

# Predicting flight delays

學號：10446021

姓名：黃于真

# 摘要

航班延誤造成航空公司及乘客困擾。在商業航空所有參與者的決策過程中，預測對他們至關重要。

此外，由於航空運輸系統的複雜性、預測方法的數量以及飛行數據的氾濫，開發航班延誤的準確預測模型變得麻煩。

在此背景下，運用機器學習算法預測特定航班的到達是否會延遲，描述航班延誤預測問題與解決這些問題的研究趨勢關係。

# 介紹

在過去的二十年裡，航空旅行在旅行者中越來越受歡迎，因為它具有高度便利性但這也導致空中交通的顯著增長。

空中交通的增長導致大量飛機在地面和空中徘徊延誤。這些延誤造成巨大的經濟和環境損失。在應對日益增長的燃料排放問題及健康的負面影響，有一個積極的航空業研究，尋找準確預測航班延誤的技術，以優化航班營運並盡量減少延誤。

算法為輸入特徵像是出發日期、出發延遲時間、兩個機場之間的距離、預定抵達時間，我們使用機器語言進行預測：如果航班到達將延遲或不延遲。

且美國聯邦航空局認為，當預定到達時間與實際到達時間不同時，航班會延誤超過15分鐘。

# 資料集介紹

- flight (day, day of the week, airline, flight number, tail number)
- origin and destination (origin airport, destination airport)
- the departure (scheduled departure, departure time, departure delay, taxi-out, wheels-off)
- the flight-journey (scheduled time, elapsed time, air time, distance)
- the arrival (wheels-on, taxi-in, scheduled arrival, arrival time, arrival delay)
- diversion, cancellation and reason of delay (air system delay, security delay, airline delay, late aircraft delay, weather delay)

# 資料預處理

## 2. 原始資料集預處理

```
In [9]: df_AIRLINE = pd.get_dummies(df['AIRLINE'])
df_AIRLINE.head()
```

Out[9]:

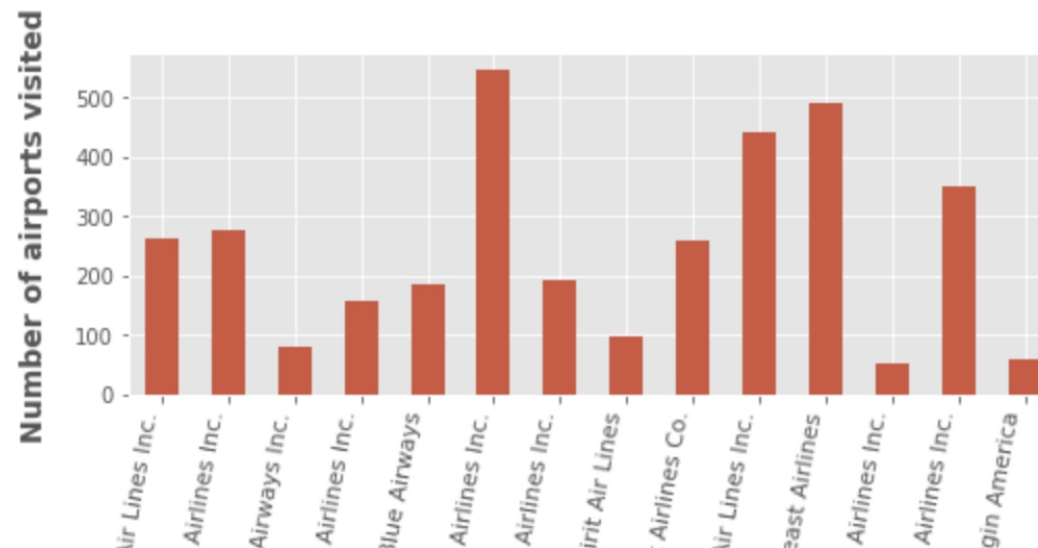
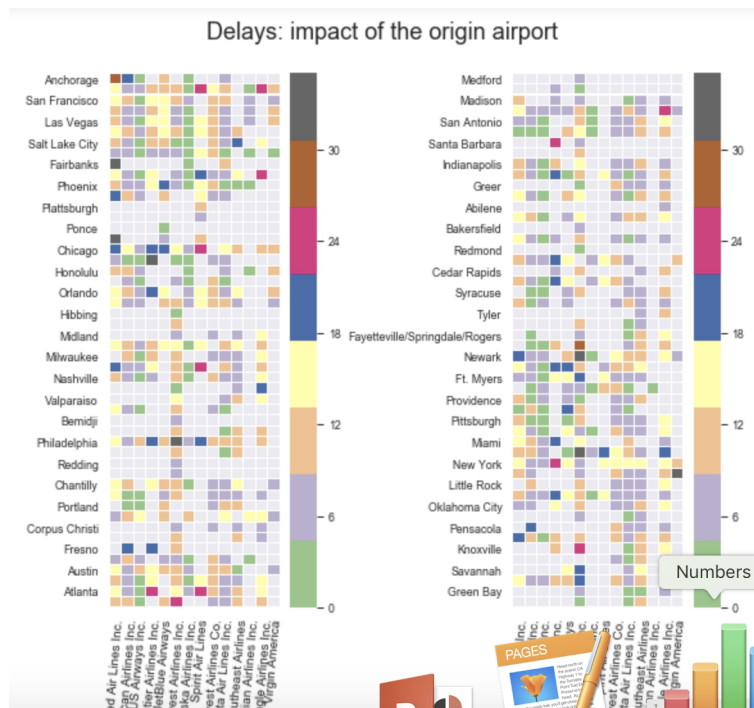
	AA	AS	B6	DL	EV	F9	HA	MQ	NK	OO	UA	US	VX	WN
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	1	0	0
3	1	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	1	0	0	0	0	0	0	0	0	0	0	0	0

```
In [28]: df2 = df[['ORIGIN_AIRPORT', 'DESTINATION_AIRPORT', 'SCHEDULED_DEPARTURE', 'DEPARTURE_TIME', 'DEPARTURE_DELAY', 'SCHEDULED_ARRIVAL',
'SCHEDULED_TIME', 'ELAPSED_TIME', 'DIVERTED', 'CANCELLED']]
df2.head()
```

