# Annexe 3.1

All data available (ALL OF/F/S other than a few removed ones due to only one value)

PCA (dimension reduction) FUNCTION ONLY SINCE BENEFITS ARE WRONG CURRENTLY

## WS

### Best models

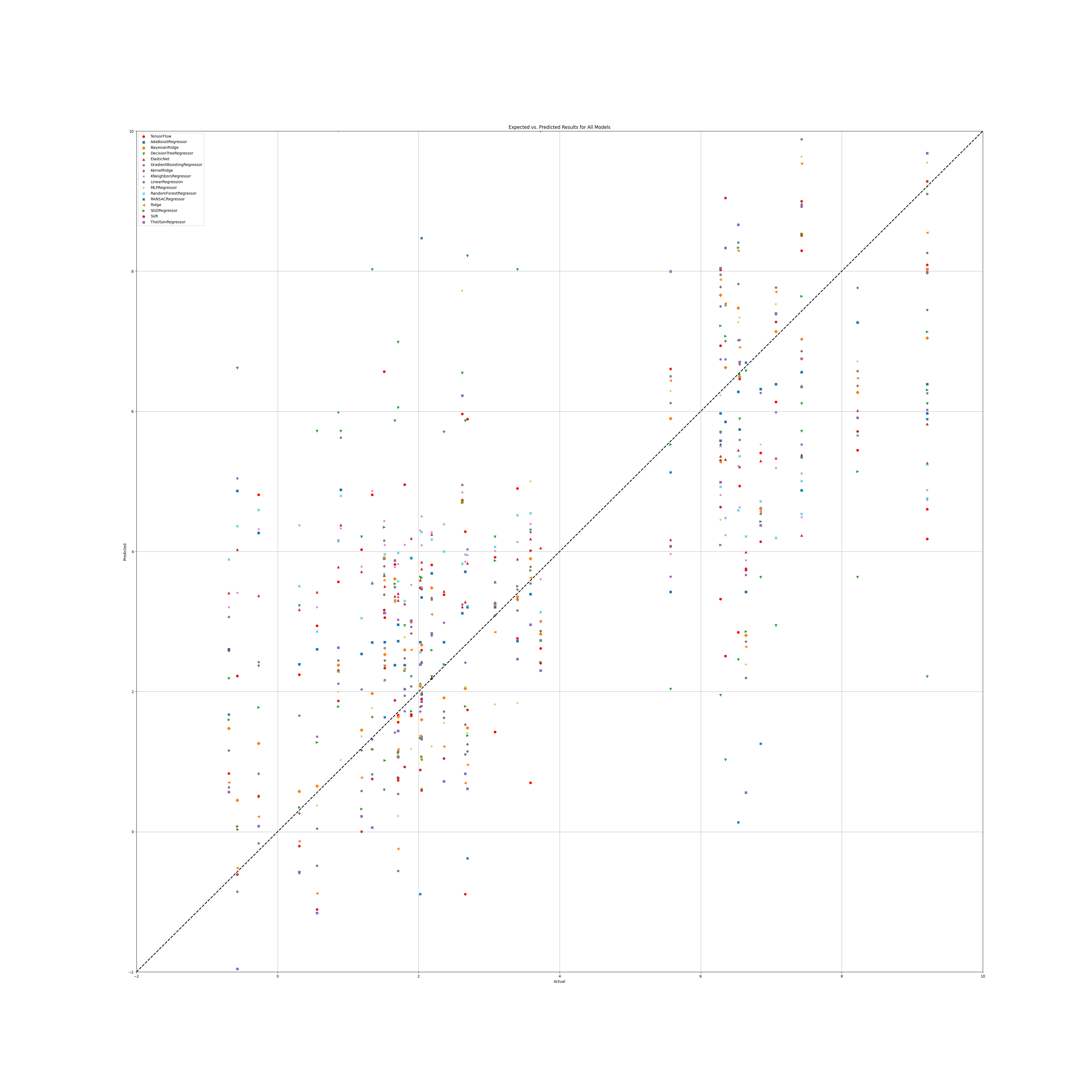
|  |  |  |  |
| --- | --- | --- | --- |
| **Ridge Regression** | **Decision Tree** | **Gradient Boosting** | **Random Forest** |
| alpha: 1.0  solver: 'saga' | criterion: 'squared\_error'  max\_features: 1  min\_samples\_split: 4  splitter: 'random' | learning\_rate: 0.1  loss: 'squared\_error'  n\_estimators: 250  warm\_start: True | criterion: 'friedman\_mse'  max\_features: 'sqrt'  min\_samples\_split: 2  n\_estimators: 50 |
| **AdaBoost** | **K-Nearest** | **MLP Regressor** | **Elastic Net** |
| learning\_rate: 1.0  loss: 'linear'  n\_estimators: 100 | algorithm: 'brute'  leaf\_size: 5  metric: 'cosine'  n\_neighbors: 25  weights: 'distance' | activation: 'logistic'  hidden\_layer\_sizes: (50, 50, 50)  learning\_rate: 'constant'  solver: 'lbfgs' | copy\_X: True  fit\_intercept: True  l1\_ratio: 0.25  positive: False  precompute: False  selection: 'random'  warm\_start: False |
| **SGD Regressor** | **Support Vector** | **Bayesian Ridge** | **Kernel Ridge** |
| learning\_rate: 'adaptive'  loss: 'huber'  penalty: 'elasticnet'  warm\_start: True | degree: 1  gamma: 'scale'  kernel: 'linear'  shrinking: False | alpha\_1: 1e-07  alpha\_2: 1e-05  lambda\_1: 1e-05  lambda\_2: 1e-07 | alpha: 0.1  coef0: 1.0  degree: 1  kernel: 'poly' |
| **Linear Regression** | **RANSAC** | **TheilSen** | **Tensorflow** |
| copy\_X: True  fit\_intercept: True  positive: False | loss: 'absolute\_error'  max\_trials: 100  min\_samples: 5 | max\_subpopulation: 1000  n\_subsamples: None |  |

