Random Forest

Random Forest Prediction Model

Open libraries.

```
library("mlbench")
library("dplyr")
library("caret")
library("randomForest")
library("lattice")
library("ggplot2")
library("rpart")
library("e1071")
library("caret", lib.loc="/Library/Frameworks/R.framework/Versions/3.4/Resources/libr
ary")
library("stats")
library(relaimpo)
library(party)
```

Get a list of Variable names.

```
Names <- names(dfNormDatAndFactor)
unlist(Names, recursive = TRUE, use.names = TRUE)
```

```
 [1] "Enrolling"
 [2] "Sex.F"
 [3] "Sex.M"
 [4] "Expel.N"
 [5] "Expel.Y"
 [6] "First.Gen.N"
 [7] "First.Gen.Y"
 [8] "Challenge.Tag.N"
 [9] "Challenge.Tag.Y"
[10] "Boettcher.Semi.N"
[11] "Boettcher.Semi.Y"
[12] "Boettcher.Final.N"
[13] "Boettcher.Final.Y"
[14] "Daniels.Semi.N"
[15] "Daniels.Semi.Y"
```

```
 [16] "Daniels.Final.N"
 [17] "Daniels.Final.Y"
 [18] "Harvey.App.N"
 [19] "Harvey.App.Y"
 [20] "Harvey.Final.N"
 [21] "Harvey.Final.Y"
 [22] "FC.App.N"
 [23] "FC.App.Y"
 [24] "Thorson.App.N"
 [25] "Thorson.App.Y"
 [26] "Thorson.Admit.N"
 [27] "Thorson.Admit.Y"
 [28] "Summet.Participant.N"
 [29] "Summet.Participant.Y"
 [30] "Mines.Medal.N"
 [31] "Mines.Medal.Y"
 [32] "SPS.N"
 [33] "SPS.Y"
 [34] "Veteran.N"
 [35] "Veteran.Y"
 [36] "Legacy.N"
 [37] "Legacy.Y"
 [38] "Athlete.N"
 [39] "Athlete.Y"
 [40] "State.CO"
 [41] "State.Other"
 [42] "Citizenship.Foreign.National.International"
 [43] "Citizenship.International"
 [44] "Citizenship.Missing"
 [45] "Citizenship.U.S..Citizen"
 [46] "Citizenship.U.S..Permanent.Resident.Green.Card.Holder"
 [47] "Citizenship.Undocumented.DACA"
 [48] "Ethnicity.American.Indian.or.Alaska.Native"
 [49] "Ethnicity.Asian"
 [50] "Ethnicity.Black.or.African.American"
 [51] "Ethnicity.Hispanic.or.Latino"
 [52] "Ethnicity.Missing"
 [53] "Ethnicity.Multiracial"
 [54] "Ethnicity.Native.Hawaiian.or.Other.Pacific.Islander"
 [55] "Ethnicity.NotDeclared"
 [56] "Ethnicity.Unknown"
 [57] "Ethnicity.White"
 [58] "Major.App.Applied.Mathematics...Statistics...Computational...Applied.Mathemati
cs"
 [59] "Major.App.Applied.Mathematics...Statistics...Statistics"
 [60] "Major.App.Chemical.Engineering"
 [61] "Major.App.Chemical.Engineering...Biological.Engineering.Specialty"
 [62] "Major.App.Chemistry"
 [63] "Major.App.Chemistry...Biochemistry.Specialty"
 [64] "Major.App.Chemistry...Environmental.Chemistry.Specialty"
```

```
 [65] "Major.App.Civil.Engineering"
 [66] "Major.App.Computer.Science"
 [67] "Major.App.Economics"
 [68] "Major.App.Electrical.Engineering"
 [69] "Major.App.Engineering.Physics"
 [70] "Major.App.Environmental.Engineering"
 [71] "Major.App.Geological.Engineering"
 [72] "Major.App.Geophysical.Engineering"
 [73] "Major.App.Mechanical.Engineering"
 [74] "Major.App.Metallurgical...Materials.Engineering"
 [75] "Major.App.Mining.Engineering"
 [76] "Major.App.Missing"
 [77] "Major.App.Petroleum.Engineering"
 [78] "Major.App.Undecided"
 [79] "First.Contact.ACT"
 [80] "First.Contact.Application"
 [81] "First.Contact.Athlete.Form"
 [82] "First.Contact.Campus.Visit"
 [83] "First.Contact.College.Fair"
 [84] "First.Contact.FUF"
 [85] "First.Contact.Girls.Lead.the.Way"
 [86] "First.Contact.GPA.Form"
 [87] "First.Contact.Inquiry.Form"
 [88] "First.Contact.Mailing"
 [89] "First.Contact.Materials"
 [90] "First.Contact.MEP"
 [91] "First.Contact.Phone"
 [92] "First.Contact.Preview.Mines"
 [93] "First.Contact.Royall.Search"
 [94] "First.Contact.SAT"
 [95] "First.Contact.TOEFL"
 [96] "First.Visit.Campus.Tour"
 [97] "First.Visit.Campus.Visit"
 [98] "First.Visit.Class.Shadow"
 [99] "First.Visit.Discover.Mines"
[100] "First.Visit.Discovery.Mines.in.Your.City"
[101] "First.Visit.Girls.Lead.the.Way"
[102] "First.Visit.Launch"
[103] "First.Visit.Making.The.Connection"
[104] "First.Visit.Meet.Me.at.Mines"
[105] "First.Visit.None"
[106] "First.Visit.Preview.Mines"
[107] "First.Visit.Preview.Mines.In.Your.City"
[108] "Sport.Men.s.Varsity.Baseball"
[109] "Sport.Men.s.Varsity.Basketball"
[110] "Sport.Men.s.Varsity.Football"
[111] "Sport.Men.s.Varsity.Golf"
[112] "Sport.Men.s.Varsity.Soccer"
[113] "Sport.Men.s.Varsity.Wrestling"
[114] "Sport.None"
```

```
[115] "Sport.Varsity.Swimming.Diving"
[116] "Sport.Varsity.Track.Field"
[117] "Sport.Varsity.XCountry"
[118] "Sport.Women.s.Varsity.Basketball"
[119] "Sport.Women.s.Varsity.Soccer"
[120] "Sport.Women.s.Varsity.Softball"
[121] "Sport.Women.s.Varsity.Volleyball"
[122] "App.Created.Days"
[123] "Age"
[124] "HS.GPA"
[125] "SATR.Converted"
[126] "Review.OutsideActivity"
[127] "Review.Leadership"
[128] "Review.WorkExp"
[129] "Review.WorkEthic"
[130] "Review.ExpDiversity"
[131] "Review.DesireAttend"
[132] "Review.Affinity"
[133] "Review.InnovEntrep"
[134] "Review.Teamwork"
[135] "Review.OverallFit"
[136] "Logins.60Days"
[137] "EventCount.All"
[138] "EventCount.Admitted"
[139] "EventCount.Campus"
```

