TFW
FINAL PROJECT

音樂特徵提取
分析音樂風格成份
比較模型與資料集差異

PRESENTER:
郭政穎

TOPIC IDEAS

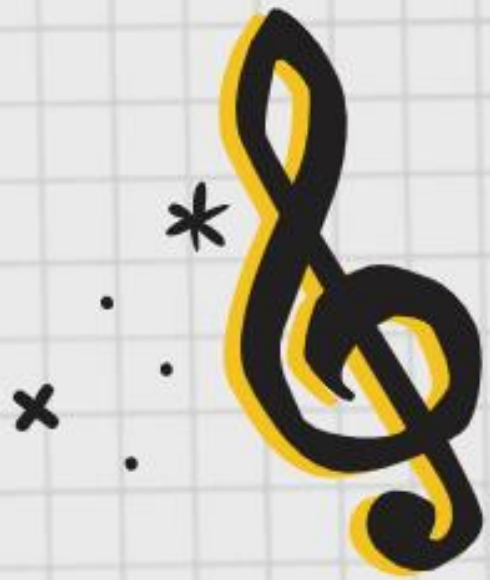
音樂推薦系統:

- 使用音頻特徵來分析歌曲的風格、節奏、音色等，從而為用戶推薦相似的音樂。
- 音樂媒體平台（如Spotify、Apple Music、YouTube）廣泛使用此技術來提供個性化的播放列表。

音頻分類與檢索:

- 自動分類音頻文件（如音樂風格分類、情感分類）。
- 在龐大的音頻資料庫中快速檢索和匹配特定類型的音頻文件。

郭政穎



PART 1-1.

FEATURE EXTRACTION

FEATURE INTRODUCTION

1.length:

音樂的長度，表示音樂文件的總持續時間。

2.chroma_stft:

表示音樂中的音高分佈，反映了十二個半音的能量分佈。

3.rms:

表示音樂訊號的能量。

郭政穎

FEATURE INTRODUCTION

4.spectral_centroid:

表示頻譜的中心位置，即頻譜質心。

5.spectral_bandwidth:

表示頻譜的寬度，反映頻率分佈的範圍。

6.rolloff:

表示頻譜能量下降到總能量的某個百分比的頻率。

7.zero_crossing_rate:

表示訊號過零點的頻率，反映訊號的振盪特性。

郭政穎

FEATURE INTRODUCTION

8.harmony:

表示音樂訊號中的和聲成分。

9.perceptr:

反映音樂訊號的感知特徵，如音調和音色。

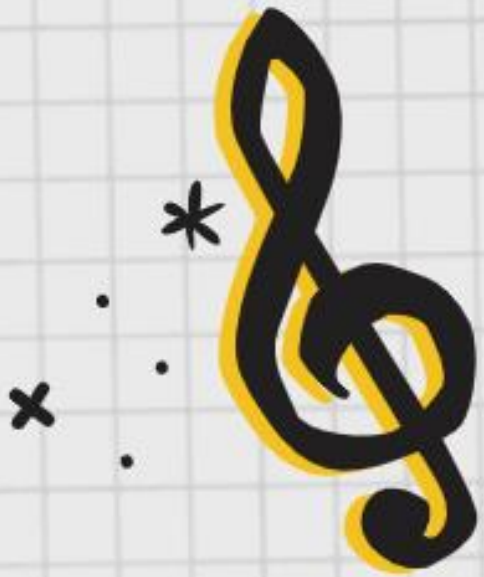
10.tempo:

表示音樂的節奏，即每分鐘的拍數（BPM）。

11.mfcc:

表示音樂訊號的倒譜特徵，廣泛用於音樂和語音分析。

郭政穎



PART 1-2.

KEY FEATURE

KEY FEATURE

MFCC calculation

Step:

1.Pre-emphasis:增強高頻成分，補償人耳感知的高頻衰減。

2.Framing:將音頻訊號分成小段，每段稱為一幀。

3.Windowing:減少訊號分段處理引入的邊界效應。

4.FFT:將時間域訊號轉換為頻域訊號。

郭政穎

KEY FEATURE

MFCC calculation

Step:

5.Mel Filter Bank:模擬人耳對不同頻率的感知，特別是對數尺度。

6.Logarithm:將能量譜轉換為對數尺度，模擬人耳對音強的感知。

7.DCT:將對數能量譜轉換為倒譜係數，去除冗餘信息。

8.Delta and Delta-Delta:捕捉訊號的動態變化。

郭政穎

KEY FEATURE

Why do we need to take so many MFCCs?

Reason:

