Predicting Games Played

Kyle Joecken

Question

Can we use data from previous seasons to predict how much time a particular NHL skater will get on the ice (TOI) in various situations (even strength, short-handed, power play) in the upcoming 2014-2015 season?

Data

As before, we load our database of season-wide NHL data into a data frame called `skaterstats`. We'll also grab our games played predictions and store it in a data frame called `skatpred15`; these two steps are done with the `setup.R` file. Finally, we will load our previous model for games played (`GPModel`).

Feature Selection

We'd like to compare a few different attempts. First, we'd like to model time on ice directly, with and without our predicted games played. Then, we'd like to model time on ice per game and multiply it by our predicted games played numbers and compare the accuracy.

We'll start again by modelling even-strength TOI using all possible predictors, then try to ascertain which are the most important factors.

```
gbmdata <- nhlShape(2012, 2013, outcome = 38)
rfdata <- nhlShape(2013, 2013, outcome = 38)
```

First, we build random forest models from several seeds.

```
esrfmod1 <- nhlBuild(data = rfdata, perc = 0.7, seed = 9112)
```

```
## [1] 0.7332
```

```
esrfmod2 <- nhlBuild(data = rfdata, perc = 0.7, seed = 2857)
```

```
## [1] 0.7851
```

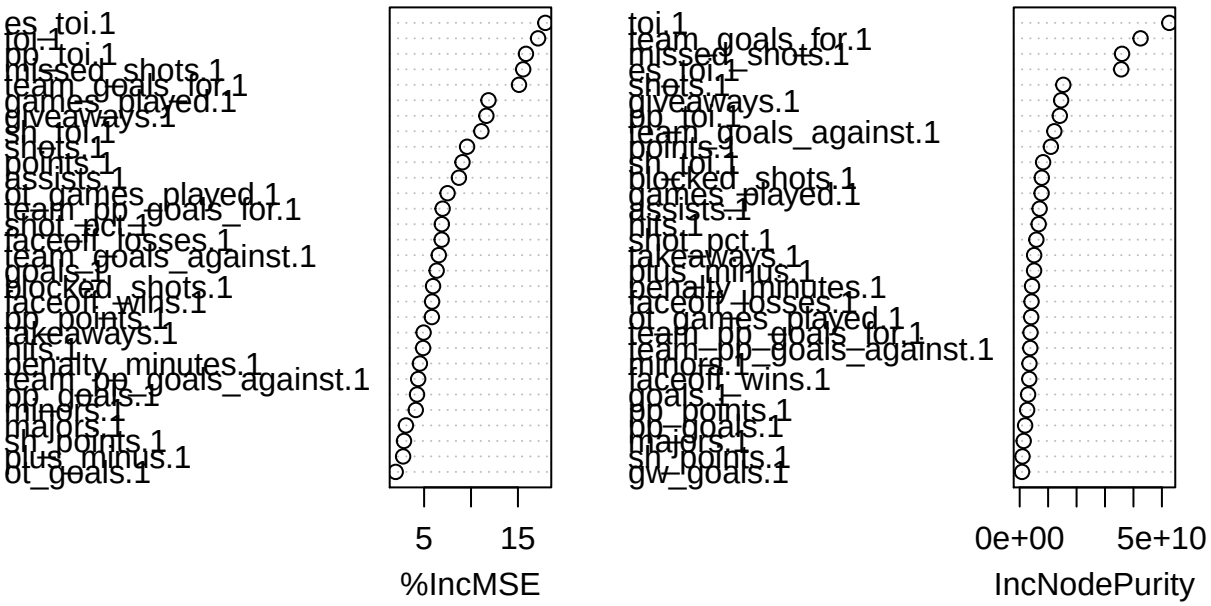
```
esrfmod3 <- nhlBuild(data = rfdata, perc = 0.7, seed = 31415)
```

```
## [1] 0.7582
```

