Predicting Recidivism

Rakesh Annamaneni

2025-01-22

Problem Statement

Determine factors that are most significant in predicting recidivism three years after being released from prison?

```r
#importing required packages
library(randomForest)
library(tidyverse)
library(readr)
library(glmnet)
library(caret)
library(ROCR)
library(pROC)
library(dplyr)

#Displaying the summary statistics and structure of the dataset
summary(recidivism_df)
```

```
##      Race           Age_at_Release     Gang_Affiliated
Supervision_Level_First
##  Length:14170       Length:14170       Mode :logical   Length:14170
##  Class :character   Class :character   FALSE:11543     Class :character
##  Mode  :character   Mode  :character   TRUE :2627      Mode  :character
##  Education_Level     Dependents         Prison_Offense     Prison_Years
##  Length:14170       Length:14170       Length:14170       Length:14170
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##  Prior_Arrest_Episodes_Felony Prior_Arrest_Episodes_Misd
##  Length:14170                 Length:14170
##  Class :character             Class :character
##  Mode  :character             Mode  :character
##  Prior_Arrest_Episodes_Drug Prior_Conviction_Episodes_Felony
##  Length:14170               Length:14170
##  Class :character           Class :character
##  Mode  :character           Mode  :character
##  Prior_Conviction_Episodes_Misd Prior_Conviction_Episodes_Drug
##  Length:14170                   Length:14170
##  Class :character               Class :character
##  Mode  :character               Mode  :character
##  Residence_Changes  Recidivism_Within_3years Drug_Test_Positive
##  Length:14170       Mode :logical            Mode :logical
##  Class :character   FALSE:5747               FALSE:13183
##  Mode  :character   TRUE :8423               TRUE :987
```

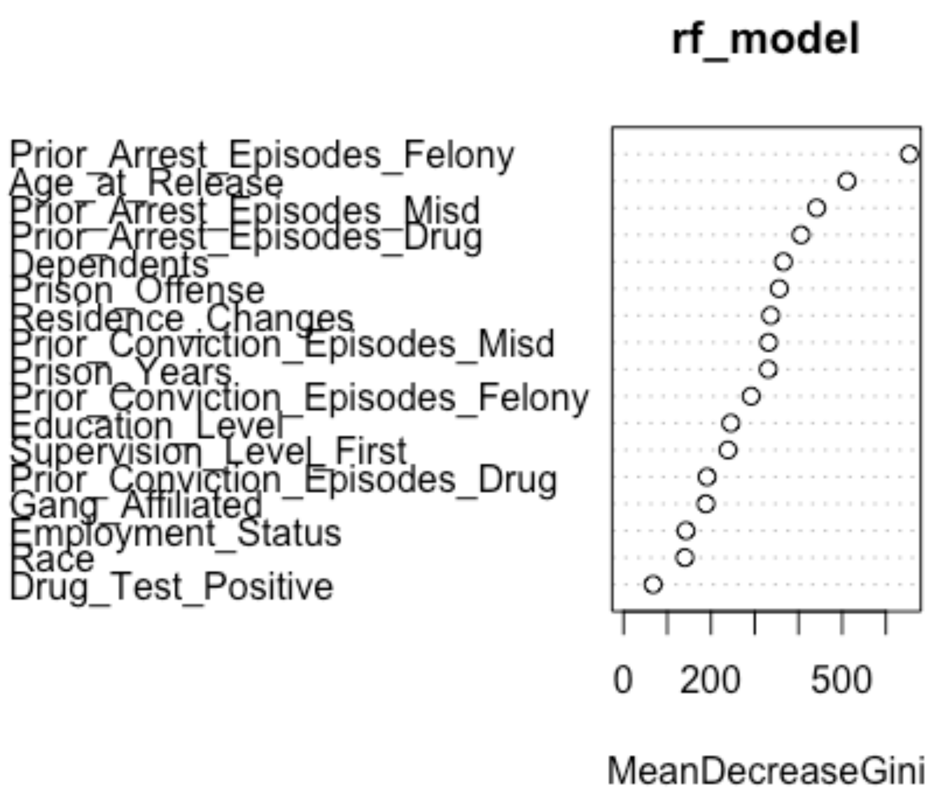
```
##  Employment_Status
##  Length:14170
##  Class :character
##  Mode  :character
```

```r
# Convert all character variables to factors
recidivism_df <- recidivism_df %>%
  mutate(across(where(is.character), as.factor))
recidivism_df$Recidivism_Within_3years <-
as.factor(recidivism_df$Recidivism_Within_3years)

str(recidivism_df)
```

```
## 'data.frame':    14170 obs. of  18 variables:
##  $ Race                           : Factor w/ 2 levels "BLACK","WHITE": 1
1 1 2 2 2 1 1 2 1 ...
##  $ Age_at_Release                 : Factor w/ 7 levels "18-22","23-
27",..: 6 4 7 5 4 5 5 6 6 4 ...
##  $ Gang_Affiliated                : logi  FALSE FALSE FALSE FALSE FALSE
FALSE ...
##  $ Supervision_Level_First        : Factor w/ 3 levels
"High","Specialized",..: 3 2 1 1 2 3 1 3 2 2 ...
##  $ Education_Level                : Factor w/ 3 levels "At least some
college",..: 1 3 1 3 3 2 2 2 3 2 ...
##  $ Dependents                     : Factor w/ 4 levels "0","1","2","3 or
more": 4 2 4 2 4 1 4 4 2 4 ...
##  $ Prison_Offense                 : Factor w/ 5 levels "Drug","Other",..:
1 4 1 3 4 3 1 3 4 2 ...
##  $ Prison_Years                   : Factor w/ 4 levels "1-2
years","Greater than 2 to 3 years",..: 4 4 1 1 1 4 2 4 1 1 ...
##  $ Prior_Arrest_Episodes_Felony    : Factor w/ 11 levels "0","1","10 or
more",..: 8 9 8 10 6 6 8 3 5 3 ...
##  $ Prior_Arrest_Episodes_Misd     : Factor w/ 7 levels
"0","1","2","3",..: 7 7 7 7 5 1 7 7 7 2 ...
##  $ Prior_Arrest_Episodes_Drug     : Factor w/ 6 levels
"0","1","2","3",..: 4 4 3 4 2 1 3 3 2 3 ...
##  $ Prior_Conviction_Episodes_Felony: Factor w/ 4 levels "0","1","2","3 or
more": 4 4 4 4 2 2 2 4 1 4 ...
##  $ Prior_Conviction_Episodes_Misd  : Factor w/ 5 levels
"0","1","2","3",..: 4 5 3 5 1 1 5 5 4 2 ...
##  $ Prior_Conviction_Episodes_Drug  : Factor w/ 3 levels "0","1","2 or
more": 3 3 3 3 2 1 3 3 1 3 ...
##  $ Residence_Changes              : Factor w/ 4 levels "0","1","2","3 or
more": 3 3 1 4 1 4 1 1 3 1 ...
##  $ Recidivism_Within_3years       : Factor w/ 2 levels "FALSE","TRUE": 1
2 2 1 2 1 1 2 2 2 ...
##  $ Drug_Test_Positive             : logi  FALSE FALSE FALSE FALSE FALSE
FALSE ...
##  $ Employment_Status              : Factor w/ 2 levels
"Employed","Unemployed": 1 1 2 1 1 1 1 1 1 2 ...
```

