AIS3

教育部先進資通安全實務人才培育計畫

# 111年度新型態資安實務暑期課程

Advanced Information Security Summer School

## Anti-fraud Recognition Using Machine Learning

跨域金融資訊安全 – 第一組

詹侑晟、王念祖、張茲涵、陳勝舢

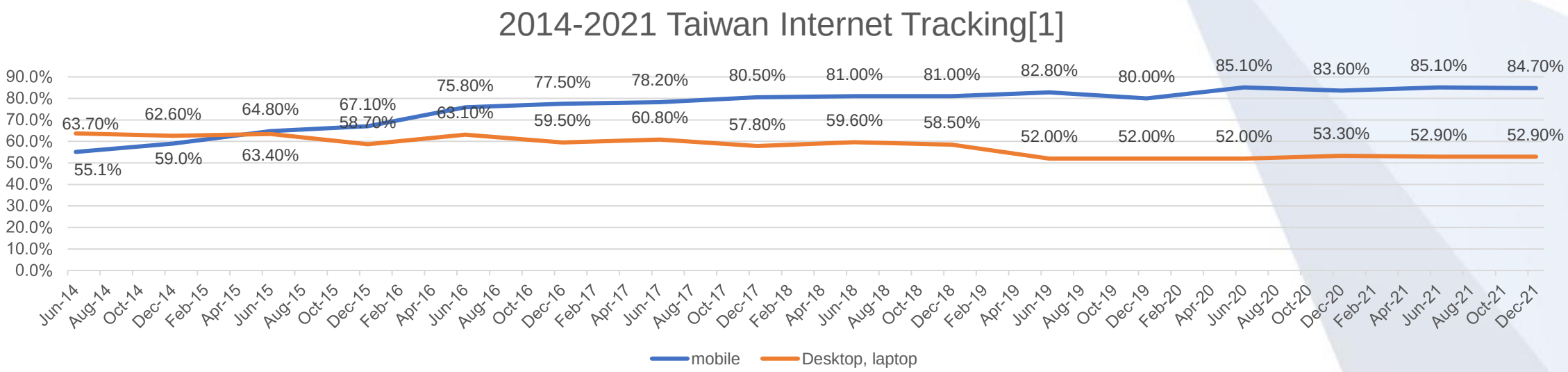
# Outline

- Motivation
- Background
  - Anti-Fraud
  - Phishing
  - Machine Learing
  - Django
- Issues
- Problem statement
  - Fraud Detection
  - Safe URL
  - Web API
- Related work
- Solution approach
  - System Structure
  - Data collection
  - Datasets
  - NLP
  - Data Preprocess
  - Data Analysis
  - SVM
  - Confusion Matrix
  - Prediction
  - Safe Browsing API
  - Web Design and Security
- Demo
- Conclusion
- Reference

# Motivation

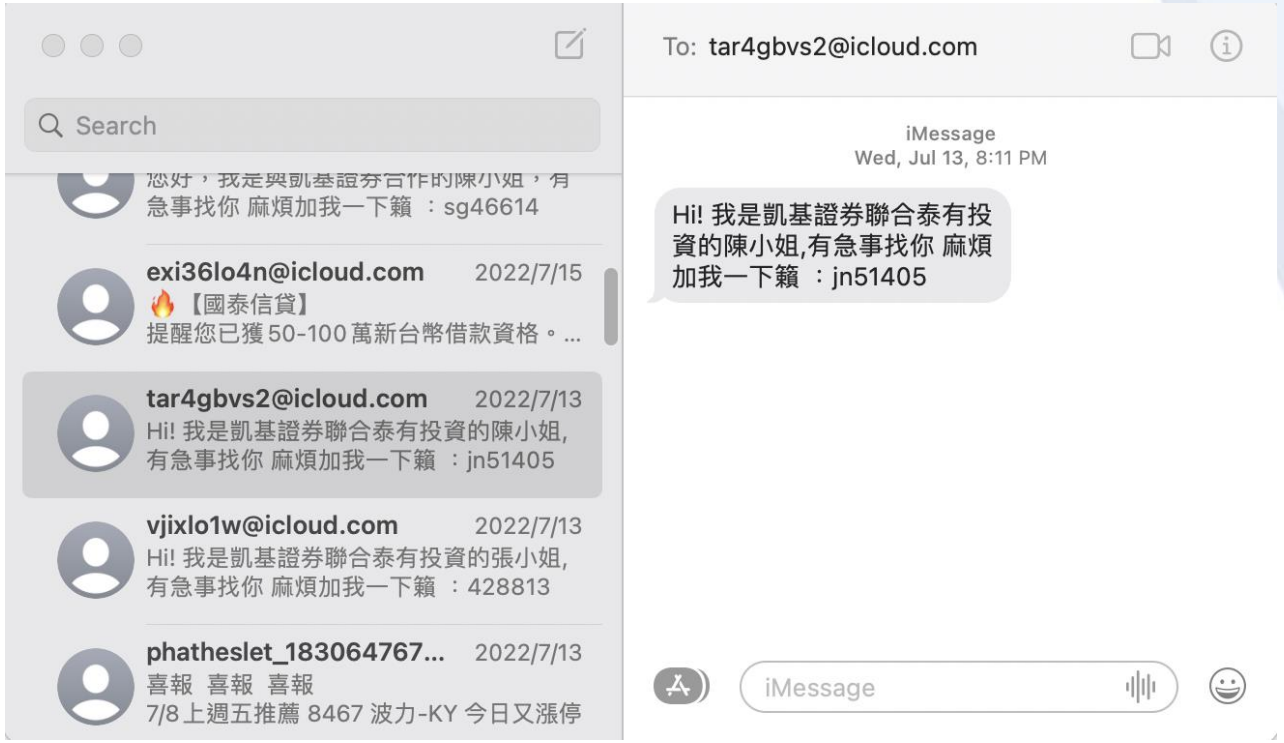
# Motivation - 1

Each year the internet grows, but 2021 is a special year. The COVID-19 continued for some time, keeping people at home and making us rely on digital technology more than we ever had before.



# Motivation - 2

"If there is an emergency, please add LINE!" The bombing of scam text messages

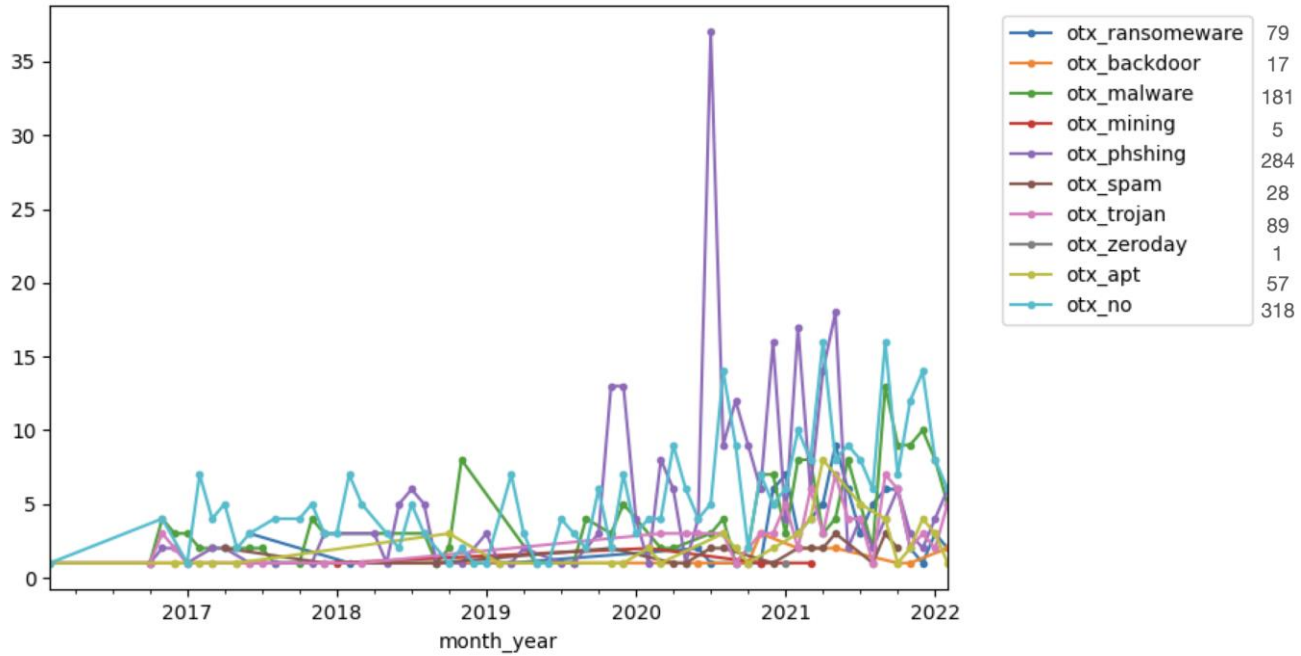



Fraud SMS source by us

# Motivation - 3

Fraud cases from January to October 2021[2]

