

B0027217 Y.D. LIN

法官演算法對毒品案件的裁量

大綱

- ▶ 動機
- ▶ 資料介紹
- ▶ 資料前處理
- ▶ 探索式資料分析
- ▶ 資料探勘演算法 Decision Tree & Naive Bayes & Random Forest
- ▶ 討論：如何讓量刑模型更好
- ▶ Q&A

動機與目的

- ▶ 機器量刑系統能夠大幅減輕司法官的工作量
- ▶ 機器量刑系統能降低社會大眾對法官自由心證的疑慮
- ▶ FUN

資料介紹

- ▶ 資料名稱：司法院各級法院裁判書104年10月
- ▶ 資料格式：JSON
- ▶ 資料來源：政府資料開放平臺

資料長相

```
{
  "JNO": "38",
  "JFULL": "臺灣高等法院花蓮分院刑事判決      104年度原上易字第38號\r\n上  訴  人\r\n即  被  告  劉祈薇",
  "JCASE": "原上易",
  "JID": "HLHM,104,原上易,38,20151014,1",
  "JTITLE": "違反毒品危害防制條例",
  "JDATE": "20151014",
  "JYEAR": "104"
}
```

判決沒因子，此風不可長

- ▶ 《毒品危害防制條例》中所有犯罪行為的定義及其刑度。

- ▶ 參照

《刑法》第6章 累犯

《刑法》第7章 數罪併罰

《刑法》第8章 刑之酌科及加減

《毒品危害防制條例》第4條 製造、販賣、運輸

- ▶製造、運輸、販賣**第一級毒品**者，處**死刑或無期徒刑**；處無期徒刑者，得併科新臺幣二千萬元以下罰金。
- ▶製造、運輸、販賣**第二級毒品**者，處**無期徒刑或七年以上有期徒刑**，得併科新臺幣一千萬元以下罰金。
- ▶製造、運輸、販賣**第三級毒品**者，處**七年以上有期徒刑**，得併科新臺幣七百萬元以下罰金。
- ▶製造、運輸、販賣**第四級毒品**者，處**五年以上十二年以下有期徒刑**，得併科新臺幣三百萬元以下罰金。
- ▶製造、運輸、販賣專供製造或施用毒品之器具者，處**一年以上七年以下有期徒刑**，得併科新臺幣一百萬元以下罰金。

《毒品危害防制條例》 第5條 意圖販賣而持有

- ▶ 意圖販賣而持有**第一級毒品者**，**處無期徒刑或十年以上有期徒刑**，得併科新臺幣七百萬元以下罰金。
- ▶ 意圖販賣而持有**第二級毒品者**，**處五年以上有期徒刑**，得併科新臺幣五百萬元以下罰金。
- ▶ 意圖販賣而持有**第三級毒品者**，**處三年以上十年以下有期徒刑**，得併科新臺幣三百萬元以下罰金。
- ▶ 意圖販賣而持有**第四級毒品或專供製造、施用毒品之器具者**，**處一年以上七年以下有期徒刑**，得併科新臺幣一百萬元以下罰金。

《毒品危害防制條例》第6條 強暴、脅迫、欺瞞或其他非法之方法使人施用

- ▶ 以強暴、脅迫、欺瞞或其他非法之方法使人施用第一級毒品者，處死刑、無期徒刑或十年以上有期徒刑；處無期徒刑或十年以上有期徒刑者，得併科新臺幣一千萬元以下罰金。
- ▶ 以前項方法使人施用第二級毒品者，處無期徒刑或七年以上有期徒刑，得併科新臺幣七百萬元以下罰金。
- ▶ 以第一項方法使人施用第三級毒品者，處五年以上有期徒刑，得併科新臺幣五百萬元以下罰金。
- ▶ 以第一項方法使人施用第四級毒品者，處三年以上十年以下有期徒刑，得併科新臺幣三百萬元以下罰金。
- ▶ 前四項之未遂犯罰之。

《毒品危害防制條例》 第7條 引誘他人施用

- ▶ 引誘他人施用第一級毒品者，處三年以上十年以下有期徒刑，得併科新臺幣三百萬元以下罰金。
- ▶ 引誘他人施用第二級毒品者，處一年以上七年以下有期徒刑，得併科新臺幣一百萬元以下罰金。
- ▶ 引誘他人施用第三級毒品者，處六月以上五年以下有期徒刑，得併科新臺幣七十萬元以下罰金。
- ▶ 引誘他人施用第四級毒品者，處三年以下有期徒刑，得併科新臺幣五十萬元以下罰金。
- ▶ 前四項之未遂犯罰之。

《毒品危害防制條例》 第8條 轉讓毒品

- ▶ 轉讓第一級毒品者，處一年以上七年以下有期徒刑，得併科新臺幣一百萬元以下罰金。
- ▶ 轉讓第二級毒品者，處六月以上五年以下有期徒刑，得併科新臺幣七十萬元以下罰金。
- ▶ 轉讓第三級毒品者，處三年以下有期徒刑，得併科新臺幣三十萬元以下罰金。
- ▶ 轉讓第四級毒品者，處一年以下有期徒刑，得併科新臺幣十萬元以下罰金。
- ▶ 前四項之未遂犯罰之。
- ▶ 轉讓**毒品達一定數量者，加重其刑至二分之一**，其標準由行政院定之。

《毒品危害防制條例》第9條 對未成年、孕婦施用毒品

- ▶ 成年人對未成年人犯前三條之罪者，依各該條項規定加重其刑至二分之一。
- ▶ 明知為懷胎婦女而對之犯前三條之罪者，亦同。

《毒品危害防制條例》 第10條 施用毒品

- ▶ 施用第一級毒品者，處六月以上五年以下有期徒刑。
- ▶ 施用第二級毒品者，處三年以下有期徒刑
- ▶ 施用K他命沒有刑責

《毒品危害防制條例》 第11條 持有毒品

- ▶ 持有第一級毒品者，處三年以下有期徒刑、拘役或新臺幣五萬元以下罰金。
- ▶ 持有第二級毒品者，處二年以下有期徒刑、拘役或新臺幣三萬元以下罰金。
- ▶ 持有第一級毒品純質淨重十公克以上者，處一年以上七年以下有期徒刑，得併科新臺幣一百萬元以下罰金。
- ▶ 持有第二級毒品純質淨重二十公克以上者，處六月以上五年以下有期徒刑，得併科新臺幣七十萬元以下罰金。
- ▶ 持有第三級毒品純質淨重二十公克以上者，處三年以下有期徒刑，得併科新臺幣三十萬元以下罰金。
- ▶ 持有第四級毒品純質淨重二十公克以上者，處一年以下有期徒刑，得併科新臺幣十萬元以下罰金。
- ▶ 持有專供製造或施用第一級、第二級毒品之器具者，處一年以下有期徒刑、拘役或新臺幣一萬元以下罰金。

《毒品危害防制條例》第11-1條 持有第三、四級毒品器具

- ▶ 第三級、第四級毒品及製造或施用毒品之器具，無正當理由，不得擅自持有。
- ▶ 無正當理由持有或施用第三級或第四級毒品者，處新臺幣一萬元以上五萬元以下罰鍰，並應限期令其接受四小時以上八小時以下之毒品危害講習。
- ▶ 少年施用第三級或第四級毒品者，應依少年事件處理法處理，不適用前項規定。
- ▶ 第二項裁罰之基準及毒品危害講習之方式、內容、時機、時數、執行單位等事項之辦法，由法務部會同內政部、行政院衛生署定之。

《毒品危害防制條例》第12條 提供罌粟、古柯

- ▶ 意圖供製造毒品之用，而栽種罌粟或古柯者，處無期徒刑或七年以上有期徒刑，得併科新臺幣七百萬元以下罰金。
- ▶ 意圖供製造毒品之用，而栽種大麻者，處五年以上有期徒刑，得併科新臺幣五百萬元以下罰金。
- ▶ 前二項之未遂犯罰之。

《毒品危害防制條例》 第13條 提供罌粟、古柯種子

- ▶ 意圖供栽種之用，而運輸或販賣罌粟種子或古柯種子者，處五年以下有期徒刑，得併科新臺幣五十萬元以下罰金。
- ▶ 意圖供栽種之用，而運輸或販賣大麻種子者，處二年以下有期徒刑，得併科新臺幣二十萬元以下罰金。

《毒品危害防制條例》第14條 販賣與轉讓罌粟、古柯種子

- ▶ 意圖販賣而持有或轉讓罌粟種子、古柯種子者，處三年以下有期徒刑。
- ▶ 意圖販賣而持有或轉讓大麻種子者，處二年以下有期徒刑。
- ▶ 持有罌粟種子、古柯種子者，處二年以下有期徒刑、拘役或新臺幣三萬元以下罰金。
- ▶ 持有大麻種子者，處一年以下有期徒刑、拘役或新臺幣一萬元以下罰金。

《毒品危害防制條例》 第15條 公務人員濫用職權

- ▶ 公務員假借職務上之權力、機會或方法犯第四條第二項或第六條第一項之罪者，處死刑或無期徒刑；處無期徒刑者，得併科新臺幣一千萬元以下罰金。犯第四條第三項至第五項、第五條、第六條第二項至第四項、第七條第一項至第四項、第八條第一項至第四項、第九條至第十四條之罪者，依各該條項規定加重其刑至二分之一。
- ▶ 公務員明知他人犯第四條至第十四條之罪而予以庇護者，處一年以上七年以下有期徒刑

《毒品危害防制條例》 第17條 供出毒品罪

- ▶ 犯第四條至第八條、第十條或第十一條之罪，供出毒品來源，因而查獲其他正犯或共犯者，減輕或免除其刑。
- ▶ 犯第四條至第八條之罪於偵查及審判中均自白者，減輕其刑。

《刑法》第6章 累犯

- ▶ 受徒刑之執行完畢，或一部之執行而赦免後，五年以內故意再犯有期徒刑以上之罪者，為累犯，加重本刑至二分之一。
- ▶ 第九十八條第二項關於因強制工作而免其刑之執行者，於受強制工作處分之執行完畢或一部之執行而免除後，五年以內故意再犯有期徒刑以上之罪者，以累犯論。

《刑法》 第7章 數罪併罰

▶ 太難了...