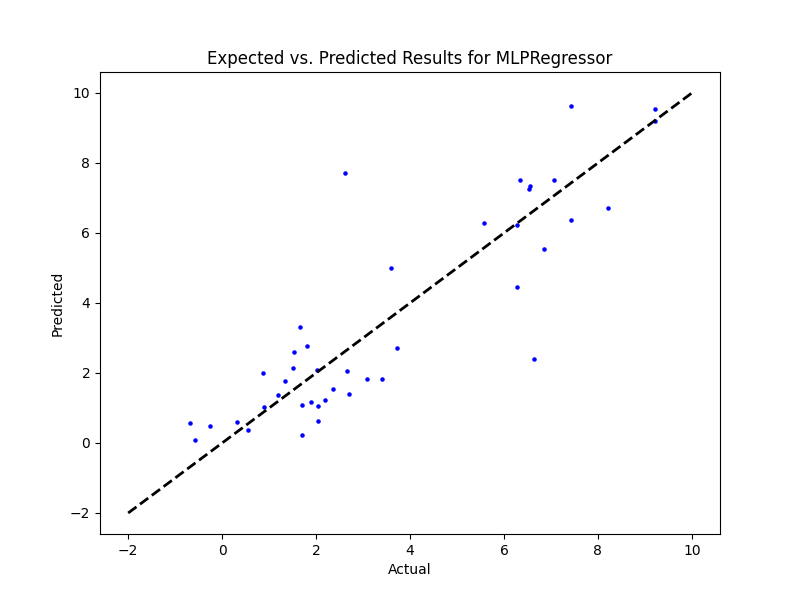
### Prediction Results

Model: Ridge  
RMSE: 4.463533162857565  
  
Model: DecisionTreeRegressor  
RMSE: 3.769559184656421  
  
Model: GradientBoostingRegressor  
RMSE: 1.8289168913755642  
  
Model: RandomForestRegressor  
RMSE: 2.4699260905896736  
  
Model: AdaBoostRegressor

RMSE: 1.9347288999067873  
  
Model: KNeighborsRegressor  
RMSE: 2.5676767810680814  
  
Model: MLPRegressor  
RMSE: 1.4401968251265855  
  
Model: ElasticNet  
RMSE: 2.2196443573402043  
  
Model: SGDRegressor  
RMSE: 2.880370013194233  
  
Model: SVR  
RMSE: 5.973673448454882  
  
Model: BayesianRidge  
RMSE: 3.271775032311624  
  
Model: KernelRidge  
RMSE: 3.432849644633296  
  
Model: LinearRegression  
RMSE: 302.25841010313803  
  
Model: RANSACRegressor  
RMSE: 35.5484694937987  
  
Model: TheilSenRegressor  
RMSE: 227.11283428149946  
  
Model: TensorFlow  
RMSE: 10.981475361698914

### Graphs





## WS Benefit

|  |  |  |  |
| --- | --- | --- | --- |
| **Ridge Regression** | **Decision Tree** | **Gradient Boosting** | **Random Forest** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
| **SGD Regressor** | **Support Vector** | **Bayesian Ridge** | **Kernel Ridge** |
|  |  |  |  |
| **Linear Regression** | **RANSAC** | **TheilSen** | **Tensorflow** |
|  |  |  |  |