- investment fraud 2,168 pieces, nearly doubling from the same period last year 94.61%




2016-2022 Finance OSINT Tag source from CyberTotal in Cycraft


Fake URL

# Background

# Background – Anti-fraud

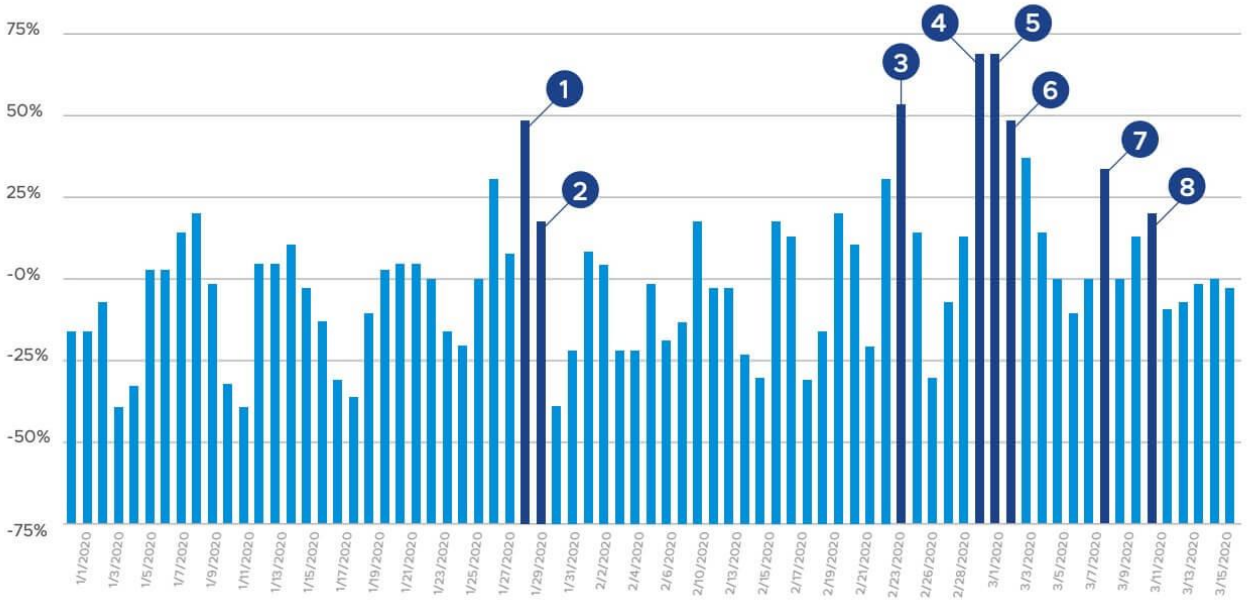
The persistently high volume of fraud may become the "new normal of fraud" in the post-pandemic era.


Fraud case ranking[3]


Fraud message trends[4]

# Background – Phishing

Phishing is a tactic where the attacker poses as a legitimate individual or service to gather sensitive information such as credentials or payment card information, or install malware on the target’ s device.




1. First United States confirmed case.

2. President Trump announced entry ban on foreign nationals.

3. Announcement that 2,400+ have died globally from COVID-19 related illnesses.

4. First death in the United States.

5. Florida declares public health emergency, followed by several other states.

6. Italy COVID-19 cases surpass 2,000.

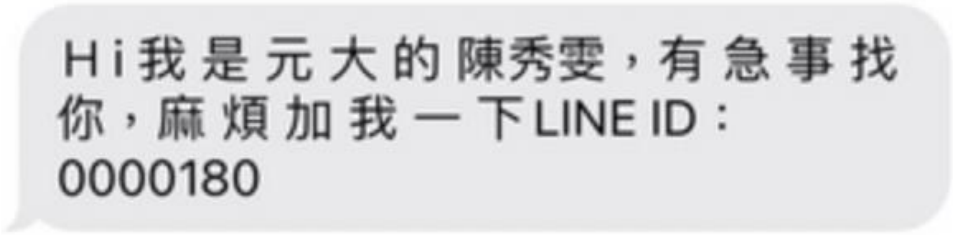
7. Italy starts lockdown.

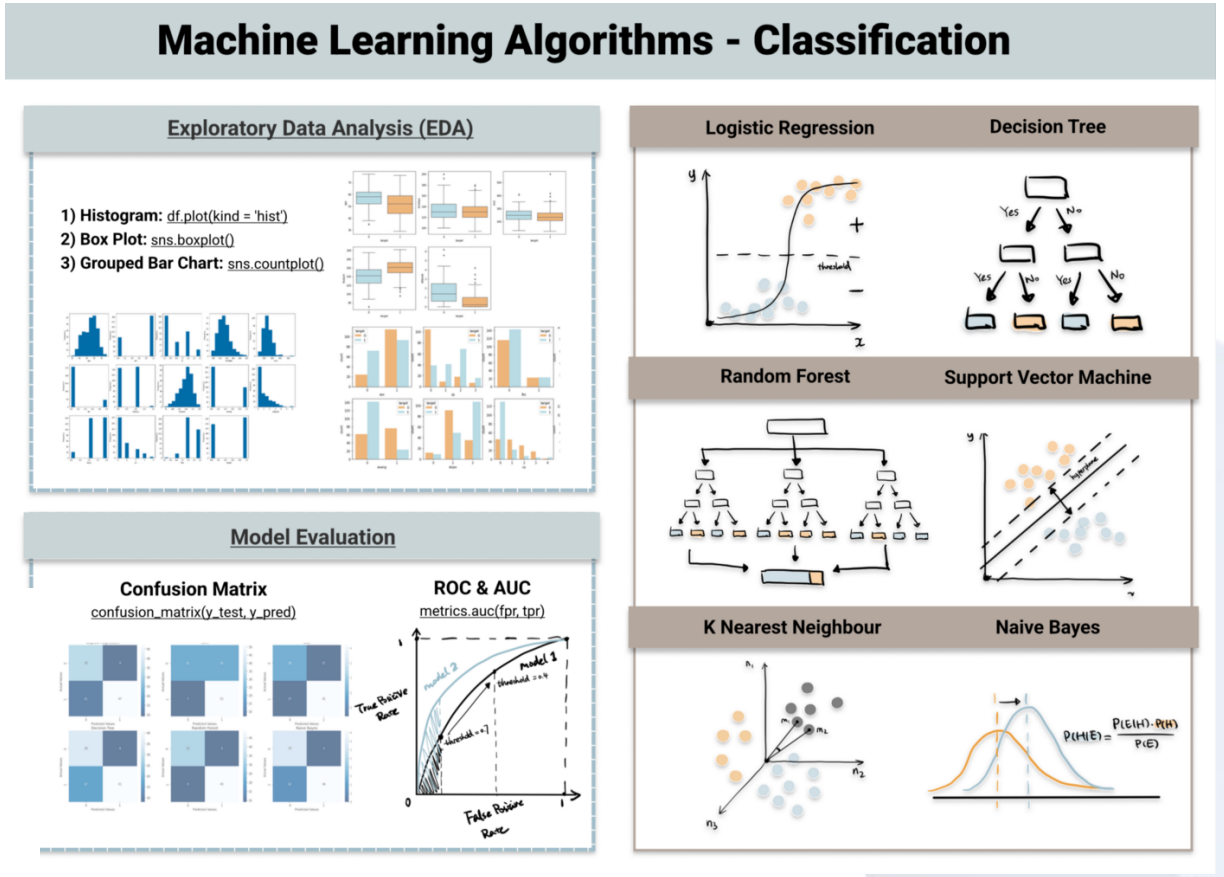
8. WHO declares a pandemic.

