

# 商管機器學習 Final Project

---

## Group 2

統計碩二 張和家、統計碩一 楊博儒  
工管四 呂啟睿、工管三 余羿辰

# Agenda



01

Introduction

02

EDA

03

Building Models -  
Thoughts & Procedure

04

Difficulties

05

Future  
Application

# Introduction



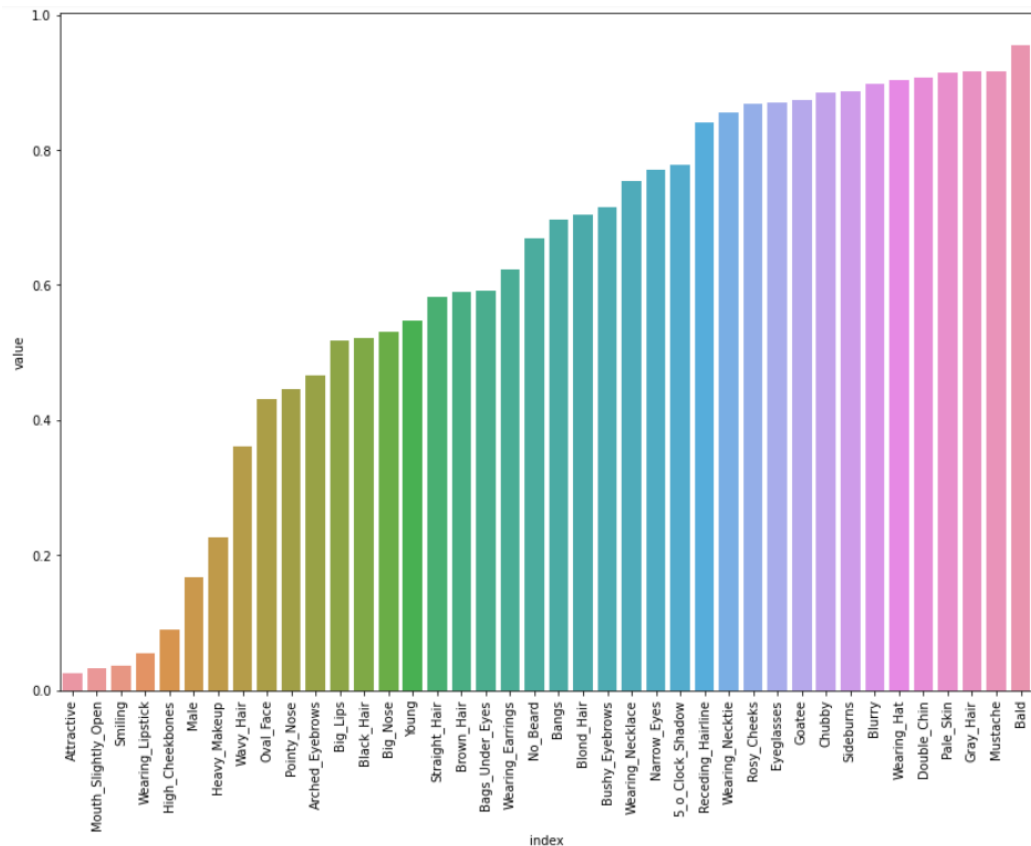
- 本組針對交友軟體之配對方式提出可改變之設計
  1. 現在→過往資料分析 or 手動輸入
  2. 未來→臉部特徵分群
    - 清晰辨認人臉特徵
    - 個人化臉部特徵推薦系統
- **Final Project** 內容：利用 **CNN** 訓練模型，來預測圖片是否擁有分類之特徵，並透過數據強化、**hidden layers** 設計等方法來提高預測準確程度。



