

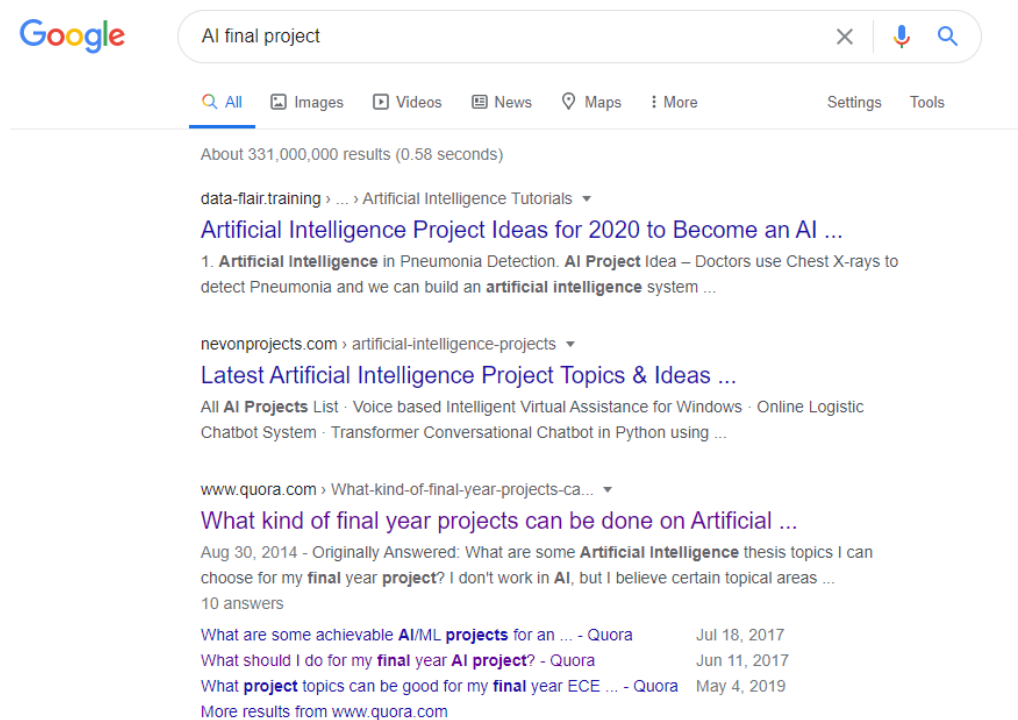
# FAQs for Final project

Po-Chih Kuo

- How to find the topic?
- How will we grade your project?
- Where do data come from?
- I don't have machine!

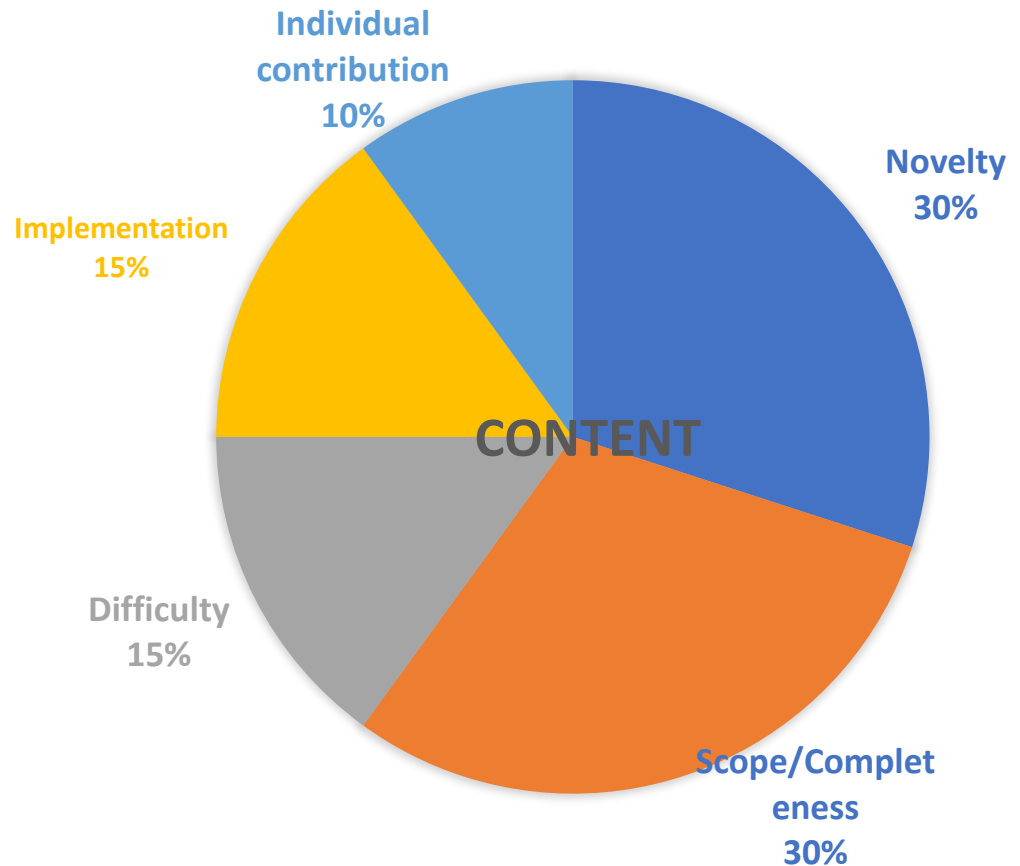
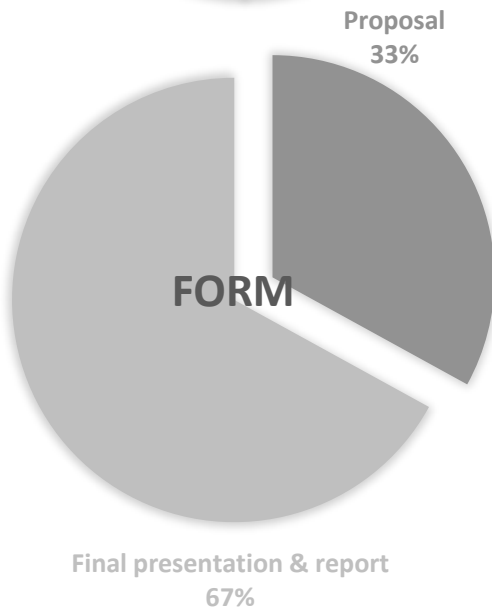
# How to find the topic?