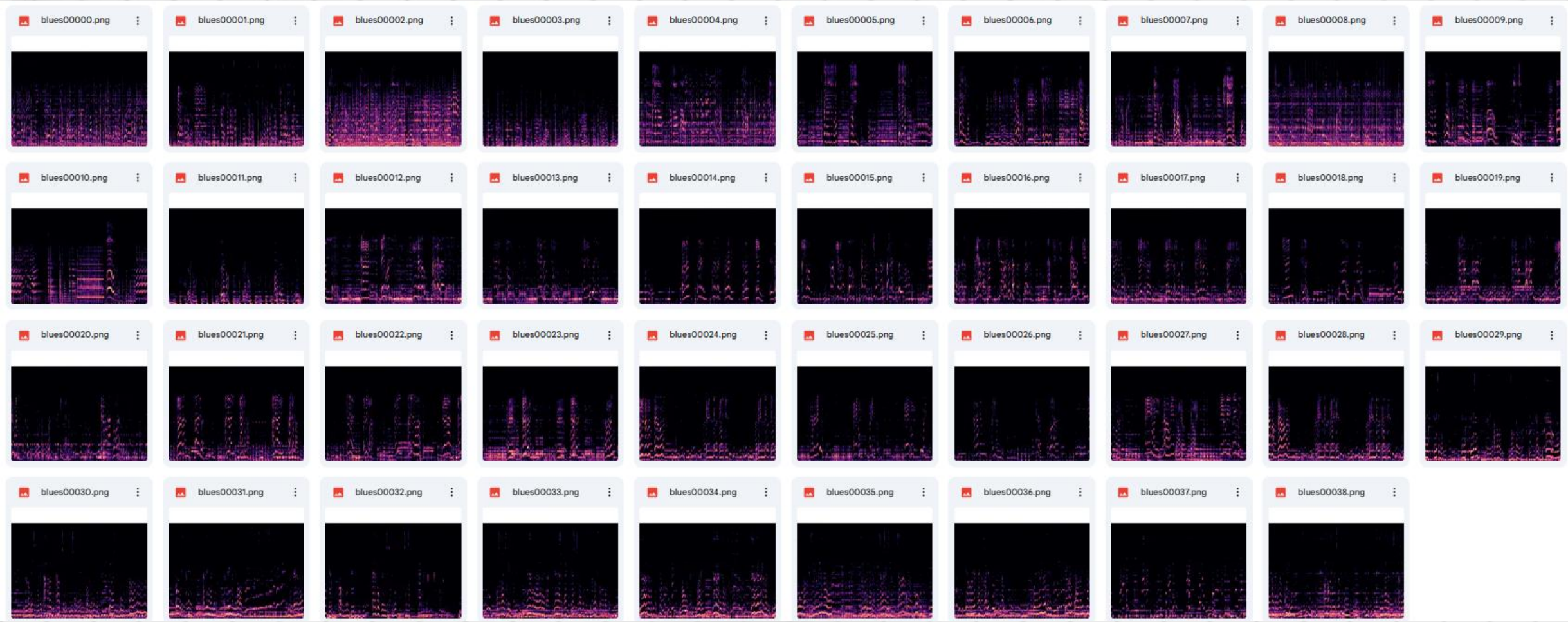
1.捕捉更多信息：每個MFCC係數代表不同的頻率特徵。
取較多的係數可以更全面地描述音頻訊號的頻率結構。

2.平衡特徵數量與計算成本：通常取13到20個係數，這樣可以在捕捉足夠信息和計算效率之間取得平衡。

3.模擬人耳聽覺系統：取較多的係數可以更好地模擬人耳對不同頻率和音調的敏感性。

郭政穎

One of the musical styles MFCC (blues)

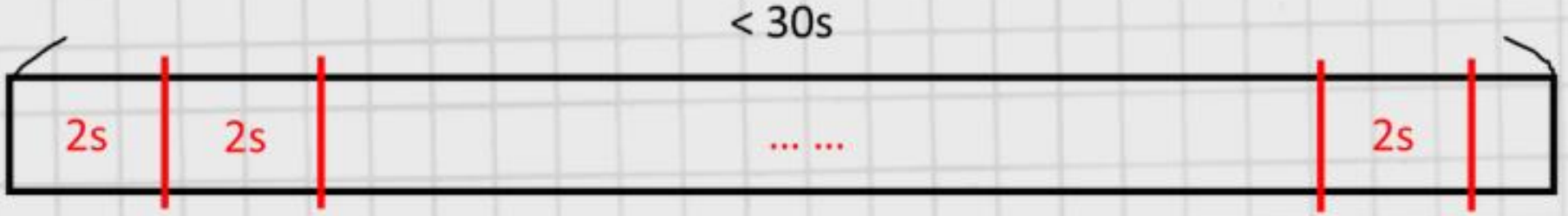


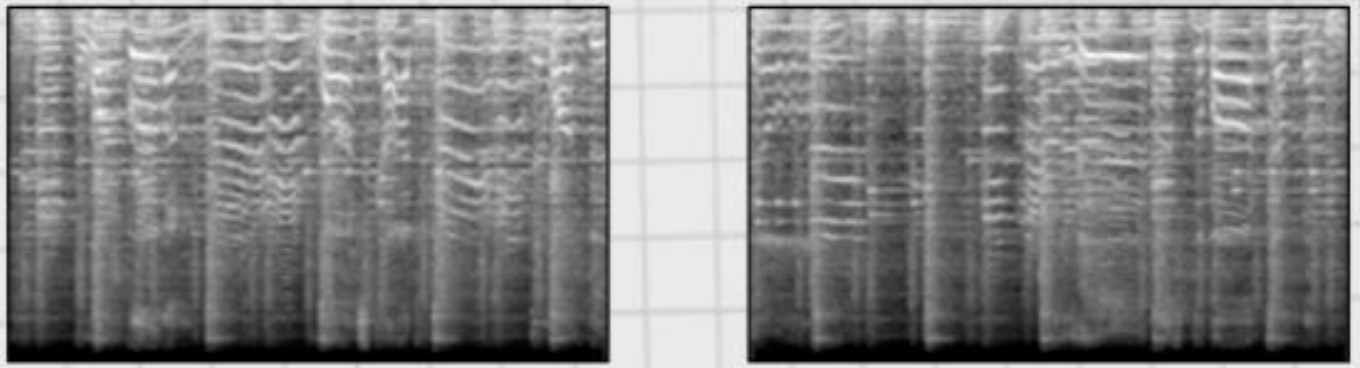
PART 2.