EDA



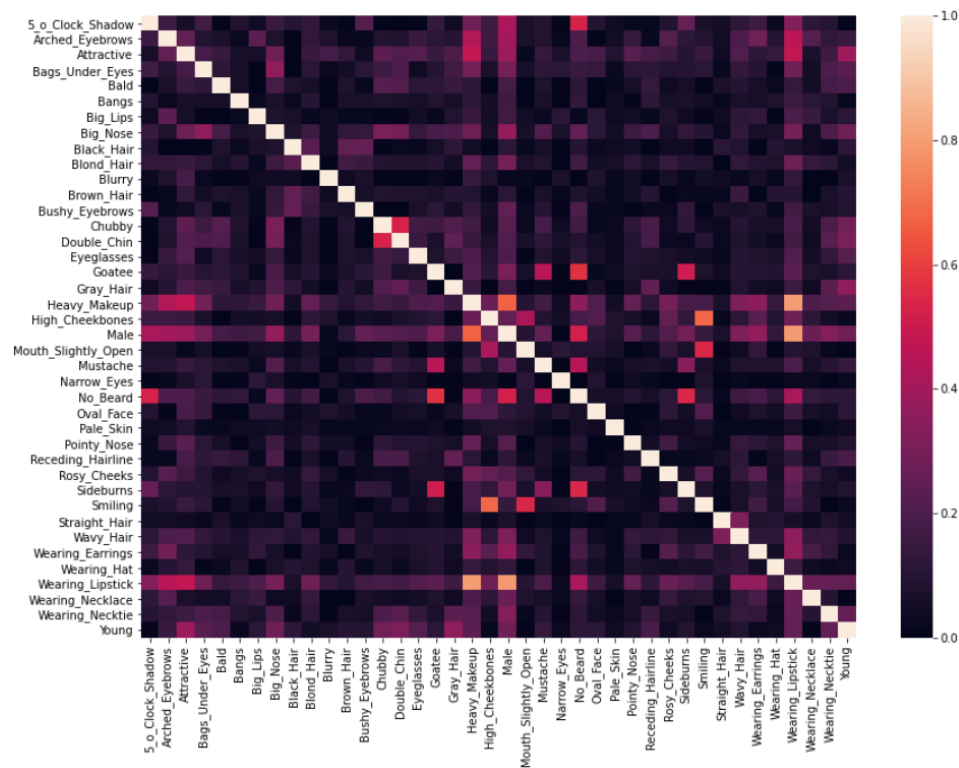
# EDA- Features Value



**Value:** 利用有無特徵之二值相減後取絕對值再除以資料總長

數量差距情形每個 feature 表現皆不同，features 如 Bald、Mustache 等差距較大；Attractive、Smiling 等則較小

# EDA- Heatmap

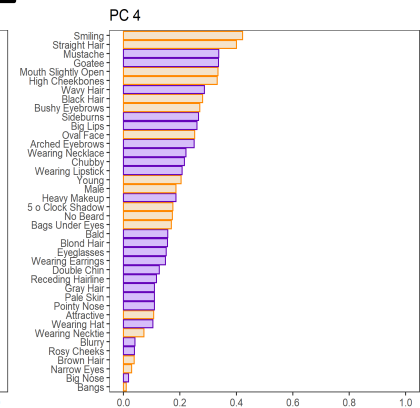
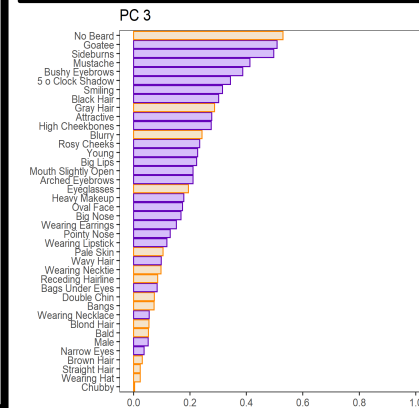
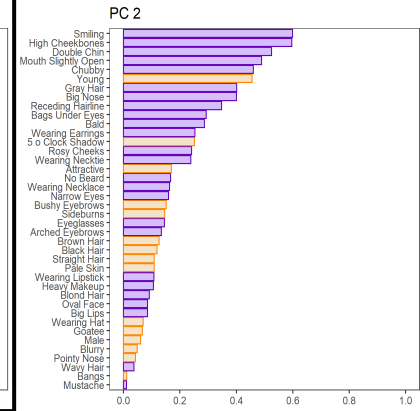
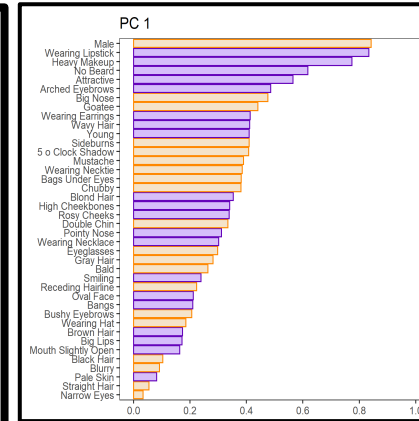
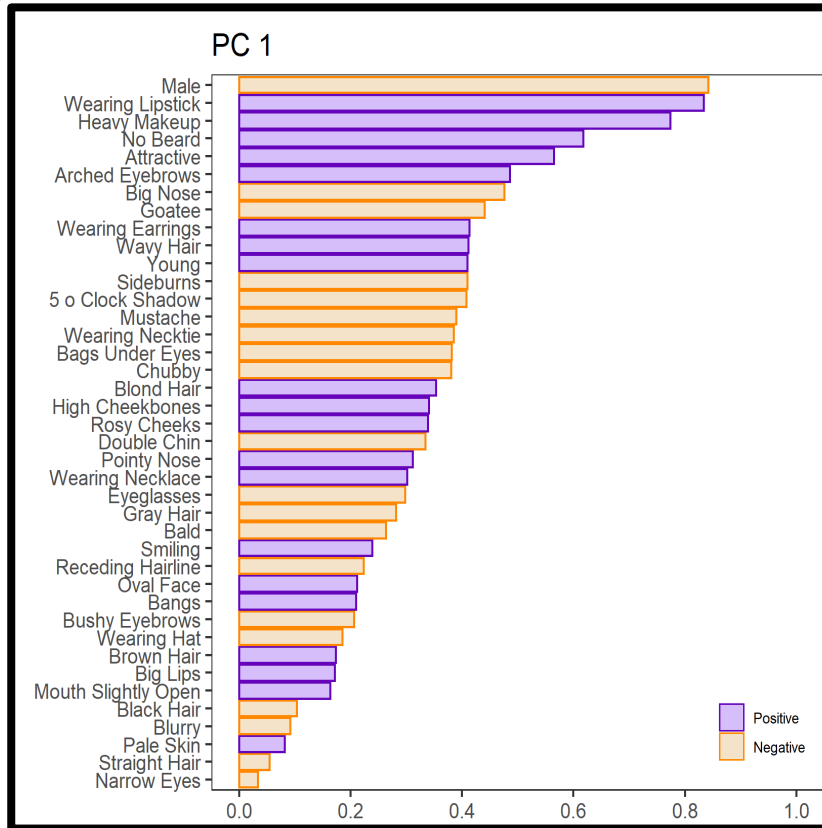


Male 0.67  
 Wearing\_Lipstick 0.80  
 Name: Heavy\_Makeup, dtype: float64

Wearing\_Lipstick 0.79  
 Name: Male, dtype: float64

Smiling 0.68  
 Name: High\_Cheekbones, dtype: float64

# EDA- PCA



# EDA- PCA

Factor	Attribute	Loading	Eigenvalue (Cumulative)
Feminine	Male	-0.842	6.078 (15.19%)
	Wearing Lipstick	0.834	
	Heavy Makeup	0.774	
	No Beard	0.617	
	Attractive	0.566	
Smiling	Smiling	0.599	2.814 (22.23%)
	High Cheekbones	0.597	
	Double Chin	0.525	
	Mouth Slightly Open	0.489	
	Chubby	0.460	
	Young	-0.456	
Bearded	No Beard	-0.530	2.208 (27.75%)
	Goatee	0.509	
	Sideburns	0.497	
Masculine	Smiling	-0.422	1.820 (32.3%)
	Straight Hair	-0.400	
	Mustache	0.337	
	Goatee	0.337	
	Mouth Slightly Open	-0.335	
	High Cheekbones	-0.332	

