Projet 8

Participez à une compétition Kaggle !

SIIM SOCIETY FOR IMAGING INFORMATICS IN MEDICINE
INNOVATING IMAGING INFORMATICS

kaggle™

Loridan Adrien

student machine learning engineer

medium.com/@adrien.loridan

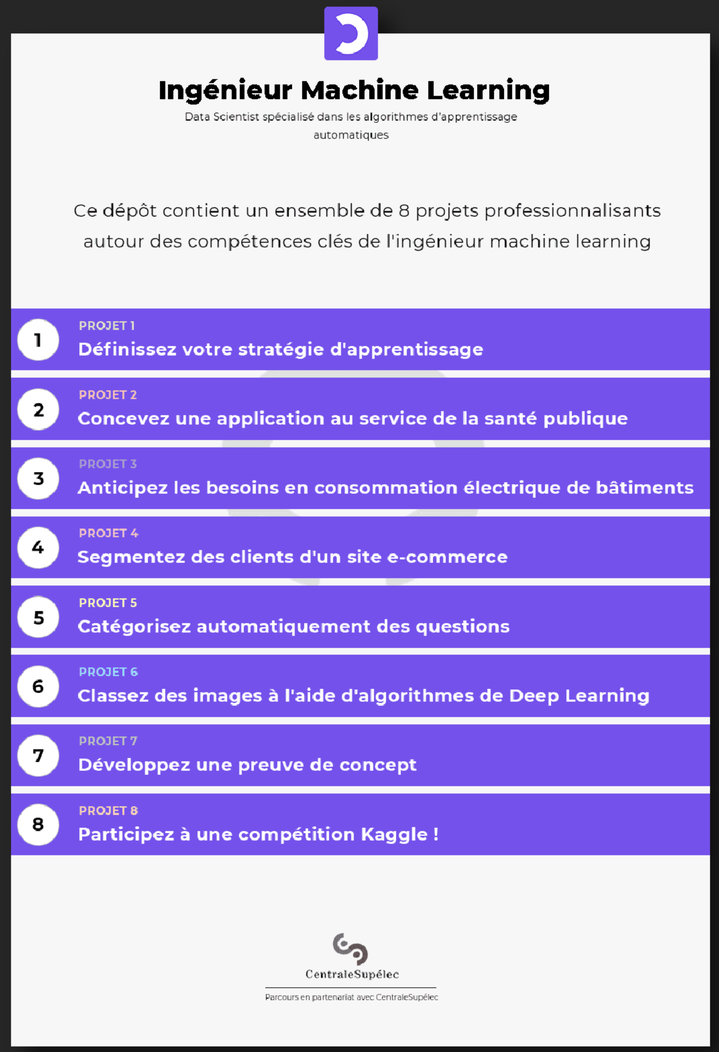
linkedin.com/in/adrien-loridan

github.com/loridan

SIIM SOCIETY FOR IMAGING INFORMATICS IN MEDICINE INNOVATING IMAGING INFORMATICS

kaggle

github.com/Loridan/openclassrooms-iml-projects



Ingénieur Machine Learning

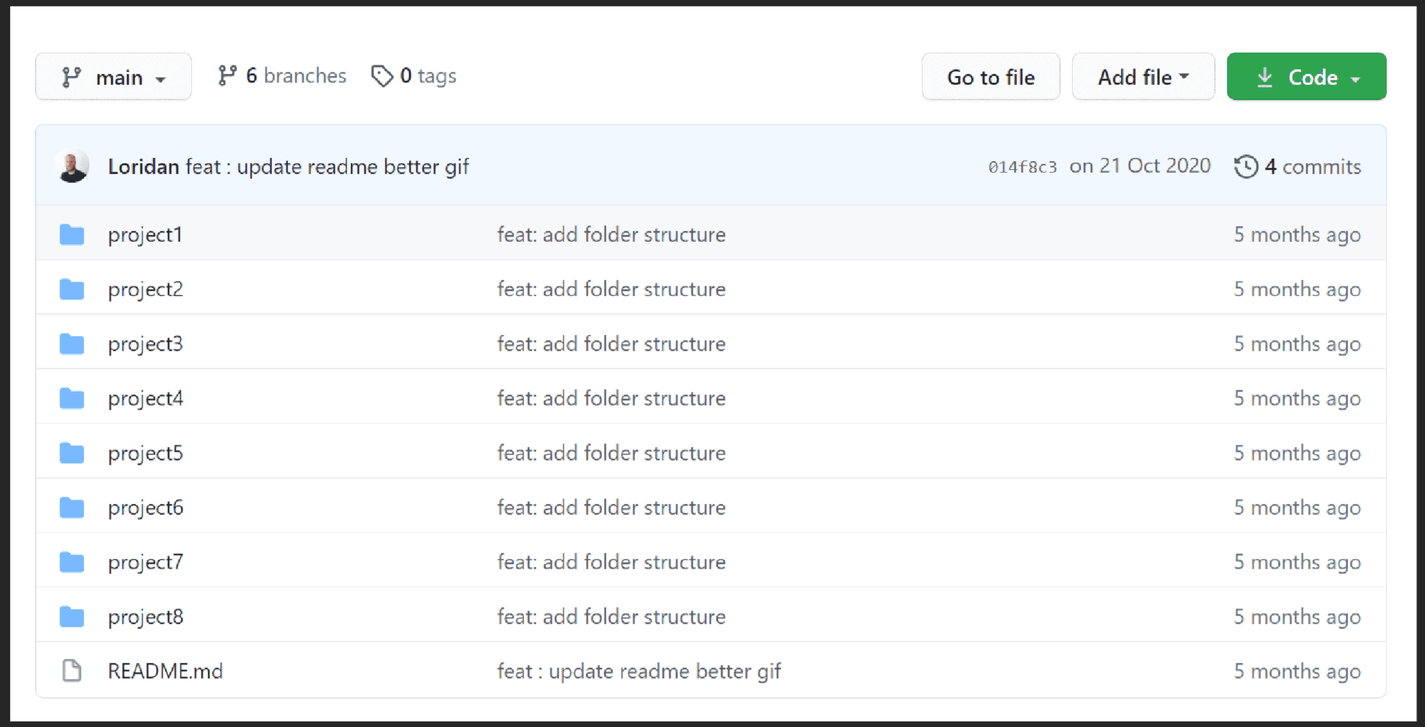
Data Scientist spécialisé dans les algorithmes d'apprentissage automatiques

Ce dépôt contient un ensemble de 8 projets professionnalisants autour des compétences clés de l'ingénieur machine learning

1. PROJET 1 — Définissez votre stratégie d'apprentissage
2. PROJET 2 — Concevez une application au service de la santé publique
3. PROJET 3 — Anticipez les besoins en consommation électrique de bâtiments
4. PROJET 4 — Segmentez des clients d'un site e-commerce
5. PROJET 5 — Catégorisez automatiquement des questions
6. PROJET 6 — Classez des images à l'aide d'algorithmes de Deep Learning
7. PROJET 7 — Développez une preuve de concept
8. PROJET 8 — Participez à une compétition Kaggle !

CentraleSupélec

Parcours en partenariat avec CentraleSupélec



main · 6 branches · 0 tags · Go to file · Add file · Code

Loridan feat : update readme better gif — 014f8c3 on 21 Oct 2020 — 4 commits

project1	feat: add folder structure	5 months ago
project2	feat: add folder structure	5 months ago
project3	feat: add folder structure	5 months ago
project4	feat: add folder structure	5 months ago
project5	feat: add folder structure	5 months ago
project6	feat: add folder structure	5 months ago
project7	feat: add folder structure	5 months ago
project8	feat: add folder structure	5 months ago
README.md	feat : update readme better gif	5 months ago

« l'ennemi est là, invisible,

insaisissable »

présentation

nous allons saisir l'opportunité de mettre à profit nos compétences acquises au cours de cette formation en participant à une compétition Kaggle active

