R\_May8\_SHAO\_weighted\_logistic\_modelselection.R

Administrator

Sat May 09 21:04:24 2015

library(data.table) ## load data in quickly with fread  
library(Metrics)  
rm(list=ls())  
setwd("E:/Dropbox/kaggle/West Nile Virus Prediction")  
x <- fread("data/train\_expand.csv")  
test <- fread("data/test.csv")  
  
## prep the species column by moving the test-only UNSPECIFIED CULEX to CULEX ERRATICUS, and re-doing the levels  
## logistic regression will complain otherwise  
vSpecies<-c(as.character(x$Species),as.character(test$Species))  
vSpecies[vSpecies=="UNSPECIFIED CULEX"]<-"CULEX ERRATICUS"  
vSpecies<-factor(vSpecies,levels=unique(vSpecies))  
  
## data.table syntax for adding a column; could overwrite the existing column as well  
x[,Species2:=factor(vSpecies[1:nrow(x)],levels=unique(vSpecies))]

## Date Address  
## 1: 2007-05-29 4100 North Oak Park Avenue, Chicago, IL 60634, USA  
## 2: 2007-05-29 4100 North Oak Park Avenue, Chicago, IL 60634, USA  
## 3: 2007-05-29 6200 North Mandell Avenue, Chicago, IL 60646, USA  
## 4: 2007-05-29 7900 West Foster Avenue, Chicago, IL 60656, USA  
## 5: 2007-05-29 7900 West Foster Avenue, Chicago, IL 60656, USA  
## ---   
## 135035: 2013-09-26 7100 North Harlem Avenue, Chicago, IL 60631, USA  
## 135036: 2013-09-26 7100 North Harlem Avenue, Chicago, IL 60631, USA  
## 135037: 2013-09-26 7100 North Harlem Avenue, Chicago, IL 60631, USA  
## 135038: 2013-09-26 7100 North Harlem Avenue, Chicago, IL 60631, USA  
## 135039: 2013-09-26 4200 West 65th Street, Chicago, IL 60621, USA  
## Species Block Street Trap  
## 1: CULEX PIPIENS/RESTUANS 41 N OAK PARK AVE T002  
## 2: CULEX RESTUANS 41 N OAK PARK AVE T002  
## 3: CULEX RESTUANS 62 N MANDELL AVE T007  
## 4: CULEX PIPIENS/RESTUANS 79 W FOSTER AVE T015  
## 5: CULEX RESTUANS 79 W FOSTER AVE T015  
## ---   
## 135035: CULEX PIPIENS/RESTUANS 71 N HARLEM AVE T233  
## 135036: CULEX PIPIENS/RESTUANS 71 N HARLEM AVE T233  
## 135037: CULEX PIPIENS/RESTUANS 71 N HARLEM AVE T233  
## 135038: CULEX PIPIENS/RESTUANS 71 N HARLEM AVE T233  
## 135039: CULEX PIPIENS/RESTUANS 42 W 65TH ST T235  
## AddressNumberAndStreet Latitude Longitude  
## 1: 4100 N OAK PARK AVE, Chicago, IL 41.95469 -87.80099  
## 2: 4100 N OAK PARK AVE, Chicago, IL 41.95469 -87.80099  
## 3: 6200 N MANDELL AVE, Chicago, IL 41.99499 -87.76928  
## 4: 7900 W FOSTER AVE, Chicago, IL 41.97409 -87.82481  
## 5: 7900 W FOSTER AVE, Chicago, IL 41.97409 -87.82481  
## ---   
## 135035: 7100 N HARLEM AVE, Chicago, IL 42.00988 -87.80728  
## 135036: 7100 N HARLEM AVE, Chicago, IL 42.00988 -87.80728  
## 135037: 7100 N HARLEM AVE, Chicago, IL 42.00988 -87.80728  
## 135038: 7100 N HARLEM AVE, Chicago, IL 42.00988 -87.80728  
## 135039: 4200 W 65TH ST, Chicago, IL 41.77643 -87.62710  
## AddressAccuracy NumMosquitos WnvPresent Species2  
## 1: 9 1 0 CULEX PIPIENS/RESTUANS  
## 2: 9 1 0 CULEX RESTUANS  
## 3: 9 1 0 CULEX RESTUANS  
## 4: 8 1 0 CULEX PIPIENS/RESTUANS  
## 5: 8 4 0 CULEX RESTUANS  
## ---   
## 135035: 9 5 0 CULEX PIPIENS/RESTUANS  
## 135036: 9 5 0 CULEX PIPIENS/RESTUANS  
## 135037: 9 5 0 CULEX PIPIENS/RESTUANS  
## 135038: 9 5 0 CULEX PIPIENS/RESTUANS  
## 135039: 8 1 0 CULEX PIPIENS/RESTUANS

