

RASSANT Nolan  
NDOMBI MOURI Caprice  
ROUSSEL Benjamin

## **RENDU ATELIER DE MACHINE LEARNING**

Dépôt Github : [https://github.com/NolanRasant/atelier\\_machine\\_learning](https://github.com/NolanRasant/atelier_machine_learning)

### Problématique

- Est-ce que l'évaluation (Synthese\_eval\_sanit) diffère selon le type d'établissement (APP\_Libelle\_activite\_etablissement) ?
- Est-ce que les inspections se passent à une certaine période (Date\_inspection) dans l'année ?
- Est-ce que le lieu (Libelle\_commune ou/et Code\_postal) joue sur la note (Synthese\_eval\_sanit) ?

### Hypothèse

Il semblerait que les dates, lieux et types d'établissement peuvent jouer sur la note d'évaluation, et donc le compte rendu de l'inspection.

On constate aussi qu'il y a certains mois pendant lesquels les inspections sont plus fréquentes pour une catégorie de types d'établissements.

### Réponse

Pour le modèle de régression linéaire, nous obtenons un coefficient de 0,054 (très faible). Ce qui signifie que ce modèle n'est pas le plus adapté. Il ne parvient pas à prédire nos données de manière significative.

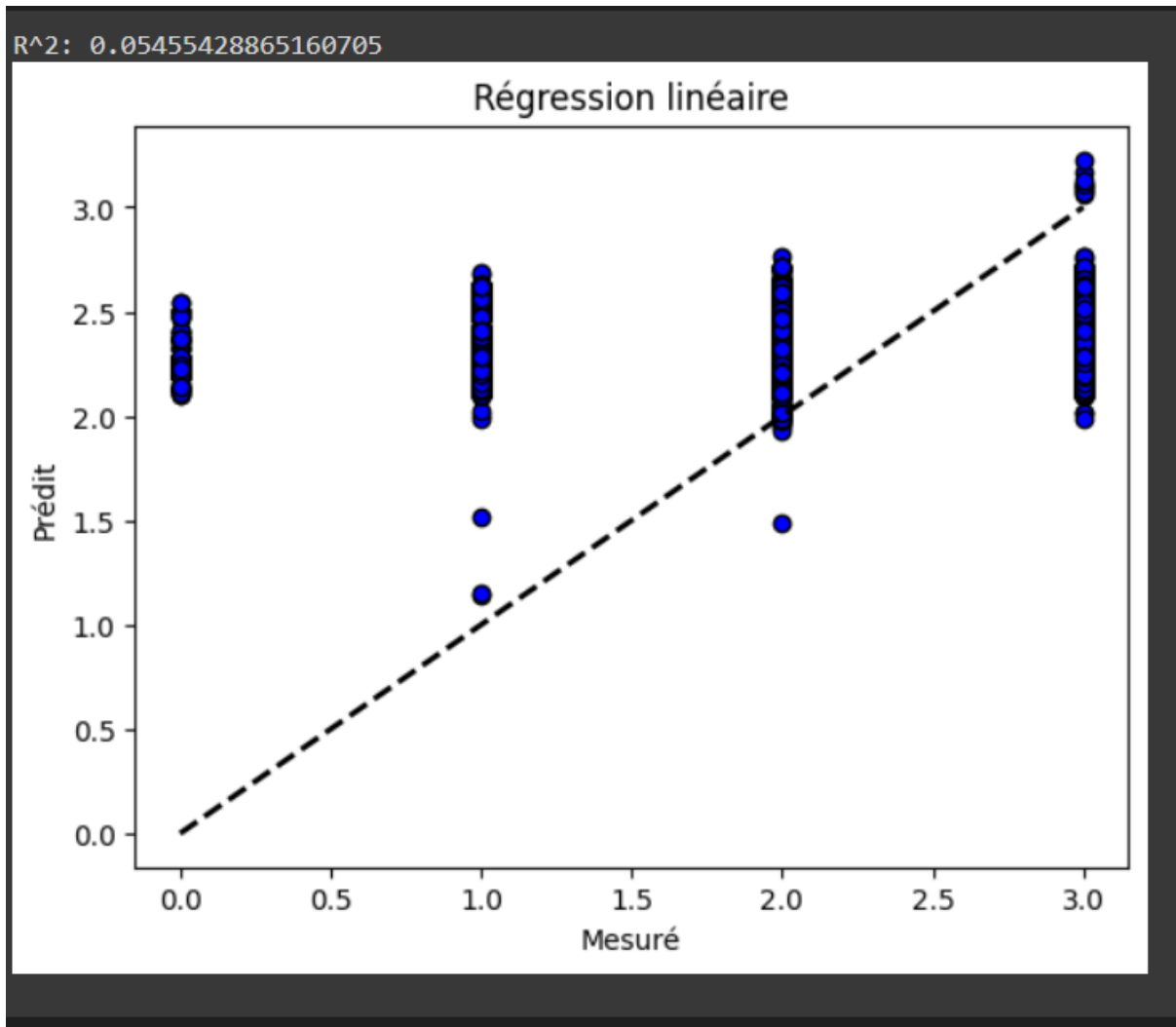
Avec le modèle de régression logistique, nous parvenons déjà à quelques meilleurs résultats car on arrive à prédire correctement sur un peu plus de la moitié de l'ensemble de nos échantillons de tests ( accuracy de 0,54).

