Predictors of Implement and Plan to Implement

Stats Group and BCHC

August 30, 2016

Analysis Plan

For each outcome we would like to identify any salient associations among demogrpahic and institutional variables. We will use random forests to both measure how good a prediction can be created and to identify which attributes are important for the prediction. We will give descriptive statistics broken down by the 6 levels of the targets.

Read Data

Data were downloaded from redcap.

```
load("~/Box Sync/bchc/bchc.Rdata")
```

basic summary of retained variables

```
summary(x)
```

```
##        imp          qi3             four         plan           set
##  no implem:18   no qi   :17   imp, no qi:17   no plan: 8   UniHosp    :17
##  implem   :23   qi team: 6    no imp    :18   plan   :10   CorpHosp   : 7
##  NA's     : 3   NA's   :21    NA's      : 9   NA's   :26   Community :14
##                                                            BirtCenter: 2
##                                                            Other     : 1
##                                                            NA's      : 3
##
##   acred         years        nonwhite                 role
##  TJC :38   Min.   : 4.00   Mode :logical   provider       : 1
##  DVN : 3   1st Qu.:20.00   FALSE:38        directcaregiver: 9
##  none: 3   Median :28.00   TRUE :6         nursemanager   :17
##            Mean   :27.17   NA's :0         gradstudent    : 2
##            3rd Qu.:36.00                   other          :12
##            Max.   :50.00                   NA's           : 3
##            NA's   :3
##     climate       culture     orgcapacity   unitcapacity  accreditation
##  noeffect:31   noeffect:23   noeffect:28   noeffect:31    noeffect:32
##  top4    :13   top4    :21   top4    :16   top4    :13    top4    :12
##
##
##
##
##
##    management    physician      nursing       patient    documentation
##  noeffect:37   noeffect:35   noeffect:37   noeffect:38   noeffect:22
##  top4    : 7   top4    : 9   top4    : 7   top4    : 6   top4    :22
```

```
##
##
##
##
##
##        emr          time         accredit       staff         admin
##  noeffect:25   noeffect:39   noeffect:31   noeffect:23   noeffect:34
##  top4    :19   top4    : 5   top4    :13   top4    :21   top4    :10
##
##
##
##
##
##         nm           md          champ        student        right
##  noeffect:37   noeffect:41   noeffect:30   noeffect:35   noeffect:12
##  top4    : 7   top4    : 3   top4    :14   top4    : 9   top4    :32
##
##
##
##
##
##    individual
##  noeffect:15
##  top4    :29
##
##
##
##
##
```

Random Forests

Random forests attemps to build a classifier by creating a large number of decision trees from bootstrapped data randomly selecting potential features. A good attribute of this method is its ability function when there are more predictors than observations. Additionally the method is robust against correlated predictors, missing values. The output contains the out of the box(oob) error rate which is the rate of error for ensemble of trees applied to data not used to train the model. That makes the oob error resistant to overfitting.

Models can be calibrated with weights to place more value on correctly identifying either level of the target. We did not do that here so models are scored based on their overall correct prediction.

