Présentation ML :

Diagramme de Pareto : On observe que 80% des interventions entre 2020 et 2023 sont des alarme automatique (défaut), des feu, inondations et des interventions qualifiés de Divers cad Odeur suspect, odeur de gaz, arbre sur la route, court-circuit etc. La majorité c’est des alarmes automatiques dû à un défaut technique cela se passe dès que les nouveaux immeubles sont construits.

La médian des interventions est 1h après il existe quelques interventions exceptionnelles qui ont duré 25h ou 49h par exemple

22 features : T['tempmax', 'tempmin', 'temp', 'feelslike', 'dew', 'humidity', 'precip', 'precipprob', 'precipcover', 'snow', 'snowdepth', 'windspeed','winddir', 'sealevelpressure', 'cloudcover', 'visibility', 'solarradiation', 'solarenergy', 'severerisk', 'Intervention', 'Weekend', 'Precip\_last\_7D', 'holidays']

Nombre de données : 1579 données entre 2020 et 2023

J’ai fait plusieurs modèles et au début hésite entre prédire si interventions ou non ou alors aller plus dans le détail cad prédire plus précisément le type d’intervention.

Je me suis rendu compte en faisant les deux que le modèle a mieux prédire si intervention ou non le score f1 et recall était plus bas pour si je prédit le type d’intervention.

* Je vous présente mes résultats des 4 modèles sur la prédiction si intervention ou non et je me base sur le score Recall et f1-score pour évaluer mon modèle.
* Pk : mon but est d’avoir le plus petit nombre de faux négatif cad de jour mentionnant qu’il n’y a pas d’intervention alors qu’il y en a 🡺 Le recall permet de savoir le pourcentage de positifs bien prédit par notre modèle est une solution ou f1-score est une bonne solution est un bon compromis entre précision et recall.
* J’ai voulu comparer les meilleures features pour un recall et un f1-score optimale pour chaque modèle et voir si même chose ainsi quel modèle donne des valeurs plus hautes pour le recall et F1-score.