Credit Card Default Detection Capstone Project - Code

Anubhav Gupta

April 07, 2021

Contents

1 Executive Summary

It is important that credit card companies are able to recognize customers credibility value by looking at its payment records as well as other demographic characteristics. This will further help companies for risk management. The datasets contains customers default payments in Taiwan from April to September, 2005

For achieving the task of classifying customer at default or not, several algorithms such as Naive Bayes Classifier, KNN , SVM (Linear/ Sigmoid Kernel), Random Forest, GBM, XGBoost and LightGBM.

2 Exploratory Data Analysis

2.1 The Dataset

This dataset presents 6 months data of 30000 customers , where we have **6636 defaulted** ones .

The dataset contains following variables:

LIMIT_BAL: Amount of the given credit (NT dollar): it includes both the individual consumer credit and his/her family (supplementary) credit.

SEX:1 = male; 2 = female.

EDUCATION: Education (1 = graduate school; 2 = university; 3 = high school; 0/4/5/6 = others).

MARRIAGE: Marital status (1 = married; 2 = single; 3 = others).

AGE: Age (year).

PAY_0 - PAY_6: History of past payment. We tracked the past monthly payment records (from April to September, 2005) as follows: PAY_0 = the repayment status in September, 2005; PAY_2 = the repayment status in August, 2005; . . .;PAY_6 = the repayment status in April, 2005. The measurement scale for the repayment status is: -1 = pay duly; 1 = payment delay for one month; 2 = payment delay for two months; . . .; 8 = payment delay for eight months; 9 = payment delay for nine months and above.

BILL_AMT1-BILL_AMT6: Amount of bill statement (NT dollar). BILL_AMT1 = amount of bill statement in September, 2005; BILL_AMT2 = amount of bill statement in August, 2005; . . .; BILL_AMT6 = amount of bill statement in April, 2005.

PAY_AMT1-PAY_AMT6: Amount of previous payment (NT dollar). PAY_AMT1 = amount paid in September, 2005; PAY_AMT2 = amount paid in August, 2005; . . .;PAY_AMT6 = amount paid in April, 2005.

class: 1: Default 0: Non-Default

Source

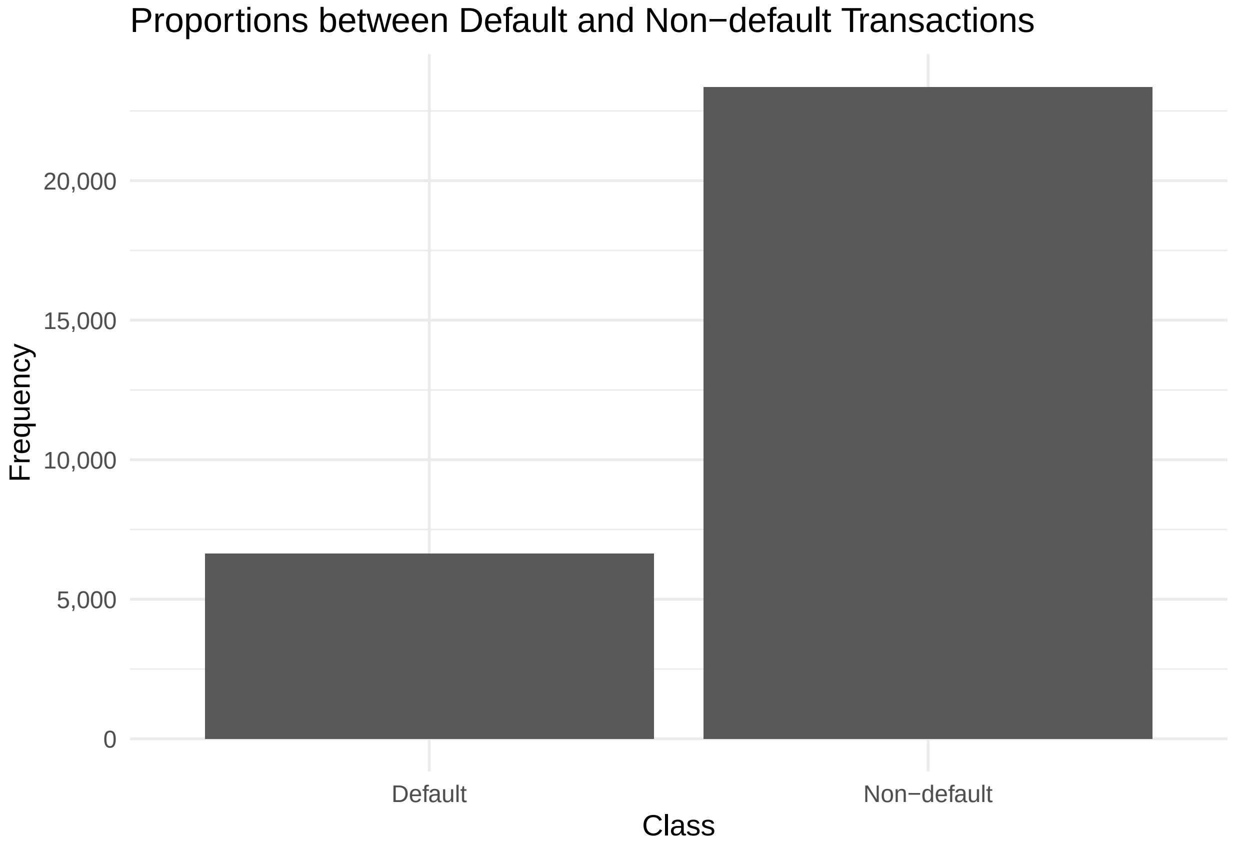
`https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients`

Dimensions

```
##   Length Columns
## 1  30000      25
```

Data Distribution

The proportion of default customers is around 22% in entire dataset.



class	Count
Default	6636
Non-default	23364

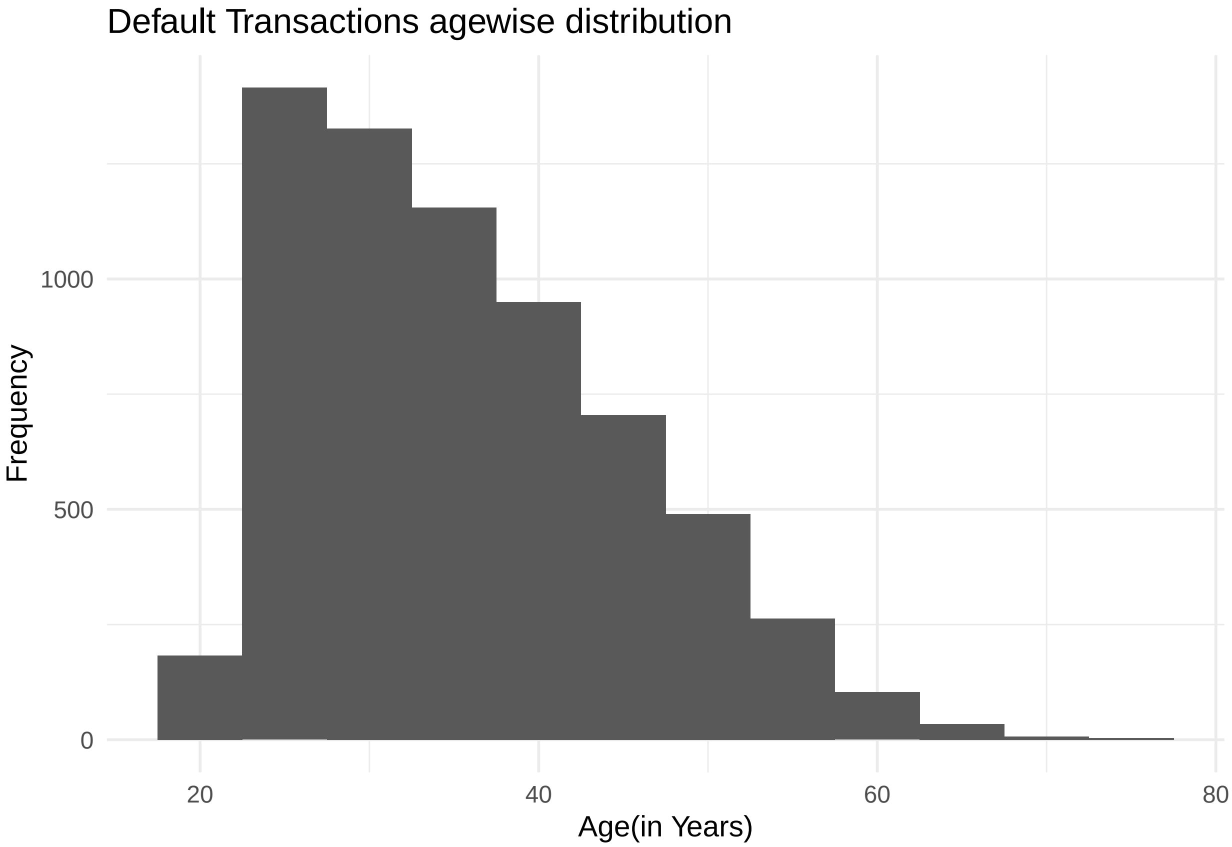
Missing Values

As the table below suggests, there aren't missing values in this dataframe.

	x
ID	0
LIMIT_BAL	0
SEX	0
EDUCATION	0
MARRIAGE	0
AGE	0
PAY_0	0
PAY_2	0
PAY_3	0
PAY_4	0
PAY_5	0
PAY_6	0
BILL_AMT1	0
BILL_AMT2	0
BILL_AMT3	0
BILL_AMT4	0
BILL_AMT5	0
BILL_AMT6	0
PAY_AMT1	0
PAY_AMT2	0
PAY_AMT3	0
PAY_AMT4	0
PAY_AMT5	0
PAY_AMT6	0
class	0