SIIM SOCIETY FOR IMAGING INFORMATICS IN MEDICINE INNOVATING IMAGING INFORMATICS

kaggle

« l'ennemi est là, invisible,
insaisissable »

présentation

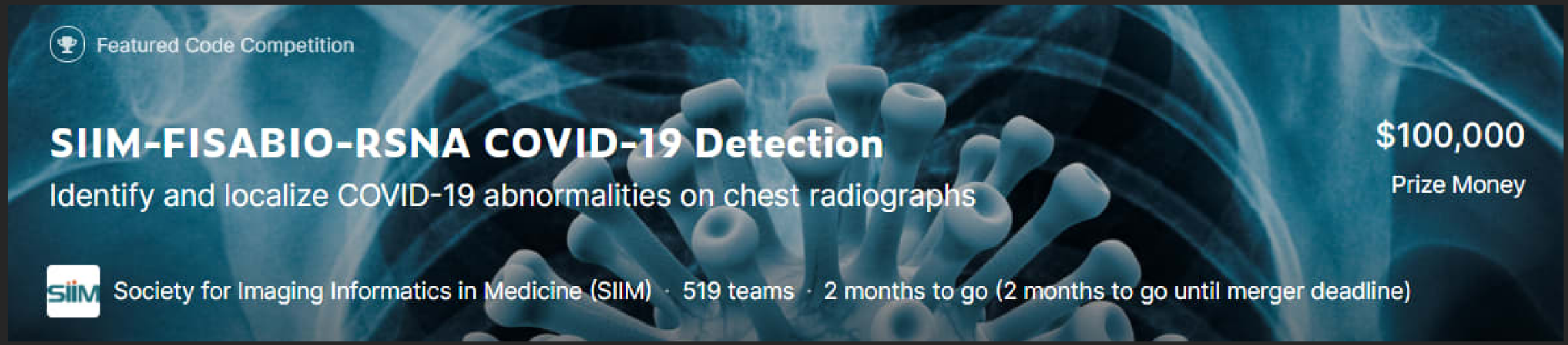
Kaggle est une plateforme collaborative proposant de résoudre des problèmes de datascience et qui dispose d'un système de rémunération sous forme de compétition

SIIM SOCIETY FOR IMAGING INFORMATICS IN MEDICINE INNOVATING IMAGING INFORMATICS

kaggle™

« l'ennemi est là, invisible,
insaisissable »

compétition



SIIM SOCIETY FOR IMAGING INFORMATICS IN MEDICINE INNOVATING IMAGING INFORMATICS kaggle

« l'ennemi est là, invisible,
insaisissable »

problème

- **on souhaite réduire la mortalité du COVID-19 en traitant plus rapidement les patients**
- **les radiographies peuvent être obtenues en quelques minutes contrairement aux tests moléculaires**

SIIM SOCIETY FOR IMAGING INFORMATICS IN MEDICINE INNOVATING IMAGING INFORMATICS kaggle™

« l'ennemi est là, invisible,
insaisissable »

problème

- **le COVID-19 ressemble beaucoup à d'autres pneumonies sur les radiographies pulmonaires**
- **l'étude d'un patient peut nécessiter plusieurs radiographies pour identifier le type de la pneumonie**

SIIM SOCIETY FOR IMAGING INFORMATICS IN MEDICINE INNOVATING IMAGING INFORMATICS
kaggle™

« l'ennemi est là, invisible,
insaisissable »

classification

nous devons classer la pneumonie du patient en 4 catégories :

1. **négative**
2. **typique**
3. **indéterminée**
4. **atypique**

SIIM SOCIETY FOR IMAGING INFORMATICS IN MEDICINE INNOVATING IMAGING INFORMATICS

kaggle

« l'ennemi est là, invisible,

insaisissable »

object detection

nous devons identifier pour chaque radiographie les zones d'interêts montrant une opacité responsable de la maladie

SIIM SOCIETY FOR IMAGING INFORMATICS IN MEDICINE INNOVATING IMAGING INFORMATICS

kaggle

« l'ennemi est là, invisible,
insaisissable »

submission

l'hôte de la compétition demande la réalisation d'un notebook de submission et sans accès à internet

la sortie doit contenir les prédictions de classification et d'object detection dans un fichier commun spécifique .csv

SIIM SOCIETY FOR IMAGING INFORMATICS IN MEDICINE INNOVATING IMAGING INFORMATICS

kaggle

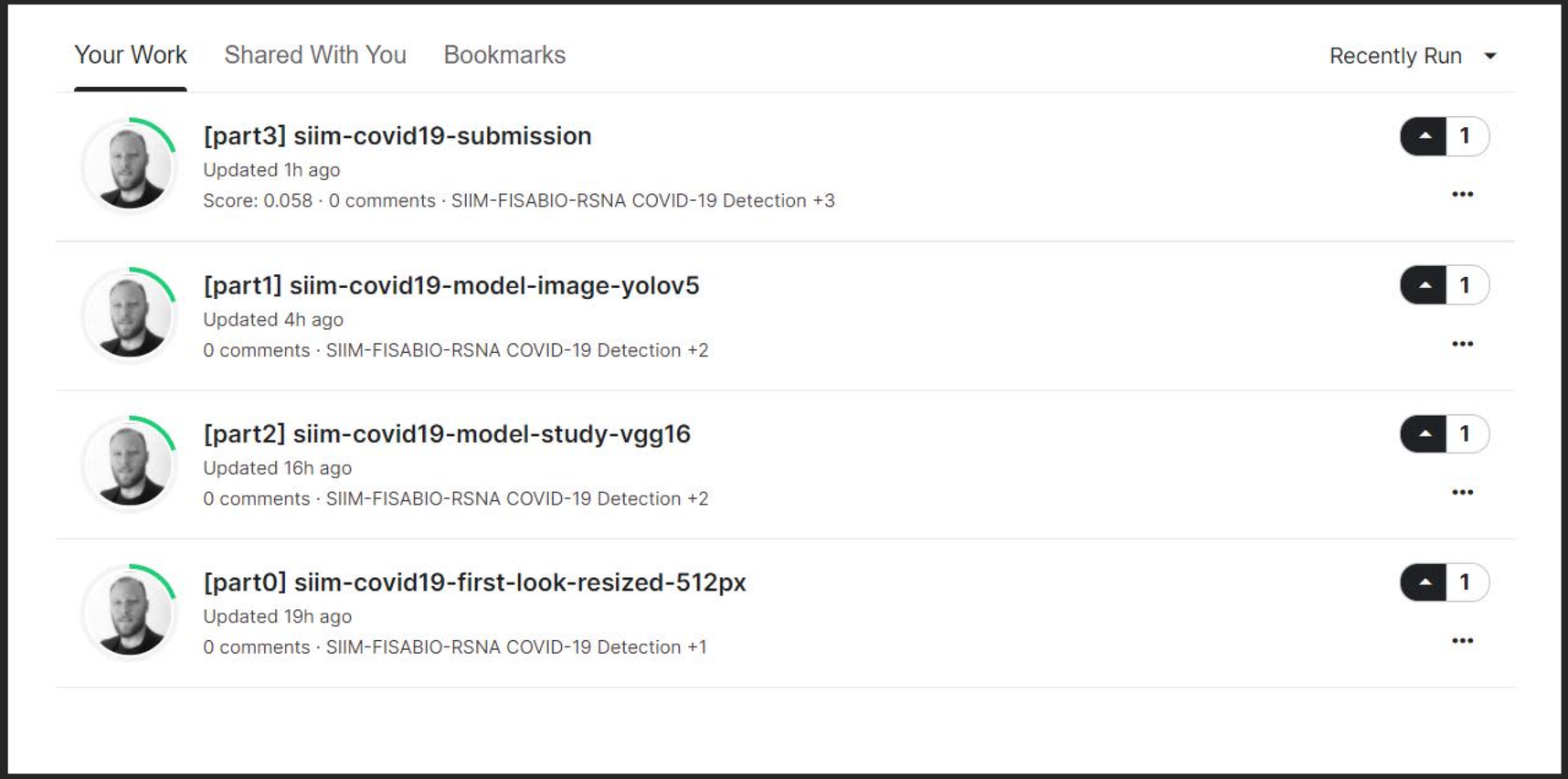
« l'ennemi est là, invisible,
insaisissable »

diverses contraintes au cours du projet (50h)

- **limitation du temps disponible gratuit en gpu sur kaggle**
- **problème en cours de l'hôte, doublon data et scoring**
- **problème d'accès au serveur notebook**

