H1N1_2

AYA

2025-03-11

CLEAR THE ENVIRONMENT

```
rm(list  =ls())
```

LOAD THE NECESSARY LIBRARIES

```
library(mlbench)
library(caret)
```

```
## Loading required package: ggplot2
```

```
## Loading required package: lattice
```

```
#install.packages("caTools")
library(caTools)
#install.packages("ranger")
library(ranger)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
#install.packages("doParallel")
library(doParallel)
```

```
## Loading required package: foreach
```

```
## Loading required package: iterators
```

```
## Loading required package: parallel
```

LOAD THE DATASET

```
#load the training features
training_features <- read.csv("training_set_features.csv", row.names = 1, header = T, stringsAsFactors =
head(training_features,5)
```

```
##   h1n1_concern h1n1_knowledge behavioral_antiviral_meds behavioral_avoidance
## 0            1              0                         0                    0
## 1            3              2                         0                    1
## 2            1              1                         0                    1
## 3            1              1                         0                    1
## 4            2              1                         0                    1
##   behavioral_face_mask behavioral_wash_hands behavioral_large_gatherings
## 0                    0                     0                           0
## 1                    0                     1                           0
## 2                    0                     0                           0
## 3                    0                     1                           1
## 4                    0                     1                           1
##   behavioral_outside_home behavioral_touch_face doctor_recc_h1n1
## 0                       1                     1               0
## 1                       1                     1               0
## 2                       0                     0              NA
## 3                       0                     0               0
## 4                       0                     1               0
##   doctor_recc_seasonal chronic_med_condition child_under_6_months health_worker
## 0                    0                     0                     0            0
## 1                    0                     0                     0            0
## 2                   NA                     1                     0            0
## 3                    1                     1                     0            0
## 4                    0                     0                     0            0
##   health_insurance opinion_h1n1_vacc_effective opinion_h1n1_risk
## 0                1                           3                 1
## 1                1                           5                 4
## 2               NA                           3                 1
## 3               NA                           3                 3
## 4               NA                           3                 3
##   opinion_h1n1_sick_from_vacc opinion_seas_vacc_effective opinion_seas_risk
## 0                           2                           2                 1
## 1                           4                           4                 2
## 2                           1                           4                 1
## 3                           5                           5                 4
## 4                           2                           3                 1
##   opinion_seas_sick_from_vacc     age_group        education  race    sex
## 0                           2 55 - 64 Years       < 12 Years White Female
## 1                           4 35 - 44 Years         12 Years White   Male
## 2                           2 18 - 34 Years College Graduate White   Male
## 3                           1     65+ Years         12 Years White Female
## 4                           4 45 - 54 Years     Some College White Female
##              income_poverty marital_status rent_or_own  employment_status
## 0             Below Poverty    Not Married         Own Not in Labor Force
## 1             Below Poverty    Not Married        Rent           Employed
## 2 <= $75,000, Above Poverty    Not Married         Own           Employed
## 3             Below Poverty    Not Married        Rent Not in Labor Force
## 4 <= $75,000, Above Poverty        Married         Own           Employed
##   hhs_geo_region               census_msa household_adults household_children
## 0       oxchjgsf                  Non-MSA               0                  0
## 1       bhuqouqj MSA, Not Principle  City               0                  0
## 2       qufhixun MSA, Not Principle  City               2                  0
```

```
## 3       lrircsnp      MSA, Principle City                0                 0
## 4       qufhixun MSA, Not Principle  City                1                 0
##   employment_industry employment_occupation
## 0
## 1            pxcmvdjn             xgwztkwe
## 2            rucpziij             xtkaffoo
## 3
## 4            wxleyezf             emcorrxb
```

```
#load the validation labels
training_labels <- read.csv("training_set_labels.csv", row.names = 1, header = T, stringsAsFactors = T)
head(training_labels,5)
```

```
##   h1n1_vaccine seasonal_vaccine
## 0            0               0
## 1            0               1
## 2            0               0
## 3            0               1
## 4            0               0
```

```
#combining labels
h1n1 <- cbind(training_features, training_labels[1:2])
```

```
head(h1n1,5)
```

```
##   h1n1_concern h1n1_knowledge behavioral_antiviral_meds behavioral_avoidance
## 0            1              0                         0                    0
## 1            3              2                         0                    1
## 2            1              1                         0                    1
## 3            1              1                         0                    1
## 4            2              1                         0                    1
##   behavioral_face_mask behavioral_wash_hands behavioral_large_gatherings
## 0                    0                     0                           0
## 1                    0                     1                           0
## 2                    0                     0                           0
## 3                    0                     1                           1
## 4                    0                     1                           1
##   behavioral_outside_home behavioral_touch_face doctor_recc_h1n1
## 0                       1                     1               0
## 1                       1                     1               0
## 2                       0                     0              NA
## 3                       0                     0               0
## 4                       0                     1               0
##   doctor_recc_seasonal chronic_med_condition child_under_6_months health_worker
## 0                    0                     0                     0             0
## 1                    0                     0                     0             0
## 2                   NA                     1                     0             0
## 3                    1                     1                     0             0
## 4                    0                     0                     0             0
##   health_insurance opinion_h1n1_vacc_effective opinion_h1n1_risk
## 0                1                           3                 1
## 1                1                           5                 4
## 2               NA                           3                 1
## 3               NA                           3                 3
## 4               NA                           3                 3
##   opinion_h1n1_sick_from_vacc opinion_seas_vacc_effective opinion_seas_risk
```

```
## 0                           2                           2                1
## 1                           4                           4                2
## 2                           1                           4                1
## 3                           5                           5                4
## 4                           2                           3                1
##   opinion_seas_sick_from_vacc     age_group        education  race    sex
## 0                           2 55 - 64 Years       < 12 Years White Female
## 1                           4 35 - 44 Years         12 Years White   Male
## 2                           2 18 - 34 Years College Graduate White   Male
## 3                           1     65+ Years         12 Years White Female
## 4                           4 45 - 54 Years     Some College White Female
##              income_poverty marital_status rent_or_own  employment_status
## 0             Below Poverty    Not Married         Own Not in Labor Force
## 1             Below Poverty    Not Married        Rent           Employed
## 2 <= $75,000, Above Poverty    Not Married         Own           Employed
## 3             Below Poverty    Not Married        Rent Not in Labor Force
## 4 <= $75,000, Above Poverty        Married         Own           Employed
##   hhs_geo_region               census_msa household_adults household_children
## 0       oxchjgsf                  Non-MSA                0                 0
## 1       bhuqouqj MSA, Not Principle  City                0                 0
## 2       qufhixun MSA, Not Principle  City                2                 0
## 3       lrircsnp      MSA, Principle City                0                 0
## 4       qufhixun MSA, Not Principle  City                1                 0
##   employment_industry employment_occupation h1n1_vaccine seasonal_vaccine
## 0                                                     0                0
## 1            pxcmvdjn              xgwztkwe            0                1
## 2            rucpziij              xtkaffoo            0                0
## 3                                                     0                1
## 4            wxleyezf              emcorrxb            0                0
```

