【Gotcha! 我抓得住你 - AML 焦點人物辨識】 API 開發說明文件

Introduction

各位玉山人工智慧公開挑戰賽的參賽者們,本次比賽將自行架設 server,並且使用 API 的方式來測試自己模型成效喔! 比起過往的比賽方式,這次變得更進階囉! 請不用擔心,跟著我們以下的操作步驟,讓我們一起展開一趟精實的資料科學旅程吧...

Requirements

在開始之前,請您先做以下準備

- 在 T-Brain (https://tbrain.trendmicro.com.tw/Competitions/Details/11) 上報名比賽
- 留存隊長於 <u>T-Brain (https://tbrain.trendmicro.com.tw/Competitions/Details/11)</u> 報名比賽時收到的手機驗證碼,此驗證碼將作為後續 API 使用之驗證資料 (弄丟也沒關係,後面會再教大家如何重新取得驗證碼)

架設 API Server

目前市面上有許多提供免費試用的雲端運算平台,例如: Google Cloud Platform、Microsoft Azure、Amazon Web Service 等等。以下我們以 Google Cloud Platform 為例做說明(參賽者要使用其他平台也可以喔!)。

• **Step 1.** 網路搜尋<u>Google Cloud Platform (https://reurl.cc/E7GVg0)</u>,會看到以下畫面(如圖 1)。



• Step 2. 填寫註冊資料(如圖2和圖3)。

個人

■ 稅務資訊 稅籍

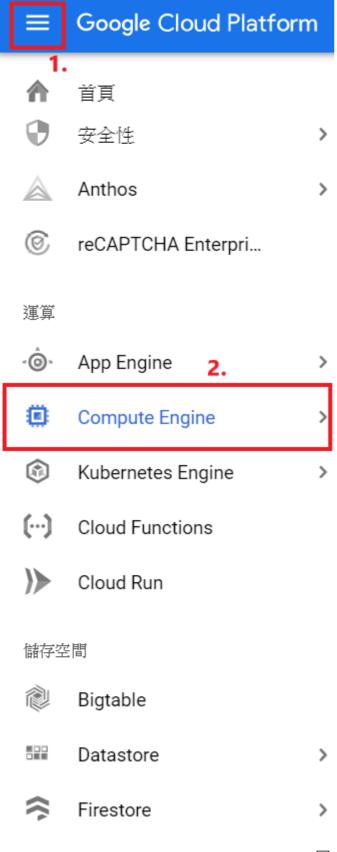
名稱和地址 ①

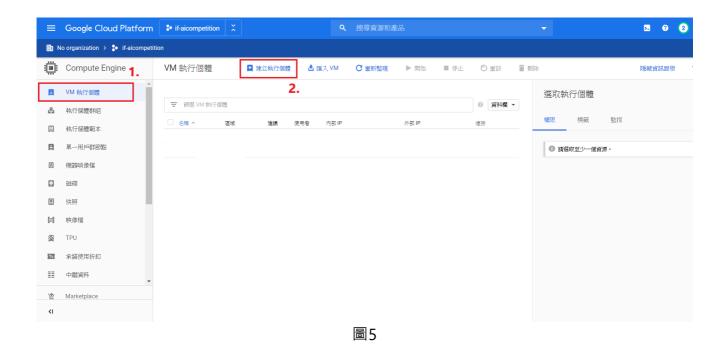


圖3

Google Cloud Platform 會請您綁定信用卡,不過 Google 對於新帳號有提供免費試用額度 \$300 美金,參賽者若於額度內使用便不會進行扣款。

• Step 3. 登入後,首先要先建立自己的 VM(如圖4),點擊選單,再點選 Compute Engine。接著會看到如圖5的畫面,請點選「VM執行個體」,再點選「建立執行個體」。





