**Pearson correlation**

|  |  |  |  |
| --- | --- | --- | --- |
| **1** | +0.703 | Posttreatment SAS 90 | Pretreatment SAS 90 |
| **2** | +0.616 | Pretreatment SAS 90 | Pretreatment SAS zero |
| **3** | +0.612 | Posttreatment SAS 90 | Pretreatment SAS zero |
| **4** | +0.540 | BMI | weight |
| **5** | +0.234 | Pretreatment SAS 90 | age |
| **6** | +0.147 | Pretreatment SAS zero | age |
| **7** | +0.113 | BMI | Posttreatment SAS 90 |
| **8** | +0.099 | Posttreatment SAS 90 | age |
| **9** | -0.091 | BMI | age |
| **10** | +0.090 | BMI | Pretreatment SAS 90 |
| **11** | -0.089 | Pretreatment SAS zero | weight |
| **12** | +0.082 | BMI | Pretreatment SAS zero |
| **13** | -0.049 | Pretreatment SAS 90 | weight |
| **14** | +0.024 | age | weight |
| **15** | -0.012 | Posttreatment SAS 90 | weight |

Here is the feature importances presented in a table format:

| **Feature** | **Importance** |
| --- | --- |
| Sex | 0.051424 |
| Weight | 0.167898 |
| BMI | 0.197436 |
| Pretreatment SAS zero | 0.195404 |
| Pretreatment SAS 90 | 0.206976 |
| Posttreatment SAS 90 | 0.180862 |

runfile('G:/Papers\_Under\_Publishing/SAS/Hyper\_Regression====Run===Final.py', wdir='G:/Papers\_Under\_Publishing/SAS')

100%|██████████| 100/100 [05:17<00:00, 3.18s/trial, best loss: 0.03537268164593922]

hyperopt Regression for Posttreatment SAS 90:

Best Parameters: {'n\_estimators': 100, 'max\_depth': 19, 'min\_samples\_split': 6, 'min\_samples\_leaf': 3}

Mean Squared Error: 0.012371330021286235

Mean Absolute Error: 0.08549733108721348

R2-score: 0.7818789985688883

Execution Time: 317.81364393234253

scikit-optimize - forest\_minimize Regression for Posttreatment SAS 90:

Best Parameters: {'n\_estimators': 145, 'max\_depth': 9, 'min\_samples\_split': 4, 'min\_samples\_leaf': 2}

Mean Squared Error: 0.00858062519252673

Mean Absolute Error: 0.07168860175794453

R2-score: 0.8487135532979366

Execution Time: 396.57281470298767

optunity Regression for Posttreatment SAS 90:

Best Parameters: {'n\_estimators': 122, 'max\_depth': 14, 'min\_samples\_split': 13, 'min\_samples\_leaf': 9}

Mean Squared Error: 0.027372781050970275

Mean Absolute Error: 0.12598663195409293

R2-score: 0.5173858910465442

Execution Time: 207.19039607048035

GPyOpt Regression for Posttreatment SAS 90:

Best Parameters: {'n\_estimators': 500, 'max\_depth': 6, 'min\_samples\_split': 2, 'min\_samples\_leaf': 2}

Mean Squared Error: 0.009601220544343729

Mean Absolute Error: 0.07750585456945257

R2-score: 0.8307192649060475

Execution Time: 220.1207865704651

Optuna Regression for Posttreatment SAS 90:

Best Parameters : {'n\_estimators': 100, 'max\_depth': 29, 'min\_samples\_split': 3, 'min\_samples\_leaf': 3}

Mean Squared Error: 0.012371330021286235

Mean Absolute Error: 0.08549733108721348

R2-score: 0.7818789985688883

Execution Time: 226.91435265541077

the hyperparameter optimization results in table format:

| **Optimizer** | **Best Parameters** | **Mean Squared Error** | **Mean Absolute Error** | **R2 Score** | **Execution Time (seconds)** |
| --- | --- | --- | --- | --- | --- |
| hyperopt | n\_estimators: 100, max\_depth: 19, min\_samples\_split: 6, min\_samples\_leaf: 3 | 0.012371330021286235 | 0.08549733108721348 | 0.7818789985688883 | 317.81364393234253 |
| scikit-optimize - forest\_minimize | n\_estimators: 145, max\_depth: 9, min\_samples\_split: 4, min\_samples\_leaf: 2 | 0.00858062519252673 | 0.07168860175794453 | 0.8487135532979366 | 396.57281470298767 |
| optunity | n\_estimators: 122, max\_depth: 14, min\_samples\_split: 13, min\_samples\_leaf: 9 | 0.027372781050970275 | 0.12598663195409293 | 0.5173858910465442 | 207.19039607048035 |
| GPyOpt | n\_estimators: 500, max\_depth: 6, min\_samples\_split: 2, min\_samples\_leaf: 2 | 0.009601220544343729 | 0.07750585456945257 | 0.8307192649060475 | 220.1207865704651 |
| Optuna | n\_estimators: 100, max\_depth: 29, min\_samples\_split: 3, min\_samples\_leaf: 3 | 0.012371330021286235 | 0.08549733108721348 | 0.7818789985688883 | 226.91435265541077 |

Sure! Here's the table with the provided data:

| **Model** | **Adjusted R-Squared** | **R-Squared** | **RMSE** | **Time Taken** |
| --- | --- | --- | --- | --- |
| GammaRegressor | 0.044256 | 0.381577 | 0.110352 | 0.014992 |
| TweedieRegressor | 0.041573 | 0.379841 | 0.110507 | 0.011990 |
| PoissonRegressor | 0.040820 | 0.379354 | 0.110550 | 0.013993 |
| LassoLarsCV | 0.001930 | 0.354190 | 0.112769 | 0.019426 |
| LarsCV | 0.001930 | 0.354190 | 0.112769 | 0.027982 |
| ElasticNetCV | -0.005359 | 0.349474 | 0.113180 | 0.110931 |
| LassoCV | -0.011805 | 0.345302 | 0.113542 | 0.087946 |
| LassoLarsIC | -0.024037 | 0.337388 | 0.114227 | 0.010993 |
| RidgeCV | -0.078959 | 0.301850 | 0.117250 | 0.008995 |
| BayesianRidge | -0.110758 | 0.281274 | 0.118965 | 0.011991 |
| OrthogonalMatchingPursuitCV | -0.111564 | 0.280753 | 0.119008 | 0.021986 |
| BaggingRegressor | -0.168617 | 0.243836 | 0.122024 | 0.026986 |
| Ridge | -0.205619 | 0.219893 | 0.123941 | 0.008994 |
| LinearRegression | -0.227601 | 0.205670 | 0.125066 | 0.010992 |
| Lars | -0.227601 | 0.205670 | 0.125066 | 0.014992 |
| TransformedTargetRegressor | -0.227601 | 0.205670 | 0.125066 | 0.008995 |
| SGDRegressor | -0.294330 | 0.162492 | 0.128420 | 0.010995 |
| RandomForestRegressor | -0.344391 | 0.130100 | 0.130880 | 0.174894 |
| ExtraTreesRegressor | -0.367700 | 0.115018 | 0.132009 | 0.129918 |
| HuberRegressor | -0.447436 | 0.063424 | 0.135803 | 0.022987 |
| OrthogonalMatchingPursuit | -0.453714 | 0.059362 | 0.136097 | 0.010993 |
| PassiveAggressiveRegressor | -0.466074 | 0.051364 | 0.136675 | 0.012993 |

F-value selector: ['age', 'BMI', 'Pretreatment SAS zero', 'Pretreatment SAS 90',

'Posttreatment SAS 90']

=================================================================

Mutual information selector: ['age', 'weight', 'Pretreatment SAS zero', 'Pretreatment SAS 90',

'Posttreatment SAS 90']

=================================================================

RFE with logistic regression: ['age', 'weight', 'BMI', 'Pretreatment SAS zero',

'Posttreatment SAS 90']

=================================================================

Select from model with random forests: ['Pretreatment SAS zero', 'Pretreatment SAS 90', 'Posttreatment SAS 90']

=================================================================

RFE with random forests: ['weight', 'BMI', 'Pretreatment SAS zero', 'Pretreatment SAS 90',

'Posttreatment SAS 90']