## 臉部特徵共同性

1. 男性與口紅呈高度負相關
2. 獨特的臉部特徵組合

## 認識用戶行為

1. 選擇外貌的動機
2. 用戶習性分類





# Building Models- Thoughts & Procedures

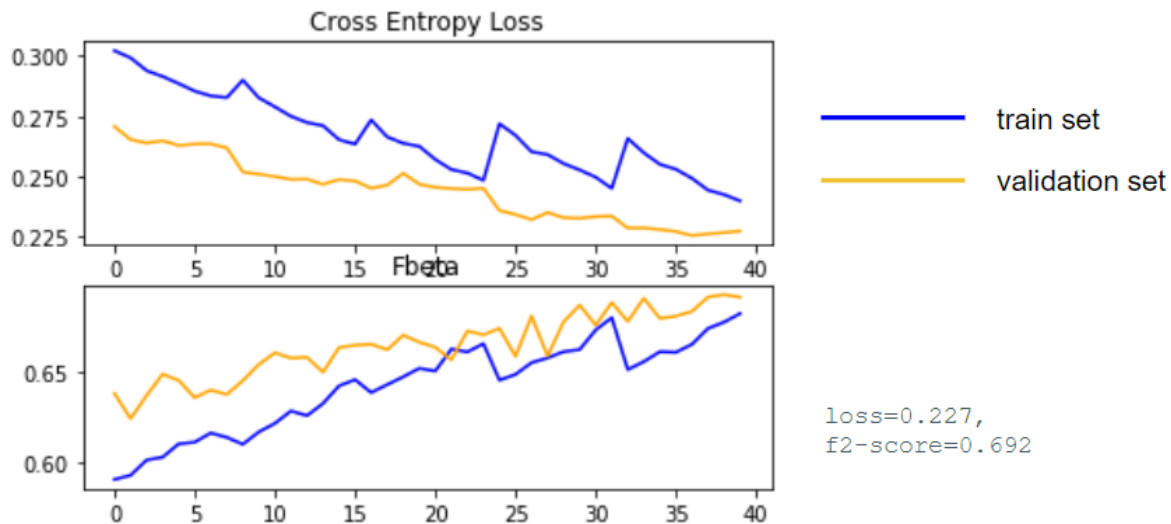


# Multiple Features: 建模

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 128, 128, 32)	896
conv2d_1 (Conv2D)	(None, 128, 128, 32)	9248
max_pooling2d (MaxPooling2D)	(None, 64, 64, 32)	0
dropout (Dropout)	(None, 64, 64, 32)	0
conv2d_2 (Conv2D)	(None, 64, 64, 64)	18496
conv2d_3 (Conv2D)	(None, 64, 64, 64)	36928
max_pooling2d_1 (MaxPooling2D)	(None, 32, 32, 64)	0
dropout_1 (Dropout)	(None, 32, 32, 64)	0
conv2d_4 (Conv2D)	(None, 32, 32, 128)	73856
conv2d_5 (Conv2D)	(None, 32, 32, 128)	147584
max_pooling2d_2 (MaxPooling2D)	(None, 16, 16, 128)	0
dropout_2 (Dropout)	(None, 16, 16, 128)	0
flatten (Flatten)	(None, 32768)	0
dense (Dense)	(None, 128)	4194432
dropout_3 (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 40)	5160

