default prediction SVM

KEVIN KANYARI

16/05/2022

loan.2 <- read.csv("C:/Users/admin/Desktop/loan 2.csv", header=TRUE)  
View(loan.2)  
summary(loan.2)

## id member\_id loan\_amnt funded\_amnt   
## Min. : 54734 Min. : 70473 Min. : 500 Min. : 500   
## 1st Qu.: 9206643 1st Qu.:10877134 1st Qu.: 8000 1st Qu.: 8000   
## Median :34433267 Median :37095283 Median :13000 Median :13000   
## Mean :32465133 Mean :35001825 Mean :14755 Mean :14742   
## 3rd Qu.:54908135 3rd Qu.:58471347 3rd Qu.:20000 3rd Qu.:20000   
## Max. :68617057 Max. :73544841 Max. :35000 Max. :35000   
##   
## funded\_amnt\_inv term int\_rate installment   
## Min. : 0 Length:887379 Min. : 5.32 Min. : 15.67   
## 1st Qu.: 8000 Class :character 1st Qu.: 9.99 1st Qu.: 260.70   
## Median :13000 Mode :character Median :12.99 Median : 382.55   
## Mean :14702 Mean :13.25 Mean : 436.72   
## 3rd Qu.:20000 3rd Qu.:16.20 3rd Qu.: 572.60   
## Max. :35000 Max. :28.99 Max. :1445.46   
##   
## grade sub\_grade emp\_length home\_ownership   
## Length:887379 Length:887379 Length:887379 Length:887379   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## annual\_inc verification\_status issue\_d loan\_status   
## Min. : 0 Length:887379 Length:887379 Length:887379   
## 1st Qu.: 45000 Class :character Class :character Class :character   
## Median : 65000 Mode :character Mode :character Mode :character   
## Mean : 75028   
## 3rd Qu.: 90000   
## Max. :9500000   
## NA's :4   
## pymnt\_plan purpose title zip\_code   
## Length:887379 Length:887379 Length:887379 Length:887379   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## addr\_state dti delinq\_2yrs earliest\_cr\_line   
## Length:887379 Min. : 0.00 Min. : 0.0000 Length:887379   
## Class :character 1st Qu.: 11.91 1st Qu.: 0.0000 Class :character   
## Mode :character Median : 17.65 Median : 0.0000 Mode :character   
## Mean : 18.16 Mean : 0.3144   
## 3rd Qu.: 23.95 3rd Qu.: 0.0000   
## Max. :9999.00 Max. :39.0000   
## NA's :29   
## inq\_last\_6mths open\_acc pub\_rec revol\_bal   
## Min. : 0.0000 Min. : 0.00 Min. : 0.0000 Min. : 0   
## 1st Qu.: 0.0000 1st Qu.: 8.00 1st Qu.: 0.0000 1st Qu.: 6443   
## Median : 0.0000 Median :11.00 Median : 0.0000 Median : 11875   
## Mean : 0.6946 Mean :11.55 Mean : 0.1953 Mean : 16921   
## 3rd Qu.: 1.0000 3rd Qu.:14.00 3rd Qu.: 0.0000 3rd Qu.: 20829   
## Max. :33.0000 Max. :90.00 Max. :86.0000 Max. :2904836   
## NA's :29 NA's :29 NA's :29   
## revol\_util total\_acc initial\_list\_status out\_prncp   
## Min. : 0.00 Min. : 1.00 Length:887379 Min. : 0   
## 1st Qu.: 37.70 1st Qu.: 17.00 Class :character 1st Qu.: 0   
## Median : 56.00 Median : 24.00 Mode :character Median : 6458   
## Mean : 55.07 Mean : 25.27 Mean : 8403   
## 3rd Qu.: 73.60 3rd Qu.: 32.00 3rd Qu.:13659   
## Max. :892.30 Max. :169.00 Max. :49373   
## NA's :502 NA's :29   
## out\_prncp\_inv total\_pymnt total\_pymnt\_inv total\_rec\_prncp  
## Min. : 0 Min. : 0 Min. : 0 Min. : 0   
## 1st Qu.: 0 1st Qu.: 1915 1st Qu.: 1900 1st Qu.: 1201   
## Median : 6456 Median : 4895 Median : 4862 Median : 3215   
## Mean : 8400 Mean : 7559 Mean : 7521 Mean : 5758   
## 3rd Qu.:13654 3rd Qu.:10617 3rd Qu.:10566 3rd Qu.: 8000   
## Max. :49373 Max. :57778 Max. :57778 Max. :35000   
##   
## total\_rec\_int total\_rec\_late\_fee recoveries   
## Min. : 0.0 Min. : 0.0000 Min. : 0.00   
## 1st Qu.: 441.5 1st Qu.: 0.0000 1st Qu.: 0.00   
## Median : 1073.3 Median : 0.0000 Median : 0.00   
## Mean : 1754.8 Mean : 0.3967 Mean : 45.92   
## 3rd Qu.: 2238.3 3rd Qu.: 0.0000 3rd Qu.: 0.00   
## Max. :24205.6 Max. :358.6800 Max. :33520.27   
##   
## collection\_recovery\_fee last\_pymnt\_d last\_pymnt\_amnt   
## Min. : 0.000 Length:887379 Min. : 0.0   
## 1st Qu.: 0.000 Class :character 1st Qu.: 280.2   
## Median : 0.000 Mode :character Median : 462.8   
## Mean : 4.881 Mean : 2164.2   
## 3rd Qu.: 0.000 3rd Qu.: 831.2   
## Max. :7002.190 Max. :36475.6   
##   
## last\_credit\_pull\_d collections\_12\_mths\_ex\_med policy\_code application\_type   
## Length:887379 Min. : 0.00000 Min. :1 Length:887379   
## Class :character 1st Qu.: 0.00000 1st Qu.:1 Class :character   
## Mode :character Median : 0.00000 Median :1 Mode :character   
## Mean : 0.01438 Mean :1   
## 3rd Qu.: 0.00000 3rd Qu.:1   
## Max. :20.00000 Max. :1   
## NA's :145   
## acc\_now\_delinq   
## Min. : 0.000000   
## 1st Qu.: 0.000000   
## Median : 0.000000   
## Mean : 0.004991   
## 3rd Qu.: 0.000000   
## Max. :14.000000   
## NA's :29