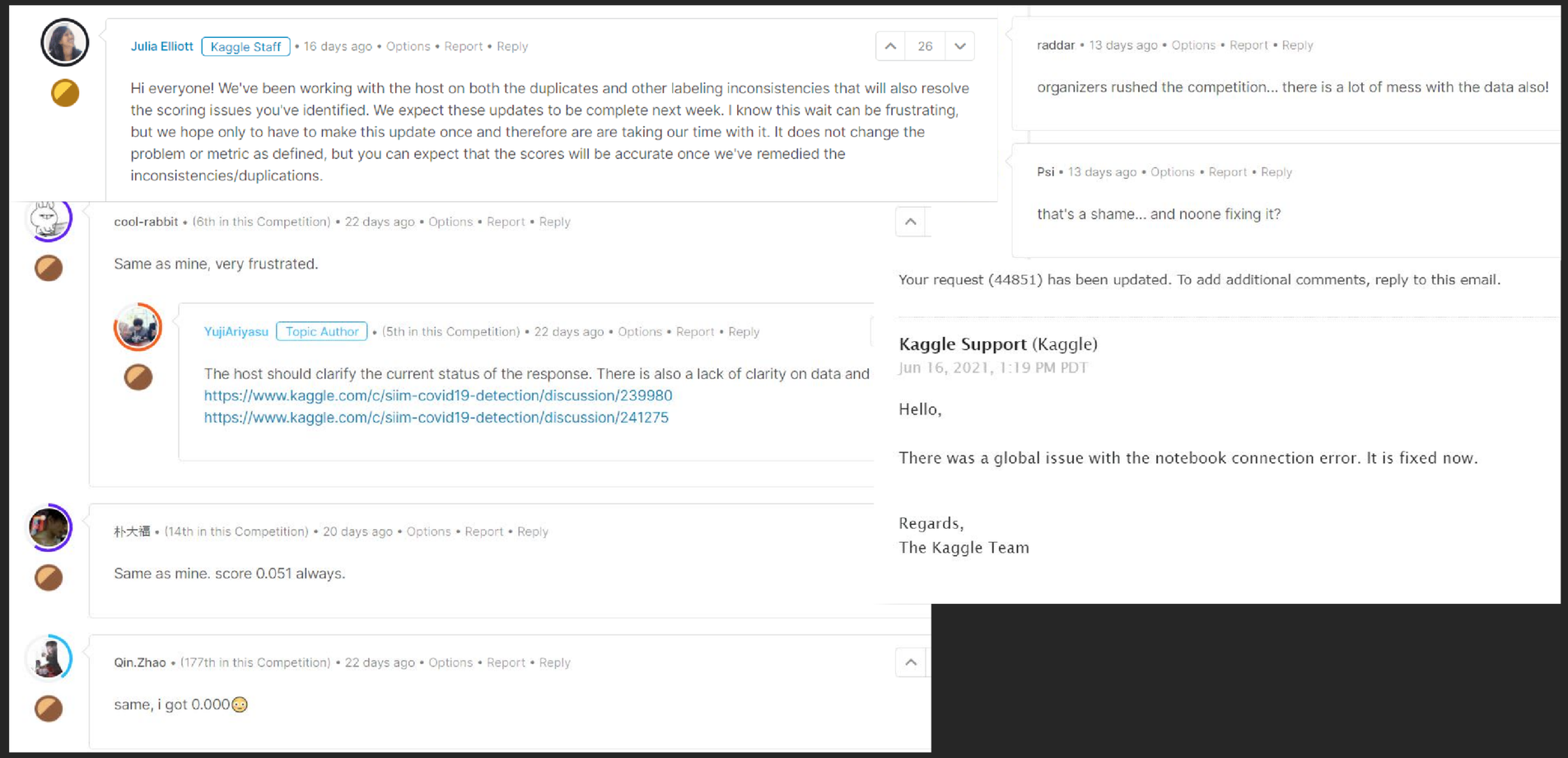
SIIM SOCIETY FOR IMAGING INFORMATICS IN MEDICINE INNOVATING IMAGING INFORMATICS
kaggle

partage à la communauté



SIIM SOCIETY FOR IMAGING INFORMATICS IN MEDICINE INNOVATING IMAGING INFORMATICS kaggle

discussions



Julia Elliott Kaggle Staff • 16 days ago • Options • Report • Reply 26

Hi everyone! We've been working with the host on both the duplicates and other labeling inconsistencies that will also resolve the scoring issues you've identified. We expect these updates to be complete next week. I know this wait can be frustrating, but we hope only to have to make this update once and therefore are are taking our time with it. It does not change the problem or metric as defined, but you can expect that the scores will be accurate once we've remedied the inconsistencies/duplications.

cool-rabbit • (6th in this Competition) • 22 days ago • Options • Report • Reply

Same as mine, very frustrated.

YujiAriyasu Topic Author • (5th in this Competition) • 22 days ago • Options • Report • Reply

The host should clarify the current status of the response. There is also a lack of clarity on data and
https://www.kaggle.com/c/siim-covid19-detection/discussion/239980
https://www.kaggle.com/c/siim-covid19-detection/discussion/241275

朴大福 • (14th in this Competition) • 20 days ago • Options • Report • Reply

Same as mine. score 0.051 always.

Qin.Zhao • (177th in this Competition) • 22 days ago • Options • Report • Reply

same, i got 0.000

raddar • 13 days ago • Options • Report • Reply

organizers rushed the competition... there is a lot of mess with the data also!

Psi • 13 days ago • Options • Report • Reply

that's a shame... and noone fixing it?

Your request (44851) has been updated. To add additional comments, reply to this email.

Kaggle Support (Kaggle)
Jun 16, 2021, 1:19 PM PDT

Hello,

There was a global issue with the notebook connection error. It is fixed now.