DATA CLEANING

```r
#sum of all NA values in each dataset
colSums(is.na(h1n1))
```

```
##                h1n1_concern             h1n1_knowledge
##                          92                        116
##   behavioral_antiviral_meds       behavioral_avoidance
##                          71                        208
##       behavioral_face_mask      behavioral_wash_hands
##                          19                         42
## behavioral_large_gatherings    behavioral_outside_home
##                          87                         82
##      behavioral_touch_face           doctor_recc_h1n1
##                         128                       2160
##       doctor_recc_seasonal      chronic_med_condition
##                       2160                        971
##       child_under_6_months             health_worker
##                        820                        804
##           health_insurance opinion_h1n1_vacc_effective
##                      12274                        391
##          opinion_h1n1_risk opinion_h1n1_sick_from_vacc
##                        388                        395
```

```
## opinion_seas_vacc_effective           opinion_seas_risk
##                          462                         514
## opinion_seas_sick_from_vacc                   age_group
##                          537                           0
##                   education                        race
##                           0                           0
##                         sex                income_poverty
##                           0                           0
##              marital_status                  rent_or_own
##                           0                           0
##           employment_status                hhs_geo_region
##                           0                           0
##                  census_msa              household_adults
##                           0                         249
##          household_children          employment_industry
##                         249                           0
##       employment_occupation                 h1n1_vaccine
##                           0                           0
##            seasonal_vaccine
##                           0
```

health insurance has the largest NA values 12274. Filling up thr data might cause inaccuracies so we will drop the column "health_insurance".

```
#removing "health_insurance" column from the dataset
h1n1$health_insurance <- NULL

#also dropping "hhs_goe_region", "employment_industry" and "employment_occupation"
h1n1$hhs_geo_region <- NULL
h1n1$employment_industry <- NULL
h1n1$employment_occupation <- NULL
```

```
#dropping all NA values in the dataset
h1n1 <- na.omit(h1n1)

#checking the dimension of NA after dropping all the values
dim(h1n1)
```

```
## [1] 22976    33
```

From 26707 entries, we now have a total of 22976(a difference of 3731).

CONVERTING NOMINAL DATA INTO NUMERIC DATA

```
#duplicating dataset
h1n1_2 <- h1n1

h1n1_2[,32:33] <- NULL
head(h1n1_2,5)
```

```
##   h1n1_concern h1n1_knowledge behavioral_antiviral_meds behavioral_avoidance
## 0            1              0                         0                    0
## 1            3              2                         0                    1
## 3            1              1                         0                    1
## 4            2              1                         0                    1
```

```
## 5            3             1                        0                    1
##   behavioral_face_mask behavioral_wash_hands behavioral_large_gatherings
## 0                    0                    0                           0
## 1                    0                    1                           0
## 3                    0                    1                           1
## 4                    0                    1                           1
## 5                    0                    1                           0
##   behavioral_outside_home behavioral_touch_face doctor_recc_h1n1
## 0                       1                    1                0
## 1                       1                    1                0
## 3                       0                    0                0
## 4                       0                    1                0
## 5                       0                    1                0
##   doctor_recc_seasonal chronic_med_condition child_under_6_months health_worker
## 0                    0                    0                     0             0
## 1                    0                    0                     0             0
## 3                    1                    1                     0             0
## 4                    0                    0                     0             0
## 5                    1                    0                     0             0
##   opinion_h1n1_vacc_effective opinion_h1n1_risk opinion_h1n1_sick_from_vacc
## 0                           3                1                           2
## 1                           5                4                           4
## 3                           3                3                           5
## 4                           3                3                           2
## 5                           5                2                           1
##   opinion_seas_vacc_effective opinion_seas_risk opinion_seas_sick_from_vacc
## 0                           2                1                           2
## 1                           4                2                           4
## 3                           5                4                           1
## 4                           3                1                           4
## 5                           5                4                           4
##       age_group    education  race    sex            income_poverty
## 0 55 - 64 Years   < 12 Years White Female             Below Poverty
## 1 35 - 44 Years     12 Years White   Male             Below Poverty
## 3     65+ Years     12 Years White Female             Below Poverty
## 4 45 - 54 Years Some College White Female <= $75,000, Above Poverty
## 5     65+ Years     12 Years White   Male <= $75,000, Above Poverty
##   marital_status rent_or_own  employment_status               census_msa
## 0    Not Married         Own Not in Labor Force                   Non-MSA
## 1    Not Married        Rent           Employed MSA, Not Principle  City
## 3    Not Married        Rent Not in Labor Force      MSA, Principle City
## 4        Married         Own           Employed MSA, Not Principle  City
## 5        Married         Own           Employed      MSA, Principle City
##   household_adults household_children
## 0                0                 0
## 1                0                 0
## 3                0                 0
## 4                1                 0
## 5                2                 3
```

```r
#binarising the nominal attributes
binary_data <- dummyVars(~., data = h1n1_2)

#View(binary_data)
```

```r
#adding the conversion to the data
new_data <- predict(binary_data, newdata = h1n1_2)
```