test[,Species2:=factor(vSpecies[(nrow(x)+1):length(vSpecies)],levels=unique(vSpecies))]

## Id Date  
## 1: 1 2008-06-11  
## 2: 2 2008-06-11  
## 3: 3 2008-06-11  
## 4: 4 2008-06-11  
## 5: 5 2008-06-11  
## ---   
## 116289: 116289 2014-10-02  
## 116290: 116290 2014-10-02  
## 116291: 116291 2014-10-02  
## 116292: 116292 2014-10-02  
## 116293: 116293 2014-10-02  
## Address  
## 1: 4100 North Oak Park Avenue, Chicago, IL 60634, USA  
## 2: 4100 North Oak Park Avenue, Chicago, IL 60634, USA  
## 3: 4100 North Oak Park Avenue, Chicago, IL 60634, USA  
## 4: 4100 North Oak Park Avenue, Chicago, IL 60634, USA  
## 5: 4100 North Oak Park Avenue, Chicago, IL 60634, USA  
## ---   
## 116289: 2100 North Cannon Drive, Chicago, IL 60614, USA  
## 116290: 2100 North Cannon Drive, Chicago, IL 60614, USA  
## 116291: 2100 North Cannon Drive, Chicago, IL 60614, USA  
## 116292: 2100 North Cannon Drive, Chicago, IL 60614, USA  
## 116293: 2100 North Cannon Drive, Chicago, IL 60614, USA  
## Species Block Street Trap  
## 1: CULEX PIPIENS/RESTUANS 41 N OAK PARK AVE T002  
## 2: CULEX RESTUANS 41 N OAK PARK AVE T002  
## 3: CULEX PIPIENS 41 N OAK PARK AVE T002  
## 4: CULEX SALINARIUS 41 N OAK PARK AVE T002  
## 5: CULEX TERRITANS 41 N OAK PARK AVE T002  
## ---   
## 116289: CULEX SALINARIUS 21 N CANNON DR T054C  
## 116290: CULEX TERRITANS 21 N CANNON DR T054C  
## 116291: CULEX TARSALIS 21 N CANNON DR T054C  
## 116292: UNSPECIFIED CULEX 21 N CANNON DR T054C  
## 116293: CULEX ERRATICUS 21 N CANNON DR T054C  
## AddressNumberAndStreet Latitude Longitude  
## 1: 4100 N OAK PARK AVE, Chicago, IL 41.95469 -87.80099  
## 2: 4100 N OAK PARK AVE, Chicago, IL 41.95469 -87.80099  
## 3: 4100 N OAK PARK AVE, Chicago, IL 41.95469 -87.80099  
## 4: 4100 N OAK PARK AVE, Chicago, IL 41.95469 -87.80099  
## 5: 4100 N OAK PARK AVE, Chicago, IL 41.95469 -87.80099  
## ---   
## 116289: 2100 N CANNON DR, Chicago, IL 41.92565 -87.63359  
## 116290: 2100 N CANNON DR, Chicago, IL 41.92565 -87.63359  
## 116291: 2100 N CANNON DR, Chicago, IL 41.92565 -87.63359  
## 116292: 2100 N CANNON DR, Chicago, IL 41.92565 -87.63359  
## 116293: 2100 N CANNON DR, Chicago, IL 41.92565 -87.63359  
## AddressAccuracy Species2  
## 1: 9 CULEX PIPIENS/RESTUANS  
## 2: 9 CULEX RESTUANS  
## 3: 9 CULEX PIPIENS  
## 4: 9 CULEX SALINARIUS  
## 5: 9 CULEX TERRITANS  
## ---   
## 116289: 8 CULEX SALINARIUS  
## 116290: 8 CULEX TERRITANS  
## 116291: 8 CULEX TARSALIS  
## 116292: 8 CULEX ERRATICUS  
## 116293: 8 CULEX ERRATICUS

