

Machine Learning : Prediction de Satisfaction des habitants en lien avec l'urbanisme vert



SOMMAIRE

1

CONTEXTE DU PROJET

2

PROBLÉMATIQUE

3

DONNÉES MOBILISÉES

- Présentation des données
- Echantillon des données
- Nettoyage des données

4

CREATION DE MODÈLE
STATISTIQUE

5

RÉSULTATS OBTENUS

6

VISUALISATION

7

TRANSPOSITION À MONTÉLIMAR

- ET SI ON APPLIQUAIT CE MODELE A MONTELIMAR ?
- Comment ce modèle peut-il être utile aux collectivités ?

8

CONCLUSION

SUJET : PREDICTION DE SATISFACTION DES HABITANTS EN LIEN AVEC L'URBANISME VERT

Contexte

- L'environnement urbain joue un rôle majeur sur la qualité de vie
- Les espaces verts sont associés à un meilleur bien-être mental, social et physique

Objectif ?

- Explorer scientifiquement ce lien à partir de données réelles

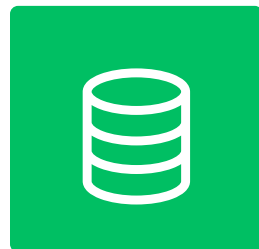




PROBLEMATIQUE :

Dans quelle mesure la présence d'espaces verts dans l'environnement immédiat influence-t-elle la satisfaction de vie des habitants ?

PRESENTATION DES DONNÉES



DONNÉES SOCIALES

- **SOURCE : SWISS HOUSEHOLD PANEL (SHP)**
- **17 000 INDIVIDUS INTERROGÉS**
- **SATISFACTION DE VIE (SCORE DE 0 À 10)**
- **DONNÉES SOCIO-DÉMOGRAPHIQUES**
- **LOCALISATION DU MÉNAGE**



DONNÉES ENVIRONNEMENTALES

- **SOURCE : DOSSIER URBAN ENVIRONMENT CH**
- **CALCULS RÉALISÉS SUR 210M AUTOUR DU DOMICILE**
- **PROPORTION D'ARBRES, PELOUSES, JARDINS...**
- **DISTANCE AU CENTRE-VILLE**
- **DENSITÉ DE POPULATION**

ECHANTILLON DES DONNÉES

idhous21	lifesat	pop_density	distance_to_center_1	trees_park_210
51	8	26266.246893	4747.95690	0.189109
51	-3	26266.246893	4747.95690	0.189109
131	7	1795.469906	1649.30439	0.012560
12 543	9	812.480663	2172.48396	0.194310
15 217	10	5720.344662	5943.020154	0.215640

Satisfaction de la vie en général



De manière générale, dans quelle mesure êtes-vous satisfait·e de votre vie, si 0 signifie "pas du tout satisfait·e" et 10 "tout à fait satisfait·e" ?



-8 autre erreur

-7 erreur de filtre

-3 inapplicable

-2 pas de réponse

-1 ne sait pas

0 pas du tout satisfait.e

1

2

3

4

5

6

7

8

9

10 tout à fait satisfait.e

```
# Supprimer les deux premières colonnes
data = data.iloc[:, 2:]
```

- Suppressions Valeurs non-importantes

- Suppressions des lignes contenant NA

```
# Supprimer les lignes contenant des NaN
data = data.dropna()
```

NETTOYAGE DES DONNÉES

```
X_train, X_test, y_train, y_test = train_test_split(X,y,test_size=0.2,random_state=42)
```

