

Projet IA

Nathan Simon, Antoine Le Boulch, Tom Lelievre

Professeures responsables :

Nesma Settouti

Nadine Abdallah Saab

Sommaire

- Clustering
 - Préparation de données
 - Clustering / Métriques
 - Anomalie
 - Carte
- Prédiction de l'âge
 - Sélection des features
 - Encoder scaler
 - Les modèles
 - Script
- Alerte pour les tempêtes

Clustering

Clustering

- Préparation des données
- Clustering et Métriques
- Anomalies
- Carte

Préparation des données

Sélection des
colonnes pour le
clustering et
l'affichage

Sélection des
colonnes pour la
détection des
anomalies

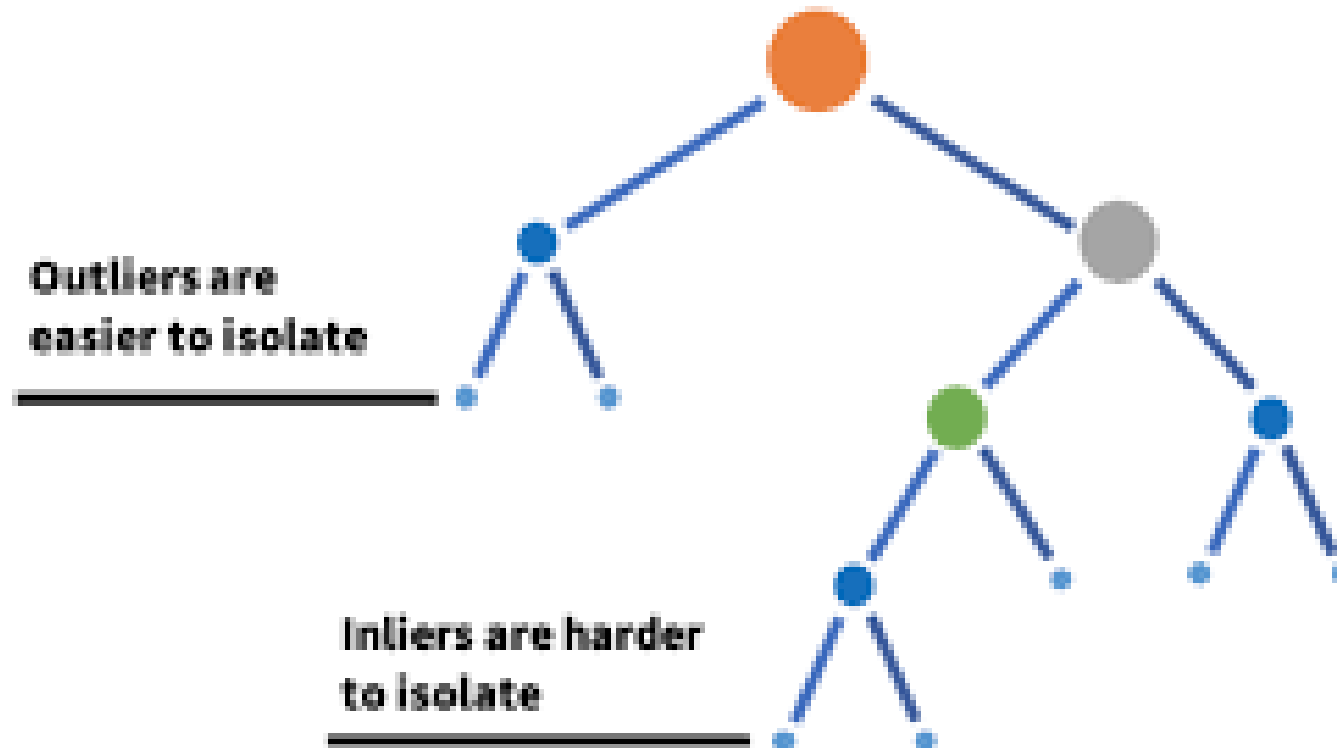
Clustering et Métriques

Pour 2 clusters

model	KMeans	Agglomerative	Spectral	GMM
silhouette	0.664	0.647	0.661	0.604
Davies-Bouldin	0.493	0.489	0.491	0.546
Calinski-Harabasz	27578.817	24068.899	26765.863	20546.308