Relative phishing attack event percentage changes for notable alerts - Data Source VMware Carbon Black Data

# Background – Machine Learning

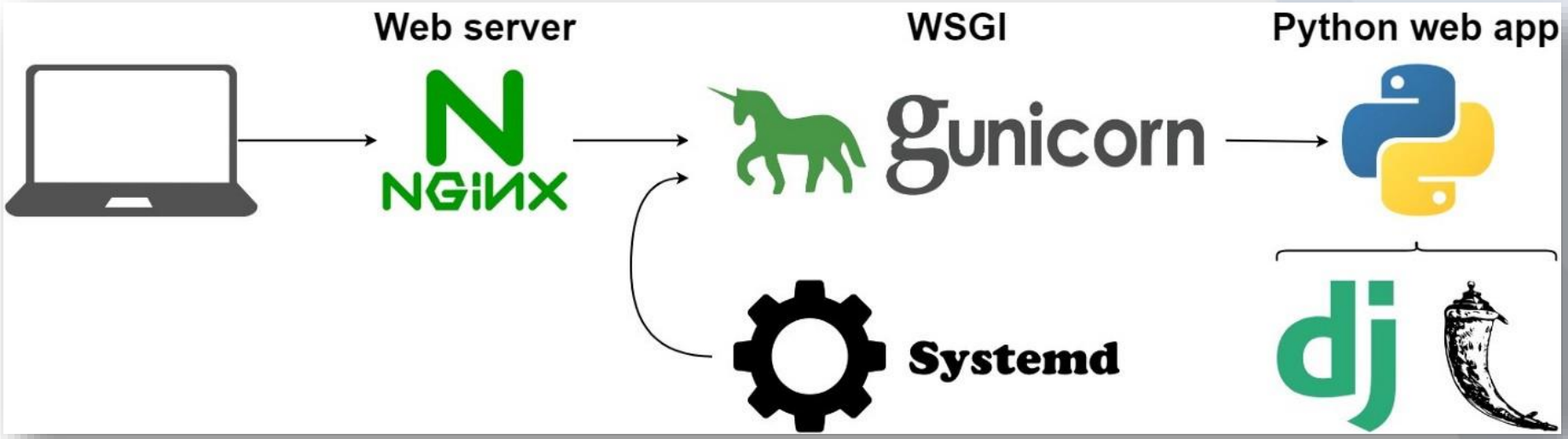
Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.



# Background – Django

Django is a high-level Python web framework that encourages rapid development and clean, pragmatic design.



# Issues

# Issues

1. How to avoid fraud messages?
2. How to improve the ineffectiveness of keyword blocking for hacker organizations?
3. What to do if there are malicious links?

# Problem statement

# Problem statement

- Input
  - Suspected message
- Objective
  - Find fraud messages
  - Scan URL whether is safe or not
- Output
  - Fraud or Not Fraud

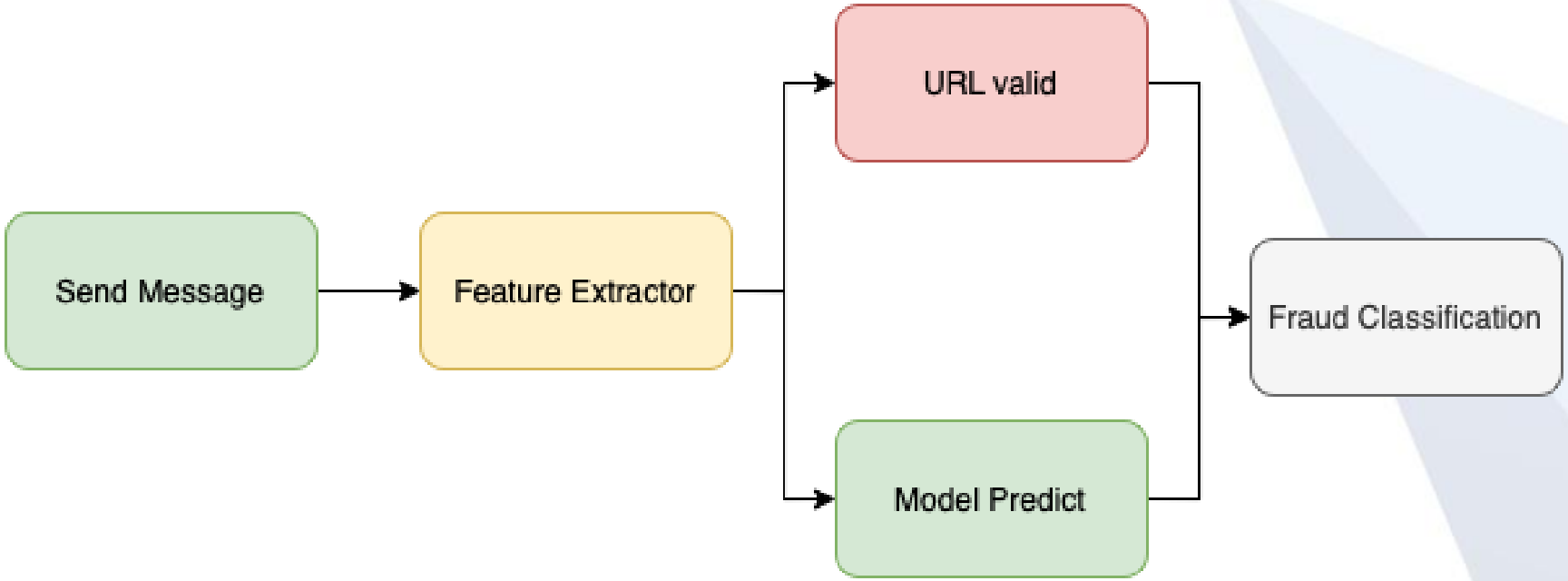
# Related work

# Relate work

| 論文、軟體或平台 | 來源 | 分類 | 技術 | 附加功能 |
|---|---|---|---|---|
| [5] | LINE、FB | 詐騙、非詐騙、待確認 | DistilBERT + RandomForest | X |
| [6] | SMS | 詐騙、非詐騙 | Keyword block | X |
| [7] | Social Community | 高風險 | 人工識別 | X |
| Our research | SMS、imessage | 詐騙、非詐騙 | NLP + SVM | 網址威脅檢查 |

# Solution approach

# System Structure



# Data collection

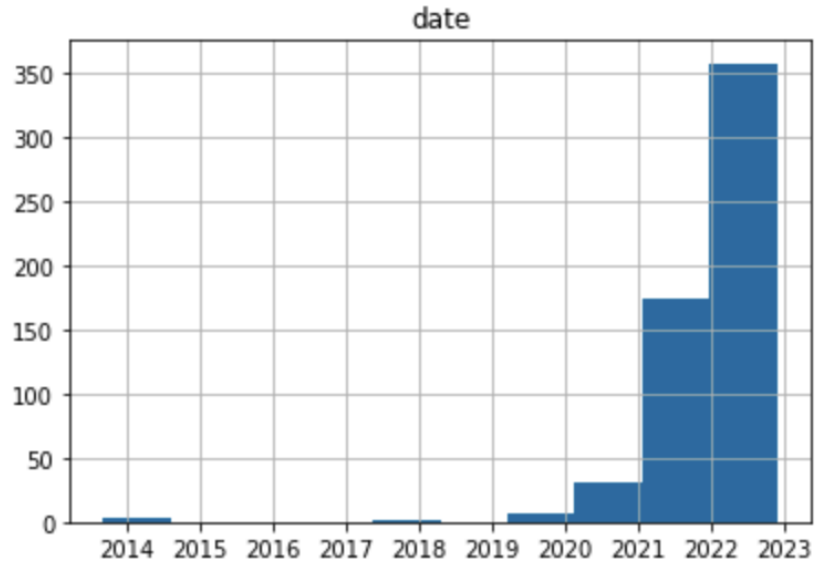
- Source:  the SMS or imessage from Team member's cell phone or computer, News or Social Community
- Method: Self-written package capture, Translators and OCR Tools

```python
for line in text:
        regex_line = re.search(r"[0-9:]{5} 陳勝舢.+", line)
        if regex_line != None:
            if message != "":
                formatted_data_set.append([sender, message, fixed_time_stamp, source, fixed_label])
                message = ""

            sender_sms_matches = re.search(r" [0-9+-]{10,}", line)
            sender_iMessage_matches = re.search(r" [a-zA-Z0-9-_]+@[a-zA-Z0-9-_]+.[a-zA-Z0-9-_]+", line)

            # sender info (+886901-234-567, abcd@mail.com, etc.)
            if sender_sms_matches != None:
                sender = sender_sms_matches.group()[1:]
                source = "簡訊"
            elif sender_iMessage_matches != None:
                sender = sender_iMessage_matches.group()[1:]
                source = "iMessage"
            else:
                message += regex_line.group()[10:]
        else:
            message += line

    return formatted_data_set
```

```python
def main():
    path = '.' + os.sep + 'image'
    lstFile = [f for f in listdir(path) if isfile(join(path, f))]

    for f in lstFile:
        if '.jpg' in f:
            idx = f.find('.jpg')
        elif '.png' in f:
            idx = f.find('.png')
        else:
            print('warning: 有檔案的副檔名不為.jpg或.png，所以不能轉譯為文字檔')
            continue
        out_name = ""
        for i in range(idx):
            out_name += f[i]
        print(out_name)
        text = ocrText('.' + os.sep + 'image' + os.sep + '{}'.format(f)) #執行OCR
        text = replaceText(text) #將轉出的文字檔做一些格式處理
        path = '.' + os.sep + 'text' + os.sep + out_name + '.txt'
        save(path, text) #將產出的txt存在text資料夾
```

