Call:

rpart(formula = Type ~ ., data = mal\_model\_clean, method = "class")

n= 1247

CP nsplit rel error xerror xstd

1 0.10191083 0 1.0000000 1.0000000 0.07461570

2 0.07006369 2 0.7961783 0.7579618 0.06608382

3 0.03821656 3 0.7261146 0.6687898 0.06245897

4 0.02547771 9 0.3949045 0.4840764 0.05380868

5 0.02229299 11 0.3439490 0.4649682 0.05280349

6 0.01000000 13 0.2993631 0.4713376 0.05314128

Variable importance

SOURCE\_APP\_BYTES REMOTE\_APP\_PACKETS APP\_PACKETS

21 19 16

REMOTE\_IPS CONTENT\_LENGTH DIST\_REMOTE\_TCP\_PORT

15 10 8

NUMBER\_SPECIAL\_CHARACTERS URL\_LENGTH

6 5

Node number 1: 1247 observations, complexity param=0.1019108

predicted class=0 expected loss=0.1259022 P(node) =1

class counts: 1090 157

probabilities: 0.874 0.126

left son=2 (1149 obs) right son=3 (98 obs)

Primary splits:

NUMBER\_SPECIAL\_CHARACTERS < 18.5 to the left, improve=25.09687, (0 missing)

DIST\_REMOTE\_TCP\_PORT < 1.5 to the right, improve=23.59567, (0 missing)

URL\_LENGTH < 92.5 to the left, improve=20.31284, (0 missing)

REMOTE\_APP\_PACKETS < 1 to the left, improve=18.32822, (0 missing)

SOURCE\_APP\_BYTES < 62 to the left, improve=18.32822, (0 missing)

Surrogate splits:

URL\_LENGTH < 92.5 to the left, agree=0.967, adj=0.582, (0 split)

Node number 2: 1149 observations, complexity param=0.03821656

predicted class=0 expected loss=0.09660574 P(node) =0.9214114

class counts: 1038 111

probabilities: 0.903 0.097

left son=4 (511 obs) right son=5 (638 obs)

Primary splits:

DIST\_REMOTE\_TCP\_PORT < 1.5 to the right, improve=12.061070, (0 missing)

REMOTE\_APP\_PACKETS < 1 to the left, improve=10.765290, (0 missing)

SOURCE\_APP\_BYTES < 62 to the left, improve=10.765290, (0 missing)

APP\_PACKETS < 8.5 to the left, improve= 7.244552, (0 missing)

REMOTE\_IPS < 0.5 to the left, improve= 5.725061, (0 missing)

Surrogate splits:

APP\_PACKETS < 1.5 to the right, agree=0.790, adj=0.528, (0 split)

REMOTE\_IPS < 0.5 to the right, agree=0.789, adj=0.526, (0 split)

REMOTE\_APP\_PACKETS < 2.5 to the right, agree=0.789, adj=0.524, (0 split)

SOURCE\_APP\_BYTES < 153 to the right, agree=0.789, adj=0.524, (0 split)

NUMBER\_SPECIAL\_CHARACTERS < 6.5 to the left, agree=0.563, adj=0.018, (0 split)

Node number 3: 98 observations, complexity param=0.1019108

predicted class=0 expected loss=0.4693878 P(node) =0.07858861

class counts: 52 46

probabilities: 0.531 0.469

left son=6 (42 obs) right son=7 (56 obs)

Primary splits:

APP\_PACKETS < 3 to the right, improve=26.14966, (0 missing)

REMOTE\_IPS < 1.5 to the right, improve=24.47129, (0 missing)

REMOTE\_APP\_PACKETS < 5.5 to the right, improve=22.64857, (0 missing)

SOURCE\_APP\_BYTES < 429 to the right, improve=22.64857, (0 missing)

CONTENT\_LENGTH < 110 to the right, improve=20.15640, (0 missing)

Surrogate splits:

REMOTE\_IPS < 1.5 to the right, agree=0.990, adj=0.976, (0 split)

REMOTE\_APP\_PACKETS < 5.5 to the right, agree=0.969, adj=0.929, (0 split)

SOURCE\_APP\_BYTES < 429 to the right, agree=0.969, adj=0.929, (0 split)

DIST\_REMOTE\_TCP\_PORT < 1.5 to the right, agree=0.857, adj=0.667, (0 split)

CONTENT\_LENGTH < 33.5 to the right, agree=0.806, adj=0.548, (0 split)

Node number 4: 511 observations

predicted class=0 expected loss=0.01565558 P(node) =0.4097835

class counts: 503 8

probabilities: 0.984 0.016

Node number 5: 638 observations, complexity param=0.03821656

predicted class=0 expected loss=0.161442 P(node) =0.5116279

class counts: 535 103

probabilities: 0.839 0.161

left son=10 (384 obs) right son=11 (254 obs)

