

SCRIPT ORAL - PRESENTATION ENERGISIGHT

Presente par : DORVI Yao Theodore

Slide 1 - Cover

Bonjour a toutes et a tous. Je suis DORVI Yao Theodore.

Je vous presente notre POC EnergiSight : un modele de prediction de la consommation energetique et des emissions CO2 pour les batiments non residentiels.

Slide 2 - Probleme et Objectif

Le probleme est simple : a Lome, les donnees energie/CO2 sont limitees et les releves terrestres sont couteux.

Notre objectif est de produire des estimations fiables pour aider la decision publique et privée.

Slide 3 - Perimetre Final

Pour ce POC, nous utilisons Seattle 2016 comme laboratoire de donnees.

Nous avons filtre strictement les batiments NonResidential, puis applique des criteres qualitatifs pour obtenir 1432 observations exploitables.

Slide 4 - Qualite des Donnees

Ici, nous montrons les variables avec valeurs manquantes.

Cette etape nous permet d'anticiper les biais et de definir une strategie d'imputation propre dans le pipeline.

Slide 5 - Skewness

Les distributions de plusieurs variables sont asymetriques.

Nous appliquons donc log1p sur les variables skewed pour stabiliser les modeles et reduire l'impact des extremes.

Slide 6 - Distributions des Cibles

On observe que CO2 et energie sont tres skewed en brut.

Apres transformation, les distributions sont plus maitrisees, ce qui ameliore la robustesse de la modelisation.

Slide 7 - Geo-cluster

Nous avons cree un geo-cluster a partir de la latitude et de la longitude avec KMeans.

Cette variable capte l'effet spatial, utile pour differencier les comportements energetiques selon les zones.

Slide 8 - Pipeline de Modelisation

Methode : split train/test 80/20, validation croisee 5 folds, puis comparaison de trois methodes : Ridge, RandomForest et GradientBoosting.

Le modele final est choisi objectivement sur le meilleur score CV RMSE.

Slide 9 - Modeles CO2

Pour la prediction CO2, le meilleur compromis est Ridge.

Il offre une performance solide avec une bonne stabilité.

Slide 10 - Modeles Energie

Pour l'energie, GradientBoosting est le plus performant.

Il capture mieux les relations non lineaires presentes dans les donnees.

Slide 11 - Feature Importances

Cette slide montre les variables qui influencent le plus les predictions.

Les surfaces et le type d'usage ressortent fortement, ainsi que certaines variables structurelles cles.

Cela nous donne des priorites claires pour la collecte locale.

Slide 12 - Resultats Train vs Test

Ici, nous comparons les performances train et test pour CO2 et energie.

Les ecart restent maitrisés, ce qui est un bon signal de generalisation pour un POC.

Slide 13 - Audit fuite / overfitting

Nous avons verifie l'absence de colonnes interdites et de doublons entre train et test.

Conclusion : aucun signal fort de fuite de donnees, et risque d'overfitting limite.

Slide 14 - Impact ENERGYSTAR Score

Le score ENERGY STAR ameliore les performances sur les deux cibles.

C'est donc une variable utile quand elle est disponible.

Slide 15 - Recommandation Finale

Notre recommandation est de conserver ce pipeline simple et robuste.

Ensuite, lancer un pilote a Lome sur 3 a 6 mois, recalibrer avec des donnees locales, puis monitorer trimestriellement.

Slide 16 - Conclusion

En conclusion, le POC valide une approche fiable, interpretable et actionnable.

La prochaine etape est le deploiement pilote local pour passer de la preuve de concept a l'impact terrain.