# Datasets

- 非詐騙 295
- 詐騙 278

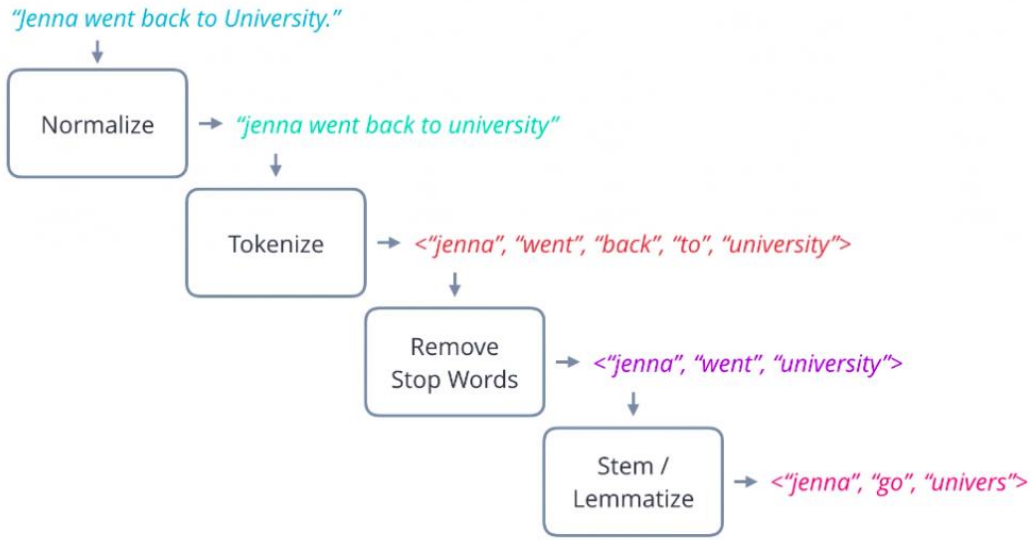



| | 寄件者 | 內文 | 時間 | 來源 | Label |
|---|---|---|---|---|---|
| 0 | f91636i@icloud.com | 受夠低薪、超時工作？還在過低薪窮忙的生活？本公司祭出超高時薪，歡迎想有效利用時間的夥伴加入我... | 2021-12-16T18:00:00 | imessage | 詐騙 |
| 1 | 885818 | [和潤企業車貸專案] 不限車齡只要您有汽機車馬上變現金... | 2022-07-15T14:05:00 | 簡訊 | 詐騙 |
| 2 | 0968382009 | 我之前跟您聯絡過的林小姐、記得加我賴喔~ https://line.me/ti/p/fp69... | 2022-05-27T20:08:00 | 簡訊 | 詐騙 |
| 3 | 0968382009 | 航海王即將啟航，添加賴領取明日飆股，名額有限，請盡快添加領取喔！https://line.m... | 2022-05-29T21:15:00 | 簡訊 | 詐騙 |
| 4 | 0903447032 | 【邀請您答題】3秒鐘答題，即可索取醫療險資料 http://imob.tw/OqN1d 免費索取 | 2022-05-26T14:15:00 | 簡訊 | 詐騙 |
| ... | ... | ... | ... | ... | ... |
| 569 | 886903448294 | 「息」比三家不吃虧，先問再借免緊張！備支票300萬內，榮泰當鋪給您【打破行情超低利專案】正派... | 2021-11-29T09:04:00 | 簡訊 | 詐騙 |
| 570 | 886903448294 | 【跟銀行一樣的利率】毅成租賃代辦銀行貸款、工程款、股票代墊，個人、企業500萬內，免押保，當... | 2021-11-29T09:04:00 | 簡訊 | 詐騙 |
| 571 | 886903448294 | 其實您還有更好選擇!【昌益租賃】新辦貸款或代還民間高利，單一窗口保密佳，500萬內備支票當日... | 2021-11-25T09:12:00 | 簡訊 | 詐騙 |
| 572 | 886903448294 | 2.88%超低利專案，大小額公司借款、個人借貸、工程款代墊，審核快速，當日放款，500萬內備... | 2021-11-23T09:27:00 | 簡訊 | 詐騙 |
| 573 | 886903448294 | 別讓債務拖垮你!【寶發理財】專辦銀行貸款、整合債務、代償高利、350萬內資金當日到位，利率最... | 2021-11-17T09:16:00 | 簡訊 | 詐騙 |

574 rows × 5 columns

# Word Segmentation – NLP

Natural language processing (NLP) refers to the branch of computer science—and more specifically, the branch of artificial intelligence or AI—concerned with giving computers the ability to understand text and spoken words in much the same way human beings can.



# Data Preprocess - 1

CKIP (Chinese Knowledge and Information Processing)

```
句子分詞後：
 ['受 夠 低薪 、 超時 工作 ？ 還 在 過 低薪 窮忙 的 生活 ？ 本 公司 祭出 超高時薪 ， 歡迎 想 有效 利用 時間 的 夥伴 加入 我',
'花旗 商務卡 : 閣下 賬號 最後 數 \n 字 為 068001 花旗 賬戶 新結\n 單現 於 Citi Manager 可']
data_vector_value:
 [[0 0 0 2 1 1 1 0 0 1 1 0 1 0 1 1 1 1 1 0 0 0 1 1 0]
 [1 1 1 0 0 0 0 1 1 0 0 1 0 1 0 0 0 0 0 2 1 1 0 0 1]]
特徵名字：
 ['068001', 'citi', 'manager', '低薪', '公司', '利用', '加入', '商務卡', '單現', '夥伴',
'有效', '歡迎', '生活', '祭出', '窮忙', '花旗', '賬戶', '賬號', '超時', '超高時薪', '閣下']
```

```python
from sklearn.feature_extraction.text import CountVectorizer
from ckiptagger import data_utils ,WS,POS, NER

def cut_word(text):
    #進行中文分詞
    ws_results = ws([text])
    temp = ""
    for i in ws_results:
       temp=temp+ " ".join(i)
    return temp

df["content_cut"] = np.nan
for i in df["content"]:
    df.loc[df.content==str(i),'content_cut']=cut_word(str(i))
```

# Data Preprocess - 2

Convert a collection of text documents to a matrix of token counts.

```
1  from sklearn.feature_extraction.text import CountVectorizer
2  count_vect = CountVectorizer()
3  X_train_counts = count_vect.fit_transform(X_train)
4  print(X_train_counts.shape)
5  pd.DataFrame(X_train_counts)[0]
```

```
(383, 3207)

0        (0, 1024)\t1\n  (0, 3098)\t1\n  (0, 2936)\t1...
1        (0, 2936)\t1\n  (0, 1696)\t1\n  (0, 2981)\t1...
2        (0, 1937)\t1\n  (0, 2175)\t1\n  (0, 331)\t1\...
3        (0, 1024)\t1\n  (0, 3098)\t1\n  (0, 1937)\t1...
4        (0, 3080)\t1\n  (0, 2432)\t1\n  (0, 1393)\t1...
                              ...
378      (0, 2936)\t1\n  (0, 2981)\t1\n  (0, 2093)\t1...
379      (0, 2056)\t1\n  (0, 2288)\t1\n  (0, 714)\t1\...
380      (0, 1024)\t1\n  (0, 3098)\t1\n  (0, 2936)\t1...
381      (0, 638)\t1\n  (0, 2112)\t1\n  (0, 2513)\t1\...
382      (0, 54)\t3\n  (0, 1810)\t1\n  (0, 1074)\t1\n...
Name: 0, Length: 383, dtype: object
```

