Report

ROBERT v 0.0.1 2023/03/28 13:17:28
Citation: ROBERT v 0.0.1, Alegre-Requena, J. V.; Dalmau, D., 2023. https://github.com/jvalegre/robert

Command line used in ROBERT: robert --curate --ignore ['level'] --y EQE --csv_name comparative-analyses-of-data-driven-machine-learning-models-for-tadf-emitters.csv

---

CURATE

o Starting data curation with the CURATE module.

o Database comparative-analyses-of-data-driven-machine-learning-models-for-tadf-emitters.csv loaded successfully, including:
- 200 datapoints
- 8 accepted descriptors
- 1 ignored descriptors
- 0 discarded descriptors

o Analyzing categorical variables
- No categorical variables were found.

o Correction filter activated with these thresholds: thres_x = 0.85, thres_y = 0.02
Excluded descriptors:
- Tp(ns): R**2 = 0.0 with the EQE values
- Td(us): R**2 = 0.0 with the EQE values
- Peak: R**2 = 0.01 with the EQE values

o 6 descriptors remaining after applying correlation filters:
- level
- PLQY
- Von
- CE
- PE
- EQE

o The curated database was stored in C:\Users\juanv\OneDrive\Escritorio\test2\CURATE\comparative-analyses-of-data-driven-machine-learning-models-for-tadf-emitters_CURATE.csv.

Time CURATE: 0.02 seconds

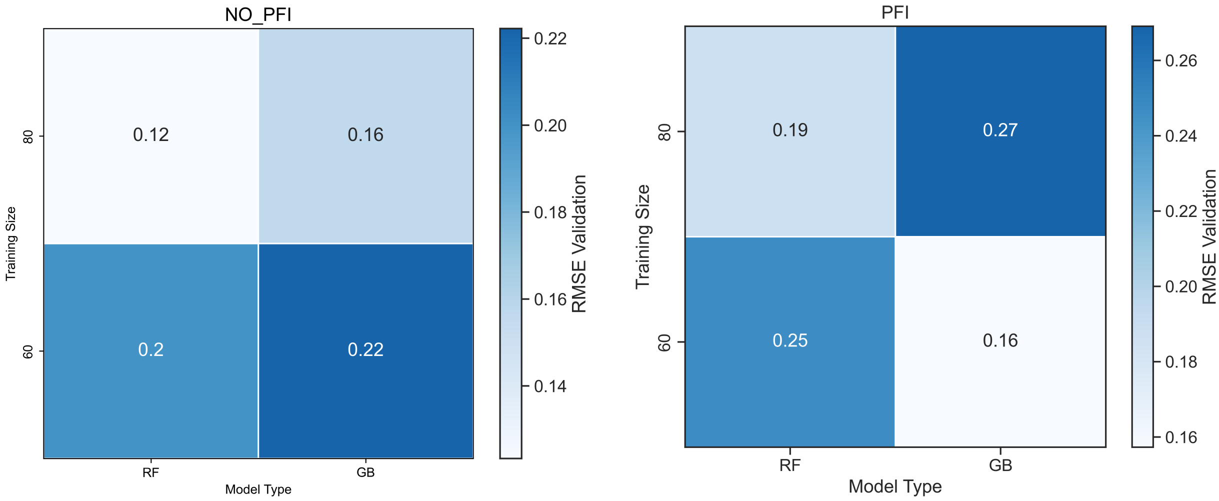
GENERATE

o Starting generation of ML models with the GENERATE module.

o Database CURATE/comparative-analyses-of-data-driven-machine-learning-models-for-tadf-emitters_CURATE.csv loaded successfully, including:

- 200 datapoints
- 5 accepted descriptors
- 1 ignored descriptors
- 0 discarded descriptors

o Starting heatmap scan with 4 ML models ['RF', 'GB', 'NN', 'VR'] and 4 training sizes [60, 70, 80, 90].
Heatmap generation:



PREDICT

o Representation of predictions and analysis of ML models with the PREDICT module

o ML model RF_90.csv (with no PFI filter) and its corresponding Xy database were loaded successfully, including:
- Target value: Target_values
- Model: RF
- Descriptors: ['x5', 'x6', 'x7', 'x8', 'x9', 'x10', 'x11', 'Csub-Csub', 'Csub-H', 'Csub-O', 'H-O']
- Training points: 33
- Validation points: 4

o Test set test.csv loaded successfully, including:
- 4 datapoints

x There are missing descriptors in the test set! Looking for categorical variables converted from CURATE
o The missing descriptors were successfully created
- Train set with predicted results: RF_90_train_No_PFI.csv
- Validation set with predicted results: RF_90_valid_No_PFI.csv
- Test set with predicted results: RF_90_test_No_PFI.csv

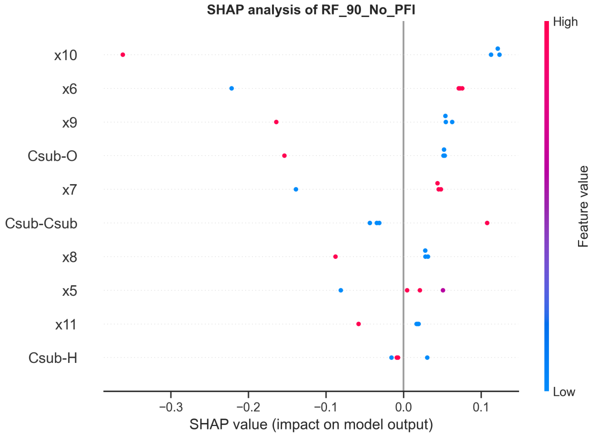
o Saving graphs and CSV databases in C:\Users\juanv\OneDrive\Escritorio\test1\PREDICT:
- Graph in: C:\Users\juanv\OneDrive\Escritorio\test1\PREDICT/Results_RF_90_No_PFI.png

o Results saved in C:\Users\juanv\OneDrive\Escritorio\test1\PREDICT/Results_RF_90_No_PFI.dat:

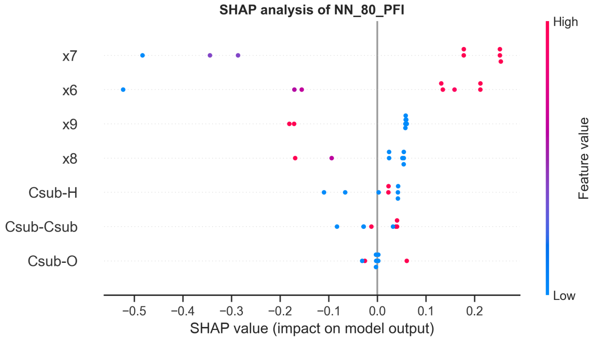
```
      -  Points Train:Validation:Test = 33:4:4
      -  Proportion Train:Validation:Test = 80:10:10
      -  Train : R2 = 0.99, MAE = 0.053, RMSE = 0.078
      -  Validation : R2 = 0.99, MAE = 0.1, RMSE = 0.11
      -  Test : R2 = 1.0, MAE = 0.022, RMSE = 0.024

   o  SHAP plot saved in C:\Users\juanv\OneDrive\Escritorio\test1\PREDICT/SHAP_R
F_90_No_PFI.png
   o  SHAP values saved in C:\Users\juanv\OneDrive\Escritorio\test1\PREDICT/SHAP
_RF_90_No_PFI.dat:
      -  x10 = min: -0.36, max: 0.12
      -  Csub-Csub = min: -0.043, max: 0.11
      -  x6 = min: -0.22, max: 0.076
      -  x9 = min: -0.16, max: 0.063
      -  Csub-O = min: -0.15, max: 0.053
      -  x5 = min: -0.081, max: 0.051
      -  x7 = min: -0.14, max: 0.048
      -  x8 = min: -0.088, max: 0.032
      -  Csub-H = min: -0.015, max: 0.031
      -  x11 = min: -0.058, max: 0.02
      -  H-O = min: 0.0, max: 0.0

   o  PFI plot saved in C:\Users\juanv\OneDrive\Escritorio\test1\PREDICT/PFI_RF_
90_No_PFI.png
   o  PFI values saved in C:\Users\juanv\OneDrive\Escritorio\test1\PREDICT/PFI_R
F_90_No_PFI.dat:
      Original score (from model.score, R2) = 0.98
      -  x10 = 0.11 +- 0.082
      -  x6 = 0.038 +- 0.034
      -  Csub-Csub = 0.028 +- 0.019
      -  x9 = 0.015 +- 0.028
      -  Csub-H = 0.012 +- 0.0076
      -  Csub-O = 0.0074 +- 0.02
      -  x7 = 0.007 +- 0.011
      -  x5 = 0.0066 +- 0.011
      -  x8 = -0.0012 +- 0.0085
      -  x11 = -0.0017 +- 0.01

o  ML model NN_80_PFI.csv (with PFI filter) and its corresponding Xy database we
re loaded successfully, including:
   - Target value: Target_values
   - Model: NN
   - Descriptors: ['x6', 'x7', 'x8', 'x9', 'Csub-Csub', 'Csub-H', 'Csub-O']
   - Training points: 29
   - Validation points: 8

   o  Test set test.csv loaded successfully, including:
      - 4 datapoints

   x  There are missing descriptors in the test set! Looking for categorical var
iables converted from CURATE
   o  The missing descriptors were successfully created
      -  Train set with predicted results: NN_80_train_PFI.csv
      -  Validation set with predicted results: NN_80_valid_PFI.csv
      -  Test set with predicted results: NN_80_test_PFI.csv

   o  Saving graphs and CSV databases in C:\Users\juanv\OneDrive\Escritorio\test
```

```
1\PREDICT:
    -  Graph in: C:\Users\juanv\OneDrive\Escritorio\test1\PREDICT/Results_NN_8
0_PFI.png

  o  Results saved in C:\Users\juanv\OneDrive\Escritorio\test1\PREDICT/Results_
NN_80_PFI.dat:
    -  Points Train:Validation:Test = 29:8:4
    -  Proportion Train:Validation:Test = 71:20:10
    -  Train : R2 = 0.96, MAE = 0.073, RMSE = 0.13
    -  Validation : R2 = 0.98, MAE = 0.1, RMSE = 0.12
    -  Test : R2 = 0.99, MAE = 0.066, RMSE = 0.094

  o  SHAP plot saved in C:\Users\juanv\OneDrive\Escritorio\test1\PREDICT/SHAP_N
N_80_PFI.png
  o  SHAP values saved in C:\Users\juanv\OneDrive\Escritorio\test1\PREDICT/SHAP
_NN_80_PFI.dat:
    -  x7 = min: -0.48, max: 0.25
    -  x6 = min: -0.52, max: 0.21
    -  x9 = min: -0.18, max: 0.061
    -  Csub-O = min: -0.031, max: 0.061
    -  x8 = min: -0.17, max: 0.054
    -  Csub-H = min: -0.11, max: 0.043
    -  Csub-Csub = min: -0.083, max: 0.041

  o  PFI plot saved in C:\Users\juanv\OneDrive\Escritorio\test1\PREDICT/PFI_NN_
80_PFI.png
  o  PFI values saved in C:\Users\juanv\OneDrive\Escritorio\test1\PREDICT/PFI_N
N_80_PFI.dat:
    Original score (from model.score, R2) = 0.97
    -  x7 = 0.34 +- 0.18
    -  x6 = 0.27 +- 0.16
    -  Csub-H = 0.069 +- 0.039
    -  x8 = 0.037 +- 0.027
    -  Csub-Csub = 0.033 +- 0.025
    -  x9 = 0.017 +- 0.017
    -  Csub-O = 0.013 +- 0.01
```

