

Téléchargement et Préparation du Dataset Covertypes

Téléchargement de la base de données Covertypes

La base de données **Covertypes** provient de l'[UCI Machine Learning Repository](#).

Vous pouvez la télécharger directement via le lien ci-dessous :

- [Télécharger le dataset Covertypes \(CSV\)](#)

Chargement et Réduction des Données

```
import pandas as pd
from sklearn.model_selection import train_test_split

# Téléchargement direct des données depuis l'URL
url = "https://archive.ics.uci.edu/ml/machine-learning-databases/covtype/covtype.data.gz"
column_names = [f'Feature_{i}' for i in range(1, 55)] + ['Cover_Type']
data = pd.read_csv(url, header=None, names=column_names)

# Réduction simple à 4% des données sans rééquilibrage des classes
sample_size = int(0.04 * len(data)) # 4% de la base d'origine
data_sampled = data.sample(n=sample_size, random_state=42).reset_index(drop=True)

# Division des données en ensembles d'entraînement, de validation et de test
train_data, temp_data = train_test_split(data_sampled, test_size=0.4, random_state=42, stratify=data_sampled['Cover_Type'])
val_data, test_data = train_test_split(temp_data, test_size=0.5, random_state=42, stratify=temp_data['Cover_Type'])

# Afficher la taille des ensembles
print(f"Entraînement : {len(train_data)} lignes")
```

```
print(f"Validation : {len(val_data)} lignes")
print(f"Test : {len(test_data)} lignes")

# Sauvegarder les ensembles en fichiers CSV
train_data.to_csv('covertime_train.csv', index=False)
val_data.to_csv('covertime_val.csv', index=False)
test_data.to_csv('covertime_test.csv', index=False)
```

Entraînement : 13944 lignes

Validation : 4648 lignes

Test : 4648 lignes

Liens pour télécharger les ensembles préparés

- [Télécharger l'ensemble d'entraînement](#)
- [Télécharger l'ensemble de validation](#)
- [Télécharger l'ensemble de test](#) ““