Last

Daniel Sandoz

Part 2 A

Step 1: Develop an understanding of the data mining project

In this project, we have a data set that is obtained from the department of transportation and we are trying to perform a data mining analysisthat predicts the number of fatalities that will result from a given accident. There are many variables in this data set, so identifying the main predictors will be important.

Step 2: Obtain the dataset to be used in the analysis

The data set is found here https://www.kaggle.com/usdot/nhtsa-traffic-fatalities

Step 3: Explore, clean, and preprocess the data & Step 4: Reduce the data dimension

I combined these two steps. I viewed the structure of the data and the number of variables involved in this data set. I then went on to delete the variables that were either irrelevant to the business question, had a large amount of missing values or could be explained better by other more significant variables. In order to run a confusion matrix, I also turned the variables into factors, so I would be able to gage the accuracy using a confusion matrix. I then brought the target variable to the front to allow for a classification model to be ran on the data set. In this data set, we started with 100 variables. It will be important to identify key predictors to allow for us to save our processing resources.

Step 5: Determine the data mining task

Predict the severity of injury to a person who was involved in an accident

Step 6: Partition the Data

For Random Forest I partitioned the data into an 80/20 split as standard for that algorithm.

Step 7: Choose the data mining techniques to be used

I decided that Random Forest would be a good technique to allow for me to understand what variables contribute to the ability to predict severity of injury, I then revisited step 4 and eliminated the variables that had a small or no impact on the model to allow for the least amount of processing resources to be used.

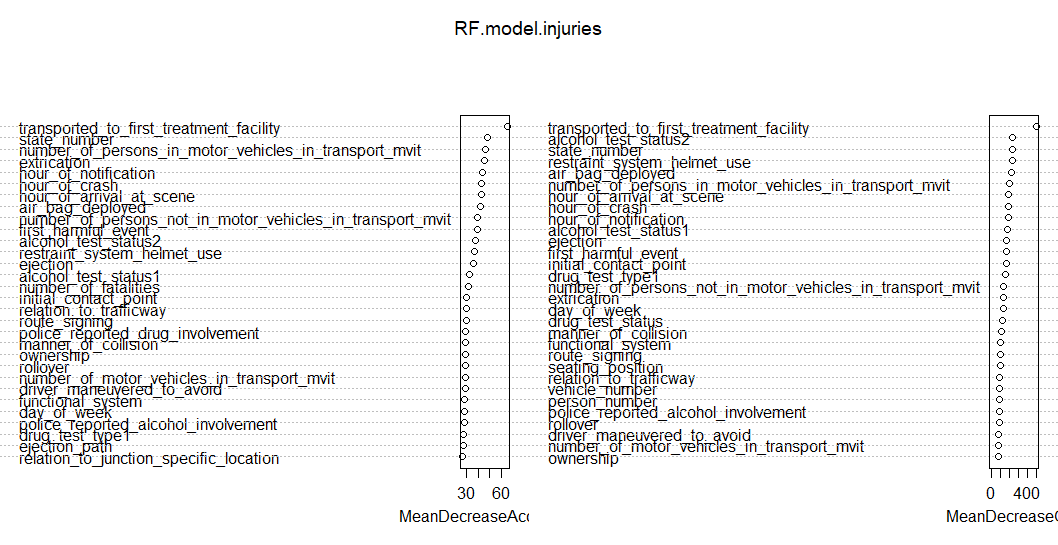
Step 8: Use algorithms to perform the task

I used Random Forest to understand what variables are important to predict the severity of injury.

Step 9: Interpret the results

With Random Forest we can now see what variables help contribute to the accuracy of predicting the severity of bodily injury. From this chart we can see that there are many variables that offer little to no explanation of severity of bodily injury, so I took them out and re-ran the algorithm. It took less time to run however with a large data set like we have, this will still take up a solid amount of processing resources. From this graph we can see that air bag deployed and the number of people in the motor vehicle in transport is a very good predictor variable when predicting the severity of bodily injury. With an accuracy of .7192 after taking out many variables, we can consider this model decently accurate based on the amount of variation that is present in data sets like this.

The Random Forest results are attached at the bottom.



Step 10: Deploy the model

Random Forest Confusion Matrix

|  |
| --- |
| Accuracy : 0.7192  95% CI : (0.7013, 0.7366)  No Information Rate : 0.4301  P-Value [Acc > NIR] : < 2.2e-16    Kappa : 0.5947  Mcnemar's Test P-Value : NA  Statistics by Class:  Class: 0 Class: 1 Class: 2 Class: 3 Class: 4 Class: 5 Class: 9  Sensitivity 0.9126 0.27309 0.47432 0.27925 0.9167 0.1111111 0.142857  Specificity 0.9288 0.98908 0.92844 0.96614 0.7844 1.0000000 0.999603  Pos Pred Value 0.7885 0.73118 0.49841 0.49007 0.7624 1.0000000 0.750000  Neg Pred Value 0.9734 0.92600 0.92176 0.92002 0.9258 0.9968479 0.992899  Prevalence 0.2253 0.09807 0.13037 0.10437 0.4301 0.0035447 0.008271  Detection Rate 0.2056 0.02678 0.06184 0.02915 0.3942 0.0003939 0.001182  Detection Prevalence 0.2607 0.03663 0.12406 0.05947 0.5171 0.0003939 0.001575  Balanced Accuracy 0.9207 0.63109 0.70138 0.62269 0.8505 0.5555556 0.571230 |
|  |
| |  | | --- | |  | |