Parrtitions created with 75% of data for training and 25% of data for testing.

Hide

Hide

```
inTraining <- createDataPartition(dfNormDatAndFactor$Enrolling, p = 0.75, list = FALSE)
training <- dfNormDatAndFactor[inTraining, ]
testing <- dfNormDatAndFactor[-inTraining, ]
```

Hide

Hide

```
training
```

Enrolling	Se...	Se...	Expel.N	Expel.Y	First.Gen.N	First.Gen.Y	Challenge.Tag.N
<fctr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1 N	0	1	1	0	1	0	1

2	N	0	1	1	0	1	0	1
3	N	0	1	1	0	1	0	1
4	N	0	1	1	0	1	0	1
5	N	0	1	1	0	1	0	1
6	N	0	1	1	0	1	0	1
8	N	0	1	1	0	1	0	1
9	N	1	0	1	0	1	0	1
10	Y	1	0	1	0	1	0	1
11	N	0	1	1	0	1	0	1

1-10 of 4,677 rows | 1-10 of 139 columns   Previous **1** 2 3 4 5 6 … 100 Next

Hide

Hide

testing

	Enrolling <fctr>	Se... <dbl>	Se... <dbl>	Expel.N <dbl>	Expel.Y <dbl>	First.Gen.N <dbl>	First.Gen.Y <dbl>	Challenge.Tag.N <dbl>
7	N	1	0	0	1	0	1	1
16	N	0	1	1	0	1	0	1
17	Y	0	1	1	0	0	1	1
19	N	0	1	1	0	1	0	1
23	N	1	0	1	0	1	0	1
26	N	0	1	1	0	1	0	1
27	Y	0	1	1	0	0	1	1
28	Y	0	1	1	0	1	0	1
31	N	0	1	1	0	1	0	1
39	N	0	1	1	0	1	0	1

1-10 of 1,558 rows | 1-10 of 139 columns   Previous **1** 2 3 4 5 6 … 100 Next

Run Random Forest Model with all variables.