# Data Preprocess - 3

$$w_{x,y} = tf_{x,y} \times \log\left(\frac{N}{df_x}\right)$$

**TF-IDF**

Term $x$ within document $y$

$tf_{x,y}$ = frequency of $x$ in $y$

$df_x$ = number of documents containing $x$

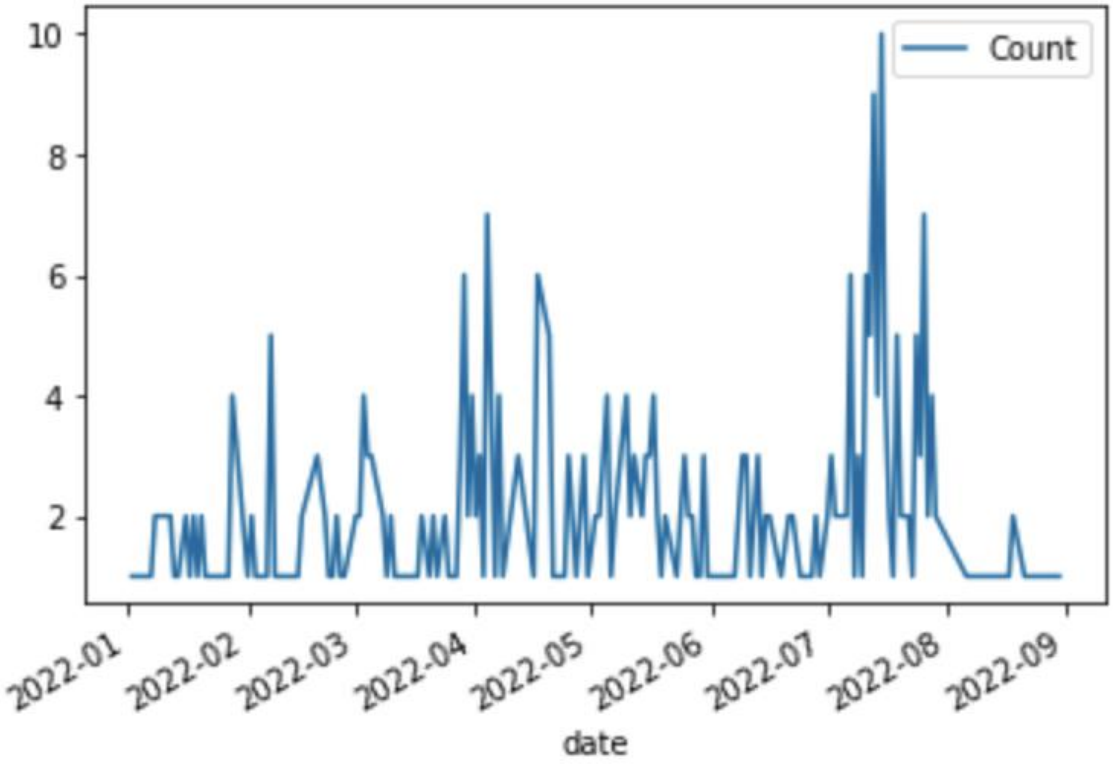
$N$ = total number of documents

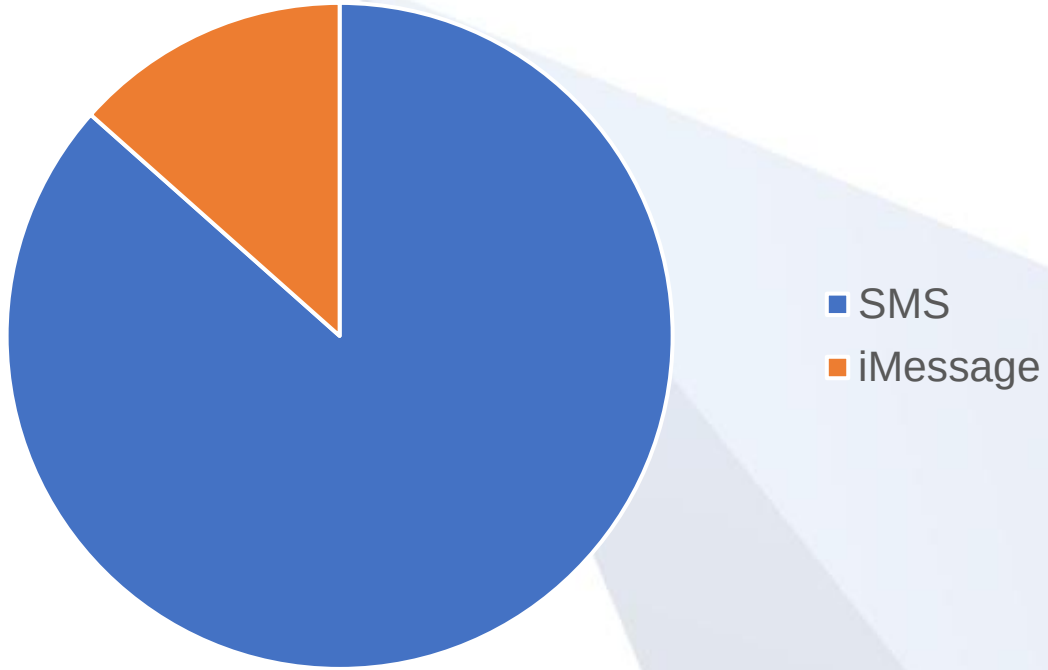
TF-IDF can quantify the importance or relevance of string representations (words, phrases, lemmas, etc.) in a document amongst a collection of documents.

```
1  from sklearn.feature_extraction.text import TfidfTransformer
2  tfidf_transformer = TfidfTransformer()
3  X_train_tfidf = tfidf_transformer.fit_transform(X_train_counts)
4  X_train_tfidf.shape
5  pd.DataFrame(X_train_tfidf)[0]
```

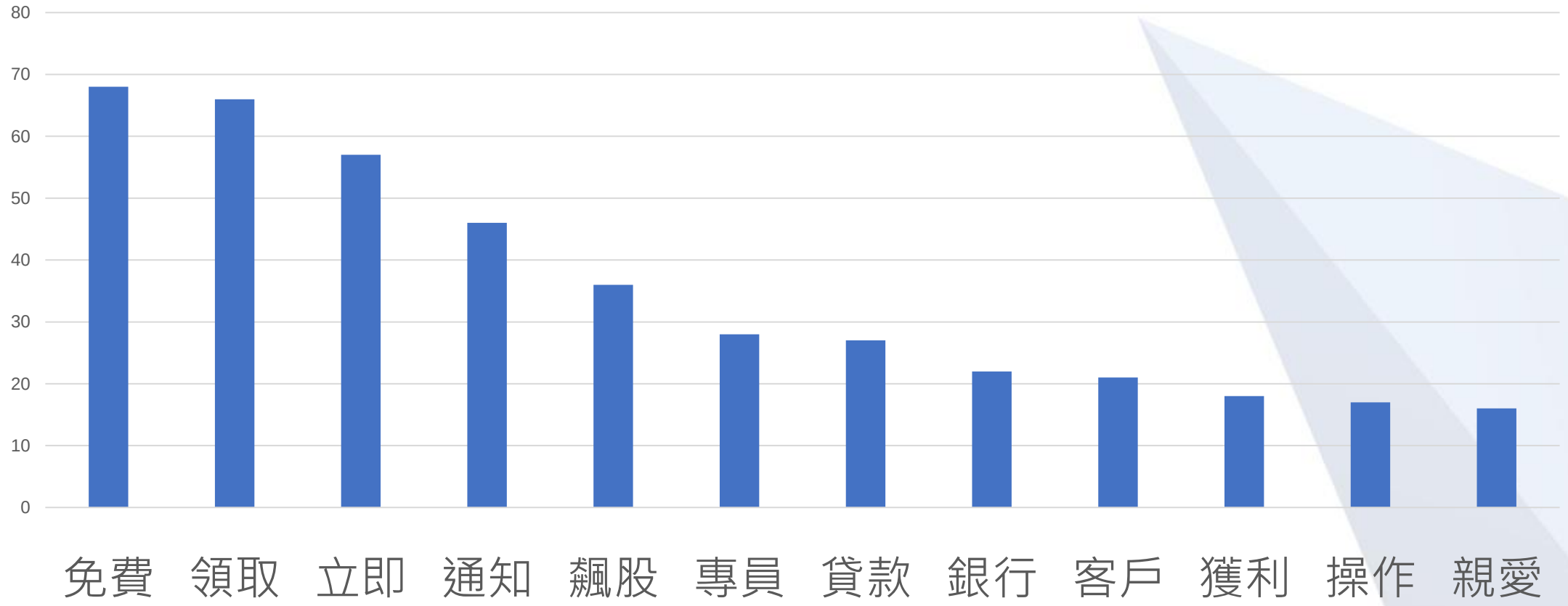
```
0         (0, 3098)\t0.15529729550102797\n  (0, 2936)\...
1         (0, 2983)\t0.19867299018078452\n  (0, 2981)\...
2         (0, 2902)\t0.15622109131890202\n  (0, 2822)\...
3         (0, 3106)\t0.2187148691164481\n   (0, 3098)\t...
4         (0, 3080)\t0.250676006508038\n    (0, 2915)\t...
                                ...
378       (0, 3150)\t0.1919441240618527\n   (0, 2981)\t...
379       (0, 2553)\t0.23987395438417522\n  (0, 2392)\...
380       (0, 3098)\t0.1649145869992665\n   (0, 2936)\t...
381       (0, 2932)\t0.1477166811304547\n   (0, 2903)\t...
382       (0, 3174)\t0.1946824365145517\n   (0, 2693)\t...
Name: 0, Length: 383, dtype: object
```