Part 2 B

Air bag deployed, and initial contact point seem to be the biggest predictors when targeting the severity of injury. Initial contact point is described as the area on this vehicle that produced the first instance of injury to non-motorists or occupants of this vehicle, or that resulted in the first instance of damage to other property or to this vehicle. This model has an accuracy of .4353 and is not as accurate as the random forest model. This is due to the nature of the complexity of the random forest model and its ability to show how influential a certain variable is on the target variable. When I took out the noninfluential variables in the random forest model, it increased the accuracy because there was less variance in the data set. However, air bag deployed , and initial contact point seem to be very good predictors from a practical standpoint. If airbags are deployed, that limits the space between the driver/passengers and their surroundings. This allows for a smaller amount of acceleration to take place due to the smaller distance which results in less injuries and deaths. Contact point also seems like a practical predictor, as head to head collisions or t-bones are more violent then a side swipe or rear ending someone.

initial\_contact\_point

Y [,1] [,2]

0 16.77756 21.57918

1 14.92890 19.82378

2 16.46419 22.42201

3 15.13319 21.39560

4 14.76601 23.20864

5 10.52941 16.10051

9 27.21333 34.16862

air\_bag\_deployed

Y [,1] [,2]

0 17.327182 19.96980

1 11.403670 19.18147

2 9.839517 18.49964

3 13.471536 25.90566

4 23.945621 36.83174

5 13.676471 27.41126

9 73.186667 41.53269

Naïve Bayes RF

Overall Statistics

Accuracy : 0.4353

95% CI : (0.4195, 0.4512)

No Information Rate : 0.4303

P-Value [Acc > NIR] : 0.2723

Kappa : 0.3249

Mcnemar's Test P-Value : <2e-16

Statistics by Class:

Class: 0 Class: 1 Class: 2 Class: 3 Class: 4 Class: 5 Class: 9

Sensitivity 0.8089 0.11796 0.09476 0.12814 0.4838 0.357143 0.774194

Specificity 0.9024 0.94761 0.97706 0.98769 0.9977 0.855599 0.732133

Pos Pred Value 0.7067 0.19643 0.38211 0.54839 0.9937 0.009042 0.023166

Neg Pred Value 0.9420 0.90823 0.87819 0.90662 0.7190 0.997236 0.997476

Prevalence 0.2253 0.09793 0.13022 0.10449 0.4303 0.003676 0.008139

Detection Rate 0.1822 0.01155 0.01234 0.01339 0.2082 0.001313 0.006301

Detection Prevalence 0.2578 0.05881 0.03229 0.02442 0.2095 0.145182 0.271987

Balanced Accuracy 0.8556 0.53279 0.53591 0.55791 0.7408 0.606371 0.753163

Random Forest Results:

Confusion matrix:

0 1 2 3 4 5 9 class.error

0 2111 12 4 2 159 0 3 0.07856831

1 249 276 221 49 201 0 0 0.72289157

2 130 66 653 113 362 0 0 0.50679758

3 10 44 296 317 397 0 0 0.70206767

4 98 30 113 90 4041 0 0 0.07570906

5 3 0 14 5 16 1 0 0.97435897

9 50 2 1 1 13 0 18 0.78823529

> print(RF.model.injuries)

Call:

randomForest(formula = injury\_severity ~ ., data = injury\_train.set, mtry = 3, ntree = 1000, importance = TRUE, na.action = na.omit)

Type of random forest: classification

Number of trees: 1000

No. of variables tried at each split: 3

OOB estimate of error rate: 27.08%

Confusion matrix:

0 1 2 3 4 5 9 class.error

0 2111 12 4 2 159 0 3 0.07856831

1 249 276 221 49 201 0 0 0.72289157

2 130 66 653 113 362 0 0 0.50679758

3 10 44 296 317 397 0 0 0.70206767

4 98 30 113 90 4041 0 0 0.07570906

5 3 0 14 5 16 1 0 0.97435897

9 50 2 1 1 13 0 18 0.78823529

> importance(RF.model.injuries)

