Project-3

Decision Tree CARAVAN Data Set

Submitted by

Dr.Shilpa Joshi

Decision tree is a graph to represent choices and their results in form of a tree. The nodes in the graph represent an event or choice and the edges of the graph represent the decision rules or conditions. It is mostly used in Machine Learning and Data Mining applications using R.

Examples of use of decision tress is − predicting an email as spam or not spam, predicting of a tumor is cancerous or predicting a loan as a good or bad credit risk based on the factors in each of these. Generally, a model is created with observed data also called training data. Then a set of validation data is used to verify and improve the model. R has packages which are used to create and visualize decision trees. For new set of predictor variable, we use this model to arrive at a decision on the category (yes/No, spam/not spam) of the data.

To build Decision Tree we will follow , the steps given below

To build your first decision trees, we will proceed as follow:

* Step 1: Import the data: we will use Caravan data which is inbuilt in ISLR library)
* > Library(ISLR)

>View(Caravan)

Showing 1 to 9 of 5,822 entries, 86 total columns

>df\_ca=Caravan ### Store in a Data Frame###

>View(df\_ca)

>head(df\_ca)

head(df\_ca)

MOSTYPE MAANTHUI MGEMOMV MGEMLEEF MOSHOOFD MGODRK MGODPR MGODOV MGODGE

1 33 1 3 2 8 0 5 1 3

2 37 1 2 2 8 1 4 1 4

3 37 1 2 2 8 0 4 2 4

4 9 1 3 3 3 2 3 2 4

5 40 1 4 2 10 1 4 1 4

6 23 1 2 1 5 0 5 0 5

MRELGE MRELSA MRELOV MFALLEEN MFGEKIND MFWEKIND MOPLHOOG MOPLMIDD MOPLLAAG

1 7 0 2 1 2 6 1 2 7

2 6 2 2 0 4 5 0 5 4

3 3 2 4 4 4 2 0 5 4

4 5 2 2 2 3 4 3 4 2

5 7 1 2 2 4 4 5 4 0

6 0 6 3 3 5 2 0 5 4

MBERHOOG MBERZELF MBERBOER MBERMIDD MBERARBG MBERARBO MSKA MSKB1 MSKB2 MSKC

1 1 0 1 2 5 2 1 1 2 6

2 0 0 0 5 0 4 0 2 3 5

3 0 0 0 7 0 2 0 5 0 4

4 4 0 0 3 1 2 3 2 1 4

5 0 5 4 0 0 0 9 0 0 0

6 2 0 0 4 2 2 2 2 2 4

MSKD MHHUUR MHKOOP MAUT1 MAUT2 MAUT0 MZFONDS MZPART MINKM30 MINK3045

1 1 1 8 8 0 1 8 1 0 4

2 0 2 7 7 1 2 6 3 2 0

3 0 7 2 7 0 2 9 0 4 5

4 0 5 4 9 0 0 7 2 1 5

5 0 4 5 6 2 1 5 4 0 0

6 2 9 0 5 3 3 9 0 5 2

MINK4575 MINK7512 MINK123M MINKGEM MKOOPKLA PWAPART PWABEDR PWALAND PPERSAUT

1 5 0 0 4 3 0 0 0 6

2 5 2 0 5 4 2 0 0 0

3 0 0 0 3 4 2 0 0 6

4 3 0 0 4 4 0 0 0 6

5 9 0 0 6 3 0 0 0 0

6 3 0 0 3 3 0 0 0 6

PBESAUT PMOTSCO PVRAAUT PAANHANG PTRACTOR PWERKT PBROM PLEVEN PPERSONG

1 0 0 0 0 0 0 0 0 0

2 0 0 0 0 0 0 0 0 0

3 0 0 0 0 0 0 0 0 0

4 0 0 0 0 0 0 0 0 0

5 0 0 0 0 0 0 0 0 0

6 0 0 0 0 0 0 0 0 0

PGEZONG PWAOREG PBRAND PZEILPL PPLEZIER PFIETS PINBOED PBYSTAND AWAPART

1 0 0 5 0 0 0 0 0 0

2 0 0 2 0 0 0 0 0 2

3 0 0 2 0 0 0 0 0 1

4 0 0 2 0 0 0 0 0 0

5 0 0 6 0 0 0 0 0 0

6 0 0 0 0 0 0 0 0 0

AWABEDR AWALAND APERSAUT ABESAUT AMOTSCO AVRAAUT AAANHANG ATRACTOR AWERKT

1 0 0 1 0 0 0 0 0 0

2 0 0 0 0 0 0 0 0 0

3 0 0 1 0 0 0 0 0 0

4 0 0 1 0 0 0 0 0 0

5 0 0 0 0 0 0 0 0 0

6 0 0 1 0 0 0 0 0 0

ABROM ALEVEN APERSONG AGEZONG AWAOREG ABRAND AZEILPL APLEZIER AFIETS AINBOED

1 0 0 0 0 0 1 0 0 0 0

2 0 0 0 0 0 1 0 0 0 0

3 0 0 0 0 0 1 0 0 0 0

4 0 0 0 0 0 1 0 0 0 0

5 0 0 0 0 0 1 0 0 0 0

6 0 0 0 0 0 0 0 0 0 0

ABYSTAND Purchase

1 0 No

2 0 No

3 0 No

4 0 No

5 0 No

6 0 No

>tail(df\_ca) ### Displays last 6 record

MOSTYPE MAANTHUI MGEMOMV MGEMLEEF MOSHOOFD MGODRK MGODPR MGODOV MGODGE

5817 3 1 2 3 1 0 6 0 3

5818 36 1 1 2 8 0 6 1 2

5819 35 1 4 4 8 1 4 1 4

5820 33 1 3 4 8 0 6 0 3

5821 34 1 3 2 8 0 7 0 2

5822 33 1 3 3 8 0 6 1 2

MRELGE MRELSA MRELOV MFALLEEN MFGEKIND MFWEKIND MOPLHOOG MOPLMIDD

5817 6 0 3 0 7 2 0 7

5818 1 2 6 5 3 2 2 5

5819 6 0 3 2 2 5 0 0

5820 5 1 4 3 3 4 0 1

5821 7 2 0 0 4 5 0 2

5822 7 1 2 1 4 4 1 2

MOPLLAAG MBERHOOG MBERZELF MBERBOER MBERMIDD MBERARBG MBERARBO MSKA MSKB1

5817 2 0 0 0 5 2 3 0 2

5818 2 2 0 0 4 1 3 2 3

5819 9 2 1 1 3 3 2 0 4

5820 8 1 0 0 2 3 5 1 1

5821 7 0 2 0 2 4 2 0 0

5822 6 1 0 1 3 2 4 1 1

MSKB2 MSKC MSKD MHHUUR MHKOOP MAUT1 MAUT2 MAUT0 MZFONDS MZPART MINKM30

5817 3 5 0 0 9 5 2 2 5 4 0

5818 3 3 0 9 0 5 1 3 5 4 4

5819 5 0 0 3 6 6 1 2 6 3 0

5820 1 4 4 7 2 4 0 5 8 1 5

5821 4 5 0 2 7 5 4 0 9 0 0

5822 2 6 1 5 4 5 2 3 6 3 2

MINK3045 MINK4575 MINK7512 MINK123M MINKGEM MKOOPKLA PWAPART PWABEDR

5817 4 4 2 0 5 6 0 0

5818 3 3 0 0 3 3 2 0

5819 9 0 0 0 4 5 0 0

5820 3 1 1 0 3 3 2 0

5821 5 4 0 0 4 6 0 0

5822 5 2 1 0 3 3 1 0

PWALAND PPERSAUT PBESAUT PMOTSCO PVRAAUT PAANHANG PTRACTOR PWERKT PBROM

5817 0 6 0 0 0 0 0 0 0

5818 0 6 0 4 0 0 0 0 0

5819 0 0 0 0 0 1 0 0 3

5820 0 6 0 0 0 0 0 0 0

5821 0 6 0 0 0 0 0 0 0

5822 0 0 0 0 0 0 0 0 0

PLEVEN PPERSONG PGEZONG PWAOREG PBRAND PZEILPL PPLEZIER PFIETS PINBOED

5817 0 0 0 0 0 0 0 0 0

5818 3 0 0 0 3 0 0 0 0

5819 0 0 0 0 5 0 0 0 0

5820 0 0 0 0 4 0 0 0 0

5821 0 0 0 0 0 0 0 0 0

5822 0 0 0 0 0 0 0 0 0

PBYSTAND AWAPART AWABEDR AWALAND APERSAUT ABESAUT AMOTSCO AVRAAUT

5817 0 0 0 0 1 0 0 0

5818 0 1 0 0 1 0 1 0

5819 0 0 0 0 0 0 0 0

5820 0 1 0 0 1 0 0 0

5821 0 0 0 0 1 0 0 0

5822 0 1 0 0 0 0 0 0

AAANHANG ATRACTOR AWERKT ABROM ALEVEN APERSONG AGEZONG AWAOREG ABRAND

5817 0 0 0 0 0 0 0 0 0

5818 0 0 0 0 2 0 0 0 1

5819 1 0 0 1 0 0 0 0 1

5820 0 0 0 0 0 0 0 0 1

5821 0 0 0 0 0 0 0 0 0

5822 0 0 0 0 0 0 0 0 0

AZEILPL APLEZIER AFIETS AINBOED ABYSTAND Purchase

5817 0 0 0 0 0 No

5818 0 0 0 0 0 No

5819 0 0 0 0 0 No

5820 0 0 0 0 0 Yes

5821 0 0 0 0 0 No

5822 0 0 0 0 0 No

We can see that output is not shuffled. This is a big issue! When you will split your data between a train set and test set, you will select **only** the few varivles which are in the top 80 percent of the observations), which means the algorithm will never see the features of rest of the data. This mistake will lead to poor prediction. To avoid this we use shuffle index

>shuffle\_index =sample(1:nrow(df\_ca))

Generate a random list of index from 1 to 5822 (i.e. the maximum number of rows).> >head(shuffle\_index)

[1] 627 1083 1270 5488 418 2474

> df\_ca= df\_ca[shuffle\_index, ]

> head(df\_ca)

MOSTYPE MAANTHUI MGEMOMV MGEMLEEF MOSHOOFD MGODRK MGODPR MGODOV MGODGE

627 33 1 3 3 8 1 4 1 5

1083 36 1 3 3 8 2 4 1 3

1270 3 1 2 4 1 0 7 0 2

5488 23 1 2 2 5 2 4 1 3

418 8 1 3 3 2 3 3 1 4

2474 24 1 4 2 5 0 5 1 4

MRELGE MRELSA MRELOV MFALLEEN MFGEKIND MFWEKIND MOPLHOOG MOPLMIDD

627 7 1 2 1 4 4 2 3

1083 6 1 3 5 0 4 3 3

1270 5 0 4 3 5 2 1 4

5488 3 2 5 5 3 2 2 6

418 6 1 3 0 3 6 3 2

2474 8 1 1 1 2 6 1 3

MOPLLAAG MBERHOOG MBERZELF MBERBOER MBERMIDD MBERARBG MBERARBO MSKA MSKB1

627 4 3 0 2 2 3 1 3 1

1083 3 3 0 2 4 0 2 3 3

1270 4 2 0 4 3 2 0 2 1

5488 2 1 0 0 3 2 4 1 3

418 4 3 0 0 2 4 0 2 0

2474 6 0 0 0 5 1 3 0 1

MSKB2 MSKC MSKD MHHUUR MHKOOP MAUT1 MAUT2 MAUT0 MZFONDS MZPART MINKM30

627 3 3 2 3 6 7 2 1 5 4 3

1083 3 2 0 0 9 6 1 3 5 4 0

1270 4 2 2 3 6 4 2 3 4 5 4

5488 3 4 1 8 1 5 0 4 8 1 6

418 0 4 3 3 6 6 1 3 6 3 0

2474 3 5 1 7 2 8 1 1 9 0 2

MINK3045 MINK4575 MINK7512 MINK123M MINKGEM MKOOPKLA PWAPART PWABEDR

627 2 3 2 0 4 3 0 0

1083 3 5 2 0 6 3 0 0

1270 3 0 2 0 4 6 2 0

5488 2 1 0 1 3 3 2 0

418 0 5 0 4 0 8 1 0

2474 6 2 0 0 3 2 2 0

PWALAND PPERSAUT PBESAUT PMOTSCO PVRAAUT PAANHANG PTRACTOR PWERKT PBROM

627 0 6 0 0 0 0 0 0 0

1083 0 0 0 0 0 0 0 0 0

1270 0 6 0 0 0 0 0 0 0

5488 0 5 0 0 0 0 0 0 0

418 0 0 0 0 0 0 0 0 0

2474 0 0 0 0 0 0 0 0 0

PLEVEN PPERSONG PGEZONG PWAOREG PBRAND PZEILPL PPLEZIER PFIETS PINBOED

627 0 0 0 0 0 0 4 0 0

1083 0 0 0 0 0 0 0 0 0

1270 0 0 0 0 4 0 0 0 0

5488 0 0 0 0 0 0 0 0 0

418 0 0 0 0 2 0 0 0 0

2474 0 0 0 0 1 0 0 0 0

PBYSTAND AWAPART AWABEDR AWALAND APERSAUT ABESAUT AMOTSCO AVRAAUT

627 0 0 0 0 1 0 0 0

1083 0 0 0 0 0 0 0 0

1270 0 1 0 0 1 0 0 0

5488 0 1 0 0 1 0 0 0

418 0 1 0 0 0 0 0 0

2474 0 1 0 0 0 0 0 0

AAANHANG ATRACTOR AWERKT ABROM ALEVEN APERSONG AGEZONG AWAOREG ABRAND

627 0 0 0 0 0 0 0 0 0

1083 0 0 0 0 0 0 0 0 0

1270 0 0 0 0 0 0 0 0 1

5488 0 0 0 0 0 0 0 0 0

418 0 0 0 0 0 0 0 0 1

2474 0 0 0 0 0 0 0 0 1

AZEILPL APLEZIER AFIETS AINBOED ABYSTAND Purchase

627 0 1 0 0 0 Yes

1083 0 0 0 0 0 No

1270 0 0 0 0 0 Yes

5488 0 0 0 0 0 No

418 0 0 0 0 0 No

2474 0 0 0 0 0 No

>

> install.packages("caret") ### Necessary to Plot Decision Tree###

* Step 2: Clean the dataset
* Step 3: Create train/test set

**Create train/test set**

Before you train your model, you need to perform two steps:

* Create a train and test set: You train the model on the train set and test the prediction on the test set (i.e. unseen data)
* Install rpart.plot from the console

The common practice is to split the data 80/20, 80 percent of the data serves to train the model, and 20 percent to make predictions. We have to create two separate data frames. Not to consider, the test set until we finish building our model.

