server.R

```
library(stringr)

library(plyr)

library(lubridate)

library(randomForest)

library(reshape2)

library(caret)

library(shiny)

library(e1071)



df <- read.csv('LoanStats.csv', h=T, stringsAsFactors=F, skip=1)


df[,'desc'] <- NULL

df[,'mths_since_last_record'] <- NULL

poor_coverage <- sapply(df, function(x) {

  coverage <- 1 - sum(is.na(x)) / length(x)

  coverage < 0.8

})

df <- df[,poor_coverage==FALSE]

bad_indicators <- c("Late (16-30 days)", "Late (31-120 days)", "Default", "Charged Off")


df$is_bad <- ifelse(df$loan_status %in% bad_indicators, 1,

                    ifelse(df$loan_status=="", NA,

                           0))

table(df$loan_status)

table(df$is_bad)
```

```
df$issue_d <- as.Date(df$issue_d,format = "%m/%d/%Y")
df$year_issued <- year(df$issue_d)
df$month_issued <- month(df$issue_d)
df$earliest_cr_line <- as.Date(df$earliest_cr_line, format = "%m/%d/%Y")

df$revol_util <- str_replace_all(df$revol_util, "[%]", "")
df$revol_util <- as.numeric(df$revol_util)
outcomes <- ddply(df, .(year_issued, month_issued), function(x) {
  c("percent_bad"=sum(x$is_bad) / nrow(x),
    "n_loans"=nrow(x))
})



df.term <- subset(df, year_issued < 2012)
df.term$home_ownership <- factor(df.term$home_ownership)
df.term$is_rent <- df.term$home_ownership=="RENT"



idx <- runif(nrow(df.term)) > 0.75
train <- df.term[idx==FALSE,]
testData <- df.term[idx==TRUE,]



df$is_bad <- ifelse(df$is_bad == 1, "X", "Y")
fitControl <- trainControl (method = "cv", number = 3)
rfFit <- train (factor(is_bad) ~ last_fico_range_high + last_fico_range_low +
```

```
              revol_util + inq_last_6mths,
            data=df[1:100,c('is_bad','last_fico_range_high','last_fico_range_low',
                        'revol_util','inq_last_6mths')],method = "rf",
            trControl = fitControl, verbose = FALSE)


shinyServer(
  function(input, output, session){

    output$loandf = renderText ({

      min1<- input$Min_FICO_Score
      max1<- input$Max_FICO_Score
      rev<- input$Revolving_Line_Utilization
      cred<- input$Credit_Inquiries_Past_6m
      hom<- input$Home_Ownership
      ann<- input$Annual_Income
      loan<- input$Loan_Amount


      testData$last_fico_range_high <- as.numeric(min1)
      testData$last_fico_range_low<-as.numeric(max1)
      testData$revol_util <- as.numeric(rev)
      testData$inq_last_6mths <- as.numeric(cred)
      testData$home_ownership<-as.character(hom)
      testData$annual_inc<-as.numeric(ann)
      testData$loan_amnt<-as.numeric(loan)
```

```
    summary(df[1,])


    round(predict (rfFit, newdata=as.data.frame(testData[1,]),type='prob')[,1],3)


  })


})
```

ui.R

```
shinyServer(
  pageWithSidebar(
    headerPanel("Predict Bad Loan Applicant"),


    sidebarPanel(
      textInput("Min_FICO_Score","Min FICO Score",500),
      textInput("Max_FICO_Score","Max FICO Score",600),
      textInput("Revolving_Line_Utilization","Revolving Line Utilization",20),
      textInput("Credit_Inquiries_Past_6m","Credit Inquiries Past 6m",1),
      selectInput("Home_Ownership","Home Ownership", choices=c("RENT", "OWN","MORTGAGE")),
      textInput("Annual_Income","Annual_Income",75000),
      textInput("Loan_Amount","Loan_Amount",6000)
    ),
    mainPanel(
      h2 ('Probability of default'),
```

```
    h3 (textOutput ('loandf')),
    tags$style ("#loandf{color: red;
                font-size: 25px;
                font-style: bold;
                }"
     ),

    #h2 ('You Entered'),

     tags$div(class="modal-footer",
              "Note: This is a capstone course project by Madhu Samudrala. The dataset used for prediction is 'Lending Club Loan Dataset' and the
              prediction algorithm is Random Forest.")
     )

))
```