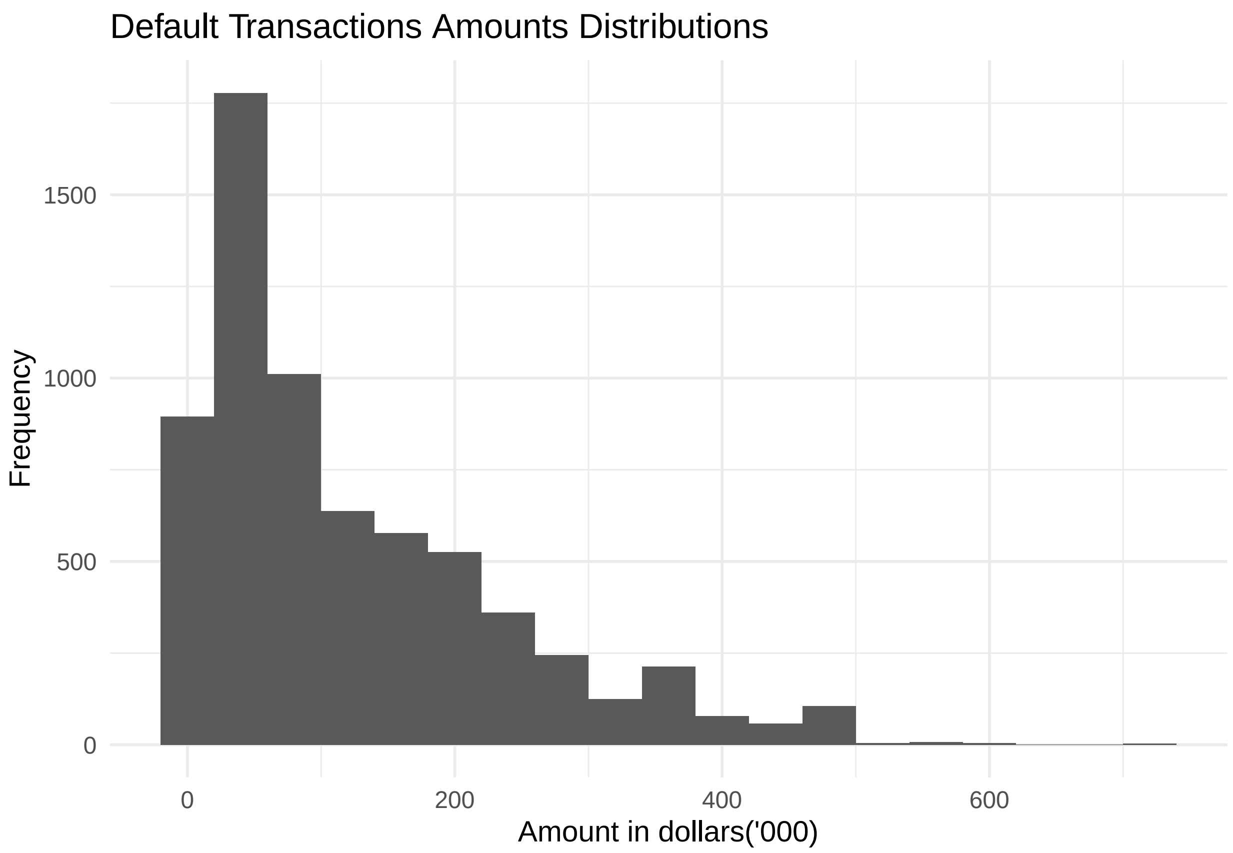
Default Distribution by age

People in age group of 25-40 are mainly defaulters. The pattern of default shows downward trend as age increases.