>s=sample(nrow(df\_ca),.80\*nrow(df\_ca))

> s

[1] 4431 1373 2932 1569 4455 3902 706 5411 1663 1928 3387 5334 180 4250

[15] 1094 3615 4743 1115 3888 2921 5715 3747 3513 4934 961 2837 107 1188

[29] 598 1449 2784 5039 2462 4552 3643 3374 957 156 3076 350 796 5817

[43] 2718 3890 3879 3451 200 795 1690 1401 3948 4802 3238 4850 4014 2471

[57] 2509 2120 69 1806 3466 2748 2723 4724 491 1845 1675 4736 3110 622

[71] 4095 4673 2002 2519 5045 5515 2716 2939 2686 4035 4121 3142 5296 1105

[85] 1748 4386 3368 1379 5051 26 185 2850 868 575 1980 1344 2559 2401

[99] 1229 189 4328 1517 4624 1524 3663 1691 923 1951 2089 2322 3794 3385

[113] 401 829 4236 4896 2440 1086 326 813 3771 5208 1079 2500 2259 2594

[127] 1235 4920 2998 2730 2133 4495 594 3843 1199 5102 1605 1706 4977 3873

[141] 4866 2160 3455 551 941 5315 2989 1557 3736 2335 2167 4062 5030 972

[155] 3610 1325 2108 4153 1870 80 2267 2846 4273 5590 3309 5520 1798 5394

[169] 2054 489 5801 4018 1854 1868 893 3307 296 1332 2153 1904 4720 5713

[183] 5195 3472 2672 3795 4958 486 2096 3431 801 5320 5190 2003 5446 3697

[197] 5109 5207 2191 2467 5409 229 1931 3712 1110 2468 5804 3790 4847 787

[211] 3611 5183 1182 3598 4926 3738 4421 1628 5780 5458 5319 199 3802 39

[225] 709 2329 2330 1045 5059 3993 333 289 1982 4093 2320 1746 3770 3407

[239] 3249 4017 4295 793 3880 5633 305 3144 4370 4655 1217 1532 4890 3618

[253] 3907 2186 4684 2439 1771 2278 4563 3918 2721 299 84 2523 1890 3341

[267] 2974 3201 1178 5452 1339 2459 5306 3813 2155 3088 100 1701 5359 3623

[281] 1942 1072 634 1395 3056 4752 974 1249 3947 2429 5813 1634 4793 419

[295] 1358 3373 1820 355 5547 1947 3285 2421 2879 1673 3958 214 3413 5134

[309] 3085 1536 4884 2390 4112 1696 948 5543 4497 3730 5640 133 1067 4290

[323] 4657 4437 1810 2579 2964 3304 1827 5648 4586 2031 4211 2075 2039 1097

[337] 1148 1222 1620 2277 1092 2239 128 4393 5389 2820 3124 4703 3713 46

[351] 3073 552 5201 4656 5497 5548 3703 2774 5174 775 1508 3871 4270 3414

[365] 3422 2603 1754 779 2706 2315 759 1121 3686 1024 5568 3490 4689 4310

[379] 5456 1900 4858 4754 3102 2546 980 3568 5723 3063 3672 2626 2266 4744

[393] 4848 5642 5191 44 3336 2887 4588 2625 700 3052 1454 1725 139 2025

[407] 1359 2688 2154 3900 1653 2768 1996 11 4176 5672 2202 3994 222 4735

[421] 4120 1594 4507 4964 208 4988 1983 5486 4001 5380 3394 4509 3648 3904

[435] 4738 3828 1493 5291 5677 5614 3502 4775 2024 2665 4472 4954 4669 3267

[449] 184 1037 3692 1767 4172 4143 2641 5588 2592 1922 2438 2612 2776 5771

[463] 149 3429 1150 4246 960 1598 213 1651 752 3980 2035 5344 4602 5012

[477] 4786 4304 770 3276 3186 3480 3485 846 283 5625 4835 4644 3384 1599

[491] 1107 2016 3129 4778 890 4186 814 5682 1069 3844 231 3044 2092 886

[505] 4142 2615 3406 1046 131 2083 2861 4660 2419 1796 4286 4559 577 4796

[519] 2020 4895 2598 2282 1770 257 4100 3495 1964 653 2316 1415 4787 2200

[533] 4245 3167 994 4533 2929 499 5656 4537 1494 5800 1487 2522 3557 176

[547] 3735 799 3117 4904 3855 4579 2904 4065 3327 934 4664 2834 1446 3243

[561] 5427 2028 3097 5476 3526 86 5161 4145 5788 1017 1823 1001 2715 3040

[575] 4170 1934 2583 5413 2619 4251 1844 141 1914 1632 3251 3875 1763 1181

[589] 3562 2134 1104 431 2826 2139 2690 3121 384 4031 1833 1790 2311 375

[603] 761 1547 4144 1504 3004 3075 4158 2892 5403 4581 847 462 219 2912

[617] 1160 3570 4384 4517 2415 1837 906 719 2678 155 1965 4836 3761 4759

[631] 3408 1906 4840 3122 585 3356 4020 1936 2638 5505 4365 1516 1338 3335

[645] 3290 4346 5173 942 3601 4401 1801 1777 3035 784 1744 2481 4233 3294

[659] 3953 2396 4505 3889 4755 5049 3379 3199 2195 4585 3789 4166 2392 3512

[673] 5300 1271 2497 5197 5106 560 2284 3931 3945 378 1568 5316 786 1664

[687] 625 4813 3968 3521 4002 5223 344 2309 1666 1019 5116 1518 5283 2842

[701] 3906 5324 3487 1867 1030 2185 3031 5349 5401 2878 1210 2841 5366 3520

[715] 3242 1971 1789 2620 2610 1141 4715 207 1745 5343 5271 919 1995 4102

[729] 5469 1948 4722 5336 3661 3282 669 3395 2800 2049 2717 1528 531 2082

[743] 1678 1227 753 2835 738 3554 3527 5514 2951 3631 5114 120 4538 4080

[757] 1180 592 3321 2698 536 2215 5691 2591 1346 4340 2084 3793 845 4380

[771] 5761 433 2040 4625 4623 3397 1034 3627 1050 1544 1575 2577 4099 3005

[785] 3065 466 449 2880 5560 3258 2314 2551 2902 4192 4489 900 2726 3345

[799] 2571 710 3940 4428 2480 5539 985 24 4325 239 376 5103 2763 1689

[813] 5210 3695 2517 3069 1153 688 672 4916 1296 2731 5807 4182 1243 3564

[827] 135 1273 3838 686 2876 1125 569 278 4944 2563 28 4326 2384 1726

[841] 3283 3624 3894 4007 817 1970 1260 5363 4309 3817 2646 3853 3381 2379

[855] 4875 611 3995 2930 5007 704 3565 1469 4733 410 3080 4413 538 588

[869] 543 2628 2461 3753 211 3681 5203 2560 5139 4856 1451 2555 1176 3551

[883] 4209 1300 5204 2326 5795 5611 63 1612 2905 2223 1099 1613 5479 4682

[897] 1974 4997 2988 5498 2609 1958 1236 3132 1604 3393 3225 2274 5137 4366

[911] 5736 2216 5493 4516 3787 2348 4530 2755 2621 2515 1883 5105 1473 1710

[925] 158 3096 5375 3197 5575 541 4039 1143 5115 4216 3136 5001 4139 3964

[939] 3375 3808 674 513 4206 3470 892 2946 4274 4826 3425 5424 519 5698

[953] 336 5369 659 1164 3259 29 3545 3987 5706 1550 2735 382 1486 258

[967] 5741 4011 5365 1040 1652 931 2260 150 2132 250 2076 3262 2171 1368

[981] 802 4774 3704 5464 4900 4269 952 3370 4156 525 2110 3751 4344 3691

[995] 3141 1051 3280 1834 4037 1244

[ reached getOption("max.print") -- omitted 3657 entries ]