## also add some fields for components of the date using simple substrings  
x[,dMonth:=substr(x$Date,6,7)]

## Date Address  
## 1: 2007-05-29 4100 North Oak Park Avenue, Chicago, IL 60634, USA  
## 2: 2007-05-29 4100 North Oak Park Avenue, Chicago, IL 60634, USA  
## 3: 2007-05-29 6200 North Mandell Avenue, Chicago, IL 60646, USA  
## 4: 2007-05-29 7900 West Foster Avenue, Chicago, IL 60656, USA  
## 5: 2007-05-29 7900 West Foster Avenue, Chicago, IL 60656, USA  
## ---   
## 135035: 2013-09-26 7100 North Harlem Avenue, Chicago, IL 60631, USA  
## 135036: 2013-09-26 7100 North Harlem Avenue, Chicago, IL 60631, USA  
## 135037: 2013-09-26 7100 North Harlem Avenue, Chicago, IL 60631, USA  
## 135038: 2013-09-26 7100 North Harlem Avenue, Chicago, IL 60631, USA  
## 135039: 2013-09-26 4200 West 65th Street, Chicago, IL 60621, USA  
## Species Block Street Trap  
## 1: CULEX PIPIENS/RESTUANS 41 N OAK PARK AVE T002  
## 2: CULEX RESTUANS 41 N OAK PARK AVE T002  
## 3: CULEX RESTUANS 62 N MANDELL AVE T007  
## 4: CULEX PIPIENS/RESTUANS 79 W FOSTER AVE T015  
## 5: CULEX RESTUANS 79 W FOSTER AVE T015  
## ---   
## 135035: CULEX PIPIENS/RESTUANS 71 N HARLEM AVE T233  
## 135036: CULEX PIPIENS/RESTUANS 71 N HARLEM AVE T233  
## 135037: CULEX PIPIENS/RESTUANS 71 N HARLEM AVE T233  
## 135038: CULEX PIPIENS/RESTUANS 71 N HARLEM AVE T233  
## 135039: CULEX PIPIENS/RESTUANS 42 W 65TH ST T235  
## AddressNumberAndStreet Latitude Longitude  
## 1: 4100 N OAK PARK AVE, Chicago, IL 41.95469 -87.80099  
## 2: 4100 N OAK PARK AVE, Chicago, IL 41.95469 -87.80099  
## 3: 6200 N MANDELL AVE, Chicago, IL 41.99499 -87.76928  
## 4: 7900 W FOSTER AVE, Chicago, IL 41.97409 -87.82481  
## 5: 7900 W FOSTER AVE, Chicago, IL 41.97409 -87.82481  
## ---   
## 135035: 7100 N HARLEM AVE, Chicago, IL 42.00988 -87.80728  
## 135036: 7100 N HARLEM AVE, Chicago, IL 42.00988 -87.80728  
## 135037: 7100 N HARLEM AVE, Chicago, IL 42.00988 -87.80728  
## 135038: 7100 N HARLEM AVE, Chicago, IL 42.00988 -87.80728  
## 135039: 4200 W 65TH ST, Chicago, IL 41.77643 -87.62710  
## AddressAccuracy NumMosquitos WnvPresent Species2  
## 1: 9 1 0 CULEX PIPIENS/RESTUANS  
## 2: 9 1 0 CULEX RESTUANS  
## 3: 9 1 0 CULEX RESTUANS  
## 4: 8 1 0 CULEX PIPIENS/RESTUANS  
## 5: 8 4 0 CULEX RESTUANS  
## ---   
## 135035: 9 5 0 CULEX PIPIENS/RESTUANS  
## 135036: 9 5 0 CULEX PIPIENS/RESTUANS  
## 135037: 9 5 0 CULEX PIPIENS/RESTUANS  
## 135038: 9 5 0 CULEX PIPIENS/RESTUANS  
## 135039: 8 1 0 CULEX PIPIENS/RESTUANS  
## dMonth  
## 1: 05  
## 2: 05  
## 3: 05  
## 4: 05  
## 5: 05  
## ---   
## 135035: 09  
## 135036: 09  
## 135037: 09  
## 135038: 09  
## 135039: 09

x[,dYear:=substr(x$Date,1,4)]