甲○○施用第一級毒品，處有期徒刑捌月；又施用第二級毒品，處有期徒刑肆月，如易科罰金，以新臺幣壹仟元折算壹日。

張寶維施用第二級毒品，累犯，處有期徒刑伍月，如易科罰金，以新臺幣壹仟元折算壹日；又施用第二級毒品，累犯，處有期徒刑伍月，如易科罰金，以新臺幣壹仟元折算壹日；應執行有期徒刑捌月，如易科罰金，以新臺幣壹仟元折算壹日。

《刑法》 第8章 刑之酌科及加減

- ▶ 一、犯罪之動機、目的。
- ▶ 二、犯罪時所受之刺激。
- ▶ 三、犯罪之手段。
- ▶ 四、**犯罪行為人之生活狀況。**
- ▶ 五、犯罪行為人之品行。
- ▶ 六、**犯罪行為人之智識程度。**
- ▶ 七、犯罪行為人與被害人之關係。
- ▶ 八、犯罪行為人違反義務之程度。
- ▶ 九、犯罪所生之危險或損害。
- ▶ 十、**犯罪後之態度。**

《刑法》 第8章 刑之酌科及加減 - 自首

- ▶ 對於未發覺之罪自首而受裁判者，得減輕其刑。但有特別規定者，依其規定。
- ▶ 法官對自首描述很多樣，需再用人工方式去校對。

犯罪行為人之生活狀況

- ▶ 勉強、貧寒、拮据、收入不固定、難謂良好
- ▶ 小康、富裕、普通
- ▶ 無

犯罪行為人之智識程度

- ▶國中畢業, 國中肄業, 高職畢業, 高中肄業, 高職肄業
- ▶高中畢業, 國小畢業, 專科畢業, 國小肄業, 大專肄業
- ▶五專畢業, 大專, 工專肄業, 專科肄業, 大專畢業, 高職休學,
- ▶無

犯罪行為人之犯後態度

- ▶ 法官的描述太多樣：

『沒有坦承犯行』、 『砌詞圖卸』、 『毫無懊悔之意』、 『坦承犯行， 卻試圖回護』

- ▶ 因此就簡化為： 『坦承犯行』、 『無』

綜合上述的各項因子

- ▶ 共有56個欄位
- ▶ 使用55個因子，推算可能的
刑度

刑度標籤之一

- ▶ 法律學術界的分法

- ▶ 短刑期：一年以下

 - 中短刑期：一年至三年

 - 中長刑期：三年至五年

 - 長刑期：五年以上

刑度標籤之二

- ▶ 法律實務界的分法

- ▶ 短刑期：一年以下

 - 中短刑期：一年至三年

 - 中刑期：三年至五年

 - 中長刑期：五年至七年

 - 長刑期：七年以上

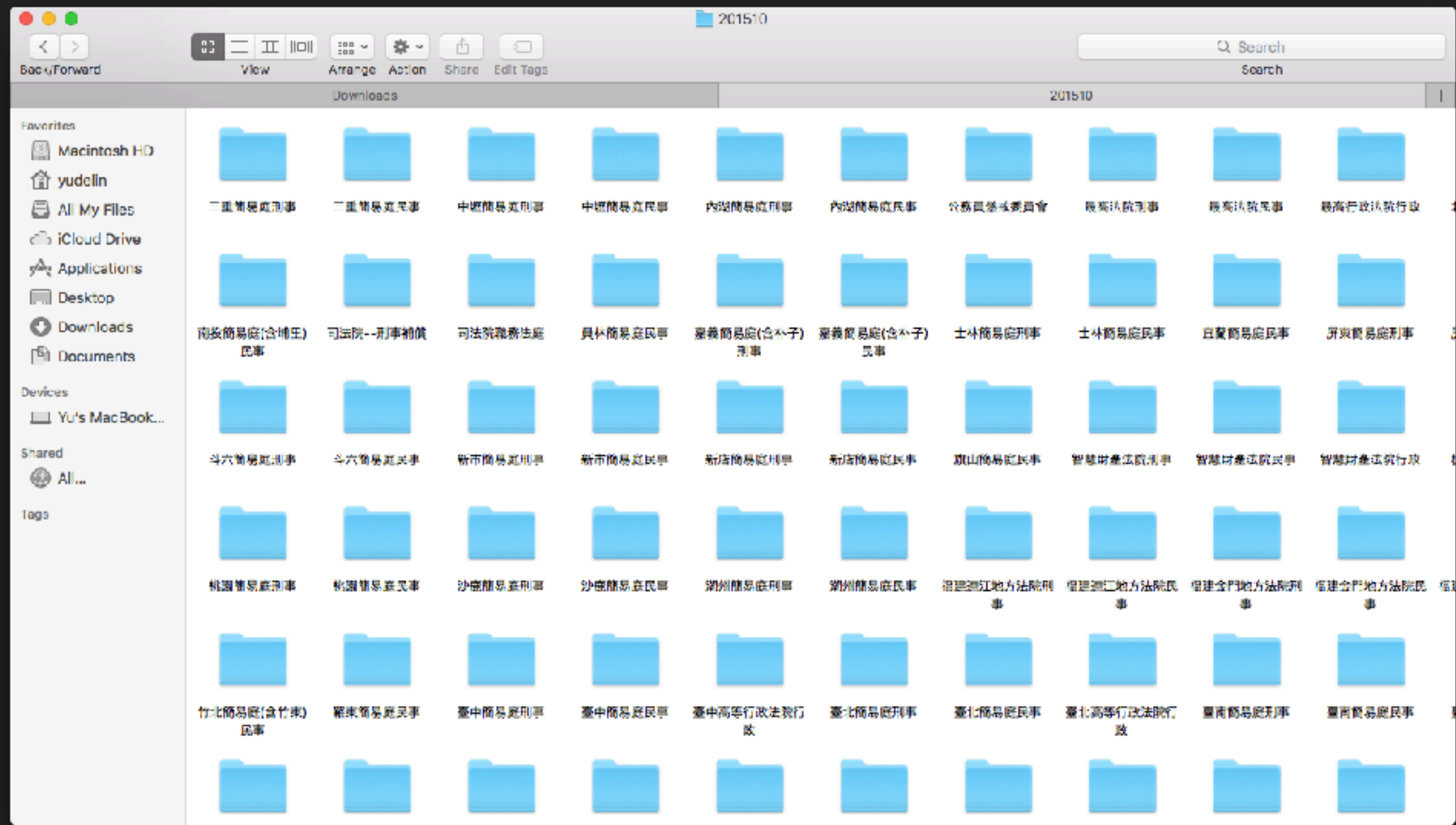
刑度標籤之三

- ▶ 參考司法院量刑統計標準
- ▶ 半年以下、半年以上一年以下、
- ▶ 一年至三年、三年至五年、五年至七年、七年至九年
- ▶ 九年至十一年、十一年至十三年、 十三年至十五年、十五年
以上

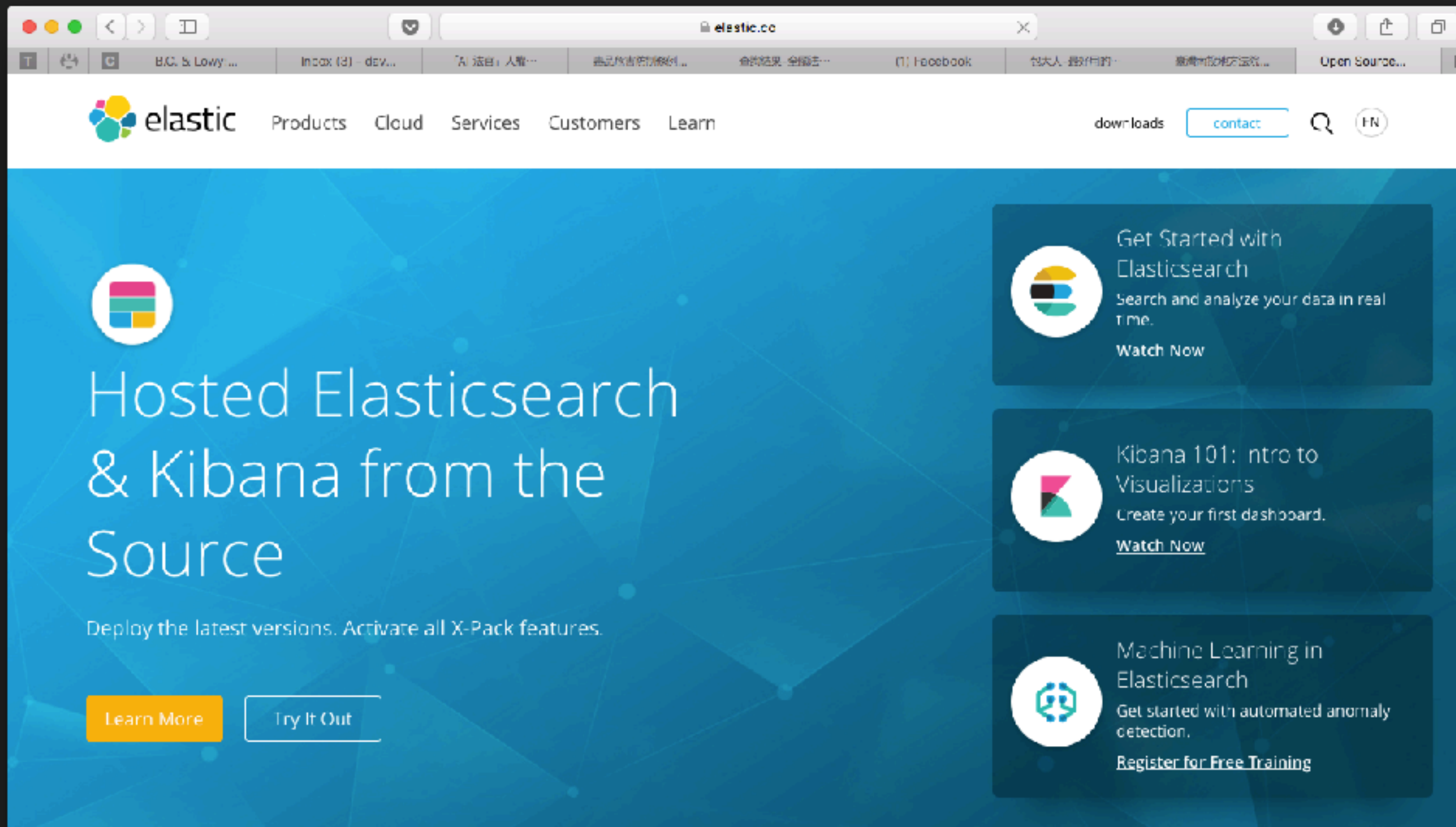
我不入地獄， 誰入地獄？

資料前處理

資料解壓縮後



全文檢索引擎



The screenshot shows the Elastic website homepage. The browser's address bar displays 'elastic.co'. The website's navigation bar includes the Elastic logo, links for 'Products', 'Cloud', 'Services', 'Customers', and 'Learn', and buttons for 'downloads', 'contact', a search icon, and a user profile icon labeled 'HN'. The main content area features a large blue background with a white Elastic logo icon and the text 'Hosted Elasticsearch & Kibana from the Source'. Below this, it states 'Deploy the latest versions. Activate all X-Pack features.' and provides two buttons: 'Learn More' and 'Try It Out'. On the right side, there are three promotional cards: 'Get Started with Elasticsearch' (Search and analyze your data in real time. Watch Now), 'Kibana 101: intro to Visualizations' (Create your first dashboard. Watch Now), and 'Machine Learning in Elasticsearch' (Get started with automated anomaly detection. Register for Free Training).

elastic.co

elastic

Products Cloud Services Customers Learn

downloads contact

Hosted Elasticsearch & Kibana from the Source

Deploy the latest versions. Activate all X-Pack features.