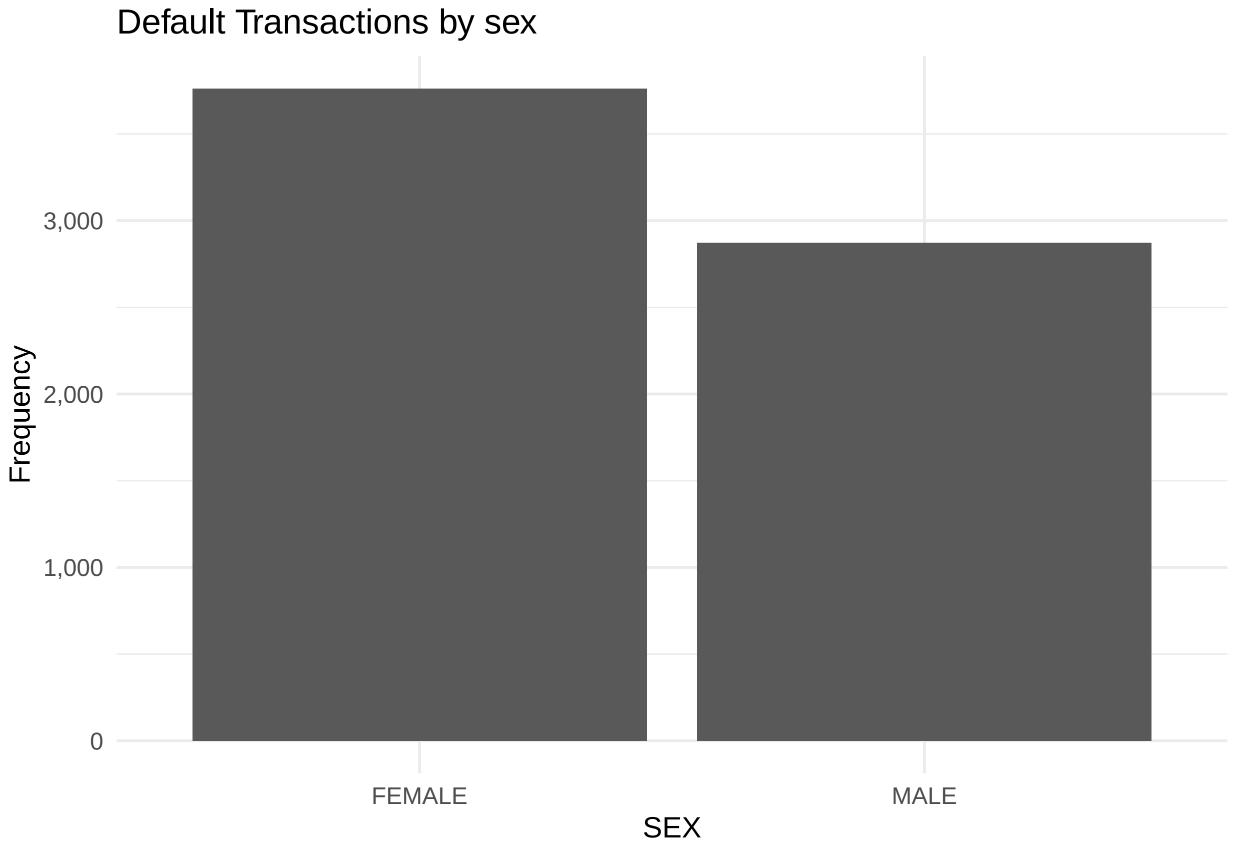
Credit Amount Distribution

The below graph shows distribution of defaulted customers credit amount



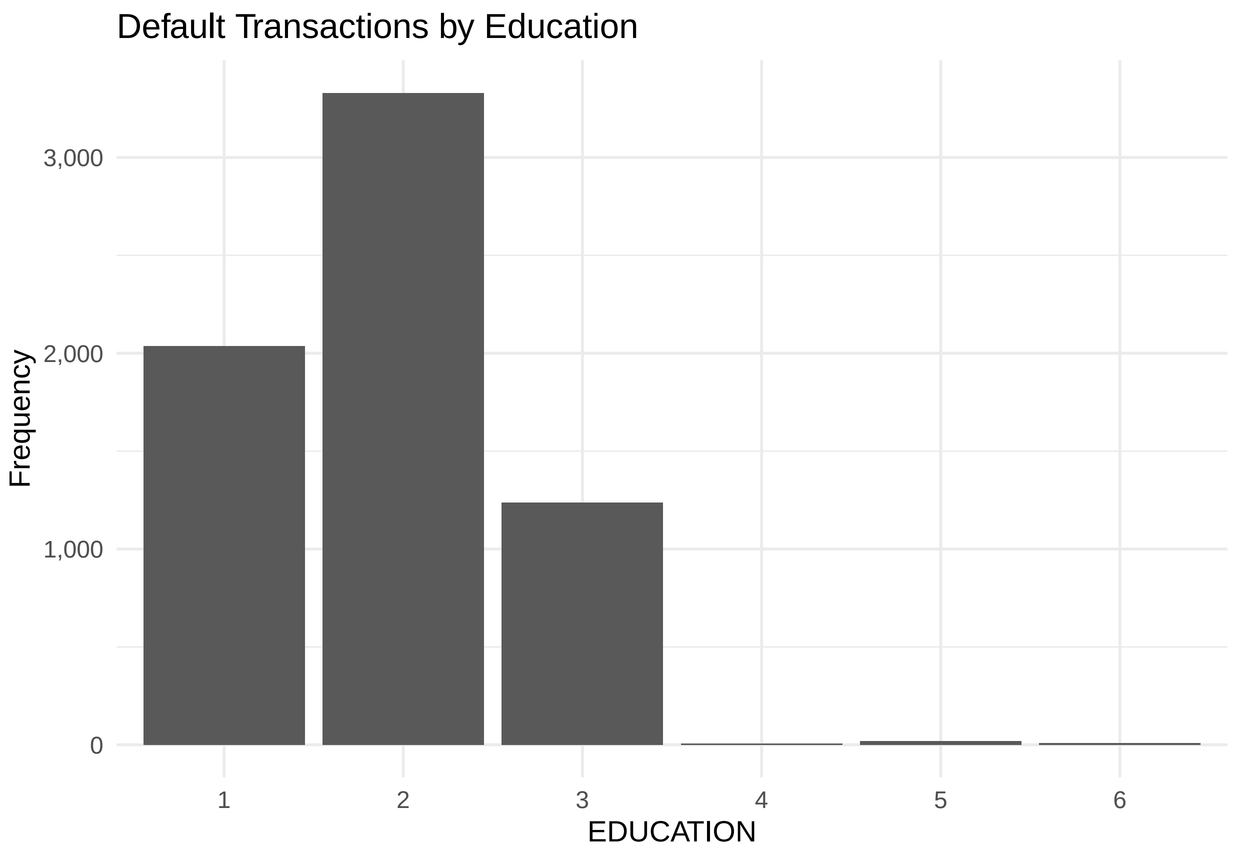
Default Distribution by SEX

The below graph suggests females are more among defaulted population



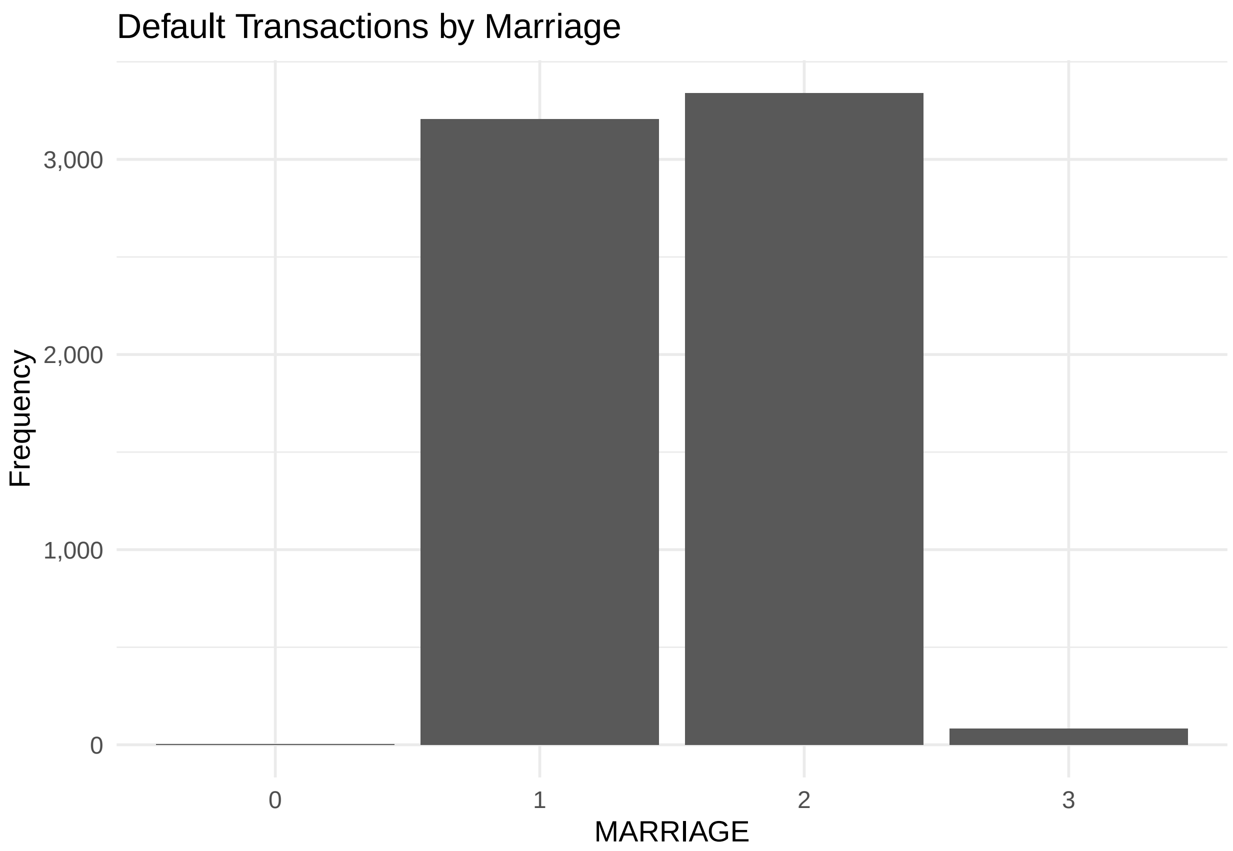
Default Distribution by Education

The below graph suggests graduated people are more among defaulted population



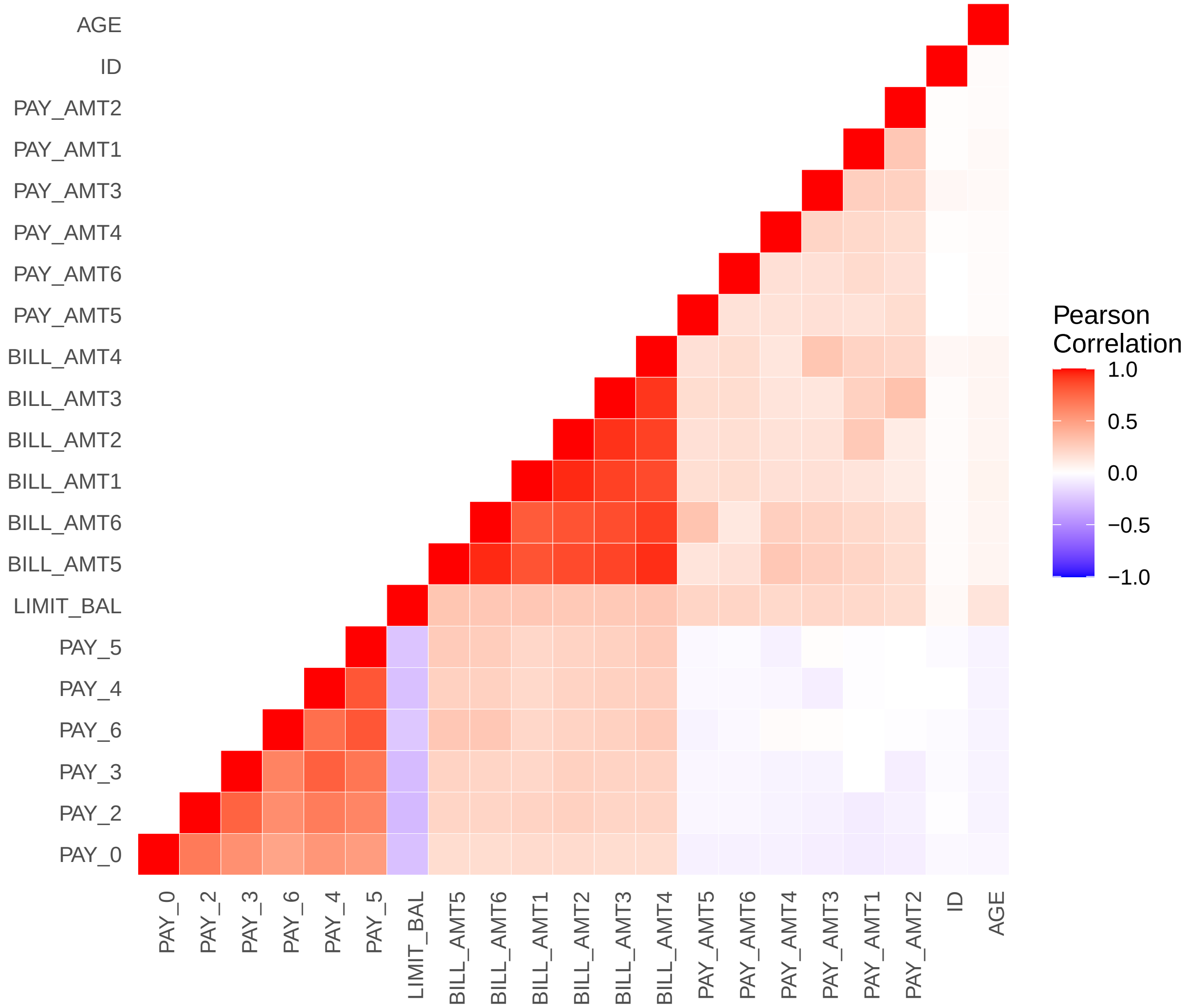
Default Distribution by Marriage

The below graph suggest martial status does not prominently distinguishes defaulted population


Default Transactions by Marriage
Frequency
3,000
2,000
1,000
0
0
1
2
3
MARRIAGE

Correlations between each variables

The correlation analysis shows that BILL__AMT1-BILL__AMT6 are highly correlated among themselves. So considering only the latest Bill amount is sufficient in analysis



3 Data Pre-Processing

Before continuing to build models, It have to do a little data pre-processing:

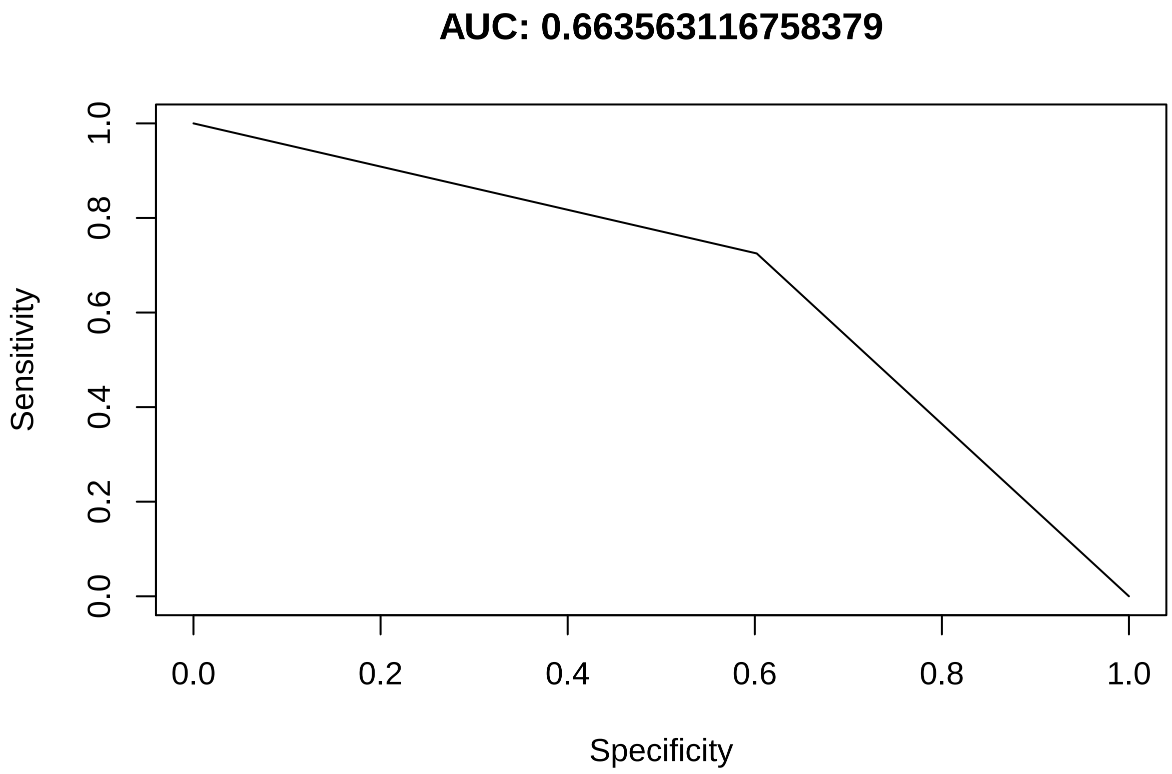
Exploratory Data Analysis and Correlation analysis suggested that the variables ID,BILL_AMT2,BILL_AMT3,BILL_AMT4,BILL_AMT5,BILL_AMT6 , MARRIAGE can be dropped from the dataset.

Split the dataset into train, test, cv dataset.

4 Analysis - Models Building and Comparison

4.1 Naive Bayes

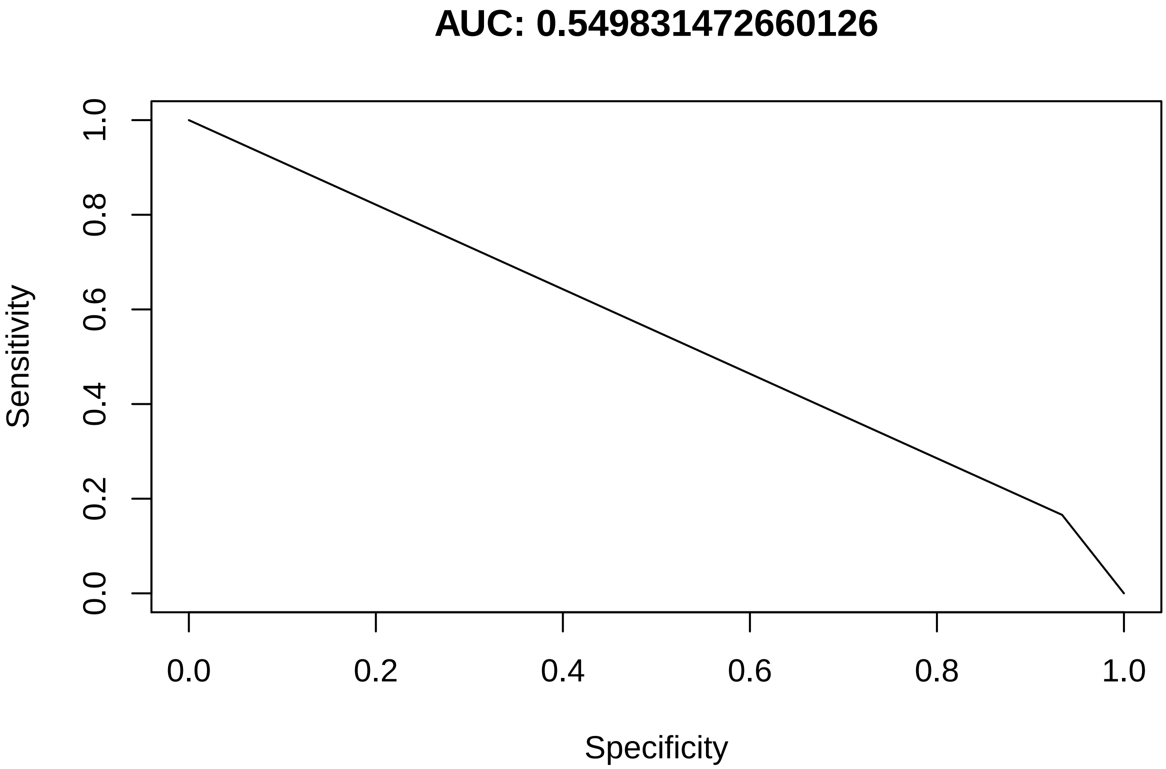
Naive Bayes Classifier is the first algorithm used in analysis. AUC is calculated which is later compared with rest of the model