• Step 4. 建立VM。

可依照個人的需求來建立 VM。記得防火牆需允許 HTTP 要打勾,如下圖6。

■ No organization > • if-aicompetition	
← 建立執行個體	
The state of the s	是分及 API 存取權 ② 服務帳戶 ② Compute Engine default service account 存取權重置 ② ① 允許預設存取權 ① 允許所有 Cloud API 的完整存取權 ② 给可以新增標紀和防火機模,允許接受來自網際網路的特定演量 》 允許 HTTP 流量 ① 允許 HTTP 流量 ① 允許 HTTP 流量 ② 管理 安全性 磁碟 網路 單獨租用 說明 (選項)
	防刑除功能 的用防删除功能 敢用刑除保護時,無法刑除執行個體。瞭解詳情

圖6

○ 省錢小撇步

- 1. VM的機器類型請先設定為最基本的規格,不用一開始就開太大,不然會很花 錢。
 - 未來可以再依照自己的模型來更改規格。
- 2. 請留意不同雲平台的計價模式 · 但非常建議不使用VM的時候請都先關閉 · 但每次關閉重啟時 · 也要留意機器的外部 IP 是否變動 · 如果有變動 · 也請重新至 Slack 註冊 IP · 並執行後續 Step 5 設定靜態 IP。

• **Step 5.** 設定靜態外部IP。

VM 的外部 IP 預設為浮動 IP。為避免重啟 VM 時被隨機賦予新的外部 IP 位址,導致參賽者接收不到題目,所以跟著以下步驟,設定靜態 IP 來確保比賽過程中的品質吧!

首先·進入 VM 後·點選右上的「編輯」(如圖7)·找到「網路介面設定」(如圖8)·找到外部 IP 後(如圖9)並設定名稱(如圖10)·新的名稱可以隨意命名。設定完成後將會看到如圖11畫面·紅框上的 IP: 35.202.143.146 會是VM的靜態外部 IP。



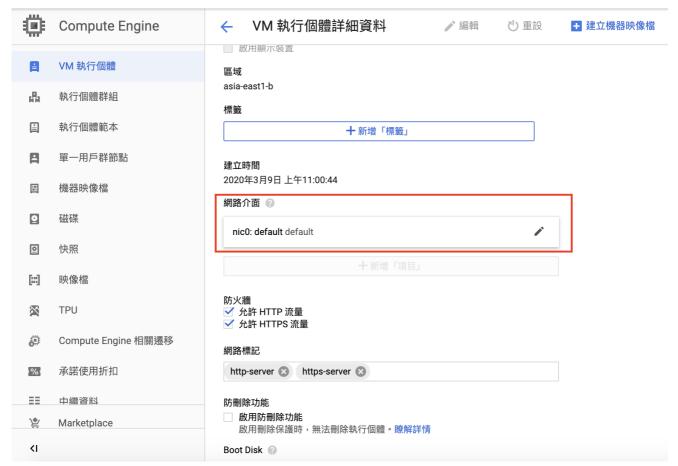


圖8



圖9



圖10



• **Step 6.** 產生金鑰並建立 SSH 連線。 打開 Terminal,輸入以下指令。

1 ssh-keygen

產生金鑰時,終端機上會顯示下列問題,參賽者可以直接**連按 Enter** 跳過直到出現以下畫面(如圖12),此圖表示金鑰產生完畢。如有設定密碼,記得連線時要輸入密碼才能連上 server 喔。