Hide

Hide

```
rfModel <- randomForest(Enrolling ~ Sex.F + Sex.M +Expel.N + Expel.Y+ First.Gen.N+ Fi
rst.Gen.Y+Challenge.Tag.N + Challenge.Tag.Y+ Boettcher.Semi.N+ Boettcher.Semi.Y+ Boet
tcher.Final.N+Boettcher.Final.Y + Daniels.Semi.N+Daniels.Semi.Y+ Daniels.Final.N+ Dan
iels.Final.Y+ Harvey.App.N+ Harvey.App.Y + Harvey.Final.N+ Harvey.Final.Y+ FC.App.N+
FC.App.Y+  Thorson.App.N+ Thorson.App.Y+ Thorson.Admit.N +Thorson.Admit.Y+Summet.Part
icipant.N+ Summet.Participant.Y + Mines.Medal.N+ Mines.Medal.Y+ SPS.N+ SPS.Y+ Veteran
.N+ Veteran.Y+ Legacy.N +  Legacy.Y+ Athlete.N +  Athlete.N + Athlete.Y + State.CO +
State.Other  + Citizenship.Foreign.National.International +Citizenship.International
+  Citizenship.Missing+ Citizenship.U.S..Citizen+ Citizenship.U.S..Permanent.Resident
.Green.Card.Holder+ Citizenship.Undocumented.DACA+ Ethnicity.American.Indian.or.Alask
a.Native+  Ethnicity.Asian+ Ethnicity.Black.or.African.American+ Ethnicity.Hispanic.o
r.Latino + Ethnicity.Missing + Ethnicity.Multiracial +Ethnicity.Native.Hawaiian.or.Ot
her.Pacific.Islander + Ethnicity.NotDeclared + Ethnicity.Unknown +Ethnicity.White+ Ma
jor.App.Applied.Mathematics...Statistics...Computational...Applied.Mathematics+ Major
.App.Applied.Mathematics...Statistics...Statistics+ Major.App.Chemical.Engineering+ M
ajor.App.Chemical.Engineering...Biological.Engineering.Specialty+ Major.App.Chemistry
+  Major.App.Chemistry...Biochemistry.Specialty + Major.App.Chemistry...Environmental
.Chemistry.Specialty + Major.App.Civil.Engineering + Major.App.Computer.Science + Maj
or.App.Economics+ Major.App.Electrical.Engineering+ Major.App.Engineering.Physics+  M
ajor.App.Environmental.Engineering+ Major.App.Geological.Engineering +  Major.App.Geo
physical.Engineering +Major.App.Mechanical.Engineering +  Major.App.Metallurgical...M
aterials.Engineering + Major.App.Mining.Engineering +  Major.App.Missing + Major.App.
Petroleum.Engineering+   Major.App.Undecided+ First.Contact.ACT +  First.Contact.Appl
ication + First.Contact.Athlete.Form + First.Contact.Campus.Visit + First.Contact.Col
lege.Fair + First.Contact.FUF+First.Contact.Girls.Lead.the.Way+First.Contact.GPA.Form
+ First.Contact.Inquiry.Form+First.Contact.Mailing+ First.Contact.Materials +  First.
Contact.MEP+ First.Contact.Phone+ First.Contact.Preview.Mines+ First.Contact.Royall.S
earch+First.Contact.TOEFL+ First.Contact.SAT+ First.Visit.Campus.Tour+ First.Visit.Ca
mpus.Visit+ First.Visit.Class.Shadow +  First.Visit.Discover.Mines+ First.Visit.Disco
very.Mines.in.Your.City+ First.Visit.Girls.Lead.the.Way+  First.Visit.Launch + First.
Visit.Making.The.Connection + First.Visit.Meet.Me.at.Mines+ First.Visit.None+  First.
Visit.Preview.Mines + First.Visit.Preview.Mines.In.Your.City + Sport.Men.s.Varsity.Ba
seball+  Sport.Men.s.Varsity.Basketball+  Sport.Men.s.Varsity.Football + Sport.Men.s.
Varsity.Golf +Sport.Men.s.Varsity.Soccer+  Sport.Men.s.Varsity.Wrestling+  Sport.None
+  Sport.Varsity.Swimming.Diving+ Sport.Varsity.Track.Field+Sport.Varsity.XCountry+
Sport.Women.s.Varsity.Basketball +  Sport.Women.s.Varsity.Soccer+  Sport.Women.s.Vars
ity.Softball + Sport.Women.s.Varsity.Volleyball+ App.Created.Days +  Age  + HS.GPA+ S
ATR.Converted +Review.OutsideActivity +Review.Leadership+Review.WorkExp+ Review.WorkE
thic +Review.ExpDiversity +Review.DesireAttend +Review.Affinity+ Review.InnovEntrep+R
eview.Teamwork +Review.OverallFit + Logins.60Days+EventCount.All+EventCount.Admitted+
EventCount.Campus, data =  training)
rfModel
```

```
Call:
 randomForest(formula = Enrolling ~ Sex.F + Sex.M + Expel.N +      Expel.Y + First.Ge
n.N + First.Gen.Y + Challenge.Tag.N + Challenge.Tag.Y +      Boettcher.Semi.N + Boett
cher.Semi.Y + Boettcher.Final.N +      Boettcher.Final.Y + Daniels.Semi.N + Daniels.S
emi.Y + Daniels.Final.N +      Daniels.Final.Y + Harvey.App.N + Harvey.App.Y + Harvey
```

```
.Final.N +        Harvey.Final.Y + FC.App.N + FC.App.Y + Thorson.App.N + Thorson.App.Y
