

Session: Shen BBB Training Test and External merged, Model: BBB Shen full set

Sun May 12 2013, 18:38

Data set: Shen BBB Training Test and External merged

Modeled property: Class

Modeling technique: Random Forests Classification

Model statistics:

	Number	Kappa	Accuracy
TRN	1288	0.9975	0.9992
VAL	275	0.8106	0.9309
TEST	275	0.8613	0.9455

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: 173

Descriptors used in the model: 173

Model details:

Data was classified into 2 classes

- $Y \leq 0$, class low
- $0 < Y$, class high

Confusion matrices

Training set

		Predicted	
		low	high
Observed	low	246	0
	high	1	1.04e+003

- Sensitivity for low : 1
- Specificity for low : 0.996
- Sensitivity for high : 0.999
- Specificity for high : 1

Validation set

		Predicted	
		low	high
Observed	low	56	18
	high	1	200

- Sensitivity for low : 0.757
- Specificity for low : 0.982
- Sensitivity for high : 0.995
- Specificity for high : 0.917

Test set

		Predicted	
		low	high
Observed	low	66	15

- Sensitivity for low : 0.815
- Specificity for low : 1
- Sensitivity for high : 1
- Specificity for high : 0.928

Descriptor	Importance
logP	1
Vx	0.298776746
NegativeCharge	0.424056888
PositiveCharge	0.0557307266
Flex	0.0278804805
AromaticRings	0.0151133426
ERTLNotPSA	0.228059411
HBA-prof	0.143198118
HBD-lip	0.242090672
HBD-prof	0.101459548
ACamideO-nh-nh2	0.018218752
ACamideO-nh0	0.00714118406
AbasicNH0	0.00475249114
AbasicNH1	0.00481515098
CH0Aa	0.0390533693
CH1Aa	0.0457995571
CH2Aa	0.0347988233
CH2hetero	0.166385382
CH2link	0.0469036549
CH2long	0.00248779124
CH3Aa	0.049481798
CH3hetero	0.000747876649
CamideNH0	0.00641507283
Ester	0.0233416129
HaloC	0.0112071214
Michael-accept	0.00164585235
NRB	0.377373546
RCamideO-nh-nh2	0.00259405305
RCamideO-nh0	0.0414924435
RSR	0.00492820051
RbasicNH0	0.0271044802
aliphOH-t6	0.0296337046
allylic-oxyd-t10	0.0283394456
aminoethanol0	0.0358181149
aminoethanol1	0.47670278
anycarbonyl	0.0769910738
aromCl	0.00146668125
aromF	5.80010919e-008
basic-NH2	-0.00657467078
benzylicOH	-0.00233136886
branchedCnotRing	0.0316530615
carbonate-carbamate	-3.86751271e-005
ch2-lipo-t9	0.00808238704
di-widhrow-cx4	0.00405127509
ertl-33	0.0130423186
ertl-35	0.00401205337
est-lact-latm-carbm-t7	0.0873119012
ether	0.0185558833
hydroxylation-t8	0.00255258335
intraHbond5	0.0857562721
intraHbond6	0.0676344708
ketone-t14	0.0107854744
ketones	0.00334838568
lipovolume	0.0208805222
nonring-at	0.406281412
p-hetero-or-halo	0.00730990386

phenol	0.0679106787
phenolic-tautomer	0.00162174972
pyridine	-0.00379042677
ring-join	0.00648633437
ring5-nH0	0.00150214462
ring5nH	0.0142871998
ringOdouble	0.0424694344
ringat	0.0753040537
sp2-carbons	0.0775187984
spiroC	0.00331422966
t-16-1	-0.000805851188
tert-amine-t11	0.0128640262
urea-thio	0.00160660164
xccn-t12	-0.00247516809
nC(sp2)	0.0737822875
nC(sp3)	0.029921839
nOH	0.128040463
nCO	0.031432189
nOS	0.0128448466
nX	0.00483777327
nNprot	0.000967597356
ssCH2	0.104754724
dsCH	0.00416236371
aaCH	0.0542247407
sssCH	0.0824759007
dssC	0.370512724
aasC	0.0721598938
aaaC	0.0158874802
ssssC	0.0505076498
sNH2	0.00149641139
ssNH	0.0079530878
aaNH	0.000852017605
dsN	0.00166015048
aaN	0.00922037009
sssN	5.80010919e-008
aasN	5.80010919e-008
ssO	0.0273774359
sF	0.0136003299
dS	0.00160322164
aaS	5.80010919e-008
sCl	0.0121791046
nNneutral	0.109751843
NnH	0.130732119
N4	0.00406677555
NbN	0.0227274708
fg5	0.0040322179
CamideNH	0.0085873045
BasicNH0R2AroRings	0.0176028647
BasicNH02AroRings	0.0175136812
BasicNH12AroRings	0.00830801483
PRX-time1	0.00907101203
PRX-time-1	0.056791462
UB	0.0542224832
HAS	0.0393816717
HAT	0.732088566
HAO	0.0181191321
AliRingAttachment	0.0505894944
C12	0.00178048911
C4	0.096890375
C10	0.0203739796
C6	0.0101834917
C3	-2.94768597e-006
C8	-0.00324516976

C1	0.0297951773
C11	0.00250160112
C2	0.0258483998
C26	-0.00327606685
N6	0.00885688141
N7	0.0614051782
N8	0.0178510249
N2	0.0895507485
N1	0.000842769572
BasicGroup	0.0958628356
AcidGroup	0.0771195665
H4	0.228644446
O3	0.0219389033
O11	0.00162994151
O9	0.043321941
O10	0.00971831754
AroRingAttachment	0.21518141
HydrophobicGroup	0.0442085266
C5	0.0762292966
C21	0.0193469711
C22	0.0120240217
C23	0.0401940942
C24	0.0065553491
S3	0.00165896944
ed70	0.000882148335
ed20	0.0223894771
ed40	0.00714292563
ed80	0.0113250725
ew10	0.00802231021
ew100	0.000731411623
f004	0.127320692
f007	0.0409349687
f015	0.0211812519
f244	0.0330165885
f245	0.00726450467
f301	0.190049142
f390	0.0299226753
f407	0.0333130397
f440	0.00076643069
f441	0.0627248064
f443	0.00406490406
q017	0.1872565
q039	0.0408775695
q040	0.097886309
q137	0.334833443
q192	0.124641627
q257	0.977241158
q300	0.0753324106
q453	0.325822115
q457	0.0684669763
q458	0.139933661
q481	0.0683779493
frg-8	0.00774159189
frg-26	0.00174861436