Out[28]:

	ORIGIN_AIRPORT	DESTINATION_AIRPORT	SCHEDULED_DEPARTURE	DEPARTURE_TIME	DEPARTURE_DELAY	SCHEDULED_ARRIVAL	ARRIVAL_TIME	ARRIV
0	ANC	SEA	5	2354.0	-11.0	430	408.0	LINE
1	LAX	PBI	10	2.0	-8.0	750	741.0	
2	SFO	CLT	20	18.0	-2.0	8	8	
3	LAX	MIA	20	15.0	-5.0	75	75	

# 機器學習



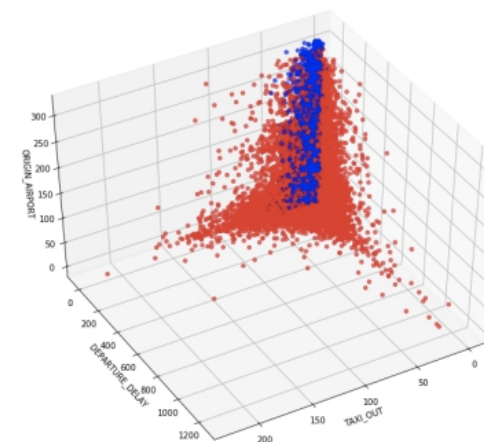
# 研究結果及討論

- precision
- accuracy
- recall

Features	Importance Score
DEPARTURE_DELAY	0.8478
TAXI_OUT	0.1438
ORIGIN_AIRPORT	0.0031

透過決策樹預測模型

藍點表示航班按時到達，紅點表示航班延誤。



# 結論

- 應用機器學習來預測飛機的延誤及取消
- 決策樹可以預測航班到達時間會延遲或不太準確
- 當預定到達時間與實際到達時間不同時，航班會延誤超過15分鐘



# 參考文獻

- Kaggle

<https://www.kaggle.com/fabiendaniel/predicting-flight-delays-tutorial/notebook>

- A Medium Corporation

<https://medium.com/@yehjames/%E8%B3%87%E6%96%99%E5%88%86%E6%9E%90-%E6%A9%9F%E5%99%A8%E5%AD%B8%E7%BF%92-%E7%AC%AC4-1%E8%AC%9B-kaggle%E7%AB%B6%E8%B3%BD-%E9%90%B5%E9%81%94%E5%B0%BC%E8%99%9F%E7%94%9F%E5%AD%98%E9%A0%90%E6%B8%AC-%E5%89%8D16-%E6%8E%92%E5%90%8D-a8842fea7077>