- Train test split

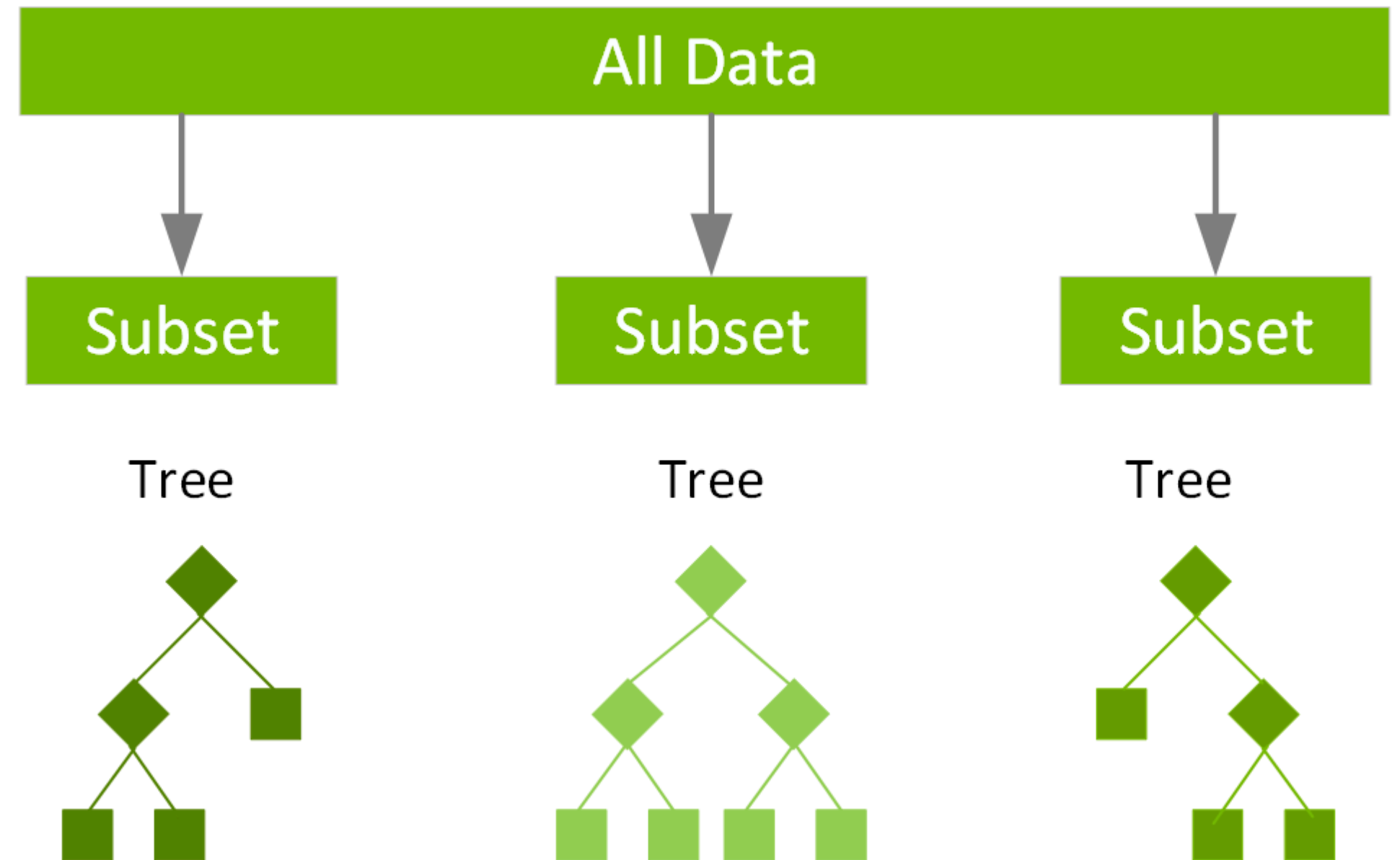
- Conversion des données non-numerique en numerique

```
# Identifier les colonnes de type string
string_columns = data.select_dtypes(include=['object']).columns

# Appliquer LabelEncoder pour transformer les strings en entiers
label_encoders = {}
for column in string_columns:
    le = LabelEncoder()
    data[column] = le.fit_transform(data[column])
    label_encoders[column] = le # Sauvegarder l'encodeur pour cette colonne
```

Création d'un modèle prédictif basé sur les caractéristiques de l'environnement

Modèle utilisé : XGBoost
(algorithme d'arbre de décision)