loan.c <- na.omit(loan.2)  
summary(loan.c)

## id member\_id loan\_amnt funded\_amnt   
## Min. : 54734 Min. : 70825 Min. : 500 Min. : 500   
## 1st Qu.: 9208870 1st Qu.:10879122 1st Qu.: 8000 1st Qu.: 8000   
## Median :34442491 Median :37096700 Median :13000 Median :13000   
## Mean :32473897 Mean :35011269 Mean :14758 Mean :14745   
## 3rd Qu.:54908928 3rd Qu.:58479058 3rd Qu.:20000 3rd Qu.:20000   
## Max. :68617057 Max. :73544841 Max. :35000 Max. :35000   
## funded\_amnt\_inv term int\_rate installment   
## Min. : 0 Length:886764 Min. : 5.32 Min. : 15.67   
## 1st Qu.: 8000 Class :character 1st Qu.: 9.99 1st Qu.: 260.85   
## Median :13000 Mode :character Median :12.99 Median : 382.55   
## Mean :14706 Mean :13.25 Mean : 436.79   
## 3rd Qu.:20000 3rd Qu.:16.20 3rd Qu.: 572.64   
## Max. :35000 Max. :28.99 Max. :1445.46   
## grade sub\_grade emp\_length home\_ownership   
## Length:886764 Length:886764 Length:886764 Length:886764   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
## annual\_inc verification\_status issue\_d loan\_status   
## Min. : 0 Length:886764 Length:886764 Length:886764   
## 1st Qu.: 45000 Class :character Class :character Class :character   
## Median : 65000 Mode :character Mode :character Mode :character   
## Mean : 75019   
## 3rd Qu.: 90000   
## Max. :9500000   
## pymnt\_plan purpose title zip\_code   
## Length:886764 Length:886764 Length:886764 Length:886764   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
## addr\_state dti delinq\_2yrs earliest\_cr\_line   
## Length:886764 Min. : 0.00 Min. : 0.0000 Length:886764   
## Class :character 1st Qu.: 11.91 1st Qu.: 0.0000 Class :character   
## Mode :character Median : 17.66 Median : 0.0000 Mode :character   
## Mean : 18.16 Mean : 0.3143   
## 3rd Qu.: 23.95 3rd Qu.: 0.0000   
## Max. :9999.00 Max. :39.0000   
## inq\_last\_6mths open\_acc pub\_rec revol\_bal   
## Min. : 0.0000 Min. : 1.00 Min. : 0.0000 Min. : 0   
## 1st Qu.: 0.0000 1st Qu.: 8.00 1st Qu.: 0.0000 1st Qu.: 6451   
## Median : 0.0000 Median :11.00 Median : 0.0000 Median : 11880   
## Mean : 0.6945 Mean :11.55 Mean : 0.1954 Mean : 16925   
## 3rd Qu.: 1.0000 3rd Qu.:14.00 3rd Qu.: 0.0000 3rd Qu.: 20834   
## Max. :33.0000 Max. :90.00 Max. :86.0000 Max. :2904836   
## revol\_util total\_acc initial\_list\_status out\_prncp   
## Min. : 0.00 Min. : 1.00 Length:886764 Min. : 0   
## 1st Qu.: 37.70 1st Qu.: 17.00 Class :character 1st Qu.: 0   
## Median : 56.00 Median : 24.00 Mode :character Median : 6466   
## Mean : 55.07 Mean : 25.27 Mean : 8406   
## 3rd Qu.: 73.60 3rd Qu.: 32.00 3rd Qu.:13661   
## Max. :892.30 Max. :169.00 Max. :49373   
## out\_prncp\_inv total\_pymnt total\_pymnt\_inv total\_rec\_prncp  
## Min. : 0 Min. : 0 Min. : 0 Min. : 0   
## 1st Qu.: 0 1st Qu.: 1915 1st Qu.: 1901 1st Qu.: 1201   
## Median : 6461 Median : 4895 Median : 4864 Median : 3215   
## Mean : 8402 Mean : 7559 Mean : 7522 Mean : 5758   
## 3rd Qu.:13657 3rd Qu.:10618 3rd Qu.:10567 3rd Qu.: 8000   
## Max. :49373 Max. :57778 Max. :57778 Max. :35000   
## total\_rec\_int total\_rec\_late\_fee recoveries   
## Min. : 0.0 Min. : 0.0000 Min. : 0.00   
## 1st Qu.: 441.5 1st Qu.: 0.0000 1st Qu.: 0.00   
## Median : 1073.6 Median : 0.0000 Median : 0.00   
## Mean : 1755.0 Mean : 0.3955 Mean : 45.91   
## 3rd Qu.: 2238.6 3rd Qu.: 0.0000 3rd Qu.: 0.00   
## Max. :24205.6 Max. :358.6800 Max. :33520.27   
## collection\_recovery\_fee last\_pymnt\_d last\_pymnt\_amnt   
## Min. : 0.000 Length:886764 Min. : 0.0   
## 1st Qu.: 0.000 Class :character 1st Qu.: 280.2   
## Median : 0.000 Mode :character Median : 462.9   
## Mean : 4.877 Mean : 2164.3   
## 3rd Qu.: 0.000 3rd Qu.: 831.2   
## Max. :7002.190 Max. :36475.6   
## last\_credit\_pull\_d collections\_12\_mths\_ex\_med policy\_code application\_type   
## Length:886764 Min. : 0.00000 Min. :1 Length:886764   
## Class :character 1st Qu.: 0.00000 1st Qu.:1 Class :character   
## Mode :character Median : 0.00000 Median :1 Mode :character   
## Mean : 0.01438 Mean :1   
## 3rd Qu.: 0.00000 3rd Qu.:1   
## Max. :20.00000 Max. :1   
## acc\_now\_delinq   
## Min. : 0.000000   
## 1st Qu.: 0.000000   
## Median : 0.000000   
## Mean : 0.004989   
## 3rd Qu.: 0.000000   
## Max. :14.000000