- Optimizer → SGD
- Loss Function → Binary Cross Entropy
- Metric → F2-score

# Multiple Features: 結果



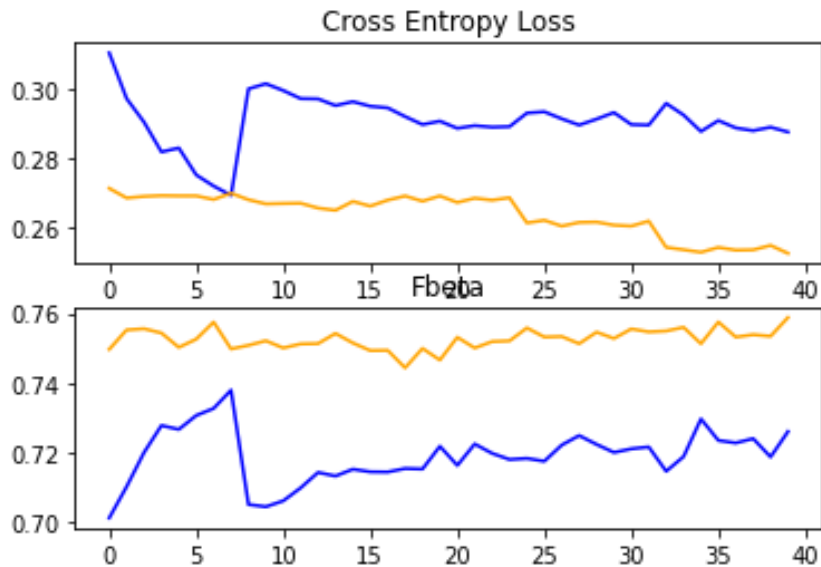
無 overfitting

F2 score 為0.692

Loss 為0.227

# Multiple Features: 結果

(with f-0.5-score)



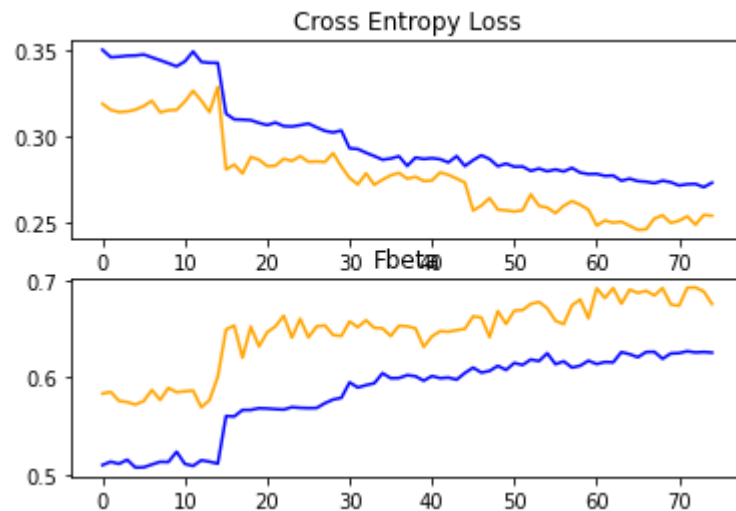
loss=0.252, f-0.5-score=0.759

- Optimizer: **SGD** → **Adam**
- Metric: **F2-score** → **F-0.5-score**  
(more emphasis on precision)

