Random Forest

Code output

Mean Squared Error: 66028825.68724419

Root Mean Squared Error: 8125.812309378318

R-squared: 0.894528730835821

|  |  |
| --- | --- |
| Feature | Importance |
| cat\_\_organ\_leaf | 0.180017 |
| cat\_\_poid\_PFHxA | 0.138394 |
| num\_\_PollutantEmission | 0.130348 |
| num\_\_IndustryValue\_Overall | 0.113913 |
| num\_\_IndustryValue | 0.113317 |
| num\_\_time | 0.103985 |
| cat\_\_poid\_PFDS | 0.0853603 |
| cat\_\_poid\_PFBA | 0.0546484 |
| cat\_\_poid\_PFBS | 0.0436508 |
| num\_\_NumWasteFactories | 0.0288827 |
| cat\_\_organ\_root | 0.00430506 |
| cat\_\_organ\_stem | 0.00317711 |
| num\_\_EnvEmerg | 1.24E-07 |
| cat\_\_organ\_liver | 1.24E-07 |
| cat\_\_poid\_PFOS | 1.14E-07 |
| cat\_\_organ\_egg | 9.99E-08 |
| cat\_\_organ\_ liver | 7.93E-08 |
| num\_\_max | 5.31E-08 |
| cat\_\_organ\_kidney | 3.95E-08 |
| cat\_\_organ\_gill | 2.92E-08 |
| num\_\_WasteGas | 1.89E-08 |
| num\_\_IndustrialEnterprise | 1.60E-08 |
| num\_\_MajorEnergyConsumption | 1.55E-08 |
| num\_\_min | 1.52E-08 |
| cat\_\_organ\_Muscle | 1.46E-08 |
| num\_\_WasteDisposalVol | 1.38E-08 |
| num\_\_RealEstateDev | 1.04E-08 |
| num\_\_NetIndustryValue | 1.03E-08 |
| cat\_\_organ\_stems、leaf | 8.24E-09 |
| cat\_\_organ\_muscle | 7.89E-09 |
| cat\_\_poid\_PFUnDA | 6.67E-09 |
| cat\_\_organ\_head | 3.38E-09 |
| cat\_\_organ\_whole (soft part) | 1.69E-09 |
| cat\_\_poid\_PFOA | 1.62E-09 |
| cat\_\_poid\_PFUdA | 1.49E-09 |
| cat\_\_poid\_PFDA | 1.37E-09 |
| cat\_\_organ\_missing | 1.10E-09 |
| cat\_\_poid\_PFTrDA | 5.65E-10 |
| cat\_\_poid\_PFDoA | 1.72E-10 |
| cat\_\_poid\_PFNA | 1.13E-10 |
| cat\_\_poid\_PFOSA | 2.31E-11 |
| cat\_\_poid\_PFHxS | 2.07E-11 |
| cat\_\_poid\_PFHpA | 1.03E-11 |
| cat\_\_poid\_PFTeDA | 5.78E-12 |
| cat\_\_organ\_whole | 4.04E-12 |
| cat\_\_organ\_gonad1.23 | 4.00E-12 |
| cat\_\_poid\_PFPeA | 1.45E-12 |
| cat\_\_organ\_muscle | 1.21E-13 |

Random Forest Removed Outliers

Code output

Mean Squared Error: 4.796971417875921

Root Mean Squared Error: 2.1901989448166392

R-squared: 0.43109431038535795

|  |  |
| --- | --- |
| Feature | Importance |
| num\_\_min | 0.145732 |
| cat\_\_poid\_PFUnDA | 0.120381 |
| cat\_\_organ\_liver | 0.0795409 |
| num\_\_EnvEmerg | 0.0541489 |
| cat\_\_poid\_PFOS | 0.0383935 |
| cat\_\_poid\_PFDoA | 0.0375885 |
| cat\_\_organ\_gill | 0.0358668 |
| cat\_\_poid\_PFDA | 0.033976 |
| num\_\_max | 0.03311 |
| cat\_\_organ\_muscle | 0.0330737 |
| cat\_\_poid\_PFNA | 0.0310601 |
| cat\_\_poid\_PFTrDA | 0.0299134 |
| cat\_\_organ\_egg | 0.0296857 |
| num\_\_MajorEnergyConsumption | 0.0249644 |
| num\_\_NetIndustryValue | 0.0234564 |
| num\_\_WasteGas | 0.0200088 |
| cat\_\_poid\_PFOA | 0.0199059 |
| cat\_\_organ\_kidney | 0.0155825 |
| num\_\_time | 0.0146027 |
| cat\_\_poid\_PFBA | 0.0144246 |
| num\_\_RealEstateDev | 0.0134357 |
| num\_\_NumWasteFactories | 0.0129943 |
| cat\_\_organ\_stems、leaf | 0.0115279 |
| cat\_\_organ\_Muscle | 0.0113859 |
| num\_\_WasteDisposalVol | 0.00989412 |
| cat\_\_poid\_PFHxA | 0.00979547 |
| num\_\_IndustryValue\_Overall | 0.00937339 |
| cat\_\_poid\_PFHpA | 0.00915617 |
| num\_\_PollutantEmission | 0.00871728 |
| cat\_\_poid\_PFHxS | 0.0084975 |
| num\_\_IndustryValue | 0.00822224 |
| cat\_\_poid\_PFBS | 0.00756973 |
| cat\_\_organ\_whole (soft part) | 0.00756568 |
| num\_\_IndustrialEnterprise | 0.00732019 |
| cat\_\_poid\_PFDS | 0.0065094 |
| cat\_\_organ\_missing | 0.0044245 |
| cat\_\_poid\_PFOSA | 0.00431702 |
| cat\_\_poid\_PFTeDA | 0.00265337 |
| cat\_\_poid\_PFPeA | 0.00255373 |
| cat\_\_organ\_ liver | 0.00253239 |
| cat\_\_organ\_head | 0.0024933 |
| cat\_\_poid\_PFUdA | 0.00151877 |
| cat\_\_organ\_whole | 0.00122235 |
| cat\_\_organ\_leaf | 0.000596411 |
| cat\_\_organ\_muscle | 0.00022417 |
| cat\_\_organ\_gonad1.23 | 8.30E-05 |

