CheatSheet LightGBM

source : [Beginner’s Guide to the Must-Know LightGBM Hyperparameters](https://towardsdatascience.com/beginners-guide-to-the-must-know-lightgbm-hyperparameters-a0005a812702#2d6f)

# Hyperparamètres

| **Hyperparamètre** | **Objectif** | **Valeurs exemple** |
| --- | --- | --- |
| Boosting\_type | Type d’algorithe | 'gbdt' (Gradient Boosting Decision Tree)  'dart' (Dropouts meet Multiple Additive Regression Trees)  'goss' (Gradient-based One-Side Sampling)  'rf' (Random Forest) |
| objective | Type d’objectif de d’algorithme | Classification :  'Binary' / 'cross\_entropy'  Classification Multiclasse :  'multiclass' / 'softmax'  Régression :  'regression' / 'regression\_L1' / 'regression\_L2'  Autres :  'quantile' / 'huber' |
| metric | Métrique à utiliser pour évaluer la performance du modèle pendant l'entraînement | Classification Binaire :  'binary\_logloss' : Perte logistique pour la classification binaire.  'binary\_error' : Taux d'erreur binaire (pourcentage d'erreurs de classification).  Classification Multiclasse  'multi\_logloss' : Perte logistique pour la classification multiclasse.  'multi\_error' : Taux d'erreur multiclasse (pourcentage d'erreurs de classification).  Régression :  'l2' : Erreur quadratique moyenne (MSE) pour la régression.  'l1' : Erreur absolue moyenne (MAE) pour la régression.  Autres :  'quantile' : Perte quantile pour la régression quantile.  'huber' : Perte de Huber pour la régression robuste.  Mesures Personnalisées :  Vous pouvez également spécifier vos propres mesures personnalisées si nécessaire. |
| seed | Choix du jeu de test pseudo-  aléatoire | [0, ∞] ⇒ 'seed': 42, |
| verbose | Quantité d’information à afficher | [0, 1] ⇒ 'verbose': 1  verbose=1 pour afficher des détails pendant l'entraînement |
| num\_iterations | Boosting round ⇒ nombre d’arbre | [1, ∞] ⇒ 'num\_iterations': 100 |
| learning\_rate | “Performance” | [0, 1] ⇒ 'learning\_rate': 0.05  Plus la valeur est petite plus la performance augmente. Pour commencer : learning\_rate=0.05 →0.005 est très petit |
| early\_stopping\_round | Arrêt anticipé (itérations) | [1, ∞] ⇒ 'early\_stopping\_round': 10  nombre d’itération sans amélioration de la métrique choisie |
| num\_leaves | Nombre maximal de feuilles | [1, ∞] ⇒ 'num\_leaves': 31  plus il est grand plus il a tendance à l’**overfiting** |
| max\_depth | Profondeur de l’arbre | [1, ∞] ⇒ 'max\_depth': -1 # arbre de profondeur illimité |
| min\_data\_in\_leaf | Données minimales par feuille | [1, ∞] ⇒ 'min\_data\_in\_leaf': 20 |
| feature\_fraction | Limite la fraction, utile pour le sous- échantillonage des features | (0, ∞] ou ]0, ∞] ⇒ 'feature\_fraction': 0.8 |
| bagging\_fraction | Introduction du bagging pour réduire la variance | (0, ∞] ou ]0, ∞] ⇒ 'bagging\_fraction': 0.8 |
| bagging\_freq | Fréquence à laquelle le bagging est effectué | [0, ∞] ⇒ 'bagging\_freq': 1 |
| lambda\_l1, lambda\_l2 | L1, L2 | [0, ∞] ⇒ 'lambda\_l1': 0.0 / 'lambda\_l2': 0.0 |
| min\_child\_samples | Echantillons mininales par feuille | [1, ∞] ⇒ 'min\_child\_samples': 20 |
| min\_split\_gain | Gain minimal pour scinder | [0, ∞] ⇒ 'min\_split\_gain': 0.0 |
| cat\_smooth | Lissage des variables catégorielles | [0, ∞] ⇒ 'cat\_smooth': 1.0 |