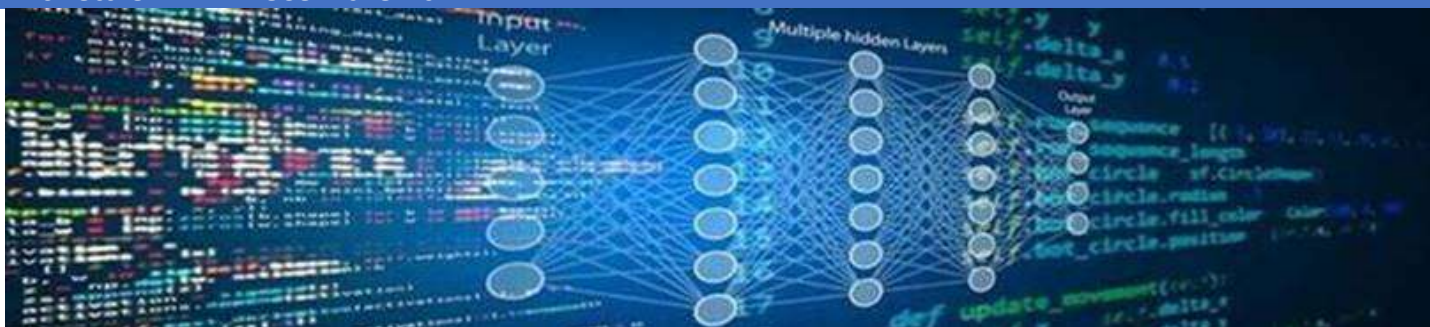




Infolettre SSP Hub

Actualités du réseau de data science
de l'INSEE et des SSM

InfoLettre n°7 – Décembre 2022



UNE MASTERCLASS DATASCIENTEST CONSACRÉE AU MLOPS DÉCOUVRIR LE SUJET DE LA MISE EN PRODUCTION DE MODÈLES DE MACHINE LEARNING

Afin de faciliter la mise en production de modèles de *machine learning*, l'**approche MLOps** consiste à automatiser et fiabiliser la préparation des données, l'estimation, la mise à disposition et l'archivage de modèles tout au long du processus de développement d'un projet. Inspiré de l'approche *DevOps*, où la production et l'amélioration des livrables d'un projet informatique est faite de manière continue, le [MLOps](#) est une méthode de travail visant à anticiper dès l'initialisation d'un projet une mise en production future. Cette approche permet de fluidifier le passage d'un prototype en production.



Le 9 décembre de 10h à 12h, l'organisme de formation [datascientest](#) proposera, en visio et en exclusivité, une **masterclass** sur le sujet. L'occasion idéale pour découvrir cette approche, les outils associés et les défis encore à relever. Cette session inaugure une série de *masterclass* exclusives qui s'annoncent passionnantes sur des sujets comme la *data engineering*, l'OCRisation...

Afin de bénéficier de cette formation exclusive, vous pouvez envoyer un **mail à ssphub-contact@insee.fr** qui vous transmettra en temps utile les informations pratiques (liens de connexion...)

👉 [L'événement sur le site web du réseau](#)

LA SAISON 1 DU PROGRAMME 10% SE TERMINE

VENEZ NOMBREUX A L'ÉVÉNEMENT DE CLÔTURE DE LA SAISON 1 !

Le programme interministériel 10% est issu des recommandations du **rapport Insee-DINUM** sur l'« [Evaluation des besoins de l'Etat en compétences et expertise en matière de donnée](#) ». L'objectif de ce programme est de permettre aux *data-scientists* de l'administration de dégager jusqu'à 10% du temps de travail pour mener des projets transversaux ou pour échanger entre pairs.

La première saison touche à sa fin et la **session de clôture est prévue ce 5 décembre de 14h à 16h** au lieu de la transformation publique :

- De 14h-15h, les *pitchs* des 4 projets portés lors de la saison 1 : [récupération facilitée de fonds de carte officiels pour la cartographie](#), [extraction d'informations depuis des PDF](#), [détection des entreprises dans le répertoire Sirene](#), [publications reproductibles au design system de l'Etat](#)
- De 15h à 16h, table ronde autour du sujet « Comment attirer et fidéliser les talents de la *data* au sein de l'Etat »

Vous pouvez consulter le programme détaillé et **vous inscrire sur [ce lien](#)**. Une version en visio sera également proposée.

Plusieurs membres du réseau ont participé activement à certains projets. N'hésitez pas à suivre leurs travaux et à rejoindre les prochaines saisons de ce programme !

- » [Le lien d'inscription](#)
- » [Le site web du programme 10%](#)

REJOIGNEZ LES GROUPES DE TRAVAIL INTERNATIONAUX !

L'UNECE RECRUTE POUR SES NOUVEAUX GROUPES DE TRAVAIL

L'[UNECE Machine Learning Group](#) est un groupe de travail de l'UNECE (commission régionale de l'ONU pour l'Europe) permettant d'alimenter la réflexion, au niveau international, sur l'usage des méthodes de *machine learning* pour la production statistique.

Un nouveau groupe prend la suite en 2023, il s'agit du **Applying Data Science and Modern Methods Group**.

Trois axes de travail sont proposés dans ce cadre :

- [Modelling](#)
- [Data Editing](#)
- [Responsible AI](#)



Si vous êtes intéressés par participer à ces groupes ou suivre les actualités d'un des thèmes en question, vous pouvez **envoyer un message à ssphub-contact@insee.fr** qui recense les participants côté statistique publique française.