MELSPECTROGRAM CNN

GTZAN dataset

blues00000.wav



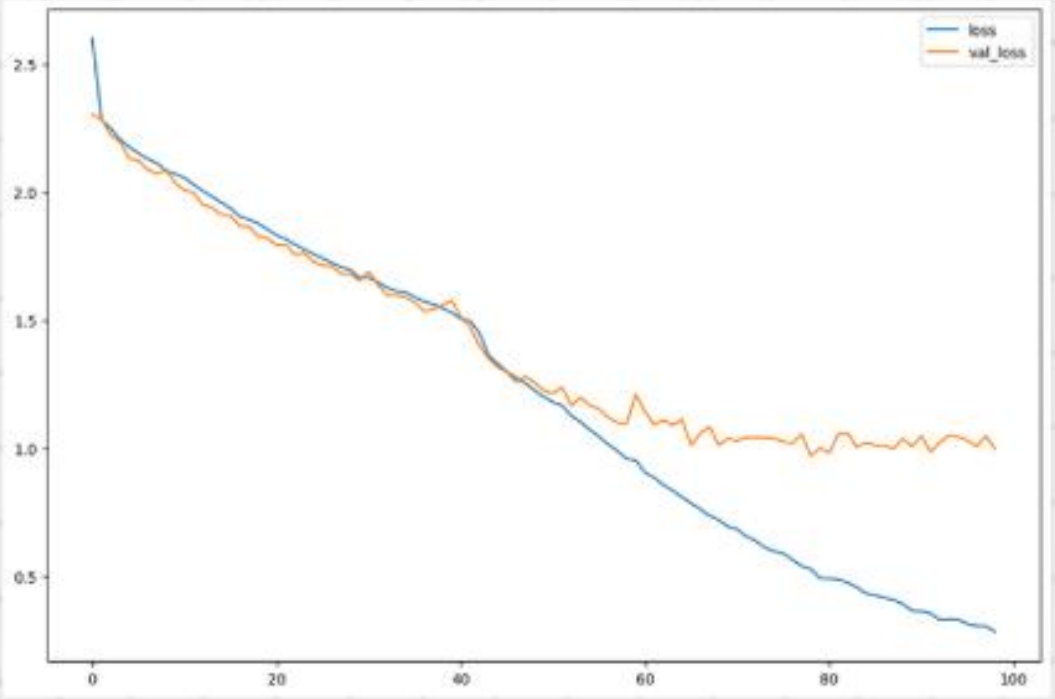


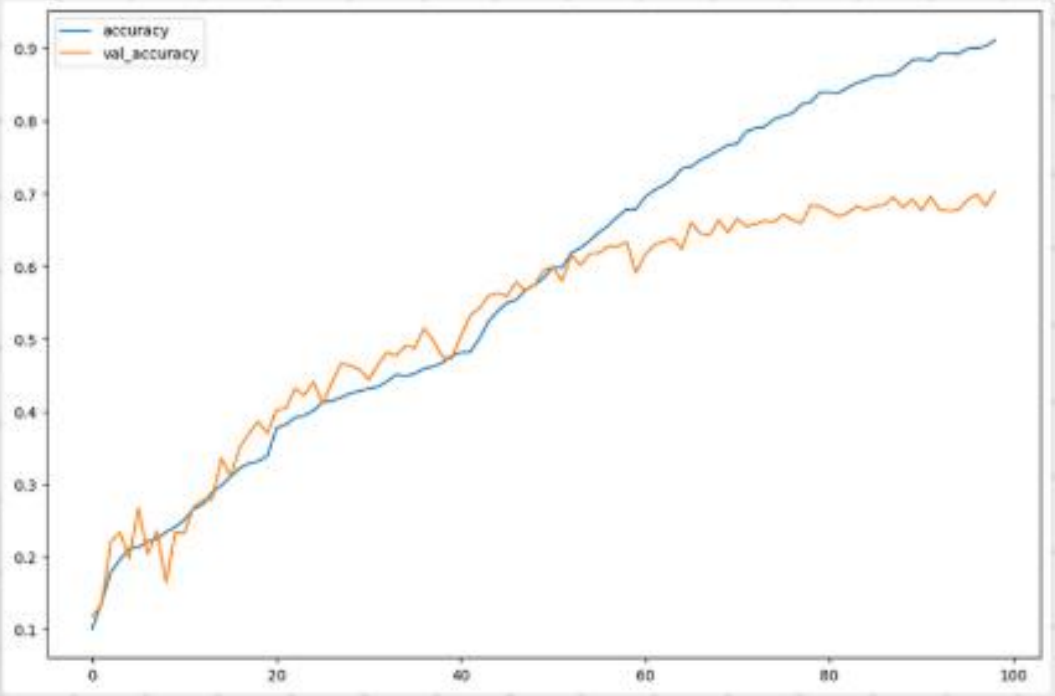