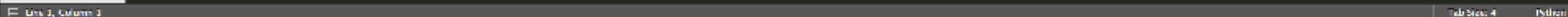
Learn More Try It Out

Get Started with Elasticsearch
Search and analyze your data in real time.
Watch Now

Kibana 101: intro to Visualizations
Create your first dashboard.
Watch Now

Machine Learning in Elasticsearch
Get started with automated anomaly detection.
Register for Free Training

PYTHON SCRIPT



PYTHON SCRIPT

```
import re
import os
import json
from datetime import datetime

from elasticsearch import Elasticsearch as ES

if __name__ == '__main__':

    ES_client = ES(['localhost:9200'])

    # The result is not good enough.
    # Every term with single keyword like '毒' or '品' are all included.
    #ES_client.search(index='test-judges', body={'query': {'multi_match': {'query': '毒品', 'fields': 'JTITLE'}}}, size=3992)
    r = ES_client.search(
        index='test-judges',
        body={
            'query': {
                'multi_match': {
                    'query': '毒品',
                    'fields': 'JTITLE'
                }
            }
        })

    for ind, hit in enumerate(r['hits']['hits']):

        file_name = ['drug_judge_', str(ind), '.json']
        file_name = ''.join(file_name)
        output_file = open(file_name, 'w')
        json.dump(hit['_source'], output_file, indent=True, ensure_ascii=False)
        output_file.close()
```


PYTHON SCRIPT

```
class JudgeParser():

    court_re = re.compile(
        '(?P<name>臺灣[\u4e00-\u9fff]+院)(?P<judgement_type>[\u4e00-\u9fff]+)(?P<judgement_type>[\u4e00-\u9fff]+)'
    )
    defendant_re = re.compile('\n\s*被\s*告(?P<name>([\u4e00-\u9fff]+\s*))\n')
    prosecutor_re = re.compile('\n[\u4e00-\u9fff]*檢察官(?P<name>[\u4e00-\u9fff]+)到庭執行職務')
    prosecutor_agent_re = re.compile('\n\s*(公|上)\s*訴\s*人\s+(?P<agent_name>[\u4e00-\u9fff]+)')

    # More Laws will be added
    laws = re.compile(
        '((刑\s*法\s*|刑\s*事\s*訴\s*訟\s*法|毒\s*品\s*危\s*害\s*防\s*制\s*條\s*例\s*))'
        '((第\s*(\d+|[一二三四五六七八九])*百*[一二三四五六七八九]*十*[一二三四五六七八九])\s*條\s*|'
        '之\s*(\d+|[一二三四五六七八九])*百*[一二三四五六七八九]*十*[一二三四五六七八九]))*\s*'
        '(第\s*(\d+|[一二三四五六七八九])*百*[一二三四五六七八九]*十*[一二三四五六七八九])\s*項)*)+、'
    )
    law_name = re.compile('(P<name>[\u4e00-\u9fff]+(法|條例))')

    # 20170614 Y.D.: Really inefficient parser
    # Verdict Result is inside 主文
    verdict_result_re = re.compile('主\s*文\s+(?P<result>(\s{0,1})+)\s+(理|事|犯)')
```

PYTHON SCRIPT

```
class JudgeParser():

    court_re = re.compile(
        '(?P<name>臺灣[\u4e00-\u9fff]+院)(?P<judgement_type>[\u4e00-\u9fff]+)(?P<judgement_type>[\u4e00-\u9fff]+)'
    )
    defendant_re = re.compile('\n\s*被\s*告(?P<name>([\u4e00-\u9fff]+\s*))\n')
    prosecutor_re = re.compile('\n[\u4e00-\u9fff]*檢察官(?P<name>[\u4e00-\u9fff]+)到庭執行職務')
    prosecutor_agent_re = re.compile('\n\s*(公|上)\s*訴\s*人\s+(?P<agent_name>[\u4e00-\u9fff]+)')

    # More Laws will be added
    laws = re.compile(
        '((刑\s*法\s*|刑\s*事\s*訴\s*訟\s*法|毒\s*品\s*危\s*害\s*防\s*制\s*條\s*例\s*))'
        '((第\s*(\d+|[一二三四五六七八九])*百*[一二三四五六七八九]*十*[一二三四五六七八九])\s*條\s*|'
        '之\s*(\d+|[一二三四五六七八九])*百*[一二三四五六七八九]*十*[一二三四五六七八九]))*\s*'
        '(第\s*(\d+|[一二三四五六七八九])*百*[一二三四五六七八九]*十*[一二三四五六七八九])\s*項)*)+、'
    )
    law_name = re.compile('(?P<name>[\u4e00-\u9fff]+(法|條例))')

    # 20170614 Y.D.: Really inefficient parser
    # Verdict Result is inside 主文
    verdict_result_re = re.compile('主\s*文\s+(?P<result>(\s{0,1})+)\s+(理|事|犯)')
```

PYTHON SCRIPT

```
CONFESS = re.compile('((符合)|(先行)|(不失為)|(合於))?自首')
NO_CONFESS = re.compile('((無|(不符合))自首)')
NO_CONFESS_2 = re.compile('與自首之要件有間')
CONTINUOUS_CRIME = re.compile('累\s*犯\s*')
EDUCATION = re.compile(
    '((?P<edu>(國\s*小\s*)|(國\s*中\s*)|(高\s*中\s*)|(高\s*職\s*)|(專\s*科\s*)|(大\s*專\s*))+'
    '(?P<grad>(畢\s*業\s*|肄\s*業\s*))?)'
    '(?P<tail>(之+\s*(知\s*識|智\s*識|教\s*育)\s*程度)|(等\s*一\s*切\s*情\s*狀\s*)*)')

ADMIT = re.compile('[\u4e00-\u9fff]+坦\s*承\s*犯\s*行\s*')

SENTENCE = re.compile('((應\s*執\s*行|處)\s*有\s*期\s*徒\s*刑\s*([\u4e00-\u9fff]\s*年)*[\u4e00-\u9fff]\s*月)')

TAKE_DRUGS = re.compile('施\s*用\s*第\s*(?P<level>一|二|\d)級\s*毒\s*品')
SELL_DRUGS = re.compile('販\s*賣\s*第\s*(?P<level>一|二|三|四|\d)級\s*毒\s*品')

# Economy Related expression
ECONOMIC = re.compile('(?P<f_desc>[\u4e00-\u9fff]\s*)+\s*經\s*濟\s*狀\s*況\s*為?\s*(?P<desc>[\u4e00-\u9fff]\s*)+')
CLOSE_POOR = re.compile('(勉|免)\s*強?\s*維?\s*持')
POOR = re.compile('貧\s*寒')
CANNT_SAY_GOOD = re.compile('難\s*謂\s*良\s*好')
LIMIT = re.compile('拮\s*据')
OKAY = re.compile('尚\s*可')
BETTER_NORMAL = re.compile('小\s*康')
NORMAL = re.compile('普\s*通')
NOT_BAD = re.compile('不\s*差')
RICH = re.compile('富\s*裕')
```


法官演算法對毒品案件的裁量

毒品危害防制條例第4條第2項	毒品危害防制條例第4條第3項	毒品危害防制條例第4條第4項	毒品危害防制條例第4條第5項	毒品危害防制條例第4條第6項	毒品危害
0	0	0	0	0	
0	0	0	0	0	
0	0	0	0	0	
0	0	0	0	0	
0	0	0	0	0	
0	0	0	0	0	
0	0	0	0	0	
0	0	0	0	0	
0	0	0	0	0	
0	0	0	0	0	
0	0	0	0	0	
0	0	0	0	0	
0	0	0	0	0	
0	0	0	0	0	
0	0	0	0	0	
0	0	0	0	0	
0	0	0	0	0	
0	0	0	0	0	
0	0	0	0	0	
0	0	0	0	0	
0	0	0	0	0	
0	1	0	0	0	
0	1	0	0	0	

法官演算法對毒品案件的裁量

毒品危害防制條例第15條第2項	毒品危害防制條例第17條第1項	毒品危害防制條例第17條第2項	自首	生活狀況	智識程度	犯罪後的態度	累犯	刑度分類
0	0	0	0	無	無	坦承犯行	1	一年至三年
0	0	0	1	無	無	坦承犯行	0	半年以上一年以下
0	0	0	0	無	無	無	1	半年以上一年以下
0	0	0	0	無	無	坦承犯行	1	半年以上一年以下
0	0	0	0	無	無	無	0	半年以上一年以下
0	0	0	0	無	無	無	0	半年以下
0	0	0	0	無	無	無	0	半年以下
0	0	0	0	無	無	坦承犯行	1	半年以上一年以下
0	0	0	0	無	無	坦承犯行	1	半年以上一年以下
0	0	0	0	勉持	國中畢業	坦承犯行	1	半年以下
0	0	0	1	無	無	無	1	一年至三年
0	0	0	0	無	無	無	1	半年以下
0	0	0	0	無	無	坦承犯行	1	半年以上一年以下
0	0	0	0	無	無	無	1	一年至三年
0	0	0	0	勉持	國中畢業	無	1	半年以下
0	0	0	1	勉持	無	坦承犯行	1	半年以下
0	0	0	0	勉持	國中畢業	坦承犯行	1	半年以下
0	0	0	0	勉持	高職畢業	坦承犯行	0	半年以下
0	0	0	0	勉持	國中畢業	無	0	半年以下
0	0	0	0	無	國中畢業	坦承犯行	1	半年以下
0	1	0	0	無	無	坦承犯行	1	半年以上一年以下
0	0	1	0	無	高職肄業	無	0	三年至五年
0	0	1	0	無	高職肄業	無	0	五年至七年