Détection des anomalies

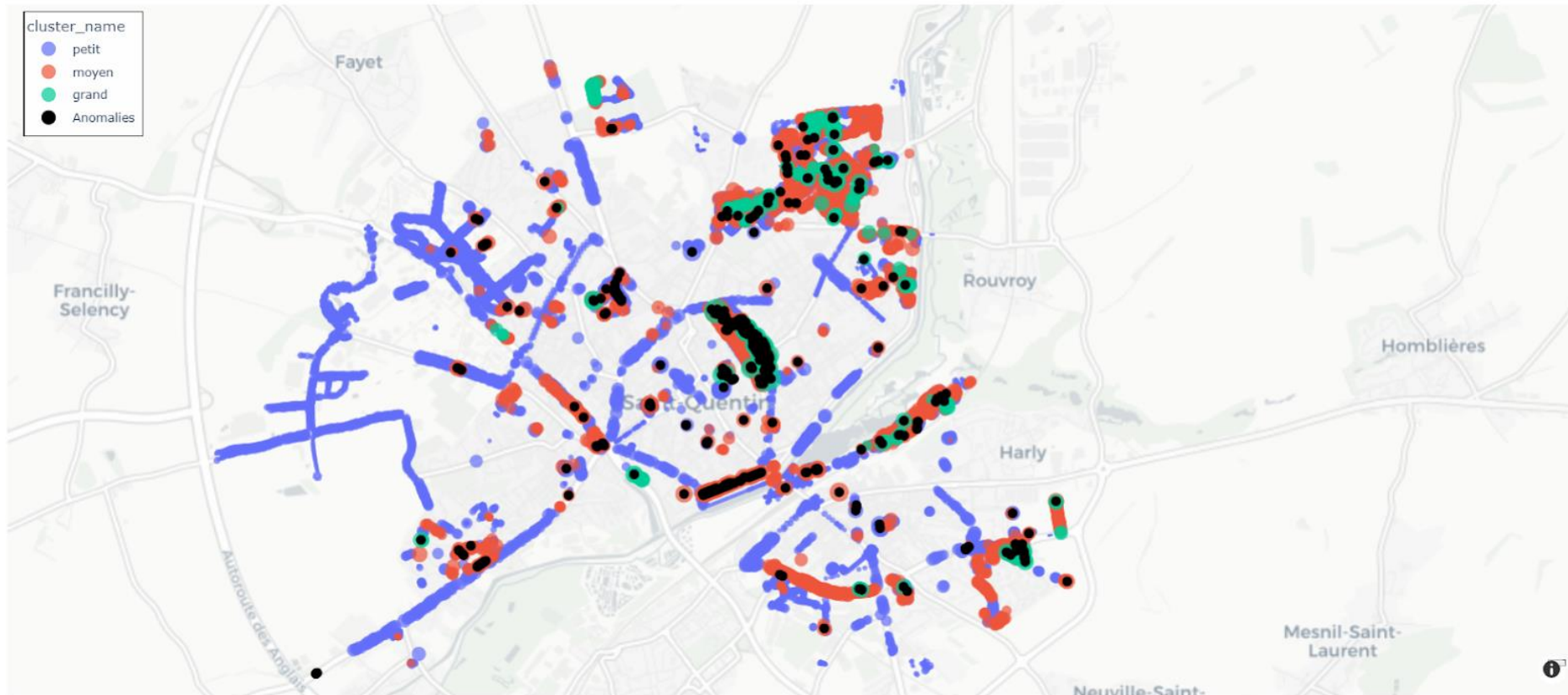
- Isolation Forest



Source : <https://www.linkedin.com/blog/engineering/data-management/isolation-forest>