Model	AUC
Naive Bayes	0.6635631

4.2 KNN - K-Nearest Neighbors

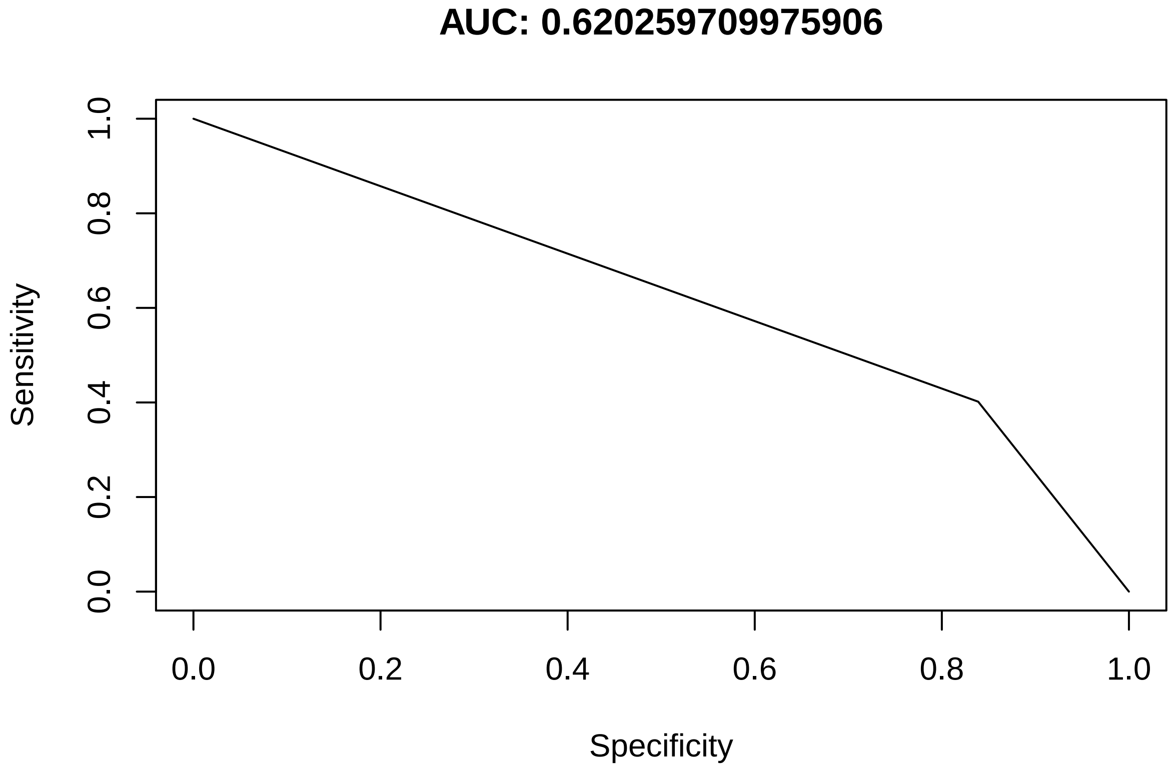
A KNN Model is fitted to data with k=5



Model	AUC
Naive Bayes	0.6635631
K-Nearest Neighbors k=5	0.5498315

4.3 SVM - Support Vector Machine (Sigmoid Kernel)

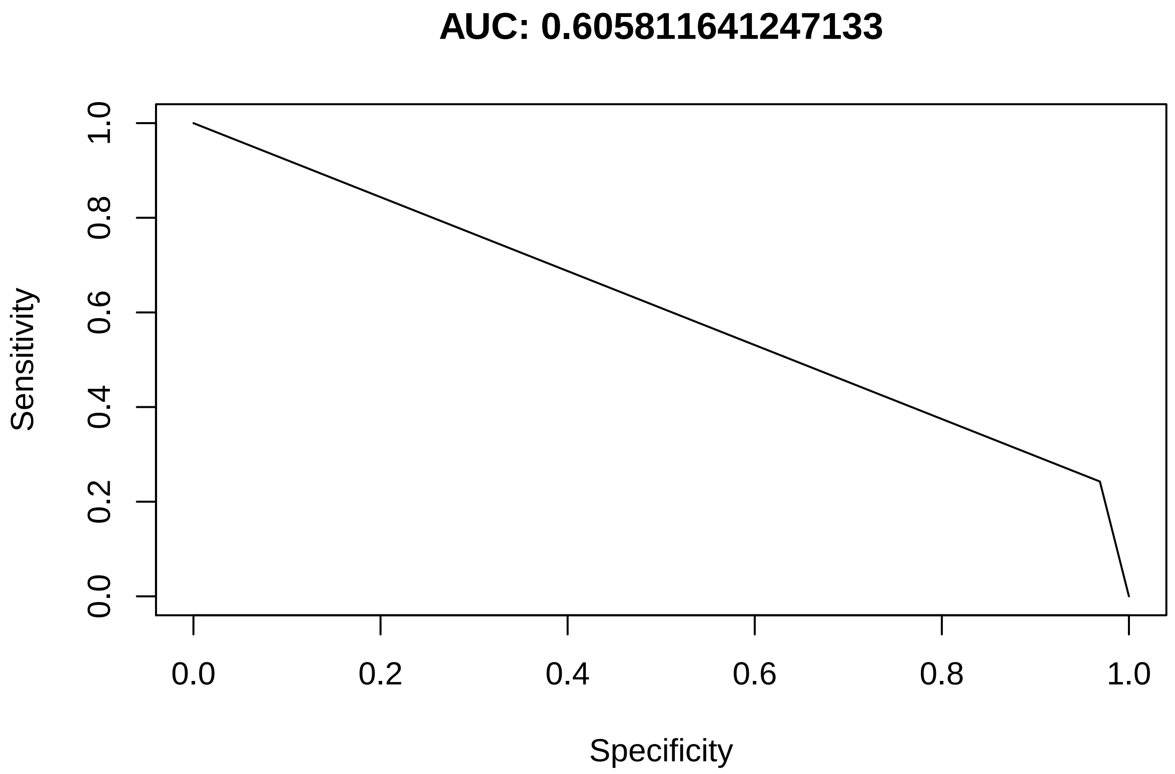
The SVM Model with a Sigmoid Kernel is fitted to data



Model	AUC
Naive Bayes	0.6635631
K-Nearest Neighbors k=5	0.5498315
SVM - Support Vector Machine-Sigmoid Kernel	0.6202597

4.4 SVM - Support Vector Machine (Linear Kernel)

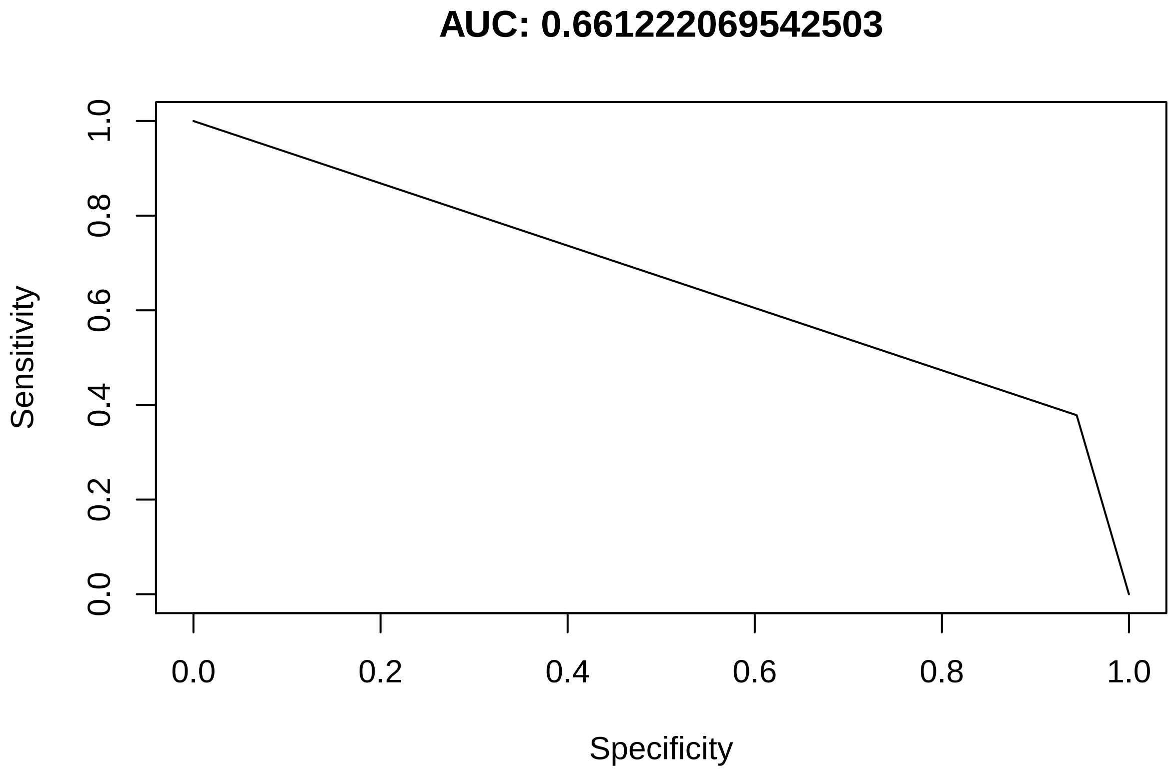
The SVM Model with a Linear Kernel is fitted to data



Model	AUC
Naive Bayes	0.6635631
K-Nearest Neighbors k=5	0.5498315
SVM - Support Vector Machine-Sigmoid Kernel	0.6202597
SVM - Support Vector Machine-Linear Kernel	0.6202597

4.5 Random Forest

The ensemble methods are capable of a significant increase in performance. Hence Random Forest is fitted to dataset