— train set  
— validation set

# Multiple Features: Data Augmentation (with f2-score)

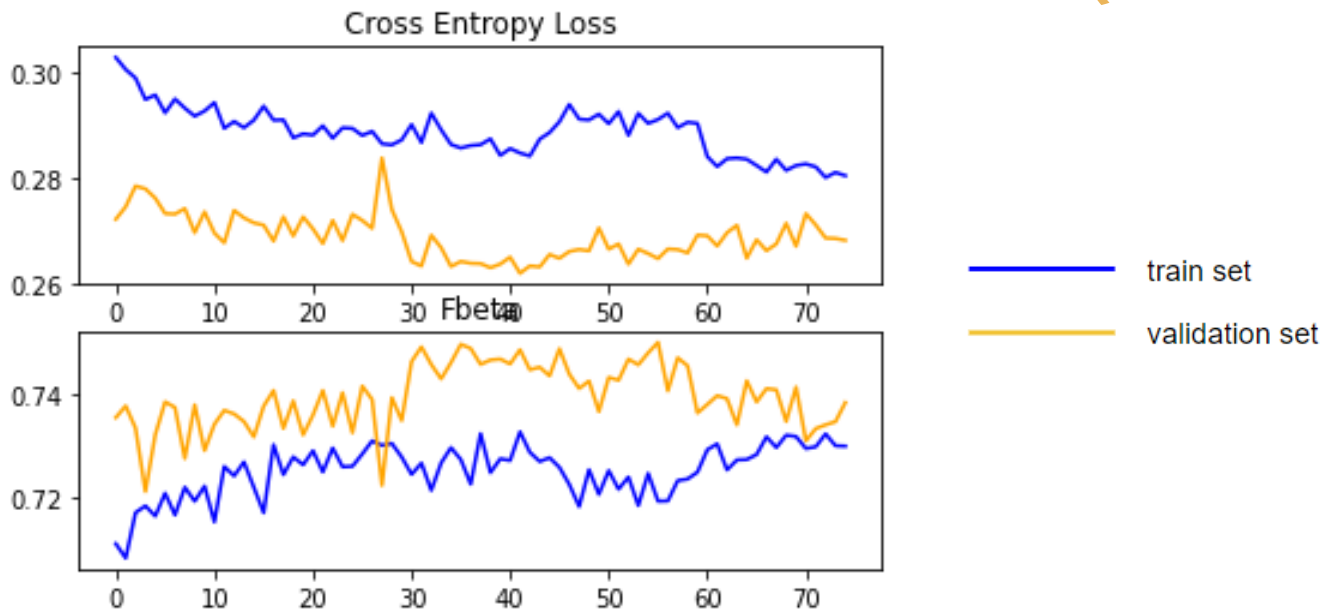
```
train_datagen = ImageDataGenerator(  
    rescale=1./255,  
    rotation_range=40,  
    width_shift_range=0.2,  
    height_shift_range=0.2,  
    shear_range=0.2,  
    zoom_range=0.2,  
    horizontal_flip=True)
```



loss=0.254, f2-score=0.676

— train set  
— validation set

# Multiple Features: Data Augmentation (with f-0.5-score)



loss=0.268, f-0.5-score=0.738

# Multiple Features: Conclusion

1. 模型經過訓練後 f2-score 有所成長，f-0.5-score 變動不大  
→ 經過訓練後 recall rate 提高，precision 變動不大
2. f score 無法對模型表現提供直覺的資訊
3. 無法針對個別特徵的特殊性處理(e.g. 某些特徵特別重要或類別數量不均衡)
4. 仍有針對單一特徵處理的必要性