+        Thorson.Admit.N + Thorson.Admit.Y + Summet.Participant.N +        Summet.Partici
pant.Y + Mines.Medal.N + Mines.Medal.Y + SPS.N +         SPS.Y + Veteran.N + Veteran.Y +
Legacy.N + Legacy.Y + Athlete.N +         Athlete.N + Athlete.Y + State.CO + State.Other
+ Citizenship.Foreign.National.International +        Citizenship.International + Citiz
enship.Missing + Citizenship.U.S..Citizen +         Citizenship.U.S..Permanent.Resident.
Green.Card.Holder + Citizenship.Undocumented.DACA +         Ethnicity.American.Indian.or
.Alaska.Native + Ethnicity.Asian +         Ethnicity.Black.or.African.American + Ethnici
ty.Hispanic.or.Latino +         Ethnicity.Missing + Ethnicity.Multiracial + Ethnicity.Na
tive.Hawaiian.or.Other.Pacific.Islander +         Ethnicity.NotDeclared + Ethnicity.Unkn
own + Ethnicity.White +         Major.App.Applied.Mathematics...Statistics...Computation
al...Applied.Mathematics +         Major.App.Applied.Mathematics...Statistics...Statisti
cs +        Major.App.Chemical.Engineering + Major.App.Chemical.Engineering...Biologica
l.Engineering.Specialty +         Major.App.Chemistry + Major.App.Chemistry...Biochemist
ry.Specialty +        Major.App.Chemistry...Environmental.Chemistry.Specialty +        Ma
jor.App.Civil.Engineering + Major.App.Computer.Science +        Major.App.Economics + M
ajor.App.Electrical.Engineering +        Major.App.Engineering.Physics + Major.App.Envi
ronmental.Engineering +         Major.App.Geological.Engineering + Major.App.Geophysical
.Engineering +        Major.App.Mechanical.Engineering + Major.App.Metallurgical...Mate
rials.Engineering +        Major.App.Mining.Engineering + Major.App.Missing + Major.App
.Petroleum.Engineering +        Major.App.Undecided + First.Contact.ACT + First.Contact
.Application +        First.Contact.Athlete.Form + First.Contact.Campus.Visit +        Fi
rst.Contact.College.Fair + First.Contact.FUF + First.Contact.Girls.Lead.the.Way +
First.Contact.GPA.Form + First.Contact.Inquiry.Form + First.Contact.Mailing +        Fi
rst.Contact.Materials + First.Contact.MEP + First.Contact.Phone +        First.Contact.
Preview.Mines + First.Contact.Royall.Search +        First.Contact.TOEFL + First.Contac
t.SAT + First.Visit.Campus.Tour +        First.Visit.Campus.Visit + First.Visit.Class.S
hadow + First.Visit.Discover.Mines +        First.Visit.Discovery.Mines.in.Your.City +
First.Visit.Girls.Lead.the.Way +        First.Visit.Launch + First.Visit.Making.The.Con
nection +        First.Visit.Meet.Me.at.Mines + First.Visit.None + First.Visit.Preview.
Mines +        First.Visit.Preview.Mines.In.Your.City + Sport.Men.s.Varsity.Baseball +
Sport.Men.s.Varsity.Basketball + Sport.Men.s.Varsity.Football +        Sport.Men.s.Vars
ity.Golf + Sport.Men.s.Varsity.Soccer + Sport.Men.s.Varsity.Wrestling +        Sport.No
ne + Sport.Varsity.Swimming.Diving + Sport.Varsity.Track.Field +        Sport.Varsity.X
Country + Sport.Women.s.Varsity.Basketball +        Sport.Women.s.Varsity.Soccer + Spor
t.Women.s.Varsity.Softball +        Sport.Women.s.Varsity.Volleyball + App.Created.Days
+ Age +        HS.GPA + SATR.Converted + Review.OutsideActivity + Review.Leadership +
Review.WorkExp + Review.WorkEthic + Review.ExpDiversity +        Review.DesireAttend +
Review.Affinity + Review.InnovEntrep +        Review.Teamwork + Review.OverallFit + Log
ins.60Days + EventCount.All +        EventCount.Admitted + EventCount.Campus, data = tr
aining)
               Type of random forest: classification
                     Number of trees: 500
No. of variables tried at each split: 11

       OOB estimate of  error rate: 8.66%
Confusion matrix:
     N    Y class.error
N 3561 129  0.03495935
Y  276 711  0.27963526
```

```
rfModel.prediction <- predict(rfModel, testing)
table(rfModel.prediction, testing$Enrolling)
```

```
rfModel.prediction    N     Y
                 N 1188    91
                 Y   41   238
```