Primary splits:

REMOTE\_APP\_PACKETS < 1 to the left, improve=50.27838, (0 missing)

SOURCE\_APP\_BYTES < 62 to the left, improve=50.27838, (0 missing)

APP\_PACKETS < 0.5 to the left, improve=38.43446, (0 missing)

REMOTE\_IPS < 0.5 to the left, improve=37.30447, (0 missing)

DIST\_REMOTE\_TCP\_PORT < 0.5 to the left, improve=14.30743, (0 missing)

Surrogate splits:

SOURCE\_APP\_BYTES < 62 to the left, agree=1.000, adj=1.000, (0 split)

APP\_PACKETS < 0.5 to the left, agree=0.983, adj=0.957, (0 split)

REMOTE\_IPS < 0.5 to the left, agree=0.981, adj=0.953, (0 split)

DIST\_REMOTE\_TCP\_PORT < 0.5 to the left, agree=0.715, adj=0.283, (0 split)

URL\_LENGTH < 24.5 to the right, agree=0.610, adj=0.020, (0 split)

Node number 6: 42 observations

predicted class=0 expected loss=0.04761905 P(node) =0.03368083

class counts: 40 2

probabilities: 0.952 0.048

Node number 7: 56 observations, complexity param=0.07006369

predicted class=1 expected loss=0.2142857 P(node) =0.04490778

class counts: 12 44

probabilities: 0.214 0.786

left son=14 (11 obs) right son=15 (45 obs)

Primary splits:

REMOTE\_APP\_PACKETS < 1 to the left, improve=16.9015900, (0 missing)

SOURCE\_APP\_BYTES < 62 to the left, improve=16.9015900, (0 missing)

CONTENT\_LENGTH < 4.5 to the right, improve= 9.8775510, (0 missing)

NUMBER\_SPECIAL\_CHARACTERS < 23.5 to the left, improve= 0.9523810, (0 missing)

URL\_LENGTH < 101 to the right, improve= 0.9494505, (0 missing)

Surrogate splits:

SOURCE\_APP\_BYTES < 62 to the left, agree=1.000, adj=1.000, (0 split)

CONTENT\_LENGTH < 4.5 to the right, agree=0.893, adj=0.455, (0 split)

Node number 10: 384 observations

predicted class=0 expected loss=0 P(node) =0.3079391

class counts: 384 0

probabilities: 1.000 0.000

Node number 11: 254 observations, complexity param=0.03821656

predicted class=0 expected loss=0.4055118 P(node) =0.2036889

class counts: 151 103

probabilities: 0.594 0.406

left son=22 (71 obs) right son=23 (183 obs)

Primary splits:

SOURCE\_APP\_BYTES < 15525 to the right, improve=12.375750, (0 missing)

CONTENT\_LENGTH < 8765.5 to the right, improve=11.535410, (0 missing)

URL\_LENGTH < 43.5 to the right, improve=10.828450, (0 missing)

APP\_PACKETS < 4.5 to the right, improve= 8.578242, (0 missing)

REMOTE\_APP\_PACKETS < 29 to the right, improve= 6.658271, (0 missing)

Surrogate splits:

REMOTE\_APP\_PACKETS < 27.5 to the right, agree=0.933, adj=0.761, (0 split)

APP\_PACKETS < 32 to the right, agree=0.898, adj=0.634, (0 split)

CONTENT\_LENGTH < 16082.5 to the right, agree=0.752, adj=0.113, (0 split)

Node number 14: 11 observations

predicted class=0 expected loss=0 P(node) =0.008821171

class counts: 11 0

probabilities: 1.000 0.000

Node number 15: 45 observations

predicted class=1 expected loss=0.02222222 P(node) =0.03608661

class counts: 1 44

probabilities: 0.022 0.978

Node number 22: 71 observations

predicted class=0 expected loss=0.1549296 P(node) =0.05693665

class counts: 60 11

probabilities: 0.845 0.155

Node number 23: 183 observations, complexity param=0.03821656

predicted class=1 expected loss=0.4972678 P(node) =0.1467522

class counts: 91 92

probabilities: 0.497 0.503

left son=46 (27 obs) right son=47 (156 obs)

Primary splits:

CONTENT\_LENGTH < 6370 to the right, improve=11.639720, (0 missing)

URL\_LENGTH < 43.5 to the right, improve=10.168800, (0 missing)

APP\_PACKETS < 4.5 to the right, improve= 5.415673, (0 missing)