Résultats de prédiction

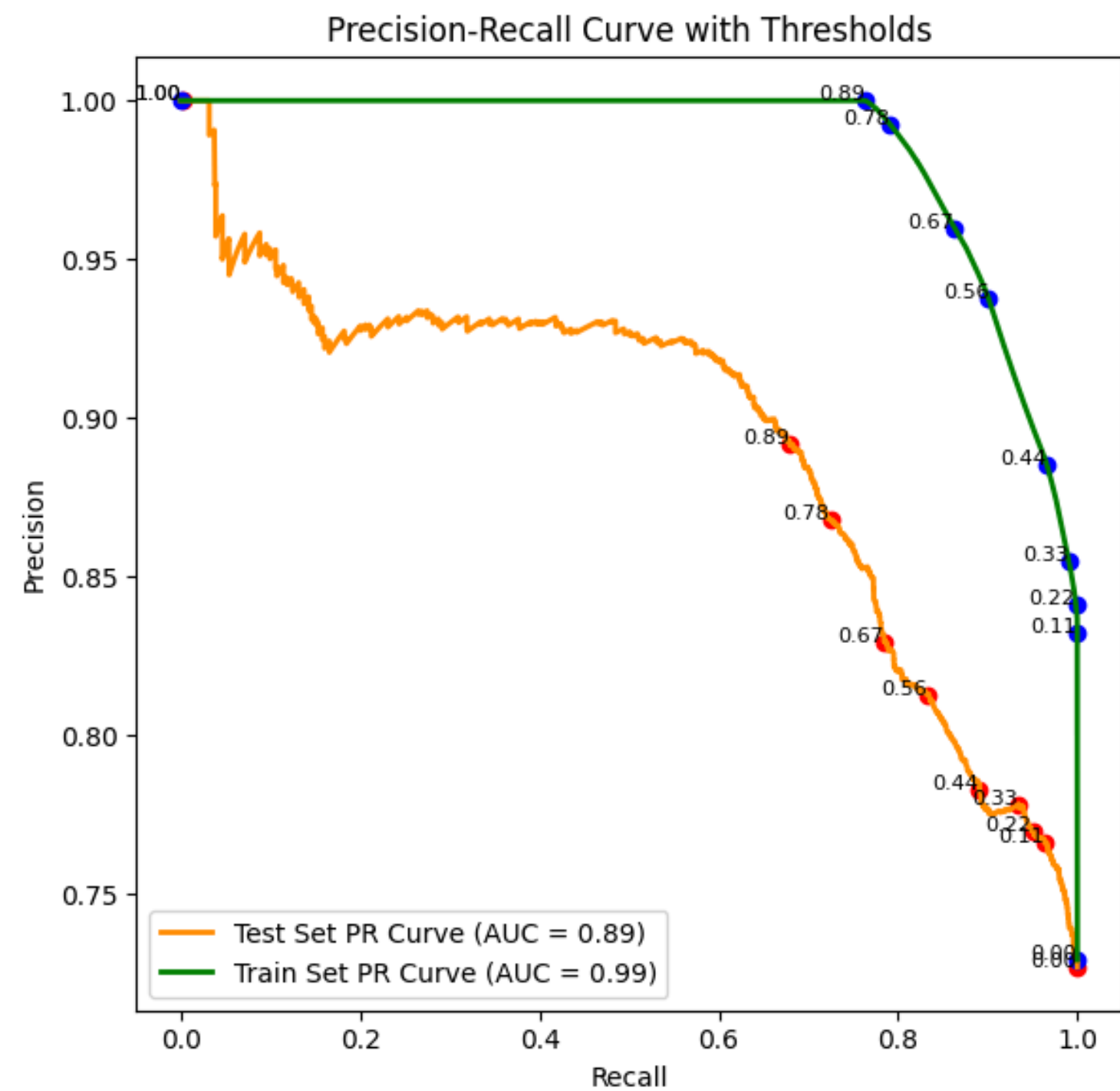
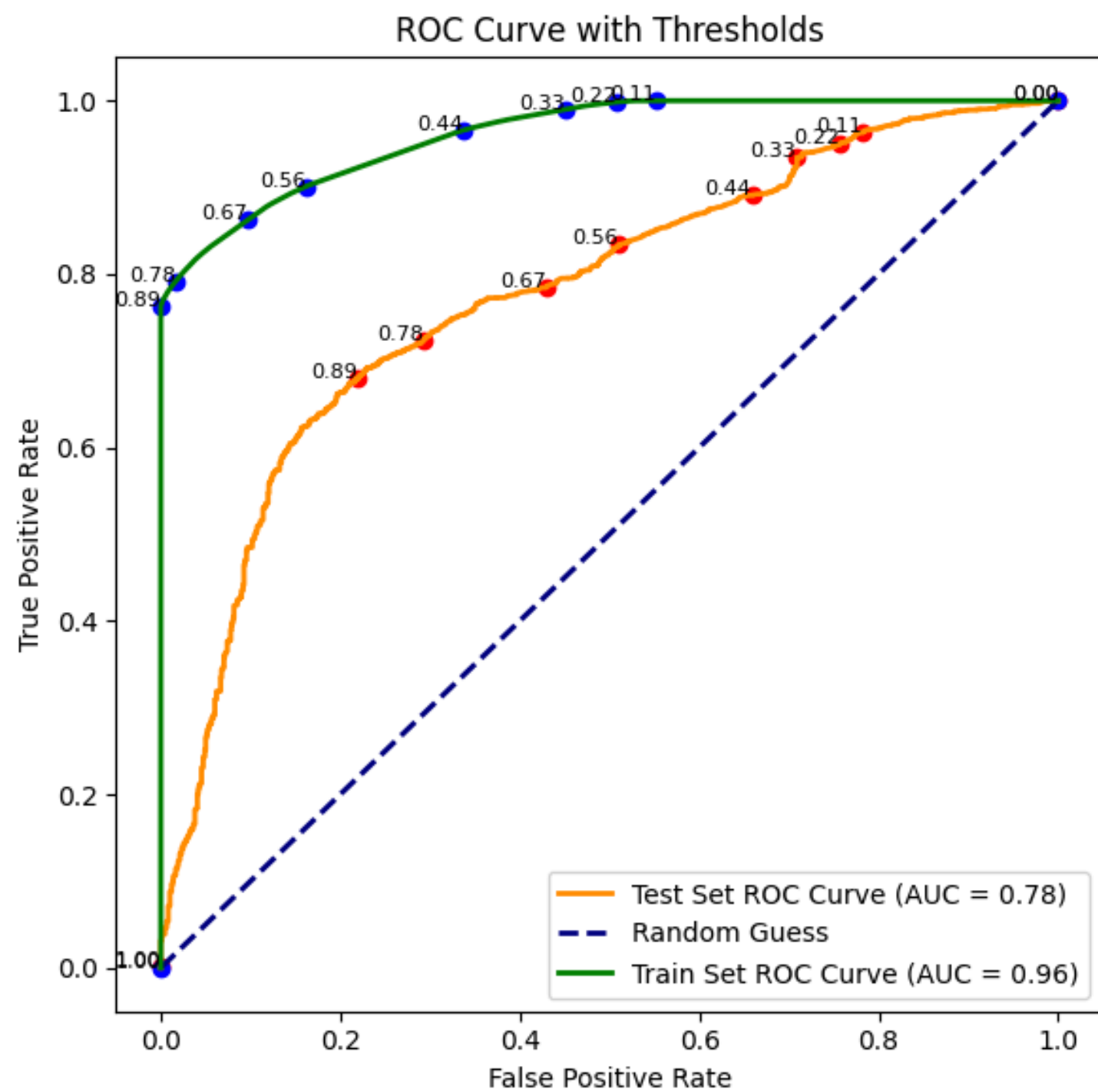
◆ Performances du modèle (test)

