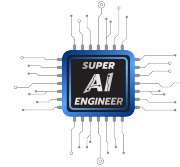


Five Final Hackathon

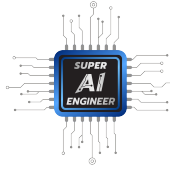
Super AI Engineer Season 5 Level 2 Individual Test



Fast Radio Burst Detection

Data

278, 278, None
279, 279, None
280, 280, None
281, 281, None
282, 282, None
283, 283, None
284, 284, None
285, 285, None
286, 286, None
287, 287, None
288, 288, None
289, 289, None
290, 290, None
291, 291, None
292, 292, None
293, 293, None
294, 294, None
295, 295, None
296, 296, None
297, 297, None
298, 298, None
299, 299, Pulse
300, 300, Pulse
301, 301, Pulse
302, 302, None
303, 303, None
304, 304, None
305, 305, None
306, 306, None
307, 307, None
308, 308, None
309, 309, None
310, 310, None
311, 311, None
312, 312, None
313, 313, None
314, 314, None
315, 315, None



Data

frequency

256

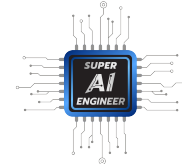
time

278, 278, None
279, 279, None
280, 280, None
281, 281, None
282, 282, None
283, 283, None
284, 284, None
285, 285, None
286, 286, None
287, 287, None
288, 288, None
289, 289, None
290, 290, None
291, 291, None
292, 292, None
293, 293, None
294, 294, None
295, 295, None
296, 296, None
297, 297, None
298, 298, None
299, 299, Pulse
300, 300, Pulse
301, 301, Pulse
302, 302, None
303, 303, None
304, 304, None
305, 305, None
306, 306, None
307, 307, None
308, 308, None
309, 309, None
310, 310, None
311, 311, None
312, 312, None
313, 313, None
314, 314, None
315, 315, None

B0531+21_2020-05-31-11_36_46_0001023.npy

	0	1	2	...	255
0	182.20207 2	175.16604 6	191.80407 7	...	308.23114 0
1	193.30155 9	186.56225 6	159.66279 6	...	319.43466 2
2	213.11222 8	184.44229 1	189.07644 7	...	288.76303 1
...
262143	187.20697 0	187.51857 0	177.88314 8		347.81857 3

Data



None

Pulse

Pulse

Pulse

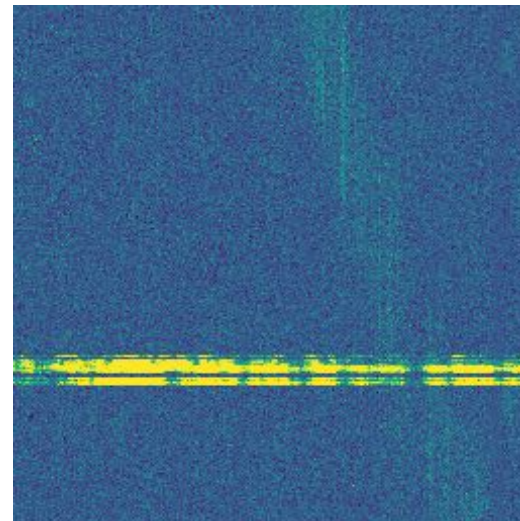
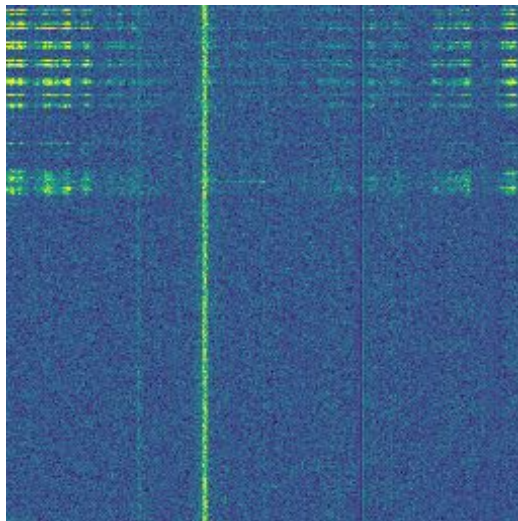
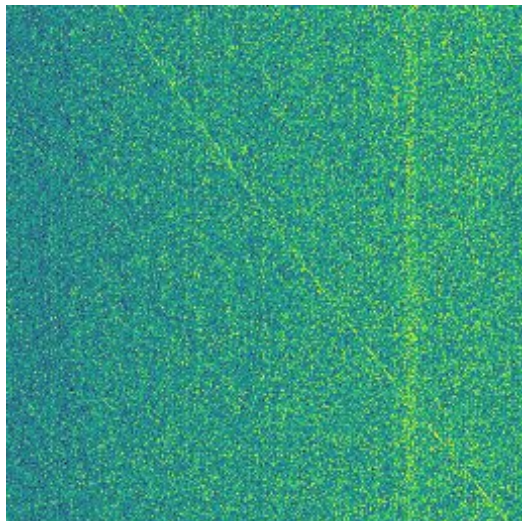
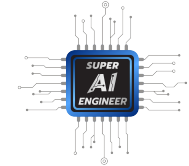
None

278, 278, None
279, 279, None
280, 280, None
281, 281, None
282, 282, None
283, 283, None
284, 284, None
285, 285, None
286, 286, None
287, 287, None
288, 288, None
289, 289, None
290, 290, None
291, 291, None
292, 292, None
293, 293, None
294, 294, None
295, 295, None
296, 296, None
297, 297, None
298, 298, None
299, 299, Pulse
300, 300, Pulse
301, 301, Pulse
302, 302, None
303, 303, None
304, 304, None
305, 305, None
306, 306, None
307, 307, None
308, 308, None
309, 309, None
310, 310, None
311, 311, None
312, 312, None
313, 313, None
314, 314, None
315, 315, None