Using Random Forest method to know the important variables and using top 10 varibles for building logistic Regression Model

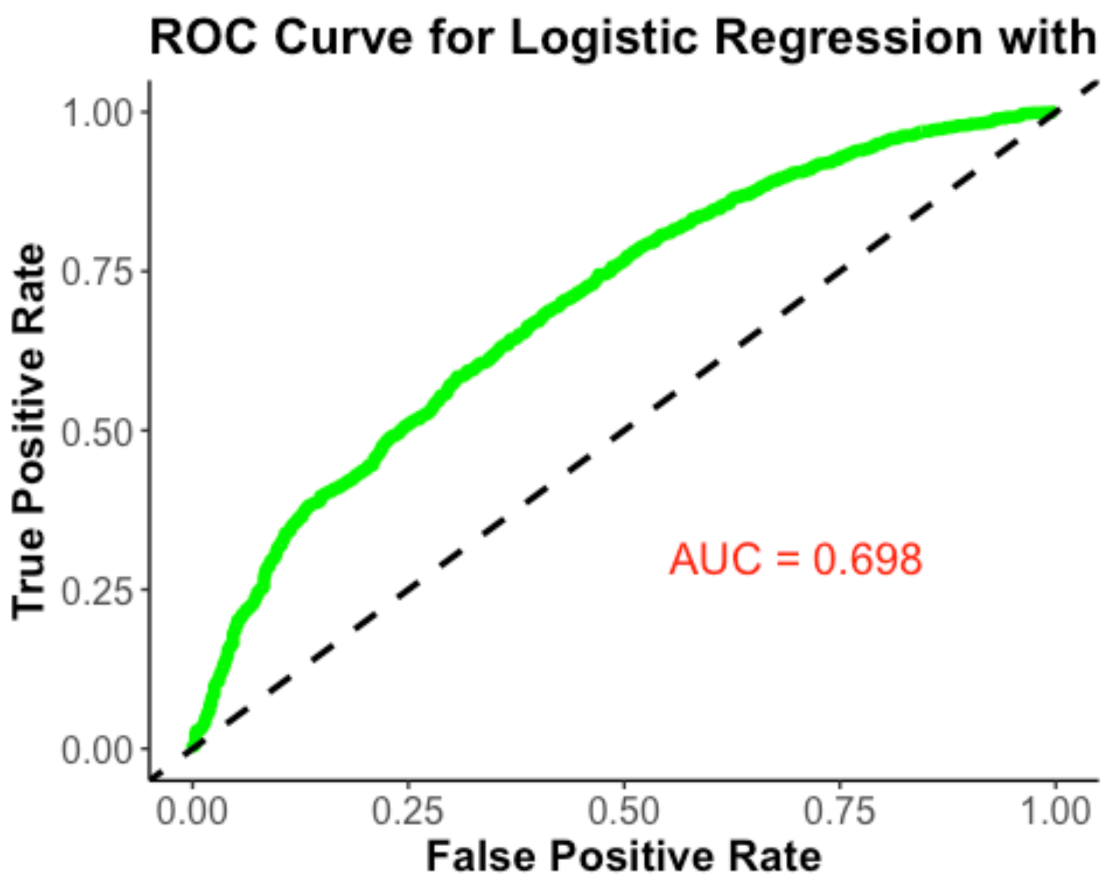
```
set.seed(1729)
trainIndex <- createDataPartition(recidivism_df$Recidivism_Within_3years, p =
0.8, list = FALSE)
training_set<- recidivism_df[trainIndex, ]
testing_set <- recidivism_df[-trainIndex, ]
set.seed(1729)
rf_model <- randomForest(Recidivism_Within_3years ~ ., data = training_set)
varImpPlot(rf_model)
```




```
important_vars <- row.names(varImp(rf_model)) %>% head(10)
log_reg_model_rf_selected <- glm(Recidivism_Within_3years ~ ., data =
training_set[, c("Recidivism_Within_3years", important_vars)], family =
binomial(link = "logit"))
glm_predicted_rf_selected <- predict(log_reg_model_rf_selected, testing_set[,
important_vars], type = "response")
glm_predicted_rf_selected_b <- rep("FALSE", nrow(testing_set))
glm_predicted_rf_selected_b[glm_predicted_rf_selected >= .5] <- "TRUE"
confusionMatrix(as.factor(glm_predicted_rf_selected_b),
testing_set$Recidivism_Within_3years)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction FALSE TRUE
##      FALSE   490  296
##      TRUE    659 1388
##
##                Accuracy : 0.6629
##                  95% CI : (0.6452, 0.6803)
##     No Information Rate : 0.5944
##     P-Value [Acc > NIR] : 3.634e-14
##
##                   Kappa : 0.2639
##
##  Mcnemar's Test P-Value : < 2.2e-16
##
##             Sensitivity : 0.4265
##             Specificity : 0.8242
##          Pos Pred Value : 0.6234
##          Neg Pred Value : 0.6781
##              Prevalence : 0.4056
##          Detection Rate : 0.1730
##    Detection Prevalence : 0.2774
##       Balanced Accuracy : 0.6253
##
##        'Positive' Class : FALSE
##

# Building ROC Curve
library(ROCR)
pred_rf <- prediction(glm_predicted_rf_selected,
testing_set$Recidivism_Within_3years)
roc_rf <- performance(pred_rf, measure = "tpr", x.measure = "fpr")
roc_obj <- roc(testing_set$Recidivism_Within_3years,
glm_predicted_rf_selected)

## Setting levels: control = FALSE, case = TRUE

## Setting direction: controls < cases

auc_rf <- auc(roc_obj)
roc_df <- data.frame(fpr = roc_rf@x.values[[1]], tpr = roc_rf@y.values[[1]])
g1<-ggplot(roc_df, aes(x = fpr, y = tpr)) +
  geom_line(color = "green", size = 2) +
  geom_abline(intercept = 0, slope = 1, linetype = "dashed", size = 1) +
  labs(title = "ROC Curve for Logistic Regression with Random Forest", x =
"False Positive Rate", y = "True Positive Rate") +
  theme_classic() +
  theme(plot.title = element_text(face = "bold", size = 16),
        axis.title = element_text(face = "bold", size = 14),
        axis.text = element_text(size = 12),
```

```
        legend.position = "bottom",
        legend.title = element_blank(),
        legend.text = element_text(face = "bold", size = 12),
        panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        panel.background = element_blank())+
  annotate("text", x = 0.7, y = 0.3, label = paste0("AUC = ", round(auc_rf,
3)), size = 5, color = "red")

## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## ℹ Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.

g1
```