## Date Address  
## 1: 2007-05-29 4100 North Oak Park Avenue, Chicago, IL 60634, USA  
## 2: 2007-05-29 4100 North Oak Park Avenue, Chicago, IL 60634, USA  
## 3: 2007-05-29 6200 North Mandell Avenue, Chicago, IL 60646, USA  
## 4: 2007-05-29 7900 West Foster Avenue, Chicago, IL 60656, USA  
## 5: 2007-05-29 7900 West Foster Avenue, Chicago, IL 60656, USA  
## ---   
## 135035: 2013-09-26 7100 North Harlem Avenue, Chicago, IL 60631, USA  
## 135036: 2013-09-26 7100 North Harlem Avenue, Chicago, IL 60631, USA  
## 135037: 2013-09-26 7100 North Harlem Avenue, Chicago, IL 60631, USA  
## 135038: 2013-09-26 7100 North Harlem Avenue, Chicago, IL 60631, USA  
## 135039: 2013-09-26 4200 West 65th Street, Chicago, IL 60621, USA  
## Species Block Street Trap  
## 1: CULEX PIPIENS/RESTUANS 41 N OAK PARK AVE T002  
## 2: CULEX RESTUANS 41 N OAK PARK AVE T002  
## 3: CULEX RESTUANS 62 N MANDELL AVE T007  
## 4: CULEX PIPIENS/RESTUANS 79 W FOSTER AVE T015  
## 5: CULEX RESTUANS 79 W FOSTER AVE T015  
## ---   
## 135035: CULEX PIPIENS/RESTUANS 71 N HARLEM AVE T233  
## 135036: CULEX PIPIENS/RESTUANS 71 N HARLEM AVE T233  
## 135037: CULEX PIPIENS/RESTUANS 71 N HARLEM AVE T233  
## 135038: CULEX PIPIENS/RESTUANS 71 N HARLEM AVE T233  
## 135039: CULEX PIPIENS/RESTUANS 42 W 65TH ST T235  
## AddressNumberAndStreet Latitude Longitude  
## 1: 4100 N OAK PARK AVE, Chicago, IL 41.95469 -87.80099  
## 2: 4100 N OAK PARK AVE, Chicago, IL 41.95469 -87.80099  
## 3: 6200 N MANDELL AVE, Chicago, IL 41.99499 -87.76928  
## 4: 7900 W FOSTER AVE, Chicago, IL 41.97409 -87.82481  
## 5: 7900 W FOSTER AVE, Chicago, IL 41.97409 -87.82481  
## ---   
## 135035: 7100 N HARLEM AVE, Chicago, IL 42.00988 -87.80728  
## 135036: 7100 N HARLEM AVE, Chicago, IL 42.00988 -87.80728  
## 135037: 7100 N HARLEM AVE, Chicago, IL 42.00988 -87.80728  
## 135038: 7100 N HARLEM AVE, Chicago, IL 42.00988 -87.80728  
## 135039: 4200 W 65TH ST, Chicago, IL 41.77643 -87.62710  
## AddressAccuracy NumMosquitos WnvPresent Species2  
## 1: 9 1 0 CULEX PIPIENS/RESTUANS  
## 2: 9 1 0 CULEX RESTUANS  
## 3: 9 1 0 CULEX RESTUANS  
## 4: 8 1 0 CULEX PIPIENS/RESTUANS  
## 5: 8 4 0 CULEX RESTUANS  
## ---   
## 135035: 9 5 0 CULEX PIPIENS/RESTUANS  
## 135036: 9 5 0 CULEX PIPIENS/RESTUANS  
## 135037: 9 5 0 CULEX PIPIENS/RESTUANS  
## 135038: 9 5 0 CULEX PIPIENS/RESTUANS  
## 135039: 8 1 0 CULEX PIPIENS/RESTUANS  
## dMonth dYear  
## 1: 05 2007  
## 2: 05 2007  
## 3: 05 2007  
## 4: 05 2007  
## 5: 05 2007  
## ---   
## 135035: 09 2013  
## 135036: 09 2013  
## 135037: 09 2013  
## 135038: 09 2013  
## 135039: 09 2013

test[,dMonth:=substr(test$Date,6,7)]