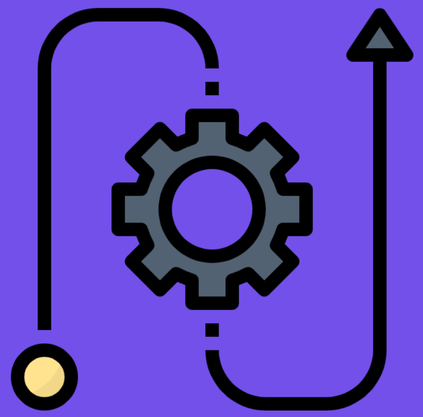
Regards,
The Kaggle Team

SIIM SOCIETY FOR IMAGING INFORMATICS IN MEDICINE INNOVATING IMAGING INFORMATICS

kaggle

la science des données

expérimentation



1. Collecter des données
2. Préparer et explorer les données
3. Modéliser
4. Comprendre

SIIM SOCIETY FOR IMAGING INFORMATICS IN MEDICINE INNOVATING IMAGING INFORMATICS

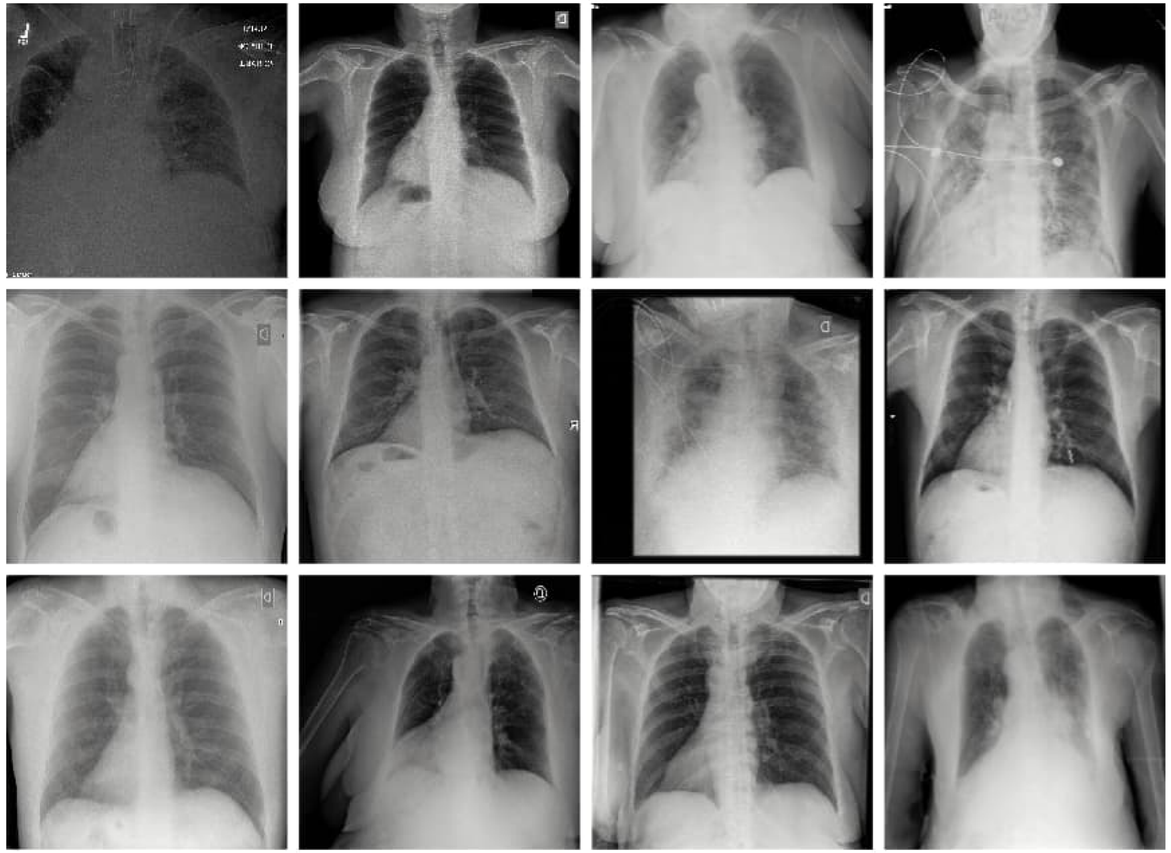
kaggle™

1

data explorer

119.68 GB

- test
- train
- sample_submission.csv
- train_image_level.csv
- train_study_level.csv



SIIM SOCIETY FOR IMAGING INFORMATICS IN MEDICINE INNOVATING IMAGING INFORMATICS

kaggle

2

P08_part0-siim-covid19-first-look-resized-512px.ipynb

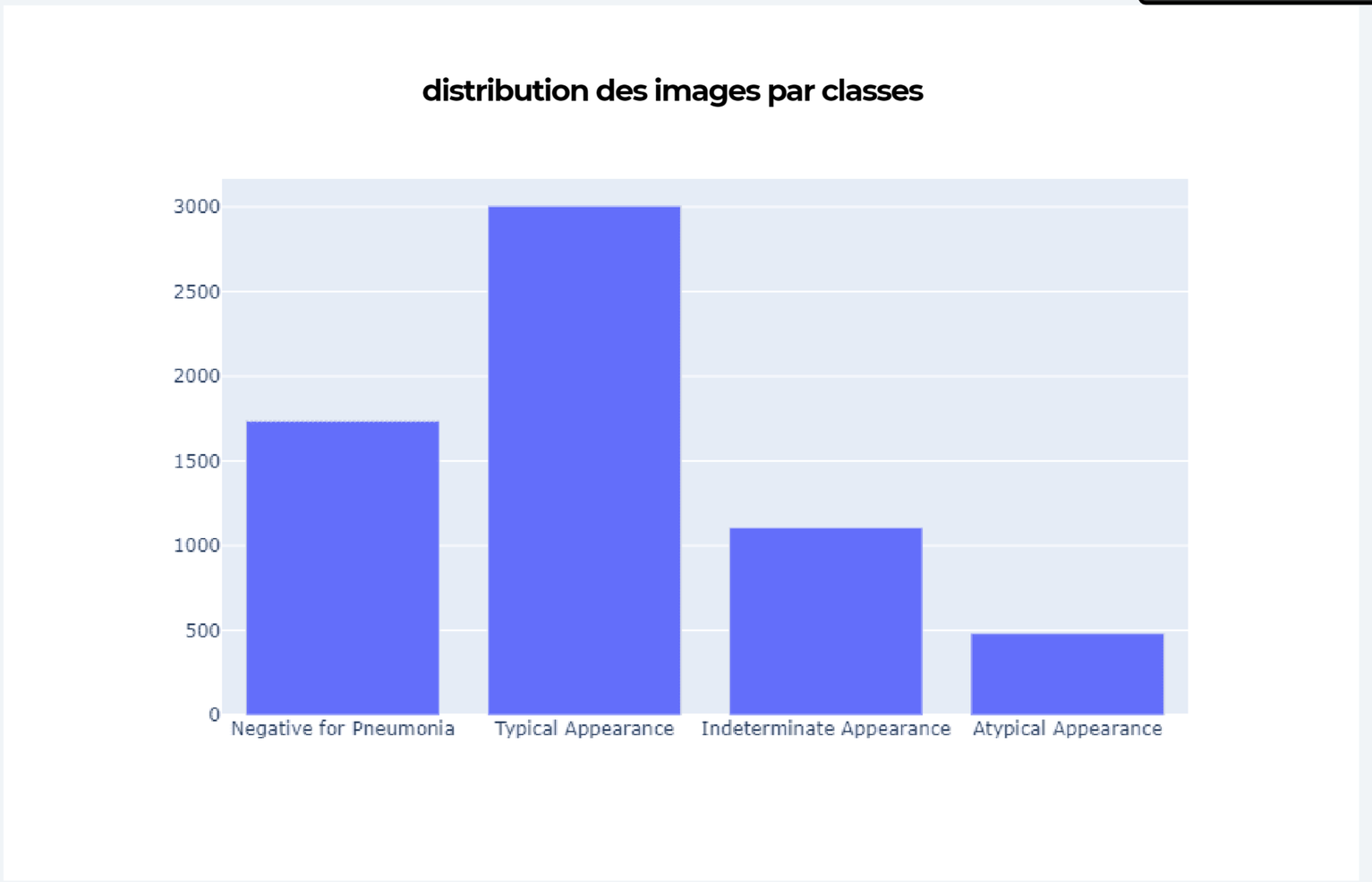
nombre d'images dans le trainset :	6334
nombre d'images dans le trainset (sans boxes) :	2040
nombre d'images dans le trainset (avec boxes) :	4294