- Good Idea is priceless!
- What if I don't have any idea?



# How will we grade your project?

## GRADING FOR FINAL PROJECT



# Where do data come from?

# Data-1

- <https://www.kaggle.com/datasets>

kaggle

Home

Compete

Data

Notebooks

Discuss

Courses

Jobs

More

Recently Viewed

Fruits 360

House Price Prediction

House Prices: Advance...

EEG-emotion\_LSTM

EEG Brainwave Datas...

Search

Datasets

Find and use datasets or complete tasks. [Learn more.](#)

+ New Dataset

Create Public Datasets

Open a dialogue, accept contributions, and get insights: improve your dataset by publishing it on Kaggle

Create Public Dataset

Search 54,471 datasets

Feedback

Filter

Public

Your Datasets

Favorites

Sort by: Hottest

Chest X-Ray Images (Pneumonia)

Paul Mooney

3 years2 GB7.55856 Files (other)1 Task

Fruits 360

Milica Obradovic

4 months760 MB8.890485 Files (other)1 Task

Fashion MNIST

Zalando Research

3 years69 MB8.510020 Files (CSV, other)1 Task

Skin Cancer MNIST: HAM10000

K Scott Mader

2 years5 GB7110020 Files (other, CSV)

CelebFaces Attributes (CelebA) Dataset

Jessica Li

2 years1 GB7.6202603 Files (other, CSV)

AV : Healthcare Analytics

shivan kumar

11 days2 MB9.410 Files (CSV, other)1 Task

Coronavirus tweets NLP - Text Classification

Aman Miglani

17 days4 MB10.02 Files (CSV)1 Task

Featured Code Competition

SIIM-FISABIO-RSNA COVID-19 Detection

Identify and localize COVID-19 abnormalities on chest radiographs

SIIM Society for Imaging Informatics in Medicine (SIIM) · 1,305 teams · 3 months ago

\$100,000

Prize Money

Overview

Data

Code

Discussion

Leaderboard

Rules

Join Competition

...

Data Description

In this competition, we are identifying and localizing COVID-19 abnormalities on chest radiographs. This is an object detection and classification problem.

For each test image, you will be predicting a bounding box and class for all findings. If you predict that there are no findings, you should create a prediction of "none 1 0 0 1 1" ("none" is the class ID for no finding, and this provides a one-pixel bounding box with a confidence of 1.0).

Further, for each test study, you should make a determination within the following labels:

'Negative for Pneumonia' 'Typical Appearance' 'Indeterminate Appearance' 'Atypical Appearance'

To make a prediction of one of the above labels, create a prediction string similar to the "none" class above: e.g. `atypical 1 0 0 1 1`

Please see the Evaluation page for more details about formatting predictions.

The images are in DICOM format, which means they contain additional data that might be useful for visualizing and classifying.