## Prediction Results

## Graphs

## NR

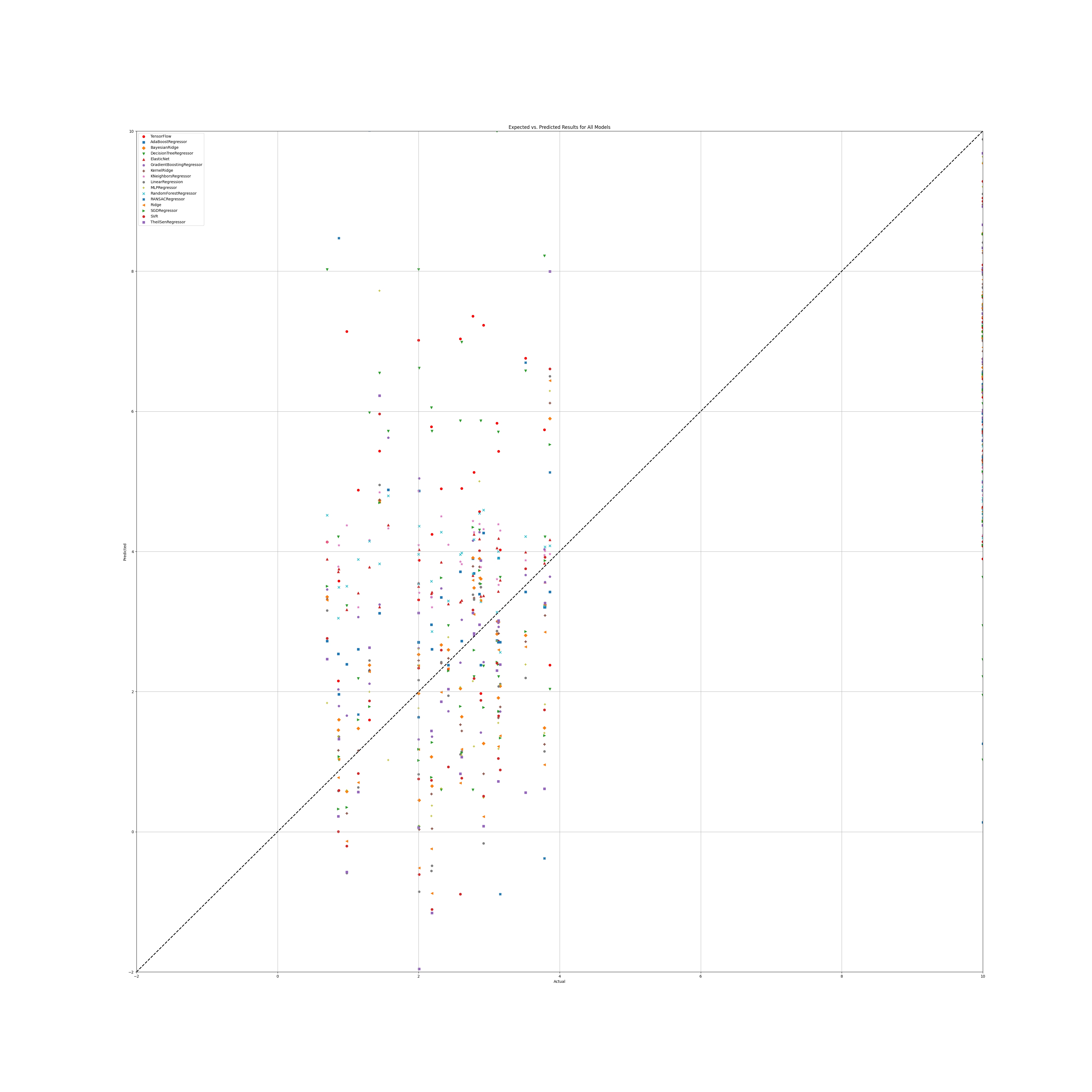
|  |  |  |  |
| --- | --- | --- | --- |
| **Ridge Regression** | **Decision Tree** | **Gradient Boosting** | **Random Forest** |
| alpha: 1.0  solver: 'saga' | criterion: 'squared\_error'  max\_features: 'log2'  min\_samples\_split: 2  splitter: 'best' | learning\_rate: 0.1  loss: 'squared\_error'  n\_estimators: 100  warm\_start: False | criterion: 'absolute\_error'  max\_features: 'sqrt'  min\_samples\_split: 2  n\_estimators: 50 |
| **AdaBoost** | **K-Nearest** | **MLP Regressor** | **Elastic Net** |
| learning\_rate: 1.0  loss: 'square'  n\_estimators: 50 | algorithm: 'brute'  leaf\_size: 5  metric: 'cosine'  n\_neighbors: 5  weights: 'distance' | activation: 'logistic'  hidden\_layer\_sizes: (100, 100, 100, 100)  learning\_rate: 'adaptive'  solver: 'lbfgs' | copy\_X: False  fit\_intercept: True  l1\_ratio: 0.25  positive: False  precompute: False  selection: 'random'  warm\_start: True |
| **SGD Regressor** | **Support Vector** | **Bayesian Ridge** | **Kernel Ridge** |
| learning\_rate: 'invscaling'  loss: 'squared\_error'  penalty: 'l1'  warm\_start: False | degree: 1  gamma: 'scale'  kernel: 'linear'  shrinking: True | alpha\_1: 1e-07  alpha\_2: 1e-05  lambda\_1: 1e-05  lambda\_2: 1e-07 | alpha: 0.1  coef0: 1.0  degree: 1  kernel: 'poly' |
| **Linear Regression** | **RANSAC** | **TheilSen** | **Tensorflow** |
| copy\_X: True  fit\_intercept: True  positive: False | loss: 'squared\_error'  max\_trials: 1  min\_samples: 50 | max\_subpopulation: 1000  n\_subsamples: None |  |