0 1 2

state\_number 30.3186888 31.5223925 27.4674082

vehicle\_number 21.3647698 22.0295867 20.1848280

person\_number 12.2640400 6.5422020 11.9554780

number\_of\_motor\_vehicle\_striking\_non\_motorist 9.5818600 6.5863359 7.8710380

vehicle\_trailing 18.2949581 11.5245452 10.3501195

special\_use 9.6596233 20.5611276 17.9640976

emergency\_motor\_vehicle\_use -0.6983641 1.7341382 5.4957487

rollover 32.0121891 21.2866288 18.6704048

initial\_contact\_point 28.7041720 26.0494185 24.7864141

fire\_occurrence 20.8074763 13.3998854 12.6271700

sex 0.6885213 5.6089257 1.8960691

person\_type 19.9668103 15.9308098 18.8259454

seating\_position 20.5993254 11.9504076 17.2640296

restraint\_system\_helmet\_use 35.7221029 30.6277696 24.7430849

indication\_of\_misuse\_of\_restraint\_system\_helmet 10.3548315 10.0782360 8.2874549

air\_bag\_deployed 41.1305671 16.9510264 20.8886479

ejection 32.4250221 31.9289754 32.5525106

ejection\_path 24.5722232 20.4836362 16.1279355

extrication 38.6101369 32.3937914 22.8735987

police\_reported\_alcohol\_involvement 23.9763933 19.9270515 12.2379673

method\_of\_alcohol\_determination\_by\_police 13.1854359 10.6121828 10.6865786

alcohol\_test\_status1 23.5734941 21.6201662 25.5928319

alcohol\_test\_status2 27.9565578 25.7871077 27.9193691

police\_reported\_drug\_involvement 23.6781810 18.0453706 11.4223796

method\_of\_drug\_determination\_by\_police 8.0624207 7.6951810 10.8294147

drug\_test\_status 20.1252427 18.8328852 20.3869532

drug\_test\_type1 21.9133993 19.0415763 21.3139863

drug\_test\_type2 11.8824306 8.1513126 9.1206691

drug\_test\_type3 8.5467415 6.4076929 5.8303716

drug\_test\_type5 11.0784854 6.8773021 9.1272329

drug\_test\_type6 8.0279778 3.8322411 5.2095575

transported\_to\_first\_treatment\_facility 64.2401674 41.5947892 56.9671457

related\_factors\_person\_level1 3.4910578 0.6181382 0.3321253

related\_factors\_person\_level2 -1.0005004 0.0000000 1.0005004

related\_factors\_person\_level3 1.4150621 0.0000000 0.0000000

non\_motorist\_location\_at\_time\_of\_crash 9.9886687 8.7594491 4.8608847

driver\_maneuvered\_to\_avoid 19.2420882 19.3889136 14.4458140

number\_of\_motor\_vehicles\_in\_transport\_mvit 17.7295809 20.7122574 16.1555547

number\_of\_parked\_working\_vehicles 3.7918956 4.1829427 6.9069879

number\_of\_persons\_not\_in\_motor\_vehicles\_in\_transport\_mvit 33.2384991 9.7125912 19.6437855

number\_of\_persons\_in\_motor\_vehicles\_in\_transport\_mvit 25.8995429 30.7575656 29.2258502