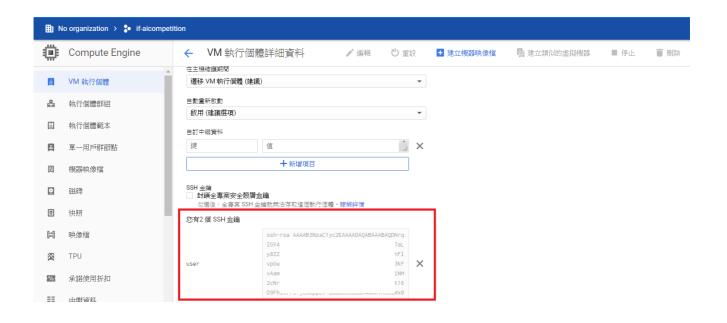
```
Terminal
檔案(F) 編輯(E) 檢視(V) 搜尋(S) 終端機(T) 求助(H)
magiclen@magiclen-linux ~ $ ssh-keygen
Generating public/private rsa key pair.
Enter file in which to save the key (/home/magiclen/.ssh/id_rsa):
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in /home/magiclen/.ssh/id_rsa.
Your public key has been saved in /home/magiclen/.ssh/id_rsa.pub.
The key fingerprint is:
The key's randomart image is:
+--[ RSA 2048]----+
       So
   Ε
       * 0 .
    +.0 .
nagiclen@magiclen-linux ~ $
```

圖12

金鑰位置預設會產生在上圖的紅框裡,金鑰檔案名稱為 id_rsa.pub。在產生自己的 RSA SSH 金鑰前,可以先確認此檔案存在與否,若已經存在,則不一定要再次產生金鑰。

點選剛剛建立好的 VM 並點選「編輯」,移動到 ssh 金鑰項目下,將 id_rsa.pub 裡的整串金鑰複製貼上(如圖13)。最後,在 Terminal 上使用如下指令,就可以從本機端連上自己建立的 server 囉!

1 ssh [external_IP_of_your_VM]



• Step 7. 環境設定。

Google Cloud Platform 的 VM 環境裡並沒有 Python,所以透過 ssh 連上 VM 後照著以下指令安裝 pip 和 Python3。

- 1 sudo apt update
- 2 sudo apt-get install python3-pip #install python3 and pip

順便安裝架設 API Server會使用到的套件吧。

- 1 sudo -H pip3 install flask
- 2 sudo -H pip3 install numpy==1.15
- 3 sudo -H pip3 install pandas

• Step 8. 架設自己的 API。

開始前,請先確定 VM 中已經安裝 flask 套件,flask 是 Python 套件中容易上手的 web 開發框架。(安裝方式請參考 Step 7)。

我們提供一支簡單易懂的 Python 範例程式,讓沒有寫過 API 的參賽者也可以輕鬆架設 比賽專用的 API。範例程式檔(api.py)可以在 T-Brain 上下載,下一步驟會簡單說明這 支程式碼。如果對架設 HTTP rest API Server 瞭若指掌的強者可以選擇自己架設,只要確保 格式相符即可。

• Step 9. 程式碼介紹。

介紹範例程式前,請參賽者先把 api.py 放到 VM 上,可參考以下步驟:

- 。 請先在 VM 上建好資料夾
 - 1 mkdir [your_directory]
- o 查看於 VM 上建好的資料夾的絕對路徑
 - 1 cd [your_directory]
 - 2 pwd # 查看目前資料夾的絕對路徑
- 在本機端,可以透過以下指令連線到VM的外部IP,並將本機端的程式碼複製到 VM 的絕對路徑下(請先確保 VM 上已建好資料夾)。
 - 1 scp [path_to_your_local_file] [external_IP_of_your_VM]:/[path_to_your_direc
 - 2 # 範例:假設VM external IP是 35.202.143.146 (step 5 設定的固定 IP)
 - 3 # scp ./api.py 35.202.143.146:./esun_ai_competition/