TRAINING H1N1_VACCINE ALONE

```r
#adding "h1n1_vaccine class" back to original dataset
new_data2 <- cbind(new_data, h1n1[32])
```

```r
head(new_data2,5)
```

```
##   h1n1_concern h1n1_knowledge behavioral_antiviral_meds behavioral_avoidance
## 0            1              0                         0                    0
## 1            3              2                         0                    1
## 3            1              1                         0                    1
## 4            2              1                         0                    1
## 5            3              1                         0                    1
##   behavioral_face_mask behavioral_wash_hands behavioral_large_gatherings
## 0                    0                    0                           0
## 1                    0                    1                           0
## 3                    0                    1                           1
## 4                    0                    1                           1
## 5                    0                    1                           0
##   behavioral_outside_home behavioral_touch_face doctor_recc_h1n1
## 0                       1                     1                0
## 1                       1                     1                0
## 3                       0                     0                0
## 4                       0                     1                0
## 5                       0                     1                0
##   doctor_recc_seasonal chronic_med_condition child_under_6_months health_worker
## 0                    0                    0                     0             0
## 1                    0                    0                     0             0
## 3                    1                    1                     0             0
## 4                    0                    0                     0             0
## 5                    1                    0                     0             0
##   opinion_h1n1_vacc_effective opinion_h1n1_risk opinion_h1n1_sick_from_vacc
## 0                           3                 1                           2
## 1                           5                 4                           4
## 3                           3                 3                           5
## 4                           3                 3                           2
## 5                           5                 2                           1
##   opinion_seas_vacc_effective opinion_seas_risk opinion_seas_sick_from_vacc
## 0                           2                 1                           2
## 1                           4                 2                           4
## 3                           5                 4                           1
## 4                           3                 1                           4
## 5                           5                 4                           4
##   age_group.18 - 34 Years age_group.35 - 44 Years age_group.45 - 54 Years
## 0                       0                       0                       0
## 1                       0                       1                       0
## 3                       0                       0                       0
## 4                       0                       0                       1
## 5                       0                       0                       0
```

```
##   age_group.55 - 64 Years age_group.65+ Years education. education.< 12 Years
## 0                       1                   0         0                     1
## 1                       0                   0         0                     0
## 3                       0                   1         0                     0
## 4                       0                   0         0                     0
## 5                       0                   1         0                     0
##   education.12 Years education.College Graduate education.Some College
## 0                  0                         0                     0
## 1                  1                         0                     0
## 3                  1                         0                     0
## 4                  0                         0                     1
## 5                  1                         0                     0
##   race.Black race.Hispanic race.Other or Multiple race.White sex.Female
## 0          0             0                      0          1          1
## 1          0             0                      0          1          0
## 3          0             0                      0          1          1
## 4          0             0                      0          1          1
## 5          0             0                      0          1          0
##   sex.Male income_poverty. income_poverty.<= $75,000, Above Poverty
## 0        0               0                                       0
## 1        1               0                                       0
## 3        0               0                                       0
## 4        0               0                                       1
## 5        1               0                                       1
##   income_poverty.> $75,000 income_poverty.Below Poverty marital_status.
## 0                        0                           1              0
## 1                        0                           1              0
## 3                        0                           1              0
## 4                        0                           0              0
## 5                        0                           0              0
##   marital_status.Married marital_status.Not Married rent_or_own.
## 0                      0                         1           0
## 1                      0                         1           0
## 3                      0                         1           0
## 4                      1                         0           0
## 5                      1                         0           0
##   rent_or_own.Own rent_or_own.Rent employment_status.
## 0               1               0                  0
## 1               0               1                  0
## 3               0               1                  0
## 4               1               0                  0
## 5               1               0                  0
##   employment_status.Employed employment_status.Not in Labor Force
## 0                          0                                    1
## 1                          1                                    0
## 3                          0                                    1
## 4                          1                                    0
## 5                          1                                    0
##   employment_status.Unemployed census_msa.MSA, Not Principle  City
## 0                            0                                  0
## 1                            0                                  1
## 3                            0                                  0
## 4                            0                                  1
## 5                            0                                  0
```

```
##   census_msa.MSA, Principle City census_msa.Non-MSA household_adults
## 0                              0                  1                0
## 1                              0                  0                0
## 3                              1                  0                0
## 4                              0                  0                1
## 5                              1                  0                2
##   household_children h1n1_vaccine
## 0                  0            0
## 1                  0            0
## 3                  0            0
## 4                  0            0
## 5                  3            0
```

```
#converting 0 and 1 to "yes" and "no" for the decision class
new_data2$h1n1_vaccine <- factor(new_data2$h1n1_vaccine, levels = c(0, 1), labels = c("no", "yes"))
head(new_data2,5)
```

```
##   h1n1_concern h1n1_knowledge behavioral_antiviral_meds behavioral_avoidance
## 0            1              0                         0                    0
## 1            3              2                         0                    1
## 3            1              1                         0                    1
## 4            2              1                         0                    1
## 5            3              1                         0                    1
##   behavioral_face_mask behavioral_wash_hands behavioral_large_gatherings
## 0                    0                     0                           0
## 1                    0                     1                           0
## 3                    0                     1                           1
## 4                    0                     1                           1
## 5                    0                     1                           0
##   behavioral_outside_home behavioral_touch_face doctor_recc_h1n1
## 0                       1                     1                0
## 1                       1                     1                0
## 3                       0                     0                0
## 4                       0                     1                0
## 5                       0                     1                0
##   doctor_recc_seasonal chronic_med_condition child_under_6_months health_worker
## 0                    0                     0                     0             0
## 1                    0                     0                     0             0
## 3                    1                     1                     0             0
## 4                    0                     0                     0             0
## 5                    1                     0                     0             0
##   opinion_h1n1_vacc_effective opinion_h1n1_risk opinion_h1n1_sick_from_vacc
## 0                           3                 1                           2
## 1                           5                 4                           4
## 3                           3                 3                           5
## 4                           3                 3                           2
## 5                           5                 2                           1
##   opinion_seas_vacc_effective opinion_seas_risk opinion_seas_sick_from_vacc
## 0                           2                 1                           2
## 1                           4                 2                           4
## 3                           5                 4                           1
## 4                           3                 1                           4
## 5                           5                 4                           4
##   age_group.18 - 34 Years age_group.35 - 44 Years age_group.45 - 54 Years
## 0                       0                       0                       0
```

```
## 1                       0                       1                       0
## 3                       0                       0                       0
## 4                       0                       0                       1
## 5                       0                       0                       0
##   age_group.55 - 64 Years age_group.65+ Years education. education.< 12 Years
## 0                       1                   0         0                     1
## 1                       0                   0         0                     0
## 3                       0                   1         0                     0
## 4                       0                   0         0                     0
## 5                       0                   1         0                     0
##   education.12 Years education.College Graduate education.Some College
## 0                  0                         0                      0
## 1                  1                         0                      0
## 3                  1                         0                      0
## 4                  0                         0                      1
## 5                  1                         0                      0
##   race.Black race.Hispanic race.Other or Multiple race.White sex.Female
## 0          0             0                       0          1          1
## 1          0             0                       0          1          0
## 3          0             0                       0          1          1
## 4          0             0                       0          1          1
## 5          0             0                       0          1          0
##   sex.Male income_poverty. income_poverty.<= $75,000, Above Poverty
## 0        0               0                                       0
## 1        1               0                                       0
## 3        0               0                                       0
## 4        0               0                                       1
## 5        1               0                                       1
##   income_poverty.> $75,000 income_poverty.Below Poverty marital_status.
## 0                        0                           1               0
## 1                        0                           1               0
## 3                        0                           1               0
## 4                        0                           0               0
## 5                        0                           0               0
##   marital_status.Married marital_status.Not Married rent_or_own.
## 0                      0                         1            0
## 1                      0                         1            0
## 3                      0                         1            0
## 4                      1                         0            0
## 5                      1                         0            0
##   rent_or_own.Own rent_or_own.Rent employment_status.
## 0               1               0                  0
## 1               0               1                  0
## 3               0               1                  0
## 4               1               0                  0
## 5               1               0                  0
##   employment_status.Employed employment_status.Not in Labor Force
## 0                          0                                    1
## 1                          1                                    0
## 3                          0                                    1
## 4                          1                                    0
## 5                          1                                    0
##   employment_status.Unemployed census_msa.MSA, Not Principle  City
## 0                            0                                  0
```

```
## 1                             0                                   1
## 3                             0                                   0
## 4                             0                                   1
## 5                             0                                   0
##   census_msa.MSA, Principle City census_msa.Non-MSA household_adults
## 0                            0                  1                 0
## 1                            0                  0                 0
## 3                            1                  0                 0
## 4                            0                  0                 1
## 5                            1                  0                 2
##   household_children h1n1_vaccine
## 0                  0          no
## 1                  0          no
## 3                  0          no
## 4                  0          no
## 5                  3          no
```