year\_of\_crash 0.0000000 0.0000000 0.0000000

day\_of\_week 18.2337981 25.5281942 20.7601235

hour\_of\_crash 31.3263316 26.9134946 24.7520822

national\_highway\_system 11.8784770 14.8638759 12.4933950

land\_use 20.6684365 18.8914612 16.1543021

functional\_system 18.6587234 20.8413869 17.1920355

ownership 16.3517135 19.6775326 16.5142211

route\_signing 17.4645203 22.4002115 18.7054806

first\_harmful\_event 31.2628856 24.1254063 20.8735271

manner\_of\_collision 21.5018773 23.2950823 20.1200784

relation\_to\_junction\_within\_interchange\_area 3.8127671 12.7968993 5.9641967

relation\_to\_junction\_specific\_location 15.4380472 18.7645777 15.9078181

type\_of\_intersection 10.5245053 14.5537812 11.3070177

work\_zone 1.5117700 12.1871953 5.9341533

relation\_to\_trafficway 27.2421990 16.0257052 16.8836412

light\_condition 15.2646589 15.9280910 14.6303716

atmospheric\_conditions 14.0886452 15.9991499 12.5628555

school\_bus\_related 5.6103326 5.3093768 13.9632401

hour\_of\_notification 31.5869922 27.5774863 25.4715889

hour\_of\_arrival\_at\_scene 29.4536959 27.0923258 24.9390595

related\_factors\_crash\_level\_1 8.9120178 13.0052585 11.0046040

related\_factors\_crash\_level\_2 1.4243017 6.3082047 2.1380932

related\_factors\_crash\_level\_3 2.5984477 2.1902220 2.3814382

number\_of\_fatalities 22.9927946 19.0988754 15.2705811

number\_of\_drunk\_drivers 15.0024867 13.1241242 11.2326974

3 4

state\_number 17.6945391 19.7463120

vehicle\_number 7.2196767 18.3584676

person\_number 17.7258152 18.7274901

number\_of\_motor\_vehicle\_striking\_non\_motorist 6.5308609 12.8706444

vehicle\_trailing 13.3843402 17.9070109

special\_use 1.4590648 12.9810833

emergency\_motor\_vehicle\_use 3.2776781 4.6658469

rollover 8.0517939 12.8524335

initial\_contact\_point 16.9420800 17.2575567

fire\_occurrence -1.6243718 11.5280135

sex -0.3357652 13.3732232

person\_type 14.2516381 15.7749779

seating\_position 11.1055140 14.9880974

restraint\_system\_helmet\_use 13.6255767 26.7380622

indication\_of\_misuse\_of\_restraint\_system\_helmet 6.6811826 10.7232360

air\_bag\_deployed 22.8047839 23.5216253

ejection 8.8559614 26.5801956

ejection\_path 3.3448078 20.3339681

extrication 16.6507279 42.3691912

police\_reported\_alcohol\_involvement 14.5547960 11.4238276

method\_of\_alcohol\_determination\_by\_police 7.2199964 13.6876154

alcohol\_test\_status1 22.5298537 27.6610321

alcohol\_test\_status2 27.2051001 30.3684639

police\_reported\_drug\_involvement 10.2201584 10.8525813

method\_of\_drug\_determination\_by\_police 4.5965625 7.7151557

drug\_test\_status 16.3400787 22.2192404

drug\_test\_type1 18.4960544 23.0804907

drug\_test\_type2 6.3705395 -1.0523613

drug\_test\_type3 1.4294524 0.2077402

drug\_test\_type5 6.5530631 -2.7110579

drug\_test\_type6 4.4388776 -4.5175326

transported\_to\_first\_treatment\_facility 58.3093119 37.5954903

related\_factors\_person\_level1 4.1831403 -4.7673743

related\_factors\_person\_level2 0.0000000 -1.7343856

related\_factors\_person\_level3 0.0000000 -1.7345339

non\_motorist\_location\_at\_time\_of\_crash 3.7887866 13.3804382

driver\_maneuvered\_to\_avoid 15.1618460 11.6347292

number\_of\_motor\_vehicles\_in\_transport\_mvit 13.1106143 23.4130927

number\_of\_parked\_working\_vehicles 3.4600373 2.0731355

number\_of\_persons\_not\_in\_motor\_vehicles\_in\_transport\_mvit 17.8041022 32.7744744

number\_of\_persons\_in\_motor\_vehicles\_in\_transport\_mvit 25.7272037 39.2872837