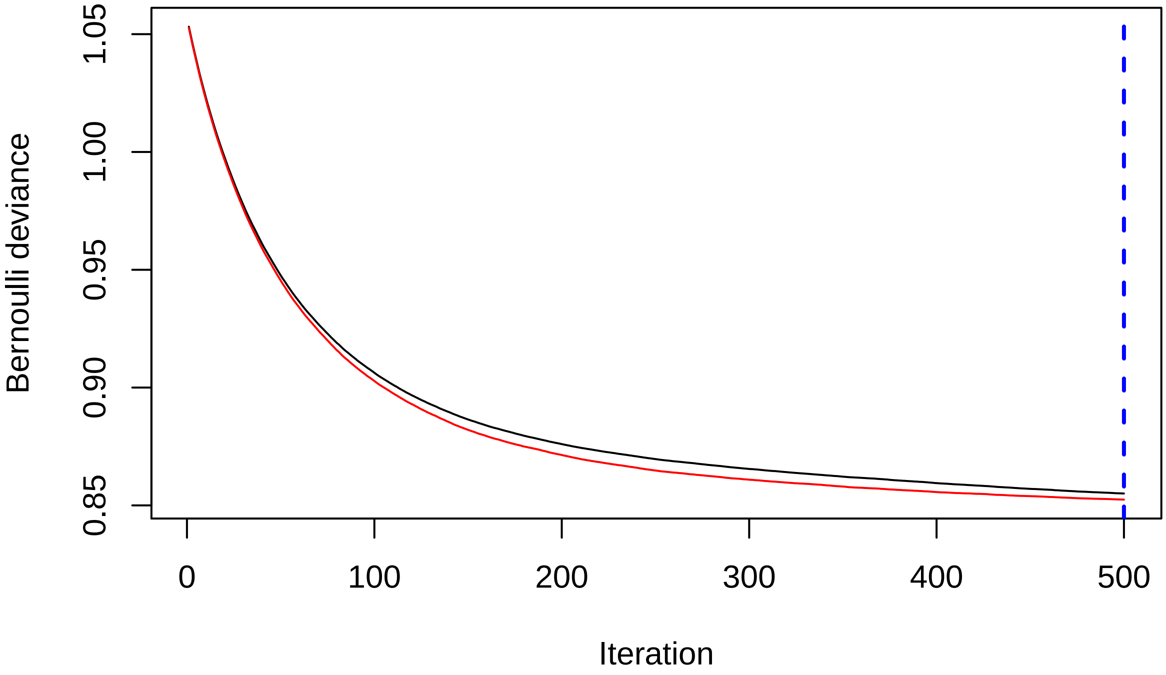


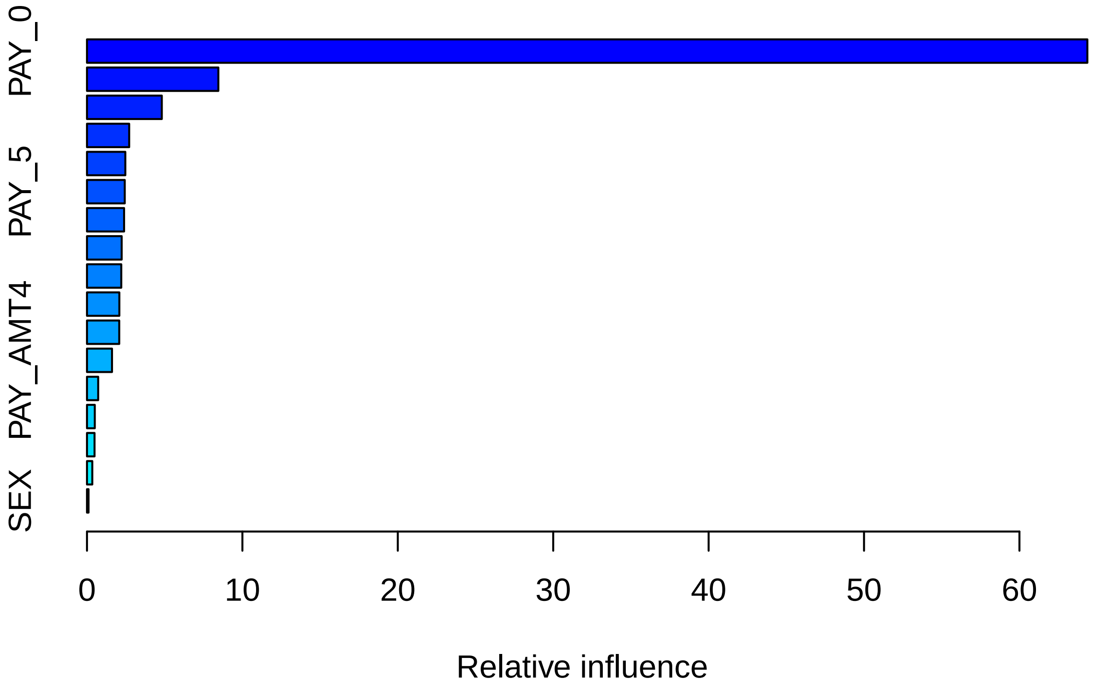
Model	AUC
Naive Bayes	0.6635631
K-Nearest Neighbors k=5	0.5498315
SVM - Support Vector Machine-Sigmoid Kernel	0.6202597
SVM - Support Vector Machine-Linear Kernel	0.6202597
Random Forest	0.6612221

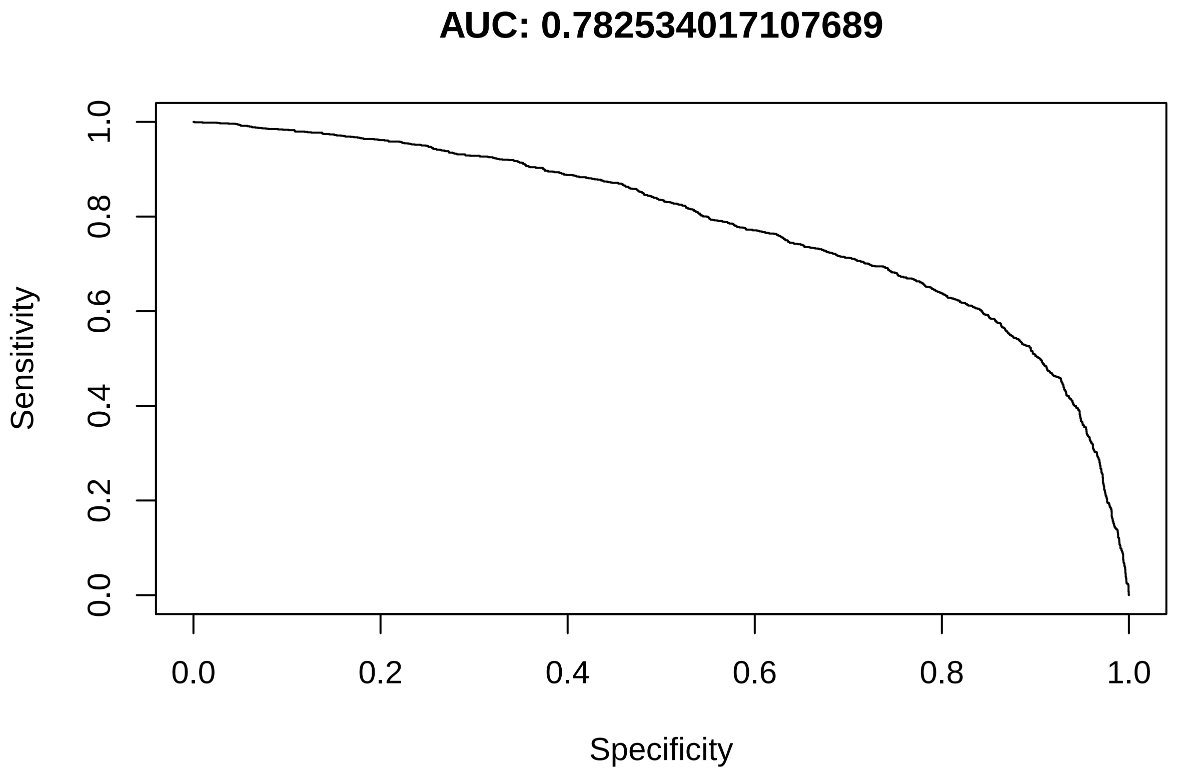
	MeanDecreaseGini
LIMIT_BAL	439.8238
SEX	86.6646
EDUCATION	168.8526
AGE	506.8629
PAY_0	682.5416
PAY_2	298.0613
PAY_3	203.9227
PAY_4	165.0536
PAY_5	151.1144
PAY_6	135.3740
BILL_AMT1	602.7844
PAY_AMT1	465.2005
PAY_AMT2	447.1369
PAY_AMT3	430.4475
PAY_AMT4	412.9895
PAY_AMT5	411.3236
PAY_AMT6	421.2530

4.6 GBM - Generalized Boosted Regression

The GBM performance are really good till now.




PAY_0
PAY_5
PAY_AMT4
SEX
0
10
20
30
40
50
60
Relative influence



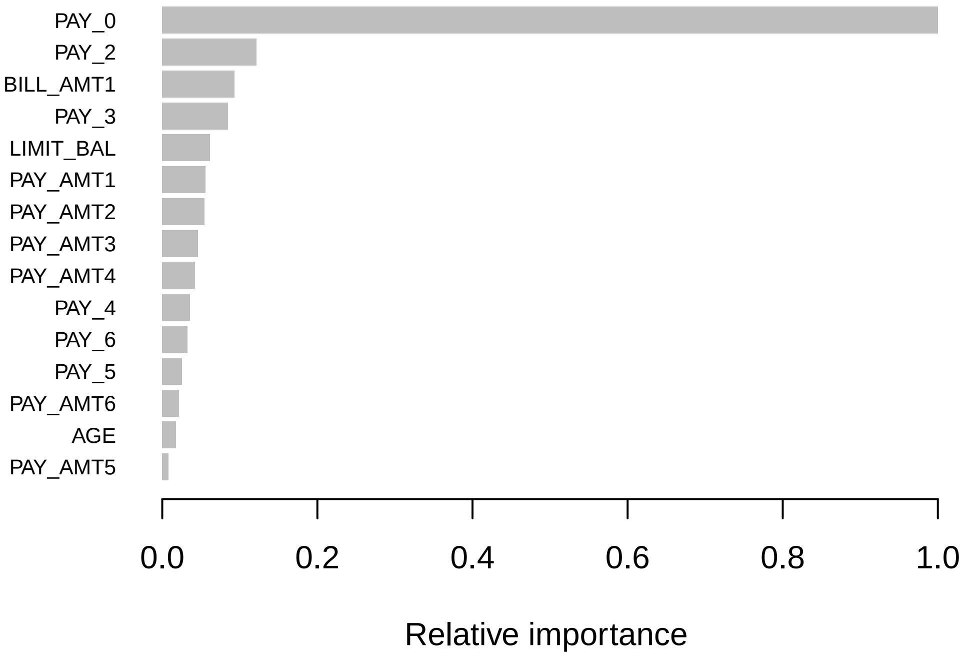
Model	AUC
Naive Bayes	0.6635631
K-Nearest Neighbors k=5	0.5498315
SVM - Support Vector Machine-Sigmoid Kernel	0.6202597
SVM - Support Vector Machine-Linear Kernel	0.6202597
Random Forest	0.6612221
GBM - Generalized Boosted Regression	0.7825340