**>df\_tr=df\_ca[s,]**

**> df\_tr**

MOSTYPE MAANTHUI MGEMOMV MGEMLEEF MOSHOOFD MGODRK MGODPR MGODOV MGODGE

2016 8 1 4 3 2 1 5 1 3

1313 35 1 2 4 8 0 5 1 3

3849 25 1 1 5 6 1 5 1 3

2296 33 2 3 3 8 0 5 0 4

205 12 1 4 2 3 0 7 2 0

5449 3 1 5 3 1 1 5 2 3

1685 23 1 2 3 5 2 4 2 3

3249 4 2 2 4 1 0 5 1 3

2711 38 1 4 3 9 0 6 0 3

136 29 1 2 3 7 1 6 0 3

4987 39 1 4 3 9 0 6 1 3

MRELGE MRELSA MRELOV MFALLEEN MFGEKIND MFWEKIND MOPLHOOG MOPLMIDD

2016 7 1 2 0 3 6 3 5

1313 7 1 2 3 4 3 0 3

3849 2 0 7 7 1 2 2 5

2296 7 0 2 2 2 6 0 2

205 9 0 0 0 0 8 2 4

5449 5 1 4 0 4 5 0 9

1685 5 2 3 2 4 3 1 3

3249 6 1 3 3 4 3 3 3

2711 7 0 2 2 2 6 0 3

136 3 2 5 4 2 3 1 4

4987 7 1 2 1 1 8 0 2

MOPLLAAG MBERHOOG MBERZELF MBERBOER MBERMIDD MBERARBG MBERARBO MSKA MSKB1

2016 1 3 1 0 4 1 2 3 3

1313 6 0 0 0 2 2 6 0 2

3849 3 1 2 1 5 1 2 2 1

2296 7 0 2 2 0 6 0 1 0

205 4 2 0 0 2 5 0 2 0

5449 0 4 2 0 2 1 1 0 4

1685 5 1 1 0 3 3 2 2 2

3249 4 3 1 1 3 2 2 3 2

2711 6 0 0 0 2 8 0 0 1

136 5 3 0 0 4 2 2 1 2

4987 7 1 0 0 1 5 3 1 1

MSKB2 MSKC MSKD MHHUUR MHKOOP MAUT1 MAUT2 MAUT0 MZFONDS MZPART MINKM30

2016 2 2 1 1 8 7 0 2 5 4 1

1313 3 3 3 5 4 9 0 0 5 4 2

3849 5 2 0 1 8 4 1 4 3 6 1

2296 2 6 0 2 7 4 4 2 8 1 5

205 3 5 0 4 5 9 0 0 7 2 0

5449 0 5 0 0 9 5 1 3 5 4 3

1685 2 5 1 6 3 7 1 2 7 2 3

3249 2 3 1 3 6 6 1 3 7 2 3

2711 2 6 0 2 7 9 0 0 9 0 5

136 4 3 2 3 6 5 1 3 7 2 3

4987 1 7 1 6 3 6 1 2 9 0 1

MINK3045 MINK4575 MINK7512 MINK123M MINKGEM MKOOPKLA PWAPART PWABEDR

2016 3 5 1 1 5 7 0 0

1313 2 3 3 0 5 5 2 0

3849 7 1 1 0 4 1 0 0

2296 2 3 0 0 3 3 0 1

205 4 5 0 0 5 7 0 0

5449 2 3 2 1 5 6 2 0

1685 4 3 1 0 3 3 0 0

3249 2 4 1 0 3 6 0 0

2711 0 4 0 0 4 3 0 0

136 2 3 2 1 5 3 2 0

4987 7 2 0 0 3 5 0 0

PWALAND PPERSAUT PBESAUT PMOTSCO PVRAAUT PAANHANG PTRACTOR PWERKT PBROM

2016 0 0 0 0 0 0 0 0 3

1313 0 6 0 0 0 0 0 0 0

3849 0 0 0 0 0 0 0 0 0

2296 0 0 0 0 0 0 0 0 0

205 0 5 0 0 0 0 0 0 0

5449 0 5 0 0 0 0 0 0 0

1685 0 0 0 0 0 0 0 0 0

3249 0 0 0 0 0 0 0 0 0

2711 0 0 0 0 0 0 0 0 0

136 0 0 0 0 0 0 0 0 0

4987 0 6 0 0 0 0 0 0 0

PLEVEN PPERSONG PGEZONG PWAOREG PBRAND PZEILPL PPLEZIER PFIETS PINBOED

2016 0 0 0 0 0 0 0 0 0

1313 0 0 0 0 3 0 0 0 0

3849 0 0 0 0 2 0 0 0 0

2296 0 0 0 0 0 0 0 0 0

205 0 0 0 0 3 0 0 0 0

5449 0 0 0 0 4 0 0 0 0

1685 0 0 0 0 0 0 0 0 0

3249 0 0 0 0 4 0 0 0 0

2711 0 0 0 0 3 0 0 0 0

136 0 0 0 0 3 0 0 0 0

4987 0 0 0 0 0 0 0 0 0

PBYSTAND AWAPART AWABEDR AWALAND APERSAUT ABESAUT AMOTSCO AVRAAUT

2016 0 0 0 0 0 0 0 0

1313 0 1 0 0 1 0 0 0

3849 0 0 0 0 0 0 0 0

2296 0 0 1 0 0 0 0 0

205 0 0 0 0 1 0 0 0

5449 0 1 0 0 1 0 0 0

1685 0 0 0 0 0 0 0 0

3249 0 0 0 0 0 0 0 0

2711 0 0 0 0 0 0 0 0

136 0 1 0 0 0 0 0 0

4987 0 0 0 0 1 0 0 0

AAANHANG ATRACTOR AWERKT ABROM ALEVEN APERSONG AGEZONG AWAOREG ABRAND

2016 0 0 0 1 0 0 0 0 0

1313 0 0 0 0 0 0 0 0 1

3849 0 0 0 0 0 0 0 0 1

2296 0 0 0 0 0 0 0 0 0

205 0 0 0 0 0 0 0 0 1

5449 0 0 0 0 0 0 0 0 1

1685 0 0 0 0 0 0 0 0 0

3249 0 0 0 0 0 0 0 0 1

2711 0 0 0 0 0 0 0 0 1

136 0 0 0 0 0 0 0 0 1

4987 0 0 0 0 0 0 0 0 0

AZEILPL APLEZIER AFIETS AINBOED ABYSTAND Purchase

2016 0 0 0 0 0 No

1313 0 0 0 0 0 No

3849 0 0 0 0 0 No

2296 0 0 0 0 0 No

205 0 0 0 0 0 No

5449 0 0 0 0 0 No

1685 0 0 0 0 0 No

3249 0 0 0 0 0 No

2711 0 0 0 0 0 No

136 0 0 0 0 0 No

4987 0 0 0 0 0 No

[ reached 'max' / getOption("max.print") -- omitted 4646 rows ]

**>df\_ts=df\_ca[-s,]**

**> df\_ts**

MOSTYPE MAANTHUI MGEMOMV MGEMLEEF MOSHOOFD MGODRK MGODPR MGODOV MGODGE

1083 36 1 3 3 8 2 4 1 3

5347 23 1 1 2 5 2 4 2 4

2385 33 1 3 3 8 0 4 1 5

2077 33 1 3 4 8 1 5 2 3

3856 36 1 3 3 8 0 6 0 3

283 3 1 1 5 1 1 5 2 2

4295 7 3 2 2 2 0 3 0 6

4396 32 1 2 3 7 0 5 1 3

1393 8 1 3 3 2 3 4 2 2

154 12 1 3 2 3 2 3 0 5

115 33 1 3 3 8 0 7 1 2

MRELGE MRELSA MRELOV MFALLEEN MFGEKIND MFWEKIND MOPLHOOG MOPLMIDD

1083 6 1 3 5 0 4 3 3

5347 7 1 2 8 0 1 5 4

2385 6 2 2 2 4 4 1 3

2077 5 1 3 0 5 4 0 3

3856 6 2 2 1 4 5 2 2

283 5 1 4 4 4 1 1 7

4295 4 0 5 4 2 3 2 3

4396 6 0 3 3 3 3 0 1

1393 8 1 0 0 5 4 4 3

154 5 2 3 0 2 7 1 7

115 8 0 1 1 4 4 0 4

MOPLLAAG MBERHOOG MBERZELF MBERBOER MBERMIDD MBERARBG MBERARBO MSKA MSKB1

1083 3 3 0 2 4 0 2 3 3

5347 0 1 0 0 3 3 3 0 5

2385 6 1 1 1 3 3 2 1 2

2077 6 0 0 4 4 0 1 0 3

3856 5 3 1 1 3 1 3 2 2

283 2 2 1 0 6 1 1 2 4

4295 5 6 0 0 2 2 0 4 0

4396 8 0 0 0 4 0 5 0 1

1393 2 5 0 0 2 1 2 4 2

154 2 2 0 0 4 0 4 2 1

115 5 1 0 0 4 2 2 1 1

MSKB2 MSKC MSKD MHHUUR MHKOOP MAUT1 MAUT2 MAUT0 MZFONDS MZPART MINKM30

1083 3 2 0 0 9 6 1 3 5 4 0

5347 1 3 0 8 1 7 1 2 6 3 3

2385 3 4 1 5 4 6 2 2 8 1 4

2077 5 1 1 2 7 6 1 3 7 2 1

3856 2 3 1 4 5 7 1 1 6 3 2

283 2 2 1 4 5 7 1 2 5 4 4

4295 0 2 4 6 3 4 2 4 4 5 4

4396 1 5 3 9 0 7 0 2 9 0 6

1393 2 1 2 0 9 9 0 0 4 5 2

154 6 2 0 4 5 7 2 0 6 3 4

115 3 4 2 0 9 7 1 2 6 3 0

MINK3045 MINK4575 MINK7512 MINK123M MINKGEM MKOOPKLA PWAPART PWABEDR

1083 3 5 2 0 6 3 0 0

5347 5 1 0 0 3 3 2 0

2385 4 2 1 1 3 3 0 0

2077 8 0 0 0 3 3 2 0

3856 4 2 2 0 4 3 0 0

283 3 3 1 0 3 6 0 0

4295 3 3 0 0 3 7 0 0

4396 2 1 0 0 2 1 2 0

1393 2 5 0 0 5 7 1 0

154 5 1 0 0 3 7 0 0

115 5 4 0 0 4 3 0 0

PWALAND PPERSAUT PBESAUT PMOTSCO PVRAAUT PAANHANG PTRACTOR PWERKT PBROM

1083 0 0 0 0 0 0 0 0 0

5347 0 0 0 0 0 0 0 0 0

2385 0 6 0 0 0 0 0 0 0

2077 0 0 0 0 0 0 0 0 0

3856 3 6 0 0 0 0 4 0 0

283 0 0 0 0 0 0 0 0 0

4295 0 6 0 0 0 0 0 0 0

4396 0 0 0 0 0 0 0 0 0

1393 0 6 0 0 0 0 0 0 0

154 0 6 0 0 0 0 0 0 0

115 0 0 0 0 0 0 0 0 0

PLEVEN PPERSONG PGEZONG PWAOREG PBRAND PZEILPL PPLEZIER PFIETS PINBOED

1083 0 0 0 0 0 0 0 0 0

5347 0 0 0 0 1 0 0 0 0

2385 0 0 0 0 0 0 0 0 0

2077 3 0 0 0 4 0 0 0 0

3856 0 0 0 0 5 0 0 0 0

283 0 0 0 0 0 0 0 0 0

4295 0 0 0 0 0 0 0 0 0

4396 0 0 0 0 2 0 0 0 0

1393 0 0 0 0 0 0 0 0 0

154 0 0 0 0 2 0 0 0 0

115 0 0 0 0 0 0 0 0 0

PBYSTAND AWAPART AWABEDR AWALAND APERSAUT ABESAUT AMOTSCO AVRAAUT

1083 0 0 0 0 0 0 0 0

5347 0 1 0 0 0 0 0 0

2385 0 0 0 0 1 0 0 0

2077 0 1 0 0 0 0 0 0

3856 0 0 0 1 1 0 0 0

283 0 0 0 0 0 0 0 0

4295 0 0 0 0 1 0 0 0

4396 0 1 0 0 0 0 0 0

1393 0 1 0 0 2 0 0 0

154 0 0 0 0 1 0 0 0

115 0 0 0 0 0 0 0 0

AAANHANG ATRACTOR AWERKT ABROM ALEVEN APERSONG AGEZONG AWAOREG ABRAND

1083 0 0 0 0 0 0 0 0 0

5347 0 0 0 0 0 0 0 0 1

2385 0 0 0 0 0 0 0 0 0

2077 0 0 0 0 2 0 0 0 1

3856 0 2 0 0 0 0 0 0 1

283 0 0 0 0 0 0 0 0 0

4295 0 0 0 0 0 0 0 0 0

4396 0 0 0 0 0 0 0 0 1

1393 0 0 0 0 0 0 0 0 0

154 0 0 0 0 0 0 0 0 1

115 0 0 0 0 0 0 0 0 0

AZEILPL APLEZIER AFIETS AINBOED ABYSTAND Purchase

1083 0 0 0 0 0 No

5347 0 0 0 0 0 No

2385 0 0 0 0 0 No

2077 0 0 0 0 0 No

3856 0 0 0 0 0 No

283 0 0 0 0 0 No

4295 0 0 0 0 0 No

4396 0 0 0 0 0 No

1393 0 0 0 0 0 No

154 0 0 0 0 0 No

115 0 0 0 0 0 No

[ reached 'max' / getOption("max.print") -- omitted 1154 rows ]

>#### Check Dimensions of traines and testing data#####

**>dim(df\_tr)**

**[1] 4657 86**

**> dim(df\_ts)**

**[1] 1165 86**

* Step 4: Build the model

dtree\_mod1=rpart(Purchase~.,df\_tr,method="class",cp=.001)

> dtree\_mod1

n= 4657

node), split, n, loss, yval, (yprob)

\* denotes terminal node

1) root 4657 271 No (0.94180803 0.05819197)

2) PPERSAUT< 5.5 2769 65 No (0.97652582 0.02347418) \*

3) PPERSAUT>=5.5 1888 206 No (0.89088983 0.10911017)

6) MOSTYPE>=8.5 1517 131 No (0.91364535 0.08635465)

12) PPLEZIER< 0.5 1503 123 No (0.91816367 0.08183633)

24) PBRAND< 2.5 767 39 No (0.94915254 0.05084746) \*

25) PBRAND>=2.5 736 84 No (0.88586957 0.11413043)

50) MOPLLAAG>=4.5 438 33 No (0.92465753 0.07534247) \*

51) MOPLLAAG< 4.5 298 51 No (0.82885906 0.17114094)

102) MZPART>=3.5 152 17 No (0.88815789 0.11184211)

204) MBERARBG< 1.5 72 3 No (0.95833333 0.04166667) \*

205) MBERARBG>=1.5 80 14 No (0.82500000 0.17500000)

410) MINKM30< 1.5 22 0 No (1.00000000 0.00000000) \*

411) MINKM30>=1.5 58 14 No (0.75862069 0.24137931)

822) MAUT1< 6.5 27 3 No (0.88888889 0.11111111) \*

823) MAUT1>=6.5 31 11 No (0.64516129 0.35483871)

1646) MFGEKIND< 2.5 11 1 No (0.90909091 0.09090909) \*

1647) MFGEKIND>=2.5 20 10 No (0.50000000 0.50000000)

3294) MINK7512< 1.5 10 3 No (0.70000000 0.30000000) \*

3295) MINK7512>=1.5 10 3 Yes (0.30000000 0.70000000) \*

103) MZPART< 3.5 146 34 No (0.76712329 0.23287671)

206) MSKB2< 3.5 104 17 No (0.83653846 0.16346154)

412) MSKD>=0.5 47 3 No (0.93617021 0.06382979) \*

413) MSKD< 0.5 57 14 No (0.75438596 0.24561404)

826) MBERARBG>=4.5 11 0 No (1.00000000 0.00000000) \*

827) MBERARBG< 4.5 46 14 No (0.69565217 0.30434783)

1654) MINKGEM< 3.5 12 1 No (0.91666667 0.08333333) \*

1655) MINKGEM>=3.5 34 13 No (0.61764706 0.38235294)

3310) MZPART>=2.5 13 2 No (0.84615385 0.15384615) \*

3311) MZPART< 2.5 21 10 Yes (0.47619048 0.52380952)

6622) MBERARBO>=2.5 8 2 No (0.75000000 0.25000000) \*

6623) MBERARBO< 2.5 13 4 Yes (0.30769231 0.69230769) \*

207) MSKB2>=3.5 42 17 No (0.59523810 0.40476190)

414) MGEMOMV< 2.5 9 0 No (1.00000000 0.00000000) \*

415) MGEMOMV>=2.5 33 16 Yes (0.48484848 0.51515152)

830) MINKGEM< 3.5 7 1 No (0.85714286 0.14285714) \*

831) MINKGEM>=3.5 26 10 Yes (0.38461538 0.61538462)

1662) MBERARBG< 0.5 13 5 No (0.61538462 0.38461538) \*

1663) MBERARBG>=0.5 13 2 Yes (0.15384615 0.84615385) \*

13) PPLEZIER>=0.5 14 6 Yes (0.42857143 0.57142857) \*

7) MOSTYPE< 8.5 371 75 No (0.79784367 0.20215633)

14) PBRAND< 3.5 187 22 No (0.88235294 0.11764706)

28) MINK123M>=0.5 44 0 No (1.00000000 0.00000000) \*

29) MINK123M< 0.5 143 22 No (0.84615385 0.15384615)

58) MBERMIDD< 1.5 35 1 No (0.97142857 0.02857143) \*

59) MBERMIDD>=1.5 108 21 No (0.80555556 0.19444444)

118) MBERHOOG< 5.5 95 14 No (0.85263158 0.14736842)