PYTHON SCRIPT

```
def change_01_to_words(row):
    if row == 0:
        return '無'
    elif row == 1:
        return '有'

for output_data, input_data in OUTPUT_INPUT:

    arff_file = open(output_data, 'w')
    drugcrimes_data = open(input_data, 'r')
    drugcrimes_dataframe = pd.read_csv(drugcrimes_data)
    attribute_names = list(drugcrimes_dataframe.columns.values)
    arff_file.write('@relation drug-crimefactor\n')
    arff_file.write('\n')

    for attribute_name in attribute_names:

        row_input = ''

        if re.match('^(毒|自|累)', attribute_name):
            classnames = ['無', '有']
            classnames = ','.join(classnames)
            drugcrimes_dataframe[attribute_name] = drugcrimes_dataframe[attribute_name].apply(change_01_to_words)
            row_input = ' '.join(['@attribute', attribute_name, '{', classnames, '}', '\n'])
        elif re.match('^(智)', attribute_name):
            classnames = [
                '國中畢業', '國中肄業', '高職畢業', '高中肄業', '高職肄業',
                '高中畢業', '國小畢業', '專科畢業', '國小肄業', '大專肄業',
                '五專畢業', '大專', '工專肄業', '專科肄業', '大專畢業', '高職休學', '無']
            classnames = ','.join(classnames)
            row_input = ' '.join([
                '@attribute', attribute_name, '{', classnames, '}', '\n'])
        elif re.match('^(生)', attribute_name):
            classnames = [
                '勉持', '小康', '貧寒', '富裕',
                '普通', '拮据', '難謂良好', '收入不固定', '無']
            classnames = ','.join(classnames)
            row_input = ' '.join([
                '@attribute', attribute_name, '{', classnames, '}', '\n'])
        elif re.match('^(犯罪後的態度', attribute_name):
```


ARFF

[illegible][illegible]

挖礦以前

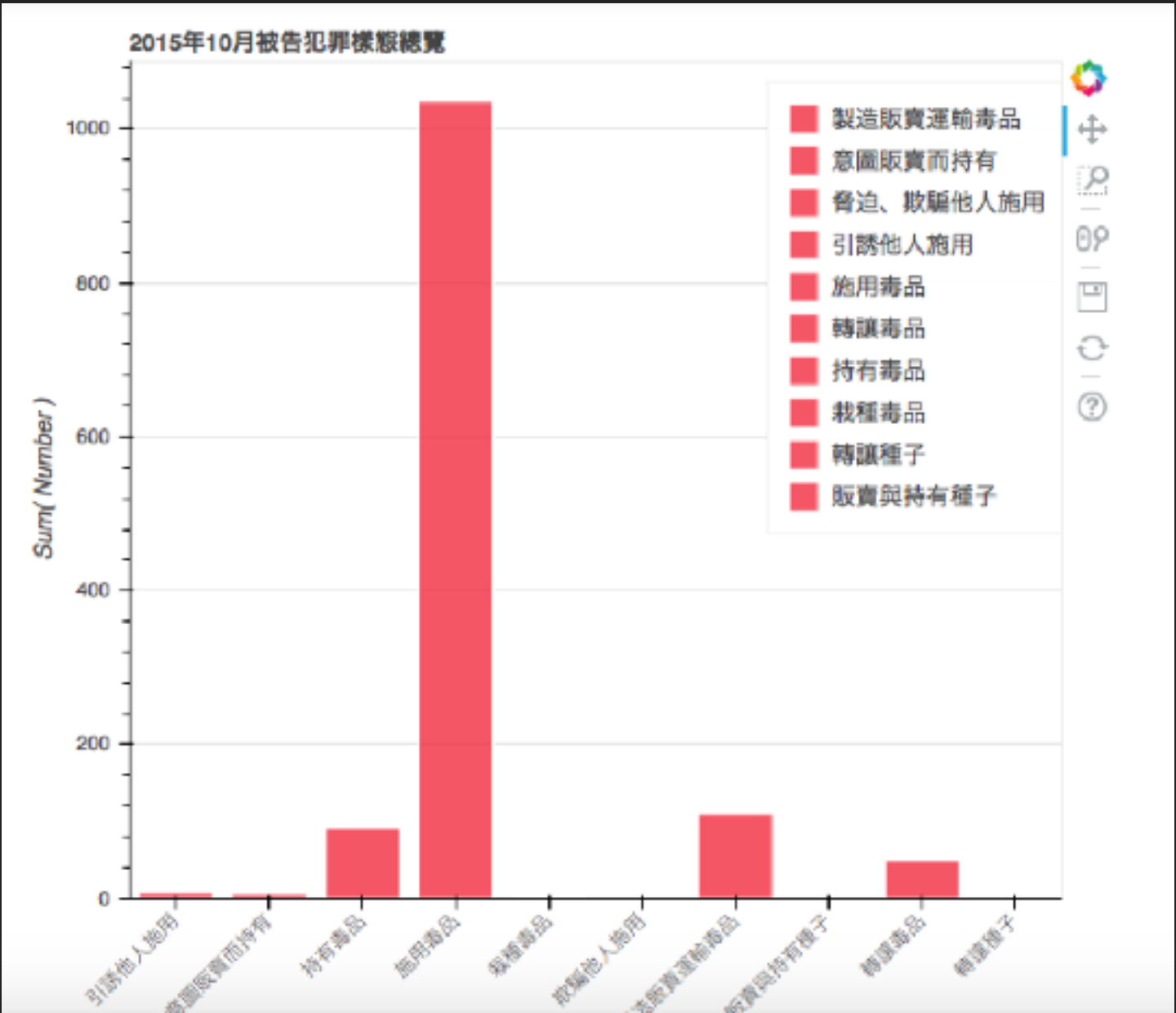
探索式分析

判決書概覽

- ▶ 本資料集內含有1018份判決,
5份為上訴到高院的裁判書;
裁判結果有16份為拘役;
刑度資料遺失(或為亂碼)的共有3份;
公訴不受理的有1份;
免刑的為1份;
資料重複的有1份
- ▶ 判處**有期徒刑**的裁判書數量為**991**份,
涉案的被告有**1009**位。

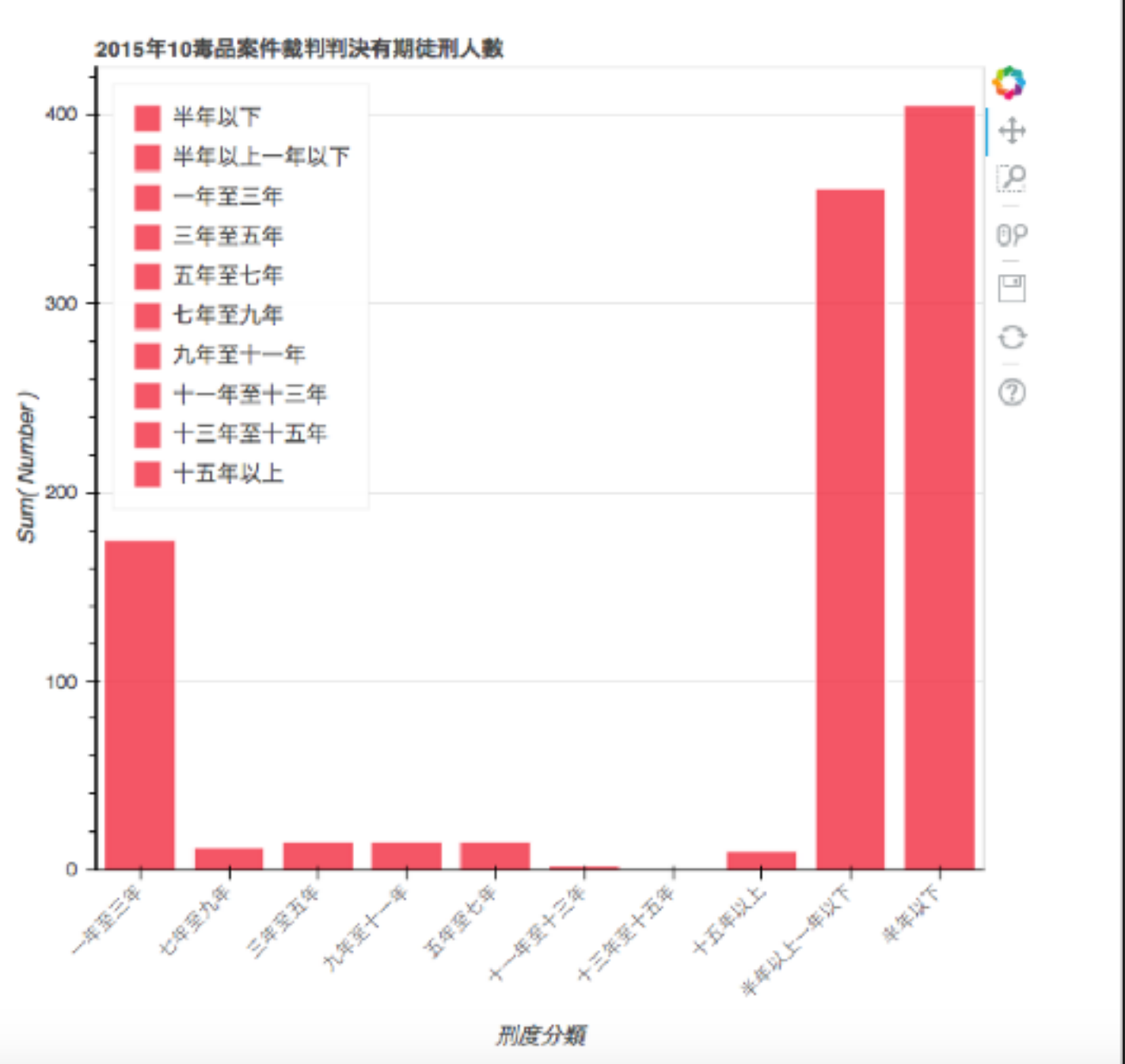
毒品犯罪樣態統計

- ▶ 製造、販賣、運輸毒品：
110位；
- ▶ 意圖販賣而持有共有6位引
誘他人的共有8位；
- ▶ 施用毒品的被告有1036位；
- ▶ 轉讓毒品的共有49位；
- ▶ 持有毒品共有91位；
- ▶ 栽種毒品植物的1位。