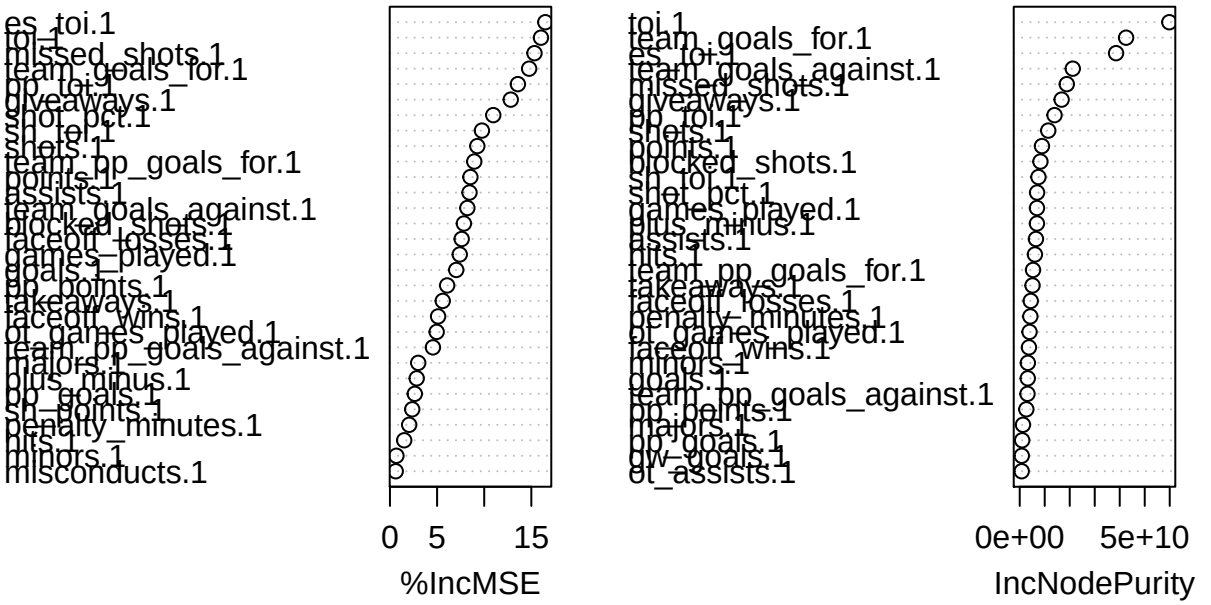
```
esrfmod4 <- nhlBuild(data = rfdata, perc = 0.7, seed = 28182)
```

```
## [1] 0.7336
```

