

[← Mon parcours](#) PROJET VALIDÉ

Anticipez les besoins en consommation de bâtiments

[Mission](#) [Cours](#) [Ressources](#) [Évaluation](#) 100 heures

Mis à jour le jeudi 9 mars 2023

Vous travaillez pour la **ville de Seattle**. Pour atteindre son objectif de ville neutre en émissions de carbone en 2050, votre équipe s'intéresse de près à la consommation et aux émissions des **bâtiments non destinés à l'habitation**.



Seattle

Des relevés minutieux ont été effectués par les agents de la ville en 2016. Voici [les données](#) et [leur source](#). Cependant, ces relevés sont coûteux à obtenir, et à partir de ceux déjà réalisés, **vous voulez tenter de prédire les émissions de CO2 et la consommation totale d'énergie** de bâtiments **non destinés à l'habitation** pour lesquels elles n'ont pas encore été mesurées.

Votre prédiction se basera sur les données structurelles des bâtiments (taille et usage des bâtiments, date de construction, situation géographique, ...)

Vous cherchez également à **évaluer l'intérêt de l'"ENERGY STAR Score" pour la prédiction d'émissions**, qui est fastidieux à calculer avec l'approche utilisée actuellement par votre équipe. Vous l'intégrerez dans la modélisation et jugerez de son intérêt.

Vous sortez tout juste d'une réunion de brief avec votre équipe. Voici un récapitulatif de votre mission :

1. Réaliser une courte analyse exploratoire.
2. Tester différents modèles de prédiction afin de répondre au mieux à la problématique.

Avant de quitter la salle de brief, **Douglas**, le **project lead**, vous donne quelques pistes et erreurs à éviter :



Douglas : L'objectif est de te passer des relevés de consommation annuels futurs (attention à la fuite de données). Nous ferons de toute façon pour tout nouveau bâtiment un premier relevé de référence la première année, donc rien ne t'interdit d'en déduire des variables structurelles aux bâtiments, par exemple la nature et proportions des sources d'énergie utilisées..

Fais bien attention au traitement des différentes variables, à la fois pour trouver de nouvelles informations (peut-on déduire des choses intéressantes d'une simple adresse ?) et optimiser les performances en appliquant des transformations simples aux variables (normalisation, passage au log, etc.).

Mets en place une évaluation rigoureuse des performances de la régression, et optimise les hyperparamètres et le choix d'algorithmes de ML à l'aide d'une validation croisée.



Livrables

- Un **notebook** de l'analyse exploratoire mis au propre et annoté.
- Un **notebook** pour chaque prédiction (émissions de CO2 et consommation totale d'énergie) des différents tests de modèles mis au propre, dans lequel vous identifierez clairement le modèle final choisi.
- Un support de **présentation** pour la soutenance (entre 15 et 25 slides).

Pour faciliter votre passage devant le jury, déposez sur la plateforme, dans un dossier zip nommé "**Titre_du_projet_nom_prénom**", votre livrable nommé comme suit :

Nom_Prénom_n° du livrable_nom du livrable_date de démarrage du projet. Cela donnera :

- *Nom_Prénom_1_notebook_exploratoire_mmaaaa*
- *Nom_Prénom_2_notebook_prediction_mmaaaa*
- *Nom_Prénom_3_notebook_prediction_mmaaaa*
- *Nom_Prénom_4_presentation_mmaaaa*

Par exemple, votre premier livrable peut être nommé comme suit :

Dupont_Jean_1_notebook_exploratoire_012022.

Soutenance

La soutenance se déroulera en visioconférence et durera 30 minutes. Le mentor évaluateur ne jouera aucun rôle particulier. Elle s'appuiera sur votre dernier livrable (votre présentation).

- **Présentation (20 minutes)**
 - Rappel de la problématique et présentation du jeu de données (5 mn).
 - Présentation du feature engineering (5 minutes).

- Explication de l'approche de modélisation et présentation des résultats (10 minutes).
- **Discussion** (5 minutes)
 - L'évaluateur vous challengera sur vos choix.
- **Débriefing** (5 minutes)
 - À la fin de la soutenance, vous pourrez débriefer ensemble.

Votre présentation devrait durer 20 minutes (+/- 5 minutes). Puisque le respect des durées des présentations est important en milieu professionnel, les présentations en dessous de 15 minutes ou au-dessus de 25 minutes peuvent être refusées.

Compétences évaluées



Mettre en place le modèle d'apprentissage supervisé adapté au problème métier



Adapter les hyperparamètres d'un algorithme d'apprentissage supervisé afin de l'améliorer



Transformer les variables pertinentes d'un modèle d'apprentissage supervisé



Évaluer les performances d'un modèle d'apprentissage supervisé

OPENCLASSROOMS



OPPORTUNITÉS



AIDE



POUR LES ENTREPRISES



EN PLUS



Français