Using Chi-Square analysis for varibale selection and buiding a Logistic Regression model

```
library(tidyr)
library(dplyr)
chi_square_results <- recidivism_df %>%
  gather(key = "Variable", value = "Value", -Recidivism_Within_3years) %>%
```

```
  group_by(Variable) %>%
  summarize(Chi_Square = chisq.test(table(Value,
Recidivism_Within_3years))$statistic,
            P_Value = chisq.test(table(Value,
Recidivism_Within_3years))$p.value) %>%
  filter(P_Value < 0.001)

## Warning: attributes are not identical across measure variables; they will
be
## dropped

significant_vars <- chi_square_results$Variable
significant_vars

##  [1] "Age_at_Release"                   "Dependents"
##  [3] "Education_Level"                  "Employment_Status"
##  [5] "Gang_Affiliated"                  "Prior_Arrest_Episodes_Drug"
##  [7] "Prior_Arrest_Episodes_Felony"     "Prior_Arrest_Episodes_Misd"
##  [9] "Prior_Conviction_Episodes_Drug"   "Prior_Conviction_Episodes_Felony"
## [11] "Prior_Conviction_Episodes_Misd"   "Prison_Offense"
## [13] "Prison_Years"                     "Residence_Changes"
## [15] "Supervision_Level_First"

log_reg_model_chi_square <- glm(Recidivism_Within_3years ~ ., data =
recidivism_df[, c("Recidivism_Within_3years", significant_vars)], family =
binomial(link = "logit"))
glm_predicted_chi_square <- predict(log_reg_model_chi_square, testing_set[,
significant_vars], type = "response")
glm_predicted_chi_square_b <- ifelse(glm_predicted_chi_square >= .5, "TRUE",
"FALSE")
confusionMatrix(as.factor(glm_predicted_chi_square_b),
testing_set$Recidivism_Within_3years)

## Confusion Matrix and Statistics
##
##           Reference
## Prediction FALSE TRUE
##      FALSE   525  300
##      TRUE    624 1384
##
##                Accuracy : 0.6738
##                  95% CI : (0.6562, 0.6911)
##     No Information Rate : 0.5944
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.2918
##
##  Mcnemar's Test P-Value : < 2.2e-16
##
##             Sensitivity : 0.4569
##             Specificity : 0.8219
```