Nous avons ensuite appliqué le modèle Random Forest sur l'ensemble des données de test. Nous obtenons un rapport beaucoup plus détaillé sur les performances fortes et faibles de notre prédiction.

## Modèles

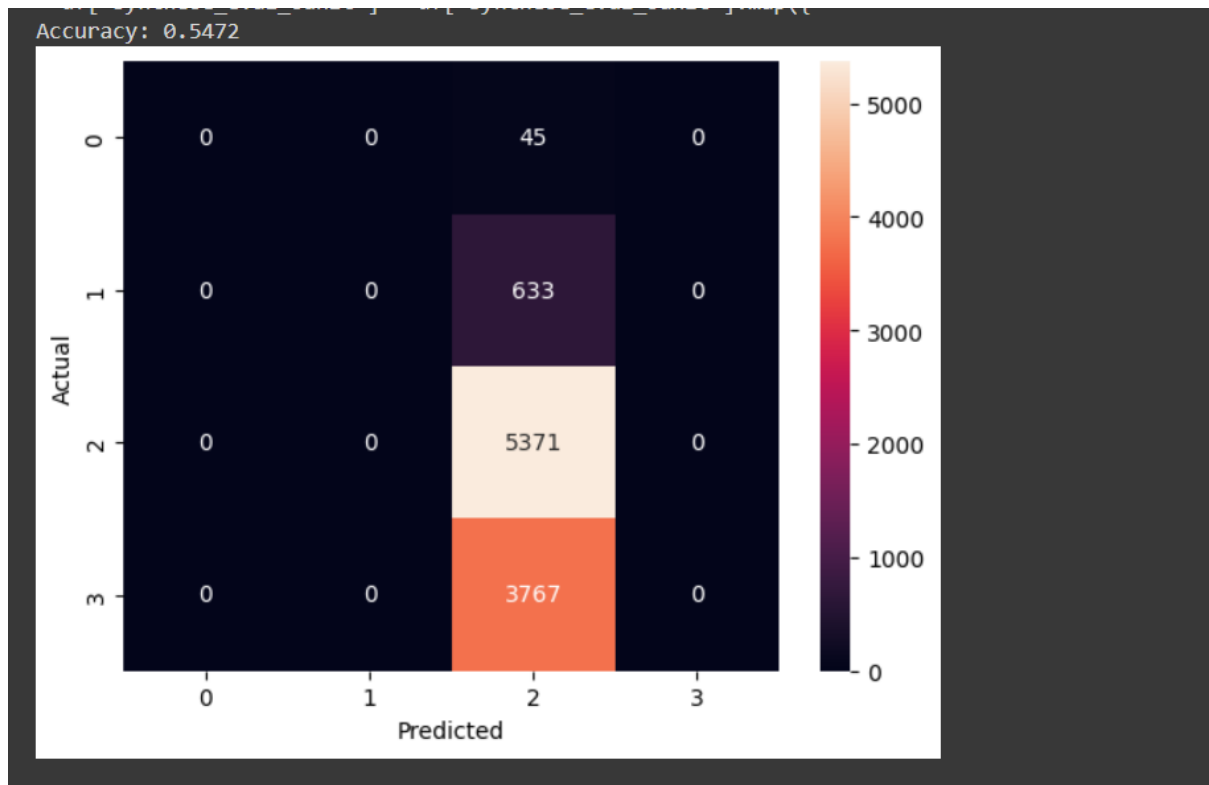
- Régression linéaire

Pour la régression linéaire, nous avons choisis de prédire la note de satisfaction selon le SIRET, le filtre et le type d'activité.



- Régression logistique

De même, nous avons choisis de prédire la note de satisfaction selon le SIRET, le filtre et le type d'activité.



- Random forest

Ici, nous avons prédit la note de satisfaction selon tous les critères

```
Accuracy: 0.4389
Matrice de confusion :
[[ 8   0  61 356]
 [ 0   0   2  22]
 [ 1   0 555 3038]
 [ 1   0 191 2309]]
Rapport de classification :
```

	precision	recall	f1-score	support
A améliorer	0.80	0.02	0.04	425
A corriger de manière urgente	1.00	0.00	0.00	24
Satisfaisant	0.69	0.15	0.25	3594
Très satisfaisant	0.40	0.92	0.56	2501
accuracy			0.44	6544
macro avg	0.72	0.27	0.21	6544
weighted avg	0.59	0.44	0.36	6544