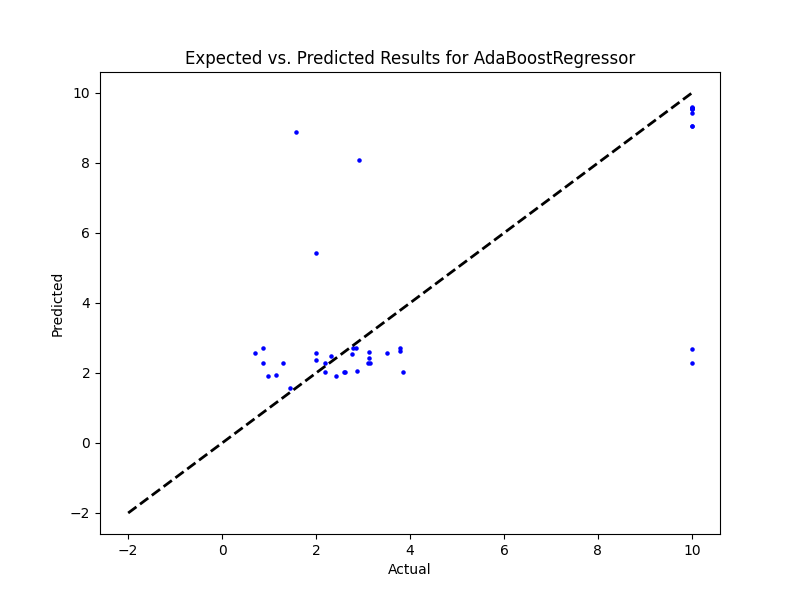
## Prediction results

Model: Ridge  
RMSE: 4.502786160345896  
  
Model: DecisionTreeRegressor  
RMSE: 5.572780414390319  
  
Model: GradientBoostingRegressor  
RMSE: 1.7864838431220031  
  
Model: RandomForestRegressor  
RMSE: 2.6335881744585308  
  
Model: AdaBoostRegressor  
RMSE: 2.341978272474304  
  
Model: KNeighborsRegressor

RMSE: 3.004760053671302  
  
Model: MLPRegressor  
RMSE: 1.769447869975285  
  
Model: ElasticNet  
RMSE: 2.4341424924551833  
  
Model: SGDRegressor  
RMSE: 4.4291122142737605  
  
Model: SVR  
RMSE: 6.198189104831161  
  
Model: BayesianRidge  
RMSE: 4.439034389572746  
  
Model: KernelRidge  
RMSE: 4.441067200757756  
  
Model: LinearRegression  
RMSE: 235.53194201784382  
  
Model: RANSACRegressor  
RMSE: 5.998286518312221  
Model: TheilSenRegressor  
RMSE: 183.01128895235414  
  
Model: TensorFlow  
RMSE: 14.08778377725483

## Graph





## NR Benefit

|  |  |  |  |
| --- | --- | --- | --- |
| **Ridge Regression** | **Decision Tree** | **Gradient Boosting** | **Random Forest** |
|  |  |  |  |
| **AdaBoost** | **K-Nearest** | **MLP Regressor** | **Elastic Net** |
|  |  |  |  |
| **SGD Regressor** | **Support Vector** | **Bayesian Ridge** | **Kernel Ridge** |
|  |  |  |  |
| **Linear Regression** | **RANSAC** | **TheilSen** | **Tensorflow** |
|  |  |  |  |