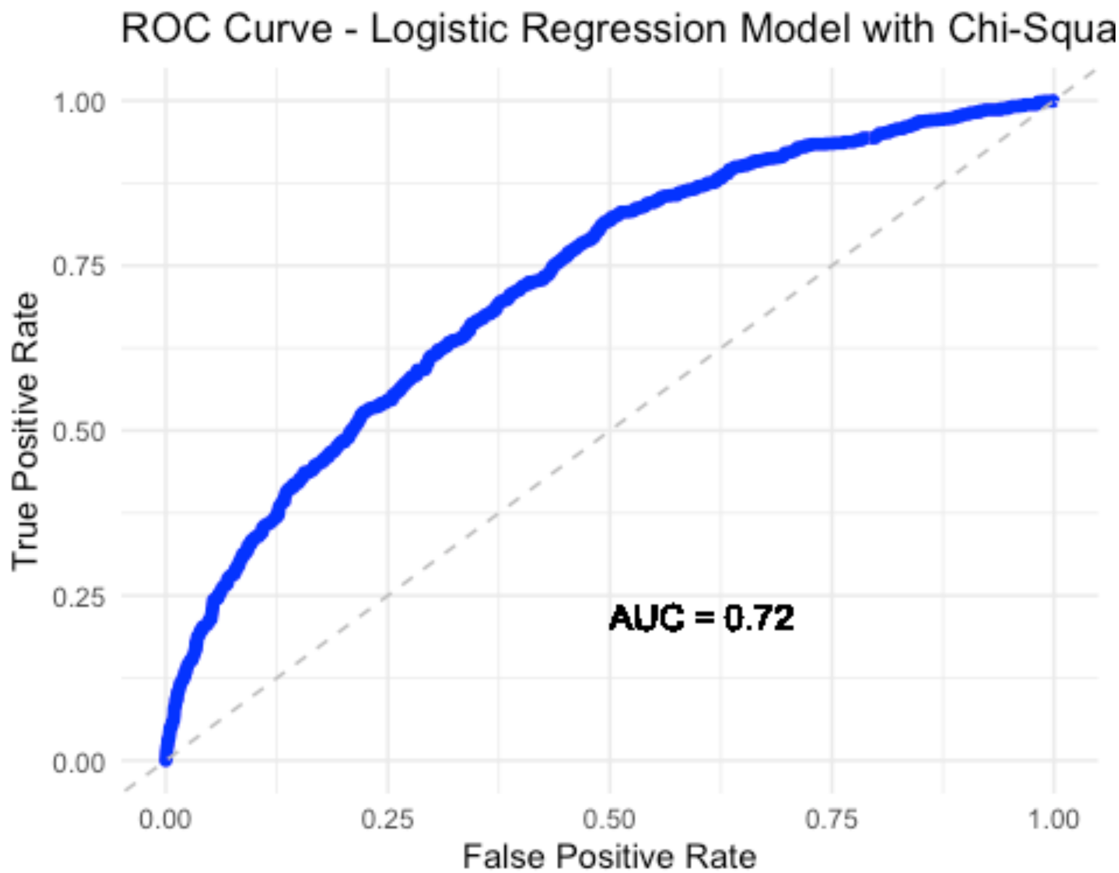
```
##          Pos Pred Value : 0.6364
##          Neg Pred Value : 0.6892
##              Prevalence : 0.4056
##          Detection Rate : 0.1853
##    Detection Prevalence : 0.2912
##       Balanced Accuracy : 0.6394
##
##        'Positive' Class : FALSE
##
```

```
library(ggplot2)
library(pROC)
roc_curve_chi_square <- roc(testing_set$Recidivism_Within_3years,
glm_predicted_chi_square, levels =
rev(levels(as.factor(testing_set$Recidivism_Within_3years))))
```

```
## Setting direction: controls > cases
```

```
auc_chi_square <- auc(roc_curve_chi_square)
pred_chi <- prediction(glm_predicted_chi_square,
testing_set$Recidivism_Within_3years)
roc_chi <- performance(pred_chi, measure = "tpr", x.measure = "fpr")
roc_data <- data.frame(
  FPR = 1 - roc_curve_chi_square$specificities,
  TPR = roc_curve_chi_square$sensitivities
)
g2<-ggplot(data = roc_data) +
  geom_line(aes(x = FPR, y = TPR), color = "blue", size = 2) +
  geom_abline(intercept = 0, slope = 1, linetype = "dashed", color = "gray")
+
  geom_text(aes(label = paste("AUC =", round(auc_chi_square, 2))),
            x = 0.5, y = 0.2, vjust = 0, hjust = 0,
            color = "black", size = 4) +
  labs(title = "ROC Curve - Logistic Regression Model with Chi-Square
Selection",
       x = "False Positive Rate",
       y = "True Positive Rate") +
  theme_minimal()
g2
```

```
## Warning in geom_text(aes(label = paste("AUC =", round(auc_chi_square,
2))), : All aesthetics have length 1, but the data has 2829 rows.
## ℹ Please consider using `annotate()` or provide this layer with data
containing
##   a single row.
```



Using Stepwise selection for varibale selection and buiding a Logistic Regression model

```
glm.predicted_stepwise <- predict(stepwise_model, testing_set,
type="response")
glm.predicted_s = rep("FALSE", nrow(testing_set))
glm.predicted_s[glm.predicted_stepwise >= 0.5] = "TRUE"
confusionMatrix(as.factor(glm.predicted_s),
testing_set$Recidivism_Within_3years)

## Confusion Matrix and Statistics
##
##             Reference
## Prediction FALSE TRUE
##      FALSE   526  312
##      TRUE    623 1372
##
##                Accuracy : 0.67
##                  95% CI : (0.6523, 0.6873)
##     No Information Rate : 0.5944
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.2848
```