## Id Date  
## 1: 1 2008-06-11  
## 2: 2 2008-06-11  
## 3: 3 2008-06-11  
## 4: 4 2008-06-11  
## 5: 5 2008-06-11  
## ---   
## 116289: 116289 2014-10-02  
## 116290: 116290 2014-10-02  
## 116291: 116291 2014-10-02  
## 116292: 116292 2014-10-02  
## 116293: 116293 2014-10-02  
## Address  
## 1: 4100 North Oak Park Avenue, Chicago, IL 60634, USA  
## 2: 4100 North Oak Park Avenue, Chicago, IL 60634, USA  
## 3: 4100 North Oak Park Avenue, Chicago, IL 60634, USA  
## 4: 4100 North Oak Park Avenue, Chicago, IL 60634, USA  
## 5: 4100 North Oak Park Avenue, Chicago, IL 60634, USA  
## ---   
## 116289: 2100 North Cannon Drive, Chicago, IL 60614, USA  
## 116290: 2100 North Cannon Drive, Chicago, IL 60614, USA  
## 116291: 2100 North Cannon Drive, Chicago, IL 60614, USA  
## 116292: 2100 North Cannon Drive, Chicago, IL 60614, USA  
## 116293: 2100 North Cannon Drive, Chicago, IL 60614, USA  
## Species Block Street Trap  
## 1: CULEX PIPIENS/RESTUANS 41 N OAK PARK AVE T002  
## 2: CULEX RESTUANS 41 N OAK PARK AVE T002  
## 3: CULEX PIPIENS 41 N OAK PARK AVE T002  
## 4: CULEX SALINARIUS 41 N OAK PARK AVE T002  
## 5: CULEX TERRITANS 41 N OAK PARK AVE T002  
## ---   
## 116289: CULEX SALINARIUS 21 N CANNON DR T054C  
## 116290: CULEX TERRITANS 21 N CANNON DR T054C  
## 116291: CULEX TARSALIS 21 N CANNON DR T054C  
## 116292: UNSPECIFIED CULEX 21 N CANNON DR T054C  
## 116293: CULEX ERRATICUS 21 N CANNON DR T054C  
## AddressNumberAndStreet Latitude Longitude  
## 1: 4100 N OAK PARK AVE, Chicago, IL 41.95469 -87.80099  
## 2: 4100 N OAK PARK AVE, Chicago, IL 41.95469 -87.80099  
## 3: 4100 N OAK PARK AVE, Chicago, IL 41.95469 -87.80099  
## 4: 4100 N OAK PARK AVE, Chicago, IL 41.95469 -87.80099  
## 5: 4100 N OAK PARK AVE, Chicago, IL 41.95469 -87.80099  
## ---   
## 116289: 2100 N CANNON DR, Chicago, IL 41.92565 -87.63359  
## 116290: 2100 N CANNON DR, Chicago, IL 41.92565 -87.63359  
## 116291: 2100 N CANNON DR, Chicago, IL 41.92565 -87.63359  
## 116292: 2100 N CANNON DR, Chicago, IL 41.92565 -87.63359  
## 116293: 2100 N CANNON DR, Chicago, IL 41.92565 -87.63359  
## AddressAccuracy Species2 dMonth  
## 1: 9 CULEX PIPIENS/RESTUANS 06  
## 2: 9 CULEX RESTUANS 06  
## 3: 9 CULEX PIPIENS 06  
## 4: 9 CULEX SALINARIUS 06  
## 5: 9 CULEX TERRITANS 06  
## ---   
## 116289: 8 CULEX SALINARIUS 10  
## 116290: 8 CULEX TERRITANS 10  
## 116291: 8 CULEX TARSALIS 10  
## 116292: 8 CULEX ERRATICUS 10  
## 116293: 8 CULEX ERRATICUS 10

### Start modeling  
## use 2011 as a cross validation year; x1 will include the other three years;x2 will include 2011  
x1<-x[dYear!=2011,]  
x2<-x[dYear==2011,]  
  
  
## fit a logistic regression model using just three key fields  
fitCv<-glm(WnvPresent ~ dMonth + Species2 + Block+Latitude+Longitude, data = x1, family = "binomial")  
summary(fitCv)