SIIM SOCIETY FOR IMAGING INFORMATICS IN MEDICINE INNOVATING IMAGING INFORMATICS

kaggle

2

P08_part0-siim-covid19-first-look-resized-512px.ipynb

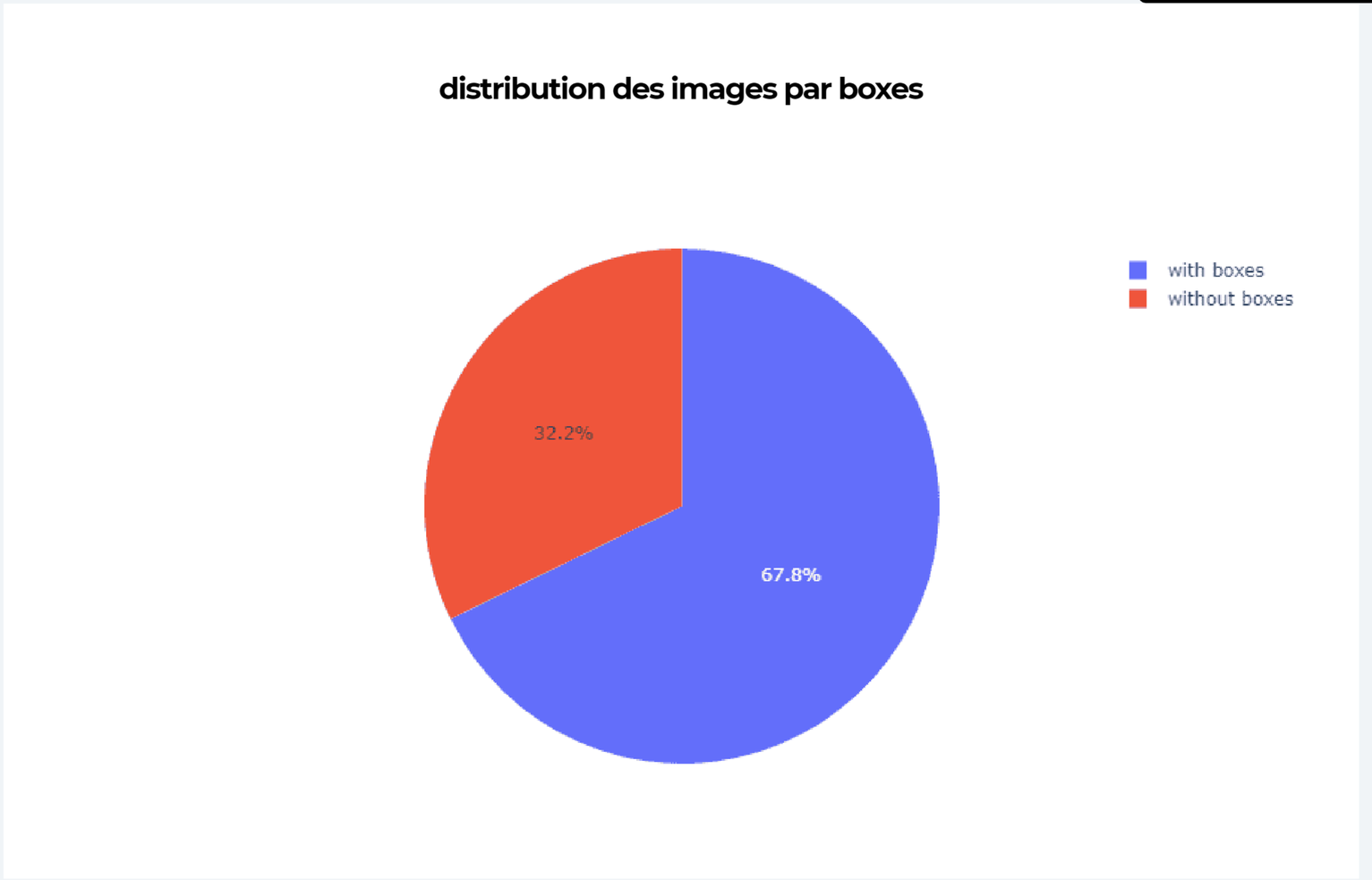


SIIM SOCIETY FOR IMAGING INFORMATICS IN MEDICINE INNOVATING IMAGING INFORMATICS

kaggle

2

P08_part0-siim-covid19-first-look-resized-512px.ipynb



SIIM SOCIETY FOR IMAGING INFORMATICS IN MEDICINE INNOVATING IMAGING INFORMATICS

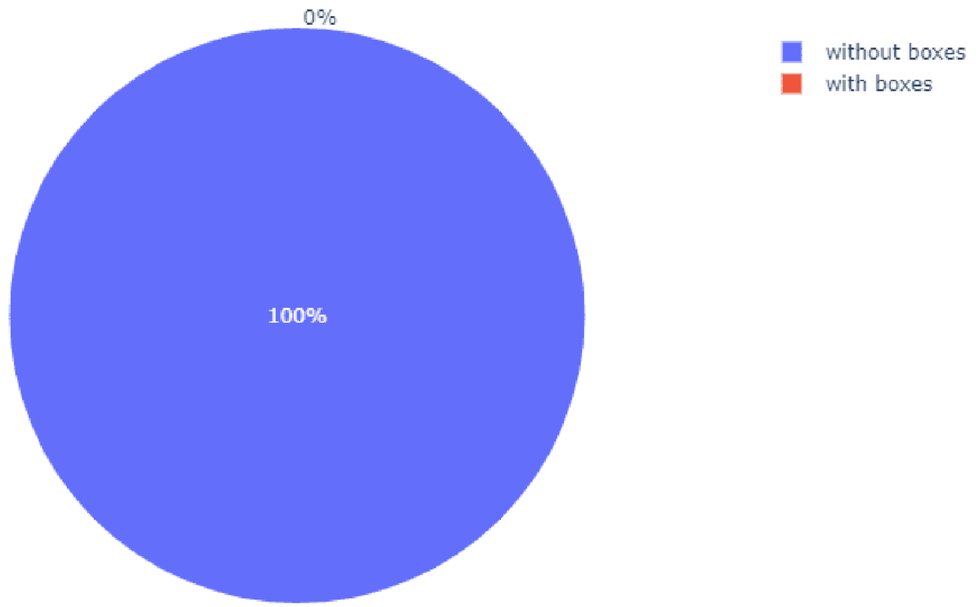
kaggle

2

P08_part0-siim-covid19-first-look-resized-512px.ipynb

distribution des images par boxes

patient négatif

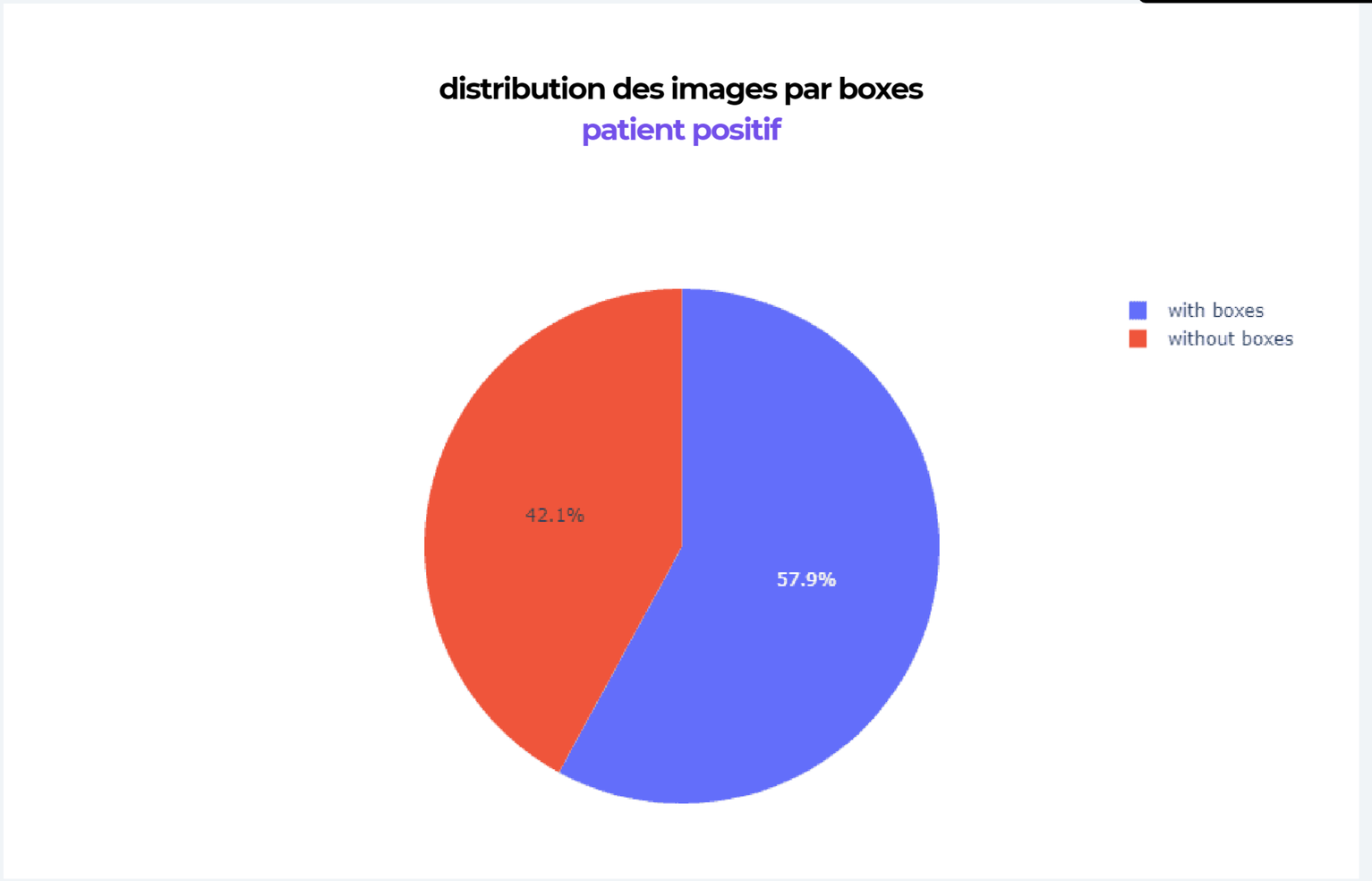


SIIM SOCIETY FOR IMAGING INFORMATICS IN MEDICINE INNOVATING IMAGING INFORMATICS

kaggle

2

P08_part0-siim-covid19-first-look-resized-512px.ipynb



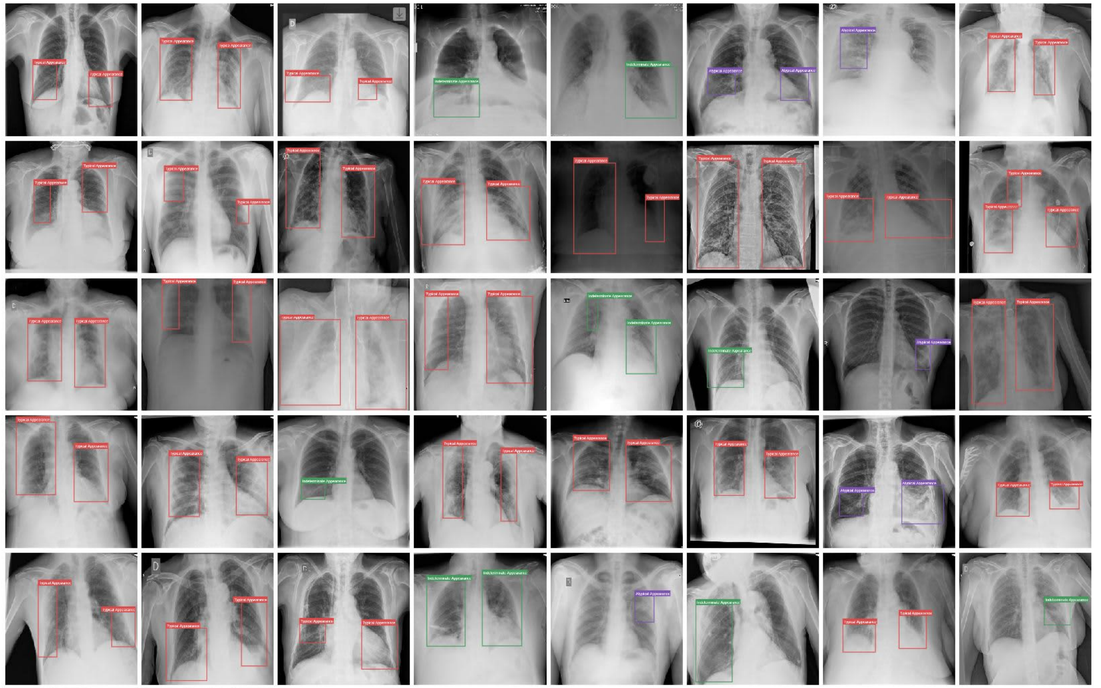
SIIM SOCIETY FOR IMAGING INFORMATICS IN MEDICINE INNOVATING IMAGING INFORMATICS

kaggle

2

P08_part0-siim-covid19-first-look-resized-512px.ipynb

visualisation des radiographies avec zones d'intérêts



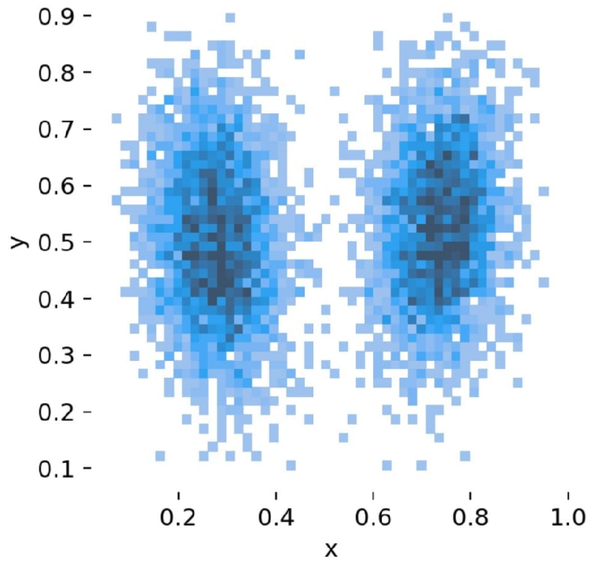
SIIM SOCIETY FOR IMAGING INFORMATICS IN MEDICINE INNOVATING IMAGING INFORMATICS

kaggle

2

P08_part0-siim-covid19-first-look-resized-512px.ipynb

