

統計學習初論 計畫書

B04705029 張家華、B05801037 李柏霆、R06722013許聖謙、R07741023鄭子萱、
R07724017 謝昕亭

分析主題：預測流浪動物領養狀況

資料及來源：Kaggle (<https://www.kaggle.com/c/petfinder-adoption-prediction/overview>)

資料筆數：14993筆

欄位數量：24個

問題：流浪動物的問題

預測動機：

1. 透過預測流浪動物的領養速度，最適化流浪動物資源配置
2. 根據寵物特徵資料，預測領養情況
3. 想了解何種寵物的樣貌較受到大家歡迎

未來展望：能主動找到更多想要養狗或者適合的人

5 至 10 筆的範例資料:

Type	Name	Age	Breed1	Breed2	Gender	Color1	Color2	Color3	MaturitySize
0	2	Nibble	3	299	0	1	1	7	0
1	2	No Name Yet	1	265	0	1	1	2	0
2	1	Brisco	1	307	0	1	2	7	0
3	1	Miko	4	307	0	2	1	2	0
4	1	Hunter	1	307	0	1	1	0	0

FurLength	Vaccinated	Dewormed	Sterilized	Health	Quantity	Fee	State	RescuerID
1	1	2	2	2	1	1	100	41326
2	2	3	3	3	1	1	0	41401
2	2	1	1	2	1	1	0	41326
2	1	1	1	2	1	1	150	41401
2	1	2	2	2	1	1	0	41326

VideoAmt	PetID	PhotoAmt	AdoptionSpeed	dataset_type
8480853f516546f6cf33aa88cd76c379	0	86e1089a3	1.0	2

3082c7125d8fb66f7dd4bff4192c8b14	0	6296e909a	2.0	0
fa90fa5b1ee11c86938398b60abc32cb	0	3422e4906	7.0	3
9238e4f44c71a75282e62f7136c6b240	0	5842f1ff5	8.0	2
95481e953f8aed9ec3d16fc4509537e8	0	850a43f90	3.0	2

欄位說明：

- PetID - 寵物資料的唯一-hash ID
- AdoptionSpeed - 被領養的速度，數值越低越快，為我們要預測的y。
- Type - 動物類型（1 =狗，2 =貓）
- Name - 寵物名稱（如果沒有命名，則為空）
- Age - 上市時寵物的年齡，以月計
- Breed1 - 寵物的主要品種（參見BreedLabels字典）
- Breed2 - 寵物的次級品種（如果寵物是混合品種）（參見BreedLabels字典）
- Gender - 寵物的性別（1 =男性，2 =女性，3 =混合，如果個人資料代表寵物群體）
- Color1 - 寵物的顏色1（參見ColorLabels字典）
- Color2 - 寵物的顏色2（參考ColorLabels字典）
- Color3 - 寵物的顏色3（參考ColorLabels字典）
- MaturitySize - 成熟時的大小（1 =小，2 =中，3 =大，4 =超大，0 =未指定）
- FurLength - 毛皮長度（1 =短，2 =中，3 =長，0 =未指定）
- Vaccinated - 寵物接種疫苗（1 =是，2 =否，3 =不確定）
- Dewormed - 寵物已被驅蟲（1 =是，2 =否，3 =不確定）
- Sterilized - 寵物已被絕育/絕育（1 =是，2 =否，3 =不確定）
- Health - 健康狀況（1 =健康，2 =輕微損傷，3 =嚴重受傷，0 =未指定）
- Quantity - 個人資料中代表的寵物數量
- fee - 領養費（0 =免費）
- State - 馬來西亞的州（參見StateLabels字典）
- RescuerID - 救援者的唯一 hash ID
- VideoAmt - 此寵物的已上傳視頻總數
- PhotoAmt - 此寵物的上傳照片總數