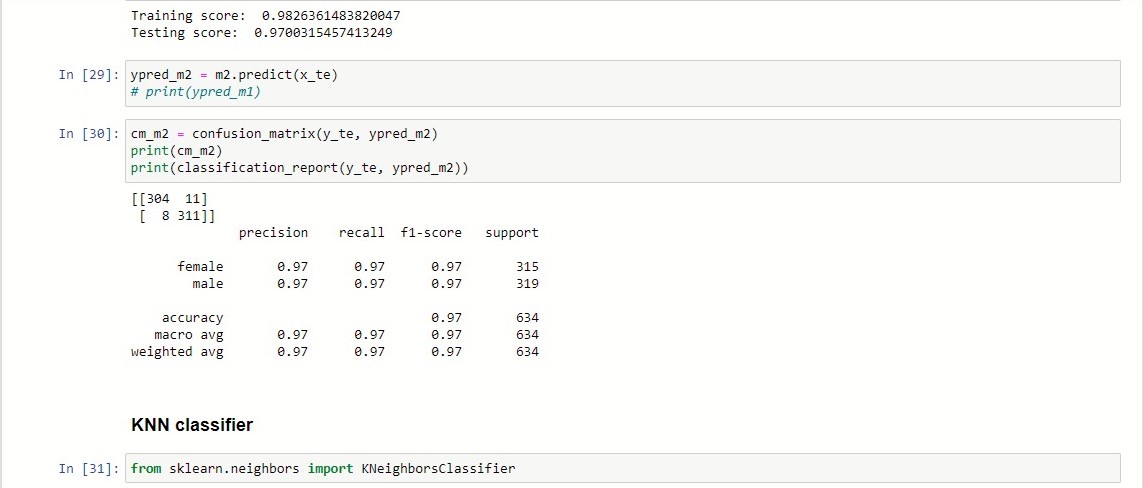
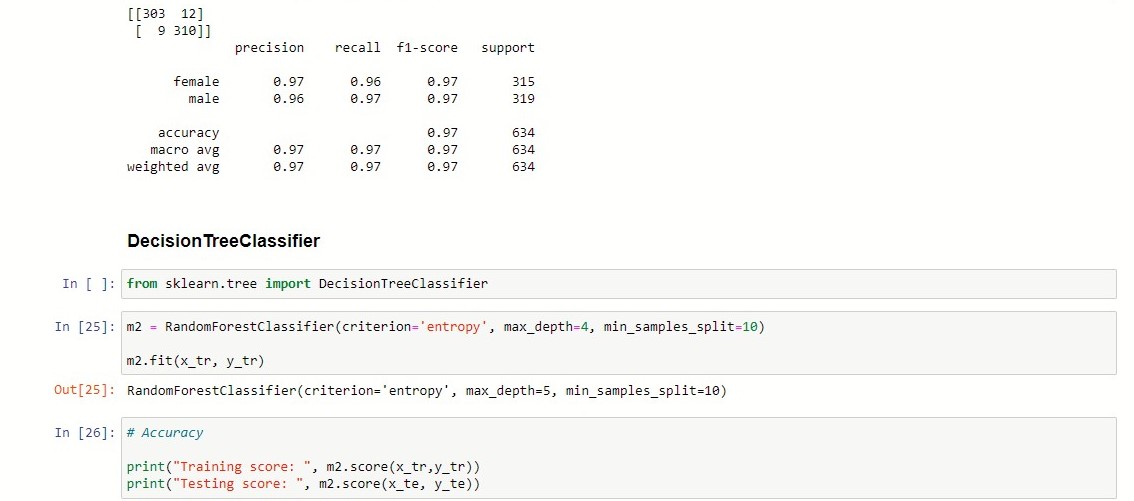
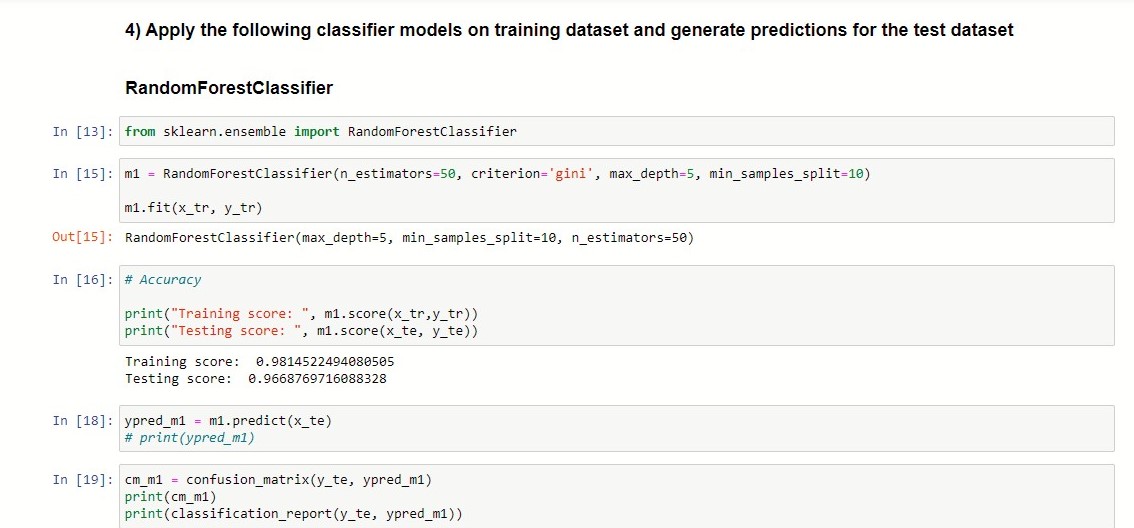
**Steps 2:**



**Steps 3:**



**Steps 4 and 5:**



**Steps 6:**



**Result:**

DecisionTree Accuracy: 0.9700315457413249

RandomForest Accuracy: 0.9668769716088328

KNNClassifer Accuracy: 0.7176656151419558

LogisticRegression Accuracy: 0.9305993690851735

SVMClassifier Accuracy: 0.9416403785488959

Best Suitable model is **RandomForestClassifier** and accuracy is **0.97.**