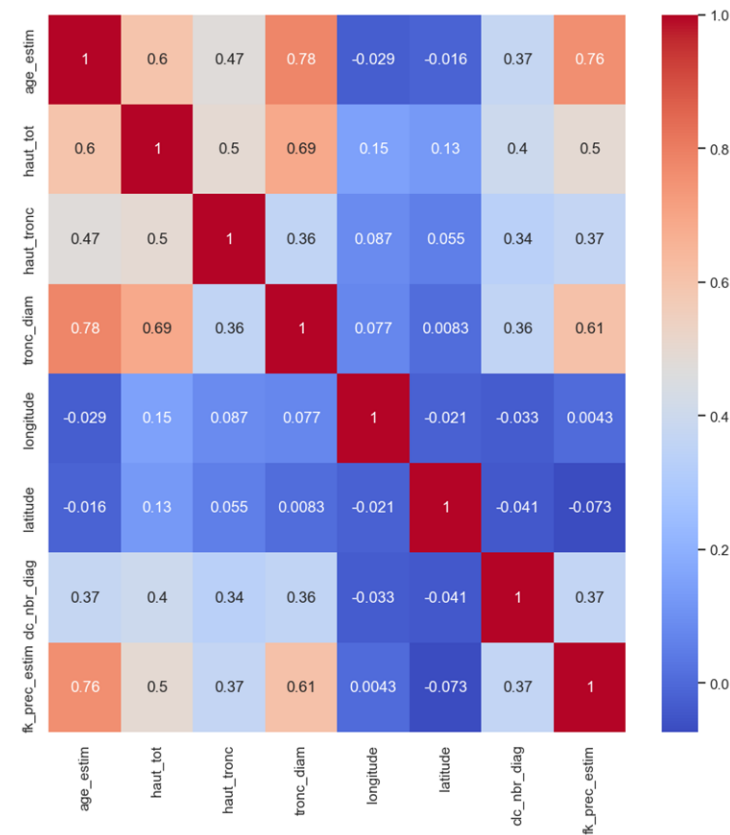
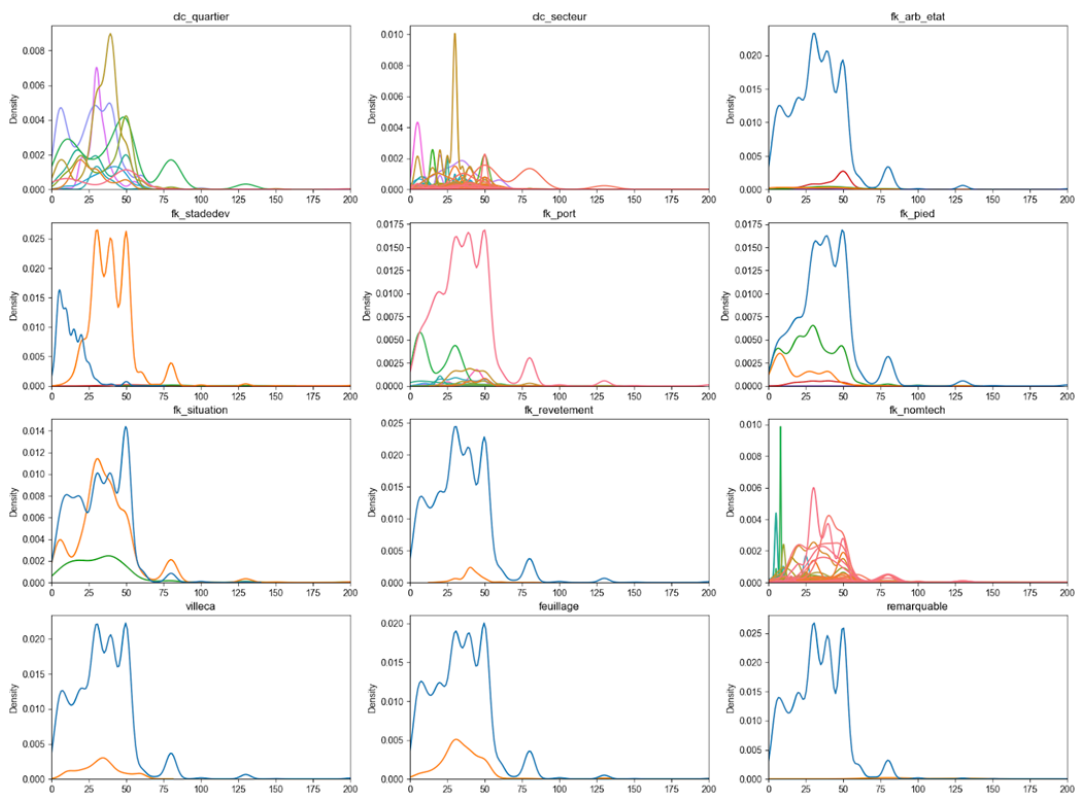
Visualisation sur une carte

