**RRegrs Flow**

**1)** Load parameters and dataset

- Load Parameters from **CSV** as data frame 🡪 Param.df

- Load dataset from **CSV** as data frame 🡪 ds

**2)** Filters – not implemented

**3)** Remove near zero variance columns using *RemNear0VarCols* 🡪 ds 🡪 **No0Var CSV**

**4)** Scaling dataset (normalization - default, standardization, etc.) using *ScalingDS* 🡪 ds 🡪 **Scaled CSV**

**5)** Remove correlated features using *RemCorrs* 🡪 ds

🡪 Dataset without correlated features: **Scaled NoCorrs CSV**

🡪 Correlation matrix: **Scaled NoCorrs CorrMAT CSV**

🡪 Correlation plot before removal of features: **Scaled NoCorrs Corrs PNG**

**🡪** Correlation plot after removal of features: **Scaled NoCorrs after RemCorr PNG**

For each dataset SPLIT (default = 10) 🡪 dfRes

**6)** Dataset splitting: Training and Test sets using *DsSplit* 🡪 ds.train, ds.test 🡪 **CSVs for train and test**

**7)** Feature selection – not implemented but wrapper functions cold be executed

**8)** REGRESSION METHODS

– executed for each cross-validation type (non-wrapper or wrapper)

– resulted **PDF plots** for each method, split and cross-validation type

– resulted **CSV** for each method with detailed statistics

8.1. Basic LM using *LMreg* 🡪 my.stats.LM appended to dfRes

8.2. GLM based on AIC regression using *GLMreg* 🡪 my.stats.GLM appended to dfRes

8.3. PLS using *PLSreg* 🡪 my.stats.PLS appended to dfRes

8.4. LASSO using *LASSOreg* 🡪 my.stats.LASSO appended to dfRes

8.5. RBF network with the DDA algorithm regression using RBF\_DDAreg 🡪 my.stats.rbfDDA appended to dfRes

8.6. SVM radial regression using *SVLMreg* 🡪 my.stats.SVLM appended to dfRes

8.7. SVM linear – not implemented

8.8. Neural Networks Regression using *NNreg* 🡪 my.stats.NN appended to dfRes

8.9. SOM – to be implemented

8.10. Recursive Feature Extraction (SVM-RFE) – to be implemented

8.11- Other methods to be implemented

**9)** RESULTS

All statistics results (not ordered) 🡪 df.res 🡪 **RRegrsResBySplit.csv**

Averaged statistics of the results by each Regression Method & CV type

All results as data table 🡪 dt.res

Averaged results 🡪 dt.mean

Ordered averaged results 🡪 dt.mean.ord 🡪 **RRegsResAvgs.csv**

**10)** Best model selection – max adjR2.ts (+/- 0.005), min RMSE

Max adjR2.ts 🡪 Best model statistics 🡪 best.dt

adjR2.ts for the best model 🡪 best.adjR2.ts

Add new conditions (max adjR2.ts (+/- 0.005), min RMSE) 🡪 best.dt

Regression method for the best model 🡪 best.reg

**11)** Best model detailed statistics

Detailed statistics for the best model 🡪**RRegrsResBest.csv**

Run the caret function with the method from the best method 🡪 my.stats.reg

**12)** Y randomization for best model – default = 100 times

Using *Yrandom* 🡪 R2Diff.Yrand 🡪 **RRegrsResBest.csv.Yrand.Hist.pdf**