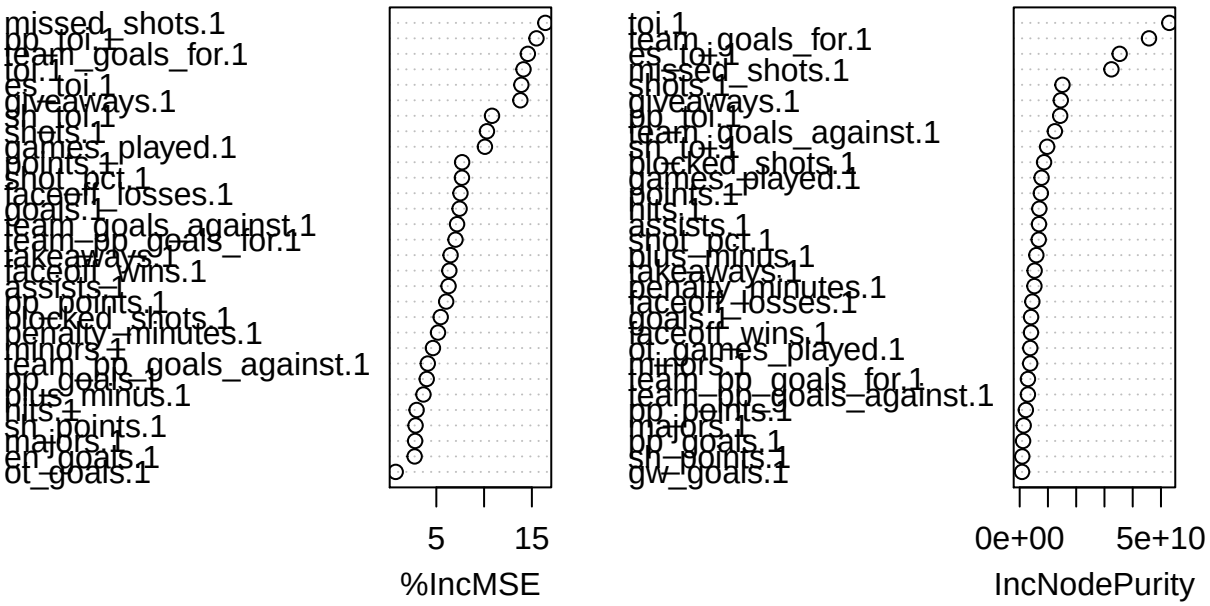
esrfmod1



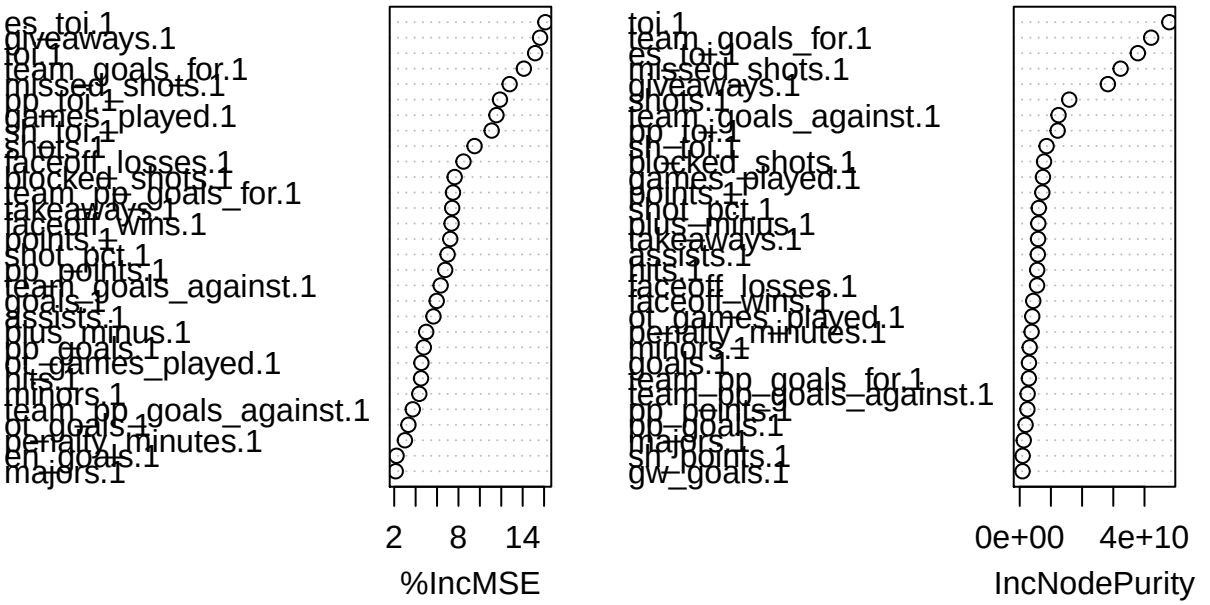
esrfmod2



esrfmod3



esrfmod4




```
esgbmmod1 <- nhlBuild(data = gbmdata, type = "gbm", perc = 0.7, seed = 9112)
esgbmmod2 <- nhlBuild(data = gbmdata, type = "gbm", perc = 0.7, seed = 2857)
esgbmmod3 <- nhlBuild(data = gbmdata, type = "gbm", perc = 0.7, seed = 31415)
```

```
esgbmmod4 <- nhlBuild(data = gbmdata, type = "gbm", perc = 0.7, seed = 28182)
```

```
##                 var var.inf.tot
## 1  team_goals_for.1   57.983263
## 2    missed_shots.1   55.538200
## 3            toi.1   50.723774
## 4         es_toi.1   36.647614
## 5  team_goals_for.2   28.477084
## 6         es_toi.2   18.466111
## 7            toi.2   16.286954
## 8         pp_toi.1   14.056329
## 9       giveaways.1   10.440052
## 10        sh_toi.1    9.118568
## 11         shots.1    7.953250
## 12        pp_toi.2    6.804834
## 13      takeaways.2    6.158915
## 14   games_played.1    5.725170
## 15      giveaways.2    5.269562
## 16  blocked_shots.1    4.489300
## 17  blocked_shots.2    4.332368
## 18          hits.2    4.273493
```

```
col1 <- c(1:3, 15, 24, 30:31, 38:41)
col2 <- c(28:29, 32)
esDataRF1 <- nhlShape(2013, 2013, cols = c(col1, col2[1]), outcome = 38, rm.nhlnum = F)
esDataRF2 <- nhlShape(2013, 2013, cols = c(col1, col2[2]), outcome = 38, rm.nhlnum = F)
esDataRF3 <- nhlShape(2013, 2013, cols = c(col1, col2[3]), outcome = 38, rm.nhlnum = F)
esDataRF4 <- nhlShape(2013, 2013, cols = c(col1, col2), outcome = 38, rm.nhlnum = F)
esDataGBM1 <- nhlShape(2012, 2013, cols = c(col1, col2[1]), outcome = 38,
                       rm.nhlnum = F, rm.NA = FALSE)
esDataGBM1 <- subset(esDataGBM1, !is.na(games_played.1))
esDataGBM2 <- nhlShape(2012, 2013, cols = c(col1, col2[2]), outcome = 38,
                       rm.nhlnum = F, rm.NA = FALSE)
esDataGBM2 <- subset(esDataGBM2, !is.na(games_played.1))
esDataGBM3 <- nhlShape(2012, 2013, cols = c(col1, col2[3]), outcome = 38,
