

資料前處理：

- 1)藉由data.info()可以看出data中沒有空值需要填補
- 2)DayofWeek特徵轉成dummy特徵，產生Monday-Sunday共七個特徵
PdDistrict特徵也切分出10個dummy特徵
Dates取出小時，轉成24個dummy特徵
- 3)X,Y資料嘗試做兩種處理，並比較效果

```
#(2-4) 對X跟Y進行處理：把X和Y資料標準化
xy_scaler = StandardScaler()
data[['X','Y']] = xy_scaler.fit_transform(data[['X','Y']])

#(2-4) 對X跟Y進行處理：透過PCA進行降維
pca = PCA(n_components=1)
data['Coordinates'] = pca.fit_transform(data[['X','Y']])
data.drop(data[['X','Y']],axis=1,inplace=True)
```

```
precision    0.23427278638745075
recall       0.2502767489852537
accuracy      0.2502767489852537
```

```
precision    0.23118882343529018
recall       0.24705142747809924
accuracy      0.24705142747809924
```

發現標準化的方法效果較佳，所以最後利用標準化處理X和Y資料

結果：

- 4)限制樹高度為3跟未限制樹高度的準確度比較

```
precision    0.08999584234254186
recall       0.23010937848783444
accuracy      0.23010937848783444
```

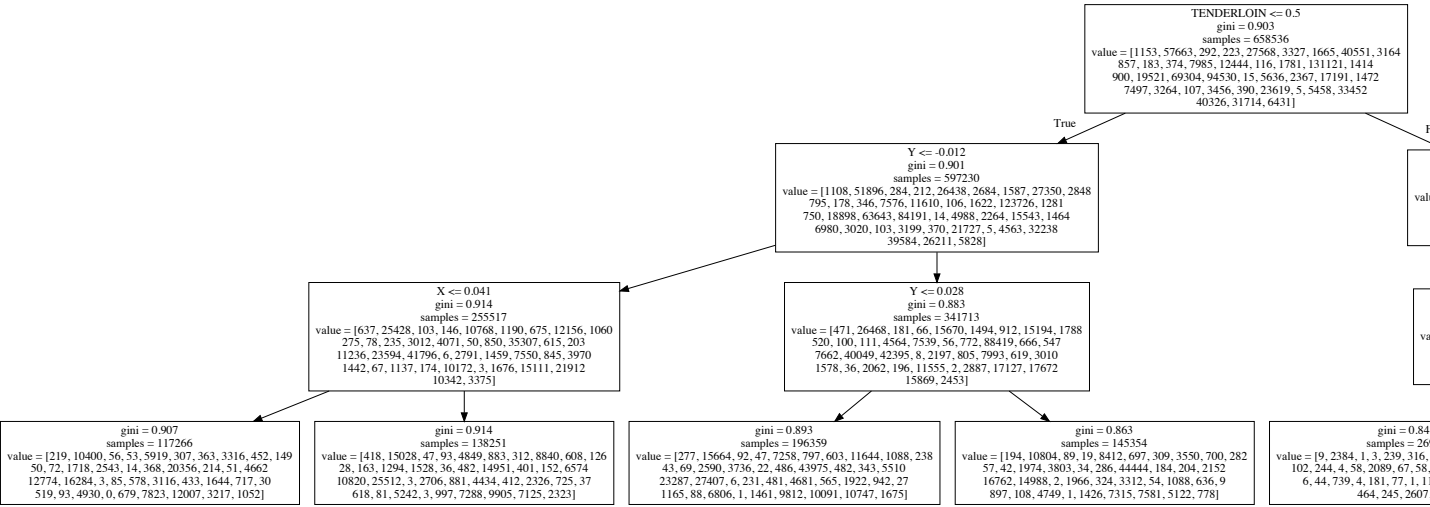
```
precision    0.23427278638745075
recall       0.2502767489852537
accuracy      0.2502767489852537
```

- 5)以sklearn.metrics.classification_report顯示針對各犯罪類型的模型預測能力

	precision	recall	f1-score	support
ARSON	0.00	0.00	0.00	366
ASSAULT	0.00	0.00	0.00	19203
BAD CHECKS	0.00	0.00	0.00	89
BRIBERY	0.00	0.00	0.00	77
BURGLARY	0.00	0.00	0.00	9199
DISORDERLY CONDUCT	0.00	0.00	0.00	1053
DRIVING UNDER THE INFLUENCE	0.00	0.00	0.00	549
DRUG/NARCOTIC	0.27	0.29	0.28	13609
DRUNKENNESS	0.00	0.00	0.00	1133
EMBEZZLEMENT	0.00	0.00	0.00	303
EXTORTION	0.00	0.00	0.00	67
FAMILY OFFENSES	0.00	0.00	0.00	130
FORGERY/COUNTERFEITING	0.00	0.00	0.00	2629
FRAUD	0.00	0.00	0.00	4207
GAMBLING	0.00	0.00	0.00	31
KIDNAPPING	0.00	0.00	0.00	581
LARCENY/THEFT	0.25	0.83	0.38	43680
LIQUOR LAWS	0.00	0.00	0.00	432
LOITERING	0.00	0.00	0.00	313
MISSING PERSON	0.00	0.00	0.00	6462
NON-CRIMINAL	0.00	0.00	0.00	22963
OTHER OFFENSES	0.18	0.34	0.24	31667
PORNOGRAPHY/OBSCENE MAT	0.00	0.00	0.00	6
PROSTITUTION	0.18	0.10	0.13	1811
RECOVERED VEHICLE	0.00	0.00	0.00	755
ROBBERY	0.00	0.00	0.00	5711
RUNAWAY	0.00	0.00	0.00	469
SECONDARY CODES	0.00	0.00	0.00	2548
SEX OFFENSES FORCIBLE	0.00	0.00	0.00	1106
SEX OFFENSES NON FORCIBLE	0.00	0.00	0.00	37
STOLEN PROPERTY	0.00	0.00	0.00	1075
SUICIDE	0.00	0.00	0.00	121
SUSPICIOUS OCC	0.00	0.00	0.00	7953
TRESPASS	0.00	0.00	0.00	1853
VANDALISM	0.00	0.00	0.00	11179
VEHICLE THEFT	0.00	0.00	0.00	13431
WARRANTS	0.00	0.00	0.00	10559
WEAPON LAWS	0.00	0.00	0.00	2156
accuracy			0.23	219513
macro avg	0.02	0.04	0.03	219513
weighted avg	0.09	0.23	0.13	219513

6)產出高度為3的決策樹

左半邊：



右半邊：