# Data-2

- <https://data.gov.tw/>

## 資料集服務分類



生育保健(509)



出生及收養(89)



求學及進修(762)



服兵役(283)



求職及就業(575)



開創事業(736)



婚姻(35)



投資理財(1716)



休閒旅遊(946)



交通及通訊(1936)



就醫(1145)



購屋及遷徙(1314)



選舉及投票(110)



生活安全及品質(2889)



退休(20)



老年安養(236)



生命禮儀(114)



公共資訊(32747)

[資料集 \(1145\)](#)[最新消息 \(12853\)](#)[活化應用 \(113\)](#)[我有話要說 \(7206\)](#)[我想要更多 \(1405\)](#)

## 中央機關

[行政院](#) 303

## 地方機關

[南投縣](#) 14  
[嘉義市](#) 18  
[基隆市](#) 22  
[宜蘭縣](#) 19  
[屏東縣](#) 8  
[彰化縣](#) 1  
[新北市](#) 360  
[新竹市](#) 10  
[新竹縣](#) 16  
[桃園市](#) 40  
[澎湖縣](#) 5  
[臺中市](#) 65  
[臺北市](#) 27  
[臺南市](#) 58  
[臺東縣](#) 2  
[花蓮縣](#) 25  
[苗栗縣](#) 55  
[連江縣](#) 3  
[金門縣](#) 2  
[雲林縣](#) 3  
[高雄市](#) 89

## 主題分類

[依申請提供資料](#) 2[其他](#) 943[政府統計](#) 197[政府預算](#) 2[政府支出](#) 1

## 提供機關

[新北市政府衛生局](#) 357[衛生福利部疾病管制署](#) 245[高雄市政府衛生局](#) 64[臺南市政府衛生局](#) 56[臺中市政府衛生局](#) 53

## 資料集列表 | Datasets

篩選條件: [✕ 就醫 \(+\)](#)排序方式: [▼ 下載次數](#) [! 多至少](#)[«](#) [<](#) [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [>](#) [»](#)

共1145筆，本頁顯示1-15筆

### 登革熱1998年起每日確定病例統計

1998年起每日登革熱確定病例統計，資料更新頻率為每日更新 (因檔案較大，資料資源中含有下載位置)

主要欄位說明: version、title、description、source、type、timestamp、column、csv、json

衛生福利部疾病管制署 / 詮釋資料更新時間: 2020/05/27 11:55

[👁 51497](#) [📄 108628](#) [💬 22](#) [🏆](#) [👑](#)[JSON](#)

### 健保門診及住院就診人次統計-腸病毒

各縣市、年齡別、年週之腸病毒門診及住院就診人次統計表

主要欄位說明: 年、週、就診類別、年齡別、縣市、腸病毒健保就診人次、健保就診總人次

衛生福利部疾病管制署 / 詮釋資料更新時間: 2020/07/28 19:34

[👁 10054](#) [📄 53302](#) [💬 0](#) [🏆](#) [👑](#)[CSV](#)[JSON](#)

### 登革熱近12個月每日確定病例統計

近12個月每日登革熱確定病例統計，資料更新頻率為每日更新

主要欄位說明:

發病日、個案研判日、通報日、性別、年齡層、居住縣市、居住鄉鎮、居住村里、最小統計區、最小統計區中心點X、最小統計區中心點Y、一級統計區、二級統計區、感染縣市、感染鄉鎮、感染村里、是否境外移入、感染國家、確定病例數、居住村里代碼、感染村里代碼、血清型、內政部居住縣市代碼、內政部居住鄉鎮代碼、內政部感染縣市代碼、內政部感染鄉鎮代碼

衛生福利部疾病管制署 / 詮釋資料更新時間: 2020/07/28 18:56

[👁 19886](#) [📄 50845](#) [💬 3](#) [🏆](#) [👑](#)[CSV](#)[JSON](#)



# Data-3

- <https://opendata.cwb.gov.tw/devManual/insrtuctionn>



The screenshot displays the 'Open Weather Data' (氣象資料開放平臺) website. The header features the platform's logo and navigation links for '登入/註冊' (Login/Register), '民衆信箱' (Public Mailbox), and '網站地圖' (Site Map). A horizontal menu bar contains links for '公告事項' (Announcements), '資料主題' (Data Topics), '開發指南' (Development Guide), '應用活化' (Application Activation), '推廣發展' (Promotion and Development), '常見問答' (FAQ), and '登入/註冊' (Login/Register). The main content area is titled '使用說明' (Usage Instructions) and includes a sidebar with links to '使用說明', '使用規範' (Usage Guidelines), and '資料擷取API線上說明文件' (Online API Data Retrieval Instructions). The main text explains the platform's purpose, login methods (via Facebook or email), and the benefits of becoming a member. It also outlines the process for obtaining data through API calls and provides a link to the '開發指南' (Development Guide).