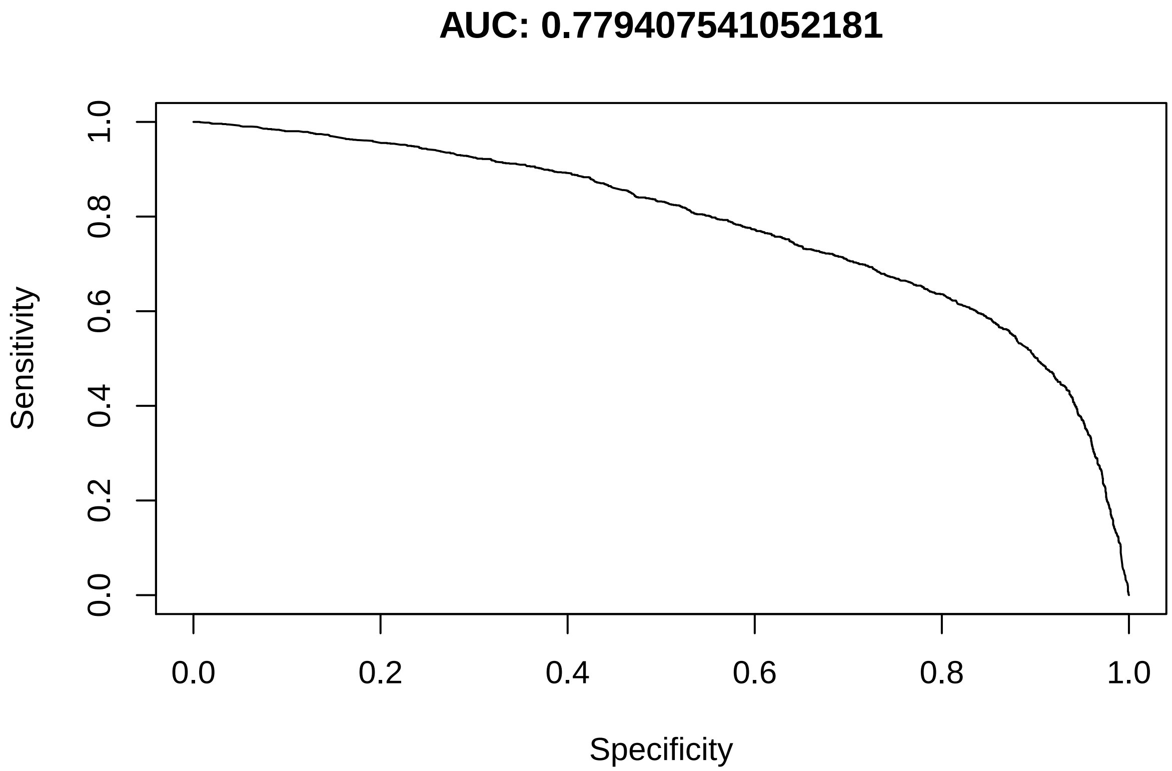
	var	rel.inf
PAY_0	PAY_0	64.3703546
PAY_2	PAY_2	8.4530186
PAY_3	PAY_3	4.8159292
PAY_6	PAY_6	2.7157006
BILL_AMT1	BILL_AMT1	2.4675059
PAY_5	PAY_5	2.4308219
PAY_AMT1	PAY_AMT1	2.3875695
LIMIT_BAL	LIMIT_BAL	2.2337888
PAY_AMT3	PAY_AMT3	2.2067436
PAY_4	PAY_4	2.0820112
PAY_AMT2	PAY_AMT2	2.0747112
PAY_AMT4	PAY_AMT4	1.6107953
PAY_AMT6	PAY_AMT6	0.7214530
AGE	AGE	0.5024765
PAY_AMT5	PAY_AMT5	0.4853345
EDUCATION	EDUCATION	0.3420405
SEX	SEX	0.0997451

4.7 XGBoost

XGBoost is the last model fitted to dataset

```
## [1]  test-aucpr:0.520352 cv-aucpr:0.511479
## Multiple eval metrics are present. Will use cv_aucpr for early stopping.
## Will train until cv_aucpr hasn’t improved in 40 rounds.
##
## [21] test-aucpr:0.544504 cv-aucpr:0.543804
## [41] test-aucpr:0.546065 cv-aucpr:0.551389
## [61] test-aucpr:0.543918 cv-aucpr:0.548323
## Stopping. Best iteration:
## [35] test-aucpr:0.544956 cv-aucpr:0.552797
```





Model	AUC
Naive Bayes	0.6635631
K-Nearest Neighbors k=5	0.5498315
SVM - Support Vector Machine-Sigmoid Kernel	0.6202597
SVM - Support Vector Machine-Linear Kernel	0.6202597
Random Forest	0.6612221
GBM - Generalized Boosted Regression	0.7825340
XGBoost	0.7794075

Feature	Gain	Cover	Frequency	Importance
PAY_0	0.5880419	0.1650563	0.0921305	0.5880419
PAY_2	0.0712929	0.0500831	0.0403071	0.0712929
BILL_AMT1	0.0549190	0.1661581	0.2015355	0.0549190
PAY_3	0.0498837	0.0572796	0.0499040	0.0498837
LIMIT_BAL	0.0362070	0.1098525	0.1113244	0.0362070
PAY_AMT1	0.0327454	0.0602381	0.0633397	0.0327454
PAY_AMT2	0.0320209	0.0754331	0.0595010	0.0320209
PAY_AMT3	0.0271323	0.0652457	0.0595010	0.0271323
PAY_AMT4	0.0247414	0.0451413	0.0710173	0.0247414
PAY_4	0.0210150	0.0329665	0.0345489	0.0210150
PAY_6	0.0191125	0.0483081	0.0403071	0.0191125
PAY_5	0.0149958	0.0250086	0.0249520	0.0149958
PAY_AMT6	0.0127862	0.0355057	0.0633397	0.0127862
AGE	0.0104809	0.0440844	0.0614203	0.0104809
PAY_AMT5	0.0046252	0.0196387	0.0268714	0.0046252

5 Results

This is the summary results for all the models built, trained and validated.

Model	AUC
Naive Bayes	0.6635631
K-Nearest Neighbors k=5	0.5498315
SVM - Support Vector Machine-Sigmoid Kernel	0.6202597
SVM - Support Vector Machine-Linear Kernel	0.6202597
Random Forest	0.6612221
GBM - Generalized Boosted Regression	0.7825340
XGBoost	0.7794075

6 Conclusion

The ensemble methods once again confirm themselves as among the best models out there. In this task, a GBM model can achieve a very good AUC of **0.78** and the others ensembe methods are very close to it. As the features importance plots and table show, there are few predictors like payment in recent months are particularly useful for identifying default.

7 Limitation and Future Scope

This analysis was restricted in using core machine learning algorithms. This can further be extended by applying deep learning techniques to see some model improvement.