DATA PREPROCESSING

Not needed since we are using random forest to train the model. And random forest is sensitive to feature scaling (but can perform normalization).

SPLIT THE TRAINING DATASET INTO TRAINING AND VALIDATION DATASET

```
#evaluation method (repeatedcv)
set.seed(42)

#splitting the data 80% training and 20% testing
trainIndex <- createDataPartition(new_data2$h1n1_vaccine, p = 0.8, list = FALSE)
trainData <- new_data2[trainIndex, ]
testData <- new_data2[-trainIndex, ]
```

TRAINING THE MODEL

```
#ensuring reproducibility
set.seed(123)

#applying training algorithms
lg_model <- glm(h1n1_vaccine~., data = trainData, family = binomial)
```

```
summary(lg_model)
```

```
##
## Call:
## glm(formula = h1n1_vaccine ~ ., family = binomial, data = trainData)
##
## Coefficients: (9 not defined because of singularities)
##                                          Estimate Std. Error z value
## (Intercept)                            -5.8597978  0.2081652 -28.150
## h1n1_concern                           -0.0851844  0.0292224  -2.915
```

```
## h1n1_knowledge                                0.1410445  0.0389592   3.620
## behavioral_antiviral_meds                     0.2311632  0.0947038   2.441
## behavioral_avoidance                         -0.0455093  0.0548163  -0.830
## behavioral_face_mask                          0.1438480  0.0817932   1.759
## behavioral_wash_hands                        -0.0432500  0.0678791  -0.637
## behavioral_large_gatherings                  -0.2176573  0.0561479  -3.877
## behavioral_outside_home                      -0.0232810  0.0568675  -0.409
## behavioral_touch_face                         0.0106934  0.0527749   0.203
## doctor_recc_h1n1                              1.9783163  0.0616410  32.094
## doctor_recc_seasonal                         -0.5506528  0.0609376  -9.036
## chronic_med_condition                         0.1381402  0.0481277   2.870
## child_under_6_months                          0.3006558  0.0721726   4.166
## health_worker                                 0.8407939  0.0626135  13.428
## opinion_h1n1_vacc_effective                   0.6659131  0.0300670  22.148
## opinion_h1n1_risk                             0.3853314  0.0201927  19.083
## opinion_h1n1_sick_from_vacc                  -0.0039885  0.0189733  -0.210
## opinion_seas_vacc_effective                   0.0850956  0.0269356   3.159
## opinion_seas_risk                             0.1596139  0.0195194   8.177
## opinion_seas_sick_from_vacc                  -0.0729428  0.0187576  -3.889
## `age_group.18 - 34 Years`                    -0.4622963  0.0826621  -5.593
## `age_group.35 - 44 Years`                    -0.5213569  0.0900479  -5.790
## `age_group.45 - 54 Years`                    -0.4218345  0.0763876  -5.522
## `age_group.55 - 64 Years`                    -0.0950131  0.0665287  -1.428
## `age_group.65+ Years`                               NA         NA      NA
## education.                                    0.3100258  0.2500841   1.240
## `education.< 12 Years`                       -0.2401206  0.0925211  -2.595
## `education.12 Years`                         -0.0641866  0.0635856  -1.009
## `education.College Graduate`                  0.1214238  0.0548541   2.214
## `education.Some College`                            NA         NA      NA
## race.Black                                   -0.3684633  0.0944458  -3.901
## race.Hispanic                                -0.1506636  0.0936963  -1.608
## `race.Other or Multiple`                      0.1830088  0.0913463   2.003
## race.White                                          NA         NA      NA
## sex.Female                                   -0.1542016  0.0458720  -3.362
## sex.Male                                            NA         NA      NA
## income_poverty.                               0.0972418  0.1030477   0.944
## `income_poverty.<= $75,000, Above Poverty`   -0.0374394  0.0829707  -0.451
## `income_poverty.> $75,000`                    0.0360077  0.0950474   0.379
## `income_poverty.Below Poverty`                      NA         NA      NA
## marital_status.                              -0.3028561  0.2749172  -1.102
## marital_status.Married                        0.1232264  0.0522205   2.360
## `marital_status.Not Married`                        NA         NA      NA
## rent_or_own.                                  0.1842787  0.1488243   1.238
## rent_or_own.Own                               0.0491395  0.0596367   0.824
## rent_or_own.Rent                                    NA         NA      NA
## employment_status.                           -0.0516821  0.2588633  -0.200
## employment_status.Employed                    0.0003749  0.1023691   0.004
## `employment_status.Not in Labor Force`        0.0370201  0.1048932   0.353
## employment_status.Unemployed                        NA         NA      NA
## `census_msa.MSA, Not Principle  City`        -0.0610905  0.0526794  -1.160
## `census_msa.MSA, Principle City`             -0.0087674  0.0581537  -0.151
## `census_msa.Non-MSA`                                NA         NA      NA
## household_adults                             -0.0029686  0.0329135  -0.090
## household_children                           -0.0246466  0.0285242  -0.864
```