                       rm.nhlnum = F, rm.NA = FALSE)
esDataGBM3 <- subset(esDataGBM3, !is.na(games_played.1))
esDataGBM4 <- nhlShape(2012, 2013, cols = c(col1, col2), outcome = 38,
                       rm.nhlnum = F, rm.NA = FALSE)
esDataGBM4 <- subset(esDataGBM4, !is.na(games_played.1))
```

Here are the low level models.

```
esrfmod1 <- nhlBuild(esDataRF1[, -1], perc = 1, seed = 77677)
esrfmod2 <- nhlBuild(esDataRF2[, -1], perc = 1)
esrfmod3 <- nhlBuild(esDataRF3[, -1], perc = 1)
esrfmod4 <- nhlBuild(esDataRF4[, -1], perc = 1)
esgbmmod1 <- nhlBuild(esDataGBM1[, -1], type = "gbm", perc = 1)
```

```
## Distribution not specified, assuming gaussian ...
```

```
esgbmmod2 <- nhlBuild(esDataGBM2[, -1], type = "gbm", perc = 1)
```

```
## Distribution not specified, assuming gaussian ...
```

```
esgbmmod3 <- nhlBuild(esDataGBM3[, -1], type = "gbm", perc = 1)
```

```
## Distribution not specified, assuming gaussian ...
```

```
esgbmmod4 <- nhlBuild(esDataGBM4[, -1], type = "gbm", perc = 1)
```

```
## Distribution not specified, assuming gaussian ...
```

We build our data frame with the low level predictions in it.

```
esrf1 <- predict(esrfmod1, esDataRF1)
esrf2 <- predict(esrfmod2, esDataRF2)
esrf3 <- predict(esrfmod3, esDataRF3)
esrf4 <- predict(esrfmod4, esDataRF4)
esrf <- as.data.frame(cbind(esDataRF1[, 1], esrf1, esrf2, esrf3, esrf4))
names(esrf) <- c("nhl_num", "esrf1", "esrf2", "esrf3", "esrf4")
esgbm1 <- predict(esgbmmod1, esDataGBM1)
```

```
## Using 9999 trees...
```

```
esgbm2 <- predict(esgbmmod2, esDataGBM2)
```

```
## Using 9991 trees...
```

```
esgbm3 <- predict(esgbmmod3, esDataGBM3)
```

```
## Using 9994 trees...
```

```
esgbm4 <- predict(esgbmmod4, esDataGBM4)
```

```
## Using 9919 trees...
```

```
esgbm <- as.data.frame(cbind(esDataGBM1[, 1], esgbm1, esgbm2, esgbm3, esgbm4))
names(esgbm) <- c("nhl_num", "esgbm1", "esgbm2", "esgbm3", "esgbm4")
esData <- merge(esrf, esgbm, all = TRUE)
esData <- merge(esData, esDataGBM1[, c(1, 22)])
```