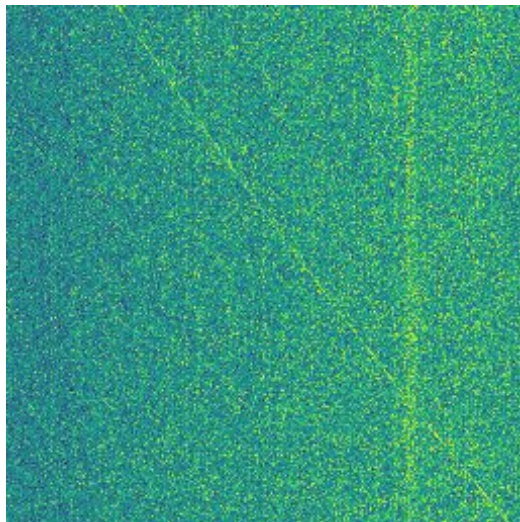
B0531+21_2020-05-31-11_36_46_0001023_labels.csv

index	labels
0	None
1	None
...	...
534	Pulse
...	...
816	Broad
...	...
908	Broad+Pulse
...	...
1023	None

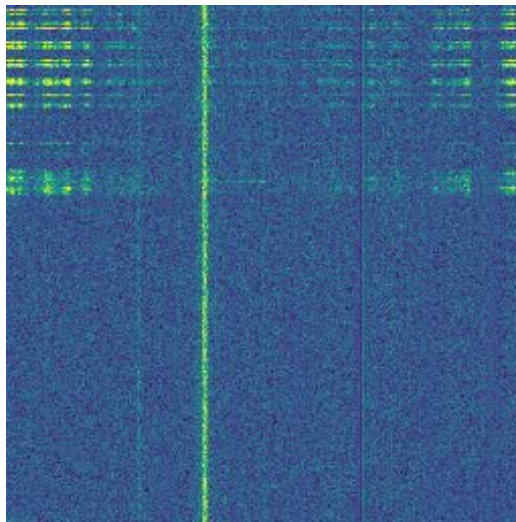
Train Label



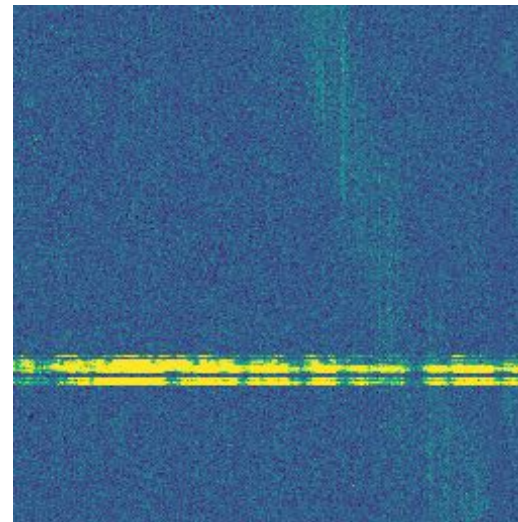
Train Label



Pulse+Narrow

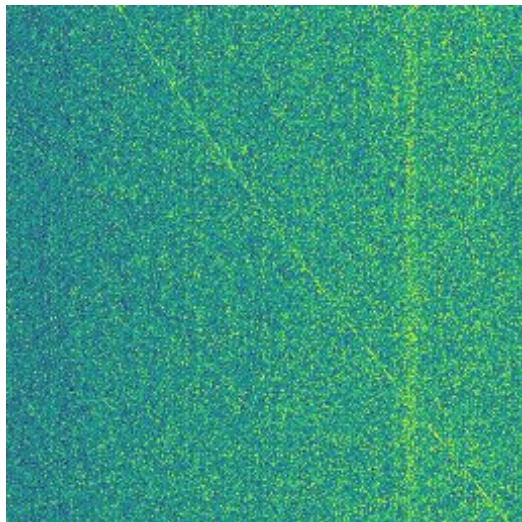


Broad+Narrow

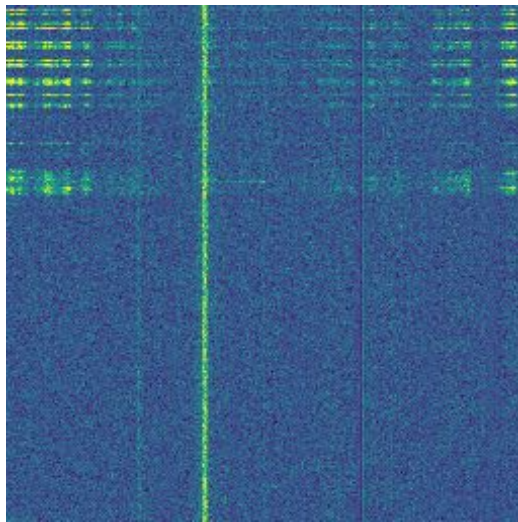


Pulse+Broad

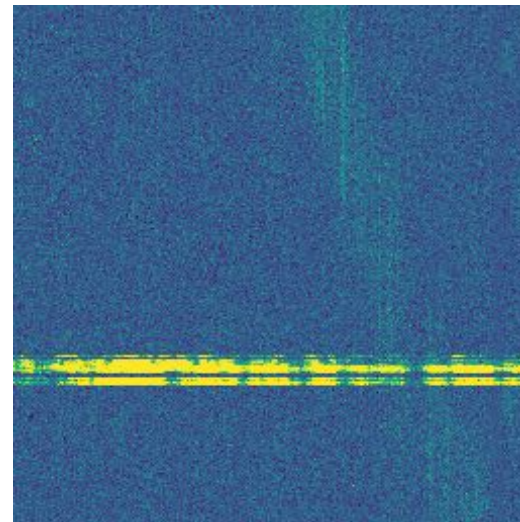
Train Label



Pulse+Narrow
Narrow+Pulse

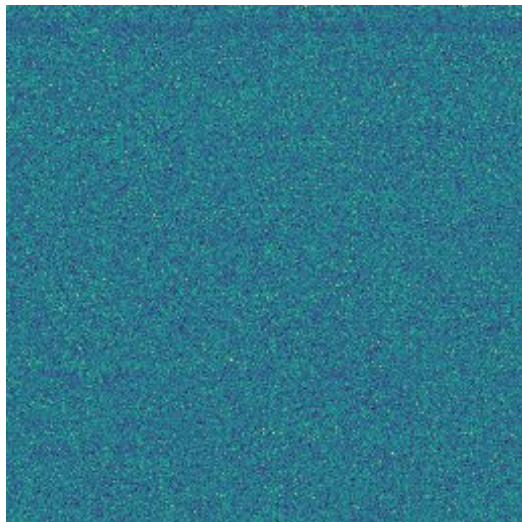
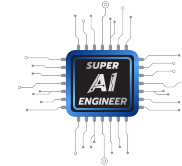