melspectrogram image

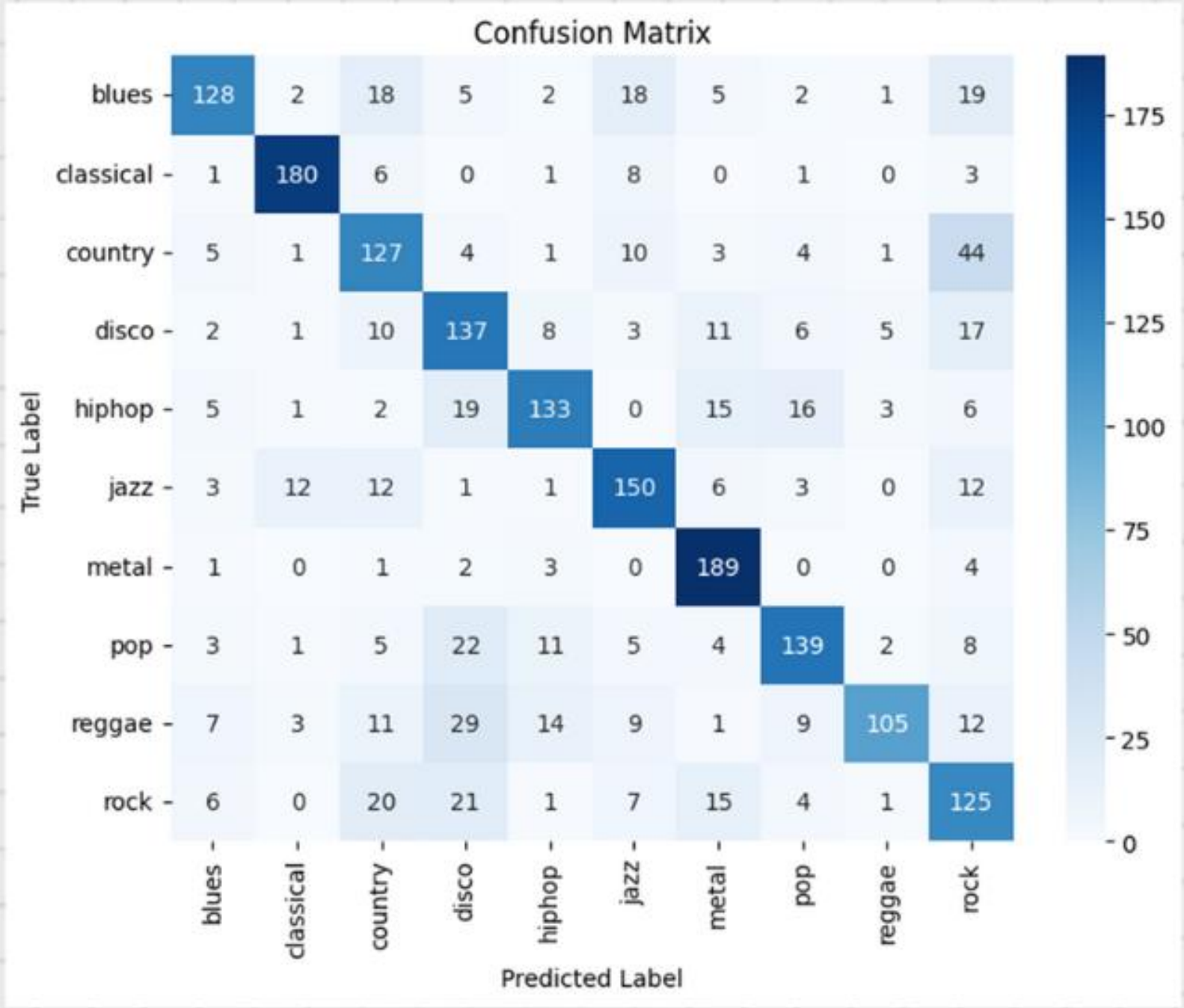
train : valid : test = 10000 : 2000 : 2000

