Imputation univariée ou multivariée[¶](https://scikit-learn.org/stable/modules/impute.html#univariate-vs-multivariate-imputation)

[6.4. Imputation des valeurs manquantes — documentation scikit-learn 1.0.2](https://scikit-learn.org/stable/modules/impute.html)

On distingue l’imputation univariée, qui impute des valeurs dans la i-ème variable en utilisant uniquement des valeurs non manquantes dans cette variable. En revanche, les algorithmes d’imputation multivariés utilisent l’ensemble des variables disponibles pour estimer les valeurs manquantes (par exemple). impute.SimpleImputerimpute.IterativeImputer

Après le processus d’imputation, les données imputées de chaque méthode ont étés utilisés pour réaliser une tâche de classification présence absence de l’occupant à l’aide de Random Forest classifier ( hyperparametres :

rf\_model=RandomForestClassifier(max\_features='auto', n\_estimators= 200, max\_depth=None, criterion='entropy')

[Imputation de données manquantes en temps réel pour les données de capteurs à court terme des systèmes marins. Une étude comparative - ScienceDirect](https://www.sciencedirect.com/science/article/pii/S0029801820311823?casa_token=tW3QTDfFykoAAAAA:OXwZWMKhhSpCqoGxdRbScXyZomXQxglarO_dy6EWfEa1GzcaSAttCup48qABgNvtWpsAJEjS_g)

<https://stefvanbuuren.name/fimd/sec-MCAR.html>

[autoimpute · PyPI](https://pypi.org/project/autoimpute/)

[Prévision de séries chronologiques  |  TensorFlow Core](https://www.tensorflow.org/tutorials/structured_data/time_series#part_2_forecast_a_multivariate_time_series)

[imputation.pdf (uaem.mx)](http://www2.uaem.mx/r-mirror/web/packages/imputation/imputation.pdf)

[impyute · PyPI](https://pypi.org/project/impyute/)

Each function in this package includes the imputation algorithm as well as a cross validatiion algorithm. The CV algorithm artificially eliminates 1/3 of the data in a dataset, and runs the imputation function. Using the completed data, the RMSE is calculated on the portion of the data that was artificially removed only. Different imputation algorithms will perform differently on different datasets, so it is important to have these functions for comparison.

Le pourcentage de valeurs manquantes prises en compte pour cette étude est de 30 %, car il englobe la gamme de valeurs manquantes perçues dans les ensembles de données analysés recueillis à partir de systèmes de machines marines, qui contiennent de 4,4 % à 26 % de valeurs manquantes.

# La visualisation des données

La visualisation des données, étape fondamentale de la science de données, elle occupe un intérêt non négligeable dans le traitement et l’analyse des données. En effet, les graphes sont des outils puissants pour bien appréhender la complexité cachée dans les données. Sans une représentation des données sous forme visuelle, il est difficile de leurs donner un sens à un humain.