氣象資料開放平臺  
OPEN WEATHER DATA

登入/註冊 民衆信箱 網站地圖

公告事項 資料主題 開發指南 應用活化 推廣發展 常見問答 登入/註冊

使用說明

使用說明

中央氣象局開放資料平臺提供各式資料，為能優化開放資料的服務，本平臺採用會員服務機制，可使用中央氣象局氣象會員之帳號或透過facebook登入。民眾或機關可在登入本平臺後透過網頁連結下載資料，也可發展程式介接進行檔案下載或資料擷取。

如無facebook帳號或尚未成為本局氣象會員，請先至本局官網申請註冊(只需有效的郵件帳號，就可免費線上申請)，再登入本平臺，成為會員除可享有本局官網提供之會員服務外，並可取得本平臺所提供之各式開放資料；當有資料集詮釋資料被更新，且資料集提供單位有填寫資料集修訂說明，會員將會收到系統以電子郵件同步通知該資料集詮釋資料內容已異動。

以下分別說明本平臺提供之資料取用方式：

一、透過網站連結下載


於氣象資料開放平臺登入後，於網站中的各資料頁面，直接點選連結下載檔案。



二、發展程式介接

本平臺提供透過URL下載檔案以及 RESTful API 資料擷取方法取用資料，惟因本平臺採用會員服務機制，需帶入資料項目代碼以及有效會員之授權碼，方可取得各式開放資料。其中，資料項目代碼可至資料清單列表查詢，會員之授權碼可於登入會員後取得。

# Data-4


- [https://aidea-web.tw/topic\\_list](https://aidea-web.tw/topic_list)

Artificial Intelligence  
Data Engineering

[About Us](#) [Topics](#) [Projects](#) [Competition](#) [Career](#)  

Portal / Topic List

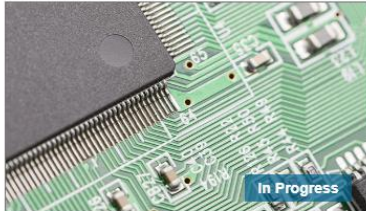
## Topic List

In Progress

### 員工離職預測

人才是企業最重要的資源，提早發現員工離職傾向並留住優秀人才，是企業持續成長的重要議題。員工離職預測是利用大數據與人工智慧，...


2020/08/05 ~ 2020/10/28

In Progress

### Defect Classifications of ...

Automated optical inspection (AOI) [1] is an automated visual inspection of printed circuit board (PCB) (or LCD,transistor) manufacture...

2018/05/24 ~ 2020/11/04

Closed

### 圖書館資源對學習成效之影...

大學圖書館之設立旨在促進大學教育目標的達成，提供了大量的資源、服務，供學生吸收知識並利用圖書館豐富資源之間的關聯性協助學生...

2020/05/28 ~ 2020/06/30

# Data-5

- <https://aidea-web.tw/moe2020>

About UsTopicsProjectsCompetitionCareer

首頁 / AI CUP 2020



## 教育部全國大專校院人工智慧競賽

### Aidea x 教育部攜手推廣人才庫

如同網際網路進入社會各個環節的強大影響力，如今，人工智慧也已經掀起另一波革命。對國家的教育政策而言，培養足夠數量研發人工智慧技術的人才，使國民具備與人工智慧共同協作之能力，進而提升整體社會的競爭力，為重中之重之任務。

本計畫之重心為「標註競賽」，著重在目標明確的題目，藉由競賽精確的題目，訓練學生解決明確目標的問題，我們希望能透過競賽，培養學生解決AI問題的經驗與能力，並導入人工智慧技術與觀念至教育體系，期望將來在投入實習領域進行實戰時，能學以致用。

#### 指導單位——教育部

教育部「結合資訊、科技與人文，引領學校邁向永續發展新世紀」之願景，規劃「前瞻人才培育」、「精緻數位學習」、「科研倫理素養」、「數位公義關懷」、「校園環境永續」五大目標，建構以「人」為中心的學習環境，培育具關懷社會與環境之現代公民，強化國家永續發展。

/ 議題入口 /



#### 醫病訊息決策與對話語料分...

隨著傳染疾病的演變，今日的全球公衛醫療環境常需要面對感染疾病大量流行時所帶來的大量醫



#### 台灣高經濟作物 - 愛文芒果...

近來隨著人工智慧的興起，智慧加值概念導入精準化的技術於農業中，不僅有助於提產產量，更



#### 和弦辨識競賽

隨著時代的進展，聆聽音樂的方式已經從以往的CD 轉移到各式各樣的智慧線上平台，如網外的

報名中

# Data-6

- <https://corelab.iiiedu.org.tw/events/hackathon/2020?fbclid=IwAR2p1zRdeaaxwrFcDOPwezhh75oY9Txm-IlyQJDdq51B8q8fAajBfrGXQRM>