To determine variable importance.

```
imprfModel <-   importance(rfModel)
imprfModel
```

```
MeanDecreaseGini
Sex.F
8.31162435
Sex.M
8.30279602
Expel.N
0.87412550
Expel.Y
1.00543550
First.Gen.N
5.15213869
First.Gen.Y
4.78919446
Challenge.Tag.N
2.83275830
Challenge.Tag.Y
2.57381977
Boettcher.Semi.N
1.37651424
Boettcher.Semi.Y
1.43482701
Boettcher.Final.N
0.33859291
Boettcher.Final.Y
0.43893491
Daniels.Semi.N
1.62110536
Daniels.Semi.Y
```

1.68219033

Daniels.Final.N

0.60787117

Daniels.Final.Y

0.54018330

Harvey.App.N

4.97117672

Harvey.App.Y

4.70471850

Harvey.Final.N

1.53181233

Harvey.Final.Y

1.54075313

FC.App.N

2.60945787

FC.App.Y

2.85827829

Thorson.App.N

4.91009047

Thorson.App.Y

6.05993438

Thorson.Admit.N

4.20177586

Thorson.Admit.Y

4.66015699

Summet.Participant.N

0.95078428

Summet.Participant.Y

1.05285091

Mines.Medal.N

0.03783863

Mines.Medal.Y

0.05204497

SPS.N

1.46168239

SPS.Y

1.47895399

Veteran.N

3.97530255

Veteran.Y

3.97406082

Legacy.N

7.77444938

Legacy.Y

7.54647949

Athlete.N

5.68213797

Athlete.Y

6.15825066

State.CO

32.09078569

State.Other

28.57987751

Citizenship.Foreign.National.International

3.02547467

Citizenship.International

1.00869728

Citizenship.Missing

2.58745909

Citizenship.U.S..Citizen

5.19523174

Citizenship.U.S..Permanent.Resident.Green.Card.Holder

1.66754622

Citizenship.Undocumented.DACA

0.03288244

Ethnicity.American.Indian.or.Alaska.Native

1.69935410

Ethnicity.Asian

5.10062410

Ethnicity.Black.or.African.American

2.24064049

Ethnicity.Hispanic.or.Latino

3.26707202

Ethnicity.Missing

0.99698980

Ethnicity.Multiracial

0.01743671

Ethnicity.Native.Hawaiian.or.Other.Pacific.Islander

0.15216737

Ethnicity.NotDeclared

3.43252597

Ethnicity.Unknown

0.02969030

Ethnicity.White

8.58428909

Major.App.Applied.Mathematics...Statistics...Computational...Applied.Mathematics

1.59490633

Major.App.Applied.Mathematics...Statistics...Statistics

1.10884453

Major.App.Chemical.Engineering

7.53180233

Major.App.Chemical.Engineering...Biological.Engineering.Specialty

5.41700331

Major.App.Chemistry

1.08667849

Major.App.Chemistry...Biochemistry.Specialty

1.88158543

Major.App.Chemistry...Environmental.Chemistry.Specialty

0.35934883

Major.App.Civil.Engineering

4.45976044
Major.App.Computer.Science
5.41948233
Major.App.Economics
1.26927741
Major.App.Electrical.Engineering
4.71005209
Major.App.Engineering.Physics
5.02828541
Major.App.Environmental.Engineering
3.46022614
Major.App.Geological.Engineering
2.35911823
Major.App.Geophysical.Engineering
1.70564568
Major.App.Mechanical.Engineering
8.26316053
Major.App.Metallurgical...Materials.Engineering
1.89466708
Major.App.Mining.Engineering
2.24874132
Major.App.Missing
2.49144008
Major.App.Petroleum.Engineering
4.57228645
Major.App.Undecided
5.85083193
First.Contact.ACT
5.32237822
First.Contact.Application
8.45047653
First.Contact.Athlete.Form
1.29534193
First.Contact.Campus.Visit
13.24551159
First.Contact.College.Fair
0.37256429
First.Contact.FUF
7.72037229
First.Contact.Girls.Lead.the.Way
0.14998241
First.Contact.GPA.Form
0.77783347
First.Contact.Inquiry.Form
2.71696161
First.Contact.Mailing
0.11418693
First.Contact.Materials
0.41322987
First.Contact.MEP