Broad+Narrow
Narrow+Broad

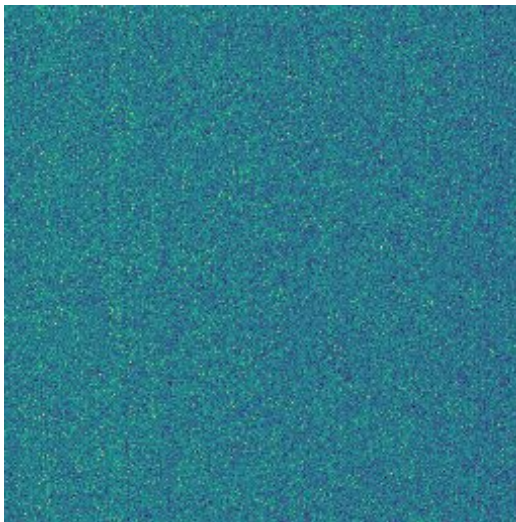


Pulse+Broad
Broad+Pulse

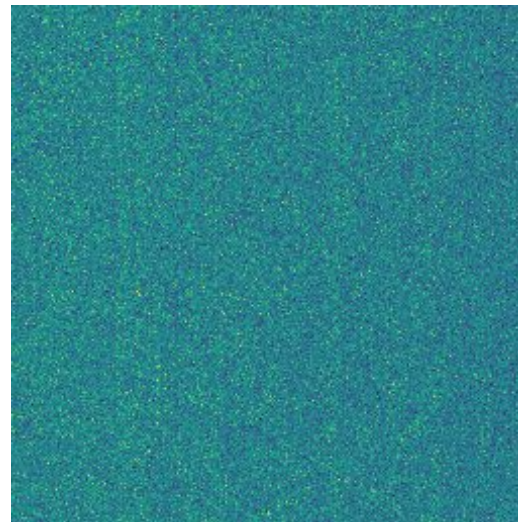
Train Label



None



None



None

Evaluation

- F1 Macro

Answer Format

Multiclass binary

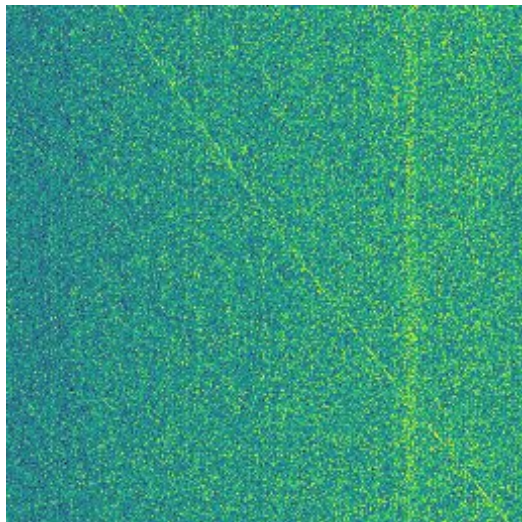
- Pulse,Broadband,Narrowband

No “Unlabeled” and “Uncertain” will be used in for scoring.