year\_of\_crash 0.0000000 0.0000000

day\_of\_week 11.6821600 -1.7523280

hour\_of\_crash 12.5141842 10.2959353

national\_highway\_system 8.1611978 3.4376041

land\_use 10.3670734 9.6186362

functional\_system 10.9004126 3.9976407

ownership 10.0553460 9.7073363

route\_signing 10.5449916 7.5496863

first\_harmful\_event 18.9007922 28.2023315

manner\_of\_collision 18.5489646 19.8530996

relation\_to\_junction\_within\_interchange\_area 1.9487335 1.1485464

relation\_to\_junction\_specific\_location 7.1878032 17.3307592

type\_of\_intersection 9.0241919 13.5086578

work\_zone 1.5788155 -0.6887963

relation\_to\_trafficway 12.9718925 23.3500951

light\_condition 7.4674922 9.2003146

atmospheric\_conditions 11.1566184 1.5684465

school\_bus\_related -0.2005630 4.2519182

hour\_of\_notification 12.6979380 9.2540985

hour\_of\_arrival\_at\_scene 12.7307448 10.4180928

related\_factors\_crash\_level\_1 8.1774046 11.2652377

related\_factors\_crash\_level\_2 5.8787864 1.8633753

related\_factors\_crash\_level\_3 3.7733734 0.9919764

number\_of\_fatalities 6.6297984 29.4764292

number\_of\_drunk\_drivers 8.3299129 7.6901582

5 9

state\_number 13.19586999 11.23612042

vehicle\_number 6.16446087 12.20466796

person\_number 0.17879947 5.74354374

number\_of\_motor\_vehicle\_striking\_non\_motorist 3.48886069 5.62018517

vehicle\_trailing 3.26195980 8.14262336

special\_use 0.00000000 10.78576737

emergency\_motor\_vehicle\_use 0.00000000 9.85216434

rollover 5.70374601 14.62758929

initial\_contact\_point 9.23463108 11.95993536

fire\_occurrence 5.55143528 8.85049449

sex -2.60704291 23.25046656

person\_type 2.04654298 10.65475340

seating\_position 2.52747357 12.87464123

restraint\_system\_helmet\_use 7.67265780 22.89293073

indication\_of\_misuse\_of\_restraint\_system\_helmet 4.56236639 5.08949943

air\_bag\_deployed 8.80538201 21.73199108

ejection 5.60958322 14.66822863

ejection\_path 4.43441990 9.40099970

extrication 5.58271117 11.18998604

police\_reported\_alcohol\_involvement 3.30808664 9.95220389

method\_of\_alcohol\_determination\_by\_police 1.39008642 5.31409794

alcohol\_test\_status1 7.27110480 11.84020194

alcohol\_test\_status2 4.57755438 14.29586032

police\_reported\_drug\_involvement 3.71748590 8.96900710

method\_of\_drug\_determination\_by\_police -0.01623222 2.52581520

drug\_test\_status 5.63094682 11.76391122

drug\_test\_type1 4.92173486 12.58979714

drug\_test\_type2 1.98199270 4.31016281

drug\_test\_type3 1.39404687 3.63734894

drug\_test\_type5 1.73386765 4.79555222

drug\_test\_type6 1.00050038 2.35529062

transported\_to\_first\_treatment\_facility 7.18789101 19.38306591

related\_factors\_person\_level1 0.00000000 6.90385605

related\_factors\_person\_level2 0.00000000 0.00000000

related\_factors\_person\_level3 0.00000000 0.00000000

non\_motorist\_location\_at\_time\_of\_crash 5.78564405 5.95239224

driver\_maneuvered\_to\_avoid 6.94836345 15.07157732

number\_of\_motor\_vehicles\_in\_transport\_mvit 8.46100812 11.17949025

number\_of\_parked\_working\_vehicles 1.00050038 5.80879207

number\_of\_persons\_not\_in\_motor\_vehicles\_in\_transport\_mvit 7.86058281 17.36283004

number\_of\_persons\_in\_motor\_vehicles\_in\_transport\_mvit 10.87279736 15.63418120