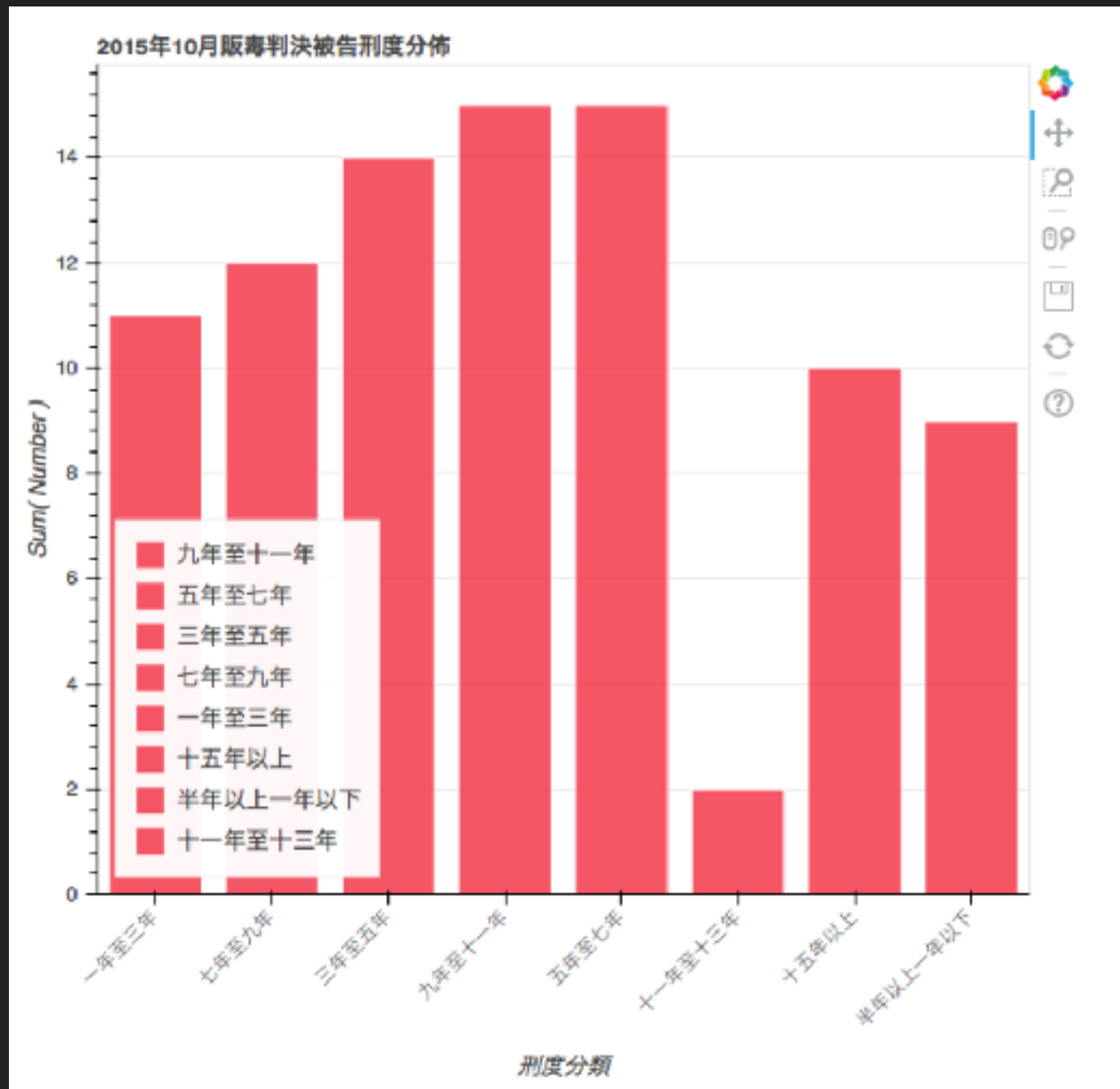
被告判決刑度分佈概覽

- ▶ 半年以下有期徒刑：405名；
- ▶ 半年以上至一年以下有期徒刑：361名
- ▶ 一年至三年：175名；
- ▶ 三年至五年：15位；
- ▶ 五年至七年：15位；
- ▶ 七至九年12年；
- ▶ 九年至十一年：15位；
- ▶ 十一至十三年：2位；
- ▶ 十五年以上：10位。



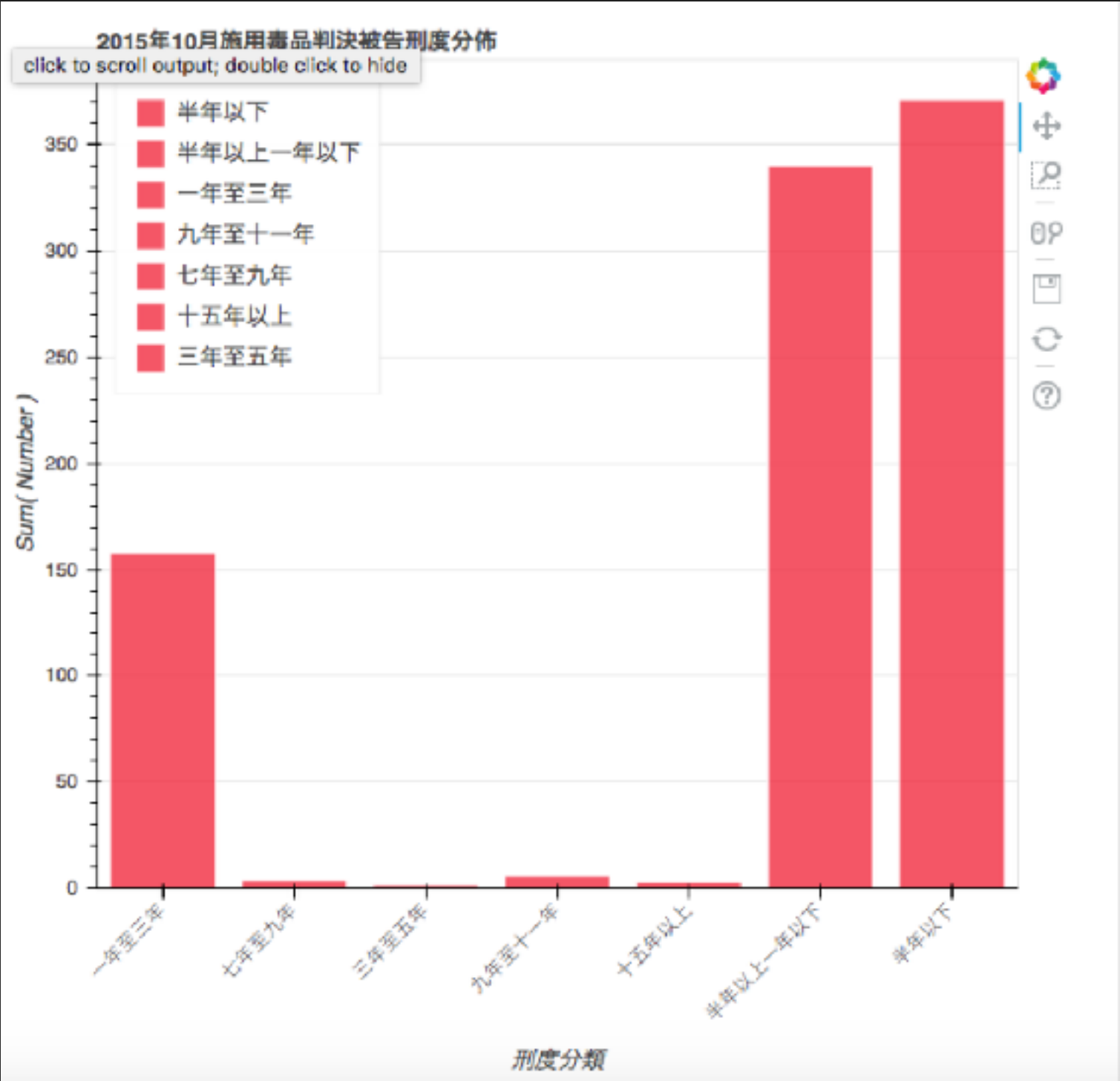
販毒被告判決刑度分佈

- ▶ 十五年以上有期徒刑者：15人；
- ▶ 十一至十三年：2人；
- ▶ 九至十一年：15人；
- ▶ 七至九年：12人；
- ▶ 五至七年：15人；
- ▶ 三至五年：14人；
- ▶ 一至三年：11人；
- ▶ 半年以上一年以下：9人。