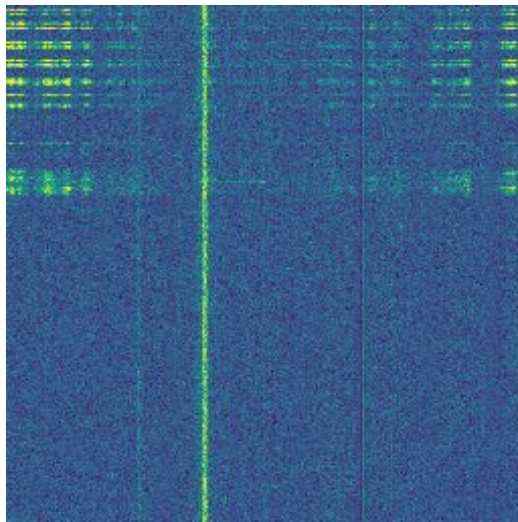
sample_submission.csv

id	pulse	broad	narrow
0_0	0	0	0
0_1	0	0	0
0_2	0	0	0
...
32_403			

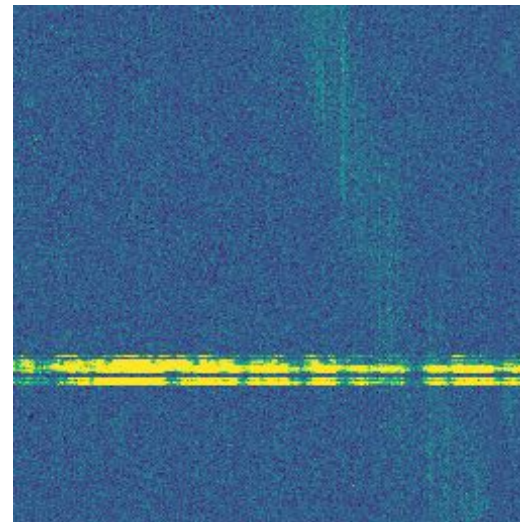
Answer Format



Pulse+Narrow

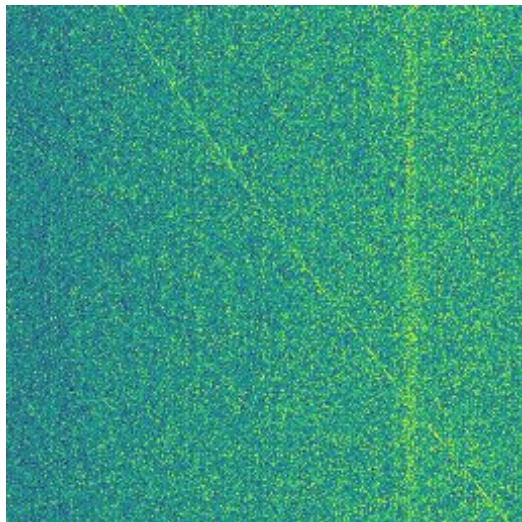


Broad+Narrow

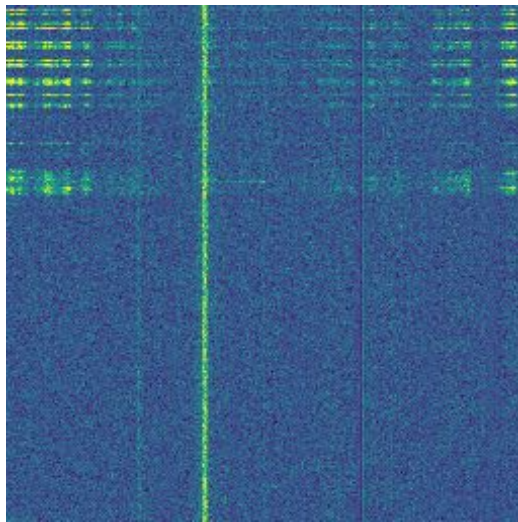


Pulse+Broad

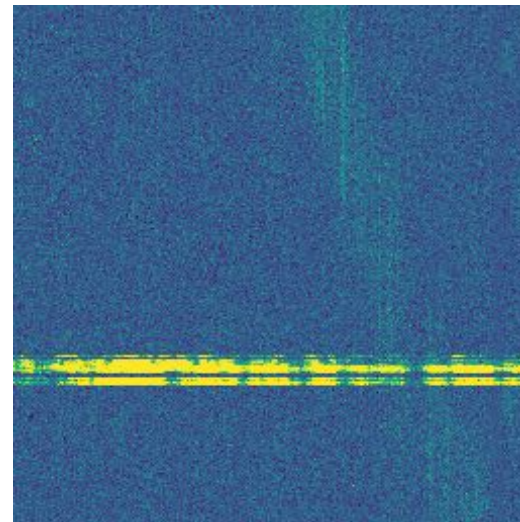
Answer Format



1,0,1

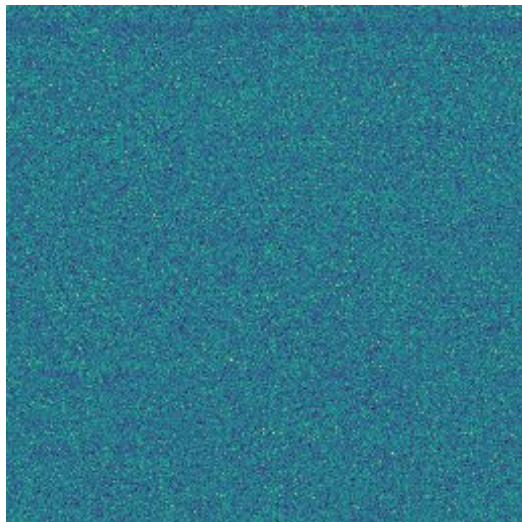


0,1,1

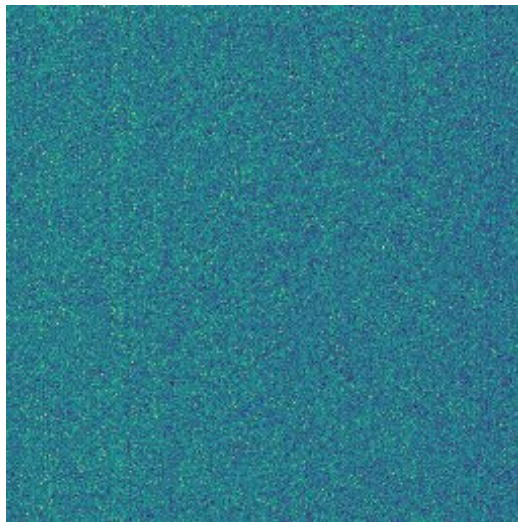


1,1,0

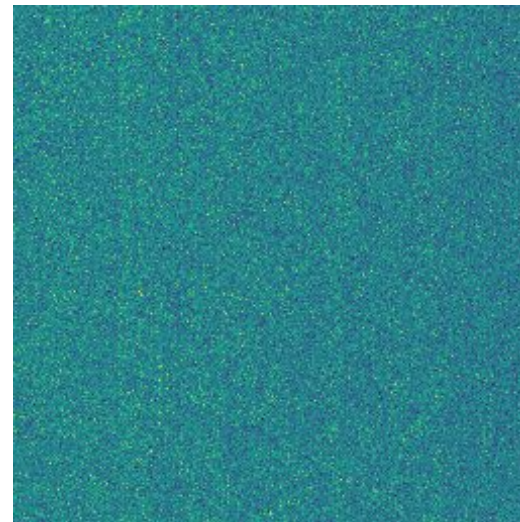
Answer Format



None

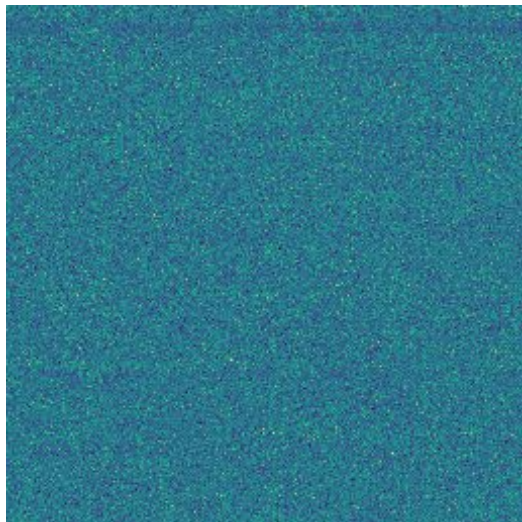


None

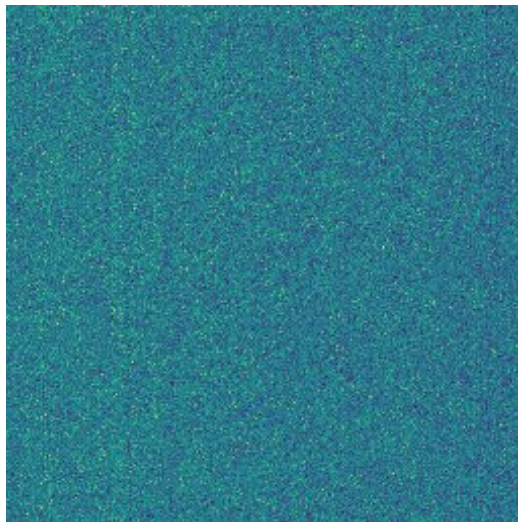


None

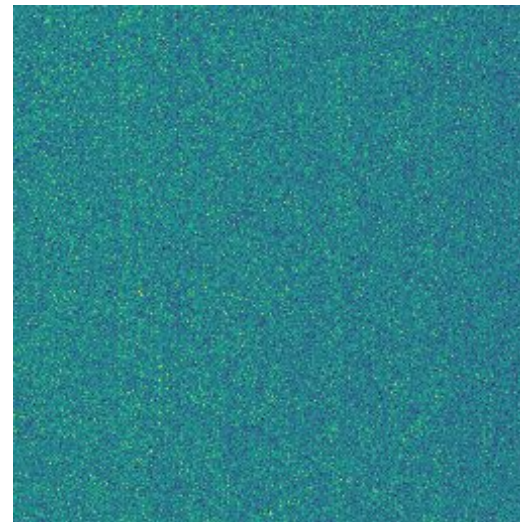
Answer Format



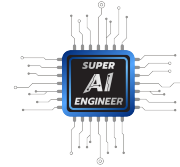
0,0,0



0,0,0



0,0,0



Temperature Prediction

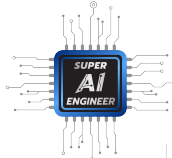
- ทำนายค่าอุณหภูมิในคอลัมน์สุดท้าย (temp) โดยนำข้อมูลที่ได้จากเซ็นเซอร์มาจากหลายๆ สถานที่
- ไฟล์ที่ต้องส่งเป็น IOT_Submit.csv โดยมีคอลัมน์ id และ temp โดย id นำมาจากคอลัมน์แรกของ IOT_test.csv และ temp คืออุณหภูมิของช่วงเวลานั้นที่ predict ได้

Output Format:

```
id,temp  
1,25.6  
2,30.1  
3,24.7
```

ไฟล์ที่เกี่ยวข้อง

1. IOT_train.CSV ชุดข้อมูลที่ใช้ในการ train จำนวน 13,635 แถว
2. IOT_test.CSV ชุดข้อมูลทดสอบ จำนวน 3,413 แถว