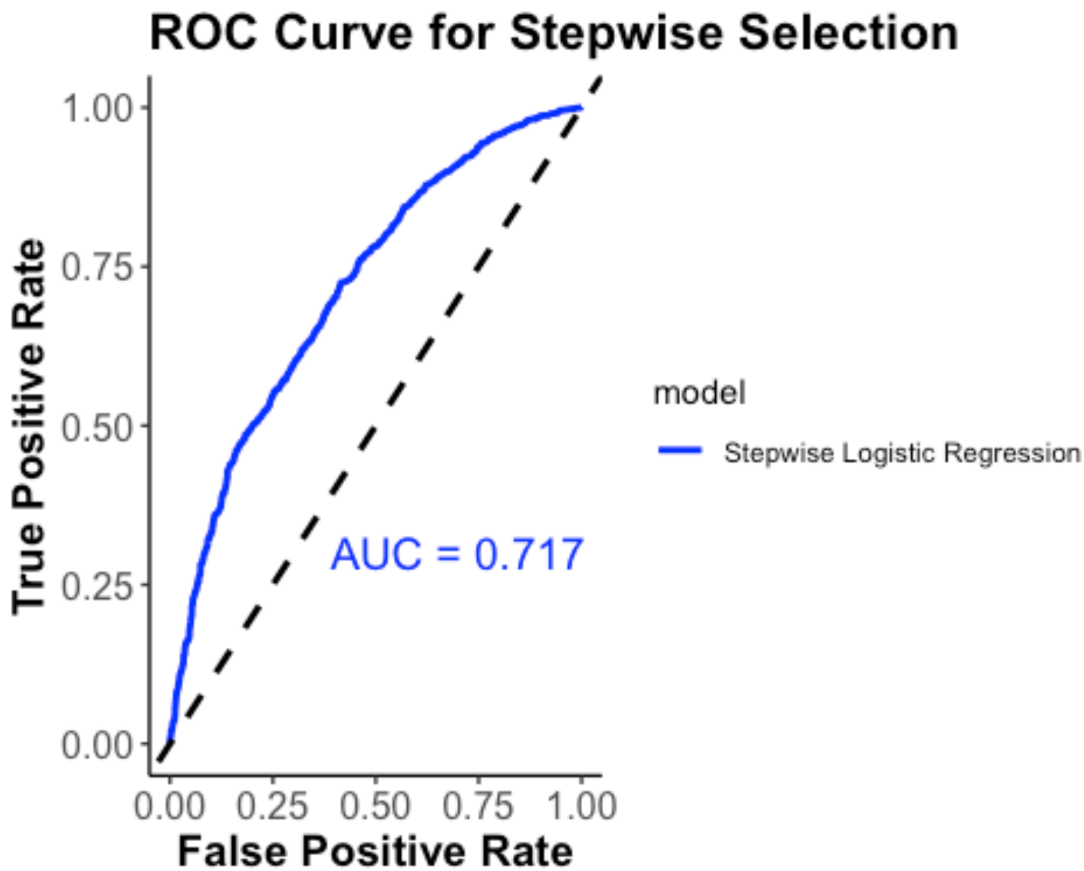
```
##
##  Mcnemar's Test P-Value : < 2.2e-16
##
##             Sensitivity : 0.4578
##             Specificity : 0.8147
##          Pos Pred Value : 0.6277
##          Neg Pred Value : 0.6877
##              Prevalence : 0.4056
##          Detection Rate : 0.1857
##    Detection Prevalence : 0.2958
##       Balanced Accuracy : 0.6363
##
##        'Positive' Class : FALSE
##
```

```
library(ROCR)
glm_prediction <- prediction(glm.predicted_stepwise,
testing_set$Recidivism_Within_3years)
roc_step <- performance(glm_prediction, measure = "tpr", x.measure = "fpr")
roc_obj_s<- roc(testing_set$Recidivism_Within_3years, glm.predicted_stepwise)
```

```
## Setting levels: control = FALSE, case = TRUE
```

```
## Setting direction: controls < cases
```

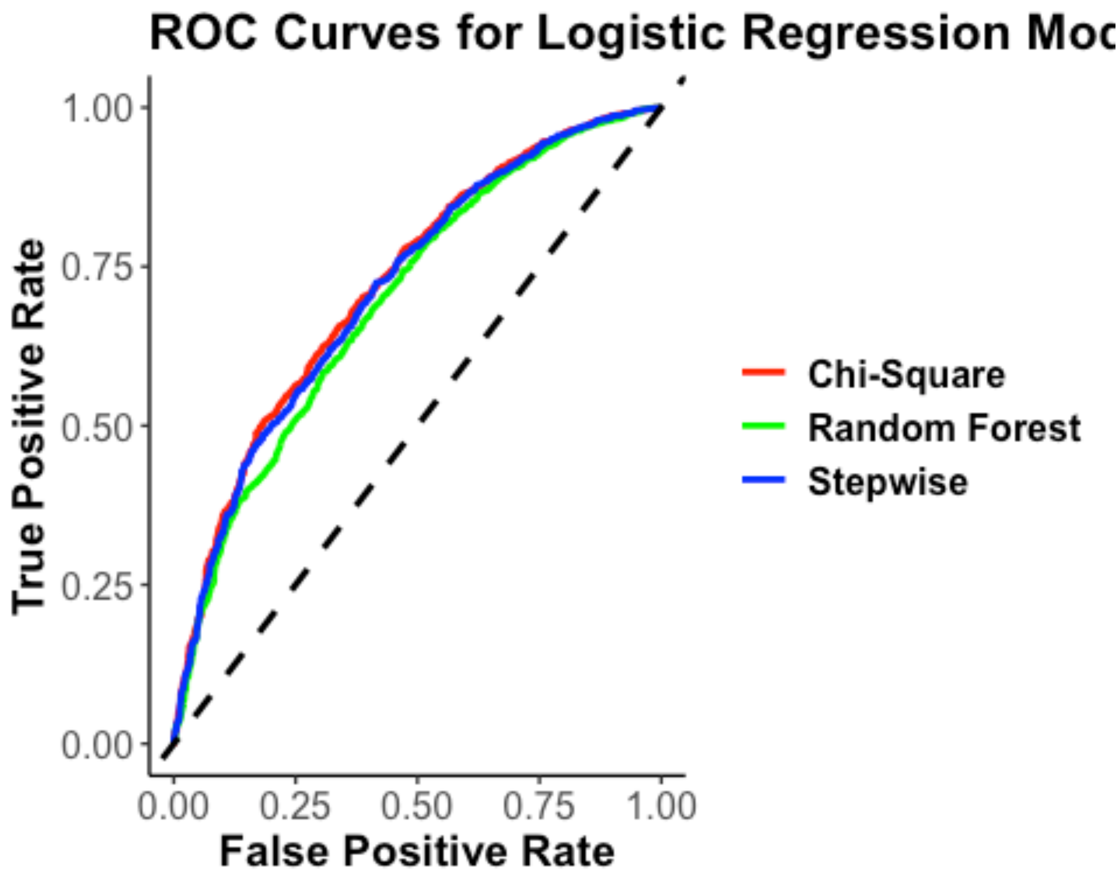
```
auc_stepwise<- auc(roc_obj_s)
#Building ROC curve
roc_data_s <- data.frame(fpr = roc_step@x.values[[1]], tpr =
roc_step@y.values[[1]])
roc_data_s$model <- "Stepwise Logistic Regression"
g3<-ggplot(roc_data_s, aes(x = fpr, y = tpr, color = model)) +
  geom_line(size = 1) +
  geom_abline(intercept = 0, slope = 1, linetype = "dashed", size = 1) +
  labs(title = "ROC Curve for Stepwise Selection", x = "False Positive Rate",
y = "True Positive Rate") +
  scale_color_manual(values = c("Stepwise Logistic Regression" = "blue")) +
  theme_classic() +
  theme(plot.title = element_text(face = "bold", size = 16),
        axis.title = element_text(face = "bold", size = 14),
        axis.text = element_text(size = 12))+
  annotate("text", x = 0.7, y = 0.3, label = paste0("AUC = ",
round(auc_stepwise, 3)), size = 5, color = "blue")
g3
```