analyse de la position des boxes



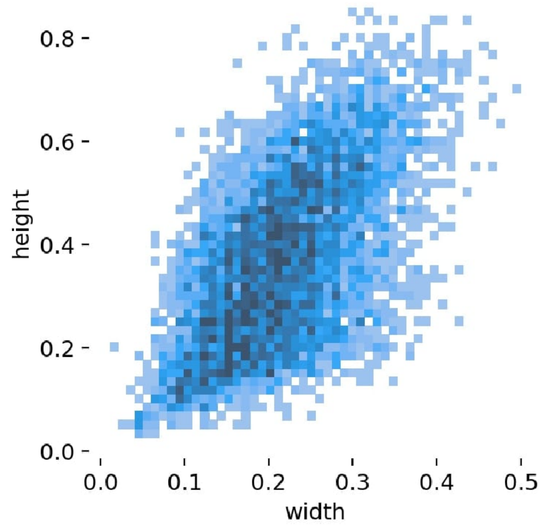
SIIM SOCIETY FOR IMAGING INFORMATICS IN MEDICINE INNOVATING IMAGING INFORMATICS

kaggle

2

P08_part0-siim-covid19-first-look-resized-512px.ipynb

analyse de la taille des boxes



SIIM SOCIETY FOR IMAGING INFORMATICS IN MEDICINE INNOVATING IMAGING INFORMATICS

kaggle

2

P08_part0-siim-covid19-first-look-resized-512px.ipynb

pré-traitement des données

- extraction des images à l'aide de pydicom
- inversion de couleur suivant profil monochrome
- redimensionnement des images avec cv2 (~2500^2 to 512^2px)

SIIM SOCIETY FOR IMAGING INFORMATICS IN MEDICINE INNOVATING IMAGING INFORMATICS

kaggle

3

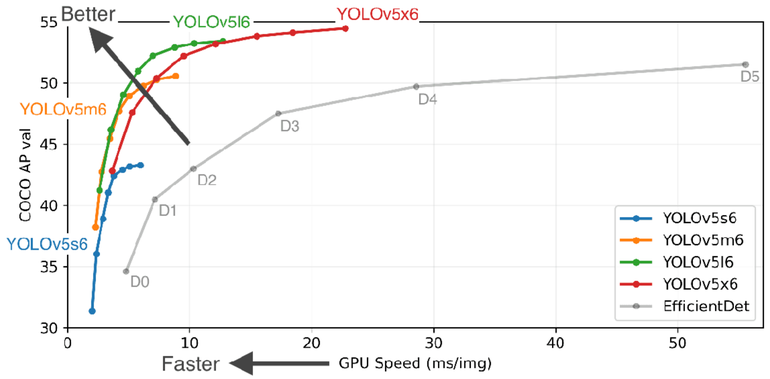
P08_part1-siim-covid19-model-image-yolov5.ipynb

object detection

yolov5s

@github.com/ultralytics/yolov5

YOLOv5 🚀 est une famille d'architectures et de modèles de détection d'objets pré-entraînés sur l'ensemble de données COCO, et représente la recherche open source d'Ultralytics sur les futures méthodes d'IA de vision, intégrant les leçons apprises et les meilleures pratiques développées au cours de milliers d'heures de recherche et développement