# Data Analysis



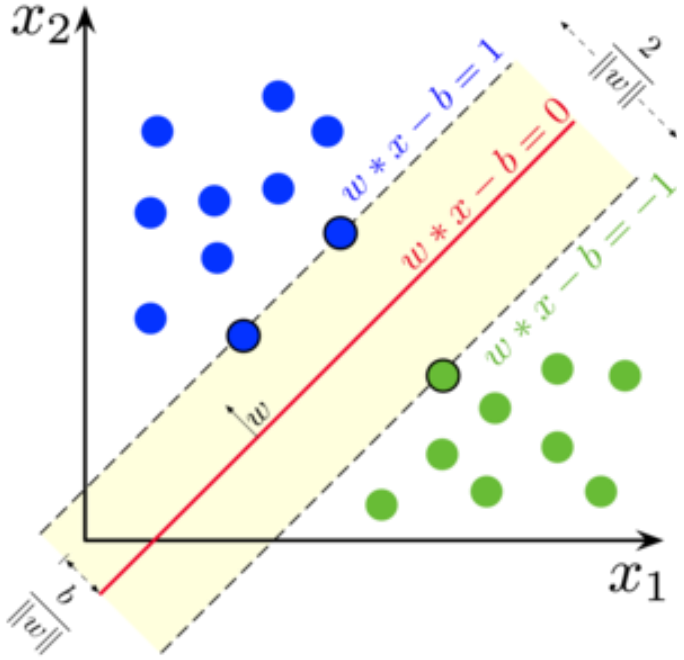



# Data Analysis

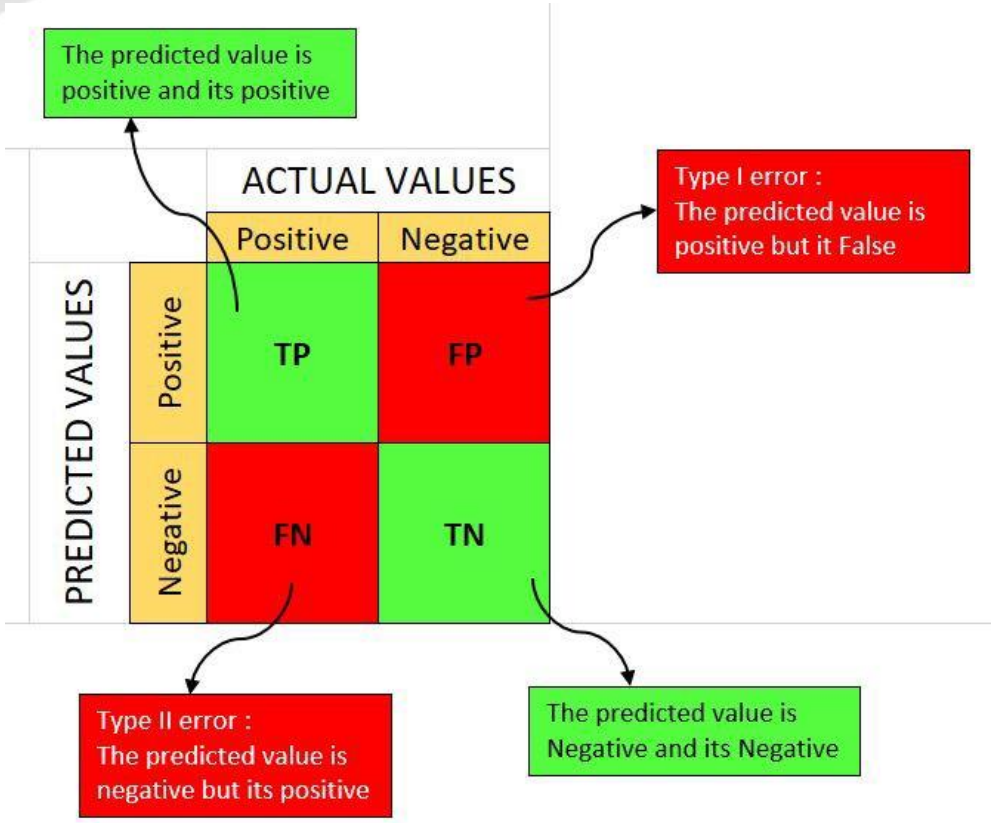


# Training model - SVM

Support vector machines (SVMs) are a set of supervised learning methods used for classification, regression and outliers detection.



# Confusion Matrix




```
from sklearn import metrics
metrics.confusion_matrix(y_test.astype(str),predictions)
```

[229] ✓ 0.1s

```
[[88  5]
 [ 7 90]]
```

```
metrics.classification_report(y_test.astype(str),predictions)
```

[230] ✓ 0.2s

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 詐騙 | 0.93 | 0.95 | 0.94 | 93 |
| 非詐騙 | 0.95 | 0.93 | 0.94 | 97 |
| accuracy | | | 0.94 | 190 |
| macro avg | 0.94 | 0.94 | 0.94 | 190 |
| weighted avg | 0.94 | 0.94 | 0.94 | 190 |

# Prediction

```
text_clf.predict(['上次去聯誼烤肉時,我閨蜜給你的感覺不錯吧?忽然消失那麼久不聯絡是怎樣?加我賴,https:!!urlzs.com/Fne7V'])[0]
```

[231] ✓ 0.7s

```
'詐騙'
```

```
text_clf.predict(['（承上則）如需查詢詳細之繳費資料，請上 123.cht.com.tw 或電子帳單系統 cht.tw/c/u400y 查詢，感謝您。'])[0]
```