Combined ROC Curve

```
library(ROCR)

# Create prediction objects for each model
pred_rf <- prediction(glm_predicted_rf_selected,
testing_set$Recidivism_Within_3years)
pred_chi <- prediction(glm_predicted_chi_square,
testing_set$Recidivism_Within_3years)
glm_prediction <- prediction(glm.predicted_stepwise,
testing_set$Recidivism_Within_3years)

# Compute ROC curves for each model
roc_rf <- performance(pred_rf, measure = "tpr", x.measure = "fpr")
roc_chi <- performance(pred_chi, measure = "tpr", x.measure = "fpr")
roc_stepwise <- performance(glm_prediction, measure = "tpr", x.measure =
"fpr")

roc_obj_rf<-
roc(testing_set$Recidivism_Within_3years,glm_predicted_rf_selected )

## Setting levels: control = FALSE, case = TRUE
```

```
## Setting direction: controls < cases

roc_obj_chi<- roc(testing_set$Recidivism_Within_3years,
glm_predicted_chi_square)

## Setting levels: control = FALSE, case = TRUE
## Setting direction: controls < cases

roc_obj_stepwise<- roc(testing_set$Recidivism_Within_3years,
glm.predicted_stepwise)

## Setting levels: control = FALSE, case = TRUE
## Setting direction: controls < cases

# Compute AUC for each model
auc_rf <- auc(roc_obj_rf)
auc_chi <- auc(roc_obj_chi)
auc_stepwise <- auc(roc_obj_stepwise)


roc_data <- data.frame(
  fpr = c(roc_rf@x.values[[1]], roc_chi@x.values[[1]],
roc_stepwise@x.values[[1]]),
  tpr = c(roc_rf@y.values[[1]], roc_chi@y.values[[1]],
roc_stepwise@y.values[[1]]),
  model = factor(rep(c("Random Forest", "Chi-Square", "Stepwise"),
c(length(roc_rf@x.values[[1]]), length(roc_chi@x.values[[1]]),
length(roc_stepwise@x.values[[1]]))))
)


g4<-ggplot(roc_data, aes(x = fpr, y = tpr, color = model)) +
  geom_line(size = 1) +
  geom_abline(intercept = 0, slope = 1, linetype = "dashed", size = 1) +
  labs(title = "ROC Curves for Logistic Regression Models", x = "False
Positive Rate", y = "True Positive Rate") +
  scale_color_manual(values = c("Random Forest" = "green", "Chi-Square" =
"red", "Stepwise" = "blue")) +
  theme_classic() +
  theme(plot.title = element_text(face = "bold", size = 16),
        axis.title = element_text(face = "bold", size = 14),
        axis.text = element_text(size = 12),
        legend.title = element_blank(),
        legend.text = element_text(face = "bold", size = 12))
g4
```



Choosing the best model from above models based on accuracy and ROC curve

```
library(shiny)
library(ggplot2)
library(ROCR)

# Load the data and create the logistic regression models here

# Create a list of the plots and their names
plots <- list(
  "ROC Curve for Logistic Regression with Chi-Square" =g2 ,

  "ROC Curve for Logistic Regression with Random Forest" = g1,

  "ROC Curve for Stepwise Selection" = g3,

  "ROC Curves for All Logistic Regression Models" =g4
)

# Define the UI for the Shiny app
ui <- fluidPage(
```

```
  selectInput("plot", "Choose a plot:", choices = names(plots)),
  plotOutput("plot_output")
)

# Define the server for the Shiny app
server <- function(input, output, session) {
  observeEvent(input$close, {
    js$closeWindow()
    stopApp()
  })
  output$plot_output <- renderPlot({
    print(plots[[input$plot]])
  })
}

# Run the Shiny app
shinyApp(ui = ui, server = server)
```