0.56274722

First.Contact.Phone

0.02108717

First.Contact.Preview.Mines

2.26815012

First.Contact.Royall.Search

12.17551414

First.Contact.TOEFL

0.76390552

First.Contact.SAT

1.28938035

First.Visit.Campus.Tour

13.33029545

First.Visit.Campus.Visit

14.16522415

First.Visit.Class.Shadow

2.20631252

First.Visit.Discover.Mines

8.47063515

First.Visit.Discovery.Mines.in.Your.City

1.55088027

First.Visit.Girls.Lead.the.Way

0.30175341

First.Visit.Launch

46.30309394

First.Visit.Making.The.Connection

4.54608849

First.Visit.Meet.Me.at.Mines

1.13989835

First.Visit.None

80.68057211

First.Visit.Preview.Mines

6.40724385

First.Visit.Preview.Mines.In.Your.City

20.47203826

Sport.Men.s.Varsity.Baseball

0.15746001

Sport.Men.s.Varsity.Basketball

0.46889078

Sport.Men.s.Varsity.Football

2.78583132

Sport.Men.s.Varsity.Golf

0.23317390

Sport.Men.s.Varsity.Soccer

0.62589857

Sport.Men.s.Varsity.Wrestling

0.65914856

Sport.None

5.10646599

Sport.Varsity.Swimming.Diving

0.00922697

Sport.Varsity.Track.Field

0.30668743

Sport.Varsity.XCountry

0.00000000

Sport.Women.s.Varsity.Basketball

0.29766657

Sport.Women.s.Varsity.Soccer

0.29437986

Sport.Women.s.Varsity.Softball

0.47338251

Sport.Women.s.Varsity.Volleyball

0.36971577

App.Created.Days

58.87385795

Age

44.68016051

HS.GPA

67.84518762

SATR.Converted

45.04786153

Review.OutsideActivity

15.53448567

Review.Leadership

20.33710006

Review.WorkExp

20.77576354

Review.WorkEthic

15.67811953

Review.ExpDiversity

17.50854333

Review.DesireAttend

22.39637717

Review.Affinity

20.41127841

Review.InnovEntrep

16.07374269

Review.Teamwork

19.42015935

Review.OverallFit

26.09393905

Logins.60Days

1.27884088

EventCount.All

164.46894875

EventCount.Admitted

208.64028819

EventCount.Campus

128.96590908

```
#format((sort(imprfModel, decreasing=TRUE)), scientific=F)
#sort(imprfModel, decreasing=TRUE)  # relative importance
```

```
dfimprfModel <-as.data.frame(imprfModel)
dfimprfModel
```

Sex.F
Sex.M
Expel.N
Expel.Y
First.Gen.N
First.Gen.Y
Challenge.Tag.N
Challenge.Tag.Y
Boettcher.Semi.N
Boettcher.Semi.Y

1-10 of 138 rows   Previous **1** 2 3 4 5 6 ... 14 Next

```
confusionMatrix(table(rfModel.prediction, testing$Enrolling))
```

```
Confusion Matrix and Statistics


rfModel.prediction     N     Y
                 N 1188    91
                 Y   41   238


               Accuracy : 0.9153
                 95% CI : (0.9003, 0.9286)
    No Information Rate : 0.7888
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.7307
 Mcnemar's Test P-Value : 2e-05

            Sensitivity : 0.9666
            Specificity : 0.7234
         Pos Pred Value : 0.9289
         Neg Pred Value : 0.8530
             Prevalence : 0.7888
         Detection Rate : 0.7625
   Detection Prevalence : 0.8209
      Balanced Accuracy : 0.8450

       'Positive' Class : N
```
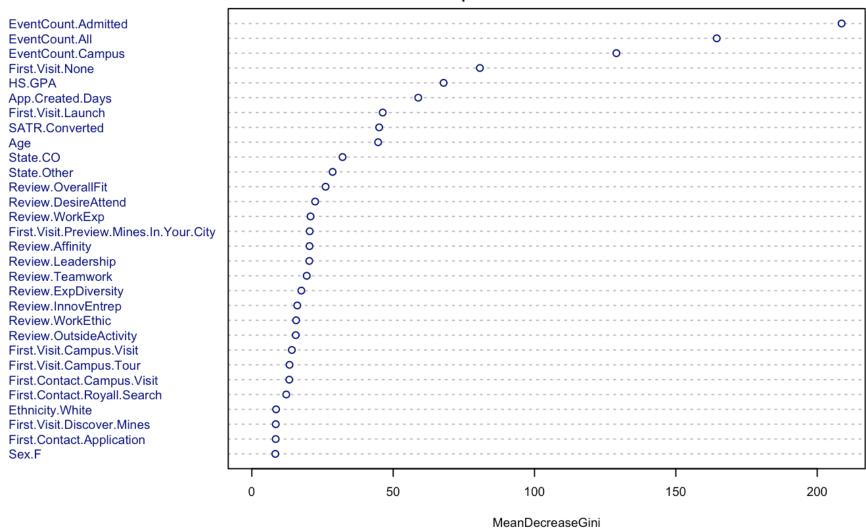
Plot of importance of Variabels.

```
varImpPlot(rfModel,cex = .6, pt.cex = .7, color = "navy blue", gcolor = par("fg"), lc
olor = "gray", main = "Variable Importance for Random Forest Model")
```