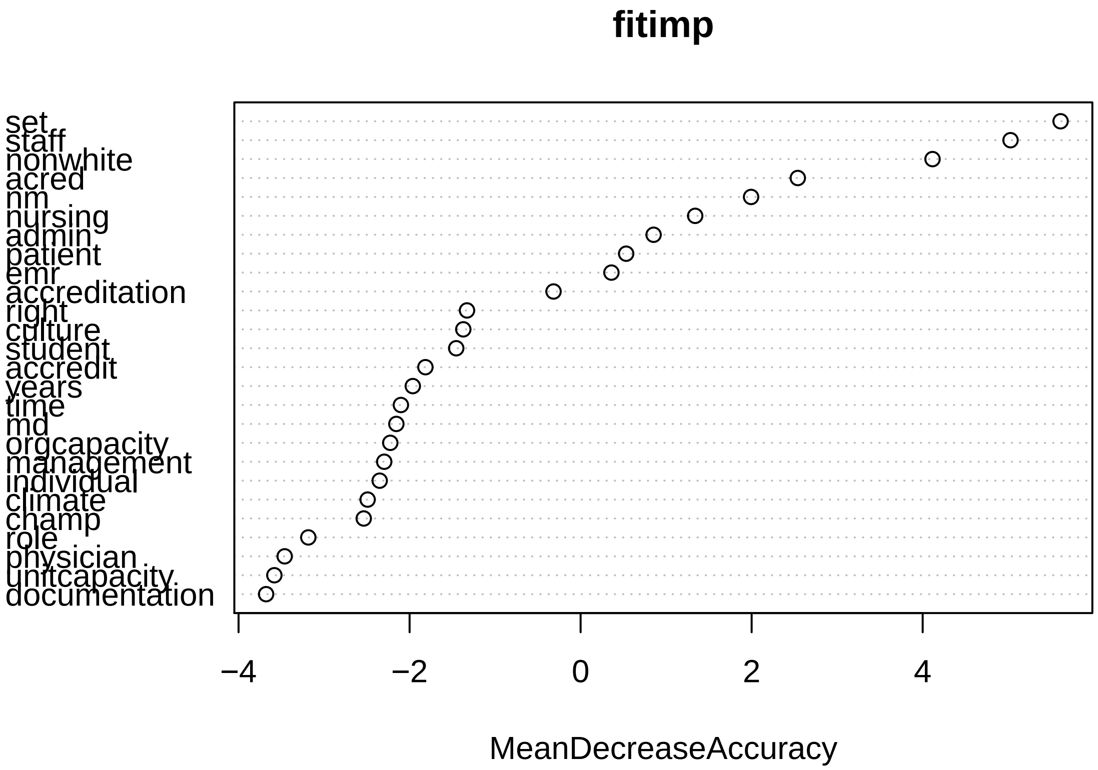
Implementation

```r
 fitimp <- randomForest(imp ~ ., data= x[,-c(2:4)],ntree=500,na.action='na.omit',importance=T)
fitimp
```

```
##
## Call:
##  randomForest(formula = imp ~ ., data = x[, -c(2:4)], ntree = 500,      importance = T, na.action =
##                Type of random forest: classification
##                      Number of trees: 500
## No. of variables tried at each split: 5
##
```

```
##         OOB estimate of  error rate: 46.34%
## Confusion matrix:
##            no implem implem class.error
## no implem          6     12   0.6666667
## implem             7     16   0.3043478
```

```r
varImpPlot(fitimp,type=1)
```