## 企業資料

<h3>學習管理系統資料</h3> <p>完整資料下載</p> <p>提供者 - 教育部教育雲</p> <p>教育雲資料記錄系統，彙集教育雲各系統服務產生的使用歷程資料，數據蒐集整合與分析環境，針對不同學習情境，以個別學生、學校或縣市等規模，進行不同層次的學習差異化分析。</p>	<h3>課程訂單資料</h3> <p>Sample 搶先看    取得方式</p> <p>提供者 - 資策會數位教育研究所</p> <p>由資策會數位教育研究所提供的歷年學員訂單資料，包含「學員訂單資料暨課程完訓薪資調查」與「就業養成班學員成績紀錄」兩個資料集。</p>	<h3>人力銀行資料</h3> <p>取得方式</p> <p>提供者 - 104資訊科技</p> <p>透過人力銀行公司以教育就業文化相關類型，提供2015年 到2020年單一個月份的職缺刊登資訊，與單月職缺所需證照、技能、工具，及此職缺所屬公司產業別、公司員工人數，共五項。</p>
<h3>程式練習資料</h3> <p>取得方式</p> <p>提供者 - CodingBar</p> <p>透過線上CodingBar平台進程式練習，針對學習者在某個程式上跑了幾次的測試資料，並各給每個使用者這次撰寫程式中的 輸入 跟預期的 輸出，關於uuid則是代表針對這題目的作答紀錄。</p>	<h3>募資課程學習資料</h3> <p>資料說明    取得方式</p> <p>提供者 - Hahow</p> <p>提供消費行為以及學習互動的數據在數據其中包含“作業討論資料集”、“問題討論資料集”、“課前問答資料集”、“課程評價資料集”與“課程資料集”共計五大項目，透過五大項目進行數據分析。</p>	<h3>語言學習平台資料</h3> <p>取得方式</p> <p>提供者 - Hi家教</p> <p>提供線上語言學習平台學員課程使用紀錄資料，希望選手能分析以下三個主題(1)預測學員購課/退課率及原因分析,(2)未付費轉為付費學員之機率分析,(3)熱門課程與學員購課/使用率之關聯性分析。</p>

# Data-7

- <https://hackmd.io/@HFW1UHlIQbGqmvTevXIVQw/rk30X7kpL#Education-Database>

## Education Database

---

### 企業數據

---

教育部教育雲

[https://drive.google.com/drive/folders/1gc33eF1m26fJiTLO\\_6C9MGoHA4F6vsbM](https://drive.google.com/drive/folders/1gc33eF1m26fJiTLO_6C9MGoHA4F6vsbM)

資策會數位教育研究所

<https://drive.google.com/drive/folders/1dnZbulqTD0LvkbAEY76VWqHPIfynvMM6>

均一平台

🔗 <https://www.kaggle.com/junyiacademy/learning-activity-public-dataset-by-junyi-academy>

### 金融資料連結

---

1. 政府資料開放平台 - 投資理財資料集 (共 1700 個資料集)

<https://data.gov.tw/datasets/search?qs=tid:254+&order=pubdate>

2. StockQ - 提供國際各項金融商品的指數

<http://www.stockq.org/>

3. 台灣證券交易所 - 證券期貨資料雲服務平台 (提供連結台灣所有公開資料集的單位)