Decision Tree Regressor

Code output

Mean Squared Error: 65133196.09480742

Root Mean Squared Error: 8070.51399198387

R-squared: 0.8959593664534032

Logistic Regression

Code output

Accuracy: 0.407514450867052

Classification Report:

precision recall f1-score support

high 0.74 0.21 0.33 123

low 0.35 0.85 0.49 105

medium 0.48 0.22 0.30 118

accuracy 0.41 346

macro avg 0.52 0.43 0.37 346

weighted avg 0.53 0.41 0.37 346

SVM

Code output

Accuracy: 0.407514450867052

Classification Report:

precision recall f1-score support

high 0.83 0.20 0.33 123

low 0.35 0.86 0.49 105

medium 0.46 0.22 0.30 118

accuracy 0.41 346

macro avg 0.55 0.43 0.37 346

weighted avg 0.56 0.41 0.37 346

Gaussian NB

Code output

Accuracy: 0.3901734104046243

Classification Report:

precision recall f1-score support

high 0.50 0.32 0.39 123

low 0.35 0.86 0.49 105

medium 0.75 0.05 0.10 118

accuracy 0.39 346

macro avg 0.53 0.41 0.33 346

weighted avg 0.54 0.39 0.32 346

XGBoost

[0] train-rmse:37476.38242 test-rmse:28196.14039

[50] train-rmse:15003.83451 test-rmse:12694.22823

[100] train-rmse:13969.62139 test-rmse:11955.66950

[108] train-rmse:13939.41493 test-rmse:11966.13093

RMSE: 11934.327940952002

R-squared: 0.8325229711218832

XGBoost No Outliers

[0] train-rmse:1.13810 test-rmse:1.31625

[50] train-rmse:0.94747 test-rmse:1.10266

[100] train-rmse:0.92561 test-rmse:1.06243

[110] train-rmse:0.92358 test-rmse:1.06256

Evaluation Metrics:

RMSE: 1.06

MAE: 0.75

R²: 0.3462

Random Forest Classifier

Average Cross-Validation Accuracy: 0.7164

Test Accuracy: 0.6649

precision recall f1-score support

high 0.77 0.75 0.76 131

low 0.69 0.67 0.68 123

medium 0.53 0.57 0.55 119

accuracy 0.66 373

macro avg 0.67 0.66 0.66 373

weighted avg 0.67 0.66 0.67 373

Random Forest Classifier

Average Cross-Validation Accuracy: 0.8354

Test Accuracy: 0.8365

precision recall f1-score support

high 0.85 0.81 0.83 183

low 0.82 0.86 0.84 190

accuracy 0.84 373

macro avg 0.84 0.84 0.84 373

weighted avg 0.84 0.84 0.84 373

feature importance

0 id 0.435964

44 poid\_PFUnDA 0.044460

2 max 0.034872

38 poid\_PFOS 0.034759

33 poid\_PFHpA 0.026853

34 poid\_PFHxA 0.024665

13 IndustryValue\_Overall 0.024050

30 poid\_PFDA 0.022896

1 time 0.021604

11 NumWasteFactories 0.021017

KNN

Cluster Summary: avg min max mean std mean std mean std cluster 0 0.482585 0.537488 0.190012 0.292815 0.971537 1.048212 1 6.295000 1.426765 2.965000 0.869569 12.180714 5.087170 2 2.466889 0.895996 0.997489 0.643699 5.992889 3.028903 3 10.296000 2.165255 4.800000 1.009950 20.420000 5.675562