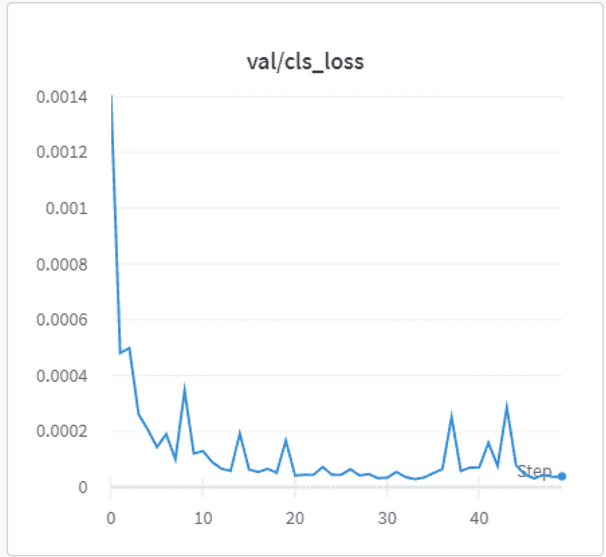
SIIM SOCIETY FOR IMAGING INFORMATICS IN MEDICINE INNOVATING IMAGING INFORMATICS

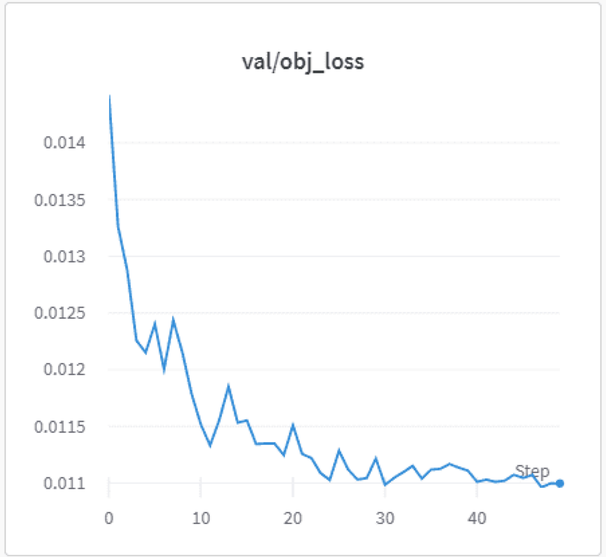
kaggle

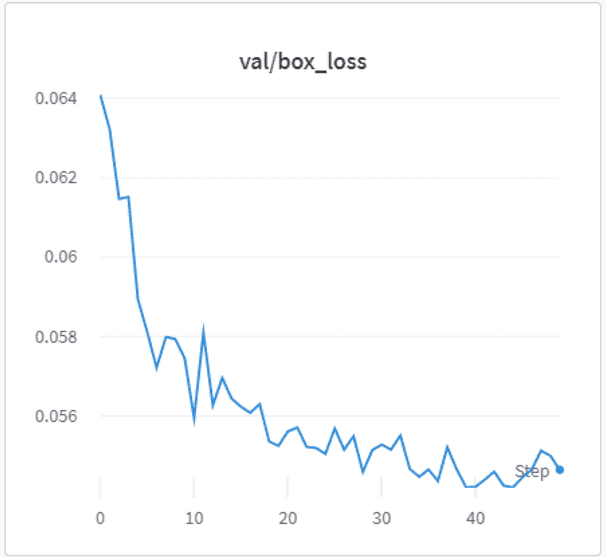
4

P08_part1-siim-covid19-model-image-yolov5.ipynb

métriques
yolov5s





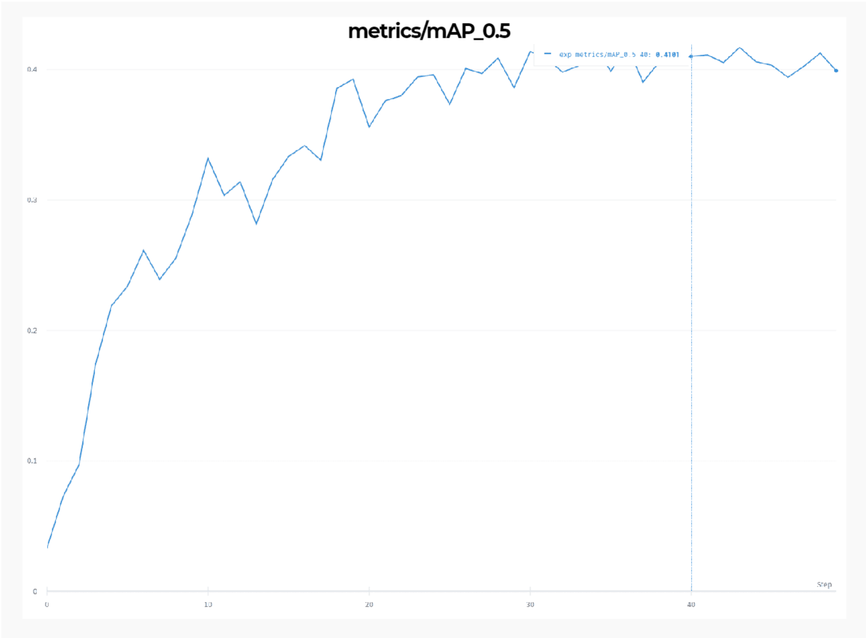


4

P08_part1-siim-covid19-model-image-yolov5.ipynb

métriques

yolov5s



SIIM SOCIETY FOR IMAGING INFORMATICS IN MEDICINE
INNOVATING IMAGING INFORMATICS

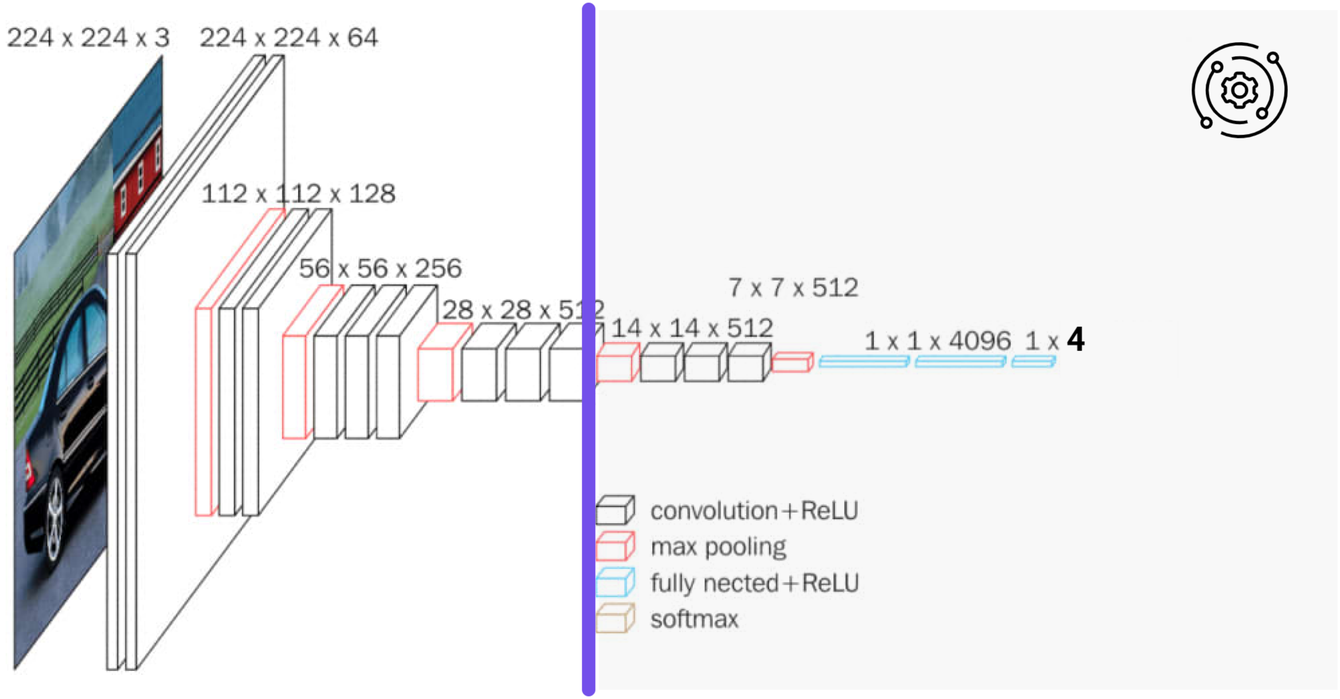
kaggle™

3

P08_part2-siim-covid19-model-study-vgg16.ipynb

classification

vgg16



SIIM SOCIETY FOR IMAGING INFORMATICS IN MEDICINE INNOVATING IMAGING INFORMATICS

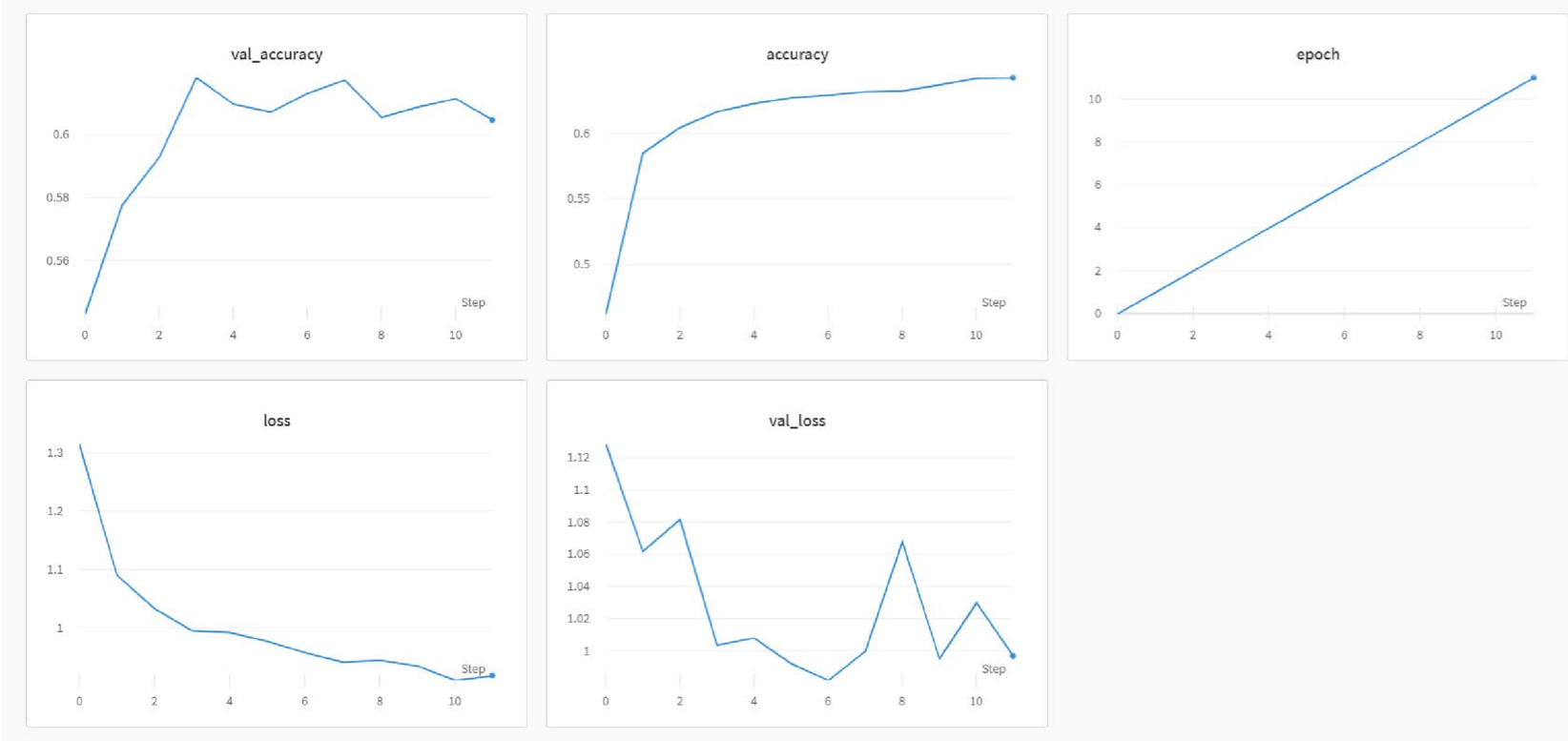
kaggle

4

P08_part2-siim-covid19-model-study-vgg16.ipynb

métriques

vgg16



SIIM SOCIETY FOR IMAGING INFORMATICS IN MEDICINE INNOVATING IMAGING INFORMATICS

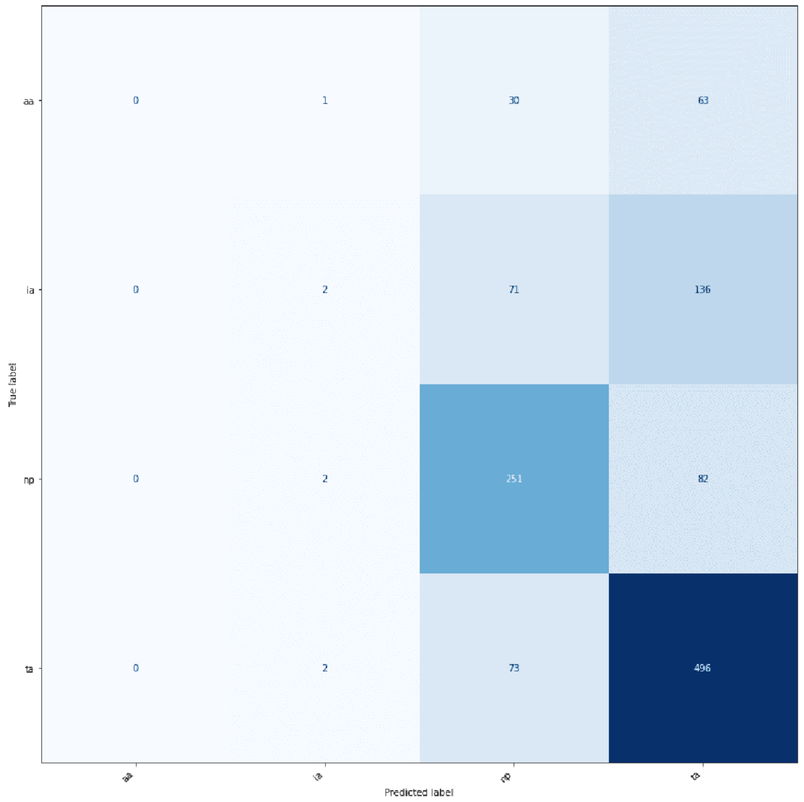
kaggle

4

P08_part2-siim-covid19-model-study-vgg16.ipynb

matrice de confusion

vgg16



SIIM SOCIETY FOR IMAGING INFORMATICS IN MEDICINE INNOVATING IMAGING INFORMATICS

kaggle™

3

P08_part3-siim-covid19-submission.ipynb

submission
leaderboard

529	Jessica Lee		0.089	7	2h
530	omelette du fromage	</> [part3] siim-covid1...	0.063	6	2h

Your Best Entry ↑