Data 數量 = 10個類別 * 每類100個音檔 * 每個音檔分成14份 = 14000

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 124, 169, 32)	832
dropout (Dropout)	(None, 124, 169, 32)	0
max_pooling2d (MaxPooling2D)	(None, 62, 84, 32)	0
conv2d_1 (Conv2D)	(None, 58, 80, 64)	51264
dropout_1 (Dropout)	(None, 58, 80, 64)	0
max_pooling2d_1 (MaxPooling2D)	(None, 29, 40, 64)	0
conv2d_2 (Conv2D)	(None, 25, 36, 128)	204928
dropout_2 (Dropout)	(None, 25, 36, 128)	0
max_pooling2d_2 (MaxPooling2D)	(None, 12, 18, 128)	0
flatten (Flatten)	(None, 27648)	0
dense (Dense)	(None, 128)	3539072
dropout_3 (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 128)	16512
dense_2 (Dense)	(None, 10)	1290

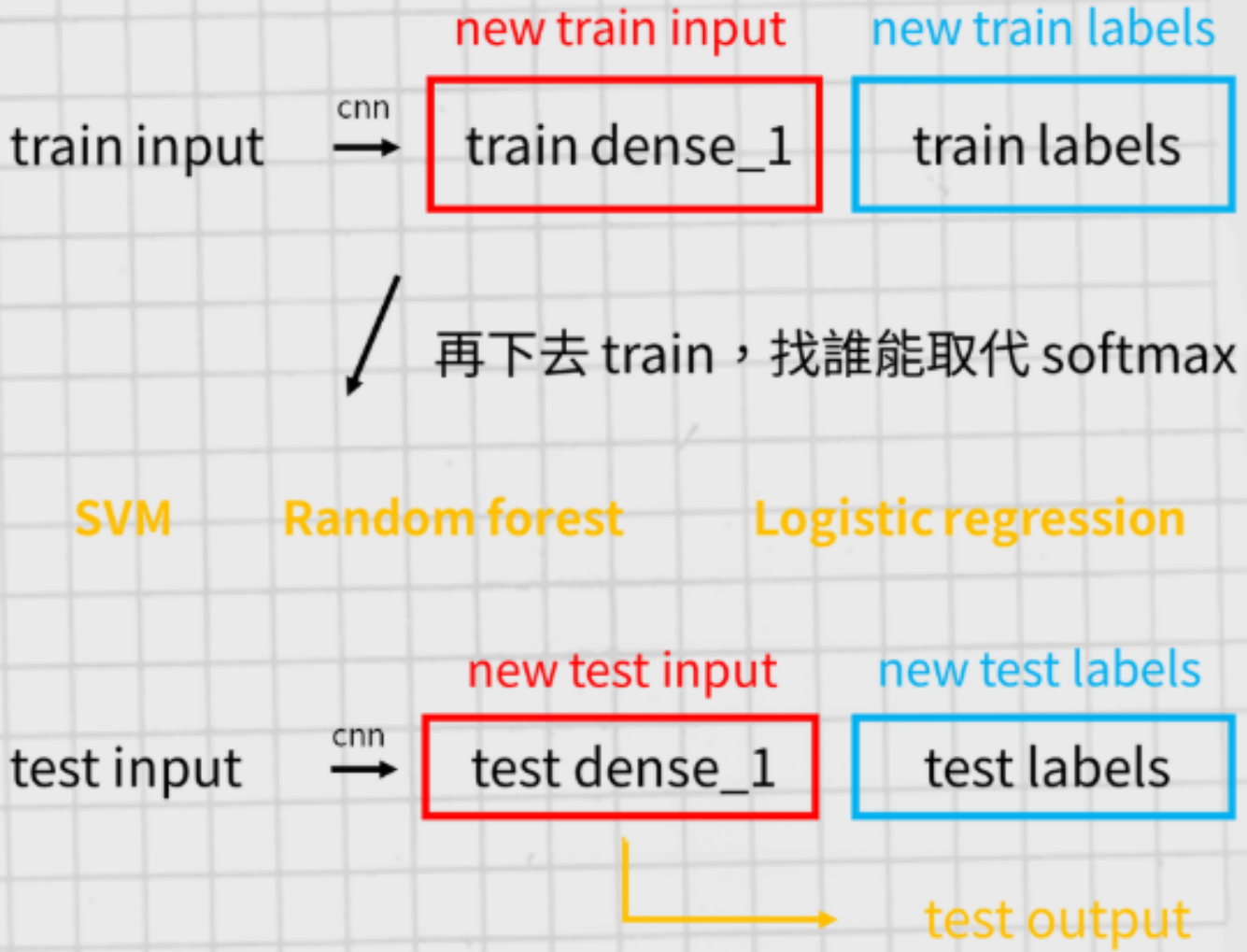






Test accuracy = 69.70%

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 124, 169, 32)	832
dropout (Dropout)	(None, 124, 169, 32)	0
max_pooling2d (MaxPooling2D)	(None, 62, 84, 32)	0
conv2d_1 (Conv2D)	(None, 58, 80, 64)	51264
dropout_1 (Dropout)	(None, 58, 80, 64)	0
max_pooling2d_1 (MaxPooling2D)	(None, 29, 40, 64)	0
conv2d_2 (Conv2D)	(None, 25, 36, 128)	204928
dropout_2 (Dropout)	(None, 25, 36, 128)	0
max_pooling2d_2 (MaxPooling2D)	(None, 12, 18, 128)	0
flatten (Flatten)	(None, 27648)	0
dense (Dense)	(None, 128)	3539072
dropout_3 (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 128)	16512
dense_2 (Dense)	(None, 10)	1290