Visualisation des Clusters et des Anomalies



Prédiction de l'âge

Sélection des features



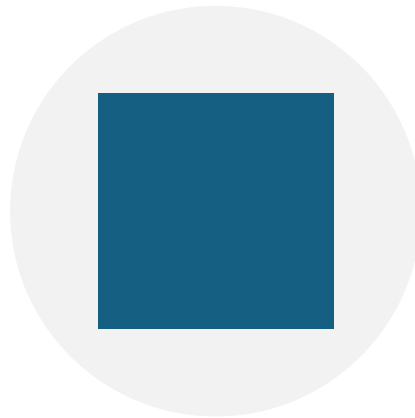
Sélection des features

haut_tronc, tronc_diam, fk_stadedev, clc_nbr_diag, fk_nomtech, haut_tot

Encodage et scaler



ORDINAL ENCODER
DE FK_STADEDEV



LABEL ENCODER



SCALER DE TOUS LES
FEATURES

Les modèles

- CART (decision tree)
- RandomForestRegressor
- Gradient Boosting
- XGBoost

Les modèles

CART :

Mean Absolute Error (MAE): 0.30149990787479575

Root Mean Squared Error (RMSE): 0.5084996857392401

R-squared (R^2): 0.7409979715286941

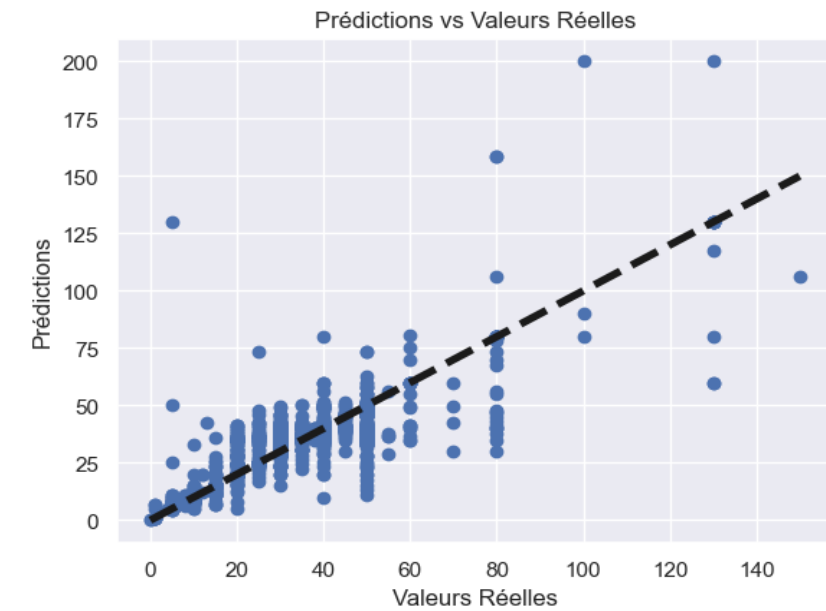
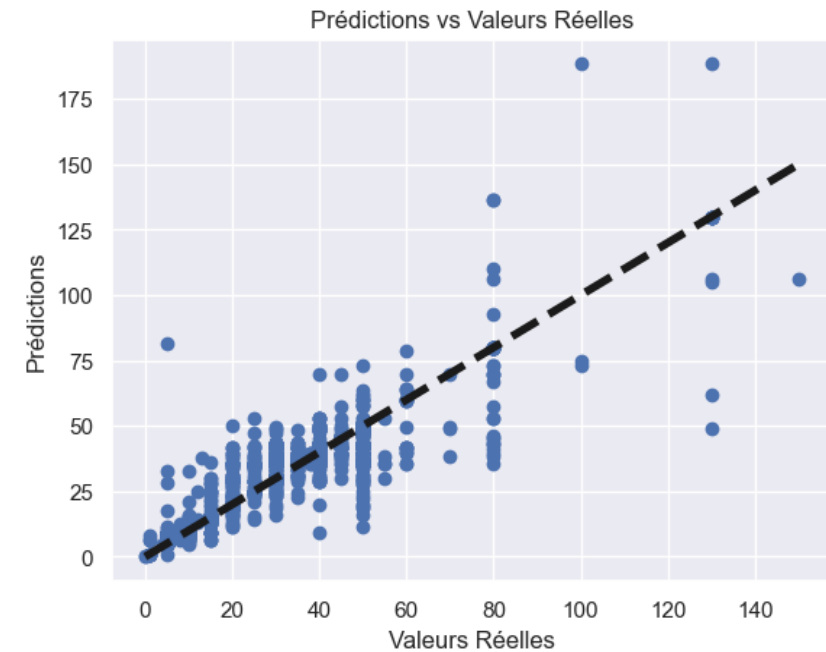
Cross Validation Score: 0.6556251672631337

