

1999問題分類器



消費支出與生活品質

Outline



- ➡ 主題：消費支出與生活品質
- ➡ 區域：新竹市
- ➡ 使用的資料集
 - － 新竹市政府社會處便民服務常見問答集
 - － 台北市政府社會處便民服務常見問答集
- ➡ 資料分析與問題觀察
 - － 資料分析方法與結果

使用的資料集

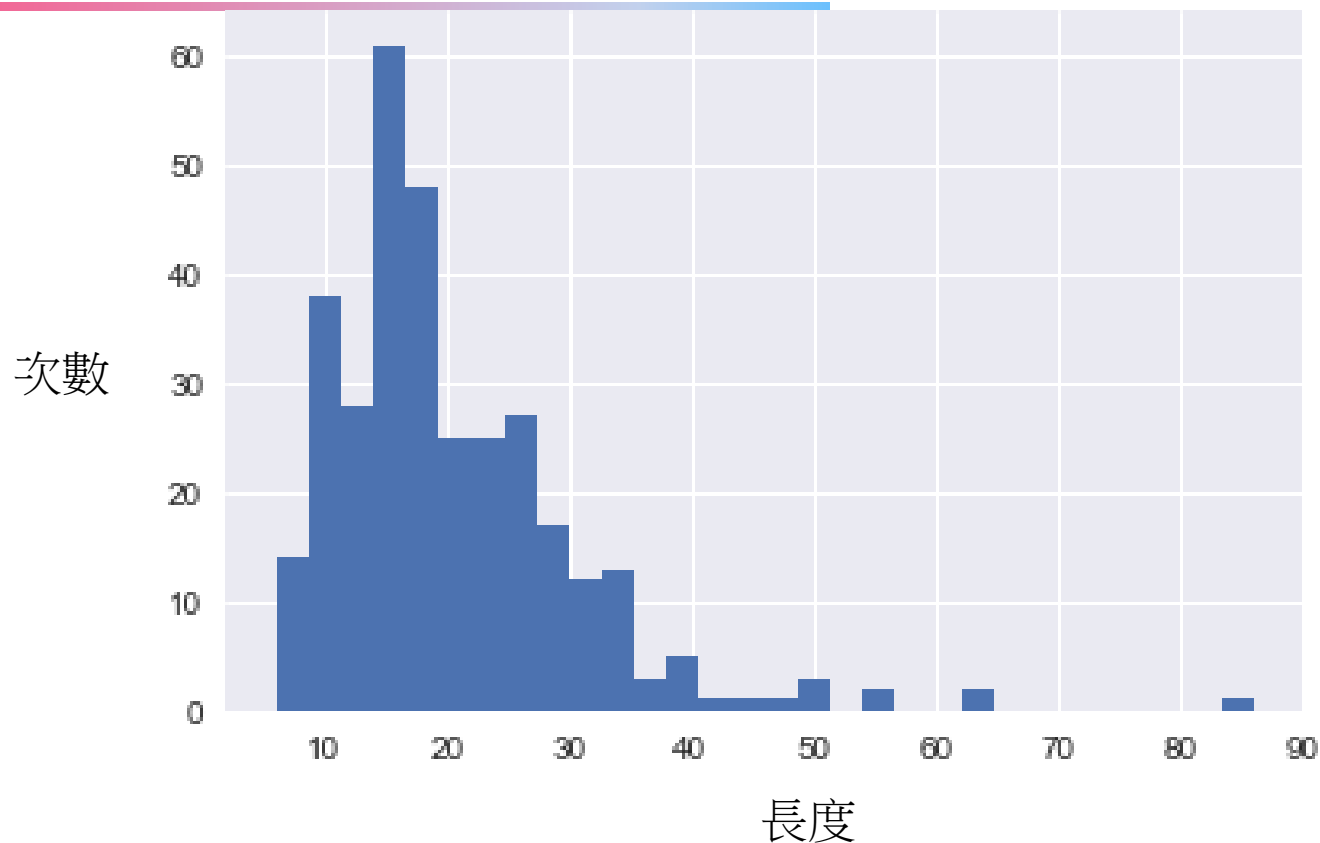


➡ 新竹市政府社會處便民服務常見問答集

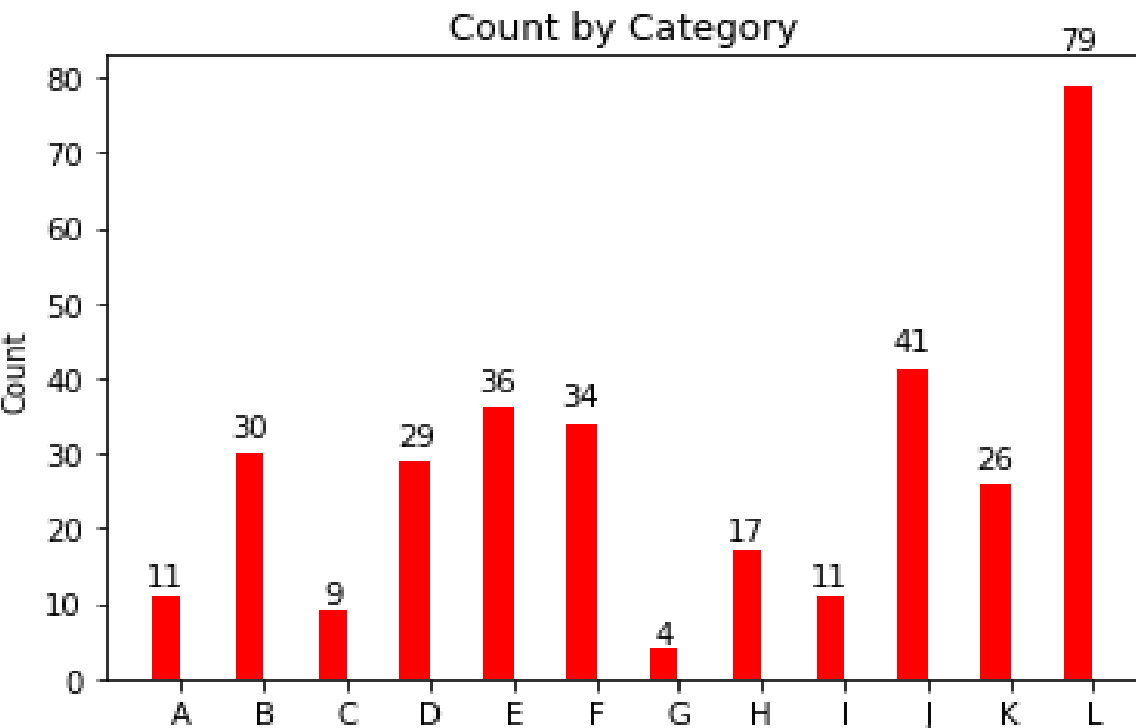
- ❑ 幼兒福利及托育服務
- ❑ 兒童少年及家庭福利
- ❑ 婦女福利
- ❑ 老人福利
- ❑ 身心障礙者福利
- ❑ 社會救助
- ❑ 家庭暴力及性侵害防治服務
- ❑ 社工專業服務
- ❑ 社區發展服務
- ❑ 人民團體及合作社場服務
- ❑ 志願服務
- ❑ 國民年金
- ❑ 公益勸募

資料分析

資料觀察-句子的長度



資料觀察-類別次數統計分析



代碼	問題集類別	筆數
A	人民團體及合作社場服務	11
B	兒童少年及家庭福利	30
C	國民年金	9
D	婦女福利	29
E	家庭暴力及性侵害防治服務	36
F	幼兒福利及托育服務	34
G	志願服務	4
H	社區發展服務	17
I	社工專業服務	11
J	社會救助	41
K	老人福利	26
L	身心障礙者福利	79

模型訓練

小孩出生後應於何時申請育兒津貼?