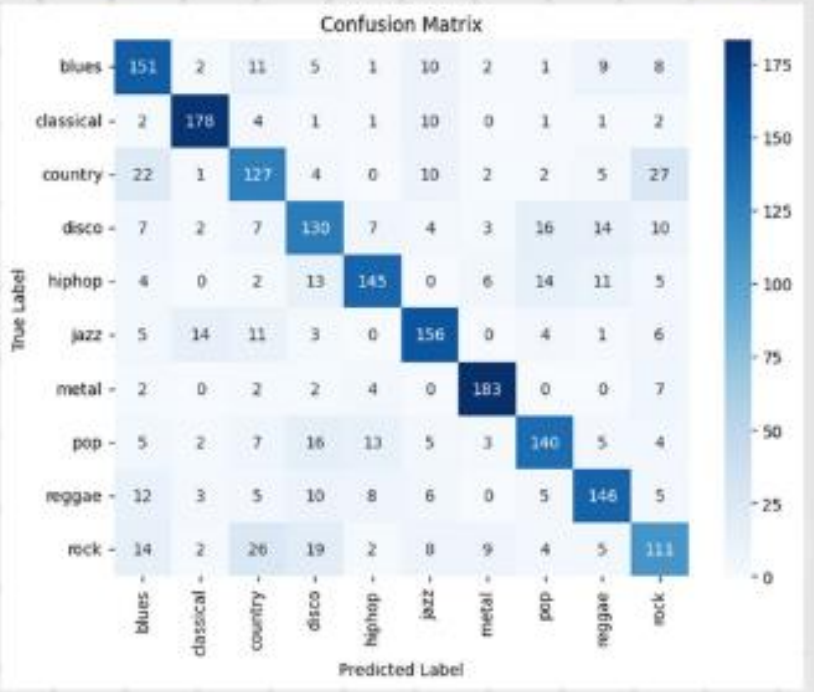


softmax

69.70%

SVM

73.35%

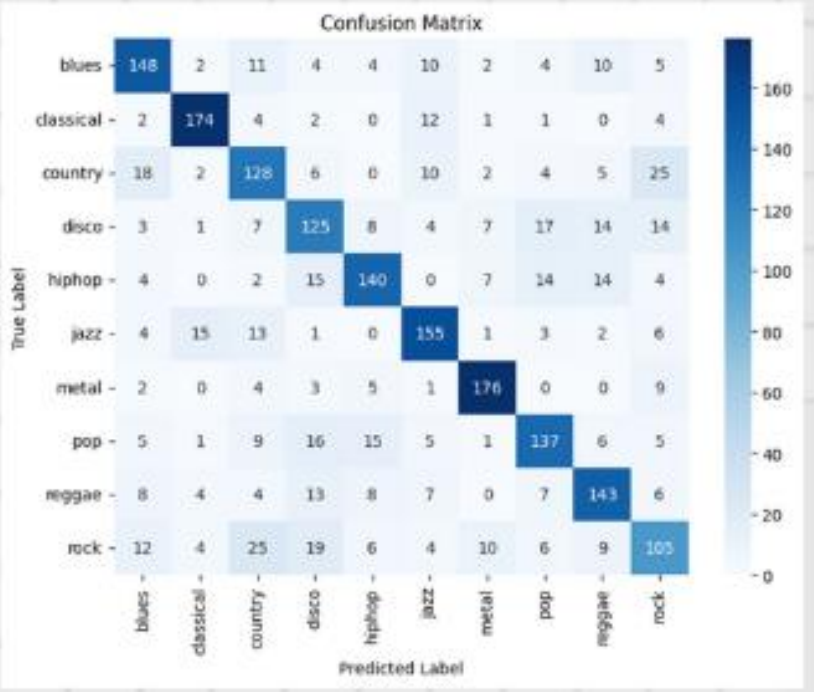


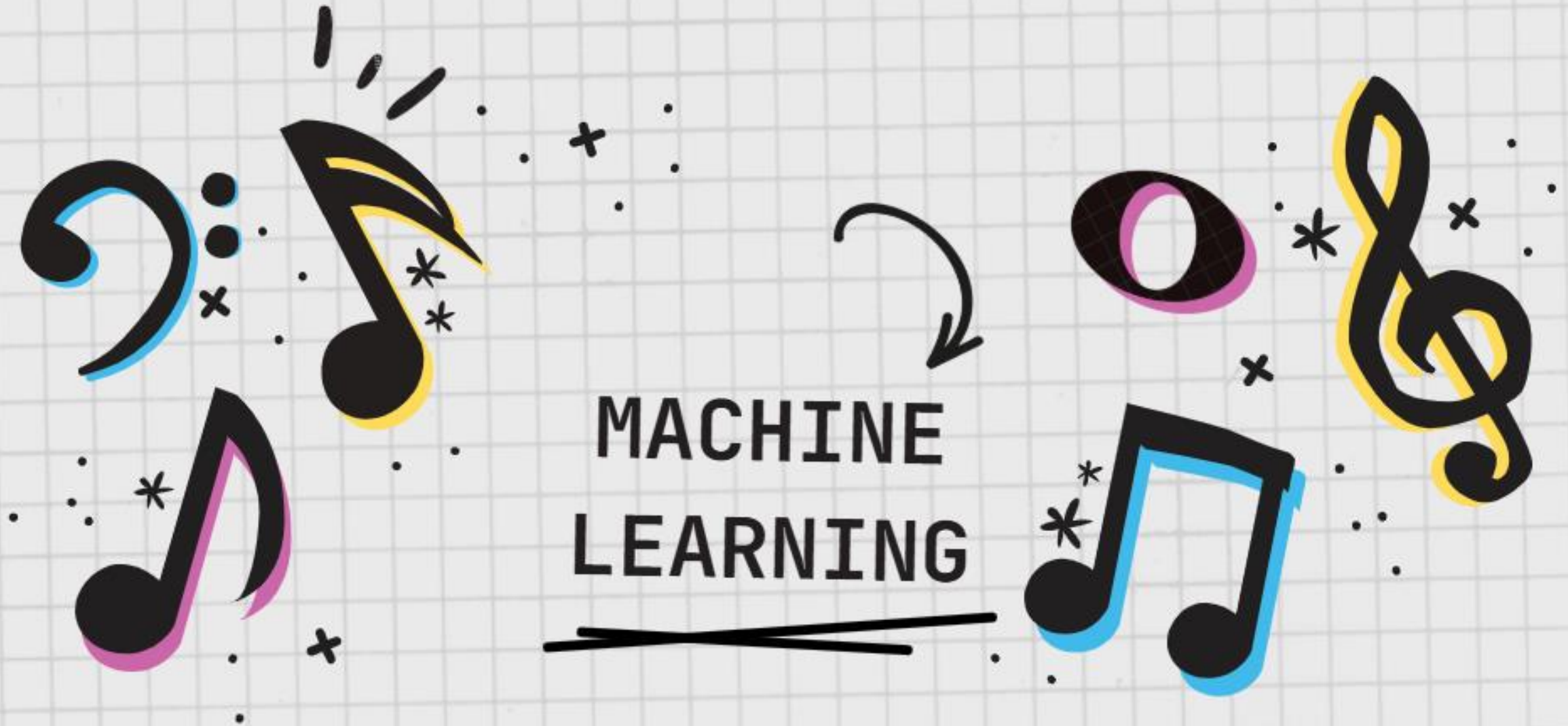
Random forest

72.25%

Logistic regression

71.55%




MACHINE
LEARNING

PART 3-1.

