

SET №1: rat_oral_LD50

The number of compounds in the dataset is 9843.

Reference statistical parameters: $Q^2_{cv}=0,59$; RMSE=0,59

Descriptors	Method	5-fold CV	
		Q^2_{cv}	RMSE
MF	GBR	0.55	0.60
MF	SVM	0.53	0.62
MF	MLP	0.52	0.62
MF	kNN	0.41	0.69
MF	catboost	0.52	0.62
MACCS	GBR	0.59	0.57
MACCS	SVM	0.58	0.58
MACCS	MLP	0.51	0.63
MACCS	kNN	0.52	0.62
MACCS	catboost	0.6	0.57
PubChem	GBR	0.6	0.57
PubChem	SVM	0.57	0.59
PubChem	MLP	0.53	0.61
PubChem	kNN	0.48	0.65
PubChem	catboost	0.6	0.57
KlekotaRoth	GBR	0,54	0,61
KlekotaRoth	SVM	0,53	0,62
KlekotaRoth	MLP	0.49	0.64
KlekotaRoth	kNN	0.53	0.62
KlekotaRoth	catboost	0.55	0.60
RDKit	GBR	0.6	0.56
RDKit	SVM	0.59	0.56
RDKit	MLP	0.49	0.64
RDKit	kNN	0.49	0.64
RDKit	catboost	0.59	0.57

SET №2: rat_intraperitoneal_LD50

The number of compounds in the dataset is 4798

Reference statistical parameters: $Q^2_{cv}=0,53$; RMSE=0,62

Descriptors	Method	5-fold CV	
		Q^2_{cv}	RMSE
MF	GBR	0.46	0.63
MF	SVM	0.39	0.66
MF	MLP	0.4	0.66
MF	kNN	0.18	0.77
MF	catboost	0.47	0.62
MACCS	GBR	0.49	0.61
MACCS	SVM	0.47	0.62
MACCS	MLP	0.38	0.67
MACCS	kNN	0.37	0.67
MACCS	catboost	0.5	0.6
PubChem	GBR	0.5	0.6
PubChem	SVM	0.47	0.62
PubChem	MLP	0.41	0.65
PubChem	kNN	0.34	0.69
PubChem	catboost	0.5	0.6
KlekotaRoth	GBR	0.42	0.65
KlekotaRoth	SVM	0.41	0.65
KlekotaRoth	MLP	0.35	0.69
KlekotaRoth	kNN	0.3	0.71
KlekotaRoth	catboost	0.43	0.64
RDKit	GBR	0.5	0.6
RDKit	SVM	0.46	0.63
RDKit	MLP	0.36	0.68
RDKit	kNN	0.37	0.68
RDKit	catboost	0.51	0.6

SET №3: rat_intravenous_LD50

The number of compounds in the dataset is 2323

Reference statistical parameters: $Q^2_{cv}=0,64$; RMSE=0,64

Descriptors	Method	5-fold CV	
		Q^2_{cv}	RMSE
MF	GBR	0.6	0.62
MF	SVM	0.59	0.63
MF	MLP	0.52	0.68
MF	kNN	0.49	0.7
MF	catboost	0.61	0.62
MACCS	GBR	0.57	0.64
MACCS	SVM	0.51	0.69
MACCS	MLP	0.51	0.69
MACCS	kNN	0.13	0.92
MACCS	catboost	0.58	0.64
PubChem	GBR	0.59	0.63
PubChem	SVM	0.56	0.65
PubChem	MLP	0.51	0.69
PubChem	kNN	0.45	0.73
PubChem	catboost	0.59	0.63
KlekotaRoth	GBR	0.53	0.68
KlekotaRoth	SVM	0.51	0.69
KlekotaRoth	MLP	0.46	0.72
KlekotaRoth	kNN	0.43	0.75
KlekotaRoth	catboost	0.54	0.67
RDKit	GBR	0.6	0.62
RDKit	SVM	0.58	0.64
RDKit	MLP	0.51	0.69
RDKit	kNN	0.52	0.68
PubChem	catboost	0.61	0.61

