

Session: 3, Model: DT_T_Half_Life

Tue November 6 2012, 11:56

Modeled property: T_half_life

Modeling technique: Decision Tree

Model statistics:

	Number	Kappa	Accuracy
TRN	998	0.7798	0.9148
VAL	244	0.4479	0.7951

Parameters used:

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: 181

Descriptors used in the model: 49

Model details:

Decision Tree technique: Decision Tree converted to ruleset with voting

Data was classified into 2 classes

- $Y \leq 15$, class low
- $15 < Y$, class high

Confusion matrices

Training set

		Predicted	
		low	high
Observed	low	218	64
	high	21	695

- Sensitivity for low : 0.773
- Specificity for low : 0.912
- Sensitivity for high : 0.971
- Specificity for high : 0.916

Validation set

		Predicted	
		low	high
Observed	low	35	28
	high	22	159

- Sensitivity for low : 0.556
- Specificity for low : 0.614
- Sensitivity for high : 0.878
- Specificity for high : 0.85

Descriptors

logP
Vx
MW
PositiveCharge
Flex
ERTLNotPSA
HBA-prof
HBD-lip
ACamideO-nh-nh2
CH1Aa
CH2link

CH3Aa
NRB
anycarbonyl
aromF
lipovolume
nonring-at
phenolic-tautomer
pyridones
ringat
nC(sp3)
nOS
ssCH2
aaCH
sssCH
dssC
aasC
sOH
NbN
PRX-time-1
HAO
C1
C2
N7
N2
BasicGroup
HydrophobicGroup
ed40
f004
f244
f440
q040
q137
q155
q192
q300
q458
q481
frg-26

Ruleset (Voting)

nC(sp3) <= 9 & q040 > 6 & ed40 > 0 & MW > 379.21 & MW <= 494.74 & q300 <= 4 Result: high(Trn: 58/0 Val:11/0 Test:0/0)
Rule: 1

HydrophobicGroup <= 1 & pyridones > 0 Result: high(Trn: 44/0 Val:14/1 Test:0/0) Rule: 2

nC(sp3) > 9 & ACamideO-nh-nh2 <= 1 & aromF > 0 & CH2link <= 7 Result: high(Trn: 41/0 Val:16/0 Test:0/0) Rule: 3

nC(sp3) > 9 & f440 <= 4 & ACamideO-nh-nh2 <= 1 & ACamideO-nh-nh2 > 0 & C2 <= 1 & N7 <= 1 Result: high(Trn: 34/0 Val:4/0 Test:0/0) Rule: 4

ACamideO-nh-nh2 <= 0 & phenolic-tautomer <= 0 & PRX-time-1 > 2 & NRB <= 26 & BasicGroup > 0 Result: high(Trn: 62/1 Val:14/2 Test:0/0) Rule: 5

nC(sp3) <= 9 & pyridones <= 0 & ed40 > 0 & aaCH <= 5 Result: high(Trn: 26/0 Val:10/0 Test:0/0) Rule: 6

ACamideO-nh-nh2 > 1 & C1 <= 5 Result: low(Trn: 22/0 Val:8/5 Test:0/0) Rule: 7

ACamideO-nh-nh2 > 1 & q137 > 20 & f004 <= 2 Result: low(Trn: 21/0 Val:10/5 Test:0/0) Rule: 8

Flex > 0.263158 & NbN > 6 & ssCH2 <= 12 Result: high(Trn: 20/0 Val:3/0 Test:0/0) Rule: 9

nC(sp3) > 9 & aromF <= 0 & Flex <= 0.263158 & HBA-prof <= 5 & CH1Aa <= 0 & nOS > 2 Result: low(Trn: 19/0 Val:5/4 Test:0/0) Rule: 10

nC(sp3) <= 9 & q040 > 6 & HAO <= 5 & HydrophobicGroup > 1 & CH3Aa > 1 & PositiveCharge <= 2 & nonring-at <= 13 Result: low(Trn: 19/0 Val:5/2 Test:0/0) Rule: 11

nC(sp3) > 9 & ACamideO-nh-nh2 <= 0 & phenolic-tautomer > 0 & Flex > 0.193548 Result: low(Trn: 38/1 Val:6/2 Test:0/0) Rule: 12

q137 <= 45 & ringat > 34 Result: high(Trn: 17/0 Val:2/1 Test:0/0) Rule: 13

nC(sp3) <= 9 & HydrophobicGroup > 1 & CH3Aa <= 1 Result: high(Trn: 50/2 Val:8/0 Test:0/0) Rule: 14

q040 > 6 & ed40 <= 0 & logP <= 4.77757 & q155 <= 10 & q192 <= 15 & anycarbonyl > 0 & sssCH <= 0 & dssC <= 0 Result: low(Trn: 11/0 Val:2/0 Test:0/0) Rule: 15

f440 <= 4 & CH1Aa > 0 & CH2link > 6 & q155 > 9 Result: high(Trn: 34/2 Val:6/3 Test:0/0) Rule: 16