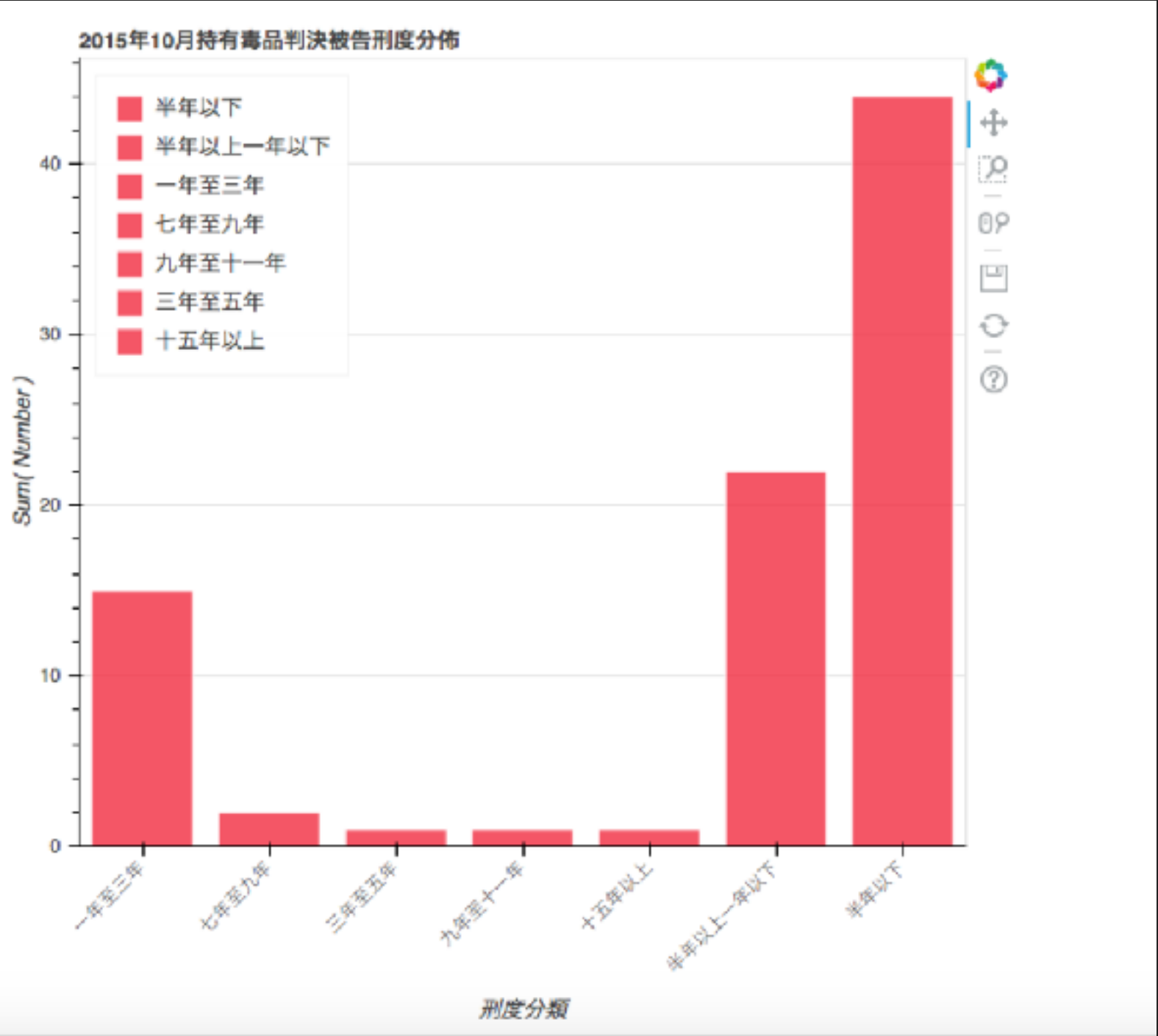
施用毒品被告判決分佈

- ▶ 半年以下共有371人,
- ▶ 半年以上一年以下：340人
- ▶ 一至三年：158人
- ▶ 三年至五年的有2人
- ▶ 七年至九年的有4人
- ▶ 九年至十一年的有6人
- ▶ 十五年以上的也有3人



持有毒品被告判決分佈

- ▶ 一年以內，半年以下：44位
- ▶ 半年以上一年以下：22位；
- ▶ 持一年至三年：15位；
- ▶ 三年至五年：1位；
- ▶ 七年至九年：2位；
- ▶ 九年至十一年：1位；
- ▶ 十五年以上：1位



決策樹

資料探勘演算法(一)

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose J48 -C 0.25 -M 2

Test options

☒ Use training set
☐ Supplied test set Set...
☐ Cross-validation Folds 10
☐ Percentage split % 66
 More options...

[None] 刑度分類4

Start Stop

Result list (right-click for options)

- 11:57:10 - rules.ZeroR
- 11:57:24 - trees.J48
- 11:57:40 - trees.J48
- 11:57:54 - trees.J48
- 11:58:08 - trees.J48

Classifier output

毒品危害防制條例第4條第6項 = 有
 累犯 = 是：一年至三年 (3.0)
 累犯 = 有：三年至五年 (2.0)
 毒品危害防制條例第11條第6項 = 有：一年以下 (2.0)

Number of Leaves : 36
 Size of the tree : 53

Time taken to build model: 0.04 seconds

--- Evaluation on training set ---

Time taken to test model on training data: 0.01 seconds

=== Summary ===

Correctly Classified Instances	884	87.5248 %
Incorrectly Classified Instances	126	12.4752 %
Kappa statistic	0.672	
Mean absolute error	0.0963	
Root mean squared error	0.2192	
Relative absolute error	46.6936 %	
Root relative squared error	70.6474 %	
Total Number of Instances	1010	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.945	0.159	0.928	0.946	0.933	0.711	0.922	0.959	一年以下
	0.594	0.648	0.759	0.594	0.667	0.613	0.894	0.845	一年至三年
	0.208	0.008	1.000	0.208	0.333	0.445	0.973	0.363	三年至五年
	0.963	0.031	0.634	0.963	0.765	0.757	0.974	0.524	五年以上
Weighted Avg.	0.675	0.264	0.678	0.675	0.669	0.693	0.921	0.876	

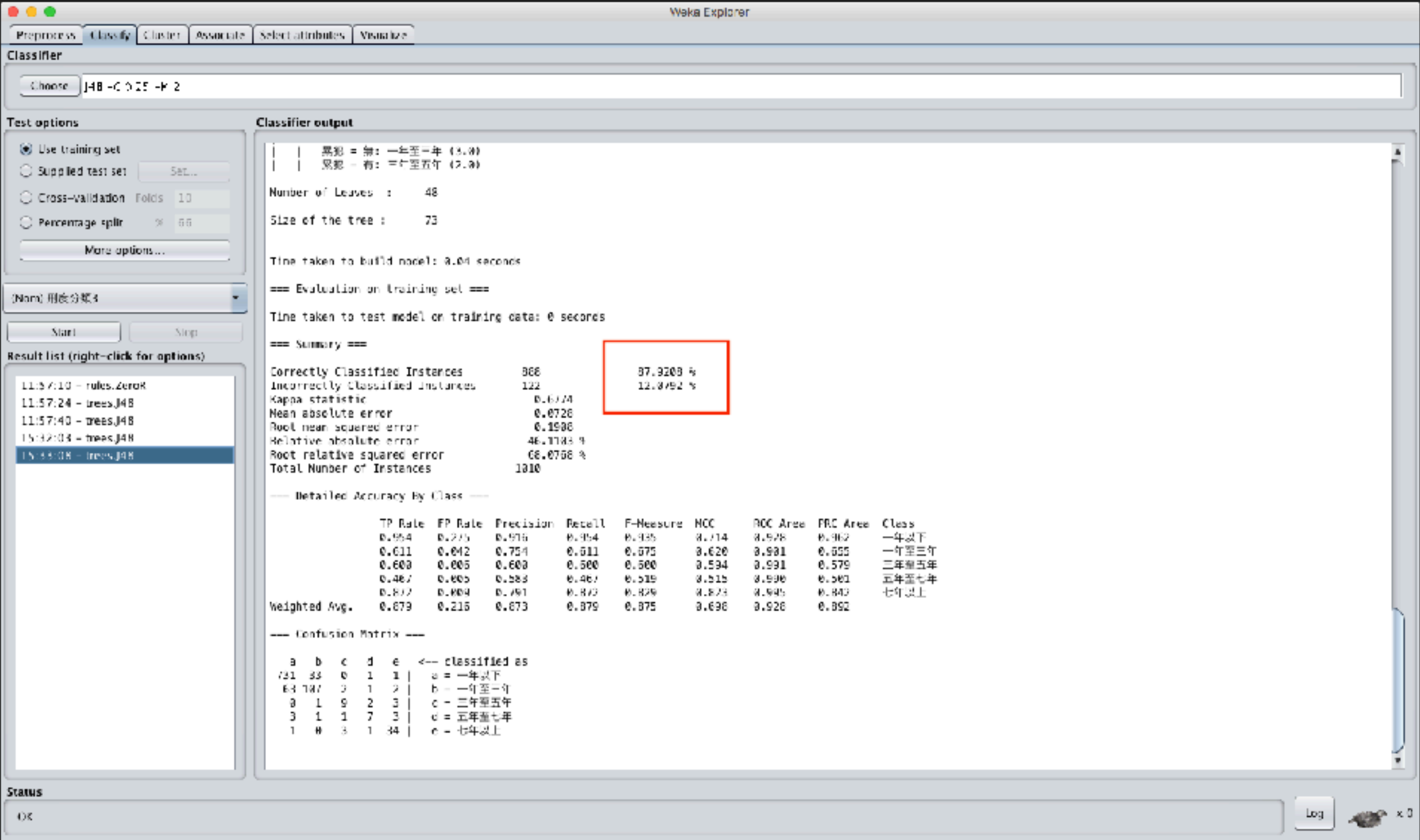
--- Confusion Matrix ---

	a	b	c	d	← classified as
73	33	0	8		a = 一年以下
61	104	0	19		b = 一年至三年
0	0	3	12		c = 三年至五年
7	0	0	57		d = 五年以上