236) MBERARBO< 2.5 72 7 No (0.90277778 0.09722222) \*

237) MBERARBO>=2.5 23 7 No (0.69565217 0.30434783)

474) PBRAND< 2.5 14 2 No (0.85714286 0.14285714) \*

475) PBRAND>=2.5 9 4 Yes (0.44444444 0.55555556) \*

119) MBERHOOG>=5.5 13 6 Yes (0.46153846 0.53846154) \*

15) PBRAND>=3.5 184 53 No (0.71195652 0.28804348)

30) MBERARBO< 3.5 176 46 No (0.73863636 0.26136364)

60) MGODGE>=1.5 150 34 No (0.77333333 0.22666667)

120) MBERMIDD< 2.5 81 13 No (0.83950617 0.16049383)

240) MFALLEEN>=0.5 51 4 No (0.92156863 0.07843137) \*

241) MFALLEEN< 0.5 30 9 No (0.70000000 0.30000000)

482) PWAPART< 0.5 8 0 No (1.00000000 0.00000000) \*

483) PWAPART>=0.5 22 9 No (0.59090909 0.40909091)

966) MGODGE>=3.5 7 1 No (0.85714286 0.14285714) \*

967) MGODGE< 3.5 15 7 Yes (0.46666667 0.53333333) \*

121) MBERMIDD>=2.5 69 21 No (0.69565217 0.30434783)

242) MKOOPKLA>=7.5 9 0 No (1.00000000 0.00000000) \*

243) MKOOPKLA< 7.5 60 21 No (0.65000000 0.35000000)

486) PBRAND>=4.5 8 0 No (1.00000000 0.00000000) \*

487) PBRAND< 4.5 52 21 No (0.59615385 0.40384615)

974) PLEVEN>=0.5 12 2 No (0.83333333 0.16666667) \*

975) PLEVEN< 0.5 40 19 No (0.52500000 0.47500000)

1950) MGEMLEEF< 3.5 33 14 No (0.57575758 0.42424242)

3900) MRELGE< 6.5 11 3 No (0.72727273 0.27272727) \*

3901) MRELGE>=6.5 22 11 No (0.50000000 0.50000000)

7802) MFGEKIND< 1.5 7 2 No (0.71428571 0.28571429) \*

7803) MFGEKIND>=1.5 15 6 Yes (0.40000000 0.60000000) \*

1951) MGEMLEEF>=3.5 7 2 Yes (0.28571429 0.71428571) \*

61) MGODGE< 1.5 26 12 No (0.53846154 0.46153846)

122) MOPLHOOG< 3 14 4 No (0.71428571 0.28571429) \*

123) MOPLHOOG>=3 12 4 Yes (0.33333333 0.66666667) \*

31) MBERARBO>=3.5 8 1 Yes (0.12500000 0.87500000) \*

> Summary(dtree\_mod1)

summary(dtree\_mod1)

Call:

rpart(formula = Purchase ~ ., data = df\_tr, method = "class",

cp = 0.001)

n= 4657

CP nsplit rel error xerror xstd

1 0.005904059 0 1.0000000 1.000000 0.05895173

2 0.003690037 7 0.9557196 1.036900 0.05996107

3 0.002952030 19 0.9040590 1.107011 0.06182043

4 0.002214022 24 0.8892989 1.143911 0.06277016

5 0.001230012 31 0.8671587 1.143911 0.06277016

6 0.001000000 39 0.8560886 1.199262 0.06415987

Variable importance

PPERSAUT PBRAND MBERARBO APERSAUT MOSTYPE MBERHOOG MZFONDS MFALLEEN

6 5 4 4 4 4 3 3

MOSHOOFD ABRAND MBERARBG MSKA MZPART MINKGEM PWAPART APLEZIER

3 3 3 3 3 3 2 2

PPLEZIER MAUT1 MOPLLAAG AWAPART MGEMOMV MHHUUR MFWEKIND MSKB2

2 2 2 2 2 2 2 2

MGODGE MGODPR MFGEKIND MHKOOP MAUT0 MKOOPKLA MINK4575 MOPLMIDD

2 2 2 2 2 2 1 1

MINKM30 MOPLHOOG MSKC MBERMIDD MRELOV MRELGE MINK7512 MSKD

1 1 1 1 1 1 1 1

MGEMLEEF MSKB1 MINK3045 MGODRK MINK123M ALEVEN PLEVEN

1 1 1 1 1 1 1

Node number 1: 4657 observations, complexity param=0.005904059

predicted class=No expected loss=0.05819197 P(node) =1

class counts: 4386 271

probabilities: 0.942 0.058

left son=2 (2769 obs) right son=3 (1888 obs)

Primary splits:

PPERSAUT < 5.5 to the left, improve=16.464990, (0 missing)

APERSAUT < 0.5 to the left, improve=11.193090, (0 missing)

PBRAND < 2.5 to the left, improve= 8.282489, (0 missing)

PPLEZIER < 0.5 to the left, improve= 8.103293, (0 missing)

APLEZIER < 0.5 to the left, improve= 8.103293, (0 missing)

Surrogate splits:

APERSAUT < 0.5 to the left, agree=0.891, adj=0.732, (0 split)

PBRAND < 3.5 to the left, agree=0.623, adj=0.070, (0 split)

PWAPART < 1.5 to the left, agree=0.605, adj=0.026, (0 split)

PTRACTOR < 1.5 to the left, agree=0.603, adj=0.020, (0 split)

PBYSTAND < 1 to the left, agree=0.603, adj=0.020, (0 split)

Node number 2: 2769 observations

predicted class=No expected loss=0.02347418 P(node) =0.5945888

class counts: 2704 65

probabilities: 0.977 0.023

Node number 3: 1888 observations, complexity param=0.005904059

predicted class=No expected loss=0.1091102 P(node) =0.4054112

class counts: 1682 206

probabilities: 0.891 0.109

left son=6 (1517 obs) right son=7 (371 obs)

Primary splits:

MOSTYPE < 8.5 to the right, improve=7.994978, (0 missing)

MOSHOOFD < 2.5 to the right, improve=7.994978, (0 missing)

PBRAND < 2.5 to the left, improve=7.704319, (0 missing)

MKOOPKLA < 5.5 to the left, improve=6.390461, (0 missing)

MOPLLAAG < 5.5 to the right, improve=6.128953, (0 missing)

Surrogate splits:

MOSHOOFD < 2.5 to the right, agree=1.000, adj=1.000, (0 split)

MKOOPKLA < 6.5 to the left, agree=0.865, adj=0.313, (0 split)

MSKA < 3.5 to the left, agree=0.837, adj=0.173, (0 split)

MBERHOOG < 5.5 to the left, agree=0.833, adj=0.151, (0 split)

MOPLLAAG < 0.5 to the right, agree=0.823, adj=0.100, (0 split)

Node number 6: 1517 observations, complexity param=0.005904059

predicted class=No expected loss=0.08635465 P(node) =0.3257462

class counts: 1386 131

probabilities: 0.914 0.086

left son=12 (1503 obs) right son=13 (14 obs)

Primary splits:

PPLEZIER < 0.5 to the left, improve=6.649676, (0 missing)

APLEZIER < 0.5 to the left, improve=6.649676, (0 missing)

PBRAND < 2.5 to the left, improve=2.730912, (0 missing)

MINKGEM < 3.5 to the left, improve=2.660027, (0 missing)

MOPLLAAG < 5.5 to the right, improve=2.093228, (0 missing)

Surrogate splits:

APLEZIER < 0.5 to the left, agree=1, adj=1, (0 split)

Node number 7: 371 observations, complexity param=0.005904059

predicted class=No expected loss=0.2021563 P(node) =0.07966502

class counts: 296 75

probabilities: 0.798 0.202

left son=14 (187 obs) right son=15 (184 obs)

Primary splits:

PBRAND < 3.5 to the left, improve=5.385629, (0 missing)

ABRAND < 0.5 to the left, improve=5.279109, (0 missing)

MBERARBO < 3.5 to the left, improve=3.356437, (0 missing)

PWAPART < 1.5 to the left, improve=3.206054, (0 missing)

AWAPART < 0.5 to the left, improve=3.105288, (0 missing)

Surrogate splits:

ABRAND < 0.5 to the left, agree=0.841, adj=0.679, (0 split)

PWAPART < 0.5 to the left, agree=0.757, adj=0.511, (0 split)

AWAPART < 0.5 to the left, agree=0.757, adj=0.511, (0 split)

MGEMOMV < 2.5 to the left, agree=0.580, adj=0.152, (0 split)

MHHUUR < 0.5 to the right, agree=0.580, adj=0.152, (0 split)

Node number 12: 1503 observations, complexity param=0.003690037

predicted class=No expected loss=0.08183633 P(node) =0.32274

class counts: 1380 123

probabilities: 0.918 0.082

left son=24 (767 obs) right son=25 (736 obs)

Primary splits:

PBRAND < 2.5 to the left, improve=3.008278, (0 missing)

PWAPART < 1.5 to the left, improve=2.844518, (0 missing)

MOPLLAAG < 4.5 to the right, improve=2.347469, (0 missing)

MINKGEM < 3.5 to the left, improve=2.334036, (0 missing)

AWAPART < 0.5 to the left, improve=2.168679, (0 missing)

Surrogate splits:

ABRAND < 0.5 to the left, agree=0.912, adj=0.821, (0 split)

PWAPART < 0.5 to the left, agree=0.792, adj=0.575, (0 split)

AWAPART < 0.5 to the left, agree=0.792, adj=0.575, (0 split)

MHHUUR < 5.5 to the right, agree=0.593, adj=0.170, (0 split)

MHKOOP < 3.5 to the left, agree=0.593, adj=0.170, (0 split)

Node number 13: 14 observations

predicted class=Yes expected loss=0.4285714 P(node) =0.003006227

class counts: 6 8

probabilities: 0.429 0.571

Node number 14: 187 observations, complexity param=0.001230012

predicted class=No expected loss=0.1176471 P(node) =0.04015461

class counts: 165 22

probabilities: 0.882 0.118

left son=28 (44 obs) right son=29 (143 obs)

Primary splits:

MINK123M < 0.5 to the right, improve=1.592760, (0 missing)

MZFONDS < 0.5 to the right, improve=1.406069, (0 missing)

MZPART < 8.5 to the left, improve=1.406069, (0 missing)

MBERARBO < 2.5 to the left, improve=1.247772, (0 missing)

MBERMIDD < 1.5 to the left, improve=1.079241, (0 missing)

Surrogate splits:

MINKGEM < 1 to the left, agree=0.786, adj=0.091, (0 split)

MOPLHOOG < 6.5 to the right, agree=0.781, adj=0.068, (0 split)

MAUT1 < 3.5 to the left, agree=0.775, adj=0.045, (0 split)

MAUT2 < 4.5 to the right, agree=0.775, adj=0.045, (0 split)

MGODRK < 4.5 to the right, agree=0.770, adj=0.023, (0 split)

Node number 15: 184 observations, complexity param=0.005904059

predicted class=No expected loss=0.2880435 P(node) =0.03951041

class counts: 131 53

probabilities: 0.712 0.288

left son=30 (176 obs) right son=31 (8 obs)

Primary splits:

MBERARBO < 3.5 to the left, improve=5.762846, (0 missing)

MFALLEEN < 1.5 to the right, improve=3.933226, (0 missing)

MOPLMIDD < 6.5 to the left, improve=2.016172, (0 missing)

MGODGE < 1.5 to the right, improve=1.822795, (0 missing)

PBRAND < 5.5 to the right, improve=1.754748, (0 missing)

Node number 24: 767 observations

predicted class=No expected loss=0.05084746 P(node) =0.1646983

class counts: 728 39

probabilities: 0.949 0.051

Node number 25: 736 observations, complexity param=0.003690037

predicted class=No expected loss=0.1141304 P(node) =0.1580417

class counts: 652 84

probabilities: 0.886 0.114

left son=50 (438 obs) right son=51 (298 obs)

Primary splits:

MOPLLAAG < 4.5 to the right, improve=3.255066, (0 missing)

MBERMIDD < 5.5 to the left, improve=1.688272, (0 missing)

MAUT1 < 6.5 to the left, improve=1.579568, (0 missing)

MRELGE < 8.5 to the left, improve=1.534625, (0 missing)

MOPLMIDD < 6.5 to the left, improve=1.420500, (0 missing)

Surrogate splits:

MOPLMIDD < 3.5 to the left, agree=0.807, adj=0.523, (0 split)

MOPLHOOG < 1.5 to the left, agree=0.776, adj=0.446, (0 split)

MSKC < 3.5 to the right, agree=0.769, adj=0.430, (0 split)

MOSTYPE < 23.5 to the right, agree=0.740, adj=0.359, (0 split)

MSKA < 1.5 to the left, agree=0.739, adj=0.356, (0 split)

Node number 28: 44 observations

predicted class=No expected loss=0 P(node) =0.009448143

class counts: 44 0

probabilities: 1.000 0.000

Node number 29: 143 observations, complexity param=0.001230012

predicted class=No expected loss=0.1538462 P(node) =0.03070646

class counts: 121 22

probabilities: 0.846 0.154

left son=58 (35 obs) right son=59 (108 obs)

Primary splits:

MBERMIDD < 1.5 to the left, improve=1.4545790, (0 missing)

MRELGE < 7.5 to the left, improve=1.3531560, (0 missing)

MBERARBO < 2.5 to the left, improve=1.2108930, (0 missing)

MRELOV < 0.5 to the right, improve=1.1282050, (0 missing)

MHHUUR < 0.5 to the right, improve=0.9323474, (0 missing)

Surrogate splits:

MBERHOOG < 6.5 to the right, agree=0.846, adj=0.371, (0 split)

MSKC < 1.5 to the left, agree=0.811, adj=0.229, (0 split)

MGODPR < 2.5 to the left, agree=0.804, adj=0.200, (0 split)

MSKA < 6.5 to the right, agree=0.804, adj=0.200, (0 split)

MAANTHUI < 2.5 to the right, agree=0.797, adj=0.171, (0 split)

Node number 30: 176 observations, complexity param=0.005904059

predicted class=No expected loss=0.2613636 P(node) =0.03779257

class counts: 130 46

probabilities: 0.739 0.261

left son=60 (150 obs) right son=61 (26 obs)

Primary splits:

MGODGE < 1.5 to the right, improve=2.444802, (0 missing)

MFALLEEN < 1.5 to the right, improve=2.369542, (0 missing)

MSKB1 < 1.5 to the left, improve=1.893939, (0 missing)

PBRAND < 4.5 to the right, improve=1.761563, (0 missing)

MOPLLAAG < 2.5 to the right, improve=1.465181, (0 missing)

Surrogate splits:

MGODPR < 6.5 to the left, agree=0.920, adj=0.462, (0 split)

MOPLMIDD < 6.5 to the left, agree=0.886, adj=0.231, (0 split)

MINK3045 < 5.5 to the left, agree=0.881, adj=0.192, (0 split)

MBERMIDD < 6.5 to the left, agree=0.869, adj=0.115, (0 split)

Node number 31: 8 observations

predicted class=Yes expected loss=0.125 P(node) =0.001717844

class counts: 1 7

probabilities: 0.125 0.875

Node number 50: 438 observations

predicted class=No expected loss=0.07534247 P(node) =0.09405196

class counts: 405 33

probabilities: 0.925 0.075

Node number 51: 298 observations, complexity param=0.003690037

predicted class=No expected loss=0.1711409 P(node) =0.06398969

class counts: 247 51

probabilities: 0.829 0.171

left son=102 (152 obs) right son=103 (146 obs)

Primary splits:

MZPART < 3.5 to the right, improve=2.181872, (0 missing)

MZFONDS < 6.5 to the left, improve=2.033018, (0 missing)

MINKGEM < 3.5 to the left, improve=1.819313, (0 missing)

PGEZONG < 1 to the left, improve=1.778107, (0 missing)

AGEZONG < 0.5 to the left, improve=1.778107, (0 missing)

Surrogate splits:

MZFONDS < 5.5 to the left, agree=0.993, adj=0.986, (0 split)

MSKA < 2.5 to the right, agree=0.695, adj=0.377, (0 split)

MBERHOOG < 2.5 to the right, agree=0.688, adj=0.363, (0 split)

MAUT0 < 1.5 to the left, agree=0.674, adj=0.336, (0 split)

MSKC < 3.5 to the left, agree=0.661, adj=0.308, (0 split)

Node number 58: 35 observations

predicted class=No expected loss=0.02857143 P(node) =0.007515568

class counts: 34 1

probabilities: 0.971 0.029

Node number 59: 108 observations, complexity param=0.001230012

predicted class=No expected loss=0.1944444 P(node) =0.0231909

class counts: 87 21

probabilities: 0.806 0.194

left son=118 (95 obs) right son=119 (13 obs)

Primary splits:

MBERHOOG < 5.5 to the left, improve=3.498111, (0 missing)

MOPLHOOG < 4.5 to the left, improve=2.560606, (0 missing)

MINK4575 < 4.5 to the left, improve=2.476861, (0 missing)

MZFONDS < 2.5 to the right, improve=1.633333, (0 missing)

MZPART < 6.5 to the left, improve=1.633333, (0 missing)

Surrogate splits:

MSKA < 5.5 to the left, agree=0.944, adj=0.538, (0 split)

MZFONDS < 1.5 to the right, agree=0.917, adj=0.308, (0 split)

MZPART < 7.5 to the left, agree=0.917, adj=0.308, (0 split)

Node number 60: 150 observations, complexity param=0.002214022

predicted class=No expected loss=0.2266667 P(node) =0.03220958

class counts: 116 34

probabilities: 0.773 0.227

left son=120 (81 obs) right son=121 (69 obs)

Primary splits:

MBERMIDD < 2.5 to the left, improve=1.5421150, (0 missing)

MFALLEEN < 1.5 to the right, improve=1.4116670, (0 missing)

MSKB1 < 1.5 to the left, improve=1.2873480, (0 missing)

PBRAND < 4.5 to the right, improve=1.1739680, (0 missing)

MBERARBG < 3.5 to the right, improve=0.9838298, (0 missing)

Surrogate splits:

MBERHOOG < 3.5 to the right, agree=0.747, adj=0.449, (0 split)

MOSTYPE < 3.5 to the left, agree=0.740, adj=0.435, (0 split)

MSKA < 3.5 to the right, agree=0.740, adj=0.435, (0 split)

MOSHOOFD < 1.5 to the left, agree=0.720, adj=0.391, (0 split)

MSKB1 < 1.5 to the left, agree=0.720, adj=0.391, (0 split)

Node number 61: 26 observations, complexity param=0.005904059

predicted class=No expected loss=0.4615385 P(node) =0.005582993

class counts: 14 12

probabilities: 0.538 0.462

left son=122 (14 obs) right son=123 (12 obs)

Primary splits:

MOPLHOOG < 3 to the left, improve=1.875458, (0 missing)

MINKGEM < 4.5 to the left, improve=1.875458, (0 missing)

MINK3045 < 3.5 to the right, improve=1.848077, (0 missing)

MOPLMIDD < 5 to the right, improve=1.230769, (0 missing)

MSKB2 < 3.5 to the right, improve=1.230769, (0 missing)

Surrogate splits:

MOPLMIDD < 5 to the right, agree=0.962, adj=0.917, (0 split)

MSKB2 < 3.5 to the right, agree=0.962, adj=0.917, (0 split)

MBERHOOG < 2.5 to the left, agree=0.923, adj=0.833, (0 split)

MZFONDS < 5.5 to the right, agree=0.923, adj=0.833, (0 split)

MINKGEM < 4.5 to the left, agree=0.923, adj=0.833, (0 split)

Node number 102: 152 observations, complexity param=0.00295203

predicted class=No expected loss=0.1118421 P(node) =0.03263904

class counts: 135 17

probabilities: 0.888 0.112

left son=204 (72 obs) right son=205 (80 obs)

Primary splits:

MBERARBG < 1.5 to the left, improve=1.3473680, (0 missing)

MINKM30 < 1.5 to the left, improve=1.2220060, (0 missing)

PWAPART < 1.5 to the left, improve=1.0140350, (0 missing)

MSKB2 < 2.5 to the left, improve=0.9800386, (0 missing)

MRELGE < 6.5 to the left, improve=0.8875771, (0 missing)

Surrogate splits:

MSKC < 2.5 to the left, agree=0.730, adj=0.431, (0 split)

MOPLLAAG < 2.5 to the left, agree=0.711, adj=0.389, (0 split)

MHHUUR < 0.5 to the left, agree=0.684, adj=0.333, (0 split)

MHKOOP < 8.5 to the right, agree=0.684, adj=0.333, (0 split)

MGEMOMV < 2.5 to the left, agree=0.651, adj=0.264, (0 split)

Node number 103: 146 observations, complexity param=0.003690037

predicted class=No expected loss=0.2328767 P(node) =0.03135065

class counts: 112 34

probabilities: 0.767 0.233

left son=206 (104 obs) right son=207 (42 obs)

Primary splits:

MSKB2 < 3.5 to the left, improve=3.483981, (0 missing)

MINKM30 < 0.5 to the right, improve=2.950399, (0 missing)

MINKGEM < 3.5 to the left, improve=2.860036, (0 missing)

MINK7512 < 2.5 to the left, improve=2.665251, (0 missing)

MSKD < 0.5 to the right, improve=2.324138, (0 missing)

Surrogate splits:

MSKC < 0.5 to the right, agree=0.747, adj=0.119, (0 split)

MINK4575 < 0.5 to the right, agree=0.747, adj=0.119, (0 split)

MINK7512 < 2.5 to the left, agree=0.747, adj=0.119, (0 split)

MINKGEM < 6.5 to the left, agree=0.747, adj=0.119, (0 split)

MAANTHUI < 1.5 to the left, agree=0.740, adj=0.095, (0 split)

Node number 118: 95 observations, complexity param=0.001230012

predicted class=No expected loss=0.1473684 P(node) =0.0203994

class counts: 81 14

probabilities: 0.853 0.147

left son=236 (72 obs) right son=237 (23 obs)

Primary splits:

MBERARBO < 2.5 to the left, improve=1.4956650, (0 missing)

PBRAND < 0.5 to the left, improve=1.0165410, (0 missing)

ABRAND < 0.5 to the left, improve=1.0165410, (0 missing)

MBERHOOG < 0.5 to the right, improve=0.8998361, (0 missing)

MAUT2 < 1.5 to the right, improve=0.7941901, (0 missing)

Surrogate splits:

MINKGEM < 2.5 to the right, agree=0.821, adj=0.261, (0 split)

MHHUUR < 5.5 to the left, agree=0.811, adj=0.217, (0 split)

MHKOOP < 2.5 to the right, agree=0.811, adj=0.217, (0 split)

MAUT0 < 3.5 to the left, agree=0.811, adj=0.217, (0 split)

MGEMOMV < 1.5 to the right, agree=0.800, adj=0.174, (0 split)

Node number 119: 13 observations

predicted class=Yes expected loss=0.4615385 P(node) =0.002791497

class counts: 6 7

probabilities: 0.462 0.538

Node number 120: 81 observations, complexity param=0.001230012

predicted class=No expected loss=0.1604938 P(node) =0.01739317

class counts: 68 13

probabilities: 0.840 0.160

left son=240 (51 obs) right son=241 (30 obs)

Primary splits:

MFALLEEN < 0.5 to the right, improve=1.854611, (0 missing)

MHHUUR < 0.5 to the right, improve=1.266185, (0 missing)

MHKOOP < 8.5 to the left, improve=1.266185, (0 missing)

MOPLHOOG < 2.5 to the left, improve=1.140447, (0 missing)

MAUT0 < 2.5 to the left, improve=1.101292, (0 missing)

Surrogate splits:

MRELOV < 0.5 to the right, agree=0.840, adj=0.567, (0 split)

MGEMLEEF < 3.5 to the right, agree=0.802, adj=0.467, (0 split)

MRELGE < 7.5 to the left, agree=0.778, adj=0.400, (0 split)

MGEMOMV < 2.5 to the left, agree=0.753, adj=0.333, (0 split)

MFWEKIND < 5.5 to the left, agree=0.753, adj=0.333, (0 split)

Node number 121: 69 observations, complexity param=0.002214022

predicted class=No expected loss=0.3043478 P(node) =0.01481641

class counts: 48 21

probabilities: 0.696 0.304

left son=242 (9 obs) right son=243 (60 obs)

Primary splits:

MKOOPKLA < 7.5 to the right, improve=1.9173910, (0 missing)

PBRAND < 4.5 to the right, improve=1.6764080, (0 missing)

PLEVEN < 0.5 to the right, improve=1.1210950, (0 missing)

ALEVEN < 0.5 to the right, improve=1.1210950, (0 missing)

MSKC < 1.5 to the left, improve=0.9767133, (0 missing)

Surrogate splits:

MOSTYPE < 2 to the left, agree=0.899, adj=0.222, (0 split)

MGODRK < 2.5 to the right, agree=0.884, adj=0.111, (0 split)

Node number 122: 14 observations

predicted class=No expected loss=0.2857143 P(node) =0.003006227

class counts: 10 4

probabilities: 0.714 0.286

Node number 123: 12 observations

predicted class=Yes expected loss=0.3333333 P(node) =0.002576766

class counts: 4 8

probabilities: 0.333 0.667

Node number 204: 72 observations

predicted class=No expected loss=0.04166667 P(node) =0.0154606

class counts: 69 3

probabilities: 0.958 0.042

Node number 205: 80 observations, complexity param=0.00295203

predicted class=No expected loss=0.175 P(node) =0.01717844

class counts: 66 14

probabilities: 0.825 0.175

left son=410 (22 obs) right son=411 (58 obs)

Primary splits:

MINKM30 < 1.5 to the left, improve=1.858621, (0 missing)

MSKB2 < 2.5 to the left, improve=1.596164, (0 missing)

PWAPART < 1.5 to the left, improve=1.039394, (0 missing)

MZFONDS < 4.5 to the left, improve=1.022921, (0 missing)

MZPART < 4.5 to the right, improve=1.022921, (0 missing)

Surrogate splits:

MINK3045 < 4.5 to the right, agree=0.850, adj=0.455, (0 split)

MFALLEEN < 0.5 to the left, agree=0.812, adj=0.318, (0 split)

MBERHOOG < 3.5 to the right, agree=0.812, adj=0.318, (0 split)

MHHUUR < 0.5 to the left, agree=0.800, adj=0.273, (0 split)

MHKOOP < 8.5 to the right, agree=0.800, adj=0.273, (0 split)

Node number 206: 104 observations, complexity param=0.003690037

predicted class=No expected loss=0.1634615 P(node) =0.02233197

class counts: 87 17

probabilities: 0.837 0.163

left son=412 (47 obs) right son=413 (57 obs)

Primary splits:

MSKD < 0.5 to the right, improve=1.7024790, (0 missing)

MBERARBG < 4.5 to the right, improve=1.0104900, (0 missing)