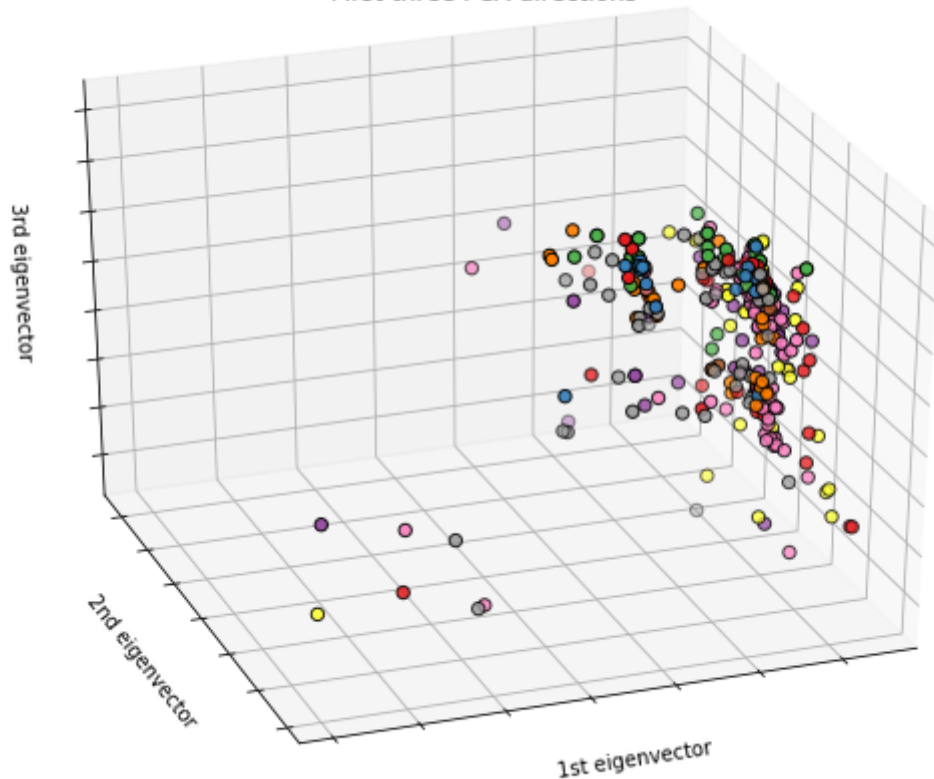
小孩，出生，後，應，於，何時，申請育，兒津，貼，？，孩子，小孩子，孩童，小女孩，之後，以後，隨後，後則，其後，過後，而後