```
##                                          Pr(>|z|)
## (Intercept)                               < 2e-16 ***
## h1n1_concern                             0.003556 **
## h1n1_knowledge                           0.000294 ***
## behavioral_antiviral_meds                0.014650 *
## behavioral_avoidance                     0.406417
## behavioral_face_mask                     0.078632 .
## behavioral_wash_hands                    0.524020
## behavioral_large_gatherings              0.000106 ***
## behavioral_outside_home                  0.682254
## behavioral_touch_face                    0.839430
## doctor_recc_h1n1                          < 2e-16 ***
## doctor_recc_seasonal                      < 2e-16 ***
## chronic_med_condition                    0.004101 **
## child_under_6_months                     3.10e-05 ***
## health_worker                             < 2e-16 ***
## opinion_h1n1_vacc_effective               < 2e-16 ***
## opinion_h1n1_risk                         < 2e-16 ***
## opinion_h1n1_sick_from_vacc              0.833499
## opinion_seas_vacc_effective              0.001582 **
## opinion_seas_risk                        2.91e-16 ***
## opinion_seas_sick_from_vacc              0.000101 ***
## `age_group.18 - 34 Years`                2.24e-08 ***
## `age_group.35 - 44 Years`                7.05e-09 ***
## `age_group.45 - 54 Years`                3.35e-08 ***
## `age_group.55 - 64 Years`                0.153248
## `age_group.65+ Years`                          NA
## education.                               0.215092
## `education.< 12 Years`                   0.009451 **
## `education.12 Years`                     0.312758
## `education.College Graduate`             0.026858 *
## `education.Some College`                       NA
## race.Black                               9.57e-05 ***
## race.Hispanic                            0.107835
## `race.Other or Multiple`                 0.045128 *
## race.White                                     NA
## sex.Female                               0.000775 ***
## sex.Male                                       NA
## income_poverty.                          0.345344
## `income_poverty.<= $75,000, Above Poverty` 0.651819
## `income_poverty.> $75,000`               0.704807
## `income_poverty.Below Poverty`                 NA
## marital_status.                          0.270624
## marital_status.Married                   0.018288 *
## `marital_status.Not Married`                   NA
## rent_or_own.                             0.215631
## rent_or_own.Own                          0.409951
## rent_or_own.Rent                               NA
## employment_status.                       0.841754
## employment_status.Employed               0.997078
## `employment_status.Not in Labor Force`   0.724140
## employment_status.Unemployed                   NA
## `census_msa.MSA, Not Principle  City`    0.246185
## `census_msa.MSA, Principle City`         0.880163
```

```
## `census_msa.Non-MSA`                              NA
## household_adults                             0.928132
## household_children                           0.387556
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 19630  on 18380  degrees of freedom
## Residual deviance: 14297  on 18334  degrees of freedom
## AIC: 14391
##
## Number of Fisher Scoring iterations: 5
```

VALIDATING THE MODEL

```
#Predict on test data
validations <- predict(lg_model, testData, type = "response")

#validations
```

```
# Convert probabilities to binary predictions
validated_labels <- ifelse(validations > 0.5, "yes", "no")

# Print results
#print(validated_labels)
```

```
# Ensure both predicted_labels and testData$h1n1_vaccine are factors with the same levels
validated_labels <- factor(validated_labels, levels = c("no", "yes"))
testData$h1n1_vaccine <- factor(testData$h1n1_vaccine, levels = c("no", "yes"))

# Now, apply confusionMatrix
confusionMatrix(validated_labels, testData$h1n1_vaccine)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   no  yes
##        no  3355  595
##        yes  203  442
##
##                Accuracy : 0.8263
##                  95% CI : (0.8151, 0.8372)
##     No Information Rate : 0.7743
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.4263
##
##  Mcnemar's Test P-Value : < 2.2e-16
##
##             Sensitivity : 0.9429
##             Specificity : 0.4262
##          Pos Pred Value : 0.8494
##          Neg Pred Value : 0.6853
```

```
##              Prevalence : 0.7743
##          Detection Rate : 0.7301
##    Detection Prevalence : 0.8596
##       Balanced Accuracy : 0.6846
##
##        'Positive' Class : no
##
```