MAUT2 < 1.5 to the left, improve=0.8981635, (0 missing)

MOSTYPE < 36.5 to the right, improve=0.8647558, (0 missing)

MOPLMIDD < 3.5 to the right, improve=0.7828874, (0 missing)

Surrogate splits:

MFALLEEN < 0.5 to the right, agree=0.702, adj=0.340, (0 split)

MRELGE < 6.5 to the left, agree=0.692, adj=0.319, (0 split)

MRELOV < 0.5 to the right, agree=0.683, adj=0.298, (0 split)

MAUT1 < 6.5 to the left, agree=0.683, adj=0.298, (0 split)

MOSTYPE < 37.5 to the right, agree=0.663, adj=0.255, (0 split)

Node number 207: 42 observations, complexity param=0.003690037

predicted class=No expected loss=0.4047619 P(node) =0.009018682

class counts: 25 17

probabilities: 0.595 0.405

left son=414 (9 obs) right son=415 (33 obs)

Primary splits:

MGEMOMV < 2.5 to the left, improve=3.753247, (0 missing)

MINKGEM < 3.5 to the left, improve=2.936042, (0 missing)

MKOOPKLA < 3.5 to the left, improve=2.936042, (0 missing)

MFWEKIND < 3.5 to the left, improve=2.880952, (0 missing)

MFALLEEN < 2.5 to the right, improve=2.752381, (0 missing)

Surrogate splits:

MFALLEEN < 2.5 to the right, agree=0.952, adj=0.778, (0 split)

MBERARBO < 3.5 to the right, agree=0.929, adj=0.667, (0 split)

MAUT1 < 5.5 to the left, agree=0.905, adj=0.556, (0 split)

MAUT0 < 2.5 to the right, agree=0.905, adj=0.556, (0 split)

MBERARBG < 2.5 to the right, agree=0.857, adj=0.333, (0 split)

Node number 236: 72 observations

predicted class=No expected loss=0.09722222 P(node) =0.0154606

class counts: 65 7

probabilities: 0.903 0.097

Node number 237: 23 observations, complexity param=0.001230012

predicted class=No expected loss=0.3043478 P(node) =0.004938802

class counts: 16 7

probabilities: 0.696 0.304

left son=474 (14 obs) right son=475 (9 obs)

Primary splits:

PBRAND < 2.5 to the left, improve=1.866115, (0 missing)

MSKB1 < 2.5 to the right, improve=1.477592, (0 missing)

ABRAND < 0.5 to the left, improve=1.354515, (0 missing)

MAUT1 < 5.5 to the left, improve=1.104210, (0 missing)

MBERMIDD < 4.5 to the right, improve=1.104210, (0 missing)

Surrogate splits:

ABRAND < 0.5 to the left, agree=0.957, adj=0.889, (0 split)

MBERHOOG < 0.5 to the right, agree=0.783, adj=0.444, (0 split)

MINKGEM < 3.5 to the right, agree=0.783, adj=0.444, (0 split)

MGEMLEEF < 2.5 to the right, agree=0.739, adj=0.333, (0 split)

MAUT0 < 3.5 to the left, agree=0.739, adj=0.333, (0 split)

Node number 240: 51 observations

predicted class=No expected loss=0.07843137 P(node) =0.01095126

class counts: 47 4

probabilities: 0.922 0.078

Node number 241: 30 observations, complexity param=0.001230012

predicted class=No expected loss=0.3 P(node) =0.006441915

class counts: 21 9

probabilities: 0.700 0.300

left son=482 (8 obs) right son=483 (22 obs)

Primary splits:

PWAPART < 0.5 to the left, improve=1.9636360, (0 missing)

AWAPART < 0.5 to the left, improve=1.9636360, (0 missing)

MGODGE < 3.5 to the right, improve=1.5186600, (0 missing)

MSKC < 2.5 to the right, improve=0.9174603, (0 missing)

MBERZELF < 0.5 to the left, improve=0.8296651, (0 missing)

Surrogate splits:

AWAPART < 0.5 to the left, agree=1.000, adj=1.000, (0 split)

MOPLLAAG < 2.5 to the right, agree=0.767, adj=0.125, (0 split)

MBERARBG < 2.5 to the right, agree=0.767, adj=0.125, (0 split)

PBRAND < 4.5 to the right, agree=0.767, adj=0.125, (0 split)

Node number 242: 9 observations

predicted class=No expected loss=0 P(node) =0.001932575

class counts: 9 0

probabilities: 1.000 0.000

Node number 243: 60 observations, complexity param=0.002214022

predicted class=No expected loss=0.35 P(node) =0.01288383

class counts: 39 21

probabilities: 0.650 0.350

left son=486 (8 obs) right son=487 (52 obs)

Primary splits:

PBRAND < 4.5 to the right, improve=2.2615380, (0 missing)

PLEVEN < 0.5 to the right, improve=1.5670810, (0 missing)

ALEVEN < 0.5 to the right, improve=1.5670810, (0 missing)

MSKB1 < 3.5 to the left, improve=0.8947712, (0 missing)

MGODRK < 0.5 to the left, improve=0.8091575, (0 missing)

Surrogate splits:

PWABEDR < 2 to the right, agree=0.917, adj=0.375, (0 split)

AWABEDR < 0.5 to the right, agree=0.917, adj=0.375, (0 split)

PWAOREG < 3 to the right, agree=0.900, adj=0.250, (0 split)

AWAOREG < 0.5 to the right, agree=0.900, adj=0.250, (0 split)

PPERSAUT < 6.5 to the right, agree=0.883, adj=0.125, (0 split)

Node number 410: 22 observations

predicted class=No expected loss=0 P(node) =0.004724071

class counts: 22 0

probabilities: 1.000 0.000

Node number 411: 58 observations, complexity param=0.00295203

predicted class=No expected loss=0.2413793 P(node) =0.01245437

class counts: 44 14

probabilities: 0.759 0.241

left son=822 (27 obs) right son=823 (31 obs)

Primary splits:

MAUT1 < 6.5 to the left, improve=1.714498, (0 missing)

MGODRK < 0.5 to the right, improve=1.536116, (0 missing)

PWAPART < 1.5 to the left, improve=1.408046, (0 missing)

AWAPART < 0.5 to the left, improve=1.241379, (0 missing)

MINK123M < 0.5 to the right, improve=1.220327, (0 missing)

Surrogate splits:

MBERZELF < 0.5 to the right, agree=0.741, adj=0.444, (0 split)

MFALLEEN < 1.5 to the right, agree=0.724, adj=0.407, (0 split)

MAUT0 < 0.5 to the right, agree=0.724, adj=0.407, (0 split)

MINK123M < 0.5 to the right, agree=0.707, adj=0.370, (0 split)

MINKM30 < 2.5 to the right, agree=0.690, adj=0.333, (0 split)

Node number 412: 47 observations

predicted class=No expected loss=0.06382979 P(node) =0.01009233

class counts: 44 3

probabilities: 0.936 0.064

Node number 413: 57 observations, complexity param=0.003690037

predicted class=No expected loss=0.245614 P(node) =0.01223964

class counts: 43 14

probabilities: 0.754 0.246

left son=826 (11 obs) right son=827 (46 obs)

Primary splits:

MBERARBG < 4.5 to the right, improve=1.6445460, (0 missing)

MKOOPKLA < 6.5 to the right, improve=1.4632330, (0 missing)

MAUT2 < 1.5 to the left, improve=1.3375130, (0 missing)

MAUT1 < 8.5 to the right, improve=1.1228070, (0 missing)

MFALLEEN < 1.5 to the left, improve=0.9981317, (0 missing)

Surrogate splits:

MAUT1 < 8.5 to the right, agree=0.912, adj=0.545, (0 split)

MSKB1 < 0.5 to the left, agree=0.895, adj=0.455, (0 split)

MGODPR < 1 to the left, agree=0.877, adj=0.364, (0 split)

PMOTSCO < 2 to the right, agree=0.842, adj=0.182, (0 split)

AMOTSCO < 0.5 to the right, agree=0.842, adj=0.182, (0 split)

Node number 414: 9 observations

predicted class=No expected loss=0 P(node) =0.001932575

class counts: 9 0

probabilities: 1.000 0.000

Node number 415: 33 observations, complexity param=0.003690037

predicted class=Yes expected loss=0.4848485 P(node) =0.007086107

class counts: 16 17

probabilities: 0.485 0.515

left son=830 (7 obs) right son=831 (26 obs)

Primary splits:

MINKGEM < 3.5 to the left, improve=2.462870, (0 missing)

MINK7512 < 1.5 to the left, improve=2.328327, (0 missing)

MBERARBG < 0.5 to the left, improve=1.845143, (0 missing)

MINK4575 < 3.5 to the left, improve=1.484848, (0 missing)

MSKA < 0.5 to the right, improve=1.484848, (0 missing)

Surrogate splits:

MHHUUR < 7.5 to the right, agree=0.879, adj=0.429, (0 split)

MHKOOP < 1.5 to the left, agree=0.879, adj=0.429, (0 split)

MINKM30 < 2.5 to the right, agree=0.879, adj=0.429, (0 split)

PBRAND < 3.5 to the left, agree=0.879, adj=0.429, (0 split)

MOSTYPE < 38.5 to the right, agree=0.818, adj=0.143, (0 split)

Node number 474: 14 observations

predicted class=No expected loss=0.1428571 P(node) =0.003006227

class counts: 12 2

probabilities: 0.857 0.143

Node number 475: 9 observations

predicted class=Yes expected loss=0.4444444 P(node) =0.001932575

class counts: 4 5

probabilities: 0.444 0.556

Node number 482: 8 observations

predicted class=No expected loss=0 P(node) =0.001717844

class counts: 8 0

probabilities: 1.000 0.000

Node number 483: 22 observations, complexity param=0.001230012

predicted class=No expected loss=0.4090909 P(node) =0.004724071

class counts: 13 9

probabilities: 0.591 0.409

left son=966 (7 obs) right son=967 (15 obs)

Primary splits:

MGODGE < 3.5 to the right, improve=1.4554110, (0 missing)

MOSTYPE < 4.5 to the right, improve=0.6363636, (0 missing)

MOSHOOFD < 1.5 to the right, improve=0.6363636, (0 missing)

MBERHOOG < 5.5 to the right, improve=0.6363636, (0 missing)

MAUT1 < 7.5 to the right, improve=0.6363636, (0 missing)

Surrogate splits:

MBERHOOG < 6.5 to the right, agree=0.909, adj=0.714, (0 split)

MINK4575 < 5.5 to the right, agree=0.864, adj=0.571, (0 split)

MGODPR < 3.5 to the left, agree=0.818, adj=0.429, (0 split)

MFGEKIND < 0.5 to the left, agree=0.773, adj=0.286, (0 split)

MBERARBG < 0.5 to the left, agree=0.773, adj=0.286, (0 split)

Node number 486: 8 observations

predicted class=No expected loss=0 P(node) =0.001717844

class counts: 8 0

probabilities: 1.000 0.000

Node number 487: 52 observations, complexity param=0.002214022

predicted class=No expected loss=0.4038462 P(node) =0.01116599

class counts: 31 21

probabilities: 0.596 0.404

left son=974 (12 obs) right son=975 (40 obs)

Primary splits:

PLEVEN < 0.5 to the right, improve=1.7551280, (0 missing)

ALEVEN < 0.5 to the right, improve=1.7551280, (0 missing)

MINK4575 < 3.5 to the right, improve=1.2672200, (0 missing)

MSKD < 0.5 to the right, improve=1.1217950, (0 missing)

MOPLHOOG < 1.5 to the right, improve=0.9527473, (0 missing)

Surrogate splits:

ALEVEN < 0.5 to the right, agree=1, adj=1, (0 split)

Node number 822: 27 observations

predicted class=No expected loss=0.1111111 P(node) =0.005797724

class counts: 24 3

probabilities: 0.889 0.111

Node number 823: 31 observations, complexity param=0.00295203

predicted class=No expected loss=0.3548387 P(node) =0.006656646

class counts: 20 11

probabilities: 0.645 0.355

left son=1646 (11 obs) right son=1647 (20 obs)

Primary splits:

MFGEKIND < 2.5 to the left, improve=2.375367, (0 missing)

MINK3045 < 3.5 to the right, improve=1.386531, (0 missing)

MFWEKIND < 6.5 to the right, improve=1.139201, (0 missing)

MOPLHOOG < 1.5 to the left, improve=1.139201, (0 missing)

MINK7512 < 1.5 to the left, improve=1.075901, (0 missing)

Surrogate splits:

MFWEKIND < 5.5 to the right, agree=0.903, adj=0.727, (0 split)

MHHUUR < 1.5 to the left, agree=0.903, adj=0.727, (0 split)

MHKOOP < 7.5 to the right, agree=0.903, adj=0.727, (0 split)

MAUT1 < 8.5 to the right, agree=0.839, adj=0.545, (0 split)

MAUT2 < 0.5 to the left, agree=0.839, adj=0.545, (0 split)

Node number 826: 11 observations

predicted class=No expected loss=0 P(node) =0.002362036

class counts: 11 0

probabilities: 1.000 0.000

Node number 827: 46 observations, complexity param=0.003690037

predicted class=No expected loss=0.3043478 P(node) =0.009877604

class counts: 32 14

probabilities: 0.696 0.304

left son=1654 (12 obs) right son=1655 (34 obs)

Primary splits:

MINKGEM < 3.5 to the left, improve=1.5861040, (0 missing)

MZPART < 2.5 to the right, improve=0.8819120, (0 missing)