- 使用 vim 指令,打開 api.py
 - 1 vim api.py

會看到以下程式碼。

```
1
     from flask import Flask
 2
     from flask import request
 3
     from flask import jsonify
 4
     import time
 5
     import hashlib
 6
     import numpy as np
 7
     import pandas as pd
8
     app = Flask(__name___)
9
10
     ###### PUT YOUR INFORMATION HERE ######
11
     CAPTAIN_EMAIL = 'my@gmail.com'
     SALT = 'my_salt'
12
     13
14
15
     def generate_server_uuid(input_string):
16
         """ Create your own server_uuid
17
         @param input_string (str): information to be encoded as server_uuid
         @returns server_uuid (str): your unique server_uuid
18
19
20
         s = hashlib.sha256()
21
         data = (input_string+SALT).encode("utf-8")
22
         s.update(data)
23
         server_uuid = s.hexdigest()
24
         return server_uuid
25
26
     def predict(article):
         """ Predict your model result
27
28
         @param article (str): a news article
29
         @returns prediction (list): a list of name
30
31
         ###### PUT YOUR MODEL INFERENCING CODE HERE ######
32
         prediction = ['aha','danny','jack']
33
34
35
         36
37
         prediction = _check_datatype_to_list(prediction)
38
         return prediction
39
40
     def _check_datatype_to_list(prediction):
41
         """ Check if your prediction is in list type or not.
42
             And then convert your prediction to list type or raise error.
43
         @param prediction (list / numpy array / pandas DataFrame): your predicti
44
         @returns prediction (list): your prediction in list type
45
46
47
         if isinstance(prediction, np.ndarray):
48
             check datatype to list(prediction.tolist())
49
         elif isinstance(prediction, pd.core.frame.DataFrame):
             _check_datatype_to_list(prediction.values)
50
51
         elif isinstance(prediction, list):
52
             return prediction
53
         raise ValueError('Prediction is not in list type.')
54
```

```
@app.route('/healthcheck', methods=['POST'])
55
56
     def healthcheck():
         """ API for health check """
57
         data = request.get_json(force=True)
58
59
         t = int(time.time())
60
         ts = str(t)
         server_uuid = generate_server_uuid(CAPTAIN_EMAIL+ts)
61
62
         server_timestamp = t
         return jsonify({'esun_uuid': data['esun_uuid'], 'server_uuid': server_uu
63
64
65
     @app.route('/inference', methods=['POST'])
     def inference():
66
         """ API that return your model predictions when E.SUN calls this API """
67
68
         data = request.get_json(force=True)
69
         esun_timestamp = data['esun_timestamp'] #自行取用
70
71
         t = int(time.time())
72
         ts = str(int(t))
73
         server_uuid = generate_server_uuid(CAPTAIN_EMAIL+ts)
74
75
         try:
76
             answer = predict(data['news'])
77
         except:
78
             raise ValueError('Model error.')
79
         server_timestamp = t
80
81
         return jsonify({'esun_timestamp': data['esun_timestamp'], 'server_uuid':
82
83
     if __name__ == "__main__":
         app.run(host='0.0.0.0', port=8080, debug=True)
84
85
```

- o 根據上述程式碼,有以下幾點需注意:
 - 1. 請在第 11 行輸入該隊隊長的 e-mail。
 - 2. 請在第 12 行的 SALT 輸入任意字元,以增加 uuid 加密的複雜度。
 - 3. 需實現 API 規格如下:

Methods	Info
healthcheck	[POST] 確認參賽者的 API 服務是否正常。
inference	[POST] 玉山提供題目,參賽者回傳模型運算結果。

4. healthcheck 資料傳輸格式說明: 資料格式皆以 **json** 格式傳送。

Request

```
1   POST /healthcheck
2   {
3      "esun_uuid":"AZE3k4jb456mdX03fvsd==",
4      "esun_timestamp":1590493614,
5      "retry":2
6   }
```

Response

```
1  {
2     "esun_uuid":"AZE3k4jb456mdX03fvsd==",
3     "server_uuid":"Rs9$revm32ofjwfs2kgrep==",
4     "captain_email":"my@gmail.com",
5     "server_timestamp":1590493615,
6  }
```

回傳資料欄位說明:

參數	型態	說明
esun_uuid	string	玉山 client 傳給開發者的 task id
server_uuid	string	由開發者自行產生的 uuid,接下來 call inference API 時,會利用此 server_uuid 作為識別。
captain_email	string	隊長 email
server_timestamp	int	healthcheck API 回傳訊息的時間 (Unix Epoch Time·從 1970-01-01T00:00:00 起算的秒數)

