

Telco Customer Churn

資料探勘報告

組員：

資財碩二-106AB8010-李欣樺

資財碩一-107AB8001-蔡雯惠

資財碩一-107AB8406-袁嘉妮

CONTENTS

1

資料集介紹

2

分析目的

3

資料清理過程

4

模型建構與驗證

5

結論與建議

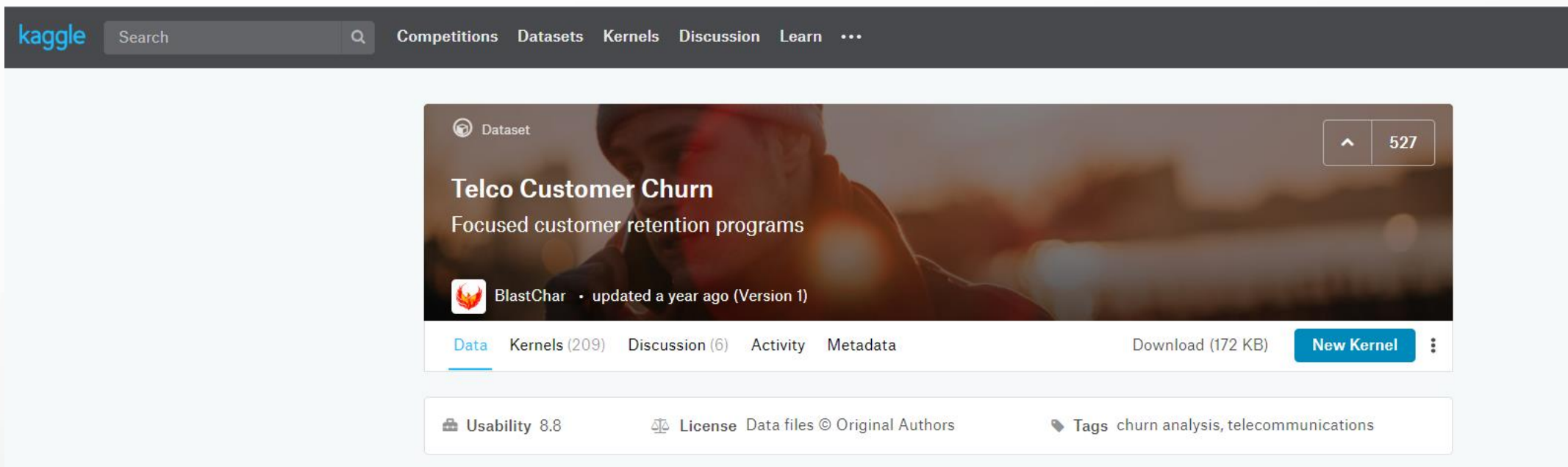


1
PART 01

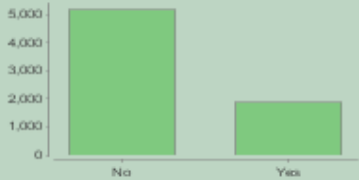


資料集介紹

► 資料集介紹

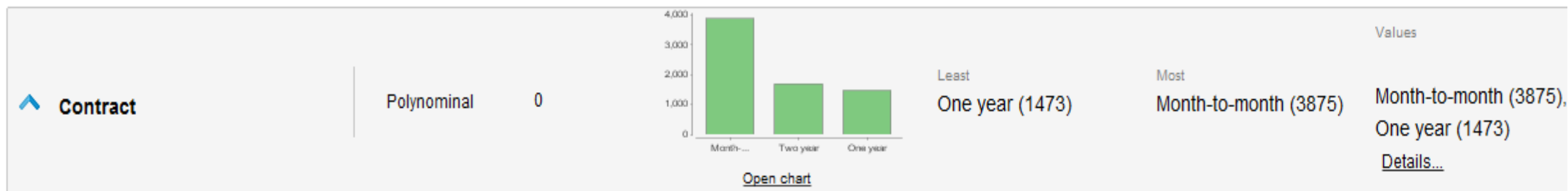
- 資料集名稱：Telco Customer Churn
- 資料集來源：Kaggle
- 資料筆數：14086筆數