## Prediction results

## Graph

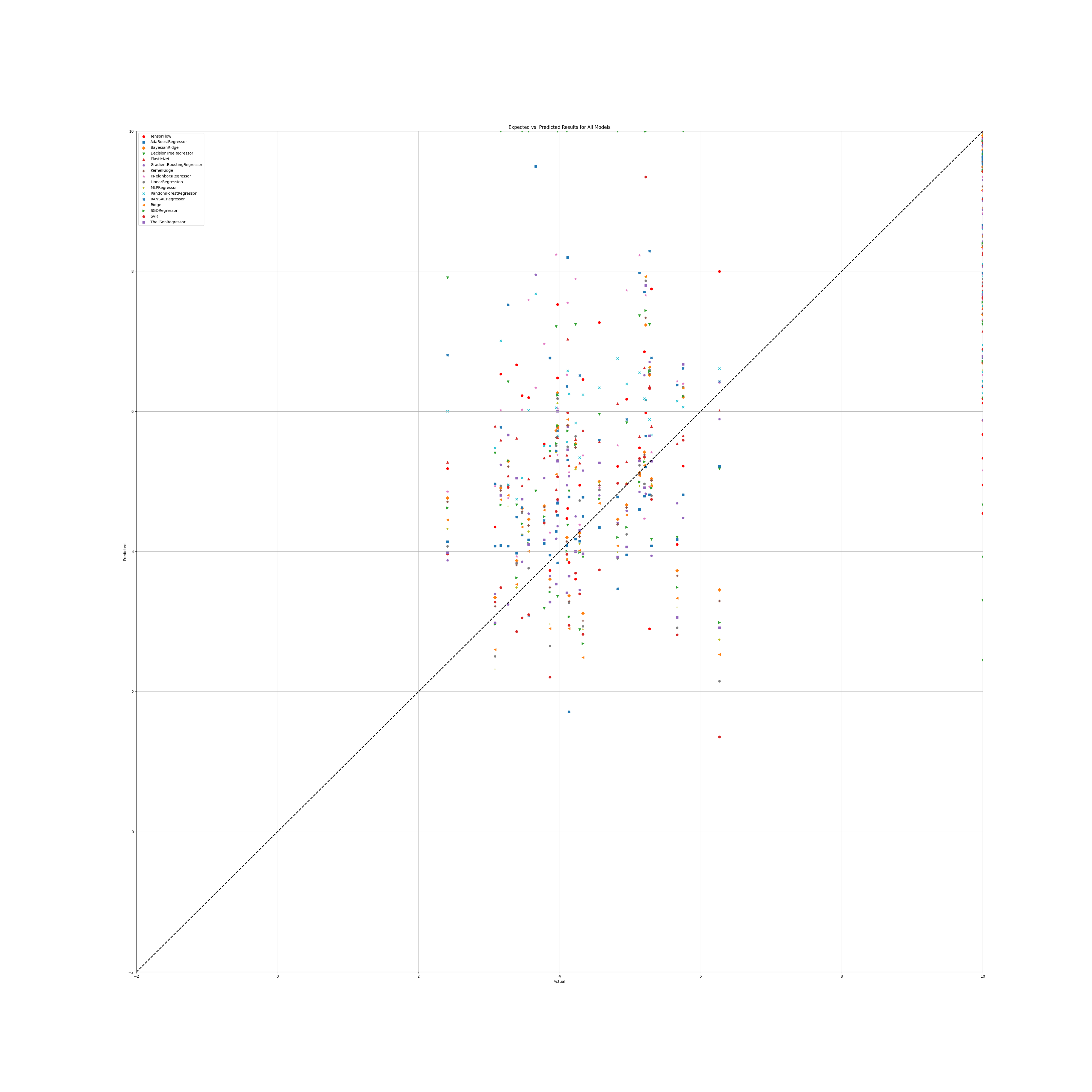
## PR

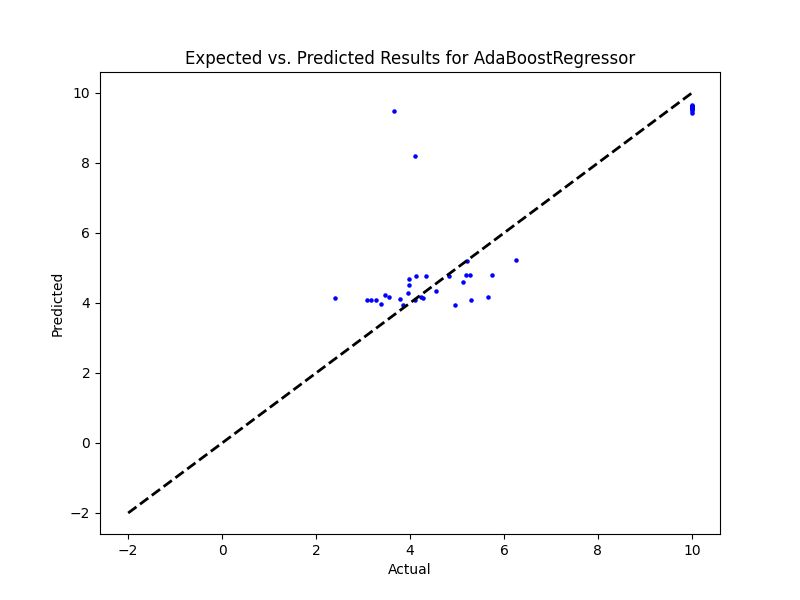
|  |  |  |  |
| --- | --- | --- | --- |
| **Ridge Regression** | **Decision Tree** | **Gradient Boosting** | **Random Forest** |
| alpha: 1.0  solver: 'saga' | criterion: 'poisson'  max\_features: 1  min\_samples\_split: 4  splitter: 'best' | learning\_rate: 0.1  loss: 'squared\_error'  n\_estimators: 250  warm\_start: False | criterion: 'friedman\_mse'  max\_features: 'sqrt'  min\_samples\_split: 2  n\_estimators: 50 |
| **AdaBoost** | **K-Nearest** | **MLP Regressor** | **Elastic Net** |
| learning\_rate: 1.0  loss: 'square'  n\_estimators: 100 | algorithm: 'brute'  leaf\_size: 5  metric: 'cosine'  n\_neighbors: 5  weights: 'distance' | activation: 'identity'  hidden\_layer\_sizes: (50, 50, 50)  learning\_rate: 'constant'  solver: 'sgd' | copy\_X: False  fit\_intercept: True  l1\_ratio: 0.25  positive: False  precompute: False  selection: 'random'  warm\_start: False |
| **SGD Regressor** | **Support Vector** | **Bayesian Ridge** | **Kernel Ridge** |
| learning\_rate: 'invscaling'  loss: 'squared\_error'  penalty: 'l1'  warm\_start: False | degree: 1  gamma: 'scale'  kernel: 'linear'  shrinking: False | alpha\_1: 1e-07  alpha\_2: 1e-05  lambda\_1: 1e-05  lambda\_2: 1e-07 | alpha: 0.1  coef0: 1.0  degree: 1  kernel: 'poly' |
| **Linear Regression** | **RANSAC** | **TheilSen** | **Tensorflow** |