The accuracy of the dataset is quite okay at 0.8263

TRAINING SEASONAL_VACCINE ALONE

```
#adding "seasonal_vaccine class" back to original dataset
new_data3 <- cbind(new_data, h1n1[33])
```

```
head(new_data3,5)
```

```
##   h1n1_concern h1n1_knowledge behavioral_antiviral_meds behavioral_avoidance
## 0            1              0                         0                    0
## 1            3              2                         0                    1
## 3            1              1                         0                    1
## 4            2              1                         0                    1
## 5            3              1                         0                    1
##   behavioral_face_mask behavioral_wash_hands behavioral_large_gatherings
## 0                    0                     0                           0
## 1                    0                     1                           0
## 3                    0                     1                           1
## 4                    0                     1                           1
## 5                    0                     1                           0
##   behavioral_outside_home behavioral_touch_face doctor_recc_h1n1
## 0                       1                     1                0
## 1                       1                     1                0
## 3                       0                     0                0
## 4                       0                     1                0
## 5                       0                     1                0
##   doctor_recc_seasonal chronic_med_condition child_under_6_months health_worker
## 0                    0                     0                     0             0
## 1                    0                     0                     0             0
## 3                    1                     1                     0             0
## 4                    0                     0                     0             0
## 5                    1                     0                     0             0
##   opinion_h1n1_vacc_effective opinion_h1n1_risk opinion_h1n1_sick_from_vacc
## 0                           3                 1                           2
## 1                           5                 4                           4
## 3                           3                 3                           5
## 4                           3                 3                           2
## 5                           5                 2                           1
##   opinion_seas_vacc_effective opinion_seas_risk opinion_seas_sick_from_vacc
## 0                           2                 1                           2
## 1                           4                 2                           4
## 3                           5                 4                           1
## 4                           3                 1                           4
## 5                           5                 4                           4
##   age_group.18 - 34 Years age_group.35 - 44 Years age_group.45 - 54 Years
```

```
## 0                       0                       0                       0
## 1                       0                       1                       0
## 3                       0                       0                       0
## 4                       0                       0                       1
## 5                       0                       0                       0
##   age_group.55 - 64 Years age_group.65+ Years education. education.< 12 Years
## 0                       1                   0         0                     1
## 1                       0                   0         0                     0
## 3                       0                   1         0                     0
## 4                       0                   0         0                     0
## 5                       0                   1         0                     0
##   education.12 Years education.College Graduate education.Some College
## 0                  0                         0                      0
## 1                  1                         0                      0
## 3                  1                         0                      0
## 4                  0                         0                      1
## 5                  1                         0                      0
##   race.Black race.Hispanic race.Other or Multiple race.White sex.Female
## 0          0             0                       0          1          1
## 1          0             0                       0          1          0
## 3          0             0                       0          1          1
## 4          0             0                       0          1          1
## 5          0             0                       0          1          0
##   sex.Male income_poverty. income_poverty.<= $75,000, Above Poverty
## 0        0               0                                       0
## 1        1               0                                       0
## 3        0               0                                       0
## 4        0               0                                       1
## 5        1               0                                       1
##   income_poverty.> $75,000 income_poverty.Below Poverty marital_status.
## 0                        0                           1               0
## 1                        0                           1               0
## 3                        0                           1               0
## 4                        0                           0               0
## 5                        0                           0               0
##   marital_status.Married marital_status.Not Married rent_or_own.
## 0                      0                         1            0
## 1                      0                         1            0
## 3                      0                         1            0
## 4                      1                         0            0
## 5                      1                         0            0
##   rent_or_own.Own rent_or_own.Rent employment_status.
## 0               1               0                  0
## 1               0               1                  0
## 3               0               1                  0
## 4               1               0                  0
## 5               1               0                  0
##   employment_status.Employed employment_status.Not in Labor Force
## 0                          0                                    1
## 1                          1                                    0
## 3                          0                                    1
## 4                          1                                    0
## 5                          1                                    0
##   employment_status.Unemployed census_msa.MSA, Not Principle  City
```

```
## 0                            0                                    0
## 1                            0                                    1
## 3                            0                                    0
## 4                            0                                    1
## 5                            0                                    0
##   census_msa.MSA, Principle City census_msa.Non-MSA household_adults
## 0                            0                   1                 0
## 1                            0                   0                 0
## 3                            1                   0                 0
## 4                            0                   0                 1
## 5                            1                   0                 2
##   household_children seasonal_vaccine
## 0                  0                0
## 1                  0                1
## 3                  0                1
## 4                  0                0
## 5                  3                0
```

