

Usage Instructions

1. Taking the DSMR-regression elimination process as an example, you first need to place the Regression-elimination.py file and the dataset file in the same folder. The data in the dataset should be stored in .csv format, with the first column as the index, the second column as the target property, and the remaining columns as features.

名称	修改日期	类型	大小
datasets	2025/5/26 18:29	文件夹	
Regression-elimination.py	2024/10/30 14:52	JetBrains PyCharm	22 KB











	Hardness	Al	Cr	Co	Ni	Cu	y	G	μC
0	486	0.207921	0.207921	0	0.168317	0.207921	1.076036	17.6869	24.57505
1	469	0.11	0.28	0	0.28	0.27	1.076256	17.40837	24.6044
2	358	0.168317	0.168317	0.168317	0.247525	0.079208	1.076224	14.5742	24.79297
3	477	0.11	0.28	0	0.25	0.29	1.076258	17.56008	24.5621
4	140	0	0.267327	0.267327	0.19802	0	1.024018	8.376908	24.74673
5	567	0.326733	0.168317	0.168317	0	0.168317	1.07572	17.96374	24.35149

2. Modify the dataset name in the .py file, and adjust the algorithm and the range of hyperparameters as needed, as well as the step size for splitting the dataset into subsets (e.g., using 10-fold cross-validation as an example). Then, run the .py file.



```
import ...
#%% 数据切分
num_splits = 10











data = pd.read_csv(r"Raw-data.csv"
                  #, encoding='gb2312'
                  )
data = data.iloc[:,1:]
df = data
```

3. The DSMR framework will then automatically split the dataset into subsets (when adding algorithms, the eliminated data will be automatically split and re-added to the original dataset).

 1 Split.csv	✓	2024/9/5 22:27	Microsoft Excel ...	5 KB
 2 Split.csv	✓	2024/9/5 22:27	Microsoft Excel ...	5 KB
 3 Split.csv	✓	2024/9/5 22:27	Microsoft Excel ...	5 KB
 4 Split.csv	✓	2024/9/5 22:27	Microsoft Excel ...	5 KB
 5 Split.csv	✓	2024/9/5 22:27	Microsoft Excel ...	5 KB
 6 Split.csv	✓	2024/9/5 22:27	Microsoft Excel ...	5 KB
 7 Split.csv	✓	2024/9/5 22:27	Microsoft Excel ...	5 KB
 8 Split.csv	✓	2024/9/5 22:27	Microsoft Excel ...	5 KB
 9 Split.csv	✓	2024/9/5 22:27	Microsoft Excel ...	5 KB
 10 Split.csv	✓	2024/9/5 22:27	Microsoft Excel ...	5 KB

4. After the computation is completed, a result folder and a file showing the model ranking will appear in the same directory. The result folder contains the accuracy of the models built by DSMR for each data subset, as well as the corresponding .pickle model files.

 results		2025/5/26 18:37	文件夹	
 Ranking_statistics.csv		2024/9/6 6:00	Microsoft Excel ...	1 KB

 1 Split	✓	2024/10/31 14:13	文件夹
 2 Split	✓	2024/10/31 14:13	文件夹
 3 Split	✓	2024/10/31 14:13	文件夹
 4 Split	✓	2024/10/31 14:13	文件夹
 5 Split	✓	2024/10/31 14:13	文件夹
 6 Split	✓	2024/10/31 14:13	文件夹
 7 Split	✓	2024/10/31 14:13	文件夹
 8 Split	✓	2024/10/31 14:13	文件夹
 9 Split	✓	2024/10/31 14:13	文件夹
 10 Split	✓	2024/10/31 14:13	文件夹

models	✓	2024/10/31 14:13	文件夹	
1 Split.csv	✓	2024/9/5 23:13	Microsoft Excel ...	5 KB
modelAcc.csv	✓	2024/9/5 23:13	Microsoft Excel ...	6 KB

model_adb.pickle	2024/5/9 21:51	PICKLE 文件	152 KB
model_bpnn.pickle	2024/5/9 21:51	PICKLE 文件	11 KB
model_br.pickle	2024/5/9 21:51	PICKLE 文件	1,011 KB
model_cbr.pickle	2024/5/9 21:51	PICKLE 文件	586 KB
model_dtr.pickle	2024/5/9 21:51	PICKLE 文件	6 KB
model_etr.pickle	2024/5/9 21:51	PICKLE 文件	807 KB
model_gbr.pickle	2024/5/9 21:51	PICKLE 文件	3,831 KB
model_lasso.pickle	2024/5/9 21:51	PICKLE 文件	1 KB
model_lgb.pickle	2024/5/9 21:51	PICKLE 文件	355 KB
model_lr.pickle	2024/5/9 21:51	PICKLE 文件	1 KB

5. The ModelACC folder contains the specific values of various evaluation metrics for each model, along with their corresponding rankings.

	R2	RMSE	MSLE	MEDAE	MAE	EVS	10kf_R2	10kf_RMSE	10kf_MSLE	10kf_MED	10kf_MAE	10kf_EVS
<catboost	0.82276	78.14536	0.031755	36.86001	51.91129	0.830186	0.627092	-69.699	0.037092	-40.6082	-52.697	0.705615
ExtraTrees	0.745814	93.58322	0.055581	34.66	63.26741	0.759101	0.614515	-71.0475	0.037893	-47.389	-54.5198	0.727856
RandomForest	0.724325	97.45874	0.064837	50.10462	67.33933	0.741261	0.616655	-73.7459	0.041541	-49.7991	-58.0914	0.709852

6. Evaluate whether the model accuracy in the current iteration meets the requirements. Then, choose to proceed to the next iteration (elimination or addition) until the accuracy can no longer improve or the data can no longer be split.