Your submission scored 0.063, which is an improvement of your previous score of 0.058. Great job! Tweet this!

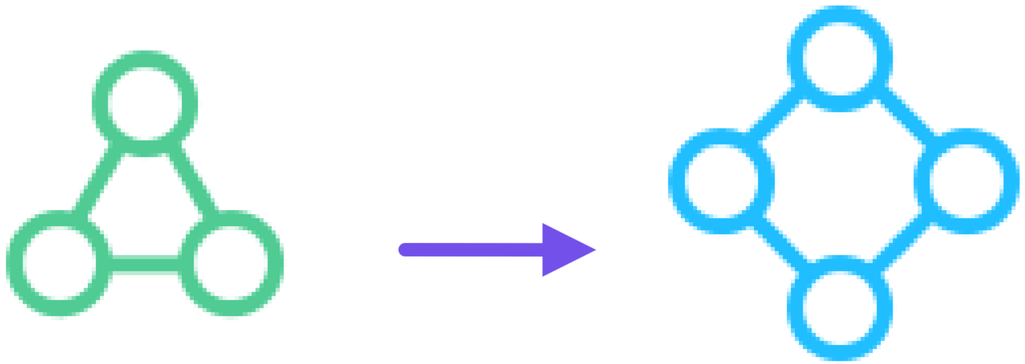
531	Naveen Rajan		0.053	2	1mo
532	Riad		0.053	1	22d
533	Carlos Picazo Montoya		0.053	1	21d
534	Howie Wu		0.053	4	19d
535	Meenakshi		0.053	2	16d
536	Bùi Nhật Trường		0.053	1	17d
537	Anton Chikin		0.050	2	12d

SIIM SOCIETY FOR IMAGING INFORMATICS IN MEDICINE INNOVATING IMAGING INFORMATICS kaggle

3

P08_part3-siim-covid19-submission.ipynb

submission
progression



novice **contributor**

SIIM SOCIETY FOR IMAGING INFORMATICS IN MEDICINE INNOVATING IMAGING INFORMATICS kaggle™

conclusion

- nous avons réussi à entrainer un modèle pour la prédiction de boxes sur les radiographies ainsi qu'un modèle de classification pour les études
- la performance observée est faible et le score de la submission par conséquence également
- l'amélioration est possible par l'utilisation de méthodes sota plus avancées dans la classification d'images, par des méthodes ensemblistes et l'entrainement de modèles plus complexes sur des serveurs puissants
- les notebooks kaggle sont disponibles publiquement pour la communauté.

SIIM SOCIETY FOR IMAGING INFORMATICS IN MEDICINE INNOVATING IMAGING INFORMATICS

kaggle

Projet 8

Thank you !