MZFONDS < 6.5 to the left, improve=0.8819120, (0 missing)

MAUT2 < 1.5 to the left, improve=0.8699275, (0 missing)

MOSTYPE < 10.5 to the left, improve=0.8356182, (0 missing)

Surrogate splits:

MINK4575 < 1.5 to the left, agree=0.870, adj=0.500, (0 split)

MRELOV < 3.5 to the right, agree=0.826, adj=0.333, (0 split)

MBERHOOG < 0.5 to the left, agree=0.826, adj=0.333, (0 split)

MRELGE < 3.5 to the left, agree=0.804, adj=0.250, (0 split)

MSKA < 0.5 to the left, agree=0.804, adj=0.250, (0 split)

Node number 830: 7 observations

predicted class=No expected loss=0.1428571 P(node) =0.001503114

class counts: 6 1

probabilities: 0.857 0.143

Node number 831: 26 observations, complexity param=0.003690037

predicted class=Yes expected loss=0.3846154 P(node) =0.005582993

class counts: 10 16

probabilities: 0.385 0.615

left son=1662 (13 obs) right son=1663 (13 obs)

Primary splits:

MBERARBG < 0.5 to the left, improve=2.769231, (0 missing)

MGODPR < 6.5 to the right, improve=2.190045, (0 missing)

MGODGE < 0.5 to the left, improve=2.190045, (0 missing)

MFGEKIND < 2.5 to the right, improve=1.557692, (0 missing)

MSKA < 0.5 to the right, improve=1.557692, (0 missing)

Surrogate splits:

MINK4575 < 3.5 to the left, agree=0.885, adj=0.769, (0 split)

MGODPR < 4.5 to the right, agree=0.846, adj=0.692, (0 split)

MGODGE < 0.5 to the left, agree=0.846, adj=0.692, (0 split)

MFWEKIND < 4.5 to the left, agree=0.808, adj=0.615, (0 split)

MGODRK < 0.5 to the left, agree=0.769, adj=0.538, (0 split)

Node number 966: 7 observations

predicted class=No expected loss=0.1428571 P(node) =0.001503114

class counts: 6 1

probabilities: 0.857 0.143

Node number 967: 15 observations

predicted class=Yes expected loss=0.4666667 P(node) =0.003220958

class counts: 7 8

probabilities: 0.467 0.533

Node number 974: 12 observations

predicted class=No expected loss=0.1666667 P(node) =0.002576766

class counts: 10 2

probabilities: 0.833 0.167

Node number 975: 40 observations, complexity param=0.002214022

predicted class=No expected loss=0.475 P(node) =0.008589221

class counts: 21 19

probabilities: 0.525 0.475

left son=1950 (33 obs) right son=1951 (7 obs)

Primary splits:

MGEMLEEF < 3.5 to the left, improve=0.9716450, (0 missing)

MOPLHOOG < 1.5 to the right, improve=0.8532258, (0 missing)

MGODOV < 0.5 to the right, improve=0.7901254, (0 missing)

MINK3045 < 3.5 to the left, improve=0.7901254, (0 missing)

MINK4575 < 3.5 to the right, improve=0.7581841, (0 missing)

Surrogate splits:

MFWEKIND < 3.5 to the right, agree=0.925, adj=0.571, (0 split)

MGEMOMV < 2.5 to the right, agree=0.900, adj=0.429, (0 split)

MRELGE < 5.5 to the right, agree=0.900, adj=0.429, (0 split)

MSKB1 < 4.5 to the left, agree=0.900, adj=0.429, (0 split)

MOSTYPE < 3.5 to the right, agree=0.875, adj=0.286, (0 split)

Node number 1646: 11 observations

predicted class=No expected loss=0.09090909 P(node) =0.002362036

class counts: 10 1

probabilities: 0.909 0.091

Node number 1647: 20 observations, complexity param=0.00295203

predicted class=No expected loss=0.5 P(node) =0.00429461

class counts: 10 10

probabilities: 0.500 0.500

left son=3294 (10 obs) right son=3295 (10 obs)

Primary splits:

MINK7512 < 1.5 to the left, improve=1.6000000, (0 missing)

MINK3045 < 3.5 to the right, improve=0.9890110, (0 missing)

MOSTYPE < 34 to the right, improve=0.4166667, (0 missing)

MSKD < 0.5 to the left, improve=0.4166667, (0 missing)

MGODGE < 3.5 to the left, improve=0.4000000, (0 missing)

Surrogate splits:

MFALLEEN < 1.5 to the right, agree=0.85, adj=0.7, (0 split)

MINK3045 < 3.5 to the right, agree=0.85, adj=0.7, (0 split)

MBERARBO < 2.5 to the right, agree=0.75, adj=0.5, (0 split)

MSKD < 1.5 to the left, agree=0.75, adj=0.5, (0 split)

MHHUUR < 2.5 to the left, agree=0.75, adj=0.5, (0 split)

Node number 1654: 12 observations

predicted class=No expected loss=0.08333333 P(node) =0.002576766

class counts: 11 1

probabilities: 0.917 0.083

Node number 1655: 34 observations, complexity param=0.003690037

predicted class=No expected loss=0.3823529 P(node) =0.007300837

class counts: 21 13

probabilities: 0.618 0.382

left son=3310 (13 obs) right son=3311 (21 obs)

Primary splits:

MZPART < 2.5 to the right, improve=2.198018, (0 missing)

MZFONDS < 6.5 to the left, improve=2.198018, (0 missing)

MFALLEEN < 1.5 to the left, improve=1.498217, (0 missing)

MAUT2 < 1.5 to the left, improve=1.498217, (0 missing)

MGODPR < 5.5 to the left, improve=1.231900, (0 missing)

Surrogate splits:

MZFONDS < 6.5 to the left, agree=1.000, adj=1.000, (0 split)

MFGEKIND < 4.5 to the right, agree=0.765, adj=0.385, (0 split)

MRELOV < 0.5 to the left, agree=0.735, adj=0.308, (0 split)

MFALLEEN < 0.5 to the left, agree=0.735, adj=0.308, (0 split)

MOPLLAAG < 3.5 to the right, agree=0.735, adj=0.308, (0 split)

Node number 1662: 13 observations

predicted class=No expected loss=0.3846154 P(node) =0.002791497

class counts: 8 5

probabilities: 0.615 0.385

Node number 1663: 13 observations

predicted class=Yes expected loss=0.1538462 P(node) =0.002791497

class counts: 2 11

probabilities: 0.154 0.846

Node number 1950: 33 observations, complexity param=0.002214022

predicted class=No expected loss=0.4242424 P(node) =0.007086107

class counts: 19 14

probabilities: 0.576 0.424

left son=3900 (11 obs) right son=3901 (22 obs)

Primary splits:

MRELGE < 6.5 to the left, improve=0.7575758, (0 missing)

MFWEKIND < 4.5 to the left, improve=0.7575758, (0 missing)

MSKD < 0.5 to the right, improve=0.5827506, (0 missing)

MSKC < 2.5 to the left, improve=0.4545455, (0 missing)

MHHUUR < 1.5 to the right, improve=0.4545455, (0 missing)

Surrogate splits:

MFALLEEN < 1.5 to the right, agree=0.909, adj=0.727, (0 split)

MRELOV < 2.5 to the right, agree=0.848, adj=0.545, (0 split)

MZFONDS < 5.5 to the right, agree=0.848, adj=0.545, (0 split)

MZPART < 3.5 to the left, agree=0.848, adj=0.545, (0 split)

MFWEKIND < 4.5 to the left, agree=0.818, adj=0.455, (0 split)

Node number 1951: 7 observations

predicted class=Yes expected loss=0.2857143 P(node) =0.001503114

class counts: 2 5

probabilities: 0.286 0.714

Node number 3294: 10 observations

predicted class=No expected loss=0.3 P(node) =0.002147305

class counts: 7 3

probabilities: 0.700 0.300

Node number 3295: 10 observations

predicted class=Yes expected loss=0.3 P(node) =0.002147305

class counts: 3 7

probabilities: 0.300 0.700

Node number 3310: 13 observations

predicted class=No expected loss=0.1538462 P(node) =0.002791497

class counts: 11 2

probabilities: 0.846 0.154

Node number 3311: 21 observations, complexity param=0.003690037

predicted class=Yes expected loss=0.4761905 P(node) =0.004509341

class counts: 10 11

probabilities: 0.476 0.524

left son=6622 (8 obs) right son=6623 (13 obs)

Primary splits:

MBERARBO < 2.5 to the right, improve=1.937729, (0 missing)

MOSTYPE < 12 to the left, improve=1.190476, (0 missing)

MSKA < 1.5 to the left, improve=1.190476, (0 missing)

MBERMIDD < 3.5 to the left, improve=1.185281, (0 missing)

MZFONDS < 7.5 to the right, improve=1.142857, (0 missing)

Surrogate splits:

MZFONDS < 7.5 to the right, agree=0.952, adj=0.875, (0 split)

MZPART < 1.5 to the left, agree=0.952, adj=0.875, (0 split)

MGODPR < 4.5 to the left, agree=0.810, adj=0.500, (0 split)

MOPLLAAG < 0.5 to the left, agree=0.810, adj=0.500, (0 split)

MSKA < 0.5 to the left, agree=0.810, adj=0.500, (0 split)

Node number 3900: 11 observations

predicted class=No expected loss=0.2727273 P(node) =0.002362036

class counts: 8 3

probabilities: 0.727 0.273

Node number 3901: 22 observations, complexity param=0.002214022

predicted class=No expected loss=0.5 P(node) =0.004724071

class counts: 11 11

probabilities: 0.500 0.500

left son=7802 (7 obs) right son=7803 (15 obs)

Primary splits:

MFGEKIND < 1.5 to the left, improve=0.9428571, (0 missing)

MFWEKIND < 7.5 to the right, improve=0.9428571, (0 missing)

MINK4575 < 4.5 to the right, improve=0.9428571, (0 missing)

MGEMOMV < 3.5 to the right, improve=0.8461538, (0 missing)

MBERARBO < 0.5 to the left, improve=0.8181818, (0 missing)

Surrogate splits:

MFWEKIND < 7.5 to the right, agree=1.000, adj=1.000, (0 split)

MSKB1 < 3.5 to the right, agree=0.909, adj=0.714, (0 split)

MSKB2 < 1.5 to the left, agree=0.909, adj=0.714, (0 split)

MINK7512 < 0.5 to the left, agree=0.909, adj=0.714, (0 split)

MGODRK < 0.5 to the left, agree=0.864, adj=0.571, (0 split)

Node number 6622: 8 observations

predicted class=No expected loss=0.25 P(node) =0.001717844

class counts: 6 2

probabilities: 0.750 0.250

Node number 6623: 13 observations

predicted class=Yes expected loss=0.3076923 P(node) =0.002791497

class counts: 4 9

probabilities: 0.308 0.692

Node number 7802: 7 observations

predicted class=No expected loss=0.2857143 P(node) =0.001503114

class counts: 5 2

probabilities: 0.714 0.286

Node number 7803: 15 observations

predicted class=Yes expected loss=0.4 P(node) =0.003220958

class counts: 6 9

probabilities: 0.400 0.600

##### Plot Decision Tree###

* library(ISLR)
* library(rpart)
* library(rattle)



fancyRpartPlot(dtree\_mod1)

et CG\_CONTEXT\_SHOW\_BACKTRACE environmental variable.

Warning message:

labs do not fit even at cex 0.15, there may be some overplotting

* Step 5: Make prediction

> pred\_mod1=predict(dtree\_mod1,df\_ts[,-86],type="class",cp=.001)

