



# DÉCOUVREZ Skrub

**LA RÉVOLUTION DE LA  
PRÉPARATION DES DONNÉES  
POUR LE MACHINE LEARNING**

Une bibliothèque Python  
pensée pour booster votre  
productivité et simplifier  
vos pipelines ML, même  
avec des données "sales"



Skrub

1

# POURQUOI Skrub ?

Pourquoi Skrub ? Moins de préparation, plus de ML

Les outils classiques (pandas, scikit-learn) sont fastidieux et chronophages pour la préparation de :

- Données hétérogènes
- Formats imprévus

👉 Skrub optimise l'étape de pré-traitement des données.

Visitez notre site internet sur:



[www.sd4hope.com](http://www.sd4hope.com)




Skrub

2

# LES SUPER-POUVOIRS

## de Skrub

Un couteau suisse pour vos données tabulaires et compatible avec **Scikit-Learn** :

 **TableReport** : analyse exploratoire interactive

 **Cleaner** : nettoyage intelligent

 **TableVectorizer** : vectorisation automatique

 **Fuzzy Join & AggJoiner** : enrichissement par jointures floues

 **Tabular\_learner** : pipelines scikit-learn simplifiés

Visitez notre site internet sur:



[www.sd4hope.com](http://www.sd4hope.com)



Skrub

3

# TableReport

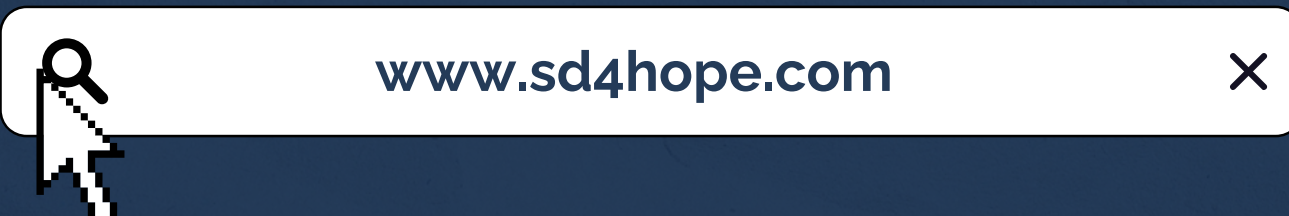
## EXPLORER VOS DONNÉES EN UN CLIC

Analyse exploratoire sans douleur :

- Statistiques descriptives
- Corrélations entre colonnes
- Détection de valeurs aberrantes
- Export HTML interactif

💡 Alternative moderne à ydata-profiling

Visitez notre site internet sur:



Skrub

4

# CLEANER

## LE MÉNAGE, VERSION ML

Fini les CSV mal formés !

- Conversion automatique des dates
- Gestion des valeurs manquantes
- Suppression des colonnes inutiles

🧠 Idéal pour les données bruyantes et non standardisées.

Visitez notre site internet sur:



[www.sd4hope.com](http://www.sd4hope.com)



Skrub

5

# TableVectorizer

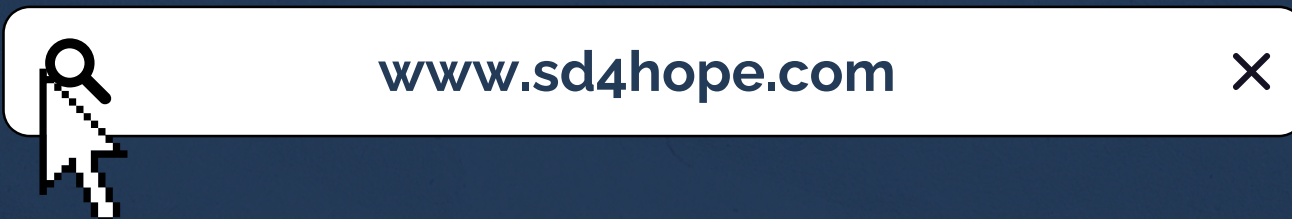
## PRÉTRAITEMENT INTELLIGENT

Encodage automatique des colonnes :

- Numériques, catégorielles, dates
- OneHot, GapEncoder, MinHash, TextEncoder...
- Détection des types et nettoyage intégré.

 Prêt pour l'entraînement en quelques lignes de code.

Visitez notre site internet sur:



Skrub

6

# ENRICHIR VOS DONNÉES AVEC DES JOINTURES FLOUES

Fuzzy Join, AggJoiner, MultiAggJoiner...

- Joignez des tables même avec des fautes de frappe
- Créez de nouvelles features à partir de data lakes
- Travaillez avec des jointures approximatives et robustes

 Boostez vos modèles avec des données enrichies.

Visitez notre site internet sur:



[www.sd4hope.com](http://www.sd4hope.com)



Skrub

7

# Expressions Skrub

## PIPELINES COMPLEXES, SIMPLIFIÉS

Nouveauté 2025 : Expressions Skrub

- Suivi des splits train/val/test
- Réglage d'hyperparamètres simplifié
- Pipelines multi-tables reproductibles

🧩 Idéal pour les projets ML complexes.

Visitez notre site internet sur:



[www.sd4hope.com](http://www.sd4hope.com)



Skrub

8


# POUR ALLER PLUS LOIN

Ressources utiles :

 Documentation officielle : <https://skrub-data.org/skrub-materials/index.html>

 Article en français :  
<https://www.stat4decision.com/fr/skrub-preparer-donnees-python-ml/>

**Prêt à booster vos projets ML ?**

 Likez, commentez, partagez si vous trouvez cela utile !

Visitez notre site internet sur:



www.sd4hope.com





**Transforming data into strategic asset**

 +226 60 02 13 13

 [www.sd4hope.com](http://www.sd4hope.com)

 [hello@sd4hope.com](mailto:hello@sd4hope.com)