- From the above three models based on AUC (Area Under ROC Curve) logistic regression with stepwise variable selection and Chi-Square variable selection have almost same Area Under Curve. When the accuracy and area under the ROC curve for both logistic regression by Chi-Square variable selection method and logistic regression by step wise variable selection method are equal, we favor the model that is easier to understand. As we know chi-square test is performed between two variables and does not account for together variation. Hence due to this uncertainity and given that Stepwise variable selection has yielded similar results, we have decided to go with Stepwise variable Selection.

```
summary(stepwise_model)

##
## Call:
## glm(formula = Recidivism_Within_3years ~ Gang_Affiliated +
Prior_Arrest_Episodes_Felony +
##     Age_at_Release + Employment_Status + Prior_Arrest_Episodes_Misd +
##     Prison_Offense + Residence_Changes + Supervision_Level_First +
##     Prison_Years + Education_Level + Prior_Conviction_Episodes_Misd +
##     Race + Dependents, family = binomial(link = "logit"), data =
training_set)
##
## Coefficients:
##                                          Estimate Std. Error z value
Pr(>|z|)
## (Intercept)                              -0.93914    0.32970  -2.848
0.004393
## Gang_AffiliatedTRUE                       0.95954    0.06491  14.783  < 2e-
16
## Prior_Arrest_Episodes_Felony1             0.24047    0.31760   0.757
0.448957
## Prior_Arrest_Episodes_Felony10 or more    1.73448    0.31720   5.468 4.55e-
```

```
08
## Prior_Arrest_Episodes_Felony2           0.62396    0.31596   1.975 
0.048286
## Prior_Arrest_Episodes_Felony3           0.87356    0.31554   2.768 
0.005633
## Prior_Arrest_Episodes_Felony4           1.05678    0.31655   3.338 
0.000843
## Prior_Arrest_Episodes_Felony5           1.09883    0.31763   3.459 
0.000541
## Prior_Arrest_Episodes_Felony6           1.19703    0.31970   3.744 
0.000181
## Prior_Arrest_Episodes_Felony7           1.00413    0.32063   3.132 
0.001737
## Prior_Arrest_Episodes_Felony8           1.29764    0.32443   4.000 6.34e-
05
## Prior_Arrest_Episodes_Felony9           1.29486    0.32719   3.958 7.57e-
05
## Age_at_Release23-27                    -0.30822    0.08865  -3.477 
0.000508
## Age_at_Release28-32                    -0.63157    0.09536  -6.623 3.52e-
11
## Age_at_Release33-37                    -0.89980    0.10174  -8.844  < 2e-
16
## Age_at_Release38-42                    -1.11938    0.11030 -10.148  < 2e-
16
## Age_at_Release43-47                    -1.27066    0.11471 -11.077  < 2e-
16
## Age_at_Release48 or older              -1.62854    0.11149 -14.608  < 2e-
16
## Employment_StatusUnemployed             0.70439    0.05160  13.651  < 2e-
16
## Prior_Arrest_Episodes_Misd1             0.12487    0.08054   1.550 
0.121027
## Prior_Arrest_Episodes_Misd2             0.23518    0.08868   2.652 
0.008001
## Prior_Arrest_Episodes_Misd3             0.20123    0.09763   2.061 
0.039282
## Prior_Arrest_Episodes_Misd4             0.41323    0.10373   3.984 6.78e-
05
## Prior_Arrest_Episodes_Misd5             0.38866    0.11119   3.495 
0.000473
## Prior_Arrest_Episodes_Misd6 or more     0.61892    0.10168   6.087 1.15e-
09
## Prison_OffenseOther                     0.09220    0.07408   1.245 
0.213310
## Prison_OffenseProperty                  0.20242    0.05914   3.423 
0.000620
## Prison_OffenseViolent/Non-Sex           0.09607    0.06605   1.454 
0.145812
## Prison_OffenseViolent/Sex              -0.48882    0.11867  -4.119 3.80e-
```

```
05
## Residence_Changes1                          0.12919    0.05197   2.486
0.012926
## Residence_Changes2                          0.11698    0.06464   1.810
0.070322
## Residence_Changes3 or more                  0.36142    0.07003   5.161 2.45e-
07
## Supervision_Level_FirstSpecialized          0.07071    0.05719   1.236
0.216363
## Supervision_Level_FirstStandard            -0.22414    0.05489  -4.083 4.44e-
05
## Prison_YearsGreater than 2 to 3 years      -0.15997    0.06153  -2.600
0.009330
## Prison_YearsLess than 1 year                0.19814    0.05827   3.400
0.000673
## Prison_YearsMore than 3 years              -0.06835    0.06161  -1.109
0.267234
## Education_LevelHigh School Diploma          0.17107    0.06345   2.696
0.007012
## Education_LevelLess than HS diploma        -0.01434    0.06517  -0.220
0.825856
## Prior_Conviction_Episodes_Misd1             0.15804    0.06774   2.333
0.019654
## Prior_Conviction_Episodes_Misd2             0.21545    0.08093   2.662
0.007766
## Prior_Conviction_Episodes_Misd3             0.09507    0.09284   1.024
0.305785
## Prior_Conviction_Episodes_Misd4 or more     0.23178    0.09269   2.501
0.012402
## RaceWHITE                                   0.08414    0.04485   1.876
0.060645
## Dependents1                                 0.12452    0.05923   2.102
0.035525
## Dependents2                                -0.01307    0.06363  -0.205
0.837209
## Dependents3 or more                         0.00506    0.05592   0.090
0.927904
##
## (Intercept)                                **
## Gang_AffiliatedTRUE                        ***
## Prior_Arrest_Episodes_Felony1
## Prior_Arrest_Episodes_Felony10 or more     ***
## Prior_Arrest_Episodes_Felony2              *
## Prior_Arrest_Episodes_Felony3              **
## Prior_Arrest_Episodes_Felony4              ***
## Prior_Arrest_Episodes_Felony5              ***
## Prior_Arrest_Episodes_Felony6              ***
## Prior_Arrest_Episodes_Felony7              **
## Prior_Arrest_Episodes_Felony8              ***
## Prior_Arrest_Episodes_Felony9              ***
```

```
## Age_at_Release23-27                        ***
## Age_at_Release28-32                        ***
## Age_at_Release33-37                        ***
## Age_at_Release38-42                        ***
## Age_at_Release43-47                        ***
## Age_at_Release48 or older                  ***
## Employment_StatusUnemployed                ***
## Prior_Arrest_Episodes_Misd1
## Prior_Arrest_Episodes_Misd2                **
## Prior_Arrest_Episodes_Misd3                *
## Prior_Arrest_Episodes_Misd4                ***
## Prior_Arrest_Episodes_Misd5                ***
## Prior_Arrest_Episodes_Misd6 or more        ***
## Prison_OffenseOther
## Prison_OffenseProperty                     ***
## Prison_OffenseViolent/Non-Sex
## Prison_OffenseViolent/Sex                  ***
## Residence_Changes1                         *
## Residence_Changes2                         .
## Residence_Changes3 or more                 ***
## Supervision_Level_FirstSpecialized
## Supervision_Level_FirstStandard            ***
## Prison_YearsGreater than 2 to 3 years      **
## Prison_YearsLess than 1 year               ***
## Prison_YearsMore than 3 years
## Education_LevelHigh School Diploma         **
## Education_LevelLess than HS diploma
## Prior_Conviction_Episodes_Misd1            *
## Prior_Conviction_Episodes_Misd2            **
## Prior_Conviction_Episodes_Misd3
## Prior_Conviction_Episodes_Misd4 or more *
## RaceWHITE                                  .
## Dependents1                                *
## Dependents2
## Dependents3 or more
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 15310  on 11336  degrees of freedom
## Residual deviance: 13318  on 11290  degrees of freedom
## AIC: 13412
##
## Number of Fisher Scoring iterations: 4
```