► 資料集介紹-參數

<div>Label</div> <div>Churn</div>	<div>Polynomial</div> <div>0</div>	<div><div>Open chart</div></div>	<div>Least</div> <div>Yes (1869)</div>	<div>Most</div> <div>No (5174)</div>	<div>Values</div> <div>No (5174), Yes (1869)</div> <div>Details...</div>
<div>SeniorCitizen</div>	<div>Binominal</div> <div>0</div>	<div><div>Open chart</div></div>	<div>Least</div> <div>true (1142)</div>	<div>Most</div> <div>false (5901)</div>	<div>Values</div> <div>false (5901), true (1142)</div> <div>Details...</div>
<div>gender</div>	<div>Polynomial</div> <div>0</div>	<div><div>Open chart</div></div>	<div>Least</div> <div>Female (3488)</div>	<div>Most</div> <div>Male (3555)</div>	<div>Values</div> <div>Male (3555), Female (3488)</div> <div>Details...</div>

► 資料集介紹-參數



Nominal values



Index	Nominal value	Absolute count	Fraction
1	Month-to-month	3875	0.550
2	Two year	1695	0.241
3	One year	1473	0.209


► 資料集介紹-參數

Dependents	Polynomial	0	 <p>Least Yes (2110)</p> <p>Most No (4933)</p> <p>Values No (4933), Yes (2110) Details...</p>
tenure	Real	0	 <p>Min 0</p> <p>Max 72</p> <p>Average 32.371</p> <p>Deviation 24.559</p>
TotalCharges	Real	11	 <p>Min 18.800</p> <p>Max 8684.800</p> <p>Average 2283.300</p> <p>Deviation 2266.771</p>

▶ 資料集介紹-參數



Nominal values



Index	Nominal value	Absolute count	Fraction
1	Electronic check	2365	0.336
2	Mailed check	1612	0.229
3	Bank transfer (automatic)	1544	0.219
4	Credit card (automatic)	1522	0.216