[232] ✓ 0.4s

```
'非詐騙'
```

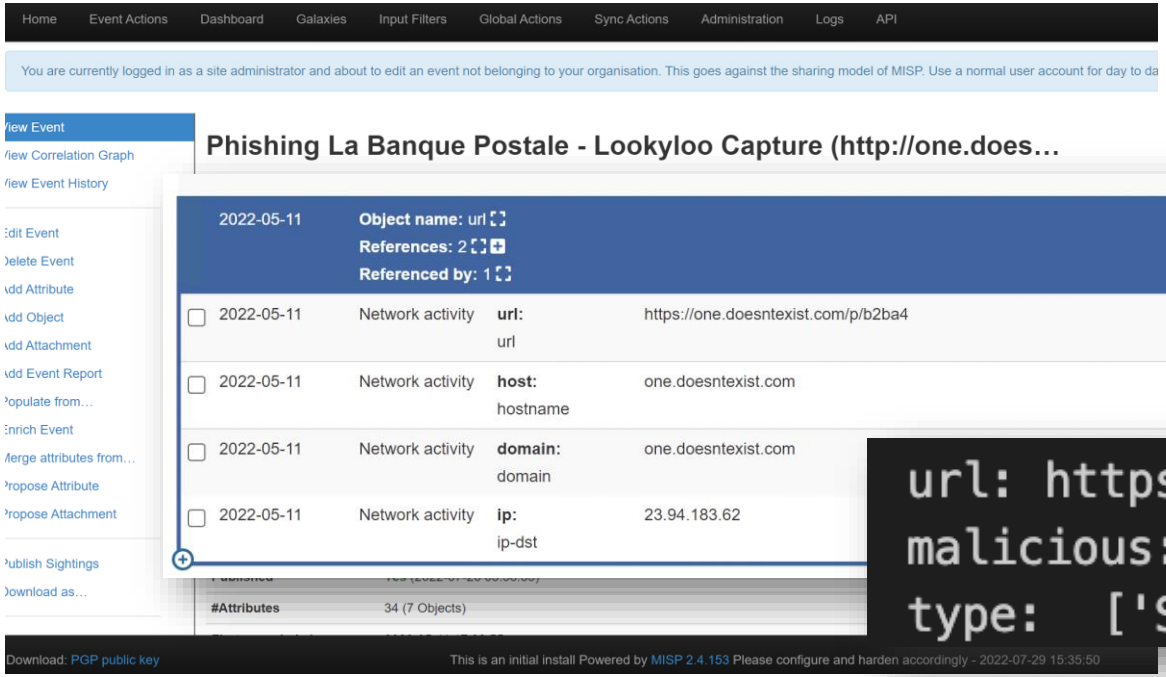
# Model Comparision

Using: Macbook Pro, 2022, Chip Apple M1, 16GB RAM

| Model | 詐騙 (F1-Score) | 非詐騙 (F1-Score) | AVG (F1-Score) | CPU times user | CPU times sys | CPU times total |
|---|---|---|---|---|---|---|
| SVM | 0.94 | 0.94 | 0.94 | 1.56 ms | 3.26 ms | 4.82 ms |
| MultinomialNB | 0.93 | 0.92 | 0.92 | 16.6 ms | 8.76 ms | 25.4 ms |
| LogisticRegression | 0.94 | 0.94 | 0.94 | 16.1 ms | 2.23 ms | 18.3 ms |
| RandomForest | 0.91 | 0.91 | 0.91 | 91.8 ms | 3.55 ms | 95.4 ms |

# Safe Browsing APIs

- The Safe Browsing APIs (v4) check URLs against Google's constantly updated lists of unsafe web resources.



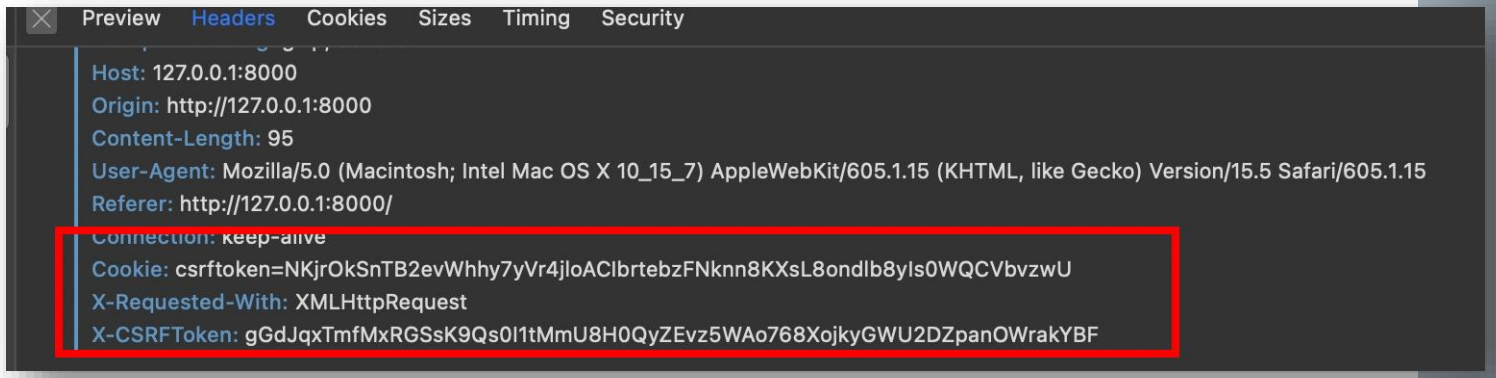

```python
from pysafebrowsing import SafeBrowsing
import validators


def malcheck(KEY, URL):
    s = SafeBrowsing(KEY)
    r = s.lookup_urls([URL])
    return r


if __name__ == "__main__":
    KEY = '[redacted]'
    while True:
        url = input("url: ").strip()
        if validators.url(url):
            res = malcheck(KEY, url)
            # print(res)
            malicious = res[url]['malicious']
            print("malicious: ", malicious)
            if malicious:
                print("type: ", res[url]['threats'])
```

```
url: https://one.doesntexist.com/p/b2ba4
malicious:  True
type:  ['SOCIAL_ENGINEERING']
```

# Web Design


沒有輕易的獲利，而是不懈的努力，防詐騙 API 即刻試用!
運用機器學習（ML）技術，以深度學習（Deep learning）方式打造簡訊中文語意分析模型，有效辨識詐騙訊息。
請輸入訊息! 承諾不保存用戶資訊!
範例：您好，我是與XX證券合作的陳小姐，有急事找你 麻煩加我一下籤 ：sectools.tw
送出檢查!

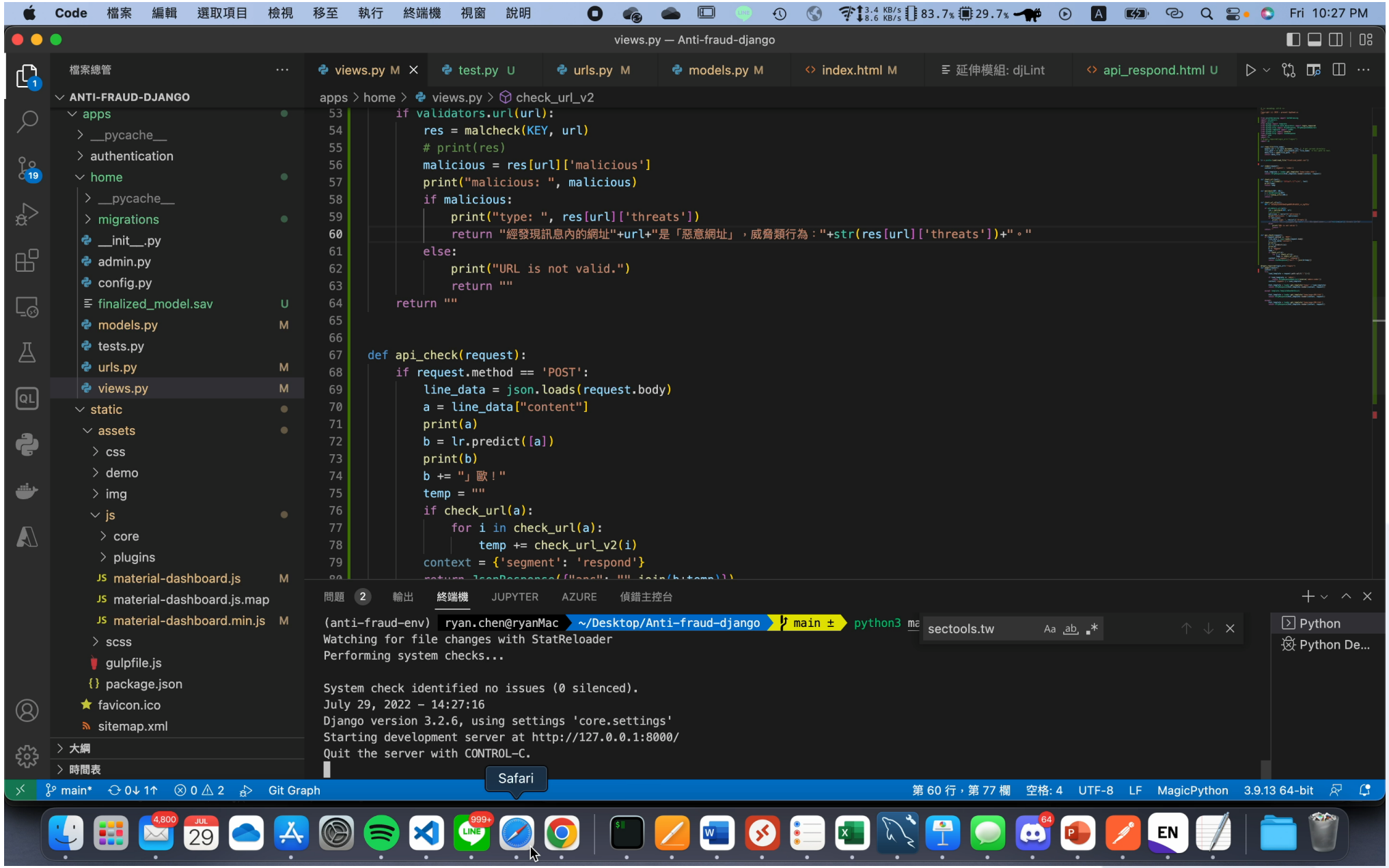
# Web Design - Security




```
<script>
    $(document).ready(function() {
      md.initDashboardPageCharts();
      $("#api_send").click(function(){
        data = {'content':$("#api_content").val()}
        $.ajax({
          headers: {
            'X-CSRFToken': "{{ csrf_token }}"
          },
          type: "POST",
          url: "/api",
          data: JSON.stringify(data),
          success: function(resultData){
              alert("Save Complete");
          }
       });
      });
    });
  </script>
```