year\_of\_crash 0.00000000 0.00000000

day\_of\_week 8.36168841 8.26221629

hour\_of\_crash 9.66433433 8.22916974

national\_highway\_system 5.07503625 5.23548902

land\_use 6.19862486 9.95445741

functional\_system 8.40292787 7.91400773

ownership 5.63803597 4.16717779

route\_signing 9.10551547 2.62658615

first\_harmful\_event 10.22393017 16.44888097

manner\_of\_collision 6.50538097 14.17865060

relation\_to\_junction\_within\_interchange\_area 0.76754615 0.05263177

relation\_to\_junction\_specific\_location 3.92721881 4.58621892

type\_of\_intersection 4.62121015 3.68154506

work\_zone -1.00050038 6.93873752

relation\_to\_trafficway 8.68377564 9.02354562

light\_condition 7.61683513 5.37029825

atmospheric\_conditions 6.65265434 5.50105338

school\_bus\_related 0.00000000 1.73123730

hour\_of\_notification 9.61427180 9.56202048

hour\_of\_arrival\_at\_scene 9.06549665 8.13292726

related\_factors\_crash\_level\_1 1.97735506 4.32583139

related\_factors\_crash\_level\_2 1.00050038 6.19969825

related\_factors\_crash\_level\_3 0.00000000 -1.21707005

number\_of\_fatalities 2.59088820 9.41730322

number\_of\_drunk\_drivers 7.13113335 9.33873528

MeanDecreaseAccuracy

state\_number 48.1256712

vehicle\_number 26.2796611

person\_number 23.5276031

number\_of\_motor\_vehicle\_striking\_non\_motorist 13.5081795

vehicle\_trailing 20.9035338

special\_use 24.6656443

emergency\_motor\_vehicle\_use 10.2470850

rollover 28.8184775

initial\_contact\_point 30.2945182

fire\_occurrence 16.9256789

sex 17.8092958

person\_type 23.7476567

seating\_position 24.1793494

restraint\_system\_helmet\_use 37.2669851

indication\_of\_misuse\_of\_restraint\_system\_helmet 11.9183091

air\_bag\_deployed 41.8790854

ejection 35.7128455

ejection\_path 27.2051404

extrication 45.8897819

police\_reported\_alcohol\_involvement 28.1045582

method\_of\_alcohol\_determination\_by\_police 22.5125816

alcohol\_test\_status1 32.1275838

alcohol\_test\_status2 37.3431147

police\_reported\_drug\_involvement 29.3855084

method\_of\_drug\_determination\_by\_police 15.5655976

drug\_test\_status 26.3026489

drug\_test\_type1 27.2812710

drug\_test\_type2 12.4166432

drug\_test\_type3 9.3064620

drug\_test\_type5 9.2997114

drug\_test\_type6 4.5726789

transported\_to\_first\_treatment\_facility 66.0602005

related\_factors\_person\_level1 1.4380468

related\_factors\_person\_level2 -1.3547764

related\_factors\_person\_level3 -0.5669554

non\_motorist\_location\_at\_time\_of\_crash 14.8213233

driver\_maneuvered\_to\_avoid 28.6991664

number\_of\_motor\_vehicles\_in\_transport\_mvit 28.7357087

number\_of\_parked\_working\_vehicles 8.6735158

number\_of\_persons\_not\_in\_motor\_vehicles\_in\_transport\_mvit 39.3522142

number\_of\_persons\_in\_motor\_vehicles\_in\_transport\_mvit 46.0723540

year\_of\_crash 0.0000000

day\_of\_week 28.2076190

hour\_of\_crash 43.3428286

national\_highway\_system 19.2760064

land\_use 25.9723874

functional\_system 28.5394909

ownership 28.8763106

route\_signing 29.5671524

first\_harmful\_event 39.0455298

manner\_of\_collision 28.8779924

relation\_to\_junction\_within\_interchange\_area 11.2322616

relation\_to\_junction\_specific\_location 26.4756919

type\_of\_intersection 20.1315615

work\_zone 10.6313061

relation\_to\_trafficway 30.0238234

light\_condition 22.7747921

atmospheric\_conditions 22.5594158

school\_bus\_related 14.5096387

hour\_of\_notification 43.9494302

hour\_of\_arrival\_at\_scene 42.8668025

related\_factors\_crash\_level\_1 21.1069370

related\_factors\_crash\_level\_2 8.1653520

related\_factors\_crash\_level\_3 4.3638763

number\_of\_fatalities 31.8780846

number\_of\_drunk\_drivers 18.8258033

MeanDecreaseGini

state\_number 233.3532170

vehicle\_number 88.8932737

person\_number 85.9139305

number\_of\_motor\_vehicle\_striking\_non\_motorist 18.6861564

vehicle\_trailing 53.4779811

special\_use 16.5280532

emergency\_motor\_vehicle\_use 3.0737893

rollover 83.1804536

initial\_contact\_point 166.1794527

fire\_occurrence 33.5750386

sex 50.7941423

person\_type 67.7622762

seating\_position 99.5644083

restraint\_system\_helmet\_use 230.2360792

indication\_of\_misuse\_of\_restraint\_system\_helmet 20.7177219

air\_bag\_deployed 222.0853816

ejection 170.3572780

ejection\_path 52.0917953

extrication 134.9278567

police\_reported\_alcohol\_involvement 83.2844724

method\_of\_alcohol\_determination\_by\_police 44.3391842

alcohol\_test\_status1 177.0859848

alcohol\_test\_status2 235.2405576

police\_reported\_drug\_involvement 69.0294032

method\_of\_drug\_determination\_by\_police 27.6507175

drug\_test\_status 127.2269973

drug\_test\_type1 150.2409623

drug\_test\_type2 16.1825809

drug\_test\_type3 6.7929400

drug\_test\_type5 21.2972739

drug\_test\_type6 9.0606472

transported\_to\_first\_treatment\_facility 508.1796689

related\_factors\_person\_level1 2.3853859

related\_factors\_person\_level2 0.1581558

related\_factors\_person\_level3 0.1778279

non\_motorist\_location\_at\_time\_of\_crash 26.0560989

driver\_maneuvered\_to\_avoid 82.2243932

number\_of\_motor\_vehicles\_in\_transport\_mvit 78.5764963

number\_of\_parked\_working\_vehicles 12.1386646

number\_of\_persons\_not\_in\_motor\_vehicles\_in\_transport\_mvit 136.5177862

number\_of\_persons\_in\_motor\_vehicles\_in\_transport\_mvit 202.8265070

year\_of\_crash 0.0000000

day\_of\_week 129.7780881

hour\_of\_crash 186.1744931

national\_highway\_system 30.9684569

land\_use 43.6373947

functional\_system 100.8024063

ownership 72.6478272

route\_signing 100.5022517

first\_harmful\_event 168.9554804

manner\_of\_collision 105.1431542

relation\_to\_junction\_within\_interchange\_area 11.5642610

relation\_to\_junction\_specific\_location 68.4468942

type\_of\_intersection 44.9490794

work\_zone 10.7951877

relation\_to\_trafficway 92.4084741

light\_condition 61.3492277

atmospheric\_conditions 67.8375649

school\_bus\_related 4.7417288

hour\_of\_notification 185.5973193

hour\_of\_arrival\_at\_scene 188.9133524

related\_factors\_crash\_level\_1 34.5029917

related\_factors\_crash\_level\_2 7.4129107

related\_factors\_crash\_level\_3 2.6523697

number\_of\_fatalities 51.1527870

number\_of\_drunk\_drivers 35.6418765

NAÏVE BAYS RESULTS:

Conditional probabilities:

vehicle\_number

Y [,1] [,2]

0 1.645387 1.0224208

1 1.653670 0.8323319

2 1.550475 0.8126787

3 1.448980 0.6974042

4 1.061961 0.6424479

5 1.588235 0.9571944

9 1.373333 0.7492940

person\_number

Y [,1] [,2]

0 1.617456 1.4109617

1 2.064220 1.9465898

2 2.365833 3.5946236

3 1.952739 1.2775651

4 1.305098 1.0431274

5 1.676471 0.9118941

9 1.400000 1.0654272

number\_of\_motor\_vehicle\_striking\_non\_motorist

Y [,1] [,2]