The final even strength model is built.

```
esModel <- nhlBuild(esData, type = "gbm", perc = 0.7, seed = 98765)
```

```
## Distribution not specified, assuming gaussian ...
## Using 10000 trees...
## [1] 0.9702938
```

We repeat the overall process for short-handed time next.

```
gbmdata <- nhlShape(2012, 2013, outcome = 39)
rfdata <- nhlShape(2013, 2013, outcome = 39)
```

```
shrfmod1 <- nhlBuild(data = rfdata, perc = 0.7, seed = 9112)
```

```
## [1] 0.7279542
```

```
shrfmod2 <- nhlBuild(data = rfdata, perc = 0.7, seed = 2857)
```

```
## [1] 0.778058
```

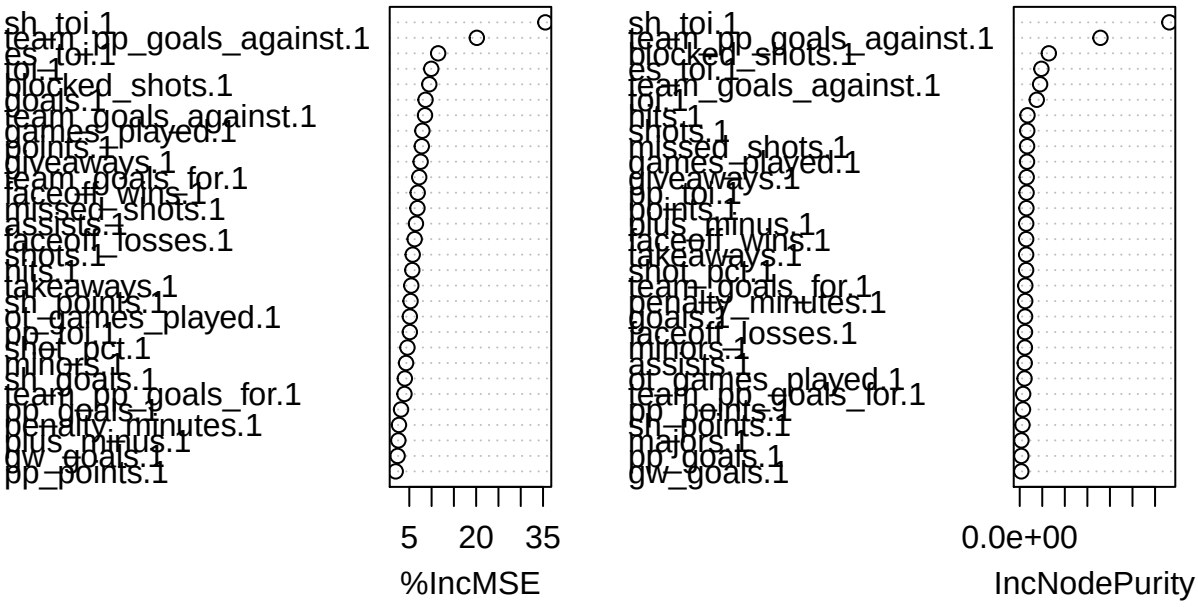
```
shrfmod3 <- nhlBuild(data = rfdata, perc = 0.7, seed = 31415)
```

```
## [1] 0.7535301
```