➤ [Site web du groupe de travail UNECE Machine Learning](#)

NOS DATA-SCIENTISTS FORMENT A L'INTERNATIONAL !

UNE FORMATION AU WEBSCRAPING PROPOSÉE AUX STATISTICIENS MAROCAINS ET TUNISIENS

Antoine Palazzolo, *data-scientist* au SSP-Lab (INSEE), proposera au début du mois de décembre une formation de deux jours au *webscraping* aux INS du Maroc et de la Tunisie dans le cadre de la coopération internationale.

L'occasion d'évoquer l'intérêt pour la production statistique du *webscraping* mais aussi les limites de cette technique, le contexte juridique, les bonnes et mauvaises pratiques et les cas d'usages dans la statistique publique française et internationale. Plusieurs mises en application, inspirées du [cours de Python de l'ENSAE](#) et s'appuyant sur le [SSP-Cloud](#), sont prévues.

Les supports de formation sont d'ores et déjà [publics](#) ! Que vous connaissiez ou non le sujet, vous apprendrez certainement beaucoup grâce à la richesse du contenu proposé !

➤ [Les supports de formation](#)

➤ [Le contenu sur Github](#)

Formation au web scraping

Rabat - Coopération internationale

Antoine Palazzolo, INSEE

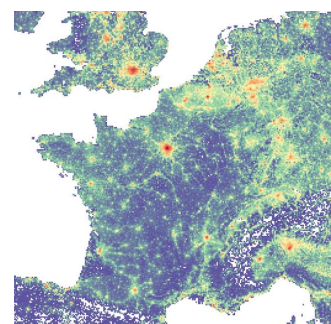
December 1, 2022

PRESENTATION DE GRIDVIZ PAR JULIEN GAFFURI

UN OUTIL POUR LA DATAVIZ CARTOGRAPHIQUE DE DONNEES CARROYEES

Après la [présentation passionnante d'observable](#) par Nicolas Lambert (CNRS), [disponible en replay sur notre site web](#), un nouveau séminaire sur un sujet de cartographie réactive est déjà prévu !

Julien Gaffuri (Eurostat) viendra nous présenter la librairie *open-source* [gridviz](#). Réservez le créneau dans votre agenda du **20 janvier 2023 de 11h à 12h30** pour découvrir cette librairie qui ouvre de nouvelles perspectives pour la mise à disposition de données carroyées.



Cette librairie permet de visualiser de manière efficace des données carroyées dans un navigateur web. Développé par Eurostat, le site web du projet regorge d'exemples de cartes issues des données carroyées Filosofi ou de la statistique publique européenne.

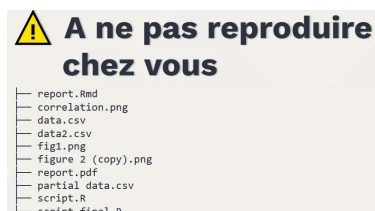
APRÈS GIT, LES BONNES PRATIQUES !

30 PERSONNES DE LA DESE (INSEE) FORMÉES AUX BONNES PRATIQUES

Après le test d'une [nouvelle formule de formation à Git](#), les statisticiens de la Direction des Etudes et des Synthèses Economiques (DESE) de l'Insee ont servi de testeurs à une [formation aux bonnes pratiques avec R](#) qui devrait prochainement figurer au plan national de formation de l'Insee. Cette formation vient compléter sa [version équivalente en Python](#) déjà existante dans le parcours de formation de l'ENSAE.

L'occasion de présenter les bons gestes à avoir dans le cadre d'un projet statistique et les outils qui permettent d'avoir des projets qui soient reproductibles, transparents et partageables à plusieurs types de publics.

Si vous êtes intéressés pour vos équipes par cette formation, ou par celle à *Git* faite le mois précédent, vous pouvez contacter romain.avouac@insee.fr et lino.galiana@insee.fr !



» [Les slides présentées](#)

» [Les slides de la session Git](#)

RETOUR SUR LE MOIS DE NOVEMBRE

UN MOIS DE NOVEMBRE RICHE EN ÉVÉNEMENTS !

Présentation d'observable pour le réseau : La [présentation très pédagogique de Nicolas Lambert \(CNRS\)](#) pour le réseau est disponible en *replay* sur notre site web. Le support présenté est disponible directement sur observable.

Participation à un hackathon de l'ONU : l'équipe *datadive*, rassemblant des *data scientists* du réseau, a participé pendant trois jours au hackathon « *big data* » organisé par l'ONU. Le résultat ? Une [application](#) qui permet de visualiser le trafic dans des zones critiques maritimes pour le commerce international.

Présentation du réseau aux nouveaux administrateurs de l'Insee : Lundi 28 novembre, Lino Galiana a présenté le réseau aux administrateurs Insee actuellement en formation à l'ENSAE et qui sont amenés à prendre un poste à la prochaine campagne. L'occasion parfaite pour faire connaître le réseau, ses objectifs et ses actions mais aussi discuter de la manière dont les *data-scientists* de différentes organisations peuvent échanger ensemble sur les innovations récentes de la *data science*.

Rédacteur : Lino Galiana

Si vous avez des actualités susceptibles d'apparaître dans [l'infolettre du SSP Hub](#), n'hésitez pas à en faire part pour l'édition suivante sur [Github](#).

N'hésitez pas à faire circuler cette infolettre auprès de vos contacts !

Vous êtes intéressés pour recevoir cette infolettre mais n'êtes pas dans la liste de diffusion ? Vous pouvez écrire à : ssphub-contact@insee.fr