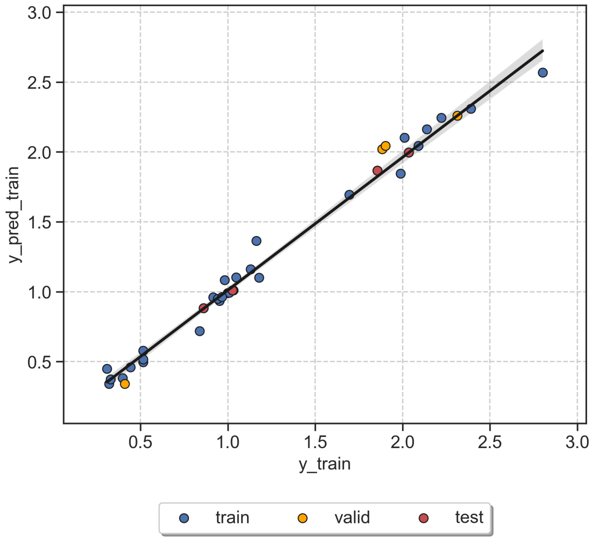
SHAP analysis of RF_90_No_PFI



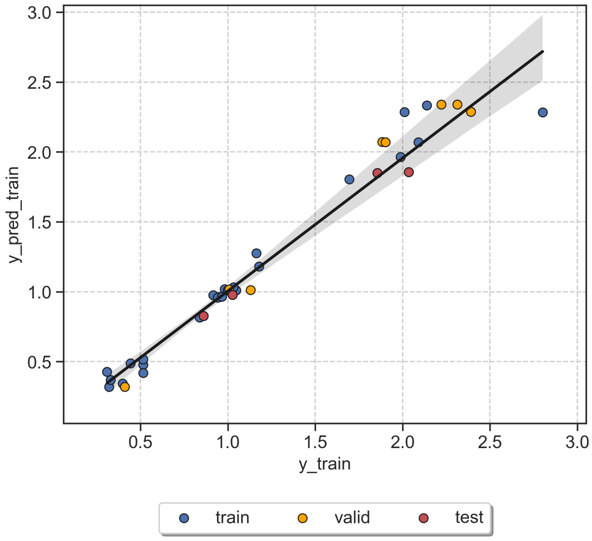
SHAP analysis of NN_80_PFI



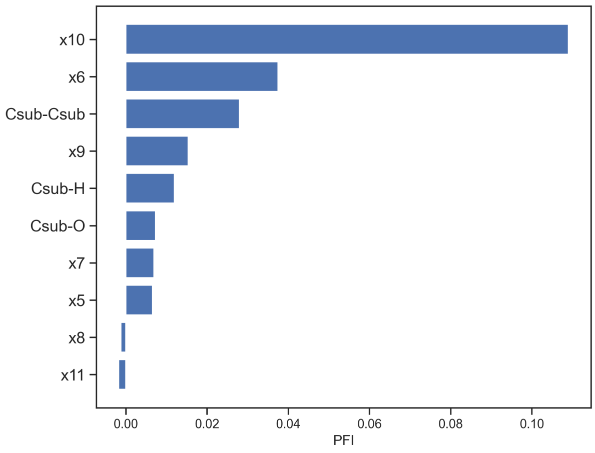
Predictions_train_valid_test of RF_90_No_PFI



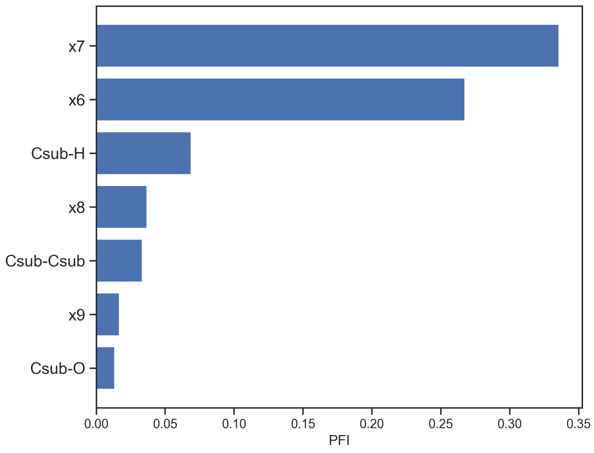
Predictions_train_valid_test of NN_80_PFI