3. IOT_Submit.csv ตัวอย่างไฟล์ที่ใช้ส่ง ต้องมีข้อมูลทั้งหมด 3,413 แถว (พร้อมเฉลย 3 แถวแรก)



18 Features

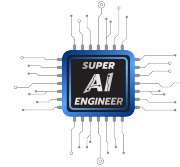
DetailCompactColumn10 of 19 cols

△ mac		△ station_name		# tambon_code		△ tambon_namt		# amphur_code	
3C71BF18CEA4	13%	โรงเรียนทองสูงสาม...	13%			หนองสูงเหนือ	13%		
3C71BF17CDBC	13%	โรงเรียนบ้านนา	13%			ตลาด	13%		
Other (10159)	75%	Other (10159)	75%	103k860k		Other (10159)	75%	10268602	
3C71BF18EA64		บ้านนา_2		300903		คำมิ่ง		3009	
3C71BF18EA64		บ้านนา_2		300903		คำมิ่ง		3009	
3C71BF18EA64		บ้านนา_2		300903		คำมิ่ง		3009	
3C71BF18EA64		บ้านนา_2		300903		คำมิ่ง		3009	
3C71BF18EA64		บ้านนา_2		300903		คำมิ่ง		3009	
3C71BF18EA64		บ้านนา_2		300903		คำมิ่ง		3009	
3C71BF18EA64		บ้านนา_2		300903		คำมิ่ง		3009	
3C71BF18EA64		บ้านนา_2		300903		คำมิ่ง		3009	
3C71BF18EA64		บ้านนา_2		300903		คำมิ่ง		3009	
3C71BF18EA64		บ้านนา_2		300903		คำมิ่ง		3009	
3C71BF18EA64		บ้านนา_2		300903		คำมิ่ง		3009	
3C71BF18EA64		บ้านนา_2		300903		คำมิ่ง		3009	
3C71BF18EA64		บ้านนา_2		300903		คำมิ่ง		3009	
3C71BF18EA64		บ้านนา_2		300903		คำมิ่ง		3009	
3C71BF18EA64		บ้านนา_2		300903		คำมิ่ง		3009	
3C71BF18EA64		บ้านนา_2		300903		คำมิ่ง		3009	
3C71BF18EA64		บ้านนา_2		300903		คำมิ่ง		3009	

DetailCompactColumn19 of 19 columns

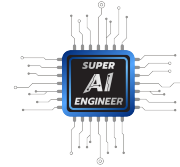
About this fileThis file does not have a description yet.