```
#converting 0 and 1 to "yes" and "no" for the decision class
new_data3$seasonal_vaccine <- factor(new_data3$seasonal_vaccine, levels = c(0, 1), labels = c("no", "yes
head(new_data3,5)
```

```
##   h1n1_concern h1n1_knowledge behavioral_antiviral_meds behavioral_avoidance
## 0            1              0                         0                    0
## 1            3              2                         0                    1
## 3            1              1                         0                    1
## 4            2              1                         0                    1
## 5            3              1                         0                    1
##   behavioral_face_mask behavioral_wash_hands behavioral_large_gatherings
## 0                    0                     0                           0
## 1                    0                     1                           0
## 3                    0                     1                           1
## 4                    0                     1                           1
## 5                    0                     1                           0
##   behavioral_outside_home behavioral_touch_face doctor_recc_h1n1
## 0                       1                     1                0
## 1                       1                     1                0
## 3                       0                     0                0
## 4                       0                     1                0
## 5                       0                     1                0
##   doctor_recc_seasonal chronic_med_condition child_under_6_months health_worker
## 0                    0                     0                    0             0
## 1                    0                     0                    0             0
## 3                    1                     1                    0             0
## 4                    0                     0                    0             0
## 5                    1                     0                    0             0
##   opinion_h1n1_vacc_effective opinion_h1n1_risk opinion_h1n1_sick_from_vacc
## 0                           3                1                           2
## 1                           5                4                           4
## 3                           3                3                           5
## 4                           3                3                           2
## 5                           5                2                           1
##   opinion_seas_vacc_effective opinion_seas_risk opinion_seas_sick_from_vacc
## 0                           2                1                           2
## 1                           4                2                           4
```

```
## 3                          5                4                          1
## 4                          3                1                          4
## 5                          5                4                          4
##   age_group.18 - 34 Years age_group.35 - 44 Years age_group.45 - 54 Years
## 0                       0                       0                       0
## 1                       0                       1                       0
## 3                       0                       0                       0
## 4                       0                       0                       1
## 5                       0                       0                       0
##   age_group.55 - 64 Years age_group.65+ Years education. education.< 12 Years
## 0                       1                   0          0                     1
## 1                       0                   0          0                     0
## 3                       0                   1          0                     0
## 4                       0                   0          0                     0
## 5                       0                   1          0                     0
##   education.12 Years education.College Graduate education.Some College
## 0                  0                          0                      0
## 1                  1                          0                      0
## 3                  1                          0                      0
## 4                  0                          0                      1
## 5                  1                          0                      0
##   race.Black race.Hispanic race.Other or Multiple race.White sex.Female
## 0          0             0                       0          1          1
## 1          0             0                       0          1          0
## 3          0             0                       0          1          1
## 4          0             0                       0          1          1
## 5          0             0                       0          1          0
##   sex.Male income_poverty. income_poverty.<= $75,000, Above Poverty
## 0        0               0                                       0
## 1        1               0                                       0
## 3        0               0                                       0
## 4        0               0                                       1
## 5        1               0                                       1
##   income_poverty.> $75,000 income_poverty.Below Poverty marital_status.
## 0                        0                           1               0
## 1                        0                           1               0
## 3                        0                           1               0
## 4                        0                           0               0
## 5                        0                           0               0
##   marital_status.Married marital_status.Not Married rent_or_own.
## 0                      0                         1            0
## 1                      0                         1            0
## 3                      0                         1            0
## 4                      1                         0            0
## 5                      1                         0            0
##   rent_or_own.Own rent_or_own.Rent employment_status.
## 0               1               0                  0
## 1               0               1                  0
## 3               0               1                  0
## 4               1               0                  0
## 5               1               0                  0
##   employment_status.Employed employment_status.Not in Labor Force
## 0                          0                                    1
## 1                          1                                    0
```

```
## 3                           0                                     1
## 4                           1                                     0
## 5                           1                                     0
##   employment_status.Unemployed census_msa.MSA, Not Principle  City
## 0                            0                                    0
## 1                            0                                    1
## 3                            0                                    0
## 4                            0                                    1
## 5                            0                                    0
##   census_msa.MSA, Principle City census_msa.Non-MSA household_adults
## 0                            0                   1                 0
## 1                            0                   0                 0
## 3                            1                   0                 0
## 4                            0                   0                 1
## 5                            1                   0                 2
##   household_children seasonal_vaccine
## 0                  0               no
## 1                  0              yes
## 3                  0              yes
## 4                  0               no
## 5                  3               no
```

