

作業一

議題：Titanic: Machine Learning from Disaster

1. 你選的這組資料為何重要
了解 Titanic 哪些人可能可以存活的機會。
2. 資料從何而來 (tips: 譬如提供者是誰、以什麼方式蒐集)
Kaggle
3. 蒐集而來的資料型態為何

Data Dictionary

Variable	Definition	Key	Data type
survival	Survival	0 = No, 1 = Yes	category
pclass	Ticket class	1 = 1st, 2 = 2nd, 3 = 3rd	ordinal
sex	Sex		category
Age	Age in years		numeric
sibsp	# of siblings / spouses aboard the Titanic		numeric
parch	# of parents / children aboard the Titanic		numeric
ticket	Ticket number		numeric

fare	Passenger fare		numeric
cabin	Cabin number		numeric
embarked	Port of Embarkation	C = Cherbourg, Q = Queenstown, S = Southampton	category

4. 這組資料想解決的問題如何評估

這組資料主要是 Binary classification, 最後評估的標準為 Accuracy。