# Demo

# Demo


views.py — Anti-fraud-django
ANTI-FRAUD-DJANGO
apps
home
static
assets
views.py
urls.py
models.py
tests.py
finalized_model.sav
admin.py
config.py
material-dashboard.js
material-dashboard.min.js
gulpfile.js
package.json
favicon.ico
sitemap.xml
def api_check(request):
if request.method == 'POST':
line_data = json.loads(request.body)
a = line_data["content"]
b = lr.predict([a])
context = {'segment': 'respond'}
(anti-fraud-env) ryan.chen@ryanMac ~/Desktop/Anti-fraud-django main ± python3 ma
Watching for file changes with StatReloader
Performing system checks...
System check identified no issues (0 silenced).
July 29, 2022 - 14:27:16
Django version 3.2.6, using settings 'core.settings'
Starting development server at http://127.0.0.1:8000/
Quit the server with CONTROL-C.
Safari

# Conclusion

# Conclusion

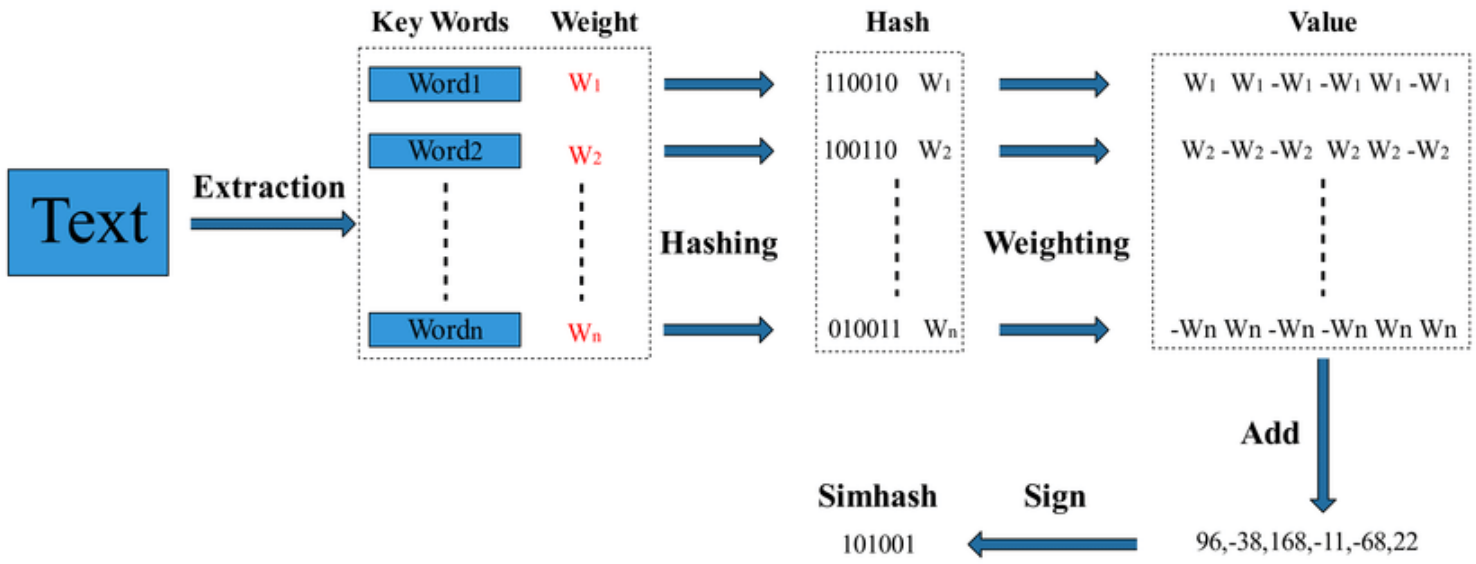
1. Our research achieves over 85% accuracy, making fraud prevention easy

2. Further analysis of URLs in messages for threats

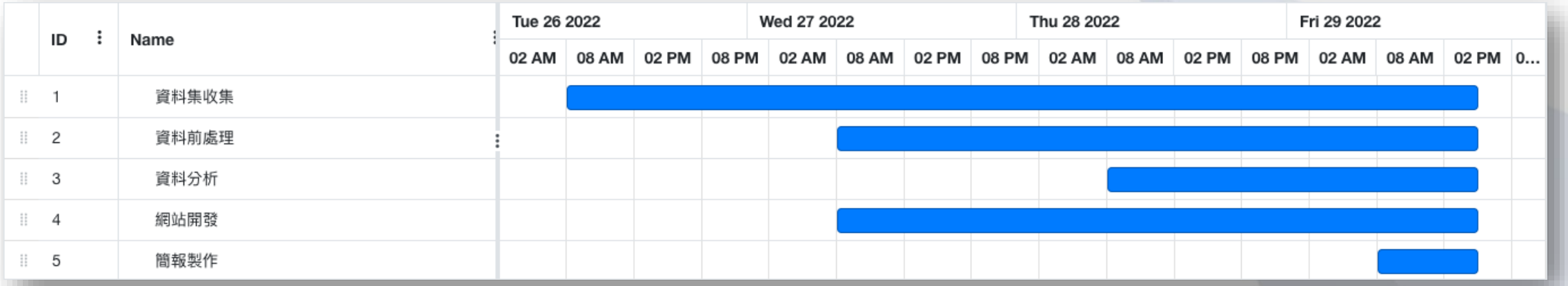
# Future Work

- Entered content contains personal privacy
  - Maybe use a Similarity algorithm to replace machine learning

    SimHash、MinHash



# Contribution



# Reference


41

# Reference


1. 創市際市場研究顧問股份有限公司 (2022). "台灣網路使用概況." from https://www.ixresearch.com/reports/cati.
2. 孫彬訓 (2022). "2021年金融活動空前熱絡 金融詐騙風暴來襲 民眾應提高警覺." from https://www.chinatimes.com/realtimenews/20211210005816-260410?chdtv
3. 沈婉玉 (2021). "高齡更難設防 先設感情圈套 再騙投資詐財." from https://orange.udn.com/orange/story/121199/5970467.
4. 電腦王阿達(2022). “Whoscall 公開詐騙電話簡訊趨勢年度報告，投資詐騙躍居榜首，股票推銷與一接就掛變多了”. from https://today.line.me/tw/v2/article/YaVyEla
5. 林秉賢（2015）。一個偵測行動裝置即時通訊訊息的反詐騙系統-以臉書即時通為例。國立臺灣科技大學資訊管理系碩士論文，台北市。 取自https://hdl.handle.net/11296/k7vdye
6. Whoscall（2022）. https://whoscall.com/en
7. 內政部警政署165 全民防騙網（2022）. https://165.npa.gov.tw

# Thanks for listening


43