0 0.001496259 0.0386622

1 0.013761468 0.1430954

2 0.014667817 0.1272488

3 0.046186896 0.3128330

4 0.160000000 0.3744173

5 0.058823529 0.2388326

9 0.000000000 0.0000000

rollover

Y [,1] [,2]

0 0.02892768 0.1676451

1 0.19724771 0.3981493

2 0.26488352 0.4414617

3 0.28678840 0.4525050

4 0.38692810 0.4871106

5 0.23529412 0.4305615

9 0.04000000 0.1972788

initial\_contact\_point

Y [,1] [,2]

0 16.77756 21.57918

1 14.92890 19.82378

2 16.46419 22.42201

3 15.13319 21.39560

4 14.76601 23.20864

5 10.52941 16.10051

9 27.21333 34.16862

fire\_occurrence

Y [,1] [,2]

0 0.002493766 0.04988775

1 0.012614679 0.11166848

2 0.020707506 0.14246479

3 0.042964554 0.20288621

4 0.040000000 0.19598480

5 0.000000000 0.00000000

9 0.000000000 0.00000000

age

Y [,1] [,2]

0 68.69277 173.97454

1 37.56078 60.06069

2 36.31320 59.90251

3 42.25671 79.44763

4 46.02954 50.79010

5 120.08824 278.04309

9 663.50667 463.71693

sex

Y [,1] [,2]

0 0.6528678 0.4761770

1 0.5699541 0.4953664

2 0.5685936 0.4954864

3 0.5757250 0.4944981

4 0.6928105 0.4613890

5 0.5882353 0.4995542

9 0.2933333 0.4583559

seating\_position

Y [,1] [,2]

0 14.61746 11.207968

1 16.35550 11.896984

2 16.13978 11.205852

3 15.65843 12.351108

4 11.02693 9.131395

5 13.44118 5.711186

9 19.05333 23.893284

indication\_of\_misuse\_of\_restraint\_system\_helmet

Y [,1] [,2]

0 0 0

1 0 0

2 0 0

3 0 0

4 0 0

5 0 0

9 0 0

air\_bag\_deployed

Y [,1] [,2]

0 17.327182 19.96980

1 11.403670 19.18147

2 9.839517 18.49964

3 13.471536 25.90566

4 23.945621 36.83174

5 13.676471 27.41126

9 73.186667 41.53269

ejection

Y [,1] [,2]

0 0.1511222 1.064115

1 0.1949541 1.219276

2 0.3218292 1.498116

3 0.8775510 2.353684

4 2.6339869 3.572676

5 1.2352941 2.944526

9 0.6133333 2.104650

ejection\_path

Y [,1] [,2]

0 0.01296758 0.3355685

1 0.07224771 0.8035858

2 0.26229508 1.4875018

3 0.83136412 2.5582513

4 1.24235294 2.9943435

5 0.02941176 0.1714986

9 0.00000000 0.0000000

extrication

Y [,1] [,2]

0 0.003491272 0.05899847

1 0.027522936 0.16369531

2 0.064710958 0.24612134

3 0.194414608 0.39596209

4 0.227189542 0.41907084

5 0.088235294 0.28790224

9 0.013333333 0.11547005

police\_reported\_alcohol\_involvement

Y [,1] [,2]

0 0.04189526 0.2003998

1 0.03555046 0.1852727

2 0.06125971 0.2399096

3 0.07196563 0.2585699

4 0.13725490 0.3441612

5 0.29411765 0.4624973

9 0.02666667 0.1621922

police\_reported\_drug\_involvement

Y [,1] [,2]

0 0.02643392 0.1604618

1 0.01834862 0.1342856

2 0.04745470 0.2127012

3 0.03437164 0.1822798

4 0.07346405 0.2609308

5 0.02941176 0.1714986

9 0.01333333 0.1154701

transported\_to\_first\_treatment\_facility

Y [,1] [,2]

0 0.000000 0.000000

1 3.564220 2.428402

2 4.162209 1.822784

3 4.102041 1.670524

4 2.181438 2.431170

5 4.000000 1.890967

9 1.360000 2.845765

race

Y [,1] [,2]

0 0.000000 0.00000

1 0.000000 0.00000

2 0.000000 0.00000

3 0.000000 0.00000

4 9.850458 27.25029

5 0.000000 0.00000

9 0.000000 0.00000

driver\_maneuvered\_to\_avoid

Y [,1] [,2]

0 38.75411 47.24662

1 43.48739 48.07225

2 50.50733 48.50775

3 51.86896 48.48221

4 43.55216 48.46622

5 12.08824 32.14679

9 66.04000 46.20063

number\_of\_vehicle\_forms\_submitted\_all

Y [,1] [,2]

0 2.099751 1.3219417

1 2.253440 1.2463447

2 2.039689 1.0415273

3 1.980666 0.9249539

4 1.614902 0.8094872

5 2.058824 1.1265710

9 1.760000 1.1722005

number\_of\_parked\_working\_vehicles

Y [,1] [,2]