# Single Feature: 建模邏輯

## 01 動機

1. 耗時、佔記憶體
2. 高度受限於電腦配備
3. 抽樣可代表母體

## 02 目的

1. 尋找卷積層與全連接層最佳的搭配
2. 調整模型的線索

## 03 驗證

1. 候選模型排名不售抽樣大小影響
2. 排名來自檢定候選模行的表現

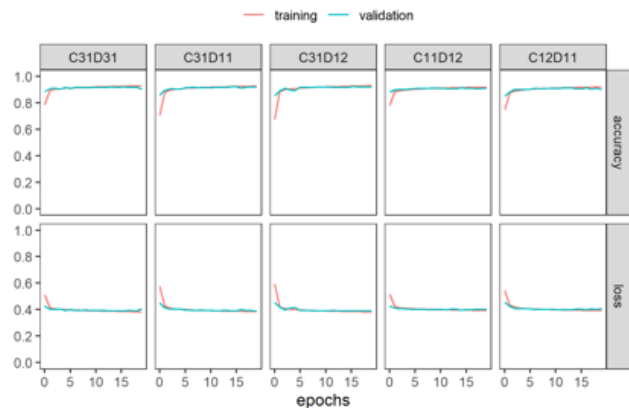
## 04 策略

1. 6種卷積層
2. 6種全連接層
3. 不同樣本大小

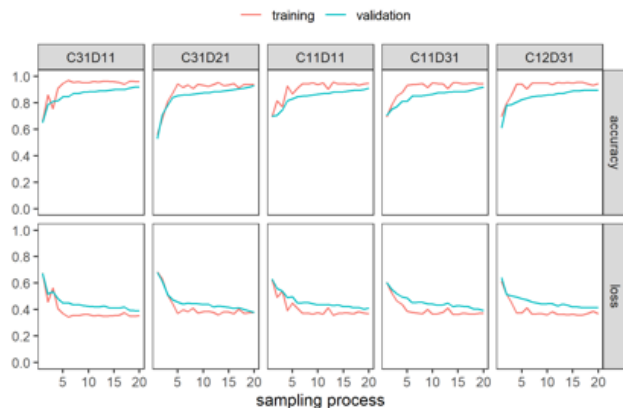


# Single Feature: Wearing Lipstick

200K sample for each epoch



Training 2K sample in each process

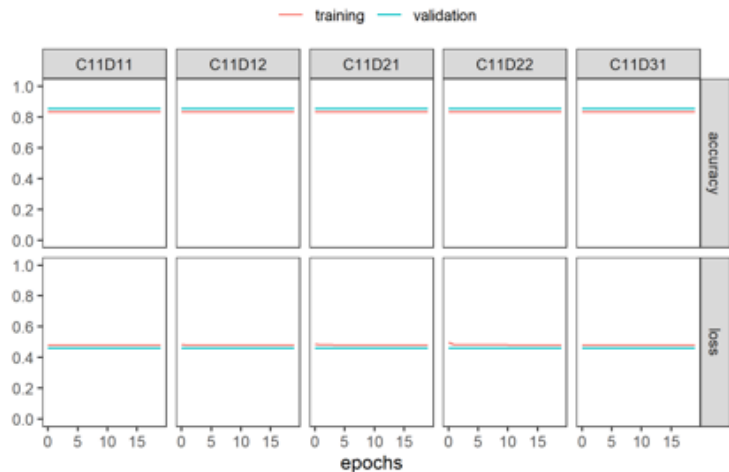


Wearing lipstick: 47.24%  
Otherwise: 56.76%

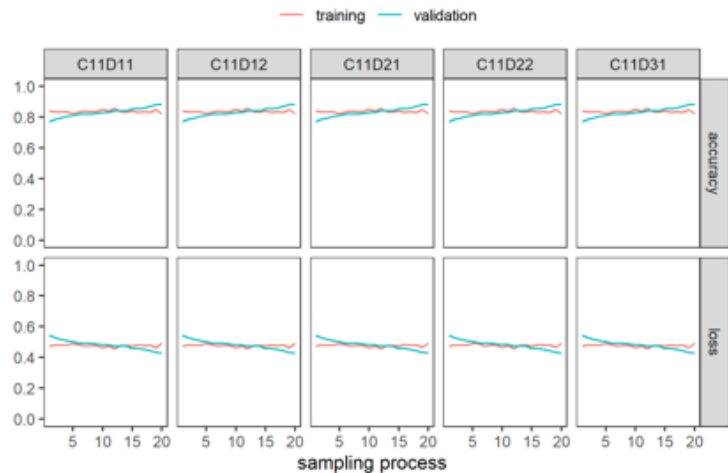
Model	C31D31	C31D11	C31D12	C11D12	C12D11	C11D31	C12D12	C12D21	C12D31	C31D21	C11D11	C31D22
200K	1	2	3	4	5	6	6	6	6	6	7	8
2K	4	1	4	5	6	3	12	7	3	2	3	11
Model	C11D32	C31D32	C11D21	C12D32	C32D11	C32D31	C12D22	C22D11	C22D31	C21D31	C21D12	C32D12
200K	9	9	10	10	11	11	12	12	12	13	14	14
2K	11	3	7	8	11	11	15	13	15	9	11	14
Model	C11D22	C21D11	C21D21	C21D32	C22D12	C22D21	C22D32	C32D21	C32D22	C32D32	C21D22	C22D22
200K	15	15	15	15	15	15	15	15	15	15	16	16
2K	10	8	16	13	11	17	19	17	20	17	18	21