- AUC ROC $\approx 0,78$
- AUC PR $\approx 0,89$
- Précision globale $\approx 73\%$
- Bon rappel sur les « satisfaits »

◆ Variables les plus influentes

- Proportion d'arbres (parcs, forêts)
- Densité de population
- Distance au centre-ville

Visualisation de données



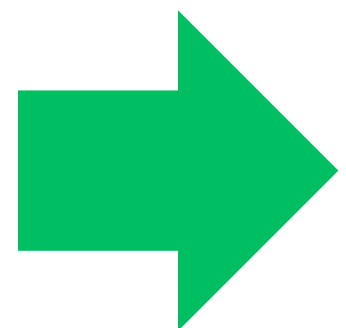
ET SI ON APPLIQUAIT CE MODELE A MONTELIMAR ?

Données locales collectées (SCot & ECOLAB)

- BD TOPO (occupation du sol, voirie)
- Base Permanente des Equipements (BPE)
- Carreaux INSEE(200m x 200m)
- Données vectorielles(.gpkg): forets, parcs...)

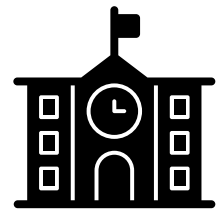
Travaux restants pour adapter le modèle:

- Harmonisation des formats et projections
- Jointures spatiales : équipements & population
- Calcul de variables comparables (buffers, densités)
- Réentraînement ou recalibrage local du modèle
- Comparaison avec avis Google



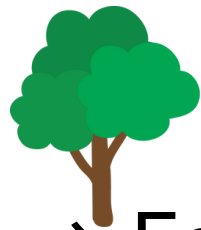
**UNE APPLICATION REALISTE MAIS QUI NECESSITE UN TRAVAIL
COMPLEMENTAIRE**

Comment ce modèle peut-il être utile aux collectivités ?



Diagnostic territorial:

- Identifier les quartiers où l'environnement végétal est le moins favorable
- Croiser bien-être perçu et inégalités territoriales



Simulation d'impact:

- Estimer les effets d'un projet de végétalisation sur la satisfaction attendue
- Aider à prioriser les investissements



Aide à la planification:

- Produire une cartographie prospective
- Intégrer des indicateurs de bien-être dans les documents d'urbanisme (PLU, SCOT...)



CONCLUSION

Conclusion & perspectives



Ce que nous avons accompli :

- *Modélisation du lien entre urbanisme vert et bien-être*
- *Bonnes performances du modèle ($AUC > 0,77$)*
- *Données locales collectées pour une future transposition*



Limites actuelles :

- *Transfert à Montélimar non finalisé*
- *Variables socio-démographiques à intégrer*
- *Travail technique restant sur les données géographiques*

Conclusion & perspectives

Perspectives :

- *Application concrète à l'échelle de Montélimar*
- *Aide au ciblage des politiques d'aménagement*



**MERCI POUR
L'ECOUTE!**