R^2 : 0.6986828945215631

MSE: 0.30081673910055573

RMSE: 0.5484676281245373

mae: 0.2949840830214086



Les modèles

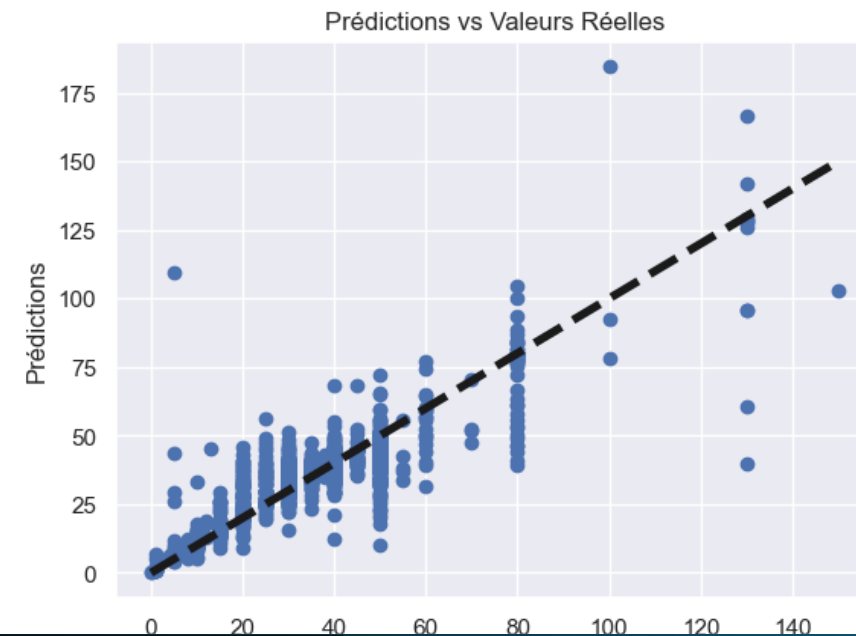
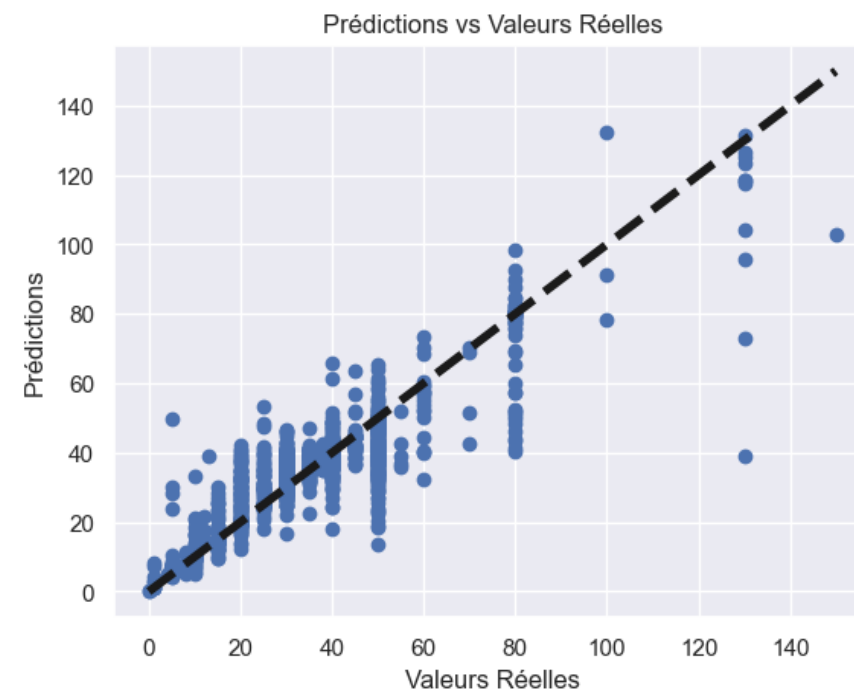
- RandomForest :
- Mean Absolute Error (MAE): 0.24500230701503742
- Root Mean Squared Error (RMSE): 0.42593542645308874
- R-squared (R^2): 0.8182772441708803
- Cross Validation Score: 0.7424395916910986
- OOB_score: 0.8220250708338028