Status

OK Log x 0

學術界的刑度

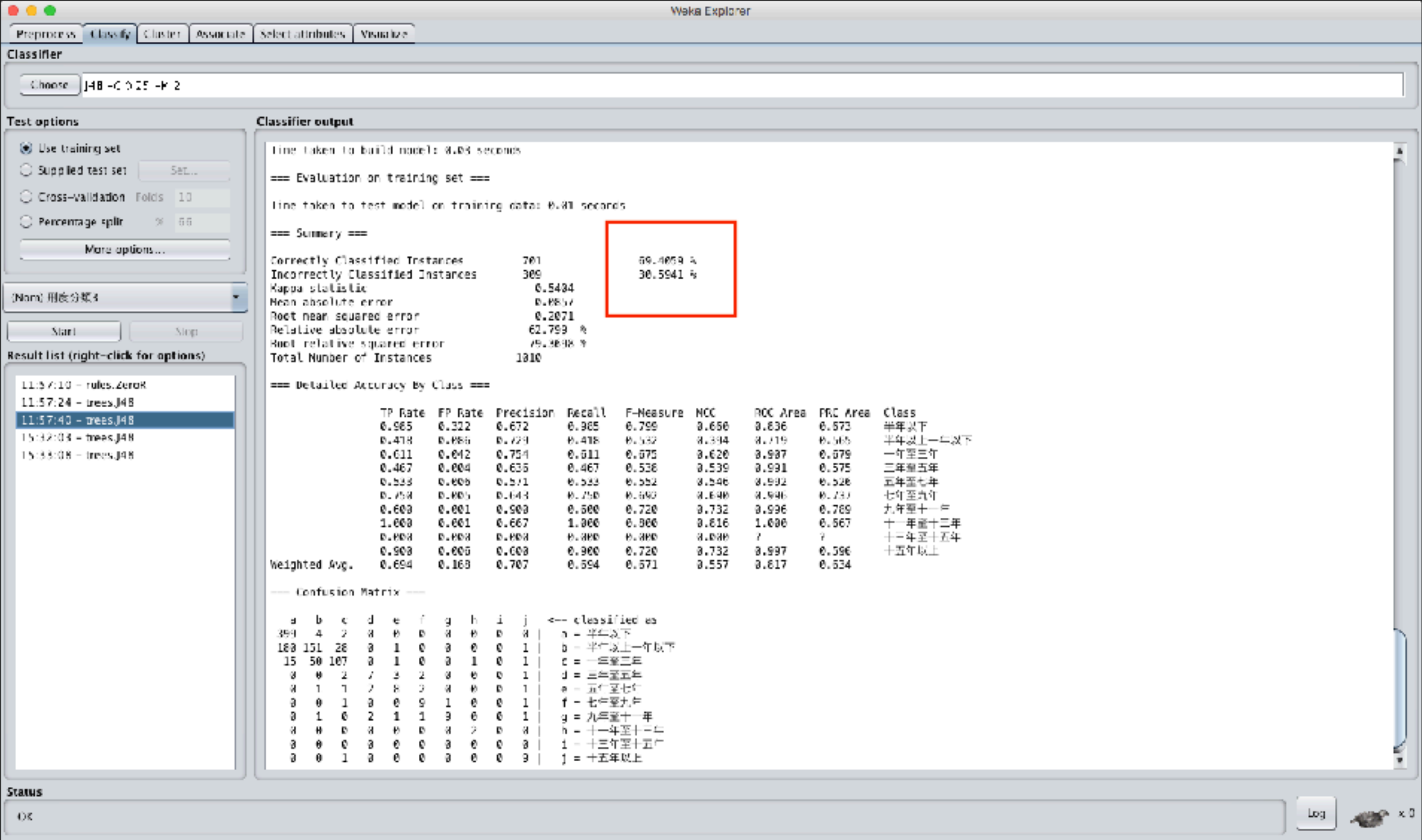


決策樹模型發現的條件規則

- ▶假如一個人吸食第一級毒品(毒品危害防制條例第10條第一項)且家境為小康，則他被判一年至三年的有期徒刑機率很高(80%)。
- ▶假如一個人吸食第一級毒品(毒品危害防制條例第10條第一項)又吸食第二級毒品(毒品危害防制條例第10條第二項)且為累犯，則有75%的機率被判一年至三年的有期徒刑。
- ▶假如一個人有在犯罪偵查中有自白(毒品危害防制條例17條第二項)且販賣行為非未遂，且是販賣第一級毒品，則他被判七年以上的有期徒刑仍高達88%。
- ▶假如一個人在犯罪偵查中有自白(毒品危害防制條例17條第二項)且販賣未遂，且非累犯，則他會被判一年至三年。

決策樹模型

參照司法法院的刑度



貝式網路

資料探勘演算法(二)

實務界的刑度

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier: Choose Naive Bayes

Test options

- ☒ Use training set
- ☐ Supplied test set
- ☐ Cross-validation Folds: 10
- ☐ Percentage split %: 66

More options...

[None] 刑度分類4

Start Stop

Result list (right-click for options)

- 11:57:10 - rules.ZeroR
- 11:57:24 - trees.J48
- 11:57:40 - trees.J48
- 15:32:04 - trees.J48
- 15:33:08 - trees.J48
- 17:31:12 - bayes.NaiveBayes
- 17:37:57 - bayes.NaiveBayes

Classifier output

高級中學
無
[total]

	1.0	1.0	1.0	2.0	1.0
	435.0	96.0	9.0	9.0	19.0
[total]	793.0	192.0	32.0	32.0	56.0

犯罪後的態度
坦承罪行
無
[total]

	447.0	105.0	14.0	11.0	31.0
	321.0	72.0	3.0	6.0	10.0
[total]	768.0	177.0	17.0	17.0	41.0

案犯
無
有
[total]

	278.0	39.0	11.0	12.0	24.0
	490.0	138.0	5.0	5.0	17.0
[total]	768.0	177.0	17.0	17.0	41.0

Time taken to build model: 0.01 seconds

— Evaluation on training set —

Time taken to test model on training data: 0.06 seconds

— Summary —

Correctly Classified Instances	787	77.9208 %
Incorrectly Classified Instances	223	22.0792 %
Kappa statistic	0.3641	
Mean absolute error	0.0957	
Root mean squared error	0.2326	
Relative absolute error	61.2456 %	
Root relative squared error	62.992 %	
Total Number of Instances	1010	

— Detailed Accuracy By Class —

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC	ROC Area	PRC Area	Class
	0.931	0.553	0.841	0.931	0.884	0.440	0.877	0.954	一年以下
	0.183	0.054	0.415	0.183	0.254	0.134	0.686	0.420	一年至三年
	0.267	0.005	0.444	0.267	0.323	0.237	0.983	0.419	三年至五年
	0.333	0.005	0.508	0.333	0.400	0.401	0.975	0.396	五年至七年
	0.645	0.034	0.509	0.645	0.529	0.633	0.986	0.777	七年以上
Weighted Avg.	0.779	0.430	0.743	0.779	0.749	0.401	0.672	0.839	

— Confusion Matrix —

Status

OK Log x 0

學術界的刑度

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier: Choose NaiveBayes

Test options

- ☒ Use training set
- ☐ Supplied test set
- ☐ Cross-validation Folds: 10
- ☐ Percentage split %: 66

More options...

[Name] 刑度分類?

Start Stop

Result list (right-click for options)

- 11:37:10 - rules.ZeroR
- 11:57:24 - trees.J48
- 11:57:40 - trees.J48
- 15:32:04 - trees.J48
- 15:33:08 - trees.J48
- 17:31:12 - bayes.NaiveBayes

Classifier output

Time taken to build model: 0.02 seconds

--- Evaluation on training set ---

Time taken to test model on training data: 0.09 seconds

=== Summary ===

Correctly Classified Instances	794	78.6134 %
Incorrectly Classified Instances	216	21.3861 %
Kappa statistic	0.364	
Mean absolute error	0.1758	
Root mean squared error	0.2591	
Relative absolute error	59.4253 %	
Root relative squared error	82.783 %	
Total Number of Instances	1010	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	NCC	ROC Area	PRC Area	Class
	0.930	0.549	0.642	0.930	0.863	0.441	0.677	0.354	一年以下
	0.183	0.053	0.421	0.183	0.255	0.197	0.881	0.418	一年至三年
	0.067	0.007	0.333	0.067	0.117	0.144	0.978	0.793	三年至五年
	0.907	0.039	0.575	0.907	0.705	0.705	0.996	0.766	五年以上
Weighted Avg.	0.785	0.428	0.747	0.786	0.753	0.407	0.671	0.642	

--- Confusion Matrix ---

	a	b	c	d	← classified as
172	42	1	11		a = 一年以下
130	32	1	12		b = 一年至三年
0	1	1	13		c = 三年至五年
4	1	0	49		d = 五年以上

Status

OK Log x 0

參照司法法院的刑度

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier: Choose NaiveBayes

Test options

- ☒ Use training set
- ☐ Supplied test set Set...
- ☐ Cross-validation Folds 10
- ☐ Percentage split % 66
- More options...

(None) 用度分類

Start Stop

Result list (right-click for options)

- 11:37:10 - rules.ZeroR
- 11:57:24 - trees.J48
- 11:57:40 - trees.J48
- 15:32:04 - trees.J48
- 15:33:08 - trees.J48
- 17:31:12 - bayes.NaiveBayes
- 17:37:57 - bayes.NaiveBayes
- 17:45:14 - bayes.NaiveBayes

Classifier output

=== Evaluation on training set ===

Time taken to test model on training data: 0.07 seconds

--- Summary ---

Correctly Classified Instances	515	50.9911 %
Incorrectly Classified Instances	495	49.1089 %
Kappa statistic	0.4	
Mean absolute error	0.0915	
Root mean squared error	0.2194	
Relative absolute error	66.9853 %	
Root relative squared error	84.1054 %	
Total Number of Instances	1010	

--- Detailed Accuracy By Class ---

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.951	0.304	0.677	0.951	0.797	0.639	0.845	0.817	半年以下
	0.465	0.228	0.542	0.465	0.512	0.264	0.730	0.566	半年以上一年以下
	0.137	0.019	0.600	0.137	0.223	0.229	0.637	0.506	一年至三年
	0.533	0.073	0.381	0.533	0.444	0.441	0.982	0.413	三年至五年
	0.000	0.014	0.391	0.000	0.174	0.475	0.982	0.424	五年至七年
	0.003	0.002	0.333	0.003	0.133	0.162	0.986	0.403	七年至九年
	0.733	0.015	0.423	0.733	0.527	0.549	0.981	0.569	九年至十一年
	0.000	0.000	0.000	0.000	0.000	0.000	1.000	1.000	十一年至十二年
	0.000	0.000	0.000	0.000	0.000	?	?	?	十二年至十五年
	0.200	0.003	0.400	0.200	0.267	0.278	0.994	0.477	十五年以上
Weighted Avg.	0.604	0.207	0.595	0.609	0.564	0.417	0.832	0.548	

=== Confusion Matrix ===

a	b	c	d	e	f	g	h	i	j	== classified as ==
365	10	1	1	0	0	0	0	0	0	a = 半年以下
171	175	6	1	3	0	3	0	0	0	b = 半年以上一年以下
73	176	24	6	3	0	1	0	0	2	c = 一年至三年
0	0	1	8	4	1	0	0	0	1	d = 三年至五年
0	2	0	1	9	0	3	0	0	0	e = 五年至七年
0	0	2	2	2	1	4	0	0	0	f = 七年至九年
0	1	0	2	1	0	11	0	0	0	g = 九年至十一年
0	0	0	0	0	0	2	0	0	0	h = 十一年至十二年
0	0	0	0	0	0	0	0	0	0	i = 十二年至十五年
0	1	4	0	0	1	2	0	0	2	j = 十五年以上

Status

OK

Log x 0

RANDOM FOREST

資料探勘演算法(三)