PRX-time-1 <= 2 & nOS <= 2 & Flex > 0.222222 & Vx > 419.43 Result: low(Trn: 8/0 Val:1/0 Test:0/0) Rule: 17

lipovolume <= 1 Result: high(Trn: 54/5 Val:14/0 Test:0/0) Rule: 18

nC(sp3) <= 9 & PositiveCharge > 2 Result: high(Trn: 136/14 Val:34/4 Test:0/0) Rule: 19

MW <= 379.21 & lipovolume <= 2 Result: high(Trn: 46/5 Val:12/0 Test:0/0) Rule: 20

nC(sp3) > 9 & f440 > 4 & lipovolume > 1 Result: low(Trn: 43/5 Val:10/2 Test:0/0) Rule: 21

ed40 > 0 & Flex > 0.138889 & q300 > 4 & logP <= 3.71378 & Vx <= 313.84 & logP > 2.99725 Result: low(Trn: 5/0 Val:0/0 Test:0/0) Rule: 22

phenolic-tautomer <= 0 & ringat <= 34 & HBA-prof > 5 Result: high(Trn: 111/16 Val:43/12 Test:0/0) Rule: 23

HBA-prof <= 5 & CH1Aa <= 0 & Flex > 0.222222 & f244 > 0 & Vx <= 419.43 Result: high(Trn: 80/12 Val:17/2 Test:0/0) Rule: 24

Vx > 382.25 & f004 <= 0 Result: low(Trn: 10/1 Val:1/0 Test:0/0) Rule: 25

Vx > 278.91 & HBD-lip <= 0 & C1 > 3 Result: low(Trn: 44/7 Val:10/4 Test:0/0) Rule: 26

PRX-time-1 <= 2 & q137 > 45 Result: low(Trn: 9/1 Val:5/1 Test:0/0) Rule: 27

MW <= 379.21 & q481 > 31 & Flex > 0.166667 Result: low(Trn: 8/1 Val:3/1 Test:0/0) Rule: 28

Flex > 0.138889 & MW <= 379.21 & q481 > 31 Result: low(Trn: 18/4 Val:5/2 Test:0/0) Rule: 29

f440 <= 4 Result: high(Trn: 901/234 Val:221/52 Test:0/0) Rule: 30

ringat <= 34 & q458 > 27 Result: low(Trn: 9/2 Val:4/1 Test:0/0) Rule: 31

aromF <= 0 & ACamideO-nh-nh2 <= 0 & CH1Aa > 0 & CH2link <= 6 & Vx > 303.26 & HBD-lip <= 2 Result: low(Trn: 39/11 Val:13/9 Test:0/0) Rule: 32

nC(sp3) <= 9 & q040 > 6 & HydrophobicGroup > 1 & CH3Aa > 1 & phenolic-tautomer <= 0 Result: low(Trn: 15/4 Val:3/2 Test:0/0) Rule: 33

HAO <= 5 & logP <= 3.71378 & Vx > 313.84 & sOH > 0 Result: low(Trn: 43/15 Val:7/4 Test:0/0) Rule: 34

sssCH <= 0 & N2 <= 0 & frg-26 > 0 Result: low(Trn: 61/22 Val:16/7 Test:0/0) Rule: 35

q040 > 6 & q192 <= 15 & anycarbonyl > 0 & sssCH > 0 & ERTLNotPSA <= 64.43 & f004 <= 2 Result: low(Trn: 43/16 Val:7/3 Test:0/0) Rule: 36

q040 > 6 & nonring-at > 16 Result: low(Trn: 74/28 Val:22/10 Test:0/0) Rule: 37

nC(sp3) > 9 & ACamideO-nh-nh2 <= 0 & BasicGroup <= 0 Result: low(Trn: 81/34 Val:31/13 Test:0/0) Rule: 38

Flex > 0.263158 & aasC > 5 Result: low(Trn: 73/32 Val:21/11 Test:0/0) Rule: 39

C1 > 5 Result: high(Trn: 129/57 Val:37/20 Test:0/0) Rule: 40

q040 > 6 & ed40 <= 0 & logP <= 4.77757 & q155 <= 10 & q192 <= 15 & f004 > 0 Result: low(Trn: 104/46 Val:24/12 Test:0/0) Rule: 41

q040 > 6 & Flex > 0.138889 & aaCH > 5 & MW <= 379.21 Result: low(Trn: 48/22 Val:12/8 Test:0/0) Rule: 42

nC(sp3) > 9 & aromF <= 0 & CH1Aa <= 0 Result: low(Trn: 201/97 Val:54/25 Test:0/0) Rule: 43

Vx > 382.25 & Flex > 0.222222 Result: low(Trn: 118/58 Val:27/16 Test:0/0) Rule: 44

f440 > 4 Result: high(Trn: 97/48 Val:23/11 Test:0/0) Rule: 45

Result: high(Trn: 998/282 Val:244/63 Test:0/0) Rule: 46