

Plus fort que la machine ?

1er Data Challenge de l'ISDM

L'institut de science des données¹ (ISDM) de Montpellier et la start-up Trustii² s'associent pour proposer le 1^{er} data challenge de l'ISDM.

Dans le cadre de ce challenge, trois problèmes de machine learning vous sont proposés par des entreprises à partir de leurs données. Serez-vous capable de proposer un modèle prédictif plus performant qu'un algorithme automatique ? Serez-vous capable de proposer le modèle le plus performant ?

Présentation synthétique des jeux de données



<https://www.sud.banquepopulaire.fr/>

anonymes composé d'ex-clients et de clients. Saurez-vous apprendre suffisamment de ces ex-clients pour prédire le comportement des clients ? A vous de relever ce défi !

Banque Populaire du Sud

Serez-vous capable d'identifier, auprès d'une clientèle de jeunes adultes, les clients sur le point de quitter leur établissement bancaire ? La Banque Populaire du Sud met à votre disposition un jeu de données

Bfore.ai

Serez-vous capable de détecter les adresses frauduleuses pour préserver les systèmes informatiques d'attaques malicieuses ? Bfore.ai vous propose l'historique de ses jeux de données de cybersécurité. Il vous faut anticiper les attaques pour ne pas les subir !



<https://www.bfore.ai/>



Montpellier Méditerranée Métropole (<https://data.montpellier3m.fr/>)

Etalab (<https://www.etalab.gouv.fr/>)

INSEE (<https://www.insee.fr/>)

Université de Montpellier – ISDM (<https://isdm.umontpellier.fr/>)

Serez-vous capable de prédire les prix de l'immobilier dans l'unité urbaine de Montpellier ? L'ISDM vous propose un jeu de données contenant les transactions immobilières sur l'année 2018 à partir des données ouvertes d'Etalab et des données sociaux-économiques ouvertes de l'INSEE. Combiné aux données de la métropole de Montpellier et de l'INSEE, est-il possible de construire un modèle prédictif performant ?

¹ <https://isdm.umontpellier.fr/>

² <https://www.trustii.io/>

Comment participer ?

Pour participer au challenge, il suffit de s'inscrire en ligne en renseignant votre adresse e-mail institutionnelle prenom.nom@etu.umontpellier.fr à l'adresse suivante (ne pas oublier de valider en cliquant sur "Envoyer") : <https://questionnaire.umontpellier.fr/index.php/188357?lang=fr>.

Vous pouvez également dès à présent vous inscrire comme *data scientist* sur la plateforme Trustii (<https://www.trustii.io>), en utilisant la même adresse e-mail. Les descriptions détaillées des problèmes seront visibles sur la plateforme, ainsi que les jeux de données associés à partir du 29/01/2021. Il est nécessaire d'être inscrit comme *data scientist* sur la plateforme Trustii pour accéder aux jeux de données.

Quels types de données ?

Les méthodes visées sont de type "supervisées". Les données d'apprentissage sont présentées sous la forme de fichiers .csv intégrant des attributs de description et la classe d'appartenance, uniquement au format numérique (pas de données textuelles).

Quel langage ?

Les modèles devront être proposés en Python. La plateforme Trustii supporte les modules Scikit-Learn, Auto-Sklearn, XGBoost et PyTorch.

Comment soumettre une solution ?

Un tutoriel en ligne explique comment soumettre une proposition sur la plateforme Trustii : <https://colab.research.google.com/gist/Trustii-team/ddf57a195205fc4ce81d54af5d7cc570/making-your-first-submission-on-trustii.ipynb>

Récompenses

Tout candidat ayant déposé correctement un modèle sera récompensé d'un lot. Tout candidat ayant proposé un modèle plus performant que l'algorithme de la plateforme Trustii sera récompensé d'un lot plus important.

Pour chaque problème, les trois meilleures propositions se verront proposer des récompenses supplémentaires.

Calendrier du challenge

Date limite d'inscription	28/01/2021 - 23h59
Publication des jeux de données	29/01/2021
Date limite de proposition de modèles	19/02/2021 - 23h59
Déclaration des résultats et remise des prix	24/02/2021

Informations et contact : <https://isdumontpellier.fr/> - isdm-challenges@umontpellier.fr