

Record Linkage Comparison Patterns

UCI Machine Learning Repository

Présentation de la base de données

Data Set Characteristics:	Multivariate	Number of Instances:	5749132	Area:	N/A
Attribute Characteristics:	Real	Number of Attributes:	12	Date Donated	2011-03-10
Associated Tasks:	Classification	Missing Values?	Yes	Number of Web Hits:	69069

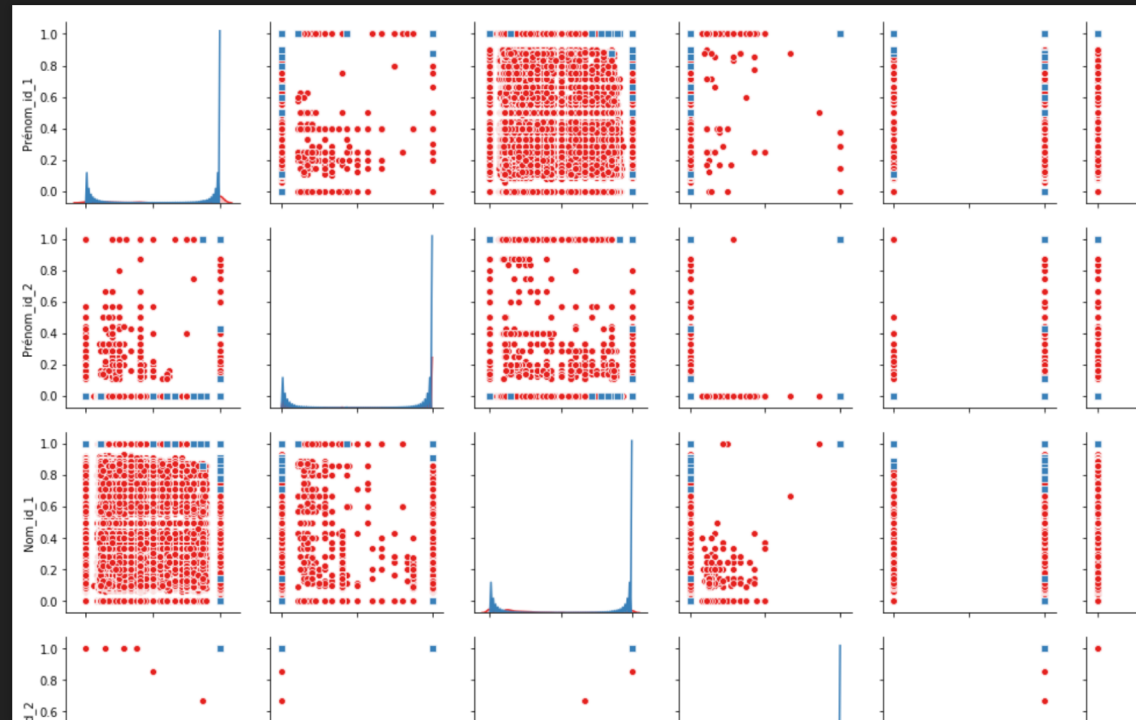
- Cet ensemble de données est une comparaison élémentaire des enregistrements avec des données personnelles provenant d'un paramètre de couplage d'enregistrements. Il peut être utilisé pour déterminer, à partir d'un modèle de comparaison, si les enregistrements sous-jacents appartiennent à une personne.

Les relations par paires dans un jeu de données

`seaborn.pairplot ()`

Par défaut, cette fonction crée une grille d'axes telle que chaque variable de données soit partagée sur l'axe des y sur une seule ligne et sur l'axe des x sur une seule colonne.

Il est également possible d'afficher un sous-ensemble de variables ou de tracer différentes variables sur les lignes et les colonnes.



Arbre de décision

Validation

Matrice de confusion :

```
[[1890322    5]
 [    19  6868]]
```

Précision :

1.0

rapport de classification :

	precision	recall	f1-score	support
0	1.00	1.00	1.00	1890327
1	1.00	1.00	1.00	6887
micro avg	1.00	1.00	1.00	1897214
macro avg	1.00	1.00	1.00	1897214
weighted avg	1.00	1.00	1.00	1897214