8

資料觀察-資料離散程度



First three PCA directions



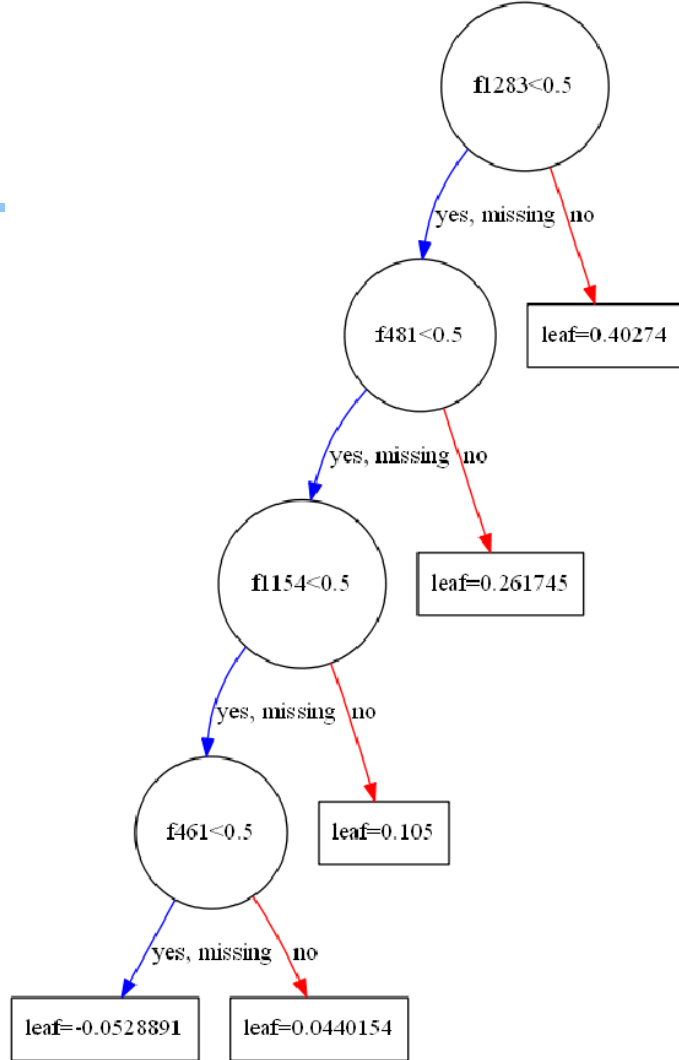
預測結果

資料分析方法與結果



資料分析方法	準確率
KNN	82.57%
SVC	75.84%
XGB	96.64%

XGB分析結果



XGB分析結果



```
In [12]: import xgboost

# Set our parameters for xgboost
params = {}
params['objective'] = 'multi:softmax'
# scale weight of positive examples
params['eta'] = 0.1
params['max_depth'] = 4
params['silent'] = 1
params['nthread'] = 4
params['num_class'] = len(set(Y))

## control over-fitting, higher depth specific to a particular sample.
d_train = xgboost.DMatrix(x_train_split, label=y_train_split)
d_valid = xgboost.DMatrix(x_valid_split, label=y_valid_split)

watchlist = [(d_train, 'train'), (d_valid, 'valid')]

bst = xgboost.train(params, d_train, 1000, watchlist, early_stopping_rounds=100, verbose_eval=10)

[0]    train-merror:0.168582    valid-merror:0.318182
Multiple eval metrics have been passed: 'valid-merror' will be used for early stopping.

Will train until valid-merror hasn't improved in 100 rounds.
[10]    train-merror:0.091954    valid-merror:0.181818
[20]    train-merror:0.065134    valid-merror:0.181818
[30]    train-merror:0.057471    valid-merror:0.181818
[40]    train-merror:0.045977    valid-merror:0.181818
[50]    train-merror:0.022989    valid-merror:0.181818
[60]    train-merror:0.019157    valid-merror:0.166667
[70]    train-merror:0.015326    valid-merror:0.166667
[80]    train-merror:0.015326    valid-merror:0.166667
[90]    train-merror:0.011494    valid-merror:0.181818
[100]   train-merror:0.003831    valid-merror:0.166667
[110]   train-merror:0.003831    valid-merror:0.181818
[120]   train-merror:0.003831    valid-merror:0.181818
Stopping. Best iteration:
[23]    train-merror:0.061303    valid-merror:0.166667
```

XGB錯誤分析



Question	Cat	Pred
請問親子館位置?館內設施及服務?會員申請方式?預約課程及活動開放時間?	幼兒福利及托育服務	婦女福利
如何查詢新竹市親子館「新手父母教室」的課程內容及上課時間?如何報名	幼兒福利及托育服務	婦女福利
離婚後孩子監護權歸我，還需要前夫的同意才能出養孩子嗎？	兒童少年及家庭福利	婦女福利
請問免費法律諮詢及心理諮商服務項目?服務地點?服務時間？申請資格?申請方式?	婦女福利	社會救助
特殊境遇家庭兒童托育津貼如何申請？補助額度如何？	婦女福利	幼兒福利及托育服務
如何申請老花眼鏡、拐杖補助？	老人福利	身心障礙者福利
如何申請安置兒童日托中心之補助？	身心障礙者福利	兒童少年及家庭福利
誰容易成為人口販運被害人？	社工專業服務	家庭暴力及性侵害防治服務
申請籌組人民團體之條件為何？	人民團體及合作社場服務	身心障礙者福利
申請人民團體應附書表及證件為何？	人民團體及合作社場服務	身心障礙者福利
如何加入志工行列？	志願服務	身心障礙者福利



DEMO



結論



項目	比重	說明
資料洞察力	40%	資料和議題之關聯性
資料分析力	40%	資料分析方法
主題的切合度	20%	資料與現象是否切合地方政府施政議題、或是民眾生活圈議題

Q & A