2
PART 02



分析目的

► 分析目的

- 分析方法

- 比較過Naïve Bayes、Decision Tree、Random Forest……等方式，發現k-NN的準確率最高

- 分析目的

- 主要是利用該資料集，將挑選出來的屬性透過k-NN的方法去預測該顧客是否會流失。

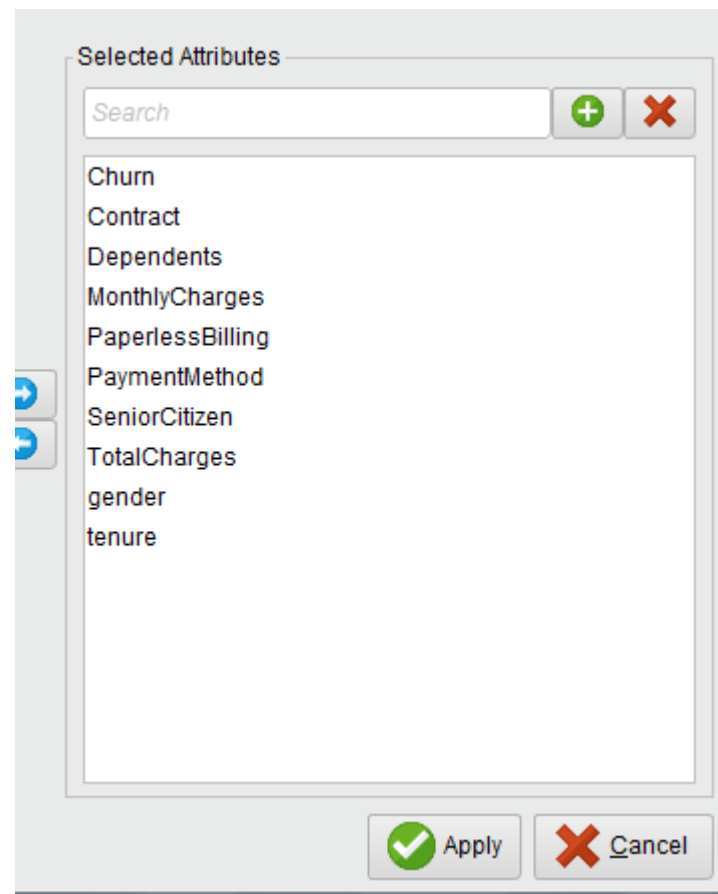
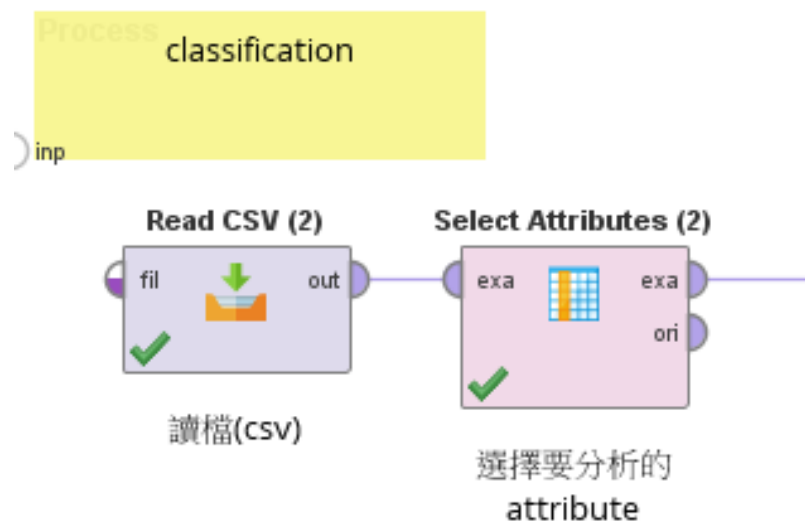