<https://data.twse.com.tw/about>

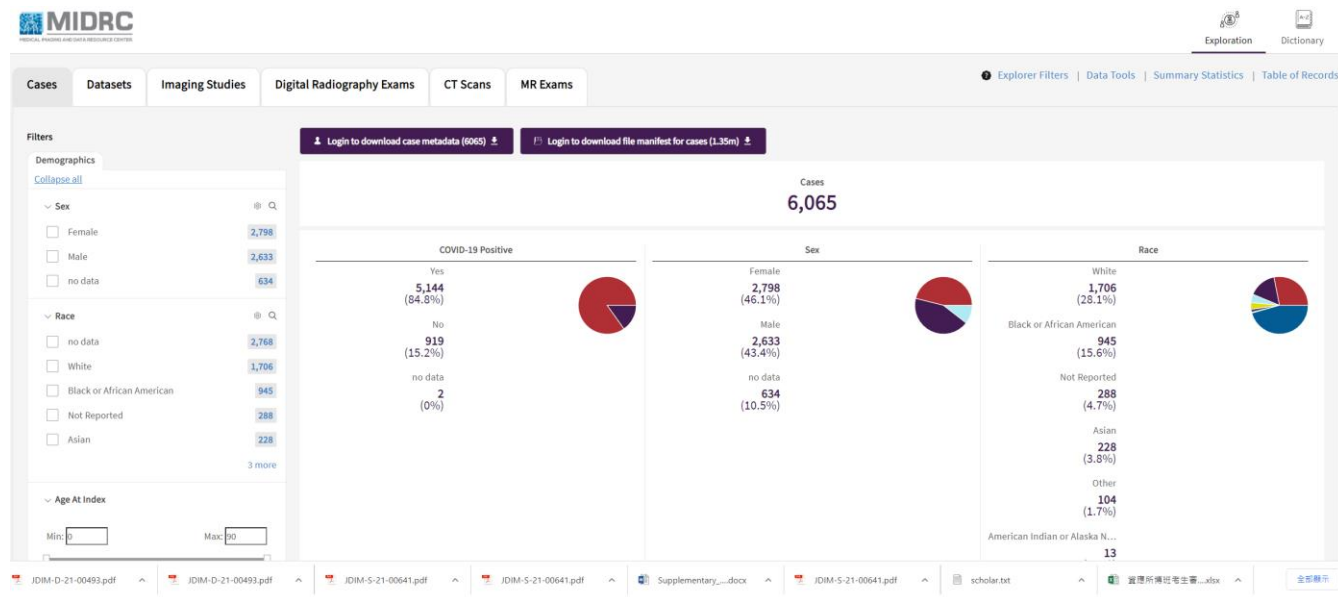
4. 財金資訊股份有限公司 - 提供新台幣、外幣、金融卡、國際卡 開放資料

<https://www.fisc.com.tw/tc/knowledge/opendata.aspx>

---

# Data-8

- <https://data.midrc.org/explorer>



# Data-9

- [https://hackmd.io/LmNy4ACZT\\_SdSBzZlDsgvA](https://hackmd.io/LmNy4ACZT_SdSBzZlDsgvA)

 PUBLISHED  CHANGED 14 DAYS AGO



可使用下列但不限於以下專案,可自行新增專案

## AI 開源專案

### 前言

為提昇台灣AI落地專案及整體AI能力,邀請各方高手以開源專案進行本專案,以開源資料及開放源代碼並以BSD授權為原則,各專案也可自行約定,未約定以則採BSD,

「讓世界看到TAIWAN,就想到AI;  
看到台灣在AI的成就,就不禁讚嘆:WA!  
這是我們對自己與台灣的期許」陳昇瑋

1. 讀取類比儀表 [deepgauge](#)
2. 瑕疵檢測
3. 預防性維護
4. [台灣開源武漢肺炎醫用支援 共筆頁面](#)
5. 4D心臟超音波動態檔,預測心衰竭
6. 選股機器人,由台股選出飆股或放空標地
7. 人員物體追蹤,偵測(電子圍籬)

# Data-10

- <https://stanfordmlgroup.github.io/competitions/chexpert/>



## What is CheXpert?

CheXpert is a large dataset of chest X-rays and competition for automated chest x-ray interpretation, which features uncertainty labels and radiologist-labeled reference standard evaluation sets.

[READ THE PAPER \(IRVIN & RAJPUKAR ET AL.\)](#)

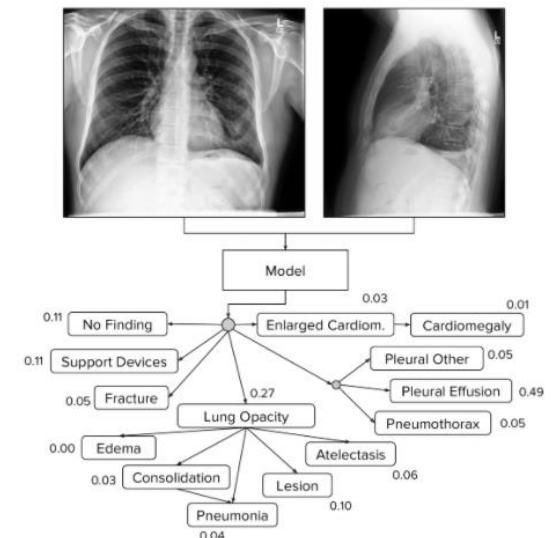
## Why CheXpert?

Chest radiography is the most common imaging examination globally, critical for screening, diagnosis, and management of many life threatening diseases. Automated chest radiograph interpretation at the level of practicing radiologists could provide substantial benefit in many medical settings, from improved workflow prioritization and clinical decision support to large-scale screening and global population health initiatives. For progress in both development and validation of automated algorithms, we realized there was a need for a labeled dataset that (1) was large, (2) had strong reference standards, and (3) provided expert human performance metrics for comparison.