SHARE



1K



127



778



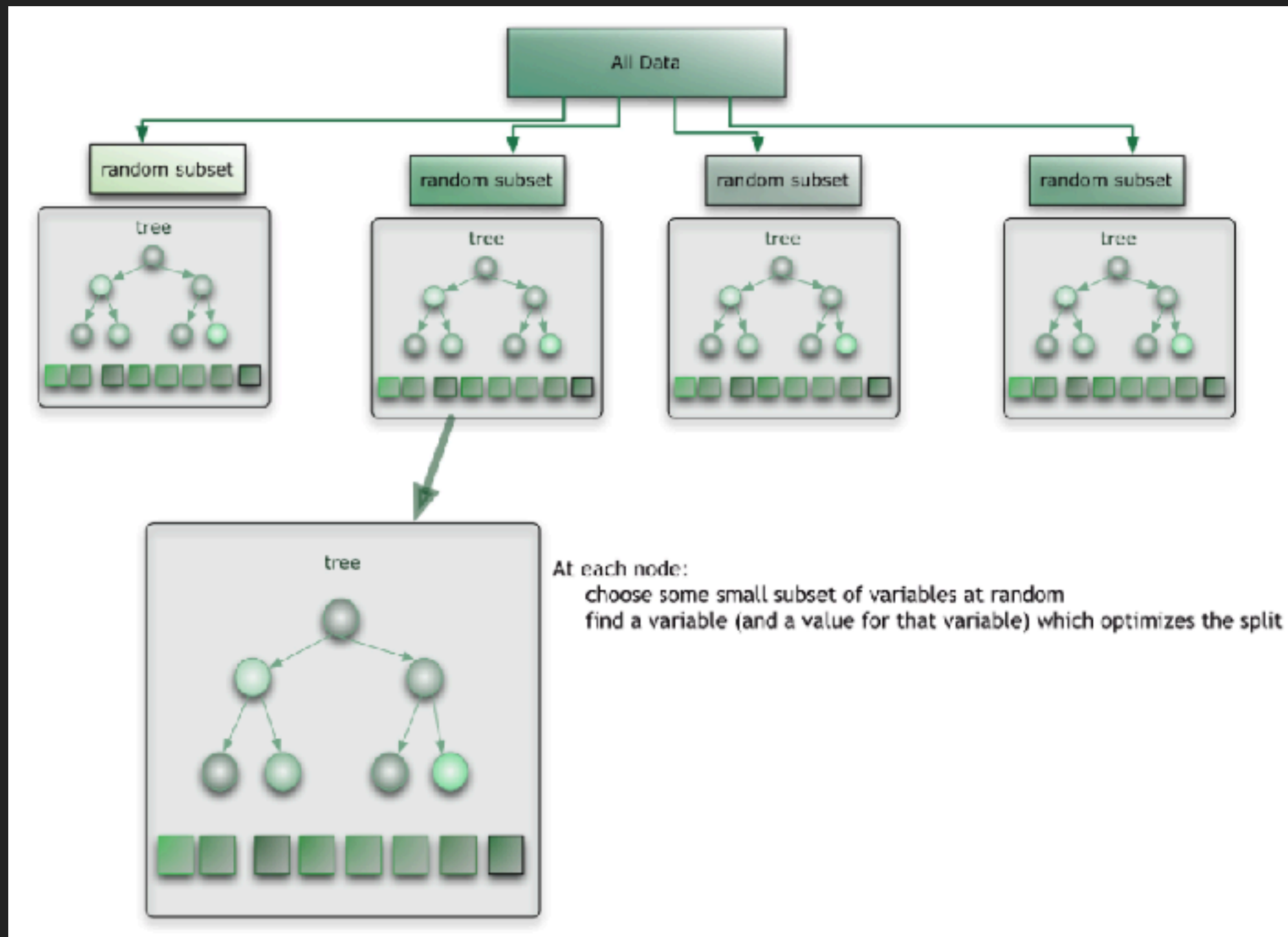
Artificial intelligence can predict Supreme Court decisions better than some experts.

davidevason/iStockphoto

Artificial intelligence prevails at predicting Supreme Court decisions

By **Matthew Hutson** | May 2, 2017, 1:45 PM

RANDOM FOREST



RANDOM FOREST

學術界的刑度

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose RandomForest -F 100 - 100 -num-slots 1 -K 0 -M 1.0 -N 0.001 -S 1

Test options

☒ Use training set
☐ Supplied test set Set...
☐ Cross-validation Folds 10
☐ Percentage split % 66
More options...

[None] 用度分類?

Start Stop

Result list (right-click for options)

- 11:37:10 - rules.ZeroR
- 11:57:24 - trees.J48
- 11:57:40 - trees.J48
- 15:12:04 - trees.J48
- 15:33:08 - trees.J48
- 17:31:12 - bayes.NaiveBayes
- 17:37:57 - bayes.NaiveBayes
- 17:45:14 - bayes.NaiveBayes
- 17:53:14 - trees.RandomForest
- 17:59:02 - trees.RandomForest

Classifier output

Test nodes: evaluate on training data

=== Classifier model (full training set) ===

RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -N 0.001 -S 1 -no-rot-check-capabilities

Time taken to build model: 0.34 seconds

--- Evaluation on training set ---

Time taken to test model on training data: 0.12 seconds

=== Summary ===

Correctly Classified Instances	435	92.5743 %
Incorrectly Classified Instances	75	7.4257 %
Kappa statistic	0.8636	
Mean absolute error	0.068%	
Root mean squared error	0.1787	
Relative absolute error	34.8381 %	
Root relative squared error	54.5519 %	
Total Number of Instances	1410	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	NCC	RCC Area	PRC Area	Class
	0.978	0.183	0.942	0.970	0.956	0.609	0.972	0.990	一年以下
	0.726	0.028	0.847	0.726	0.782	0.743	0.963	0.951	一年至三年
	0.804	0.001	0.903	0.800	0.857	0.857	0.999	0.803	三年至五年
	0.961	0.005	0.914	0.961	0.946	0.944	1.000	0.992	五年以上
Weighted Avg.	0.925	0.148	0.923	0.926	0.924	0.806	0.972	0.964	

--- Confusion Matrix ---

a	b	c	d	← classified as
143	73	0	4	a = 一年以下
45	127	0	2	b = 一年至三年
0	0	12	3	c = 三年至五年
4	0	7	54	d = 五年以上

Status

OK Log x 0

RANDOM FOREST

實務界的刑度

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose RandomForest -F 100 - 100 -num-slots 1 -K 0 -M 1.0 -N 0.001 -S 1

Test options

☒ Use training set
☐ Supplied test set Set...
☐ Cross-validation Folds 10
☐ Percentage split % 66
More options...

[None] 刑度分類4

Start Stop

Result list (right-click for options)

- 11:37:24 - trees.J48
- 11:57:40 - trees.J48
- 15:32:03 - trees.J48
- 15:33:08 - trees.J48
- 17:31:12 - bayes.NaiveBayes
- 17:37:57 - bayes.NaiveBayes
- 17:45:13 - bayes.NaiveBayes
- 17:53:14 - trees.RandomForest
- 17:59:07 - trees.RandomForest
- 18:01:25 - trees.RandomForest
- 18:02:13 - trees.RandomForest

Classifier output

=== Classifier model (full training set) ===

RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -N 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.3 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0.13 seconds

=== Summary ===

Correctly Classified Instances	932	92.2772 %
Incorrectly Classified Instances	78	7.7228 %
Kappa statistic	0.796	
Mean absolute error	0.057	
Root mean squared error	0.1553	
Relative absolute error	36.8657 %	
Root relative squared error	55.398 %	
Total Number of Instances	1010	

--- Detailed Accuracy By Class ---

	TP Rate	FP Rate	Precision	Recall	F-Measure	NCC	ROC Area	FRC Area	Class
	0.971	0.154	0.941	0.971	0.956	0.889	0.972	0.990	一年以下
	0.728	0.025	0.851	0.720	0.780	0.742	0.953	0.852	一年至三年
	0.608	0.001	0.923	0.600	0.857	0.657	0.999	0.863	三年至五年
	0.808	0.002	0.857	0.800	0.828	0.826	0.999	0.934	五年至七年
	0.974	0.006	0.864	0.974	0.916	0.914	0.999	0.984	七年以上
Weighted Avg.	0.923	0.151	0.921	0.923	0.920	0.602	0.973	0.963	

--- Confusion Matrix ---

a	b	c	d	e	← classified as
744	22	0	0	0	a = 一年以下
47	726	0	1	1	b = 一年至三年
0	0	12	1	2	c = 三年至五年
0	0	0	12	3	d = 五年至七年
0	0	1	0	38	e = 七年以上

Status

OK Log x 0

RANDOM FOREST

參照司法法院的刑度

Preprocess

Classify

Cluster

Associate

Select Attributes

Visualize

Classifier

Choose RandomForest -F 0.0 -I 100 -n-imp-dim 1 -K 0 -M 1.0 -N 0.001 -S 1

Test options

☒ Use training set

☐ Supplied test set

☐ Cross-validation Folds 10

☐ Percentage split % 55

More options...

(None) 迅速分類

Start Stop

Result list (right-click for options)

11:57:10 - rules.ZeroK

11:57:24 - trees.J48

11:57:40 - trees.J48

15:32:03 - trees.J48

15:33:08 - trees.J48

17:31:12 - bayes.NaiveBayes

17:37:57 - bayes.NaiveBayes

17:45:13 - bayes.NaiveBayes

17:53:14 - trees.RandomForest

Classifier output

Classifier model (full training set)

RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -N 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.8 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0.25 seconds

=== Summary ===

Correctly Classified Instances 791 78.3166 %

Incorrectly Classified Instances 219 21.6832 %

Kappa statistic 0.6803

Mean absolute error 0.0544

Root mean squared error 0.1596

Relative absolute error 47.1319 %

Root relative squared error 65.8737 %

Total Number of Instances 1010

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.778	0.088	0.855	0.778	0.815	0.793	0.949	0.917	半年以下
	0.889	0.202	0.689	0.889	0.742	0.585	0.891	0.812	半年以上一年以下
	0.726	0.029	0.947	0.726	0.782	0.743	0.963	0.851	一年至三年
	0.880	0.001	0.923	0.880	0.897	0.857	0.999	0.863	三年至五年
	0.857	0.004	0.765	0.857	0.813	0.811	0.999	0.934	五年至七年
	0.843	0.002	0.833	0.843	0.833	0.841	0.999	0.945	七年至九年
	0.857	0.004	0.765	0.857	0.813	0.811	0.999	0.923	九年至十一年
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	十一年至十五年
	0.000	0.000	0.000	0.000	0.000	0.000	?	?	十二年至十五年
	0.900	0.001	0.900	0.900	0.900	0.899	1.000	0.973	十五年以上
Weighted Avg.	0.783	0.112	0.793	0.783	0.785	0.677	0.935	0.868	

=== Confusion Matrix ===

a	b	c	d	e	f	g	h	i	j	s← classified as
315	89	1	0	0	0	0	0	0	0	a = 半年以下
49	290	22	0	0	0	0	0	0	0	b = 半年以上一年以下

Status

OK

Log

x 0

如何讓模型變更好

討論

決策樹模型發現的條件規則

▶一、提高文字探勘特徵值的質量：

使用正規表示式非最佳的IR方式，理想上是需要搭配NLP進行語意分析，並訓練一個模型能夠抓取判決書的量刑因子。

▶二、應納入更多量刑的因素：

毒品犯罪的量刑除了從犯罪行為、智識程度、生活狀況、累犯、自首、犯後態度以外，還有許多關鍵的要素例如**犯罪人的品行**、**犯罪人的年齡**、該次犯罪查獲的毒品重量、犯罪人是**正犯**、**共犯**還是**教唆犯**、犯罪人除了觸犯毒品罪是否還有觸犯其他法律如槍械等等。

▶三、犯罪行為的吸收：

高度的犯罪行為會吸收低度的犯罪行為，例如販毒行為，在同一案件中，會吸收其他低度的犯罪行為如持有毒品。

Q&A