3
PART 03

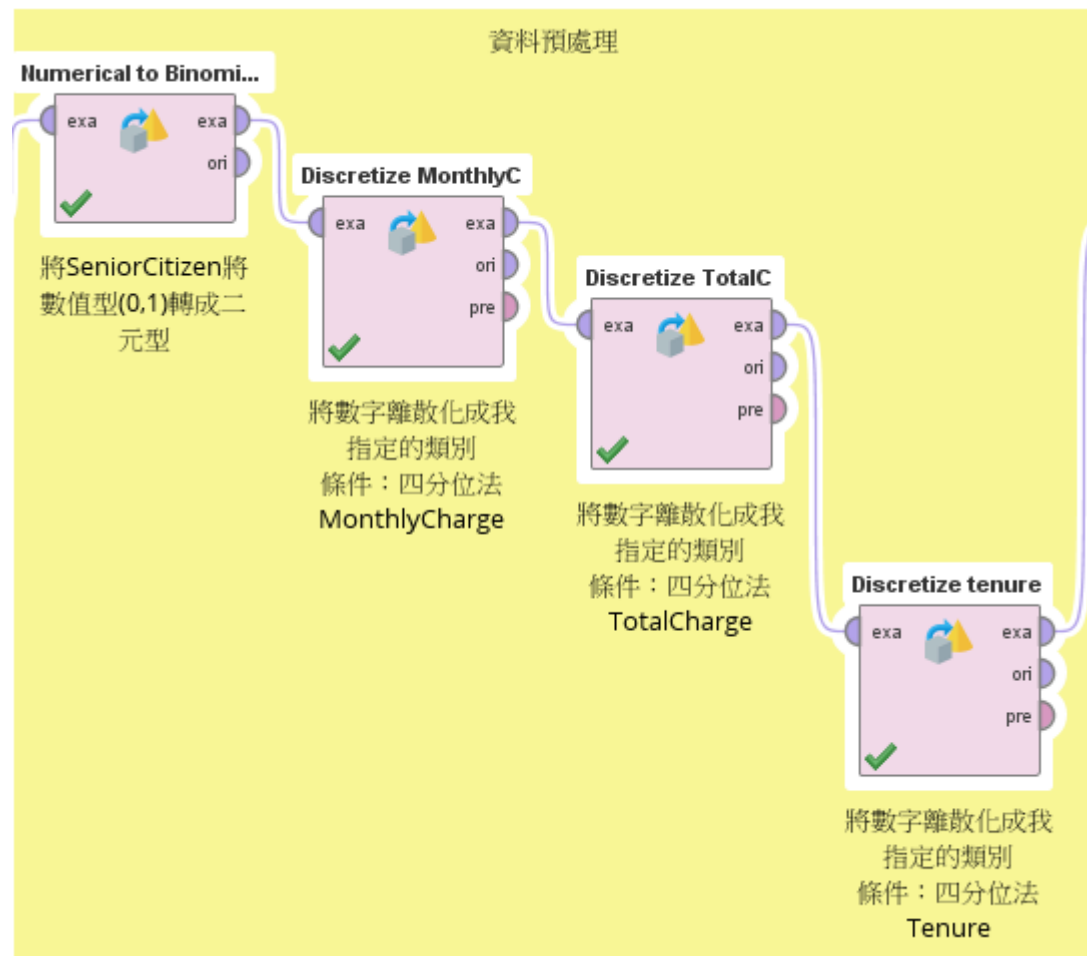


資料清理過程

► 讀檔與屬性挑選



► 預處理



Parameters ✕

🔗 **Discretize tenure (Discretize by User Specification)**

attribute filter type: single ⓘ ⓘ

attribute: tenure ⓘ ⓘ

☐ invert selection ⓘ ⓘ

☐ include special attributes ⓘ ⓘ

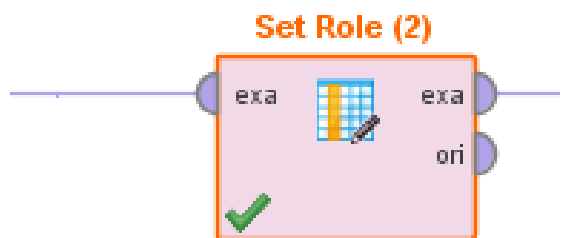
classes: [Edit List \(4\)...](#) ⓘ ⓘ

Edit Parameter List: classes

📝 Edit Parameter List: **classes**
Defines the classes and the upper limits of each class.