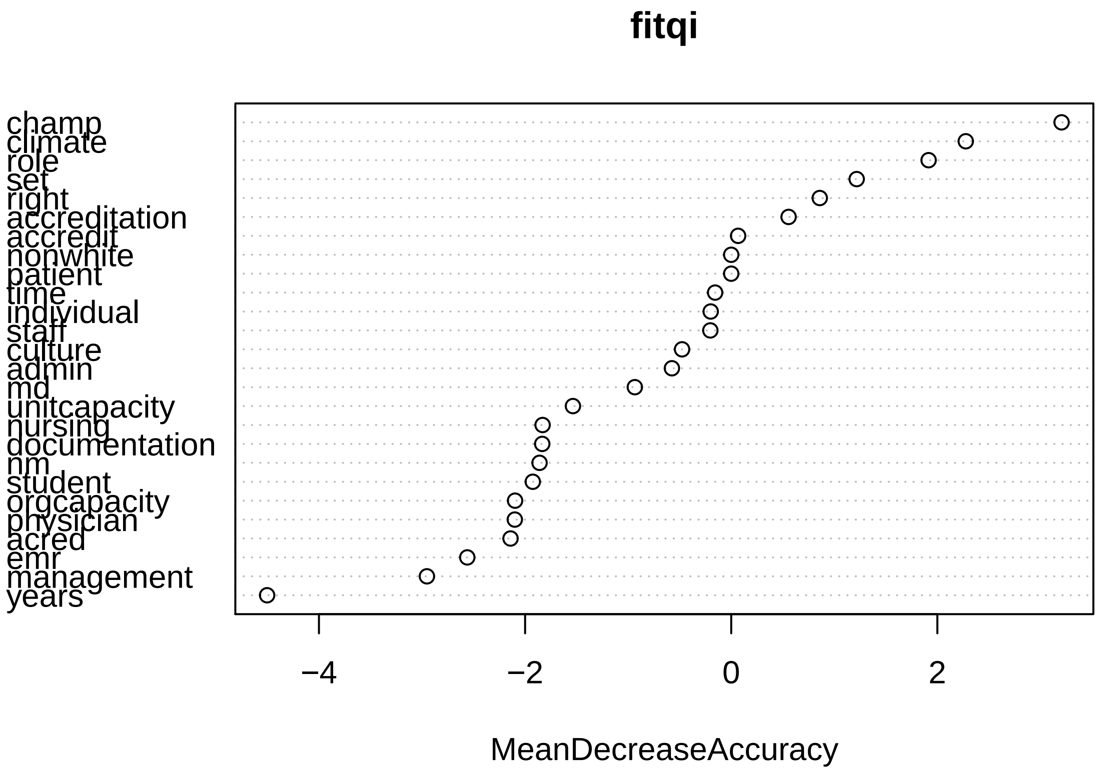


QI

```r
 fitqi <- randomForest(qi3 ~ ., data= x[,-c(1,3,4)],ntree=500,na.action='na.omit',importance=T)
fitqi
```

```
##
## Call:
##  randomForest(formula = qi3 ~ ., data = x[, -c(1, 3, 4)], ntree = 500,      importance = T, na.actio
##                Type of random forest: classification
##                      Number of trees: 500
## No. of variables tried at each split: 5
##
##         OOB estimate of  error rate: 26.09%
## Confusion matrix:
##         no qi qi team class.error
## no qi      17       0           0
## qi team     6       0           1
```

```r
varImpPlot(fitqi,type=1)
```

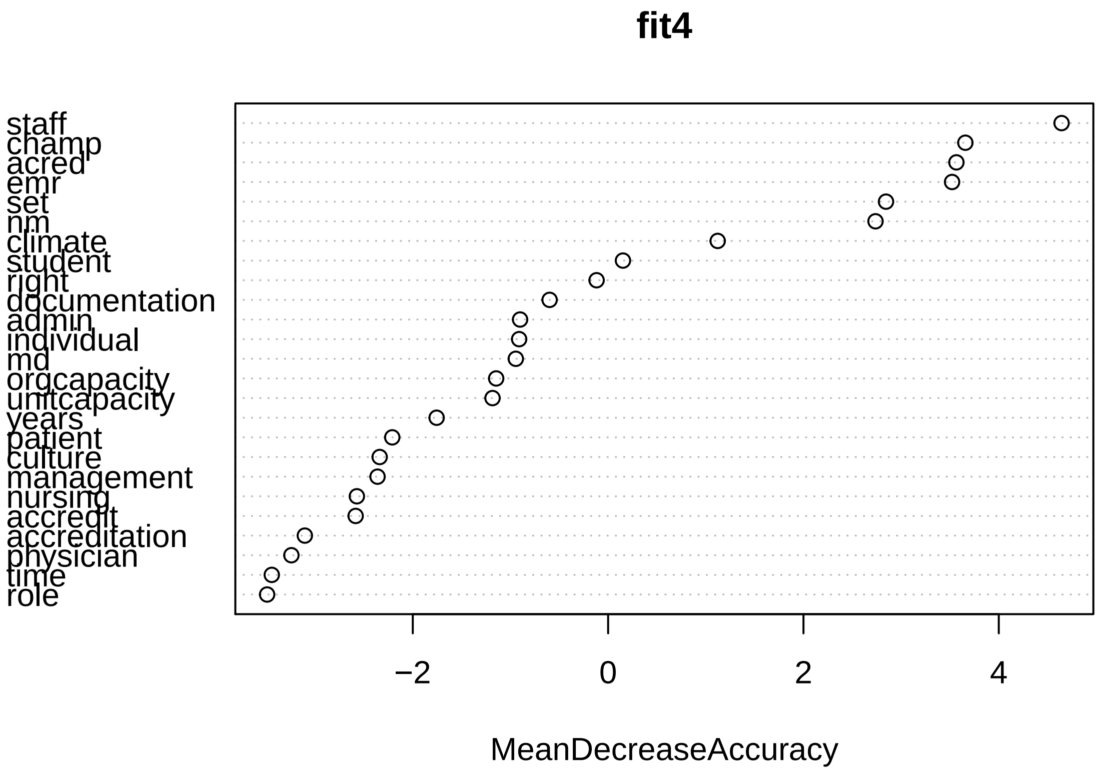


four

```
fit4 <- randomForest(four ~ ., data= x[,-c(1:2,4,8)],ntree=500,na.action='na.omit',importance=T)
fit4
```

```
##
## Call:
##  randomForest(formula = four ~ ., data = x[, -c(1:2, 4, 8)], ntree = 500,      importance = T, na.ac
##                Type of random forest: classification
##                      Number of trees: 500
## No. of variables tried at each split: 5
##
##         OOB estimate of  error rate: 51.43%
## Confusion matrix:
##             imp, no qi no imp class.error
## imp, no qi           9      8   0.4705882
## no imp              10      8   0.5555556
```

```
varImpPlot(fit4,type=1)
```

