

Session: Gombar_Hall_Suppl_Vdss, Model: log(VDss) RF

Sun May 12 2013, 21:10

Data set: Gombar_Hall_Suppl_Vdss

Modeled property: log(VDss)

Modeling technique: Random Forests Regression

Model statistics:

	Number	Rsqr	RMSE
TRN	399	0.908	0.19
VAL	85	0.6233	0.3776
TEST	85	0.6347	0.3773

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:

Descriptors used:

logP
Vx
MW
NegativeCharge
PositiveCharge
Flex
AromaticRings
OverallCharge
ERTLNotPSA
HBA-lip
HBA-prof
HBD-lip
HBD-prof
ACamideO-nh-nh2
ACamideO-nh0
ASamideO-nh-nh2
AbasicNH0
AbasicNH1
CH0Aa
CH1Aa
CH2Aa
CH2hetero
CH2link
CH3Aa
CH3hetero
CamideNH0
Ester
HaloC
NRB
RCamideO-nh-nh2

RCamideO-nh0
RSR
RbasicNH0
aliphOH-t6
allylic-oxyd-t10
aminoethanol0
aminoethanol1
anycarbonyl
aromCl
aromF
arylNHCO
benzylicOH
branchedCnotRing
ch2-lipo-t9
ertl-33
ertl-35
est-lact-latm-carbm-t7
ether
hydroxylation-t8
intraHbond5
intraHbond6
ketone-t14
ketones
lipovolume
nHindole-like
nonring-at
p-hetero-or-halo
phenol
phenolic-tautomer
pyridine
pyridones
ring-join
ring5-nH0
ring5nH
ringOdouble
ringat
sp2-carbons
t-16-1
tert-amine-t11
xccn-t12
nC(sp2)
nC(sp3)
nOH
nCO
nOS
nX
nNprot
ssCH2
dsCH
aaCH
sssCH
dssC
aasC
aaaC
ssssC
sNH2
ssNH
aaNH
dsN
aaN
sssN
sOH
ssO

sF
sCl
nNneutral
NnH
N4
NbN
fg5
CamideNH
BasicNH0R2AroRings
BasicNH02AroRings
BasicNH12AroRings
PRX-time1
PRX-time-1
UB
HAS
HAO
AliRingAttachment
C12
C4
C10
C6
C3
C8
C1
C11
C2
N6
N7
N8
N2
N1
BasicGroup
AcidGroup
H4
H2
O3
O11
O5
O9
O10
AroRingAttachment
HydrophobicGroup
C5
C21
C22
C23
C24
S3
ed70
ed20
ed40
ed80
ew75
ew10
ew100
f004
f007
f015
f244
f301
f393
f407
f440

f441
f443
f444
q017
q039
q040
q137
q155
q192
q257
q300
q358
q453
q457
q458
frg-8
frg-26