Variable Importance for Random Forest Model

Model with less variables

```
rfModelLess <- randomForest(Enrolling ~ State.CO +  State.Other  +  Major.App.Mechani
cal.Engineering +  First.Contact.Application + First.Contact.Campus.Visit + First.Con
tact.FUF + First.Contact.Royall.Search+ First.Visit.Campus.Tour+ First.Visit.Campus.V
isit+ First.Visit.Discover.Mines+ First.Visit.Launch + First.Visit.None+  First.Visit
.Preview.Mines.In.Your.City + App.Created.Days +  Age  + HS.GPA+ SATR.Converted +Revi
ew.OutsideActivity +Review.Leadership+Review.WorkExp+ Review.WorkEthic +Review.ExpDiv
ersity +Review.DesireAttend +Review.Affinity+ Review.InnovEntrep+Review.Teamwork +Rev
iew.OverallFit +Ethnicity.White +EventCount.All+EventCount.Admitted+EventCount.Campus
, data =  training)
rfModelLess
```

```
Call:
 randomForest(formula = Enrolling ~ State.CO + State.Other + Major.App.Mechanical.Eng
ineering +       First.Contact.Application + First.Contact.Campus.Visit +       First.C
ontact.FUF + First.Contact.Royall.Search + First.Visit.Campus.Tour +       First.Visit
.Campus.Visit + First.Visit.Discover.Mines + First.Visit.Launch +       First.Visit.No
ne + First.Visit.Preview.Mines.In.Your.City +       App.Created.Days + Age + HS.GPA +
SATR.Converted + Review.OutsideActivity +       Review.Leadership + Review.WorkExp + R
eview.WorkEthic + Review.ExpDiversity +       Review.DesireAttend + Review.Affinity +
Review.InnovEntrep +       Review.Teamwork + Review.OverallFit + Ethnicity.White + Eve
ntCount.All +       EventCount.Admitted + EventCount.Campus, data = training)
               Type of random forest: classification
                     Number of trees: 500
No. of variables tried at each split: 5

        OOB estimate of  error rate: 8.66%
Confusion matrix:
     N    Y class.error
N 3566 124  0.03360434
Y  281 706  0.28470111
```

```
rfModelLess.prediction <- predict(rfModelLess, testing)
table(rfModelLess.prediction, testing$Enrolling)
```

```
rfModelLess.prediction    N     Y
                  N 1186    94
                  Y    43   235
```

To determine variable imortance.

```
imprfModelLess <-    importance(rfModelLess)
#format(imprfModelLess, scientific=F)
imprfModelLess
```

	MeanDecreaseGini
State.CO	34.05031
State.Other	31.51954
Major.App.Mechanical.Engineering	12.13023
First.Contact.Application	11.15855
First.Contact.Campus.Visit	14.34067
First.Contact.FUF	11.00115
First.Contact.Royall.Search	15.02042
First.Visit.Campus.Tour	15.55050
First.Visit.Campus.Visit	16.38629
First.Visit.Discover.Mines	10.34385
First.Visit.Launch	49.91835
First.Visit.None	68.70999
First.Visit.Preview.Mines.In.Your.City	18.98535
App.Created.Days	91.68313
Age	68.96971
HS.GPA	102.51358
SATR.Converted	70.75703
Review.OutsideActivity	21.27905
Review.Leadership	29.41100
Review.WorkExp	31.62337
Review.WorkEthic	21.41111
Review.ExpDiversity	24.14377
Review.DesireAttend	30.97731
Review.Affinity	29.91587
Review.InnovEntrep	23.73693
Review.Teamwork	27.96879
Review.OverallFit	40.28370
Ethnicity.White	14.95796
EventCount.All	179.77399
EventCount.Admitted	265.99776
EventCount.Campus	142.44319

Hide

Hide

```
confusionMatrix(table(rfModelLess.prediction, testing$Enrolling))
```

```
Confusion Matrix and Statistics


rfModelLess.prediction    N     Y
                     N 1186    94
                     Y   43   235

               Accuracy : 0.9121
                 95% CI : (0.8969, 0.9257)
    No Information Rate : 0.7888
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.7202
 Mcnemar's Test P-Value : 1.939e-05

            Sensitivity : 0.9650
            Specificity : 0.7143
         Pos Pred Value : 0.9266
         Neg Pred Value : 0.8453
             Prevalence : 0.7888
         Detection Rate : 0.7612
   Detection Prevalence : 0.8216
      Balanced Accuracy : 0.8396

       'Positive' Class : N
```