| copy\_X: True  fit\_intercept: True  positive: False | loss: 'squared\_error'  max\_trials: 1  min\_samples: 50 | max\_subpopulation: 1000  n\_subsamples: None |  |

## Prediction Results

Model: Ridge  
RMSE: 6.923037540220166  
  
Model: DecisionTreeRegressor  
RMSE: 3.8697621424969704  
  
Model: GradientBoostingRegressor  
RMSE: 1.321340345396886  
  
Model: RandomForestRegressor  
RMSE: 2.2039303602842413  
  
Model: AdaBoostRegressor  
RMSE: 1.2807909634958359  
  
Model: KNeighborsRegressor  
RMSE: 2.298001014830678  
  
Model: MLPRegressor  
RMSE: 6.117177470889234  
  
Model: ElasticNet  
RMSE: 3.549001453184378  
  
Model: SGDRegressor  
RMSE: 6.391009139931659  
  
Model: SVR  
RMSE: 8.839395178331733  
  
Model: BayesianRidge  
RMSE: 6.3694919478125795  
  
Model: KernelRidge  
RMSE: 6.412290432125347  
  
Model: LinearRegression  
RMSE: 282.72326900379227  
  
Model: RANSACRegressor  
RMSE: 8.270247892935153  
  
Model: TheilSenRegressor  
RMSE: 181.75746324270537  
  
Model: TensorFlow  
RMSE: 15.420255381824443

## Graph





## PR Benefit

|  |  |  |  |
| --- | --- | --- | --- |
| **Ridge Regression** | **Decision Tree** | **Gradient Boosting** | **Random Forest** |
|  |  |  |  |
| **AdaBoost** | **K-Nearest** | **MLP Regressor** | **Elastic Net** |
|  |  |  |  |
| **SGD Regressor** | **Support Vector** | **Bayesian Ridge** | **Kernel Ridge** |
|  |  |  |  |
| **Linear Regression** | **RANSAC** | **TheilSen** | **Tensorflow** |
|  |  |  |  |