5. inference 資料傳輸格式說明: 資料格式皆以 **json** 格式傳送。

Request

```
1    POST /inference
2    {
3         "esun_uuid":"AZE3k4jb456mdX03fvsd==",
4         "server_uuid":"Rs9$revm32ofjwfs2kgrep==",
5         "esun_timestamp":1590493849,
6         "news":"人工智慧公開挑戰賽題目...",
7         "retry":2
8    }
```

Response

```
1  {
2     "esun_uuid":"AZE3k4jb456mdX03fvsd==",
3     "server_uuid":"Rs9$revm32ofsdfs2kgrsd==",
4     "server_timestamp":1590493852,
5     "answer":["Danny","Jack"]
6  }
```

回傳資料欄位說明:

參數	型態	說明
esun_uuid	string	玉山 client 傳給開發者的 task id
server_uuid	string	由開發者自行產生的 uuid。
answer	list	找出來的黑名單 string list,如果結果為空, 請回傳空的 list,例:[]
server_timestamp	int	inference API 回傳的時間 (Unix Epoch Time, 從 1970-01-01T00:00:00 起算的秒數)

預測出來的黑名單一定要存成 list 格式,不然提交答案時會出錯失去分數喔。

6. 時間限制:

healthcheck:

■ timeout: 1 秒。 ■ retry 次數: 2次。

注意

在 healthcheck 的階段,參賽者的 API 需要在 timeout 的秒數內回傳資訊,若第一次未成功回傳,玉山 server 會再呼叫 2 次。如果參賽者的 API 仍未於規定秒數內回應,將視為放棄該題,玉山 server 也不會發送該輪題目。

- inference:
 - timeout: 5 秒。
 - retry 次數: 2次。

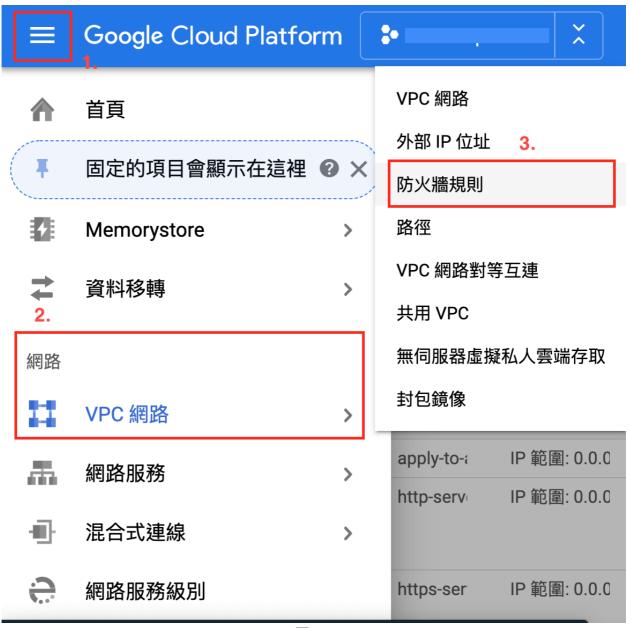
注意

在 inference 的階段·玉山 server 每輪發送一題題目後·參賽者會有 5秒的時間回覆答案。超過 5 秒後·玉山 server 會重試 2 次·若仍沒有收到參賽者該輪答案,將視為參賽者放棄此輪作答。