Model with top 25 variables

```
rfModelTop25 <- randomForest(Enrolling ~ State.CO +  State.Other  +  First.Contact.Ca
mpus.Visit + First.Visit.Campus.Tour+ First.Visit.Campus.Visit+ First.Visit.Discover.
Mines+ First.Visit.Launch + First.Visit.None+  First.Visit.Preview.Mines.In.Your.City
+ App.Created.Days +  Age  + HS.GPA+ SATR.Converted +Review.OutsideActivity +Review.L
eadership+Review.WorkExp+ Review.WorkEthic +Review.ExpDiversity +Review.DesireAttend
+Review.Affinity+ Review.InnovEntrep+Review.Teamwork +Review.OverallFit +EventCount.A
ll+EventCount.Admitted+EventCount.Campus, data =  training)
rfModelTop25
```

```
Call:
 randomForest(formula = Enrolling ~ State.CO + State.Other + First.Contact.Campus.Vis
it +       First.Visit.Campus.Tour + First.Visit.Campus.Visit + First.Visit.Discover.M
ines +       First.Visit.Launch + First.Visit.None + First.Visit.Preview.Mines.In.Your
.City +       App.Created.Days + Age + HS.GPA + SATR.Converted + Review.OutsideActivit
y +       Review.Leadership + Review.WorkExp + Review.WorkEthic + Review.ExpDiversity
+       Review.DesireAttend + Review.Affinity + Review.InnovEntrep +       Review.Teamw
ork + Review.OverallFit + EventCount.All + EventCount.Admitted +       EventCount.Camp
us, data = training)
               Type of random forest: classification
                     Number of trees: 500
No. of variables tried at each split: 5

        OOB estimate of  error rate: 8.85%
Confusion matrix:
      N    Y class.error
N 3554 136   0.03685637
Y  278 709   0.28166160
```

Hide

Hide

```
rfModelTop25.prediction <- predict(rfModelTop25, testing)
table(rfModelTop25.prediction, testing$Enrolling)
```

```
rfModelTop25.prediction    N    Y
                     N 1186   97
                     Y   43  232
```

To determine variable imortance.

Hide

Hide

```
imprfModelTop25 <-    importance(rfModelTop25)
#format(imprfModelLess, scientific=F)
imprfModelTop25
```

	MeanDecreaseGini
State.CO	35.97326
State.Other	28.98525
First.Contact.Campus.Visit	17.00886
First.Visit.Campus.Tour	14.44899
First.Visit.Campus.Visit	18.73459
First.Visit.Discover.Mines	10.37029
First.Visit.Launch	57.00928
First.Visit.None	61.66092
First.Visit.Preview.Mines.In.Your.City	19.87106
App.Created.Days	102.40801
Age	75.44019
HS.GPA	110.14023
SATR.Converted	77.62143
Review.OutsideActivity	21.85060
Review.Leadership	32.18853
Review.WorkExp	32.56727
Review.WorkEthic	23.16799
Review.ExpDiversity	25.53950
Review.DesireAttend	32.35349
Review.Affinity	30.98644
Review.InnovEntrep	26.51481
Review.Teamwork	29.93572
Review.OverallFit	42.95109
EventCount.All	216.77550
EventCount.Admitted	261.31786
EventCount.Campus	124.37093

Hide

Hide

```
confusionMatrix(table(rfModelTop25.prediction, testing$Enrolling))
```

```
Confusion Matrix and Statistics


rfModelTop25.prediction    N     Y
                     N  1186    97
                     Y    43   232

            Accuracy : 0.9101
              95% CI : (0.8948, 0.9239)
  No Information Rate : 0.7888
  P-Value [Acc > NIR] : < 2.2e-16

               Kappa : 0.713
Mcnemar's Test P-Value : 7.488e-06

         Sensitivity : 0.9650
         Specificity : 0.7052
      Pos Pred Value : 0.9244
      Neg Pred Value : 0.8436
          Prevalence : 0.7888
      Detection Rate : 0.7612
Detection Prevalence : 0.8235
    Balanced Accuracy : 0.8351

      'Positive' Class : N
```

Hide

Hide

```
PredictionTop25 <- cbind(rfModelTop25.prediction,testing)
PredictionTop25
```

	rfModelTop25.prediction	Enrolling	Se...	Se...	Expel.N	Expel.Y	First.Gen.N	First
	<fctr>	<fctr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	
7	N	N	1	0	0	1	0	
16	N	N	0	1	1	0	1	
17	N	Y	0	1	1	0	0	
19	N	N	0	1	1	0	1	
23	Y	N	1	0	1	0	1	
26	N	N	0	1	1	0	1	
27	Y	Y	0	1	1	0	0	
28	Y	Y	0	1	1	0	1	

| 31 | N | N | 0 | 1 | 1 | 0 | 1 |
| 39 | N | N | 0 | 1 | 1 | 0 | 1 |