## Prediction Results

## Graphs

## SR

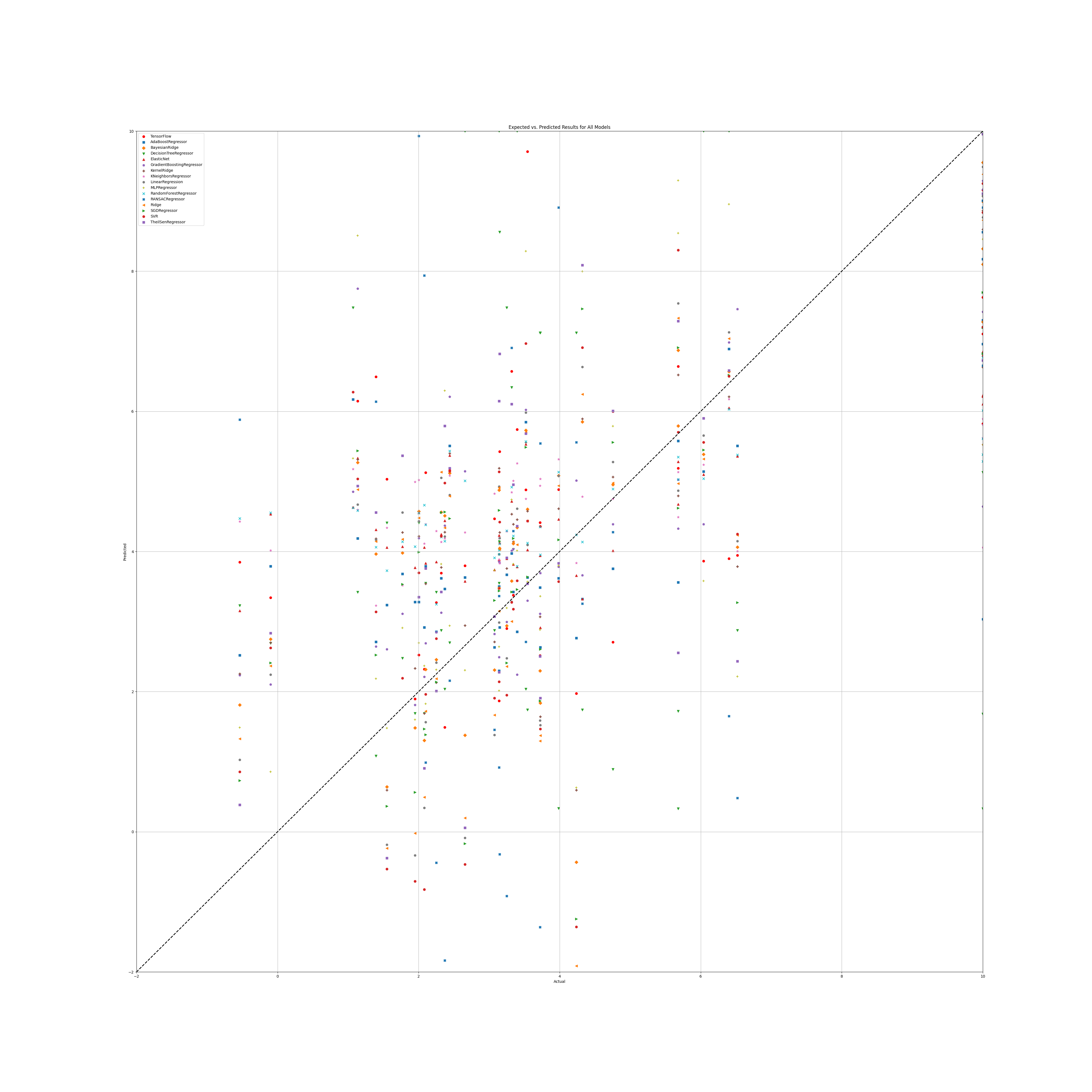
|  |  |  |  |
| --- | --- | --- | --- |
| **Ridge Regression** | **Decision Tree** | **Gradient Boosting** | **Random Forest** |
| alpha: 1.0  solver: 'saga' | criterion: 'absolute\_error'  max\_features: 'sqrt'  min\_samples\_split: 5  splitter: 'random' | learning\_rate: 0.1  loss: 'huber'  n\_estimators: 250  warm\_start: False | criterion: 'friedman\_mse'  max\_features: 'sqrt'  min\_samples\_split: 2  n\_estimators: 100 |
| **AdaBoost** | **K-Nearest** | **MLP Regressor** | **Elastic Net** |
| learning\_rate: 1.0  loss: 'linear'  n\_estimators: 100 | algorithm: 'brute'  leaf\_size: 5  metric: 'cosine'  n\_neighbors: 25  weights: 'distance' | activation: 'logistic'  hidden\_layer\_sizes: (50, 50, 50)  learning\_rate: 'constant'  solver: 'lbfgs' | copy\_X: True  fit\_intercept: True  l1\_ratio: 0.25  positive: False  precompute: True  selection: 'random'  warm\_start: False |
| **SGD Regressor** | **Support Vector** | **Bayesian Ridge** | **Kernel Ridge** |
| learning\_rate: 'constant'  loss: 'huber'  penalty: 'l1'  warm\_start: True | degree: 1  gamma: 'scale'  kernel: 'linear'  shrinking: True | alpha\_1: 1e-07  alpha\_2: 1e-05  lambda\_1: 1e-05  lambda\_2: 1e-07 | alpha: 1e-05  coef0: 1.0  degree: 2  kernel: 'poly' |
| **Linear Regression** | **RANSAC** | **TheilSen** | **Tensorflow** |
| copy\_X: True  fit\_intercept: True  positive: False | loss: 'squared\_error'  max\_trials: 100  min\_samples: 50 | max\_subpopulation: 1000  n\_subsamples: None |  |