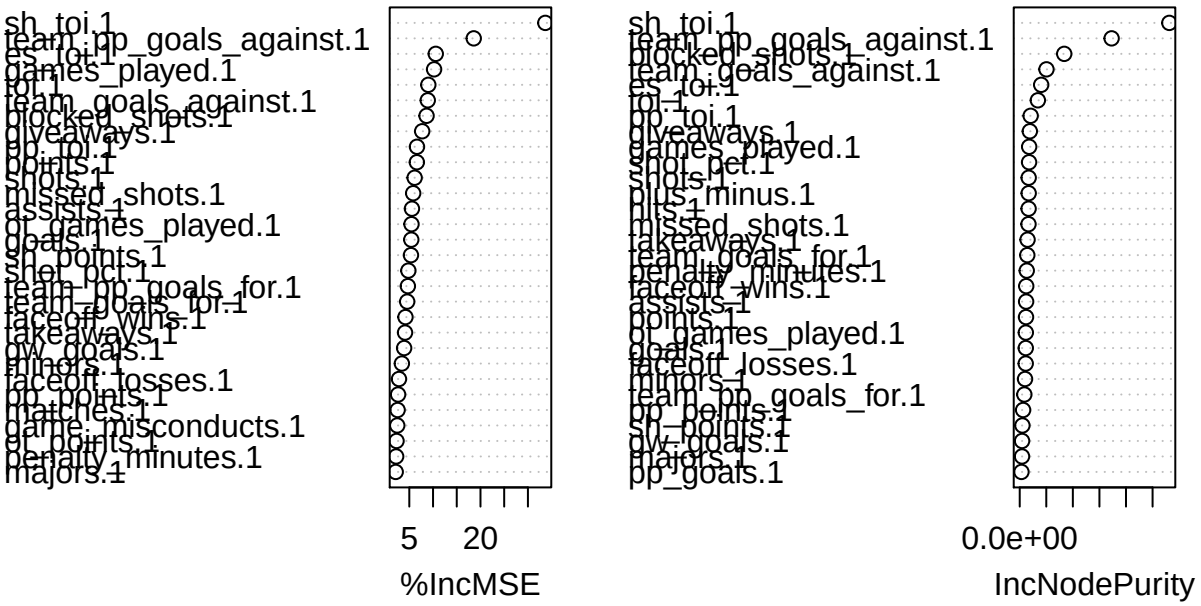
```
shrfmod4 <- nhlBuild(data = rfdata, perc = 0.7, seed = 28182)
```

```
## [1] 0.7803814
```