From Scratch :

R^2 : 0.781358967348436

MSE: 0.21827795793866195

OOB Score: 0.8222008225196206



Boosting

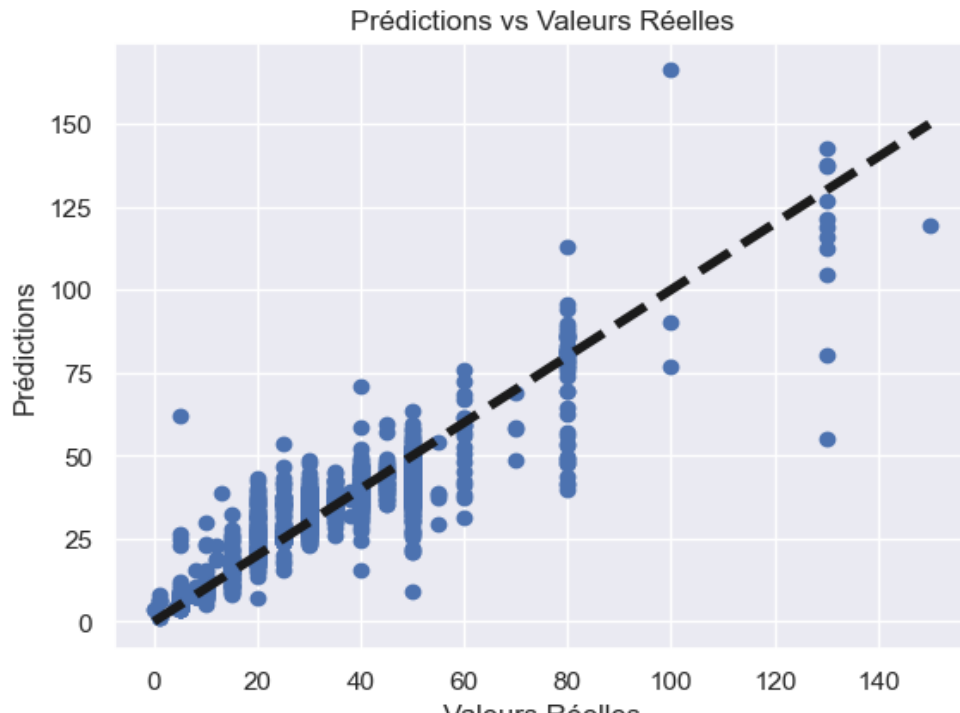
Gradient Boosting:

Mean Absolute Error (MAE): 0.2704448495050102

Root Mean Squared Error (RMSE): 0.4324043833542429

R-squared (R^2): 0.8127154450259364

Cross Validation Score: 0.7255052298199345



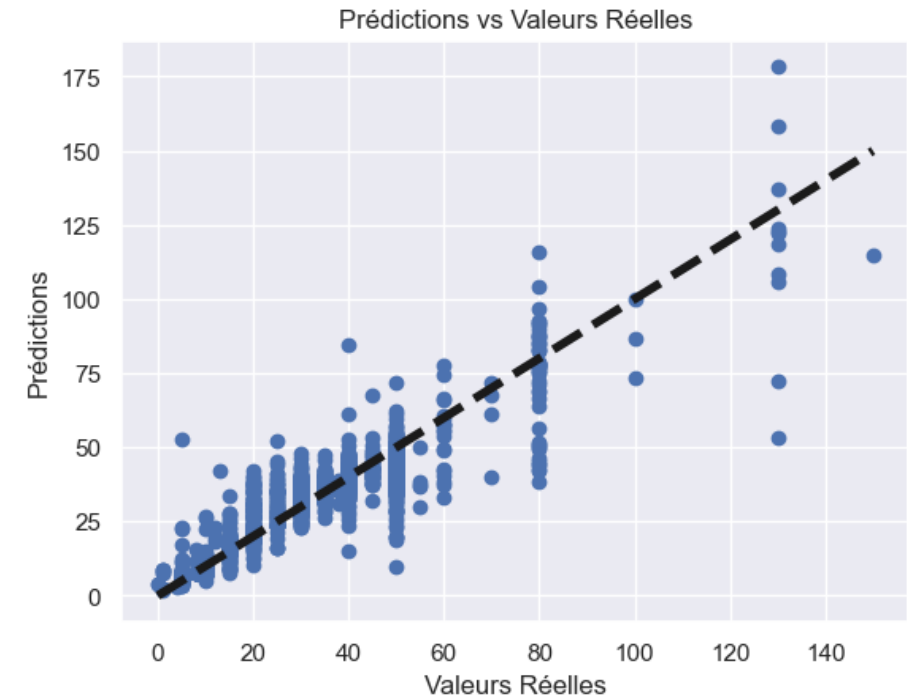
XGBoost :