	# pm2.5	# rainfall	# wind_direct	# wind_speed	# temp
368	0368	034.1	045.0338	015.1k	17.251.1
		0.0	45.0	4.3	32.6
		0.0	45.0	1.9	37.9
		0.0	135.0	3.2	26.4
		0.0	157.5	0.0	25.6
		0.0	180.0	2.3	36.2
		0.0	180.0	5.1	36.2
		0.0	157.5	0.5	29.1
		0.0	135.0	3.4	30.8



Evaluation Metric:

Mean Absolute Error (MAE)



Extraction Question Answering

Extractive Long-Form Question Answering

Q: ราคาต้นงวดคืออะไร

A: ราคาต้นทุนหรือราคาซื้อ

Impossible Question

Q: ลักษณะแนวทางในการ
ระดมทุนมีอะไรบ้าง

A:

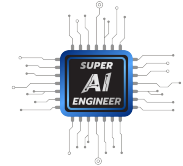
ผลตอบแทนและความเสี่ยงของกองทุนรวม

2. ผลตอบแทนของกองทุนรวม

ผลตอบแทนของกองทุนรวมประกอบด้วย 1. ผลตอบแทนจากส่วนต่างราคา (Capital gain) 2. เงินปันผล (Dividend) ซึ่งมีสูตรการคำนวณอัตราผลตอบแทนดังนี้

- * $\text{Capital gain} = (\text{ราคาปลายงวด} - \text{ราคาต้นงวด}) / \text{ราคาต้นงวด}$
- * $\text{เงินปันผล} = \text{เงินปันผล} / \text{ราคาต้นงวด}$
- * $\text{อัตราผลตอบแทนรวม} = (\text{เงินปันผล} + \text{ราคาปลายงวด} - \text{ราคาต้นงวด}) / \text{ราคาต้นงวด}$
- ราคาของกองทุน หมายถึง NAV per unit (มูลค่าสินทรัพย์สุทธิต่อหน่วย)
- ราคาปลายงวดมีความหมายรวมถึงราคาขาย
- ราคาต้นงวดมีความหมายรวมถึงราคาต้นทุนหรือราคาซื้อ
- ราคาต้นทุน คือ ราคาที่รวมค่าใช้จ่ายในการซื้อหลักทรัพย์

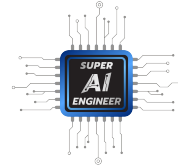
Data



Files

- context/* - Documents
- train.csv - Training data
- test.csv - Test data
- sample_submission.csv - ตัวอย่างไฟล์ที่ใช้ส่ง ต้องมีข้อมูลทั้งหมด 275 แถว (พร้อมเฉลย 3 แถวแรก)

Data



▼ context

- 1345136.txt
- 1345137.txt
- 1345138.txt
- 1345139.txt
- 1345140.txt
- 1345141.txt
- 1345142.txt
- 1345143.txt
- 1345144.txt
- 1345145.txt
- 1345146.txt
- 1345147.txt
- 1345148.txt
- 1345149.txt
- 1345150.txt

ตัวอย่าง context : 1345170.txt

ความรู้พื้นฐานของตราสารหนี้

2. ความแตกต่างของตราสารทุนและตราสารหนี้
ตราสารทุน

ลักษณะผลตอบแทน: เงินปันผล

สถานะผู้ลงทุน: เจ้าของ

ระดับความเสี่ยง: สูง

ระยะเวลา (อายุ): ไม่มีกำหนด

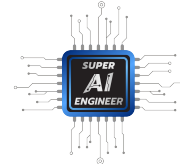
สภาพคล่อง: สูง

สิทธิการเรียกร้อง: เมื่อผู้ออกตราสารล้มละลาย เงินที่ได้รับจากการขายสินทรัพย์จะนำมาชำระหนี้คืนให้ผู้ถือตราสารหนี้ก่อนผู้ถือตราสารทุน

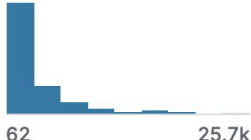
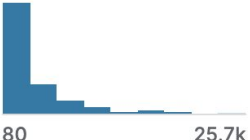
...

...

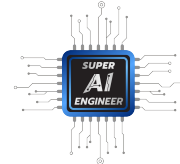
Data




train.csv

file		question	answer	answer_start	answer_end
1345160.txt	13%	410 unique values	400 unique values	 62 25.7k	 80 25.7k
1345173.txt	9%				
Other (318)	78%				
1345136.txt		ใครเป็นผู้ออกตราสารหนี้ภาคเอกชน ไร่ใบตราสาร	ไม่มีการรวมศูนย์อย่างตราสารภาครัฐ แต่ส่วนใหญ่ดำเนินการโดยศูนย์รับฝากหลักทรัพย์ (TSD)	7716	7799
1345195.txt		Penny Stock หุ้นเหรียญมีลักษณะอย่างไร	* หุ้นที่มีราคาซื้อขายต่อหน่วยต่ำ (ในสหรัฐฯ คือ ราคาต่ำกว่า 5 ดอลลาร์, ในสหราชอาณาจักร United Kingdom...	3019	3138
1345200.txt		ราคาดันจวดคืออะไร	ราคาดันทุนหรือราคาซื้อ	501	522

Data



sample_submission.csv

id	answer
	[null] 99% สูงสุด +30% ต่ำสุด -... 0% Other (1) 0%
1	สูงสุด +30% ต่ำสุด -30% ของราคาปิดสินทรัพย์อ้างอิงวัน ก่อนหน้า x สิทธิการซื้อ สินทรัพย์อ้างอิงของ 1 Wa...
2	
3	จะจ่ายเป็น รายเดือน, ราย ไตรมาส, ทุกครึ่งปี หรือ ราย ปี และแม้ว่าผู้ซื้อจะพึงข้อ ตราสารหนี้มาก่อนวันจ...