## Leaderboard

Will your model perform as well as radiologists in detecting different pathologies in chest X-rays?

Rank	Date	Model	AUC	Num Rads Below Curve
1	Aug 31, 2020	SuperCNN ensemble	0.930	2.8
2	Sep 01, 2019	Hierarchical-Learning-V1 (ensemble) <i>Vingroup Big Data Institute</i> <a href="https://arxiv.org/abs/1911.06475">https://arxiv.org/abs/1911.06475</a>	0.930	2.6
3	Oct 15, 2019	Conditional-Training-LSR ensemble	0.929	2.6
4	Dec 04, 2019	Hierarchical-Learning-V4 (ensemble) <i>Vingroup Big Data Institute</i> <a href="https://arxiv.org/abs/1911.06475">https://arxiv.org/abs/1911.06475</a>	0.929	2.6





# Data-11

- <https://openneuro.org/>



**OpenNEURO**

A free and open platform for sharing MRI,  
MEG, EEG, iEEG, and ECoG data

Search Datasets



[Browse All Public Datasets](#)

420

Public Datasets

13606

Participants



Get Data



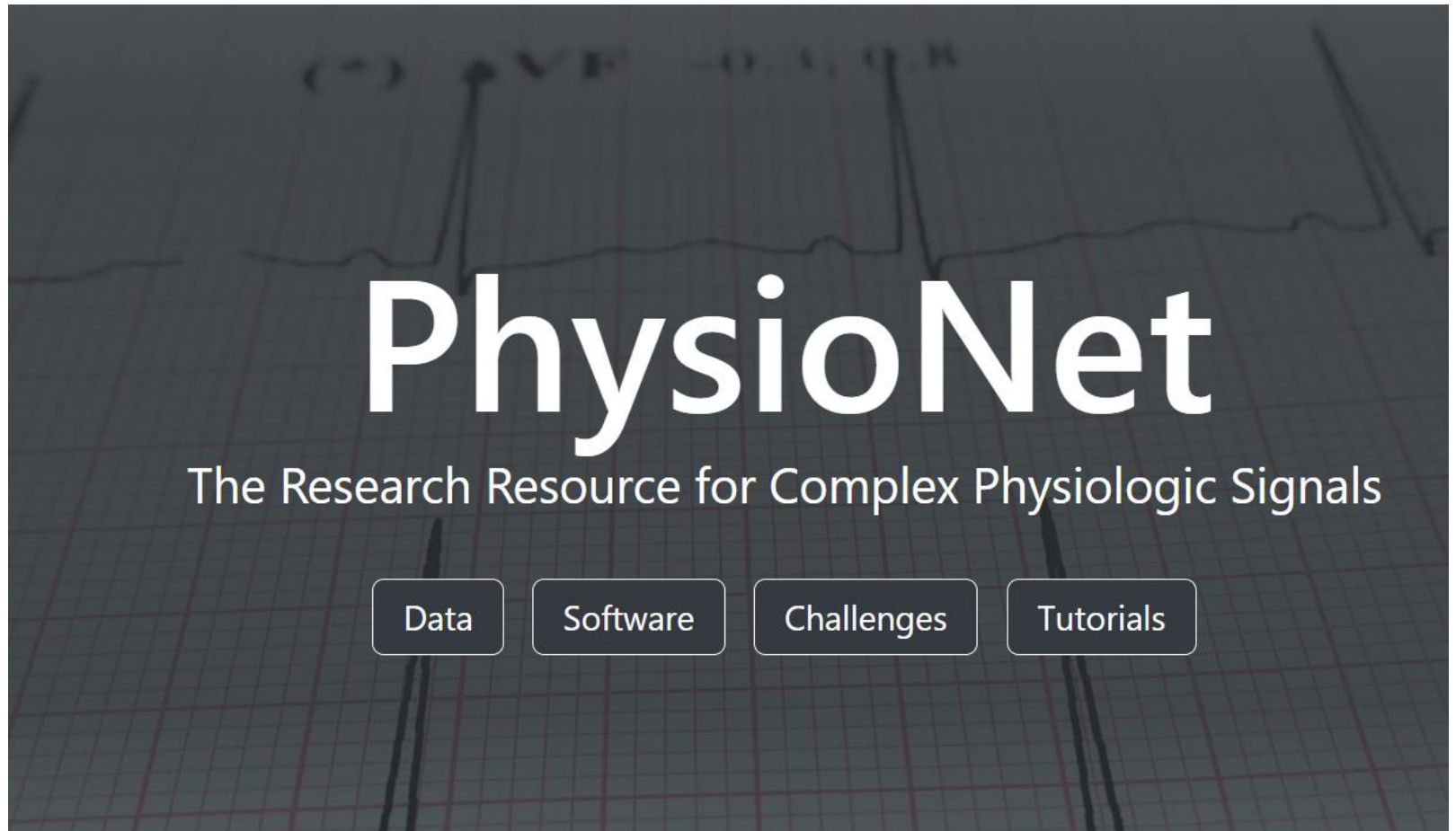
Share Data



Use Data

# Data-12

- <https://physionet.org/>



# Project-1: TBrain

- <https://tbrain.trendmicro.com.tw/Competitions/Details/18>

The screenshot shows the TBrain AI Competition website. The header includes the TBrain logo and navigation links: Home, Competitions, Discussion, Datasets, Success Story, and a Sign In button. A notice box on the left contains two announcements: one dated 2021/11/02 about a new Q&A section, and another dated 2021/11/01 about a change in training data compression. The main title is '玉山人工智能公開挑戰賽2021冬季賽 - 信用卡消費類別推薦', with a '進行中' (In Progress) status. Below the title is a navigation bar with 'Overview', 'Leaderboard', and 'Download Dataset'. The 'Overview' section, titled '競賽說明', provides details about the competition, including the number of participants (362 teams) and the total prize pool (230,000 NTD). It also mentions the competition dates from 10/27/2021 to 1/5/2022.