# Single Feature: No Beard

200K sample for each epoch



Training 2K sample in each process



No beard: 83.49%

Otherwise: 16.51%



# Single Feature: Conclusion

1. 候選模型皆無法有效處理佔比懸殊的情況
  2. 不同特徵可能適用不同的 CNN 架構
  3. 適當的抽樣方法有助於預測候選模型表現
  4. 調整抽樣方法，使抽樣的結構與原始樣本結構吻合
  5. 調整假設檢定的方法或檢定的變數
-



# Difficulties



# Difficulties



1. Time Consuming



2. Memory Consuming



3. Highly restricted  
to PC's specification



# Future Application

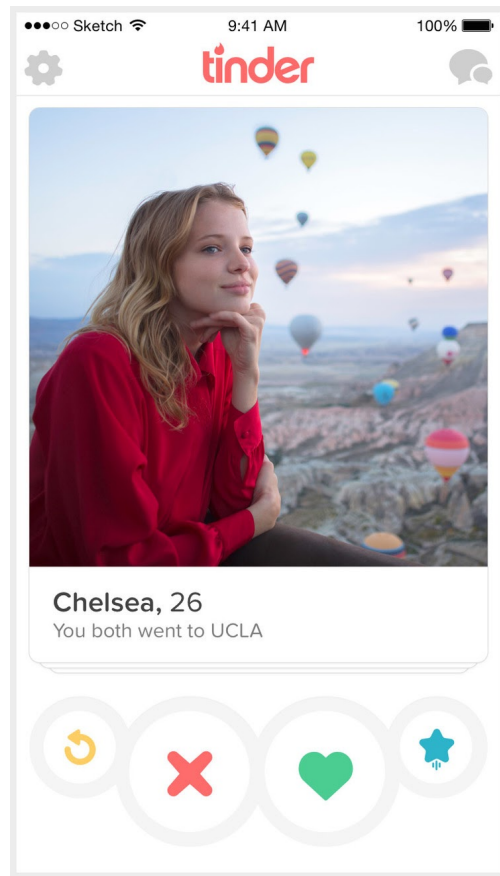
# Future Application

## 1. 交友軟體的蓬勃

→ 使用人數遽增，成為生活的一部分

## 2. 外貌為主要依據

- 網路交友大多以貌取人
- 只能對興趣、地區、學校、年齡做分類
- 需要花非常多時間挑選對象
- 花錢成為會員誘因過小



# Future Application



1. 利用特徵篩選選擇，協助找到相符合特徵的配對對象推薦
2. 推薦機制
  - 相符特徵配對
  - 推薦用戶給他可能有興趣的對象
3. 新用戶
  - 自訂喜好特徵條件
  - 用戶面相特徵推廣指定客群
  - 物以類聚，人以群分
4. 用戶分級
  - 針對普通用戶，藉由其所屬市場區隔做為推薦機制，並提供每日10次的對象推薦體驗
  - 針對高級用戶，提供個人化的推薦機制，並且推廣他感興趣的用戶





# Thank You

---