• Step 10. 防火牆規則設定。

Google Cloud Platform 預設 port 為 80,請各位參賽者跟著以下步驟開通一個 1024 以上的 port number。

1. 如圖 14,點選選單,並且找到 VPC 網路中的防火牆規則。



2. 如圖 15,點選建立防火牆規則。

	VPC 網路	防火牆規則	+ 莲	全立防火牆規則	C重新整理	▶ 啟用記錄功能	❷ 停用記錄功
=	VPC 網路	防火牆規則會管	制通往執行個體	的連入流量或源	自執行個體的外送	流量。根據預設,系統	
먑	外部 IP 位址	會封鎖來自您網					
88	防火牆規則	注意:如要管理	App Engine 防)	火牆,請按 <u>這裡</u> 。			
×	路徑	〒 篩選表格					
۵	VDO /何晚料放下注	□ 名稱	類型	目標	篩選器	通訊協定/通訊埠	動作
Ŷ	VPC 網路對等互連	allow500	輸入	apply-to-a	IP 範圍: 0.0.0	tcp:5000	允許
₩ 🕸	共用 VPC 無伺服器虛擬私人雲端存取	default- allow- http	輸入	http-serv	IP 範圍: 0.0.0	tcp:80	允許
1 8 1	封包鏡像	default- allow- https	輸入	https-ser	IP 範圍: 0.0.0	tcp:443	允許
		http	輸入	http-serv	IP 範圍: 0.0.0	tcp:8080	允許
		default- allow- icmp	輸入	全部套用	IP 範圍: 0.0.0	icmp	允許
		default- allow- internal	輸入	全部套用	IP 範圍: 10.1;	tcp:0-65535 udp:0-65535 icmp	允許
< I		default-	輸入	全部套用	IP 範圍: 0.0.0	tcp:3389	允許

圖 15

3. 如圖 16,設定目標標記、來源 IP 範圍和通訊埠設為 8080,最後點選儲存,就設定好啦!



圖 16

• Step 11. 啟動 API 服務。

輸入以下指令,並且看到圖 17 畫面就表示成功啟動 API 服務囉!

```
user@tbrain-test:~$ python3 api.py
* Serving Flask app "api" (lazy loading)
 * Environment: production
    WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.
 * Debug mode: on
 * Running on <a href="http://0.0.0.0.0">http://0.0.0.0</a>:8080/ (Press CTRL+C to quit)
 * Restarting with stat
 * Debugger is active!
   Debugger PIN: 169-167-242
```

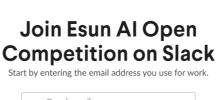
圖17

Tips:完成此處啟動 API 服務後,便可以順利執行下一章節的 API 驗證,建議您可以 先完成下一步的註冊與驗證流程,後續再將模型內容置換成您的模型。

● 想了解更多 flask 的相關應用可以點選此連結 (https://flask.palletsprojects.com/en/1.1.x/)。

於 Slack Workspace 註冊正式賽

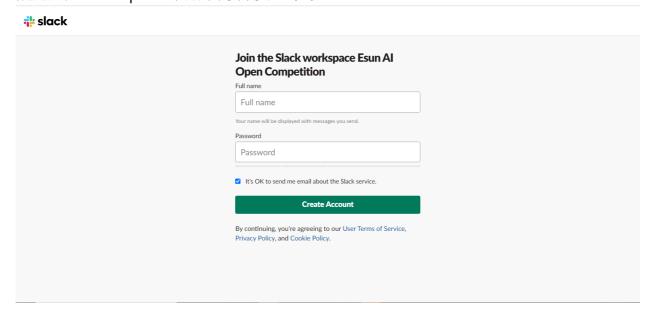
- Step 1:點<u>此連結 (https://esunaiopencompetition.slack.com/join/shared_invite/zt-ed3evsfq-</u> o3IUFB80duxGjshACB~t A#/)加入玉山人工智慧公開挑戰賽 Slack workspace
 - 。 請您輸入加入此 workspace 欲使用的 email。