Mean Absolute Error (MAE): 0.28883691011283663

Root Mean Squared Error (RMSE): 0.446949637808049

R-squared (R^2): 0.0.7999037419533639

Cross Validation Score: 0.739092603410952



Script :



IMPORTATION JSON DE
LA DATA DE TEST



IMPORTATION PKL



LECTURE
SELECTION ENCODAGE SCA
LER




PREDICTION
VIA RANDOMFOREST



IMPORTATION EN JSON

Alerte pour les tempêtes

Alerte pour les tempêtes

- Choix du modèle
 - RandomForestClassifier 
 - (XGBClassifier)
 - (MLPClassifier)
- Points clés du RF :
 - Classifieur pouvant faire du multi label ou binaire
 - Algorithme très puissant
 - Excellente accuracy
 - Overfit maîtrisé avec le bagging/bootstrapping aléatoire (les échantillons)
 - Résiste bien aux données bruitées grâce à son aléatoire (comme celles introduites par un SMOTE)
 - Supporte bien les données déséquilibrées

Choix des variables

- On ne garde que les arbre Éssôuçhé et Nộêssôuçhé
- Feature Selection avec `rf.feature_importances_`
 - On les trie par ordre décroissant d'importance
 - On garde les n premiers

	Feature	Importance
4	tronc_diam	0.064776
2	haut_tot	0.063204
1	latitude	0.053898
6	age_estim	0.047848
31	clc_secteur_E...	0.042228
0	longitude	0.040946
87	clc_secteur_S...	0.040637
3	haut_tronc	0.036857
159	villeca_VILLE	0.036351
13	clc_quantien	0.032500

SMOTE

- Utilisation de SMOTE pour équilibrer la proportion de Non essouché dans la population

Grid Search

- Optimisation « brute force » des paramètres

	÷	<u>123</u> 0	÷	<u>123</u> 1	÷
0			1.0		0.0
1			0.4		0.6

Export du modèle

- Export pour une utilisation dans le script sans avoir à faire de fit

```
prep_data_dict = {  
    "encoders" : dict_encoders,  
    "cat_cols_dist" : categorical_data_cols,  
    "bool_cols_list" : boolean_cols,  
    "top_features" : top_features  
}  
  
dict = {  
    "model" : gscv,  
    "prep_data_dict" : prep_data_dict  
}
```

Script python

- Utilisation de l'API de Visualcrossing pour obtenir des données météo.
 - Requêtes via http avec retour au format JSON
 - Donne le vent moyen et la vitesse des rafales pour le jour sélectionné

`https://weather.visualcrossing.com/VisualCrossingWebServices/rest/services/timeline/saint%20quentin/{date}/{date}?unitGroup=metric&elements=datetime%2Cwindgust%2Cwindspeedmean&key={API_KEY}&contentType=json"`

Lieu

Date

Vitesse en
rafale

Vitesse
moyenne

Clé de l'API

Type de retour

Eléments demandés à l'API

Choix des arbres à afficher



Choix des arbres à afficher en fonction de leur probabilité Essouché et de la vitesse du vent donnée par l'API



Echelle de Beaufort choisie comme référence pour apprécier l'effet du vent et choisir les paliers