8 Appendix

8.1 1a - Code used in this report

```
# Credit Card Default Detection
# Author: Anubhav Gupta
# ----------------------------------------------------------
# Description: This is the final assignment
# for the Harvard Data Science Professional Program
# In this capstone project, we
# have to choose a dataset and we have to analyze it and
# perform our machine learning tasks in complete autonomy
# without external help.


# Install all required libraries if it is not present

if(!require(tidyverse)) install.packages("tidyverse")
if(!require(kableExtra)) install.packages("kableExtra")
if(!require(tidyr)) install.packages("tidyr")
if(!require(tidyverse)) install.packages("tidyverse")
if(!require(stringr)) install.packages("stringr")
if(!require(ggplot2)) install.packages("ggplot2")
if(!require(gbm)) install.packages("gbm")
if(!require(dplyr)) install.packages("dplyr")
if(!require(caret)) install.packages("caret")
if(!require(xgboost)) install.packages("xgboost")
if(!require(e1071)) install.packages("e1071")
if(!require(class)) install.packages("class")
if(!require(ROCR)) install.packages("ROCR")
if(!require(randomForest)) install.packages("randomForest")
if(!require(PRROC)) install.packages("PRROC")
if(!require(reshape2)) install.packages("reshape2")
if(!require(lightgbm)) install.packages("lightgbm")

# Loading all required libraries

library(dplyr)
library(tidyverse)
library(kableExtra)
library(tidyr)
library(ggplot2)
library(gbm)
library(caret)
library(xgboost)
library(e1071)
library(class)
library(lightgbm)
library(ROCR)
library(randomForest)
library(PRROC)
library(reshape2)
library(lightgbm)
```

```
## Loading the dataset
setwd("C:/Users/TRUPTI/Documents/capstone") #Set your own working directory
creditcard <- read.csv("default of credit card clients.csv")
names(creditcard) <- creditcard[1,]
creditcard=creditcard[-1,]
creditcard <- data.frame(lapply(creditcard, function(x) as.integer(as.character(x))))
names(creditcard)[names(creditcard) == "default.payment.next.month"] <- "class"
creditcard$SEX=as.factor(creditcard$SEX)
creditcard$EDUCATION=as.factor(creditcard$EDUCATION)
creditcard$MARRIAGE=as.factor(creditcard$MARRIAGE)

#Data checks

data.frame("Length" = nrow(creditcard), "Columns" = ncol(creditcard))

imbalanced <- data.frame(creditcard)
imbalanced$class = ifelse(imbalanced$class == 0, 'Non-default', 'Default') %>% as.factor()

# Visualize the proportion between classes

imbalanced %>%
  ggplot(aes(class)) +
  theme_minimal()  +
  geom_bar() +
  scale_x_discrete() +
  scale_y_continuous(labels = scales::comma) +
  labs(title = "Proportions between Default and Non-default Transactions",
       x = "Class",
       y = "Frequency")

# Find missing values

data.frame(sapply(creditcard, function(x) sum(is.na(x))) )

# Credit Age Distribution

creditcard[creditcard$class == 1,] %>%
  ggplot(aes(AGE)) +
  theme_minimal()  +
  geom_histogram(binwidth = 5) +
  labs(title = "Default Transactions agewise distribution",
       x = "Age(in Years)",
       y = "Frequency")

# Credit Amount Distribution

creditcard[creditcard$class == 1,] %>%
  ggplot(aes(LIMIT_BAL/1000)) +
  theme_minimal()  +
  geom_histogram(binwidth = 40) +
  labs(title = "Default Transactions Amounts Distributions",
       x = "Amount in dollars('000)",
       y = "Frequency")
```

```
# Default distribution by SEX

imbalanced$SEX = ifelse(imbalanced$SEX == 1, 'MALE', 'FEMALE') %>% as.factor()

imbalanced[imbalanced$class == 'Default',] %>%
  ggplot(aes(SEX)) +
  theme_minimal()  +
  geom_bar() +
  scale_x_discrete() +
  scale_y_continuous(labels = scales::comma) +
  labs(title = "Default Transactions by sex",
       x = "SEX",
       y = "Frequency")

data.frame(creditcard[creditcard$class == 1,] %>%
  group_by(SEX) %>%
  summarise(count = n()) )

# Default distribution by Education

creditcard[creditcard$class == 1,] %>%
  ggplot(aes(EDUCATION)) +
  theme_minimal()  +
  geom_bar() +
  scale_x_discrete() +
  scale_y_continuous(labels = scales::comma) +
  labs(title = "Default Transactions by Education",
       x = "EDUCATION",
       y = "Frequency")

data.frame(creditcard[creditcard$class == 1,] %>%
             group_by(EDUCATION) %>%
             summarise(count = n()) )

# Default distribution by Marriage

creditcard[creditcard$class == 1,] %>%
  ggplot(aes(MARRIAGE)) +
  theme_minimal()  +
  geom_bar() +
  scale_x_discrete() +
  scale_y_continuous(labels = scales::comma) +
  labs(title = "Default Transactions by Marriage",
       x = "MARRIAGE",
       y = "Frequency")

data.frame(creditcard[creditcard$class == 1,] %>%
             group_by(MARRIAGE) %>%
             summarise(count = n()) )


# Get lower triangle of the correlation matrix

get_lower_tri<-function(cormat){
  cormat[upper.tri(cormat)] <- NA
  return(cormat)
}

# Get upper triangle of the correlation matrix

get_upper_tri <- function(cormat){
  cormat[lower.tri(cormat)]<- NA
  return(cormat)
}

reorder_cormat <- function(cormat){
  # Use correlation between variables as distance
  dd <- as.dist((1-cormat)/2)
  hc <- hclust(dd)
  cormat <-cormat[hc$order, hc$order]
}
creditcard1=creditcard %>% select(-c(SEX,EDUCATION,MARRIAGE,class))
corr_matrix <- round(cor(creditcard1),2)
corr_matrix <- reorder_cormat(corr_matrix)

upper_tri <- get_upper_tri(corr_matrix)

melted_corr_matrix <- melt(upper_tri, na.rm = TRUE)

ggplot(melted_corr_matrix, aes(Var2, Var1, fill = value)) +
  geom_tile(color = "white") +
  scale_fill_gradient2(low = "blue", high = "red", mid = "white",
                       midpoint = 0, limit = c(-1,1), space = "Lab",
                       name="Pearson\nCorrelation") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, vjust = 1,
                                   size = 9, hjust = 1),
        axis.text.y = element_text(size = 9),
        axis.title.x = element_blank(),
        axis.title.y = element_blank(),
        panel.grid.major = element_blank(),
        panel.border = element_blank(),
        panel.background = element_blank(),
        axis.ticks = element_blank()) +
  coord_fixed()

# Set seed for reproducibility

set.seed(1234)

# Remove unnecessary columns from the dataset

creditcard$class <- as.factor(creditcard$class)
creditcard <- creditcard %>% select(-c(ID,BILL_AMT2,BILL_AMT3,BILL_AMT4,BILL_AMT5,BILL_AMT6,MARRIAGE))

# Split the dataset into train, test dataset and cv

train_index <- createDataPartition(
  y = creditcard$class,
  p = .6,
  list = F
)

train <- creditcard[train_index,]

test_cv <- creditcard[-train_index,]

test_index <- createDataPartition(
  y = test_cv$class,
  p = .5,
  list = F)

test <- test_cv[test_index,]
cv <- test_cv[-test_index,]

rm(train_index, test_index, test_cv)


###############################################
#MODELLING
###############################################

# Set seed 123 for reproducibility

set.seed(123)

# Build the model with Class as target and all other variables
# as predictors

naive_model <- naiveBayes(class ~ ., data = train, laplace=1)

# Predict

predictions <- predict(naive_model, newdata=test)

# Compute the AUC and AUCPR for the Naive Model

pred <- prediction(as.numeric(predictions) , test$class)

auc_val_naive <- performance(pred, "auc")

auc_plot_naive <- performance(pred, ’sens’, ’spec’)

# Make the relative plot


plot(auc_plot_naive, main=paste("AUC:", auc_val_naive@y.values[[1]]))


# Adding the respective metrics to the results dataset

results <-  data.frame(
  Model = "Naive Bayes",
  AUC = auc_val_naive@y.values[[1]]
)
# Show results on a table

results %>%
  kable() %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed",           "responsive"),
                position = "center",
                font_size = 10,
                full_width = FALSE)

# Set seed 123 for reproducibility

set.seed(123)

# Build a KNN Model with Class as Target and all other
# variables as predictors. k is set to 5

knn_model <- knn(train[,-30], test[,-30], train$class, k=5, prob = TRUE)

# Compute the AUC and AUCPR for the KNN Model

pred <- prediction(
  as.numeric(as.character(knn_model)),  as.numeric(as.character(test$class))
)

auc_val_knn <- performance(pred, "auc")

auc_plot_knn <- performance(pred, 'sens', 'spec')


# Make the relative plot


plot(auc_plot_knn, main=paste("AUC:", auc_val_knn@y.values[[1]]))

# Adding the respective metrics to the results dataset

results <- results %>% add_row(
  Model = "K-Nearest Neighbors k=5",
  AUC = auc_val_knn@y.values[[1]]

)

# Show results on a table

results %>%
  kable() %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed",           "responsive"),
                position = "center",
                font_size = 10,
                full_width = FALSE)
```

```
# Set seed 123 for reproducibility

set.seed(123)

# Build a SVM Model with Class as Target and all other
# variables as predictors. The kernel is set to sigmoid

svm_model <- svm(class ~ ., data = train, kernel='sigmoid')

# Make predictions based on this model

predictions <- predict(svm_model, newdata=test)

# Compute AUC and AUCPR

pred <- prediction(
  as.numeric(as.character(predictions)), as.numeric(as.character(test$class))
)

auc_val_svm <- performance(pred, "auc")

auc_plot_svm <- performance(pred, 'sens', 'spec')

# Make the relative plot


plot(auc_plot_svm, main=paste("AUC:", auc_val_svm@y.values[[1]]))


# Adding the respective metrics to the results dataset

results <- results %>% add_row(
  Model = "SVM - Support Vector Machine-Sigmoid Kernel",
  AUC = auc_val_svm@y.values[[1]]
  )

# Show results on a table

results %>%
  kable() %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed",           "responsive"),
                position = "center",
                font_size = 10,
                full_width = FALSE)

# Set seed 123 for reproducibility

set.seed(123)

# Build a SVM Model with Class as Target and all other
# variables as predictors. The kernel is set to linear

svm_model_lk <- svm(class ~ ., data = train, kernel='linear')
# Make predictions based on this model
```

```
predictions <- predict(svm_model_lk, newdata=test)

# Compute AUC and AUCPR

pred <- prediction(
  as.numeric(as.character(predictions)), as.numeric(as.character(test$class))
)

auc_val_svm_lk <- performance(pred, "auc")

auc_plot_svm_lk <- performance(pred, 'sens', 'spec')


# Make the relative plot


plot(auc_plot_svm_lk, main=paste("AUC:", auc_val_svm_lk@y.values[[1]]))


# Adding the respective metrics to the results dataset

results <- results %>% add_row(
  Model = "SVM - Support Vector Machine-Linear Kernel",
  AUC = auc_val_svm@y.values[[1]]
)

# Show results on a table

results %>%
  kable() %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed",           "responsive"),
                position = "center",
                font_size = 10,
                full_width = FALSE)

# Set seed 123 for reproducibility

set.seed(123)

# Build a Random Forest Model with Class as Target and all other
# variables as predictors. The number of trees is set to 500

rf_model <- randomForest(class ~ ., data = train, ntree = 500)

# Get the feature importance

feature_imp_rf <- data.frame(importance(rf_model))

# Make predictions based on this model

predictions <- predict(rf_model, newdata=test)

# Compute the AUC and AUPCR
```

```
pred <- prediction(
  as.numeric(as.character(predictions)), as.numeric(as.character(test$class))
)

auc_val_rf <- performance(pred, "auc")

auc_plot_rf <- performance(pred, 'sens', 'spec')

# make the relative plot

plot(auc_plot_rf, main=paste("AUC:", auc_val_rf@y.values[[1]]))


# Adding the respective metrics to the results dataset

results <- results %>% add_row(
  Model = "Random Forest",
  AUC = auc_val_rf@y.values[[1]]
  )

# Show results on a table

results %>%
  kable() %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed",           "responsive"),
                position = "center",
                font_size = 10,
                full_width = FALSE)

# Show feature importance on a table

feature_imp_rf %>%
  kable() %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed",           "responsive"),
                position = "center",
                font_size = 10,
                full_width = FALSE)

# Set seed 123 for reproducibility

set.seed(123)

# Build a GBM Model with Class as Target and all other
# variables as predictors. Distribution is bernoully,
# number of tree is 500

gbm_model <- gbm(as.character(class) ~ .,
                 distribution = "bernoulli",
                 data = rbind(train, test),
                 n.trees = 500,
                 interaction.depth = 3,
                 n.minobsinnode = 100,
                 shrinkage = 0.01,
                 train.fraction = 0.7,
)

# Determine the best iteration based on test data

best_iter = gbm.perf(gbm_model, method = "test")

# Make predictions based on this model

predictions = predict.gbm(
  gbm_model,
  newdata = test,
  n.trees = best_iter,
  type="response"
)

# Get feature importance

feature_imp_gbm = summary(gbm_model, n.trees = best_iter)

# Compute the AUC and AUPCR

pred <- prediction(
  as.numeric(as.character(predictions)),as.numeric(as.character(test$class))
)

auc_val_gbm <- performance(pred, "auc")

auc_plot_gbm <- performance(pred, 'sens', 'spec')


# Make the relative plot


plot(auc_plot_gbm, main=paste("AUC:", auc_val_gbm@y.values[[1]]))


# Adding the respective metrics to the results dataset

results <- results %>% add_row(
  Model = "GBM - Generalized Boosted Regression",
  AUC = auc_val_gbm@y.values[[1]]
 )

# Show results on a table

results %>%
  kable() %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed",           "responsive"),
                position = "center",
                font_size = 10,
                full_width = FALSE)

# Show feature importance on a table
```

```
feature_imp_gbm %>%
  kable() %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed",           "responsive"),
                position = "center",
                font_size = 10,
                full_width = FALSE)

# Set seed 123 for reproducibility

set.seed(123)

# Prepare the training dataset
train=train %>% select(-c(SEX,EDUCATION))
test=test %>% select(-c(SEX,EDUCATION))
cv=cv %>% select(-c(SEX,EDUCATION))

xgb_train <- xgb.DMatrix(
  as.matrix(train[, colnames(train) != "class"]),
  label = as.numeric(as.character(train$class))
)

# Prepare the test dataset

xgb_test <- xgb.DMatrix(
  as.matrix(test[, colnames(test) != "class"]),
  label = as.numeric(as.character(test$class))
)

# Prepare the cv dataset

xgb_cv <- xgb.DMatrix(
  as.matrix(cv[, colnames(cv) != "class"]),
  label = as.numeric(as.character(cv$class))
)

# Prepare the parameters list.

xgb_params <- list(
  objective = "binary:logistic",
  eta = 0.1,
  max.depth = 3,
  nthread = 6,
  eval_metric = "aucpr"
)

# Train the XGBoost Model

xgb_model <- xgb.train(
  data = xgb_train,
  params = xgb_params,
  watchlist = list(test = xgb_test, cv = xgb_cv),
  nrounds = 500,
  early_stopping_rounds = 40,
  print_every_n = 20
)

# Get feature importance

feature_imp_xgb <- xgb.importance(colnames(train), model = xgb_model)

xgb.plot.importance(feature_imp_xgb, rel_to_first = TRUE, xlab = "Relative importance")

# Make predictions based on this model

predictions = predict(
  xgb_model,
  newdata = as.matrix(test[, colnames(test) != "class"]),
  ntreelimit = xgb_model$bestInd
)

# Compute the AUC and AUPCR

pred <- prediction(
  as.numeric(as.character(predictions)), as.numeric(as.character(test$class))
)

auc_val_xgb <- performance(pred, "auc")

auc_plot_xgb <- performance(pred, 'sens', 'spec')


# Make the relative plot

plot(auc_plot_xgb, main=paste("AUC:", auc_val_xgb@y.values[[1]]))


# Adding the respective metrics to the results dataset

results <- results %>% add_row(
  Model = "XGBoost",
  AUC = auc_val_xgb@y.values[[1]]
  )

# Show results on a table

results %>%
  kable() %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed",           "responsive"),
                position = "center",
                font_size = 10,
                full_width = FALSE)

# Show feature importance on a table

feature_imp_xgb %>%
  kable() %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed",           "responsive"),
                position = "center",
                font_size = 10,
                full_width = FALSE)
```