DATA PREPROCESSING

Not needed since we are using random forest to train the model. And random forest is sensitive to feature scaling (but can perform normalization).

SPLIT THE TRAINING DATASET INTO TRAINING AND VALIDATION DATASET

```
#evaluation method (repeatedcv)
set.seed(42)

#splitting the data 80% training and 20% testing
trainIndex <- createDataPartition(new_data3$seasonal_vaccine, p = 0.8, list = FALSE)
trainData2 <- new_data3[trainIndex, ]
testData2 <- new_data3[-trainIndex, ]
```

TRAINING THE MODEL

```
#ensuring reproducibility
set.seed(123)

#applying training algorithms
seasonal_lg_model <- glm(seasonal_vaccine~., data = trainData2, family = binomial)
```

```
summary(seasonal_lg_model)
```

```
##
## Call:
## glm(formula = seasonal_vaccine ~ ., family = binomial, data = trainData2)
```

```
##
## Coefficients: (9 not defined because of singularities)
##                                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)                              -4.32832    0.17198 -25.168  < 2e-16
## h1n1_concern                              0.00528    0.02587   0.204 0.838280
## h1n1_knowledge                            0.19988    0.03430   5.827 5.65e-09
## behavioral_antiviral_meds                 0.11616    0.08863   1.311 0.189958
## behavioral_avoidance                      0.01053    0.04792   0.220 0.826069
## behavioral_face_mask                      0.04915    0.07918   0.621 0.534799
## behavioral_wash_hands                     0.02988    0.05763   0.518 0.604140
## behavioral_large_gatherings              -0.02948    0.05028  -0.586 0.557603
## behavioral_outside_home                  -0.02556    0.05116  -0.500 0.617346
## behavioral_touch_face                     0.16016    0.04603   3.480 0.000502
## doctor_recc_h1n1                         -0.28849    0.05971  -4.832 1.35e-06
## doctor_recc_seasonal                      1.45381    0.05235  27.771  < 2e-16
## chronic_med_condition                     0.21470    0.04436   4.840 1.30e-06
## child_under_6_months                      0.07865    0.07004   1.123 0.261489
## health_worker                             0.86094    0.06328  13.605  < 2e-16
## opinion_h1n1_vacc_effective               0.02632    0.02269   1.160 0.245940
## opinion_h1n1_risk                         0.02798    0.01922   1.456 0.145517
## opinion_h1n1_sick_from_vacc              -0.04299    0.01746  -2.462 0.013822
## opinion_seas_vacc_effective               0.57178    0.02337  24.467  < 2e-16
## opinion_seas_risk                         0.57388    0.01769  32.447  < 2e-16
## opinion_seas_sick_from_vacc              -0.19243    0.01710 -11.255  < 2e-16
## `age_group.18 - 34 Years`                -1.53654    0.07359 -20.879  < 2e-16
## `age_group.35 - 44 Years`                -1.30413    0.07890 -16.529  < 2e-16
## `age_group.45 - 54 Years`                -1.13147    0.06753 -16.756  < 2e-16
## `age_group.55 - 64 Years`                -0.82698    0.06068 -13.628  < 2e-16
## `age_group.65+ Years`                         NA         NA      NA       NA
## education.                                0.05198    0.21318   0.244 0.807350
## `education.< 12 Years`                   -0.28214    0.07869  -3.585 0.000336
## `education.12 Years`                     -0.06635    0.05497  -1.207 0.227371
## `education.College Graduate`              0.12516    0.04889   2.560 0.010463
## `education.Some College`                      NA         NA      NA       NA
## race.Black                               -0.28421    0.07615  -3.732 0.000190
## race.Hispanic                            -0.21313    0.08124  -2.623 0.008707
## `race.Other or Multiple`                  0.10794    0.08137   1.327 0.184638
## race.White                                    NA         NA      NA       NA
## sex.Female                               -0.03152    0.04034  -0.782 0.434471
## sex.Male                                      NA         NA      NA       NA
## income_poverty.                           0.27027    0.08945   3.021 0.002516
## `income_poverty.<= $75,000, Above Poverty` 0.15241   0.07281   2.093 0.036327
## `income_poverty.> $75,000`                0.32520    0.08374   3.883 0.000103
## `income_poverty.Below Poverty`                NA         NA      NA       NA
## marital_status.                          -0.02002    0.22546  -0.089 0.929229
## marital_status.Married                    0.07785    0.04597   1.694 0.090351
## `marital_status.Not Married`                  NA         NA      NA       NA
## rent_or_own.                              0.26795    0.12756   2.101 0.035671
## rent_or_own.Own                           0.17865    0.05241   3.409 0.000653
## rent_or_own.Rent                              NA         NA      NA       NA
## employment_status.                        0.25757    0.21778   1.183 0.236918
## employment_status.Employed                0.16958    0.08901   1.905 0.056750
## `employment_status.Not in Labor Force`    0.26723    0.09164   2.916 0.003544
## employment_status.Unemployed                  NA         NA      NA       NA
```

```
## `census_msa.MSA, Not Principle  City`       0.16000    0.04680   3.419 0.000628
## `census_msa.MSA, Principle City`            0.14792    0.05216   2.836 0.004569
## `census_msa.Non-MSA`                             NA         NA      NA       NA
## household_adults                           -0.04367    0.02867  -1.523 0.127717
## household_children                         -0.05780    0.02501  -2.311 0.020838
##
## (Intercept)                                ***
## h1n1_concern
## h1n1_knowledge                             ***
## behavioral_antiviral_meds
## behavioral_avoidance
## behavioral_face_mask
## behavioral_wash_hands
## behavioral_large_gatherings
## behavioral_outside_home
## behavioral_touch_face                      ***
## doctor_recc_h1n1                           ***
## doctor_recc_seasonal                       ***
## chronic_med_condition                      ***
## child_under_6_months
## health_worker                              ***
## opinion_h1n1_vacc_effective
## opinion_h1n1_risk
## opinion_h1n1_sick_from_vacc                *
## opinion_seas_vacc_effective                ***
## opinion_seas_risk                          ***
## opinion_seas_sick_from_vacc                ***
## `age_group.18 - 34 Years`                  ***
## `age_group.35 - 44 Years`                  ***
## `age_group.45 - 54 Years`                  ***
## `age_group.55 - 64 Years`                  ***
## `age_group.65+ Years`
## education.
## `education.< 12 Years`                     ***
## `education.12 Years`
## `education.College Graduate`               *
## `education.Some College`
## race.Black                                 ***
## race.Hispanic                              **
## `race.Other or Multiple`
## race.White
## sex.Female
## sex.Male
## income_poverty.                            **
## `income_poverty.<= $75,000, Above Poverty` *
## `income_poverty.> $75,000`                 ***
## `income_poverty.Below Poverty`
## marital_status.
## marital_status.Married                     .
## `marital_status.Not Married`
## rent_or_own.                               *
## rent_or_own.Own                            ***
## rent_or_own.Rent
## employment_status.
```

```
## employment_status.Employed                  .
## `employment_status.Not in Labor Force`     **
## employment_status.Unemployed
## `census_msa.MSA, Not Principle  City`      ***
## `census_msa.MSA, Principle City`           **
## `census_msa.Non-MSA`
## household_adults
## household_children                         *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 25446  on 18380  degrees of freedom
## Residual deviance: 17366  on 18334  degrees of freedom
## AIC: 17460
##
## Number of Fisher Scoring iterations: 5
```

VALIDATING THE MODEL

```
#Predict on test data
validations2 <- predict(seasonal_lg_model, testData2, type = "response")

#validations2
```

```
# Convert probabilities to binary predictions
validated_labels2 <- ifelse(validations2 > 0.5, "yes", "no")

# Print results
#print(validated_labels2)
```

```
# Ensure both predicted_labels and testData$h1n1_vaccine are factors with the same levels
validated_labels2 <- factor(validated_labels2, levels = c("no", "yes"))
testData2$seasonal_vaccine <- factor(testData2$seasonal_vaccine, levels = c("no", "yes"))

# Now, apply confusionMatrix
confusionMatrix(validated_labels2, testData2$seasonal_vaccine)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   no  yes
##        no  1916  519
##        yes  482 1678
##
##                Accuracy : 0.7822
##                  95% CI : (0.7699, 0.794)
##     No Information Rate : 0.5219
##     P-Value [Acc > NIR] : <2e-16
##
##                   Kappa : 0.5632
##
##  Mcnemar's Test P-Value : 0.2552
```

```
##
##             Sensitivity : 0.7990
##             Specificity : 0.7638
##          Pos Pred Value : 0.7869
##          Neg Pred Value : 0.7769
##              Prevalence : 0.5219
##          Detection Rate : 0.4170
##    Detection Prevalence : 0.5299
##       Balanced Accuracy : 0.7814
##
##        'Positive' Class : no
##
```

The accuracy of the dataset is quite okay at 0.7822