Submission File Format

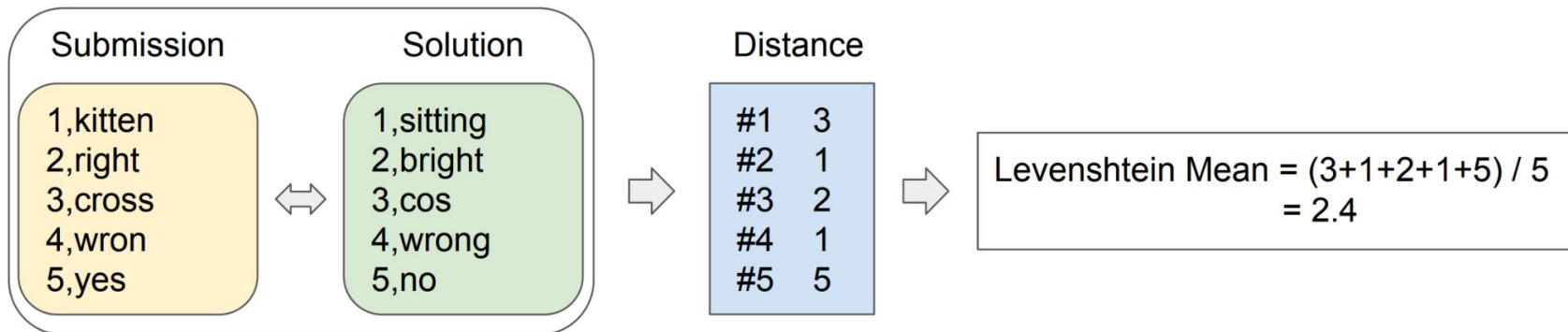
`id`	`answer`
รหัสคำถาม	คำตอบ (ถ้าเป็นคำถามที่ไม่มีคำตอบใน context ต้องไม่ใส่ ตามตัวอย่าง sample submission id=2)

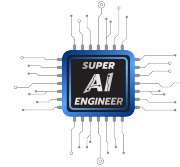
QA Evaluation - Levenshtein Distance

“kitten” vs “sitting” Levenshtein Distance (Edit Distance) = 3

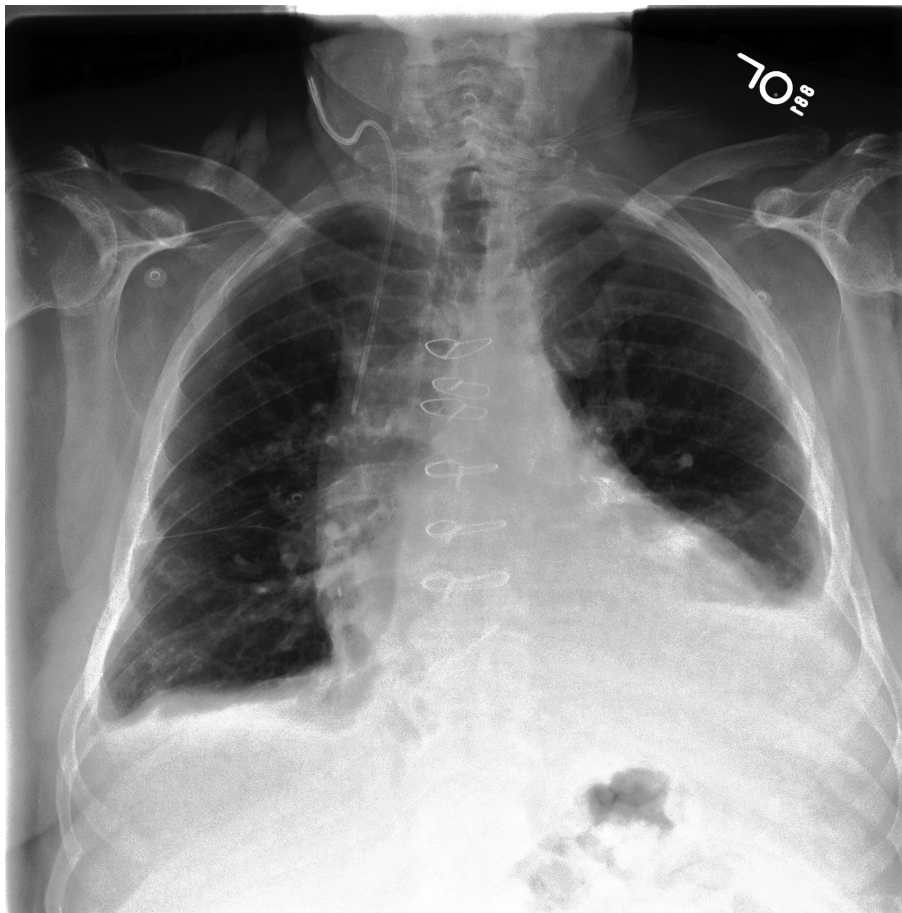
1. kitten → sitten (substitution of "s" for "k"),
2. sitten → sittin (substitution of "i" for "e"),
3. sittin → sitting (insertion of "g" at the end).

Levenshtein Mean





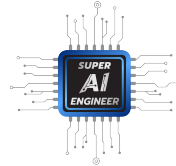
Chest Disease Detection



Predict multiple chest X-ray conditions from images across 9 possible classes

- Atelectasis : ภาวะปอดแฟบ
- Cardiomegaly : โรคหัวใจโต
- Consolidation
- Edema : อาการบวมน้ำในปอด
- Enlarged Cardio mediastinum
- Fracture : กระดูกหัก
- Lung Lesion : รอยโรคในปอด
- Lung Opacity
- No Finding : ไม่พบความผิดปกติ (หากเป็น 1, อื่นๆ จะเป็น 0)

Data



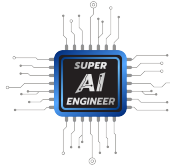
Files

- image/: Training and Test images.
- train.csv: Contains image_id and label for each training image.
- sample_submission.csv: Format for submission with image_id and predicted label.

Training Set with 9963 samples

Test set with 2506 samples

Data



train.csv (647.78 kB)

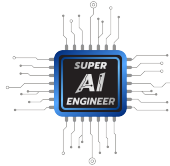


Detail Compact Column

10 of 14 columns ▾

▲ filename	# Atelectasis	# Cardiomegaly	# Consolidation	# Edema	# Enlarged Cardio...
9963 unique values					
cxr00030.jpg	1.0	0.0	0.0	0.0	0.0
cxr00033.jpg	0.0	0.0	0.0	0.0	0.0
cxr00034.jpg	0.0	0.0	0.0	0.0	0.0
cxr00035.jpg	0.0	0.0	0.0	0.0	0.0
cxr00038.jpg	0.0	0.0	0.0	0.0	1.0
cxr00042.jpg	0.0	0.0	0.0	0.0	1.0
cxr00044.jpg	0.0	0.0	0.0	0.0	0.0
cxr00047.jpg	0.0	1.0	0.0	0.0	0.0
cxr00048.jpg	0.0	1.0	0.0	1.0	0.0
cxr00049.jpg	0.0	0.0	0.0	1.0	0.0
cxr00051.jpg	0.0	0.0	0.0	1.0	0.0
cxr00052.jpg	0.0	0.0	0.0	0.0	0.0
cxr00057.jpg	0.0	1.0	0.0	1.0	0.0

Data



test_submission.csv (65.45 kB)



Detail Compact Column

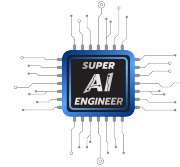
10 of 14 columns ▾

About this file



This file does not have a description yet.

