電子商務技術 HW7

資料集：gender_recog.csv

任務：根據處理過的聲音數據判斷此聲音來自 male/female

以下題目請使用 Python 完成

1. 將資料集讀取為 DataFrame 格式，並將原始順序打亂
2. 使用 scikit-learn 建立 RandomForest 分類器，並以 10 cross validation 評估模型在此資料集的分類表現 (評估指標：F1-score，請印出平均 F1-score)
3. 使用 scikit-learn 建立 MLP 分類器，並以 10 cross validation 評估模型在此資料集的分類表現 (評估指標：F1-score，請印出平均 F1-score)
4. 使用 Keras 建立 MLP 分類器，並以 10 cross validation 評估模型在此資料集的分類表現 (評估指標：F1-score，請印出平均 F1-score)
5. 請問你在上述 3 題所使用的 F1-score 是以哪一類別為依據 (male、female、兩者平均)
6. 以 t-test 比較上述三個模型的表現，並簡述結論
7. 有某新資料各屬性值如下，請判斷此聲音的性別：(可從 new_data.txt 中直接複製)

meanfreq	sd	median	Q25	Q75	IQR	skew	kurt
0.1528	0.0735	0.1490	0.0479	0.2095	0.1416	1.5325	7.3388
sp.ent	sfm	mode	centroid	meanfun	minfun	maxfun	Meandom
0.9631	0.7383	0.1325	0.1427	0.1101	0.0111	0.2539	0.2982
mindom	maxdom	dfrange	modindx				
0.0078	2.7235	2.7184	0.1251				

8. 哪一個屬性在 RandomForest 的分類中最重要?

作業繳交說明

- 繳交期限：6/1 (三) 中午 12:00
- Python 題請繳交.ipynb 檔，檔名 ECT_HW7_學號。
    - 程式中請以註解或文字方塊標示題號

- 上傳至 ee-class 作業區，遲交一天扣該次作業得分 **5%**，最多扣 **50%**。

附件

- 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
- 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