

Input data

Visualization

Descriptors

Model Training

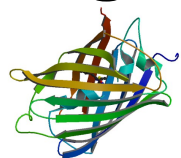
New Molecules

Compounds



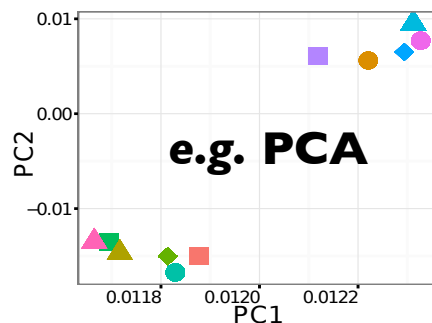
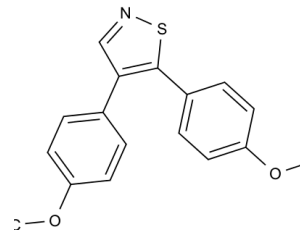
+

Targets



=

Bioactivities / Properties



Compounds (after Standardization)

Morgan fps
- hashed
- unhashed
PaDEL (905 2D)
PaDEL fps (13)

Targets

Amino acid (8)
Sequence (13)

1. Data preprocessing

Feature selection
Centering & scaling
Data partitioning



2. Training

e.g. RF, SVM, GP, ..
Ensemble Modelling



3. Validation

Metrics: R_0^2 , RMSE, ..
Max. model performance

Preprocessing



Predictions