TBrain AI實戰吧

Home Competitions Discussion Datasets Success Story Sign In

【公告】  
2021/11/02 :  
新增Q&A於比賽辦法網站。  
2021/11/01 :  
調整訓練資料集壓縮方式，資料內容不變。

## 玉山人工智能公開挑戰賽2021冬季賽 - 信用卡消費類別推薦

進行中

Overview Leaderboard Download Dataset

### 競賽說明

【聰明消費來預3—信用卡消費類別推薦】

你知道嗎，全台灣信用卡流通量竟然遠遠勝過全國總人口，高達5100多萬張！而將近每三個人就有一張玉山信用卡！

在台灣，多數人都有用信用卡消費的經驗，銀行也推出各式各樣的行銷活動，期望吸引消費者使用該銀行信用卡，但！這些行銷真的都是用戶想要的嗎？過多的行銷內容不會打擾顧客嗎？

362 參賽隊伍

總獎金 新台幣 23 萬元

開始 10/27/2021 結束 1/5/2022

# Project-2&3: Aldea

- <https://aidea-web.tw/topic/ca331d1c-a506-43bd-8cf2-d86f9a39e1c2>
- <https://aidea-web.tw/topic/e4639c0b-f28d-4f8e-aaba-eeba47fea3e2>



清華大學 / 機器學習概論 (郭柏志 / 電影情緒評論分類)



[Introduction](#) [Rules](#) [Data](#) [Upload](#) [Forum](#) [Team up](#) [Leaderboard](#)

### Introduction

情緒分析是現今自然語言處理 (NLP) 最重要的應用之一，藉此可以理解人們文字與言論中的情緒，理解評論的真實走向，幫助資料庫建立更加準確的評論系統。

此議題中蒐集了許多簡短的電影文字評論，參賽者需使用現有的評論，訓練人工智慧模型來分析、判斷類似的文字評論帶有正面或是負面的情緒。

Registration

68

days

10

hrs

01

min

44

secs

Time to team-up deadline



清華大學 / 機器學習概論 (郭柏志 / 尋找病媒蚊孳生源-積水容器影像物件辨識)



[Introduction](#) [Rules](#) [Data](#) [Upload](#) [Forum](#) [Team up](#) [Leaderboard](#)

### Introduction

登革熱，是一種藉由病媒蚊叮咬而感染的急性傳染病。每年台灣入夏後，容易引發登革熱感染潮，典型登革熱的症狀會有突發性的高燒 ( $\geq 38^{\circ}\text{C}$ )，頭痛、後眼痠痛、肌肉痛、關節痛及出疹等現象；然而，若是先後感染不同型別之登革病毒，有更高機率導致較嚴重的臨床症狀，如果沒有及時就醫，死亡率高達20%以上。

要有效防範登革熱，清除病媒蚊產卵、幼蟲孳生的地方為根本之道，病媒蚊孳生源泛指所有「積水

Registration

68

days

10

hrs

01

min

25

secs

Time to team-up deadline

# I don't have machine.

- Computational resources
  - Google Colab
  - NCHC
    - Weekend: 11/15~1/15
    - 1T storage space
    - Contact TA



I don't have time.

- Drop this course as soon as possible.