=================================================================

Feature importance with random forests: ['Posttreatment SAS 90', 'Pretreatment SAS zero', 'Pretreatment SAS 90', 'BMI', 'weight']

| **Feature Selection Method** | **Selected Features** |
| --- | --- |
| F-value selector | age, BMI, Pretreatment SAS zero, Pretreatment SAS 90, Posttreatment SAS 90 |
| Mutual information selector | age, weight, Pretreatment SAS zero, Pretreatment SAS 90, Posttreatment SAS 90 |
| RFE with logistic regression | age, weight, BMI, Pretreatment SAS zero, Posttreatment SAS 90 |
| Select from model with random forests | Pretreatment SAS zero, Pretreatment SAS 90, Posttreatment SAS 90 |
| RFE with random forests | weight, BMI, Pretreatment SAS zero, Pretreatment SAS 90, Posttreatment SAS 90 |
| Feature importance with random forests | Posttreatment SAS 90, Pretreatment SAS zero, Pretreatment SAS 90, BMI, weight |

correlation Matrix

sex age weight BMI Pretreatment SAS zero Pretreatment SAS 90

sex 1.0 0.15243118406003386 -0.0006031239128078113 -0.0774848933710429 0.08115205705125378 0.0527500758986282

age 0.15243118406003386 1.0 0.03260159417678021 -0.08135313950109901 0.1447268499285029 0.2171021523146969

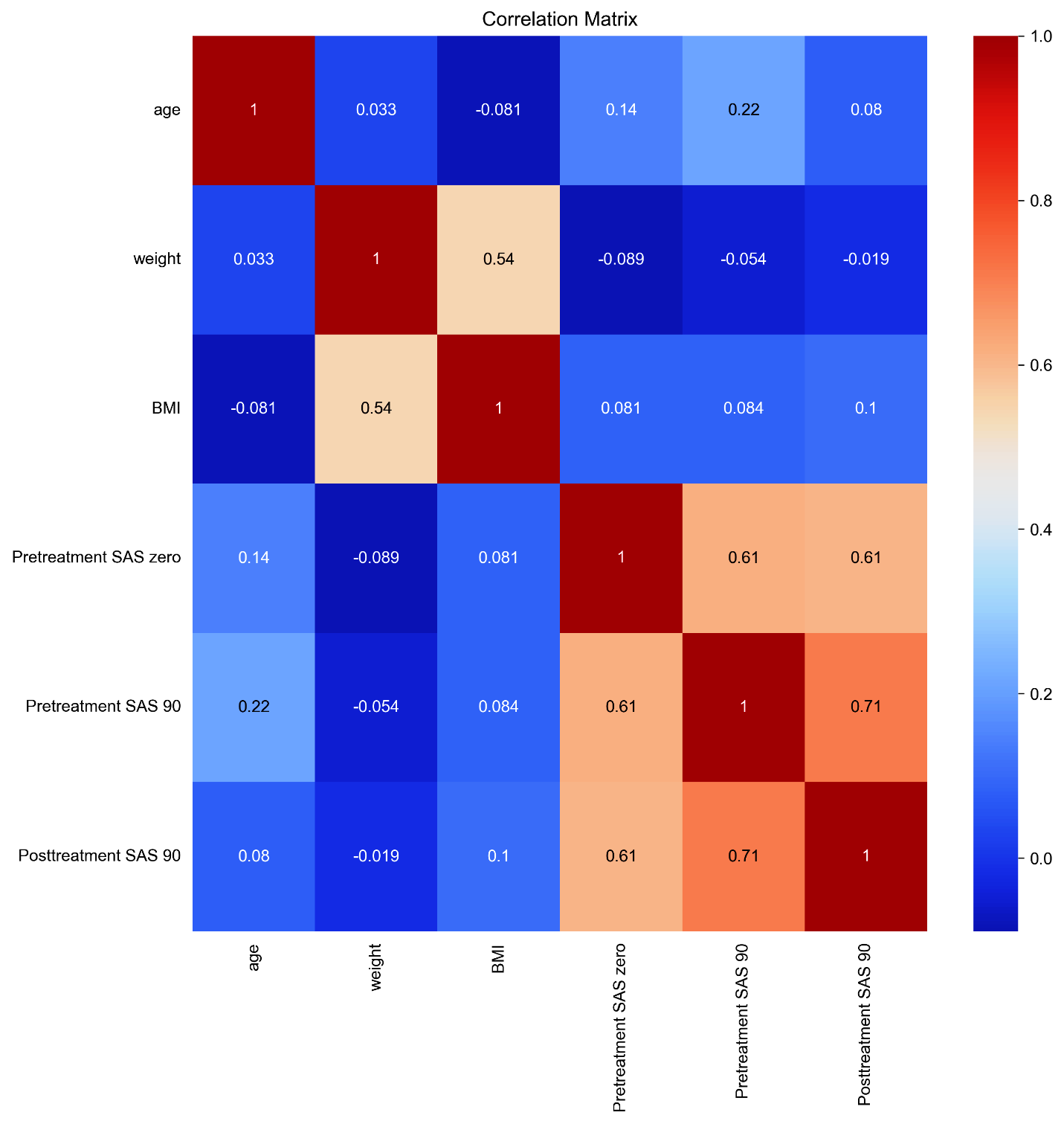
weight -0.0006031239128078113 0.03260159417678021 1.0 0.5417375083027863 -0.0890268697698943 -0.05403493690864452