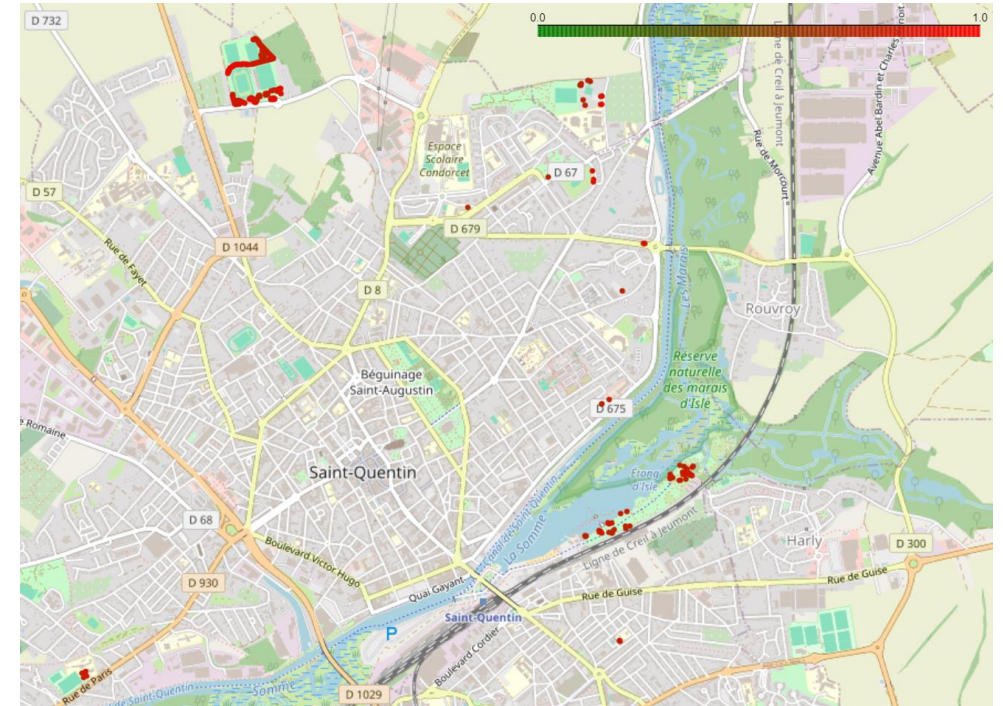
ECHELLE DE BEAUFORT				
ECHELLE	TERME	VENT MOYEN EN KM/H	EFFETS EN MER	EFFETS SUR TERRE
0	CALME	<1 km/h	Mer lisse, sans vague	Fumées verticales
1-2-3-4-5	BRISES	1 à 38 km/h	Quelques rides à quelques vagues (vagues: 0.2 à 2.5 m)	La fumée indique la direction du vent, les drapeaux flottent
6	VENT FRAIS	39 à 49 km/h	Crêtes d'écumes blanches, embruns (vagues: 2.5 à 4 m)	Les branches des arbres s'agitent, parapluies qui se retournent
7	GRAND FRAIS	50 à 61 km/h	Trainées d'écumes, lames déferlantes (vagues: 4 à 5.5 m)	Les arbres se balancent, la marche contre le vent est freinée
8	COUP DE VENT	62 à 74 km/h	Tourbillons d'écumes à la crête des lames (vagues: 5.5 à 7 m)	Les branches d'arbres peuvent casser, la marche est difficile
9	FORT COUP DE VENT	75 à 88 km/h	Lames déferlantes grosses à énormes, visibilité réduite (vagues: 7 à 9 m)	Le vent peut endommager les bâtiments, envol de tuiles
10	TEMPÊTE	89 à 102 km/h	Très grosses lames à longue crête en panache (vagues: 9 à 12 m)	Dégâts importants aux bâtiments, arbres parfois déracinés
11	VIOLENTE TEMPÊTE	103 à 117 km/h	Lames exceptionnellement hautes (vagues: 12 à 16 m)	Ravages étendus et importants
12	OURAGAN	>128 km/h	L'air est plein d'écume et d'embruns (vagues >16 m)	Dégâts majeurs et catastrophiques

Création de la carte

- Finalité du Besoin Client 3
- Librairie Folium
 - Popups pour afficher les informations en cliquant sur un arbre
 - Taille des cercles en fonction du diamètre de l'arbre
 - Couleur en fonction de la probabilité Essouché (dégradé vert (0) → rouge (1))



Folium



Démo des scripts

[illegible]

Fin

- Merci pour votre attention