##   
## Call:  
## glm(formula = WnvPresent ~ dMonth + Species2 + Block + Latitude +   
## Longitude, family = "binomial", data = x1)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.1062 -0.5403 -0.4183 -0.1573 3.5783   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -3.934e+02 9.460e+01 -4.158 3.21e-05 \*\*\*  
## dMonth06 9.358e+00 9.408e+01 0.099 0.921   
## dMonth07 1.235e+01 9.408e+01 0.131 0.896   
## dMonth08 1.393e+01 9.408e+01 0.148 0.882   
## dMonth09 1.339e+01 9.408e+01 0.142 0.887   
## dMonth10 1.063e+01 9.408e+01 0.113 0.910   
## Species2CULEX RESTUANS -7.315e-01 4.727e-02 -15.477 < 2e-16 \*\*\*  
## Species2CULEX PIPIENS 3.041e-01 2.135e-02 14.243 < 2e-16 \*\*\*  
## Species2CULEX SALINARIUS -1.319e+01 1.431e+02 -0.092 0.927   
## Species2CULEX TERRITANS -1.328e+01 7.262e+01 -0.183 0.855   
## Species2CULEX TARSALIS -1.331e+01 1.029e+03 -0.013 0.990   
## Species2CULEX ERRATICUS -1.434e+01 5.501e+02 -0.026 0.979   
## Block 3.467e-03 4.065e-04 8.531 < 2e-16 \*\*\*  
## Latitude 1.411e+00 1.346e-01 10.486 < 2e-16 \*\*\*  
## Longitude -3.635e+00 1.559e-01 -23.311 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 82789 on 116239 degrees of freedom  
## Residual deviance: 72175 on 116225 degrees of freedom  
## AIC: 72205  
##   
## Number of Fisher Scoring iterations: 14

upper=formula(WnvPresent ~dMonth + Species2 + Block+Latitude+Longitude+Latitude^2+Longitude^2+Latitude:Longitude+Block:Longitude:Latitude+Species2:dMonth)

fitCvselection=step(fitCv, scope=list(lower=formula(WnvPresent ~1), upper=upper), trace=TRUE, direction="both")

## Start: AIC=72205.36  
## WnvPresent ~ dMonth + Species2 + Block + Latitude + Longitude  
##   
## Df Deviance AIC  
## + dMonth:Species2 19 71917 71985  
## <none> 72175 72205  
## + Latitude:Longitude 1 72175 72207  
## - Block 1 72247 72275  
## - Latitude 1 72286 72314  
## - Longitude 1 72719 72747  
## - Species2 6 72858 72876  
## - dMonth 5 78291 78311  
##   
## Step: AIC=71985.23  
## WnvPresent ~ dMonth + Species2 + Block + Latitude + Longitude +   
## dMonth:Species2  
##   
## Df Deviance AIC  
## <none> 71917 71985  
## + Latitude:Longitude 1 71917 71987  
## - Block 1 71987 72053  
## - Latitude 1 72019 72085  
## - dMonth:Species2 19 72175 72205  
## - Longitude 1 72475 72541

fitCvselection$anova

## Step Df Deviance Resid. Df Resid. Dev AIC  
## 1 NA NA 116225 72175.36 72205.36  
## 2 + dMonth:Species2 -19 258.1213 116206 71917.23 71985.23

p2<-predict(fitCvselection, newdata = x2, type = "response")

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type =  
## ifelse(type == : prediction from a rank-deficient fit may be misleading

#p2<-predict(fitCv, newdata = x2, type = "response")  
auc(x2$WnvPresent,p2)

## [1] 0.8365235

fitSubmit<-glm(WnvPresent ~ dMonth + Species2 + Block+Latitude+Longitude+dMonth:Species2, data = x, family = "binomial")  
#final=step(fitSubmit, scope=list(lower=formula(fitSubmit), upper= ~ .^2), trace=TRUE, direction="both")  
  
pSubmit<-predict(fitSubmit, newdata = test, type = "response")

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type =  
## ifelse(type == : prediction from a rank-deficient fit may be misleading

submissionFile<-cbind(test$Id,pSubmit)  
colnames(submissionFile)<-c("Id","WnvPresent")  
options("scipen"=100, "digits"=8)  
write.csv(submissionFile,"exapnd\_logistic\_selection1.csv",row.names=FALSE,quote=FALSE)