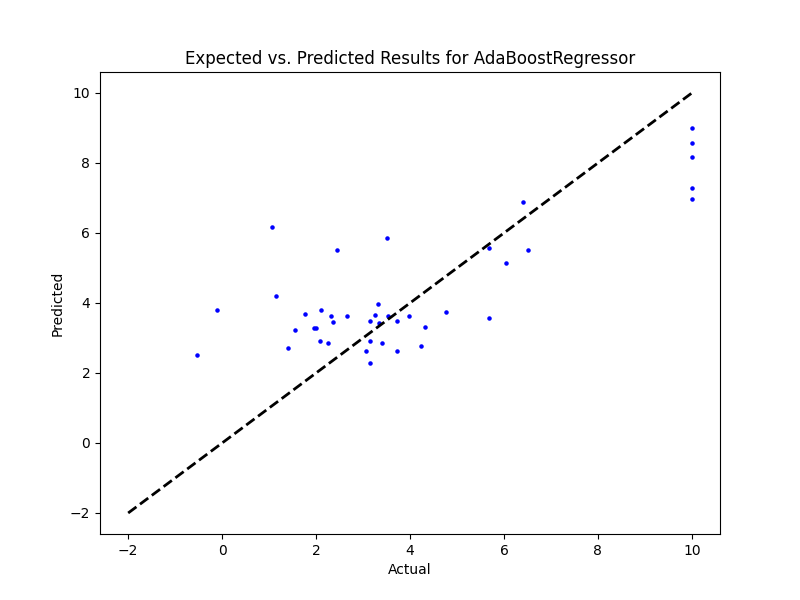
## Prediction Results

Model: Ridge  
RMSE: 2.7416689134378043  
  
Model: DecisionTreeRegressor  
RMSE: 3.8820690426033755  
  
Model: GradientBoostingRegressor  
RMSE: 1.952565262079683  
  
Model: RandomForestRegressor  
RMSE: 2.390240991583788  
  
Model: AdaBoostRegressor  
RMSE: 1.7662717602904512  
  
Model: KNeighborsRegressor  
RMSE: 2.5483244821008255  
  
Model: MLPRegressor  
RMSE: 2.3380229576739384  
  
Model: ElasticNet  
RMSE: 2.1932454571632563  
  
Model: SGDRegressor  
RMSE: 2.863368512482077  
  
Model: SVR  
RMSE: 2.184281599358465  
  
Model: BayesianRidge  
RMSE: 2.5230986637158646  
  
Model: KernelRidge  
RMSE: 12.71682107465145  
  
Model: LinearRegression  
RMSE: 196.9838158074395  
  
Model: RANSACRegressor  
RMSE: 749.9827284514041  
  
Model: TheilSenRegressor  
RMSE: 86.80863812783315

Model: TensorFlow  
RMSE: 18.266800915905105

## Graph





## SR Benefit

|  |  |  |  |
| --- | --- | --- | --- |
| **Ridge Regression** | **Decision Tree** | **Gradient Boosting** | **Random Forest** |
|  |  |  |  |
| **AdaBoost** | **K-Nearest** | **MLP Regressor** | **Elastic Net** |
|  |  |  |  |
| **SGD Regressor** | **Support Vector** | **Bayesian Ridge** | **Kernel Ridge** |
|  |  |  |  |
| **Linear Regression** | **RANSAC** | **TheilSen** | **Tensorflow** |
|  |  |  |  |

## Prediction Results

## Graph