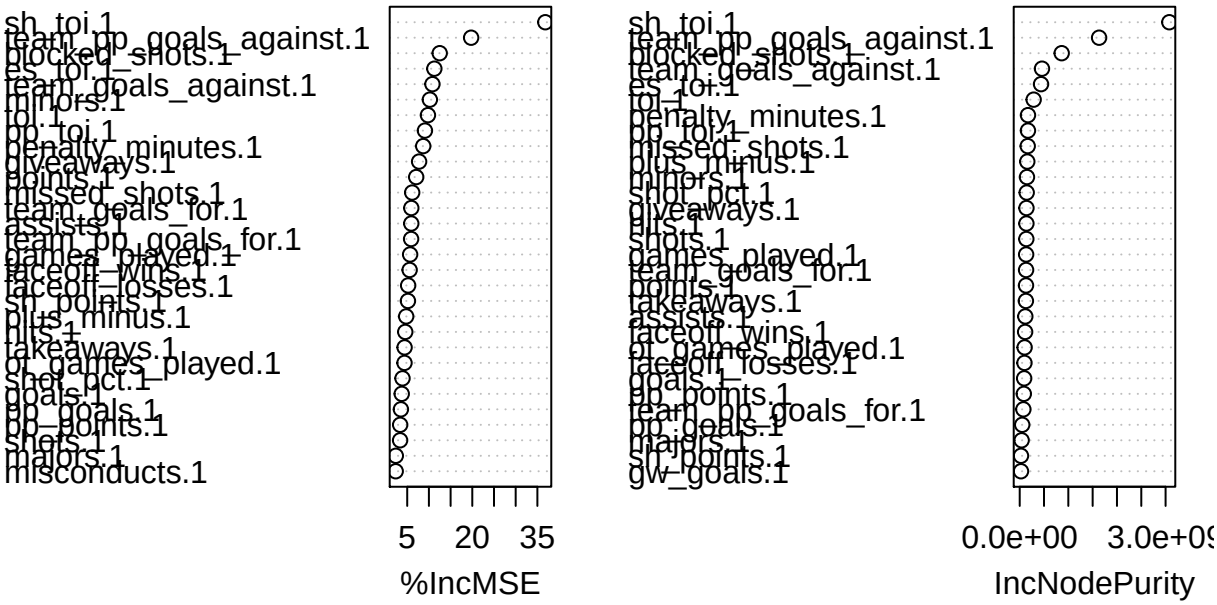
shrfmod1



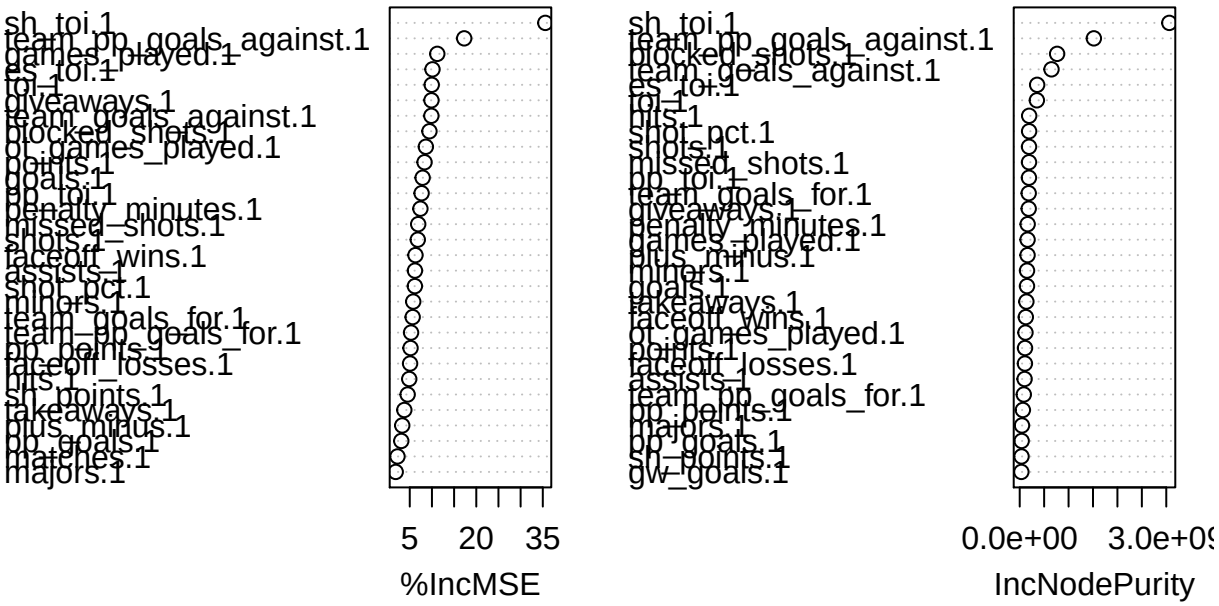
shrfmod2



shrfmod3



shrfmod4




```
shgbmmod1 <- nhlBuild(data = gbmdata, type = "gbm", perc = 0.7, seed = 9112)
shgbmmod2 <- nhlBuild(data = gbmdata, type = "gbm", perc = 0.7, seed = 2857)
shgbmmod3 <- nhlBuild(data = gbmdata, type = "gbm", perc = 0.7, seed = 31415)
shgbmmod4 <- nhlBuild(data = gbmdata, type = "gbm", perc = 0.7, seed = 28182)
```

```
##                         var var.inf.tot
## 1                 sh_toi.1  235.904451
## 2                 sh_toi.2   52.866135
## 3 team_pp_goals_against.1   19.982456
## 4         blocked_shots.1    8.464144
## 5 team_pp_goals_against.2    6.716257
## 6                   toi.1    5.656339
## 7         blocked_shots.2    5.498642
## 8                 es_toi.1    5.282992
```