BMI -0.0774848933710429 -0.08135313950109901 0.5417375083027863 1.0 0.08135272088184767 0.08378296838534889

Pretreatment SAS zero 0.08115205705125378 0.1447268499285029 -0.0890268697698943 0.08135272088184767 1.0 0.613368626885486

Pretreatment SAS 90 0.0527500758986282 0.2171021523146969 -0.05403493690864452 0.08378296838534889 0.613368626885486 1.0

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | sex | age | weight | BMI | Pretreatment SAS zero | Pretreatment SAS 90 |
| sex | 1 | 0.152431 | -0.000603 | -0.077485 | 0.081152 | 0.05275 |
| age | 0.152 | 1 | 0.032602 | -0.081353 | 0.144727 | 0.217102 |
| weight | -0.001 | 0.032602 | 1 | 0.541738 | -0.089027 | -0.054035 |
| BMI | -0.077 | -0.0813 | 0.541738 | 1 | 0.081353 | 0.083783 |
| Pretreatment SAS zero | 0.081152 | 0.144727 | -0.089027 | 0.081353 | 1 | 0.613369 |
| Pretreatment SAS 90 | 0.05275 | 0.217102 | -0.054035 | 0.083783 | 0.613369 | 1 |



Number of samples: 85

Number of features: 6

sex statistics:

mean: 0.45

median: 0.00

std\_dev: 0.50

min: 0.00

25%: 0.00

50%: 0.00

75%: 1.00

max: 1.00

age statistics:

mean: 20.96

median: 20.00

std\_dev: 2.26

min: 18.00

25%: 19.00

50%: 20.00

75%: 23.00

max: 25.00

weight statistics:

mean: 60.69

median: 60.00

std\_dev: 2.96

min: 55.00

25%: 59.00

50%: 60.00

75%: 63.00

max: 68.00

BMI statistics:

mean: 21.36

median: 21.00

std\_dev: 1.36

min: 19.00

25%: 20.10

50%: 21.00

75%: 22.60

max: 25.00

Pretreatment SAS zero statistics:

mean: 1.10

median: 1.06

std\_dev: 0.18

min: 0.72

25%: 0.98

50%: 1.06

75%: 1.20

max: 1.82

Pretreatment SAS 90 statistics:

mean: 1.07

median: 1.07

std\_dev: 0.22

min: 0.63

25%: 0.91

50%: 1.07

75%: 1.19

max: 1.90

| **Feature** | **mean** | **median** | **std\_dev** | **min** | **25%** | **50%** | **75%** | **max** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| sex | 0.45 | 0.00 | 0.50 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 |
| age | 20.96 | 20.00 | 2.26 | 18.00 | 19.00 | 20.00 | 23.00 | 25.00 |
| weight | 60.69 | 60.00 | 2.96 | 55.00 | 59.00 | 60.00 | 63.00 | 68.00 |
| BMI | 21.36 | 21.00 | 1.36 | 19.00 | 20.10 | 21.00 | 22.60 | 25.00 |
| Pretreatment SAS zero | 1.10 | 1.06 | 0.18 | 0.72 | 0.98 | 1.06 | 1.20 | 1.82 |
| Pretreatment SAS 90 | 1.07 | 1.07 | 0.22 | 0.63 | 0.91 | 1.07 | 1.19 | 1.90 |