# We begin by loading the required packages  
library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5 v purrr 0.3.4  
## v tibble 3.1.0 v dplyr 1.0.5  
## v tidyr 1.1.4 v stringr 1.4.0  
## v readr 2.1.1 v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(tidyr)  
library(ggplot2)  
library(purrr)  
library(tibble)  
library(dplyr)  
library(stringr)  
library(forcats)  
library(ggthemes)  
library(corrplot)

## corrplot 0.92 loaded

library(GGally)

## Registered S3 method overwritten by 'GGally':  
## method from   
## +.gg ggplot2

library(DT)  
library(caret)

## Loading required package: lattice

##   
## Attaching package: 'caret'

## The following object is masked from 'package:purrr':  
##   
## lift

library(sqldf)

## Loading required package: gsubfn

## Loading required package: proto

## Loading required package: RSQLite

library(plotly)

##   
## Attaching package: 'plotly'

## The following object is masked from 'package:ggplot2':  
##   
## last\_plot

## The following object is masked from 'package:stats':  
##   
## filter

## The following object is masked from 'package:graphics':  
##   
## layout

library(DescTools)

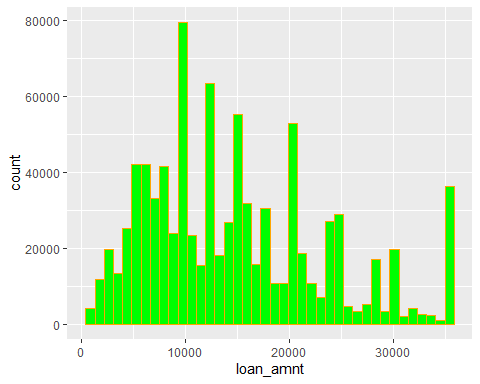
##   
## Attaching package: 'DescTools'

## The following objects are masked from 'package:caret':  
##   
## MAE, RMSE

######Data visualization  
dim(loan.c)

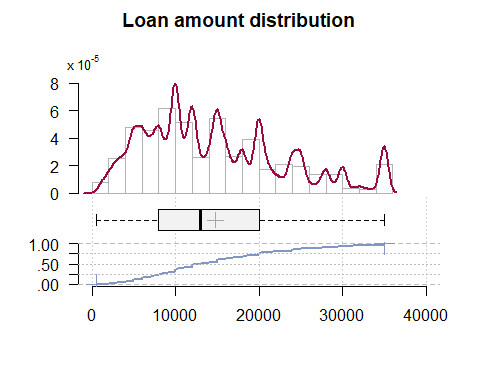
## [1] 886764 47

ggplot(data=loan.c, aes(loan\_amnt))+geom\_histogram(bins = 40,color="orange",fill="green")



Desc(loan.c$loan\_amnt, main = "Loan amount distribution", plotit = TRUE)

## ------------------------------------------------------------------------------   
## Loan amount distribution  
##   
## length n NAs unique 0s mean meanCI'  
## 886'764 886'764 0 1'372 0 14'758.02 14'740.46  
## 100.0% 0.0% 0.0% 14'775.57  
##   
## .05 .10 .25 median .75 .90 .95  
## 3'600.00 5'000.00 8'000.00 13'000.00 20'000.00 28'000.00 32'000.00  
##   
## range sd vcoef mad IQR skew kurt  
## 34'500.00 8'434.27 0.57 8'599.08 12'000.00 0.68 -0.26  
##   
## lowest : 500 (11), 550, 600 (6), 700 (2), 725  
## highest: 34'900 (14), 34'925 (9), 34'950 (18), 34'975 (31), 35'000 (36'341)  
##   
## heap(?): remarkable frequency (7.0%) for the mode(s) (= 10000)  
##   
## ' 95%-CI (classic)



#correlogram of the numeric variables  
loan.num<- data.frame (loan.c$loan\_amnt,loan.c$funded\_amnt\_inv,loan.c$int\_rate,loan.c$installment,loan.c$annual\_inc,loan.c$total\_pymnt)  
corrplot(cor(loan.num))