REMOTE\_IPS < 0.5 to the right, improve= 4.400777, (0 missing)

NUMBER\_SPECIAL\_CHARACTERS < 8.5 to the left, improve= 4.382292, (0 missing)

Surrogate splits:

SOURCE\_APP\_BYTES < 13338 to the right, agree=0.858, adj=0.037, (0 split)

Node number 46: 27 observations

predicted class=0 expected loss=0.07407407 P(node) =0.02165196

class counts: 25 2

probabilities: 0.926 0.074

Node number 47: 156 observations, complexity param=0.03821656

predicted class=1 expected loss=0.4230769 P(node) =0.1251002

class counts: 66 90

probabilities: 0.423 0.577

left son=94 (98 obs) right son=95 (58 obs)

Primary splits:

CONTENT\_LENGTH < 629.5 to the left, improve=13.253070, (0 missing)

URL\_LENGTH < 43.5 to the right, improve= 7.949195, (0 missing)

APP\_PACKETS < 27.5 to the left, improve= 5.076923, (0 missing)

NUMBER\_SPECIAL\_CHARACTERS < 8.5 to the left, improve= 4.940171, (0 missing)

SOURCE\_APP\_BYTES < 11610 to the left, improve= 3.419152, (0 missing)

Surrogate splits:

APP\_PACKETS < 23.5 to the left, agree=0.763, adj=0.362, (0 split)

REMOTE\_APP\_PACKETS < 19.5 to the left, agree=0.724, adj=0.259, (0 split)

SOURCE\_APP\_BYTES < 2794 to the left, agree=0.699, adj=0.190, (0 split)

URL\_LENGTH < 43.5 to the right, agree=0.692, adj=0.172, (0 split)

DIST\_REMOTE\_TCP\_PORT < 0.5 to the left, agree=0.673, adj=0.121, (0 split)

Node number 94: 98 observations, complexity param=0.03821656

predicted class=0 expected loss=0.4183673 P(node) =0.07858861

class counts: 57 41

probabilities: 0.582 0.418

left son=188 (84 obs) right son=189 (14 obs)

Primary splits:

APP\_PACKETS < 4.5 to the right, improve=8.503401, (0 missing)

SOURCE\_APP\_BYTES < 1272.5 to the right, improve=7.252498, (0 missing)

REMOTE\_IPS < 0.5 to the right, improve=6.790777, (0 missing)

REMOTE\_APP\_PACKETS < 5.5 to the right, improve=5.940829, (0 missing)

NUMBER\_SPECIAL\_CHARACTERS < 16.5 to the left, improve=3.243878, (0 missing)

Surrogate splits:

REMOTE\_IPS < 0.5 to the right, agree=0.980, adj=0.857, (0 split)

REMOTE\_APP\_PACKETS < 5.5 to the right, agree=0.980, adj=0.857, (0 split)

SOURCE\_APP\_BYTES < 449 to the right, agree=0.980, adj=0.857, (0 split)

URL\_LENGTH < 81.5 to the left, agree=0.918, adj=0.429, (0 split)

NUMBER\_SPECIAL\_CHARACTERS < 16.5 to the left, agree=0.918, adj=0.429, (0 split)

Node number 95: 58 observations, complexity param=0.02229299

predicted class=1 expected loss=0.1551724 P(node) =0.04651163

class counts: 9 49

probabilities: 0.155 0.845

left son=190 (21 obs) right son=191 (37 obs)

Primary splits:

CONTENT\_LENGTH < 979 to the right, improve=4.921182, (0 missing)

URL\_LENGTH < 48.5 to the right, improve=2.758717, (0 missing)

APP\_PACKETS < 20.5 to the left, improve=1.689655, (0 missing)

REMOTE\_IPS < 3.5 to the left, improve=1.525524, (0 missing)

NUMBER\_SPECIAL\_CHARACTERS < 8.5 to the left, improve=1.448563, (0 missing)

Surrogate splits:

REMOTE\_IPS < 2.5 to the left, agree=0.879, adj=0.667, (0 split)

SOURCE\_APP\_BYTES < 4338 to the right, agree=0.879, adj=0.667, (0 split)

DIST\_REMOTE\_TCP\_PORT < 0.5 to the left, agree=0.793, adj=0.429, (0 split)

NUMBER\_SPECIAL\_CHARACTERS < 10.5 to the left, agree=0.741, adj=0.286, (0 split)

URL\_LENGTH < 47 to the right, agree=0.724, adj=0.238, (0 split)

Node number 188: 84 observations, complexity param=0.02547771

predicted class=0 expected loss=0.3333333 P(node) =0.06736167

class counts: 56 28

probabilities: 0.667 0.333

left son=376 (57 obs) right son=377 (27 obs)