slack

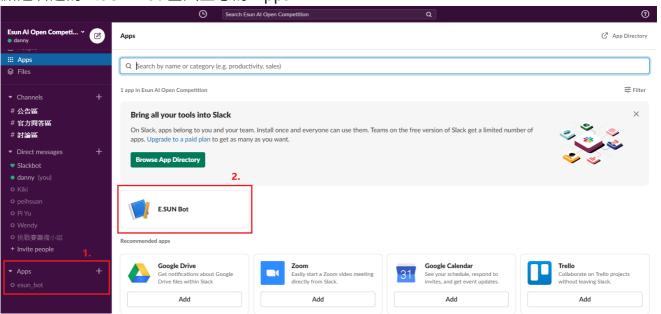
Continu	ue

。 設定於此 workspace 中顯示的名字與密碼。



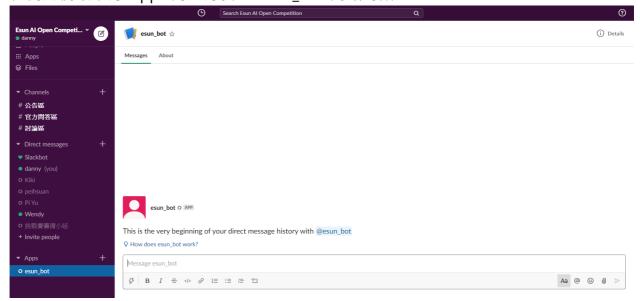
注意:以下步驟所使用到的 esun_bot 會在7/1開放。

• **Step 2**: 檢查是否有 esun_bot 在 Apps 裡,若沒有請依照下圖點選 Apps 旁邊的+號,再 點選右邊的 E.SUN Bot 匯入至您的 Apps。

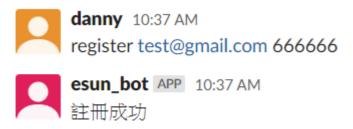


• Step 3: 與E.SUN Bot 對話註冊正式賽。

o 在側邊欄最下方 Apps 的地方找到 esun_bot 開啟對話。



- (Optional)若隊長忘記於 T-Brain 註冊時收到的手機驗證碼(若還留存請跳至下一步),可於對話框輸入以下指令,需注意 forget_token 與 captain_email 之間是以空格分隔。
 - forget_token [captain_email]
 - 2 # [captain_email] 請置換成隊長 email
- 。 於對話框中輸入以下註冊指令,需注意, register 、 captain_email 、 phone_token_of_captain 之間均是以空格分隔。
 - 1 register [captain_email] [phone_token_of_captain]
 - 2 # [captain_email] 請置換成隊長 email
 - 3 # [phone_token_of_captain] 請置換成隊長手機驗證碼
- 註冊成功會看到如下圖



- **Step 4**:請先確保您已完成架設 API 章節並啟動您的 API 服務,接下來就可以在 Slack 測試 API 是否能順利獲得題目與回傳答案。
 - o 於對話框中輸入以下驗證指令,需注意, verification 、 captain_email 、 phone_token_of_captain 、 api_url 之間均是以空格分隔。

- verification [captain_email] [phone_token_of_captain] http://[external_IP_of
- 2 # [captain email] 請置換成隊長 email
- 3 # [phone_token_of_captain] 請置換成隊長手機驗證碼
- 4 # [external_IP_of_your_VM] 請置換成您 VM 的外部 IP