From the Model Summary We can Infer that

- The most important factors in predicting recidivism three years after being released from jail are, according to the p-values in the summary:

- Age at release: The likelihood that a person will commit another crime decreases with age. Age groups 28 to 32, 33 to 37, 38 to 42, and 43 to 47 all exhibit extremely low p-values, pointing to a significant correlation with recidivism.
- Gang affiliation: People who are members of gangs are more likely to commit crimes again. This variable has a very low p-value, which suggests a high correlation with recidivism.
- Prior arrests Episodes : Having more arrests in the past, especially for misdemeanours and felonies, is linked to a higher risk of re-offending.
- Prison Offense: Prison_Offenses related to Property and Prison_Offense related Violent/Sex are significant predictors of recidivism, with property offenses strongly linked to re-offending, and violent/sex offenses being negatively correlated with re-offending. Prison_OffenseOther and Prison_OffenseViolent/Non-Sex are not significant predictors.
- Employment Status: Employment_Status is highly significant in predicting recidivism, with a coefficient of 0.70439 and a very low p-value (< 2e-16). This suggests that being unemployed increases the likelihood of re-offending. The large positive coefficient indicates a strong association between unemployment and higher recidivism risk.
- Prison Years - The Prison_Years reveals that individuals serving less than 1 year in prison have a significantly higher likelihood of re-offending, as indicated by the positive coefficient and a very low p-value. In contrast, those serving 2 to 3 years show a slight reduction in recidivism risk, though the effect is modest and significant at the 5% level.
- Resident Changes - The Residence_Changes variable shows a clear pattern in predicting recidivism. Individuals with one residence change have a moderate but statistically significant increase in re-offending risk (p-value = 0.0129). Those with two residence changes show a weaker effect, with a p-value of 0.0703, which is marginally significant at the 10% level. However, individuals with three or more residence changes exhibit the strongest association with recidivism, as indicated by a large positive coefficient and a very low p-value (2.45e-07), making it highly significant in predicting re-offending.
- Super Vision Level - The Supervision_Level variable indicates a significant relationship with recidivism. Individuals with FirstStandard supervision level show a strong negative association with re-offending, as the coefficient is -0.22414 and the p-value is very low (4.44e-05), making it highly significant in predicting lower recidivism risk

###Summary

The most significant factors in predicting recidivism three years after being released from prison include:

Age at Release: Younger individuals are more likely to re-offend, with those in older age groups showing a lower chance of re-offending.

Gang Affiliation: People who are part of gangs have a higher chance of committing crimes again after being released.

Prior Arrests: Individuals with a history of more arrests, especially for serious crimes, are more likely to re-offend.

Prison Offense Type: Those convicted of property crimes are more likely to re-offend, while those convicted of violent or sexual offenses tend to have a lower chance of re-offending.

Employment Status: Unemployed individuals are more likely to re-offend, as having a job seems to lower the risk of re-offending.

Prison Time Served: Those who served less than a year in prison are more likely to re-offend, while those who served between 2 to 3 years show a slightly reduced risk.

Residence Changes: People who changed residences multiple times after release are more likely to re-offend, with those having three or more moves showing the highest risk.

Supervision Level: Individuals under standard supervision have a lower chance of re-offending, suggesting that more supervision reduces the likelihood of recidivism.