Primary splits:

URL\_LENGTH < 38.5 to the right, improve=2.729045, (0 missing)

REMOTE\_APP\_PACKETS < 8.5 to the left, improve=2.597101, (0 missing)

SOURCE\_APP\_BYTES < 672.5 to the left, improve=2.240000, (0 missing)

NUMBER\_SPECIAL\_CHARACTERS < 14.5 to the right, improve=1.964912, (0 missing)

APP\_PACKETS < 11.5 to the right, improve=1.555556, (0 missing)

Surrogate splits:

NUMBER\_SPECIAL\_CHARACTERS < 7.5 to the right, agree=0.869, adj=0.593, (0 split)

REMOTE\_IPS < 7.5 to the left, agree=0.702, adj=0.074, (0 split)

CONTENT\_LENGTH < 257 to the left, agree=0.690, adj=0.037, (0 split)

APP\_PACKETS < 24 to the left, agree=0.690, adj=0.037, (0 split)

Node number 189: 14 observations

predicted class=1 expected loss=0.07142857 P(node) =0.01122694

class counts: 1 13

probabilities: 0.071 0.929

Node number 190: 21 observations, complexity param=0.02229299

predicted class=1 expected loss=0.4285714 P(node) =0.01684042

class counts: 9 12

probabilities: 0.429 0.571

left son=380 (7 obs) right son=381 (14 obs)

Primary splits:

CONTENT\_LENGTH < 2174 to the left, improve=6.857143, (0 missing)

SOURCE\_APP\_BYTES < 5613.5 to the left, improve=5.151099, (0 missing)

APP\_PACKETS < 20.5 to the left, improve=4.122078, (0 missing)

REMOTE\_APP\_PACKETS < 21 to the left, improve=3.174603, (0 missing)

URL\_LENGTH < 38.5 to the left, improve=1.714286, (0 missing)

Surrogate splits:

SOURCE\_APP\_BYTES < 4962.5 to the left, agree=0.905, adj=0.714, (0 split)

URL\_LENGTH < 48.5 to the right, agree=0.762, adj=0.286, (0 split)

DIST\_REMOTE\_TCP\_PORT < 0.5 to the right, agree=0.762, adj=0.286, (0 split)

REMOTE\_IPS < 2.5 to the right, agree=0.762, adj=0.286, (0 split)

REMOTE\_APP\_PACKETS < 10.5 to the left, agree=0.762, adj=0.286, (0 split)

Node number 191: 37 observations

predicted class=1 expected loss=0 P(node) =0.02967121

class counts: 0 37

probabilities: 0.000 1.000

Node number 376: 57 observations

predicted class=0 expected loss=0.245614 P(node) =0.0457097

class counts: 43 14

probabilities: 0.754 0.246

Node number 377: 27 observations, complexity param=0.02547771

predicted class=1 expected loss=0.4814815 P(node) =0.02165196

class counts: 13 14

probabilities: 0.481 0.519

left son=754 (17 obs) right son=755 (10 obs)

Primary splits:

NUMBER\_SPECIAL\_CHARACTERS < 7.5 to the left, improve=4.6226580, (0 missing)

SOURCE\_APP\_BYTES < 1399.5 to the right, improve=1.5167760, (0 missing)

REMOTE\_IPS < 2.5 to the right, improve=0.8990639, (0 missing)

REMOTE\_APP\_PACKETS < 10.5 to the right, improve=0.4704925, (0 missing)

URL\_LENGTH < 25 to the left, improve=0.4683236, (0 missing)

Surrogate splits:

URL\_LENGTH < 30.5 to the left, agree=0.741, adj=0.3, (0 split)

SOURCE\_APP\_BYTES < 1112.5 to the right, agree=0.704, adj=0.2, (0 split)

CONTENT\_LENGTH < 190.5 to the right, agree=0.667, adj=0.1, (0 split)

REMOTE\_APP\_PACKETS < 13.5 to the left, agree=0.667, adj=0.1, (0 split)

Node number 380: 7 observations

predicted class=0 expected loss=0 P(node) =0.005613472

class counts: 7 0

probabilities: 1.000 0.000

Node number 381: 14 observations

predicted class=1 expected loss=0.1428571 P(node) =0.01122694

class counts: 2 12

probabilities: 0.143 0.857

Node number 754: 17 observations

predicted class=0 expected loss=0.2941176 P(node) =0.01363272

class counts: 12 5

probabilities: 0.706 0.294

Node number 755: 10 observations

predicted class=1 expected loss=0.1 P(node) =0.008019246

class counts: 1 9

probabilities: 0.100 0.900