

```
!pip install missingpy
```

```
Collecting missingpy
  Downloading missingpy-0.2.0-py3-none-any.whl (49 kB)
    |████████████████████| 49 kB 4.5 MB/s
Installing collected packages: missingpy
Successfully installed missingpy-0.2.0
```

```
import pandas as pd
import sys
import sklearn.neighbors._base
sys.modules['sklearn.neighbors.base'] = sklearn.neighbors._base
from sklearn.impute import KNNImputer
from missingpy import MissForest
import numpy as np
```

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

```
cd /content/drive/My Drive
```

```
/content/drive/My Drive
```

```
df = pd.read_csv('/content/drive/MyDrive/Dat_file/KL_research/02102022/missing_data.csv')
```

```
df.shape
```

```
(21599, 56)
```

```
imputer = KNNImputer(n_neighbors=5)
imputer2 = MissForest()
```

```
X = df.iloc[0:21599,0:56]
```

```
X_new = imputer.fit_transform(X)
```

```
X_new2 = imputer2.fit_transform(X)
```

```
/usr/local/lib/python3.7/dist-packages/sklearn/ensemble/_forest.py:400: FutureWarning:
FutureWarning,
/usr/local/lib/python3.7/dist-packages/sklearn/ensemble/_forest.py:400: FutureWarning
FutureWarning,
/usr/local/lib/python3.7/dist-packages/sklearn/ensemble/_forest.py:400: FutureWarning
FutureWarning,
/usr/local/lib/python3.7/dist-packages/sklearn/ensemble/_forest.py:400: FutureWarning
FutureWarning,
```

```

/usr/local/lib/python3.7/dist-packages/sklearn/ensemble/_forest.py:400: FutureWarning
FutureWarning,
/usr/local/lib/python3.7/dist-packages/sklearn/ensemble/_forest.py:400: FutureWarning
FutureWarning,
/usr/local/lib/python3.7/dist-packages/sklearn/ensemble/_forest.py:400: FutureWarning
FutureWarning,
/usr/local/lib/python3.7/dist-packages/sklearn/ensemble/_forest.py:400: FutureWarning
FutureWarning,
/usr/local/lib/python3.7/dist-packages/sklearn/ensemble/_forest.py:400: FutureWarning
FutureWarning,
/usr/local/lib/python3.7/dist-packages/sklearn/ensemble/_forest.py:400: FutureWarning
FutureWarning,
/usr/local/lib/python3.7/dist-packages/sklearn/ensemble/_forest.py:400: FutureWarning
FutureWarning,
/usr/local/lib/python3.7/dist-packages/sklearn/ensemble/_forest.py:400: FutureWarning
FutureWarning,
/usr/local/lib/python3.7/dist-packages/sklearn/ensemble/_forest.py:400: FutureWarning
FutureWarning,
/usr/local/lib/python3.7/dist-packages/sklearn/ensemble/_forest.py:400: FutureWarning
FutureWarning,
/usr/local/lib/python3.7/dist-packages/sklearn/ensemble/_forest.py:400: FutureWarning
FutureWarning,
/usr/local/lib/python3.7/dist-packages/sklearn/ensemble/_forest.py:400: FutureWarning
FutureWarning,
/usr/local/lib/python3.7/dist-packages/sklearn/ensemble/_forest.py:400: FutureWarning
FutureWarning,
/usr/local/lib/python3.7/dist-packages/sklearn/ensemble/_forest.py:400: FutureWarning
FutureWarning,
/usr/local/lib/python3.7/dist-packages/sklearn/ensemble/_forest.py:400: FutureWarning
FutureWarning,
/usr/local/lib/python3.7/dist-packages/sklearn/ensemble/_forest.py:400: FutureWarning
FutureWarning,
/usr/local/lib/python3.7/dist-packages/sklearn/ensemble/_forest.py:400: FutureWarning
FutureWarning,
/usr/local/lib/python3.7/dist-packages/sklearn/ensemble/_forest.py:400: FutureWarning
FutureWarning,
/usr/local/lib/python3.7/dist-packages/sklearn/ensemble/_forest.py:400: FutureWarning
FutureWarning,
Iteration: 0
/usr/local/lib/python3.7/dist-packages/sklearn/ensemble/_forest.py:400: FutureWarning
FutureWarning,
/usr/local/lib/python3.7/dist-packages/sklearn/ensemble/_forest.py:400: FutureWarning
FutureWarning,

```

```
col = list(df.columns)
```

```
sgb_ran = pd.DataFrame(X_new2, columns= col)
```

```
sgb_knn = pd.DataFrame(X_new, columns= col)
```

```
sgb_knn.shape
```

```
(21599, 56)
```

```
sgb_knn.head()
```

	LAISUAT.logarithm_base_bal.2	LAISUAT.2.logarithm_base_bal	LAISUAT.logarithm_
0	1163.116447	2182.687556	12
1	1173.772549	2192.663303	12
2	1143.756894	2164.446405	11
3	1270.863526	2281.547064	12
4	1112.186335	1925.815482	12

```
sgb_ran.head()
```

	LAISUAT.logarithm_base_bal.2	LAISUAT.2.logarithm_base_bal	LAISUAT.logarithm_
0	1163.116447	2182.687556	12
1	1173.772549	2192.663303	12
2	1143.756894	2164.446405	11
3	1270.863526	2281.547064	12
4	1112.186335	1925.815482	12

```
from sklearn.impute import SimpleImputer
imputer3 = SimpleImputer(strategy='median')
```

```
X_new3 = imputer3.fit_transform(X)
```

```
X_new3
```

```
array([[1.16311645e+03, 2.18268756e+03, 1.23905907e+03, ...,
        1.00000000e+00, 1.00000000e+00, 1.00000000e+00],
       [1.17377255e+03, 2.19266330e+03, 1.24307601e+03, ...,
        1.00000000e+00, 1.00000000e+00, 1.00000000e+00],
       [1.14375689e+03, 2.16444641e+03, 1.13783082e+03, ...,
        1.00000000e+00, 1.00000000e+00, 1.00000000e+00],
```

```
...,  
[8.24710548e+02, 1.40866580e+03, 8.24710548e+02, ...,  
 0.00000000e+00, 0.00000000e+00, 1.00000000e+00],  
[8.21349341e+02, 1.05985981e+03, 8.21349341e+02, ...,  
 1.00000000e+00, 0.00000000e+00, 1.00000000e+00],  
[5.37145950e+02, 5.84512535e+02, 5.37145950e+02, ...,  
 0.00000000e+00, 1.00000000e+00, 1.00000000e+00]])
```

```
sgb_median = pd.DataFrame(X_new3, columns= col)
```

```
sgb_median.shape
```

```
↳ (21599, 56)
```

```
sgb_median.to_csv('Kien_L_med.csv')
```