class names	upper limit
0%-25%	9.0
26%-50%	29.0
51%-75%	55.0
76%-100%	72.0

► 設定label



分類需要指定label
因此指定churn

Parameters ✕

 **Set Role (2) (Set Role)**


attribute name

Churn ▼ ⓘ

target role

label ▼ ⓘ

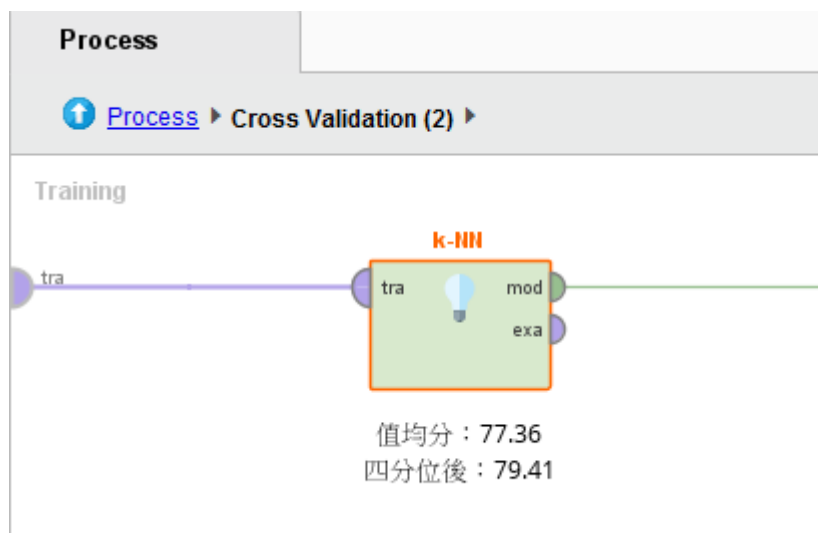
set additional roles

 Edit List (0)... ⓘ



模型建構與 驗證

► 模型建構與驗證



Parameters

✕

💡 k-NN

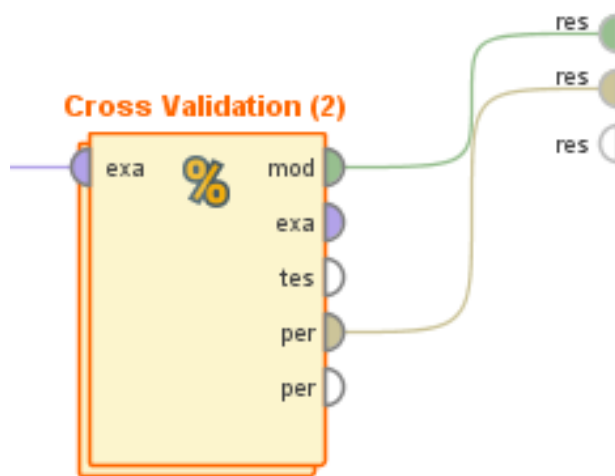
k 60 ⓘ

☒ weighted vote ⓘ

measure types MixedMeasures ⓘ

mixed measure MixedEuclideanDistance ⓘ

► 模型建構與驗證



Parameters ✕

% Cross Validation (2) (Cross Validation)

☐ split on batch attribute ⓘ

☐ leave one out ⓘ

number of folds ⓘ

sampling type ⓘ

☐ use local random seed ⓘ

☒ enable parallel execution ⓘ



PART 05

結論與建議

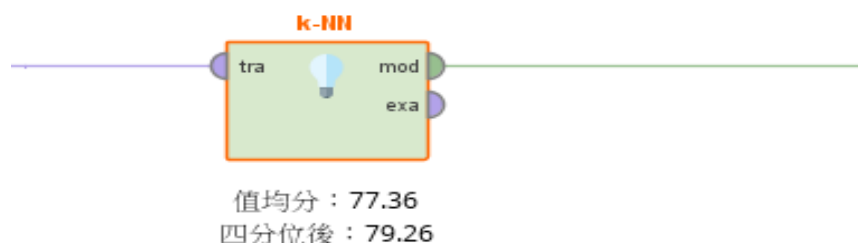
► 結論與建議

accuracy: 79.41% +/- 1.47% (micro average: 79.41%)

	true No	true Yes	class precision
pred. No	4671 TN	947 FN	83.14%
pred. Yes	503 FP	922 TP	64.70%
class recall	90.28%	49.33%	

► 結論與建議

- 值分成四等份做離散化→四分位離散化：每個方法的準確率都有些微提升1~2%



Decision Tree (2)



值均分：75.58
四分位後：77.71

Random Forest



值均分：76.0
四分位後：77.1

Deep Learning



值均分：75.44
四分位後：76.56

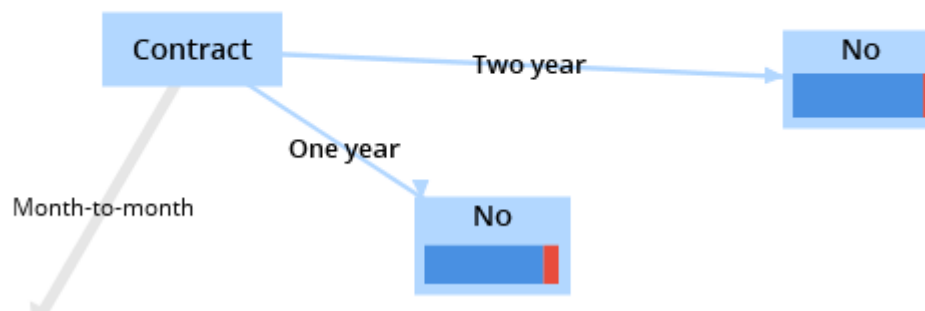
Naive Bayes (2)