SET №4: rat_skin_LD50

The number of compounds in the dataset is 816

Reference statistical parameters: $Q^2_{cv}=0,33$; RMSE=0,67

Descriptors	Method	5-fold CV	
		Q^2_{cv}	RMSE
MF	GBR	0.29	0.67
	SVM	0.28	0.67
	MLP	0.25	0.69
	kNN	0.18	0.72
	catboost	0.31	0.66
MACCS	GBR	0.34	0.34
	SVM	0.32	0.65
	MLP	0.24	0.69
	kNN	0.24	0.69
	catboost	0.36	0.63
PubChem	GBR	0.32	0.65
	SVM	0.27	0.68
	MLP	0.13	0.74
	kNN	0.22	0.7
	catboost	0.33	0.65
KlekotaRoth	GBR	0.21	0.71
	SVM	0.21	0.71
	MLP	0.07	0.76
	kNN	0.08	0.76
	catboost	0.23	0.7
RDKit	GBR	0.34	0.64
	SVM	0.33	0.65
	MLP	0.09	0.76
	kNN	0.24	0.69
	catboost	0.37	0.63

SET №5: rat_intramuscular_LD50

The number of compounds in the dataset is 274

Reference statistical parameters: $Q^2_{cv}=0.43$; RMSE=0.9

Descriptors	Method	5-fold CV	
		Q^2_{cv}	RMSE
MF	GBR	0.27	0.94
MF	SVM	0.26	0.94
MF	MLP	0.26	0.94
MF	kNN	0.08	1.05
MF	catboost	0.26	0.94
MACCS	GBR	0.29	0.92
MACCS	SVM	0.29	0.92
MACCS	MLP	0.31	0.9
MACCS	kNN	0.27	0.93
MACCS	catboost	0.39	0.85
PubChem	GBR	0.27	0.93
PubChem	SVM	0.26	0.94
PubChem	MLP	0.24	0.95
PubChem	kNN	0.18	0.99
PubChem	catboost	0.32	0.9
KlekotaRoth	GBR	0.28	0.93
KlekotaRoth	SVM	0.25	0.95
KlekotaRoth	MLP	0.26	0.94
KlekotaRoth	kNN	0.06	1.06
KlekotaRoth	catboost	0.31	0.91
RDKit	GBR	0.27	0.93
RDKit	SVM	0.28	0.92
RDKit	MLP	0.11	1.03
RDKit	kNN	0.18	0.99
RDKit	catboost	0.34	0.89

SET №6: rat_subcutaneous_LD50

The number of compounds in the dataset is 1761

Reference statistical parameters: $Q^2_{cv}=0.48$; RMSE=0.76

Descriptors	Method	5-fold CV	
		Q^2_{cv}	RMSE
MF	GBR	0.44	0.76
MF	SVM	0.39	0.79
MF	MLP	0.4	0.79
MF	kNN	0.23	0.89
MF	catboost	0.44	0.76
MACCS	GBR	0.5	0.72
MACCS	SVM	0.5	0.72
MACCS	MLP	0.43	0.77
MACCS	kNN	0.42	0.77
MACCS	catboost	0.51	0.71
PubChem	GBR	0.49	0.72
PubChem	SVM	0.45	0.75
PubChem	MLP	0.4	0.79
PubChem	kNN	0.35	0.82
PubChem	catboost	0.51	0.71
KlekotaRoth	GBR	0.43	0.76
KlekotaRoth	SVM	0.43	0.77
KlekotaRoth	MLP	0.35	0.82
KlekotaRoth	kNN	0.3	0.85
KlekotaRoth	catboost	0.45	0.75
RDKit	GBR	0.47	0.74
RDKit	SVM	0.42	0.77
RDKit	MLP	0.29	0.86
RDKit	kNN	0.32	0.84
RDKit	catboost	0.48	0.73