DATASET

DATASET

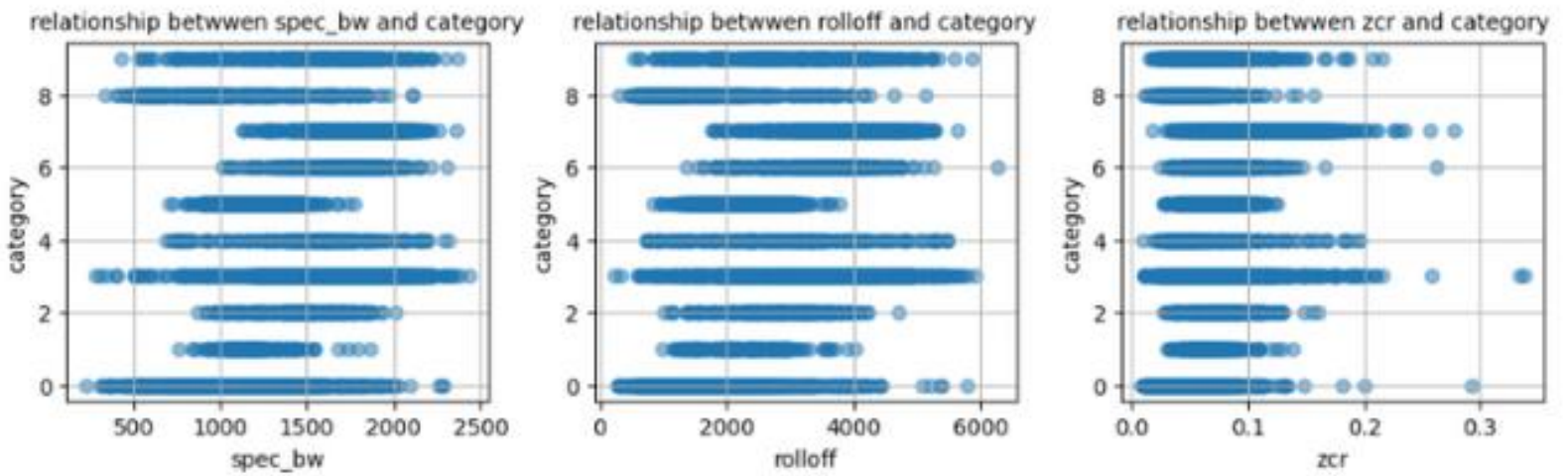
MAGNATAGATUNE

Length of Dataset	6964
Features Num	27
Class Num	10
Audio Length	30s

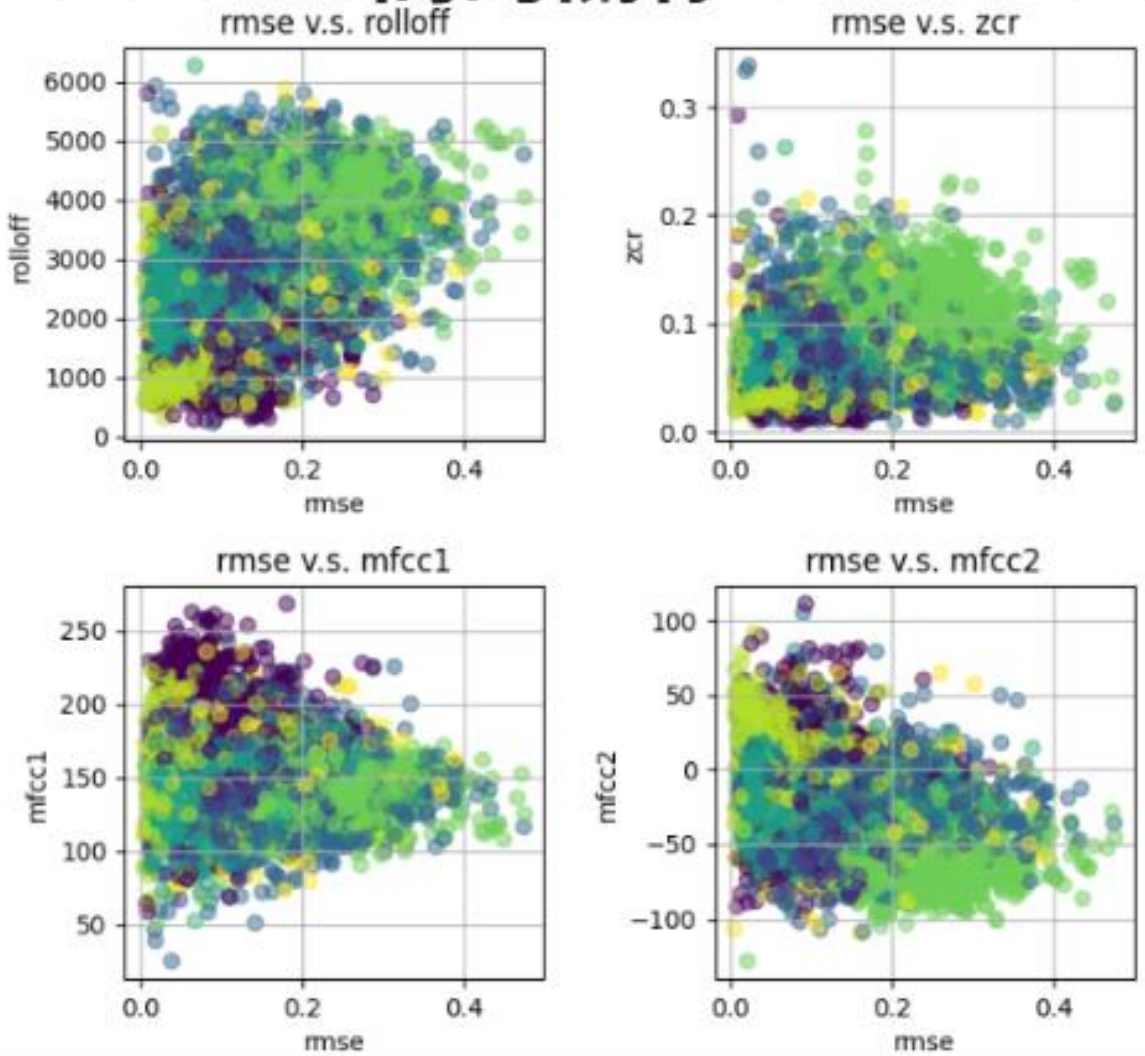
GTZAN_30

Length of Dataset	1000
Features Num	58
Class Num	10
Audio Length	30s

特徵與分類間的關係


relationship betwwen spec_bw and category
category
spec_bw
relationship betwwen rolloff and category
rolloff
relationship betwwen zcr and category
zcr
relationship betwwen tempogram and category
tempogram
relationship betwwen mfcc0 and category
mfcc0
relationship betwwen mfcc1 and category
mfcc1

特徵與分類間的關係


rmse v.s. rolloff
rolloff
rmse
rmse v.s. zcr
zcr
rmse
rmse v.s. mfcc1
mfcc1
rmse
rmse v.s. mfcc2
mfcc2
rmse

PART 3-2.

SVM

MAGNATAGATUNE

GTZAN_30

```
SVM_model = SVC(probability=True, kernel='rbf', gamma=1/58)
```

SVM Model

	precision	recall	f1-score	support
0	0.71	0.80	0.75	304
1	0.77	0.66	0.71	77
2	0.79	0.48	0.60	105
3	0.52	0.78	0.63	311
4	0.58	0.44	0.50	81
5	0.83	0.91	0.87	288
6	0.49	0.28	0.36	129
7	0.82	0.90	0.86	500
8	0.77	0.68	0.72	173
9	0.00	0.00	0.00	122
accuracy			0.71	2090
macro avg	0.63	0.59	0.60	2090
weighted avg	0.68	0.71	0.69	2090

Train Accuracy: 0.7570783750512926
Test Accuracy: 0.7124401913875598

SVM Model

	precision	recall	f1-score	support
0	0.78	0.60	0.68	35
1	0.77	1.00	0.87	20
2	0.79	0.70	0.74	37
3	0.64	0.62	0.63	34
4	0.63	0.71	0.67	24
5	0.93	0.79	0.85	33
6	0.60	0.93	0.73	30
7	0.89	0.74	0.81	23
8	0.60	0.52	0.56	29
9	0.49	0.49	0.49	35
accuracy			0.69	300
macro avg	0.71	0.71	0.70	300
weighted avg	0.71	0.69	0.69	300

Train Accuracy: 0.8971428571428571
Test Accuracy: 0.6933333333333334

- rbf: $\exp(-\gamma\|x - x'\|^2)$, where γ is specified by parameter `gamma`, must be greater than 0.

PART 3-3.