值均分：76.73
四分位後：77.48

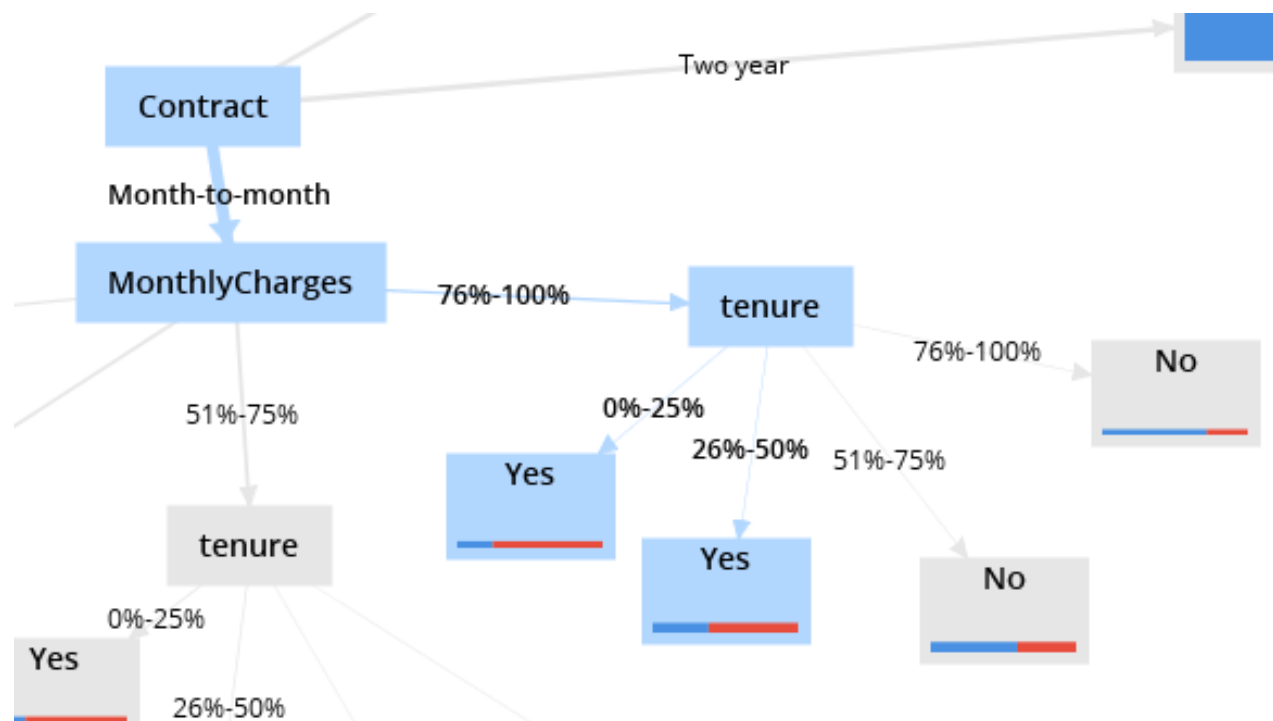
► 結論與建議

- 比較不同的預測模型，k-NN的正確率較高，較能準確地預測客戶是否流失。
- 從決策樹的模型可以看到，合約期限較長的(1、2 years)的客戶，都傾向於待在原有的電信公司。



► 結論與建議

- 合約期限以月計算，且月繳費用中高的客戶，如果待在該電信公司時間越短就越容易出走流失。



► 結論與建議

- 我們認為應該推出一些促銷方案，讓客戶能長期待在我們的電信公司，使其成為我們長期客戶，降低流失率。
- 但在台灣因為各家電信不斷在推出低價格吃到飽的方案，因此我們認為電信公司還需要有自己的特色，才能脫穎而出。

Thanks for your listening.