ML Project Trial

Judy Fordjuoh

April 14, 2022

We aimed to evaluate the association between a broad range of prenatal and postnatal lifestyle and environmental exposures and lung function in children.

Question:Can the quality of the built environment during a mothers pregnancy predict whether an individuals develops asthma within their life?

set.seed(100)  
  
#Trying three different values of mtry (square root, half)  
# since we are not specifying our cross validation, the default is a bootstrap. R is bootstrapping 25 times.  
  
mtry.vals <- c(ncol(train\_asthma) - 1, sqrt(ncol(train\_asthma) - 1), 0.5\*ncol(train\_asthma) - 1)  
  
mtry.grid <- expand.grid(.mtry = mtry.vals)  
rf\_asthma <- train(asthma ~., data = train\_asthma, method = "rf", metric = "Accuracy", tuneGrid = mtry.grid, ntree = 100)  
  
confusionMatrix(rf\_asthma)

## Bootstrapped (25 reps) Confusion Matrix   
##   
## (entries are percentual average cell counts across resamples)  
##   
## Reference  
## Prediction 0 1  
## 0 88.5 10.6  
## 1 0.7 0.2  
##   
## Accuracy (average) : 0.8868

rf\_asthma$results

## mtry Accuracy Kappa AccuracySD KappaSD  
## 1 2.828427 0.8865768 0.005352124 0.01540246 0.03557021  
## 2 3.500000 0.8866975 0.016551433 0.01572857 0.04255275  
## 3 8.000000 0.8788789 0.014351772 0.01668224 0.04911192

rf\_asthma$bestTune

## mtry  
## 2 3.5

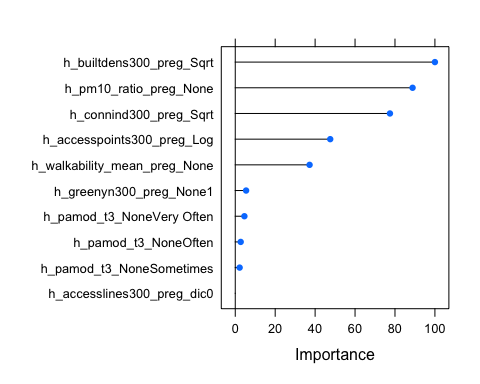
rf\_asthma$finalModel

##   
## Call:  
## randomForest(x = x, y = y, ntree = 100, mtry = min(param$mtry, ncol(x)))   
## Type of random forest: classification  
## Number of trees: 100  
## No. of variables tried at each split: 4  
##   
## OOB estimate of error rate: 11.51%  
## Confusion matrix:  
## 0 1 class.error  
## 0 806 6 0.007389163  
## 1 99 1 0.990000000

varImp(rf\_asthma)

## rf variable importance  
##   
## Overall  
## h\_builtdens300\_preg\_Sqrt 100.000  
## h\_pm10\_ratio\_preg\_None 88.815  
## h\_connind300\_preg\_Sqrt 77.481  
## h\_accesspoints300\_preg\_Log 47.566  
## h\_walkability\_mean\_preg\_None 37.225  
## h\_greenyn300\_preg\_None1 5.430  
## h\_pamod\_t3\_NoneVery Often 4.578  
## h\_pamod\_t3\_NoneOften 2.724  
## h\_pamod\_t3\_NoneSometimes 2.194  
## h\_accesslines300\_preg\_dic0 0.000

plot(varImp(rf\_asthma))



varImpPlot(rf\_asthma$finalModel)