LOGISTIC REGRESSION

MAGNATAGATUNE

GTZAN_30

```
LR_model = linear_model.LogisticRegression(solver='saga', penalty='l1', max_iter=10000)
```

Logistic Regression Model

	precision	recall	f1-score	support
0	0.70	0.75	0.73	304
1	0.65	0.58	0.62	77
2	0.62	0.45	0.52	105
3	0.52	0.72	0.60	311
4	0.39	0.32	0.35	81
5	0.81	0.89	0.85	288
6	0.41	0.25	0.31	129
7	0.80	0.89	0.84	500
8	0.73	0.66	0.70	173
9	0.45	0.04	0.08	122
accuracy			0.68	2090
macro avg	0.61	0.56	0.56	2090
weighted avg	0.66	0.68	0.66	2090

Train Accuracy: 0.6879359868691014
Test Accuracy: 0.6813397129186602

Logistic Regression Model

	precision	recall	f1-score	support
0	0.62	0.71	0.67	35
1	0.95	1.00	0.98	20
2	0.68	0.68	0.68	37
3	0.75	0.71	0.73	34
4	0.65	0.71	0.68	24
5	0.93	0.85	0.89	33
6	0.70	0.93	0.80	30
7	0.86	0.78	0.82	23
8	0.64	0.55	0.59	29
9	0.57	0.46	0.51	35
accuracy			0.72	300
macro avg	0.74	0.74	0.73	300
weighted avg	0.72	0.72	0.72	300

Train Accuracy: 0.9014285714285715
Test Accuracy: 0.7233333333333334

DATASET

GTZAN_30

Length of Dataset	1000
Features Num	58
Class Num	10
Audio Length	30s

GTZAN_3

Length of Dataset	9990
Features Num	58
Class Num	10
Audio Length	3s

GTZAN_3

SVM Model

	precision	recall	f1-score	support
0	0.87	0.87	0.87	319
1	0.86	0.98	0.91	308
2	0.79	0.78	0.79	286
3	0.81	0.80	0.80	301
4	0.93	0.84	0.88	311
5	0.84	0.85	0.85	286
6	0.89	0.92	0.91	303
7	0.87	0.91	0.89	267
8	0.88	0.88	0.88	316
9	0.79	0.70	0.74	300
accuracy			0.85	2997
macro avg	0.85	0.85	0.85	2997
weighted avg	0.85	0.85	0.85	2997

Train Accuracy: 0.9121979121979122
Test Accuracy: 0.8538538538538538

SVM

Logistic Regression Model

	precision	recall	f1-score	support
0	0.75	0.73	0.74	319
1	0.89	0.94	0.92	308
2	0.64	0.64	0.64	286
3	0.65	0.63	0.64	301
4	0.73	0.69	0.71	311
5	0.78	0.81	0.80	286
6	0.83	0.84	0.84	303
7	0.75	0.87	0.80	267
8	0.70	0.66	0.68	316
9	0.56	0.52	0.54	300
accuracy			0.73	2997
macro avg	0.73	0.73	0.73	2997
weighted avg	0.73	0.73	0.73	2997

Train Accuracy: 0.7437437437437437
Test Accuracy: 0.7323990657323991

Logistic Regression

PART 3-4.

SGD

GTZAN_30

GTZAN_3

```
SGD_model = linear_model.SGDClassifier(loss='log_loss', penalty='l1', max_iter=10000, random_state=42)
```

SGD Model

	precision	recall	f1-score	support
0	0.69	0.71	0.70	35
1	0.91	1.00	0.95	20
2	0.55	0.57	0.56	37
3	0.49	0.74	0.59	34
4	0.79	0.62	0.70	24
5	0.91	0.61	0.73	33
6	0.74	0.93	0.82	30
7	0.86	0.78	0.82	23
8	0.56	0.52	0.54	29
9	0.42	0.31	0.36	35
accuracy			0.66	300
macro avg	0.69	0.68	0.68	300
weighted avg	0.67	0.66	0.66	300

Train Accuracy: 0.8985714285714286
Test Accuracy: 0.66

SGD Model

	precision	recall	f1-score	support
0	0.64	0.67	0.66	319
1	0.92	0.92	0.92	308
2	0.51	0.62	0.56	286
3	0.50	0.61	0.55	301
4	0.62	0.67	0.64	311
5	0.76	0.70	0.73	286
6	0.84	0.79	0.81	303
7	0.74	0.79	0.76	267
8	0.46	0.39	0.42	316
9	0.30	0.20	0.24	300
accuracy			0.63	2997
macro avg	0.63	0.64	0.63	2997
weighted avg	0.63	0.63	0.63	2997

Train Accuracy: 0.6582296582296582
Test Accuracy: 0.6349683016349683


PART 3-5.
RANDOM FOREST

GTZAN_30

GTZAN_3

```
RF_model = RandomForestClassifier(criterion='entropy')
```

Random Forest Model

	precision	recall	f1-score	support
0	0.75	0.69	0.72	35
1	0.83	0.95	0.88	20
2	0.69	0.68	0.68	37
3	0.85	0.68	0.75	34
4	0.67	0.83	0.74	24
5	0.84	0.82	0.83	33
6	0.72	0.97	0.83	30
7	0.81	0.91	0.86	23
8	0.70	0.66	0.68	29
9	0.70	0.54	0.61	35
accuracy			0.75	300
macro avg	0.76	0.77	0.76	300
weighted avg	0.76	0.75	0.75	300

Train Accuracy: 1.0
Test Accuracy: 0.7533333333333333

Random Forest Model

	precision	recall	f1-score	support
0	0.89	0.86	0.87	208
1	0.93	0.98	0.95	203
2	0.76	0.82	0.79	186
3	0.83	0.83	0.83	199
4	0.93	0.86	0.90	218
5	0.83	0.91	0.87	192
6	0.89	0.96	0.93	204
7	0.92	0.92	0.92	180
8	0.89	0.88	0.88	211
9	0.88	0.73	0.79	197
accuracy			0.87	1998
macro avg	0.87	0.87	0.87	1998
weighted avg	0.88	0.87	0.87	1998

Train Accuracy: 0.9992492492492493
Test Accuracy: 0.8743743743743744

PART 3-6.

DEMO

PART 3-.

CONCLUSION


AND THAT'S A WRAP.
THANKS!