> pred\_mod1

1083 5347 2385 2077 3856 283 4295 4396 1393 154 115 5143 3838 2518 4168

No No No No No No Yes No No No No No No No No

192 30 3913 1359 3520 4674 4816 2555 2046 2906 3923 1193 4064 2090 76

No No No No No No No No No No No No No Yes No

931 3939 3954 234 5080 5382 941 3042 1493 1150 1296 356 1633 3812 5545

No No No No No No No No No No No No No No No

1160 746 4269 2883 3235 2162 1163 845 5778 2611 4580 2728 2867 3446 3059

No No No No No No No No No No No No No No No

2089 5006 184 301 5683 2945 2843 921 3163 3081 771 659 4527 6 5177

No No Yes No No No No No No No No No No No No

3135 5762 1603 3705 3468 481 143 5031 2508 2513 229 2612 3494 2682 4698

No No No No No No No No No No No No No No No

4302 1218 1379 785 5215 3355 38 5597 5405 4547 5149 221 5498 340 3474

No No No No No No No No No No No Yes No No No

1774 4487 2404 4446 3543 101 1900 1244 2258 1737 3804 5062 5708 2842 4686

No No No No No No No No No No No No No No No

1801 2704 3276 993 4133 4536 4990 3280 4353 2649 2566 2985 4113 4318 4732

No No No No No No No No No No No No No No No

1556 396 1684 4919 5654 2004 5621 2179 5765 5607 1700 2316 2483 2793 5231

No No No No No No No No No No No No Yes No No

251 1987 5067 5674 1128 5264 311 4757 286 2208 3689 5106 5518 1112 5416

No No No No No No No No No No No No No No No

2746 2756 5089 829 4013 2172 2785 4121 5348 3833 1970 3775 5554 3430 5424

No No No No No No Yes No No No No No No No No

2365 3619 2722 3673 5241 1560 2772 187 3400 4421 5063 5110 675 2731 3049

No No No No No No No No No No No No No No No

3861 5567 3892 3525 3097 2161 5338 3514 4195 2107 5234 360 3141 173 4550

No No No No No No No No No No No No No No No

1716 485 2701 1185 2688 2869 4402 4650 1928 1327 1391 3533 5702 4689 2631

No No No No No No No No No No No No No No No

5359 46 3807 117 2113 2657 1271 4591 2076 4850 3902 2145 1582 4197 1349

No No No No No No No No No No No No No No No

2243 3031 3133 3521 460 3917 3695 3562 274 3230 3218 3583 1399 5465 3505

No No No No No No No No No No No No No No No

1154 2777 3365 4291 2324 5350 581 156 1166 2354 4615 3091 4411 5295 1151

No No No No No No No No No No No No No No No

5423 5680 2897 5811 5627 1554 4009 1823 4803 63 5057 3095 2423 4184 660

No No No No No No No No No No No No No No No

4238 959 5748 701 2983 3440 5812 5468 793 4965 5764 5180 1597 2050 850

No No No No No No No No No No No No No No No

1812 4703 4932 854 4875 2420 3195 4107 5615 3387 2153 5783 190 2907 5549

No No No No No No No No No No No No No No No

5235 2386 3451 2521 1418 3204 3877 508 5458 4278 1941 4408 4942 2417 4872

No No Yes No No No No No No No No No No No No

3254 2219 5727 3889 2942 4211 3946 3774 3319 2911 918 4590 4625 625 1161

No No No No No No No No No No No No No No No

564 5436 4490 3166 1950 3701 4392 3813 2158 5483 4877 3507 236 814 4457

No No No No No No No No Yes No No No No No No

1795 3960 2256 3222 3497 1127 4067 5305 3580 3726 2428 664 5444 5090 1194

No No No No No No No No No No No No No No No

2931 1947 1246 3942 2120 3579 4730 1445 4000 3024 2442 3891 5004 2174 578

No No No No No No No No No No No No No No No

3409 4277 4911 4132 1925 4754 5570 3092 3816 3205 1113 1307 3160 4283 5268

No No No No No No No No No No No No No No No

3996 3967 2625 44 2414 1132 3604 4209 4671 2593 2196 2118 562 1073 3653

No No No No No No No No No No No No No No No

5481 20 128 1984 1778 2470 1793 2710 3788 4204 2456 2157 4474 5503 4568

No No No No No No No No No No No No No No No

1337 2812 3056 5 646 5138 2659 1604 983 401 590 2982 1820 1790 5469

No No No No No No No No No No No No No No No

2493 1618 5581 3364 1300 5595 2670 4648 1308 722 5233 1422 1593 4182 621

No No No No No No No No No No No No No No No

4070 3515 3075 4988 3539 4974 4518 1849 880 3601 4504 4325 1273 1626 5315

No No No No No No No No No No No No No No No

1595 4187 507 702 5415 5304 4384 519 4610 5200 4961 4726 2340 5626 3485

No No No No No No No No No No No No No No No

2263 2245 3318 5752 2504 3781 4497 433 3665 3036 365 2463 488 2695 3345

No No No No No No No No No No No No No No No

5077 4137 4899 2956 4001 1497 731 1558 2721 3592 2141 3506 178 537 758

Yes No No No No No No No No No No No No No No

2810 1711 2788 5406 2719 5353 4883 897 904 1588 5369 3773 1010 710 1077

No No No No No No No No No No No No No No No

328 2849 5167 3363 4215 5343 964 3770 4807 1813 4364 3082 5462 5194 3293

No No No No No No No No No No No No No No No

3737 860 4420 2159 2726 4921 1057 3564 3593 4986 37 1020 1361 1707 425

No No No No No No No No No No No No No No No

3307 5505 3388 4898 4136 1453 2472 4078 801 258 2503 1131 1111 2257 5279

No No No No No No No No No No No No No No No

1498 4316 2353 2305 2491 4660 3599 835 65 4257 1095 4312 1637 364 4239

No No No No No No No No No No No No No No No

142 1238 5647 5364 1896 2656 3636 1395 3084 4409 986 3492 1714 185 571

No No No No No No No No Yes No No No No No No

4758 4793 5653 3979 2600 3019 5401 408 2605 4120 2151 1888 97 4289 1459

No No No No No No No No No No No No No No No

882 2627 4240 1868 4827 3823 292 696 728 2803 2723 2301 5577 5393 2140

No No No No No No No No No No No Yes No No No

1245 1107 1863 3226 2327 2712 2851 4071 3391 3158 1473 358 698 1330 3573

No No No No No No No No No No No No No No No

3424 2236 1404 529 604 2150 2123 3460 2599 1144 1976 943 5796 4844 3569

No No No No No No No No No No No No No No No

335 2984 5763 3523 5713 995 3147 3028 2402 517 5625 5362 2581 5565 5365

No No No No No No No No No No No No No No No

5563 2425 4166 427 5303 4356 3890 2534 3211 5552 94 5114 2771 3389 3551

No No No No No No No No No No No No No No No

3032 2979 1655 1721 879 933 2792 2307 3964 1419 5329 3392 2831 1598 5719

No No No No No No No No No No No No No No No

4545 4560 1084 3969 1409 5689 775 5725 4342 5758 4052 3190 1365 4805 5410

No No No No No No No No No No No Yes No No No

4603 5690 3641 3935 512 2965 1475 145 5457 3383 381 4026 3399 932 3415

No No No No No No No No No No No No No No No

3620 1781 3074 1919 177 3787 4526 2446 2900 4451 5760 1265 343 1013 2606

No No No No No No Yes No No No No No No No No

839 439 141 3038 3679 3030 859 3894 1318 912 1770 5508 5705 885 3443

No No No No No No No No No No No No No No No

2093 324 5782 246 1628 4460 3860 2001 1326 466 5717 1586 2195 2333 5169

No No No No No No No No No No No No No No No

168 1450 1045 828 5472 3428 1967 4762 3961 1442 2248 1730 841 3382 5313

No No No No Yes No No No No No No No No No No

4750 1477 3847 5308 2382 4463 5022 4815 1096 2706 3600 2037 4777 181 354

No No No No No No No No No No No No No No No

4073 1577 4301 4596 3783 5544 5407 1509 5771 2371 669 2753 3648 4929 1610

No No No No No No No No No No No No No No No

1783 915 4427 2038 4393 4829 2 1306 825 1028 2490 3362 5286 2227 791

No No No No No No No No No No No No No No No

5714 111 120 969 2233 4734 487 1815 228 974 2921 617 3089 981 1414

No No No No No No No No No No No No No No No

4021 1789 2197 2915 275 1648 4179 4149 2318 4020 3542 3566 309 2303 5775

No No No No No No No No No No No No No No No

5726 1223 2684 3231 2336 1287 5259 3962 5664 2395 5170 4981 2130 3149 609

No No Yes No No No No No No No No No No No No

1534 4371 5479 5253 3760 5164 3025 2271 5676 1336 5033 4174 5357 1474 5232

No No No No No No No No No No No No No No No

2881 3199 61 4210 684 600 1173 4566 5334 789 2418 3779 3924 4177 2289

No No No No No No No No No No No No No No No

5041 2235 759 3811 5620 2574 5800 533 3776 2056 5596 3397 125 1742 5396

No No No No Yes No No No No No No No Yes No No

4945 1695 2660 477 4173 5649 4255 977 3216 3367 3337 3605 293 1749 1066

No No No No No No No No Yes No No No No No No

5278 3613 5133 3331 4957 1788 2567 4864 911 5070 5288 5122 98 935 824

No No No No No No No No No No No No No No No

2468 1668 1121 58 3360 5341 4529 3270 2445 4502 2690 2786 5456 4909 5082

No No No No No No No No No No No No No No No

5230 5246 157 1787 2576 1579 762 3901 4216 2207

No No No No No No No No No No

[ reached getOption("max.print") -- omitted 165 entries ]

Levels: No Yes

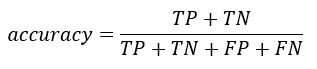
> ### -5 because 5th is species column##

**> table(df\_ts$Purchase,pred\_mod1)### Confusion Matrix**##



The confusion matrix is a better choice to evaluate the classification performance. The general idea is to count the number of times True instances are classified are False.

The confusion matrix is a better choice to evaluate the classification performance. The general idea is to count the number of times True instances are classified are False. Each row in a confusion matrix represents an actual target, while each column represents a predicted target.

[](https://www.guru99.com/images/r_programming/032918_0938_DecisionTre3.jpg)### Accuracy test from the confusion matrix:

**> pred\_mod1**

**No Yes**

**No 1072 16**

**Yes 72 5**

> ac\_tab=table(df\_ts$Purchase,pred\_mod1)

> sum(diag(ac\_tab))/sum(ac\_tab)

**[1] 0.9244635 …. Accuracy 92%**

>

**Step 7) Tune the hyper-parameters**

Decision tree has various parameters that control aspects of the fit. In rpart library, you can control the parameters using the rpart.control() function.

**Find out important variable---**

>varImp(dtree\_mod1)

Overall

ABRAND 7.6501654

AGEZONG 1.7781069

ALEVEN 4.4433040

APERSAUT 11.1930907

APLEZIER 14.7529694

AWAPART 8.4789826

MAUT0 1.1012918

MAUT1 6.1574458

MAUT2 5.3980115

MBERARBG 9.6006072

MBERARBO 15.8295233

MBERHOOG 5.0343104

MBERMIDD 8.0536981

MBERZELF 0.8296651

MFALLEEN 14.8177766

MFGEKIND 4.8759160

MFWEKIND 5.7205858

MGEMLEEF 0.9716450

MGEMOMV 4.5994006

MGODGE 9.8317141

MGODOV 0.7901254

MGODPR 3.4219457

MGODRK 2.3452737

MHHUUR 2.6530778

MHKOOP 1.2661849

MINK123M 2.8130869

MINK3045 5.0137441

MINK4575 6.9299700

MINK7512 7.6694792

MINKGEM 18.5338873

MINKM30 6.0310257

MKOOPKLA 12.7071278

MOPLHOOG 8.5216848

MOPLLAAG 15.2898969

MOPLMIDD 5.4503288

MOSHOOFD 8.6313415

MOSTYPE 11.9388584

MRELGE 4.5329340

MRELOV 1.1282051

MSKA 4.2330170

MSKB1 5.5536508

MSKB2 7.2909522

MSKC 2.3487191

MSKD 6.1478293

MZFONDS 10.3181282

MZPART 9.3241252

**PBRAND 38.6225090**

PGEZONG 1.7781069

PLEVEN 4.4433040

**PPERSAUT 16.4649858**

PPLEZIER 14.7529694

PWAPART 11.4756836

MAANTHUI 0.0000000

MRELSA 0.0000000

MBERBOER 0.0000000

PWABEDR 0.0000000

PWALAND 0.0000000

PBESAUT 0.0000000

PMOTSCO 0.0000000

PVRAAUT 0.0000000

PAANHANG 0.0000000

PTRACTOR 0.0000000

PWERKT 0.0000000

PBROM 0.0000000

PPERSONG 0.0000000

PWAOREG 0.0000000

PZEILPL 0.0000000

PFIETS 0.0000000

PINBOED 0.0000000

PBYSTAND 0.0000000

AWABEDR 0.0000000

AWALAND 0.0000000

ABESAUT 0.0000000

AMOTSCO 0.0000000

AVRAAUT 0.0000000

AAANHANG 0.0000000

ATRACTOR 0.0000000

AWERKT 0.0000000

ABROM 0.0000000

APERSONG 0.0000000

AWAOREG 0.0000000

AZEILPL 0.0000000

AFIETS 0.0000000

AINBOED 0.0000000

ABYSTAND 0.0000000

### Make a another model with only important variables###

dtree\_mod2=rpart(Purchase~PBRAND+PPERSAUT,df\_tr,method="class",control=rpart.control(minsplit=1,minbucket=1,maxdepth=30,cp=0.001))

> dtree\_mod2

n= 4657

node), split, n, loss, yval, (yprob)

\* denotes terminal node

1. root 4657 271 No (0.94180803 0.05819197) \*

>summary(dtree\_mod2)

Call:

rpart(formula = Purchase ~ PBRAND + PPERSAUT, data = df\_tr, method = "class",

control = rpart.control(minsplit = 1, minbucket = 1, maxdepth = 30,

cp = 0.001))

n= 4657

CP nsplit rel error xerror xstd

1 0 0 1 0 0

Node number 1: 4657 observations

predicted class=No expected loss=0.05819197 P(node) =1

class counts: 4386 271

probabilities: 0.942 0.058

>library(caret)##### Cross Validation####

###cross validation..Traiing data divided into 50%model and rest for testing###

### 1,2,3,45,..ist 1 for traing and other for testing##

### tr\_cv1=traing cross validation##

r\_cv1=trainControl(method="LOOCV")

> model\_loocv=train(Purchase~+PBRAND+MINKGEM+PPERSAUT+MGODGE+APLEZIER,data=df\_tr,method="rpart",trControl=tr\_cv1)

> model\_cv10

CART

112 samples

4 predictor

3 classes: 'setosa', 'versicolor', 'virginica'

No pre-processing

Resampling: Cross-Validated (10 fold)

Summary of sample sizes: 101, 100, 101, 101, 100, 101, ...

Resampling results across tuning parameters:

cp Accuracy Kappa

0.0000000 0.9462121 0.9174828

0.4571429 0.7674242 0.6474828

0.4857143 0.5113636 0.2250000

Accuracy was used to select the optimal model using the largest value.

The final value used for the model was cp = 0.

Repeated Cross Validation is not giving proper results.

**Summary: Decision tree is completed sucessfully**

.