0 0.04588529 0.2978454

1 0.03899083 0.2639328

2 0.03106126 0.1878930

3 0.03222342 0.2644396

4 0.03503268 0.2943442

5 0.05882353 0.3429972

9 0.05333333 0.2262105

number\_of\_forms\_submitted\_for\_persons\_not\_in\_motor\_vehicles

Y [,1] [,2]

0 0.40748130 0.6231110

1 0.07798165 0.3532647

2 0.08714409 0.4497154

3 0.10955961 0.5928108

4 0.18248366 0.4643759

5 0.17647059 0.7164977

9 0.61333333 0.5170951

number\_of\_persons\_not\_in\_motor\_vehicles\_in\_transport\_mvit

Y [,1] [,2]

0 0.4389027 0.6811423

1 0.0940367 0.4235992

2 0.1138913 0.5583147

3 0.1181525 0.6090741

4 0.1911111 0.4811558

5 0.1764706 0.7164977

9 0.7066667 0.6930803

number\_of\_persons\_in\_motor\_vehicles\_in\_transport\_mvit

Y [,1] [,2]

0 3.848379 3.374046

1 5.069954 4.039520

2 5.168248 5.552248

3 4.288937 2.851386

4 2.548235 2.284850

5 3.470588 1.727413

9 2.746667 3.145539

number\_of\_forms\_submitted\_for\_persons\_in\_motor\_vehicles

Y [,1] [,2]

0 3.879800 3.373907

1 5.086009 4.030388

2 5.194996 5.540706

3 4.297530 2.850502

4 2.556863 2.287397

5 3.470588 1.727413

9 2.840000 3.119425

month\_of\_crash

Y [,1] [,2]

0 6.826434 3.284695

1 7.066514 3.310239

2 6.914582 3.181953

3 6.771214 3.224948

4 6.722092 3.309583

5 6.852941 3.173110

9 7.533333 3.402437

national\_highway\_system

Y [,1] [,2]

0 0.4723192 0.5341220

1 0.4931193 0.5002396

2 0.4495255 0.7045358

3 0.4371643 0.4963026

4 0.4005229 0.5781949

5 0.5000000 0.5075192

9 0.4666667 0.5022472

land\_use

Y [,1] [,2]

0 1.655860 0.5688801

1 1.516055 0.5440158

2 1.460742 0.5321801

3 1.411386 0.5137249

4 1.490458 0.5629564

5 1.500000 0.5075192

9 1.640000 0.4832324

ownership

Y [,1] [,2]

0 17.451870 35.53089

1 16.769495 34.99776

2 16.498706 34.72798

3 19.678840 37.63242

4 17.105359 35.11400

5 4.205882 16.76898

9 17.186667 35.56368

manner\_of\_collision

Y [,1] [,2]

0 2.792519 6.381176

1 3.426606 5.388394

2 3.177739 6.258630

3 3.029001 6.048213

4 2.057516 5.571646

5 2.470588 2.765976

9 2.533333 11.555195

relation\_to\_junction\_within\_interchange\_area

Y [,1] [,2]

0 0.04438903 0.2060092

1 0.05045872 0.2190152

2 0.04745470 0.2127012

3 0.04511278 0.2076630

4 0.03790850 0.1909999

5 0.11764706 0.3270350

9 0.02666667 0.1621922

light\_condition

Y [,1] [,2]

0 1.769576 1.0073443

1 1.631881 1.0044869

2 1.629853 1.0009525

3 1.661654 1.0151651

4 1.774118 1.0446163

5 1.470588 0.7876045

9 2.013333 0.9514332

atmospheric\_conditions\_1

Y [,1] [,2]

0 9.030424 24.039712

1 6.258028 18.497617

2 9.157032 23.799019

3 8.496241 22.291971

4 8.701961 23.302687

5 1.411765 1.559286

9 5.186667 15.836746

minute\_of\_notification

Y [,1] [,2]

0 29.70175 17.08307

1 29.60206 16.37551

2 29.17774 16.60569

3 28.94522 16.89640

4 29.55739 17.16047

5 34.67647 14.48033

9 31.74667 18.17341

minute\_of\_arrival\_at\_scene

Y [,1] [,2]

0 29.59202 17.36327

1 29.88991 17.28247

2 30.13719 16.83869

3 29.70999 16.40468

4 29.91529 17.30053

5 30.05882 17.28538

9 29.00000 18.10032

number\_of\_fatalities

Y [,1] [,2]

0 1.064339 0.2952554

1 1.103211 0.3597279

2 1.179465 0.6705444

3 1.211600 0.5499899

4 1.211765 0.5920654

5 1.235294 0.6059715

9 1.026667 0.1621922

number\_of\_drunk\_drivers

Y [,1] [,2]

0 0.1690773 0.4080544

1 0.2018349 0.4211365

2 0.2459016 0.4654914

3 0.2964554 0.5124035

4 0.2677124 0.4652862

5 0.4411765 0.6601734

9 0.1200000 0.3661339