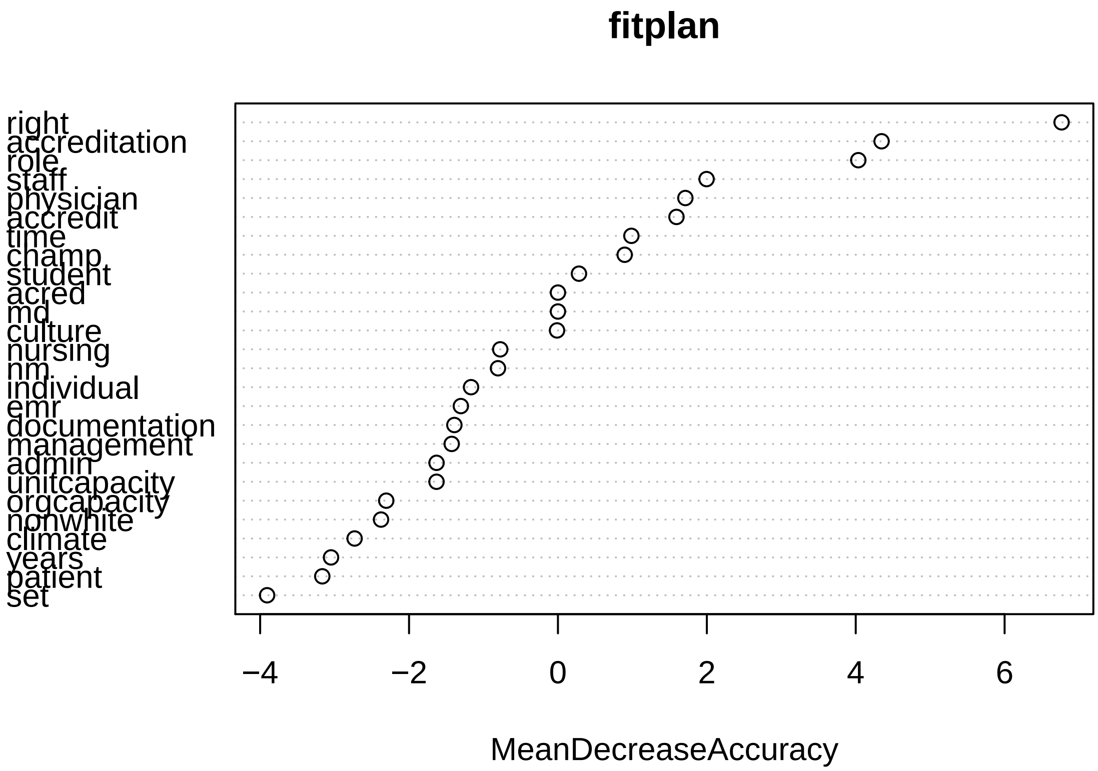


plan

```
fitplan <- randomForest(plan ~ ., data= x[,-c(1:3)],ntree=500,na.action='na.omit',importance=T)
fitplan
```

```
##
## Call:
##  randomForest(formula = plan ~ ., data = x[, -c(1:3)], ntree = 500,      importance = T, na.action =
##                Type of random forest: classification
##                      Number of trees: 500
## No. of variables tried at each split: 5
##
##         OOB estimate of  error rate: 44.44%
## Confusion matrix:
##         no plan plan class.error
## no plan       2    6        0.75
## plan          2    8        0.20
```

```
varImpPlot(fitplan,type=1)
```


fitplan
right
accreditation
role
staff
physician
accredit
time
champ
student
acred
md
culture
nursing
nm
individual
emr
documentation
management
admin
unitcapacity
orgcapacity
nonwhite
climate
years
patient
set
−4
−2
0
2
4
6
MeanDecreaseAccuracy