▲ filename	# Atelectasis	# Cardiomegaly	# Consolidation	# Edema	# Enlarged Cardio...
2506 unique values	2506 total values	2506 total values	2506 total values	2506 total values	2506 total values
cxr00001.jpg	0.0	0.0	0.0	0.0	0.0
cxr00002.jpg	0.0	0.0	0.0	0.0	0.0
cxr00003.jpg	0.0	1.0	0.0	0.0	0.0
cxr00007.jpg					
cxr00008.jpg					
cxr00010.jpg					
cxr00011.jpg					



Evaluation Metric:

sample-average F1


SPAI Sorting Hat



Objective

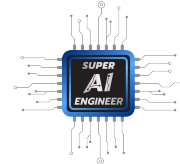


Predict the **house** each user belongs to based on their **attendance logs** during SPAI Lv2 Onsite.

user_id	datetime
170 unique values	 2024-04-19 2024-05-17
5931fa6a-af6b-43b8-babb-28349854d406	2024-04-19 17:30:08.827555
5f1a2ecf-943d-446b-b94d-d7d1ad64213d	2024-04-19 17:30:11.405020
6e009a96-2b8b-4048-8b0c-20a7924a4b4f	2024-04-19 17:30:14.433098
cdd098ee-4295-4839-b1a3-4d44840feacd	2024-04-19 17:30:17.214453
f922a5d4-3651-46f7-aa98-307c045dd4b5	2024-04-19 17:30:21.556805
2ad18754-2522-4fa9-af12-bccdf42d3905	2024-04-19 17:30:24.008379
3bd6dcac-904c-4614-be50-9c3949cb3405	2024-04-19 17:30:26.610108

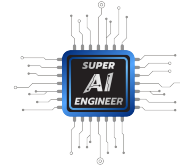
user_id	house
102 unique values	6 unique values
4b2569be-cd32-40aa-9cc4-9e13e55903de	house1
cdd098ee-4295-4839-b1a3-4d44840feacd	house3
226c662f-b1b3-40bc-9bfc-4f49a80b9041	house6
7889a31f-7351-4b4d-b872-e04e5e35b9dd	house2
0492584f-5dda-4ad6-ab2f-79c8d43a8753	house1
33a88b0d-39e1-4000-ab90-7d7079b57a46	house3
ce506c75-9d99-4d9d-b925-4f95384da3d0	house1

Data



Files

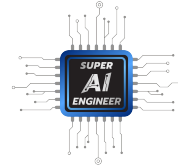
- attendance.csv: Contains logs of users' onsite attendance during SPAI Lv2 Onsite (user_id, datetime).
- train.csv: Training set with house assignments for users (user_id, house).
- submission.csv: Sample submission format for predictions (user_id, house).



Evaluation Metric:

weighted-average F1

Rules



- ผู้เข้าอบรมต้องใช้ คอมพิวเตอร์ส่วนตัวในการสอบ
- สามารถใช้เครื่องมือ เช่น VSCode, Google Colab, Kaggle หรือแพลตฟอร์มอื่นๆ โดยต้องปิดฟังก์ชันการเชื่อมต่อ AI
- ผู้เข้าอบรมจะได้รับทรัพยากร LANTA 100 SHr ต่อคน
- ห้ามใช้ Generative AI ทุกชนิด (เช่น ChatGPT, GitHub Copilot ฯลฯ) ในการเขียนโค้ดหรือสอบถามข้อมูล
- ห้ามใช้โซเชียลมีเดียแพลตฟอร์ม (เช่น Facebook, Discord ฯลฯ) ตลอดการแข่งขัน
- อนุญาตให้ ค้นหาข้อมูลบนอินเทอร์เน็ต หรือเปิด แหล่งข้อมูลที่เตรียมไว้ล่วงหน้า
- ห้ามแชร์ข้อมูลหรือโค้ดกับผู้อื่นระหว่างการสอบ
- ห้ามติดต่อกับบุคคลภายนอก เว้นแต่ได้รับอนุญาตจากกรรมการ (เช่น ผู้ปกครองในกรณีจำเป็น)
- ห้ามนำอุปกรณ์คอมพิวเตอร์ออกจากห้องสอบตลอดระยะเวลาการสอบ
- อนุญาตให้ใช้ AutoML ในการทดสอบ
- ห้ามใช้ Third-party API
- ห้ามใช้ Private Dataset

Important Information

ระยะเวลาการแข่งขัน : Sat 5 Jul 14:00 - Sun 6 Jul 14:00

จำนวนครั้งต่อ hac ในการส่ง : 4 ครั้ง/วัน, ตัดรอบ 07.00 น. (ต่อ hackathon)

Kaggle link

Individual Test: Fast Radio Burst Detection : <https://www.kaggle.com/t/aa4ecc875124416cad5b24cbc850b9a>

Individual Test: Temperature Prediction : <https://www.kaggle.com/t/5d7c91fb91984b7181208fc85a581688>

Individual Test: Chest Disease Detection : <https://www.kaggle.com/t/f08951ef80b145d99f6344651771b9af>

Individual Test: Extraction Question Answering : <https://www.kaggle.com/t/142d23deb8d747cf9139688be84ef838>

Individual Test: SPAI Sorting Hat : <https://www.kaggle.com/t/a79e1d7dbe084f49a81dd384ecbe3519>

Good Luck!