Permutation feature importances (PFIs) of RF_90_No_PFI



Permutation feature importances (PFIs) of NN_80_PFI



VERIFY

o Starting tests to verify the prediction ability of the ML models with the VER
IFY module

o ML model RF_90.csv (with no PFI filter) and its corresponding Xy database wer
e loaded successfully, including:

- Target value: Target_values
- Model: RF
- Descriptors: ['x2', 'x5', 'x6', 'x7', 'x8', 'x9', 'x10', 'x11', 'Csub-Csub', 'Csub-H', 'Csub-O', 'H-O']
- Training points: 33
- Validation points: 4

Results of the verify tests. Original score: RMSE = 0.1, with a +- threshold (thres_test option) of 0.6:

- 5-fold CV: NOT DETERMINED, data splitting was done with k-neighbours (KN). CV result : RMSE = 0.29

x X_shuffle: FAILED, RMSE = 0.15 is lower than the threshold (0.16)
o y_shuffle: PASSED, RMSE = 0.83 is higher than the threshold (0.16)
x onehot: FAILED, RMSE = 0.16 is lower than the threshold (0.16)

o ML model RF_80_PFI.csv (with PFI filter) and its corresponding Xy database were loaded successfully, including:

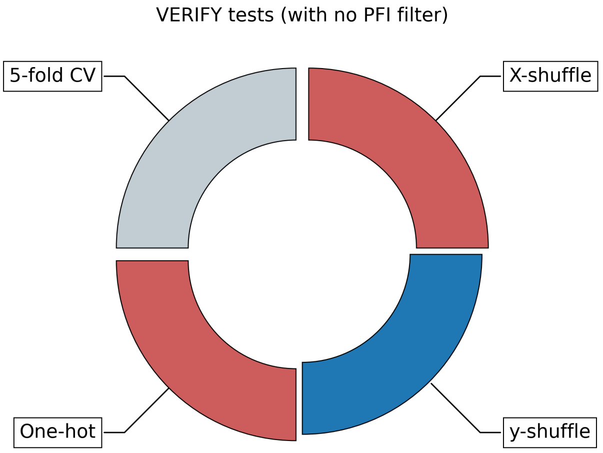
- Target value: Target_values
- Model: RF
- Descriptors: ['x6', 'x7', 'x10']
- Training points: 29
- Validation points: 8

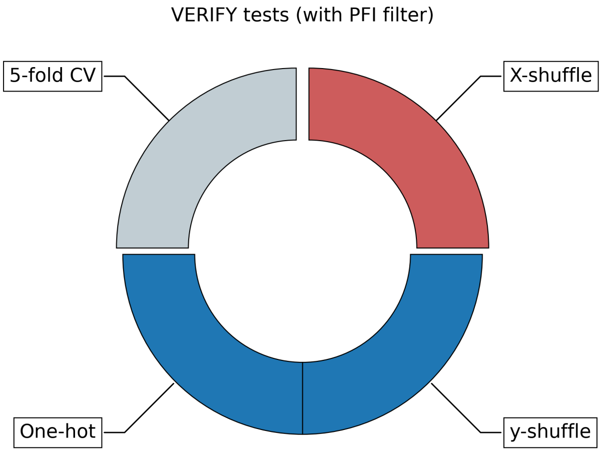
Results of the verify tests. Original score: RMSE = 0.12, with a +- threshold (thres_test option) of 0.6:

- 5-fold CV: NOT DETERMINED, data splitting was done with k-neighbours (KN). CV result : RMSE = 0.22

x X_shuffle: FAILED, RMSE = 0.11 is lower than the threshold (0.19)
o y_shuffle: PASSED, RMSE = 0.95 is higher than the threshold (0.19)
o onehot: PASSED, RMSE = 0.28 is higher than the threshold (0.19)

Time VERIFY: 1.25 seconds





AQME-ROBERT