Session: minf_201200059_sm_dataset_198cpds

Created: Mon October 8 2012, 12:24

Continuous model session

Data set: minf_201200059_sm_dataset_198cpds (198 compounds)

Modeled property: pEC50

Summary of model results

	Trn		Val		Test	
Model	Rsqr	RMSE	Rsqr	RMSE	Rsqr	RMSE
PLS Model	0.79	0.5007	0.2471	0.9785	0.665	0.4613
RBF Model	1	1.629e-06	0.4817	0.8119	0.7152	0.4253
Random Forest Regression Model	0.9613	0.2149	0.6365	0.6799	0.6444	0.4752
GPFixed	0.8983	0.3484	0.3666	0.8975	0.6634	0.4624
GP2DSearch	0.8876	0.3664	0.3505	0.9088	0.6642	0.4618
GPFVS	0.8604	0.4083	0.3818	0.8866	0.6743	0.4548
GPRFVS	0.8804	0.3779	0.2909	0.9496	0.6869	0.4459
GPOPT	0.9258	0.2977	0.7036	0.614	0.6033	0.5019
GPNEST	0.8644	0.4024	0.5535	0.7535	0.6612	0.4638

Parameters used:

Set split:

- Training set size: 70%
- Validation set size: 15%
- Clustering with tanimoto level: 0.7

Descriptor pre-selection:

- Threshold for minimum occurrence: 4%
- Threshold for minimum standard deviation: 0.0005
- Threshold for maximum correlation between descriptors: 0.95

Descriptors remaining after pre-selection: 132

logP, Vx, PositiveCharge, Flex, AromaticRings, ERTLNotPSA, ERTLNoSPtPSA, HBA-lip, HBA-prof, HBD-lip, HBD-prof, ACamideO-nh-nh2, AbasicNH0, AbasicNH1, CH1Aa, CH2Aa, CH2hetero, CH2link, CH2long, CH3Aa, CH3hetero, CamideNH0, NRB, RCamideO-nh0, RbasicNH0, RbasicNH1, aliphOH-t6, aminoethanol0, aminoethanol1, anycarbonyl, aromCl, aromF, arylNHCO, benzylicOH, branchedCnotRing, ch2-lipo-t9, ether, hydroxylation-t8, intraHbond6, ketone-t14, lipovolume, nonring-at, ohccn-t17, p-hetero-or-halo, phenolic-tautomer, ring-join, ring5-nH0, ringOdouble, ringat, sp2-carbons, tert-amine-t11, xccn-t12, nC(sp2), nC(sp3), nOH, nCO, nOS, nX, nNprot, ssCH2, sssCH, dssC, aasC, aaaC, ssssC, ssNH, dsN, aaN, sssN, ssO, sF, nNneutral, N4, NbN, BasicNH0R2AroRings, BasicNH12AroRings, PRX-time1, PRX-time-1, UB, HAS, HAO, AliRingAttachment, C12, C4, C10, C3, C8, C1, C11, C2, N7, N8, N2, O3, O9, O10, AroRingAttachment, HydrophobicGroup, H1a, C5, C21, C22, C23, ed70, ed80, ew10, f004, f007, f015, f244, f245, f301, f407, f440, f443, f444, q017, q039, q040, q137, q139, q155, q192, q257, q300, q358, q453, q457, q458, q481, frg-8, frg-26

Compound IDs for split sets for session minf_201200059_sm_dataset_198cpds

Training (140 compounds)

StarDropID 1, StarDropID 4, StarDropID 5, StarDropID 6, StarDropID 7, StarDropID 8, StarDropID 9, StarDropID 10, StarDropID 11, StarDropID 12, StarDropID 13, StarDropID 14, StarDropID 15, StarDropID 16, StarDropID 17, StarDropID 18, StarDropID 19, StarDropID 20, StarDropID 21, StarDropID 23, StarDropID 26, StarDropID 28, StarDropID 31, StarDropID 33, StarDropID 36, StarDropID 37, StarDropID 38, StarDropID 39, StarDropID 41, StarDropID 42, StarDropID 43, StarDropID 45, StarDropID 46, StarDropID 48, StarDropID 49, StarDropID 50, StarDropID 51, StarDropID 52, StarDropID 53, StarDropID 56, StarDropID 57, StarDropID 58, StarDropID 59, StarDropID 61, StarDropID 65, StarDropID 67, StarDropID 68, StarDropID 69, StarDropID 70, StarDropID 71, StarDropID 72, StarDropID 73, StarDropID 74, StarDropID 75, StarDropID 76, StarDropID 77, StarDropID 78, StarDropID 82, StarDropID 83, StarDropID 84, StarDropID 86, StarDropID 87, StarDropID 88, StarDropID 89, StarDropID 90, StarDropID 92, StarDropID 93, StarDropID 94, StarDropID 95, StarDropID 97, StarDropID 99, StarDropID 100, StarDropID 102, StarDropID 106, StarDropID 107, StarDropID 108, StarDropID 109, StarDropID 111, StarDropID 112, StarDropID 113, StarDropID 116, StarDropID 117, StarDropID 119, StarDropID 120, StarDropID 121, StarDropID 122, StarDropID 123, StarDropID 124, StarDropID 125, StarDropID 126, StarDropID 129, StarDropID 130, StarDropID 131, StarDropID 132, StarDropID 134, StarDropID 135, StarDropID 136, StarDropID 138, StarDropID 139, StarDropID 140, StarDropID 145, StarDropID 146, StarDropID 147, StarDropID 152, StarDropID 155, StarDropID 156, StarDropID 157, StarDropID 158, StarDropID 159, StarDropID 160, StarDropID 161, StarDropID 163, StarDropID 164, StarDropID 165, StarDropID 166, StarDropID 167, StarDropID 168, StarDropID 171,

StarDropID 172, StarDropID 173, StarDropID 175, StarDropID 176, StarDropID 178, StarDropID 179, StarDropID 180, StarDropID 181, StarDropID 182, StarDropID 183, StarDropID 184, StarDropID 185, StarDropID 187, StarDropID 188, StarDropID 189, StarDropID 190, StarDropID 191, StarDropID 193, StarDropID 194, StarDropID 195, StarDropID 196, StarDropID 197

Validation (29 compounds)

StarDropID 2, StarDropID 3, StarDropID 22, StarDropID 40, StarDropID 44, StarDropID 54, StarDropID 62, StarDropID 63, StarDropID 64, StarDropID 66, StarDropID 81, StarDropID 96, StarDropID 98, StarDropID 103, StarDropID 104, StarDropID 115, StarDropID 118, StarDropID 127, StarDropID 133, StarDropID 141, StarDropID 143, StarDropID 150, StarDropID 151, StarDropID 169, StarDropID 170, StarDropID 177, StarDropID 186, StarDropID 192, StarDropID 198

Test (29 compounds)

StarDropID 24, StarDropID 25, StarDropID 27, StarDropID 29, StarDropID 30, StarDropID 32, StarDropID 34, StarDropID 35, StarDropID 47, StarDropID 55, StarDropID 60, StarDropID 79, StarDropID 80, StarDropID 85, StarDropID 91, StarDropID 101, StarDropID 105, StarDropID 110, StarDropID 114, StarDropID 128, StarDropID 137, StarDropID 142, StarDropID 144, StarDropID 148, StarDropID 149, StarDropID 153, StarDropID 154, StarDropID 162, StarDropID 174