1. 需注意,最後一次驗證的 API 將成為該隊伍正式賽時玉山發送題目的 API,故每個隊伍都必須完成驗證,否則會影響到您參與正式賽的權益。

- 2. 在比賽結束前,您可以隨時運用 verification 指令來驗證您調整後 API 的輸出入格式是否正確,但需注意,您在每場正式賽開始前,最後一次驗證的 API url 會自動被視為正式賽呼叫之 API url。
- 如下圖·驗證時 esun_bot 會發送一題正式比賽題目至參賽者的API·同時確認 API 位址和回傳的答案格式是否正確。
 - danny 10:43 AM verification test@gmail.com 666666 http://34.80.231.179:8080
 - esun_bot APP 10:43 AM 已發送驗證題目給您,請至您的 API 服務做確認,後續可用 get_status 確認驗證結果
- **Step 5**:取得最後一次驗證狀態是否成功以及驗證的 API url 為何,了解調整後的 API 是否仍可正確接收題目與回傳答案。
 - o 於對話框中輸入以下指令,需注意, get_status 、captain_email 、 phone_token_of_captain 之間均是以空格分隔。
 - 1 get_status [captain_email] [phone_token_of_captain]
 - 2 # [captain_email] 請置換成隊長 email
 - 3 # [phone_token_of_captain] 請置換成隊長手機驗證碼
 - 若驗證失敗會看到以下圖示,請修正錯誤並確認驗證成功再進行正式比賽。
 - danny 10:49 AM get_status test@gmail.com 666666
 - esun_bot APP 10:49 AM 最近一次驗證時間為:2020-05-26 10:43:31.979240, 驗證api url為:http://34.80.231.179:8080, 驗證狀態為:失敗
 - o 若驗證成功會看到如下畫面。
 - danny 11:49 AM get_status test@gmail.com 666666
 - esun_bot APP 11:49 AM 最近一次驗證時間為:2020-05-26 11:40:14.322685, 驗證api url為:http://34.80.231.179:8080, 驗證狀態為:成功

開始訓練自己的模型吧!

完成以上流程,便可以開始訓練您的模型,並將範例 code 中的模型區塊置換成您訓練好的模型。

測試賽與正式賽

- 若我想查詢 API log 我可以怎麼做?
 - 。 (此功能將於 7/22 測試賽後開放) 於 Slack esun_bot 對話框輸入以下指令,需注意,get_log、captain_email、 phone_token_of_captain、 date_for_query 之間均是以空格分隔
 - get_log [captain_email] [phone_token_of_captain] [date_for_query]
 - 2 # [captain_email] 請置換成隊長 email
 - 3 ♯ [phone_token_of_captain] 請置換成隊長手機驗證碼
 - 1. [date_for_query] 的格式為 YYYY-MM-DD
 - 2. 若您無指定日期 (無輸入 date_for_query),則預設回傳給您查詢當天的答題狀況

常見 Q&A

1. Windows 沒有沒有 Linux 的命令提示字元怎麼辦?

其中一種解法為下載 MobaXterm

官方網址: https://mobaxterm.mobatek.net/ (https://mobaxterm.mobatek.net/)

2. HTTP Status code的意義為何?

200: OK

API請求已成功

400: Bad request

由於明顯的客戶端錯誤(例如·格式錯誤的請求語法·太大的大小·無效的請求訊息或欺騙性路由請求)·伺服器不能或不會處理該請求。

404: Not found

請求失敗,請求所希望得到的資源未被在伺服器上發現(例如,post 的 url 輸入錯誤),但允許用戶的後續請求。

500: Internal server error

通用錯誤訊息,伺服器遇到了一個未曾預料的狀況,導致了它無法完成對請求的處理。沒有給出具體錯誤資訊。

3. 使用 Terminal 連線到 server 很容易斷線該怎麼辦?

tmux 指令會是你的小幫手。

tmux 會在同一個 session (連線狀態下的執行環境)保存這些視窗和視窗區塊。 我們可以在任何時間離開這個連線·tmux 會繼續維持這個 session 直到您把 tmux server 砍掉或者重開機的時候。

4. Linux 指令怎麼使用?

以下介紹幾個常用的 Linux 指令:

- o vim [檔名]: 打開並且進入此[檔名]的檔案。
 - 先按 esc,再打 :q 為不儲存並離開。
 - 先按 esc,再打 :w 為儲存不離開
 - 先按 esc,再打:wq 為儲存並離開
- o cd [路徑]:轉換工作目錄至此[路徑]中。
- Is:列出此路徑下的所有檔案。
- o **mkdir**: 創建新的目錄。

如果想更深入了解 Linux 指令,可以點選**鳥哥的Linux指令基本介紹**

(https://linux.vbird.org/linux_basic/redhat6.1/linux_06command.php) °

5. 我不想一直開著使用者手冊,如何在 Slack 上查詢玉山 Slack Bot 有哪些功能資訊呢

您可以開始開啟Slack與esun_bot 的對話框,輸入

1 command_list

即可看到所有功能選單與使用方式