8.2 1b- Enviroment

```
print("Operating System:")
```

```
## [1] "Operating System:"
```

```
version
```

```
##                _
## platform       x86_64-w64-mingw32
## arch           x86_64
## os             mingw32
## system         x86_64, mingw32
## status
## major          4
## minor          0.4
## year           2021
## month          02
## day            15
## svn rev        80002
## language       R
## version.string R version 4.0.4 (2021-02-15)
## nickname       Lost Library Book
```

```
print("All installed packages")
```

```
## [1] "All installed packages"
```

```
installed.packages()
```

```
##              Package        LibPath                                       Version      Priority
## askpass      "askpass"      "C:/Users/TRUPTI/Documents/R/win-library/4.0" "1.1"        NA
## assertthat   "assertthat"   "C:/Users/TRUPTI/Documents/R/win-library/4.0" "0.2.1"      NA
## backports    "backports"    "C:/Users/TRUPTI/Documents/R/win-library/4.0" "1.2.1"      NA
## base64enc    "base64enc"    "C:/Users/TRUPTI/Documents/R/win-library/4.0" "0.1-3"      NA
## BH           "BH"           "C:/Users/TRUPTI/Documents/R/win-library/4.0" "1.75.0-0"   NA
## bitops       "bitops"       "C:/Users/TRUPTI/Documents/R/win-library/4.0" "1.0-6"      NA
## blob         "blob"         "C:/Users/TRUPTI/Documents/R/win-library/4.0" "1.2.1"      NA
## bookdown     "bookdown"     "C:/Users/TRUPTI/Documents/R/win-library/4.0" "0.21"       NA
## brew         "brew"         "C:/Users/TRUPTI/Documents/R/win-library/4.0" "1.0-6"      NA
## brio         "brio"         "C:/Users/TRUPTI/Documents/R/win-library/4.0" "1.1.1"      NA
## broom        "broom"        "C:/Users/TRUPTI/Documents/R/win-library/4.0" "0.7.5"      NA
## cachem       "cachem"       "C:/Users/TRUPTI/Documents/R/win-library/4.0" "1.0.4"      NA
## callr        "callr"        "C:/Users/TRUPTI/Documents/R/win-library/4.0" "3.6.0"      NA
## caret        "caret"        "C:/Users/TRUPTI/Documents/R/win-library/4.0" "6.0-86"     NA
## caTools      "caTools"      "C:/Users/TRUPTI/Documents/R/win-library/4.0" "1.18.2"     NA
## cellranger   "cellranger"   "C:/Users/TRUPTI/Documents/R/win-library/4.0" "1.1.0"      NA
## cli          "cli"          "C:/Users/TRUPTI/Documents/R/win-library/4.0" "2.3.1"      NA
## clipr        "clipr"        "C:/Users/TRUPTI/Documents/R/win-library/4.0" "0.7.1"      NA
## colorspace   "colorspace"   "C:/Users/TRUPTI/Documents/R/win-library/4.0" "2.0-0"      NA
## commonmark   "commonmark"   "C:/Users/TRUPTI/Documents/R/win-library/4.0" "1.7"        NA
## covr         "covr"         "C:/Users/TRUPTI/Documents/R/win-library/4.0" "3.5.1"      NA
## cpp11        "cpp11"        "C:/Users/TRUPTI/Documents/R/win-library/4.0" "0.2.7"      NA
## crayon       "crayon"       "C:/Users/TRUPTI/Documents/R/win-library/4.0" "1.4.1"      NA
## credentials  "credentials"  "C:/Users/TRUPTI/Documents/R/win-library/4.0" "1.3.0"      NA
## crosstalk    "crosstalk"    "C:/Users/TRUPTI/Documents/R/win-library/4.0" "1.1.1"      NA
## curl         "curl"         "C:/Users/TRUPTI/Documents/R/win-library/4.0" "4.3"        NA
## data.table   "data.table"   "C:/Users/TRUPTI/Documents/R/win-library/4.0" "1.14.0"     NA
## DBI          "DBI"          "C:/Users/TRUPTI/Documents/R/win-library/4.0" "1.1.1"      NA
## dbplyr       "dbplyr"       "C:/Users/TRUPTI/Documents/R/win-library/4.0" "2.1.0"      NA
## desc         "desc"         "C:/Users/TRUPTI/Documents/R/win-library/4.0" "1.3.0"      NA
## devtools     "devtools"     "C:/Users/TRUPTI/Documents/R/win-library/4.0" "2.3.2"      NA
## diffobj      "diffobj"      "C:/Users/TRUPTI/Documents/R/win-library/4.0" "0.3.4"      NA
## digest       "digest"       "C:/Users/TRUPTI/Documents/R/win-library/4.0" "0.6.27"     NA
## dplyr        "dplyr"        "C:/Users/TRUPTI/Documents/R/win-library/4.0" "1.0.5"      NA
## DT           "DT"           "C:/Users/TRUPTI/Documents/R/win-library/4.0" "0.17"       NA
## e1071        "e1071"        "C:/Users/TRUPTI/Documents/R/win-library/4.0" "1.7-6"      NA
## ellipsis     "ellipsis"     "C:/Users/TRUPTI/Documents/R/win-library/4.0" "0.3.1"      NA
## evaluate     "evaluate"     "C:/Users/TRUPTI/Documents/R/win-library/4.0" "0.14"       NA
## fansi        "fansi"        "C:/Users/TRUPTI/Documents/R/win-library/4.0" "0.4.2"      NA
## farver       "farver"       "C:/Users/TRUPTI/Documents/R/win-library/4.0" "2.1.0"      NA
## fastmap      "fastmap"      "C:/Users/TRUPTI/Documents/R/win-library/4.0" "1.1.0"      NA
## float        "float"        "C:/Users/TRUPTI/Documents/R/win-library/4.0" "0.2-4"      NA
## forcats      "forcats"      "C:/Users/TRUPTI/Documents/R/win-library/4.0" "0.5.1"      NA
## foreach      "foreach"      "C:/Users/TRUPTI/Documents/R/win-library/4.0" "1.5.1"      NA
## fs           "fs"           "C:/Users/TRUPTI/Documents/R/win-library/4.0" "1.5.0"      NA
## gbm          "gbm"          "C:/Users/TRUPTI/Documents/R/win-library/4.0" "2.1.8"      NA
## gdata        "gdata"        "C:/Users/TRUPTI/Documents/R/win-library/4.0" "2.18.0"     NA
## generics     "generics"     "C:/Users/TRUPTI/Documents/R/win-library/4.0" "0.1.0"      NA
## gert         "gert"         "C:/Users/TRUPTI/Documents/R/win-library/4.0" "1.3.0"      NA
## ggplot2      "ggplot2"      "C:/Users/TRUPTI/Documents/R/win-library/4.0" "3.3.3"      NA
## gh           "gh"           "C:/Users/TRUPTI/Documents/R/win-library/4.0" "1.2.1"      NA
## gitcreds     "gitcreds"     "C:/Users/TRUPTI/Documents/R/win-library/4.0" "0.1.1"      NA
## glue         "glue"         "C:/Users/TRUPTI/Documents/R/win-library/4.0" "1.4.2"      NA
## gower        "gower"        "C:/Users/TRUPTI/Documents/R/win-library/4.0" "0.2.2"      NA
## gplots       "gplots"       "C:/Users/TRUPTI/Documents/R/win-library/4.0" "3.1.1"      NA
## gtable       "gtable"       "C:/Users/TRUPTI/Documents/R/win-library/4.0" "0.3.0"      NA
## gtools       "gtools"       "C:/Users/TRUPTI/Documents/R/win-library/4.0" "3.8.2"      NA
## haven        "haven"        "C:/Users/TRUPTI/Documents/R/win-library/4.0" "2.3.1"      NA
## highr        "highr"        "C:/Users/TRUPTI/Documents/R/win-library/4.0" "0.8"        NA
## hms          "hms"          "C:/Users/TRUPTI/Documents/R/win-library/4.0" "1.0.0"      NA
## htmltools    "htmltools"    "C:/Users/TRUPTI/Documents/R/win-library/4.0" "0.5.1.1"    NA
## htmlwidgets  "htmlwidgets"  "C:/Users/TRUPTI/Documents/R/win-library/4.0" "1.5.3"      NA
##              Depends
## askpass      NA
## assertthat   NA
## backports    "R (>= 3.0.0)"
## base64enc    "R (>= 2.9.0)"
## BH           NA
## bitops       NA
## blob         NA
## bookdown     NA
## brew         NA
## brio         NA
## broom        "R (>= 3.1)"
## cachem       NA
## callr        NA
## caret        "R (>= 3.2.0), lattice (>= 0.20), ggplot2"
## caTools      "R (>= 3.6.0)"
## cellranger   "R (>= 3.0.0)"
## cli          "R (>= 2.10)"
## clipr        NA
## colorspace   "R (>= 3.0.0), methods"
## commonmark   NA
## covr         "R (>= 3.1.0), methods"
## cpp11        NA
## crayon       NA
## credentials  NA
## crosstalk    NA
## curl         "R (>= 3.0.0)"
## data.table   "R (>= 3.1.0)"
## DBI          "methods, R (>= 3.0.0)"
## dbplyr       "R (>= 3.1)"
## desc         "R (>= 3.1.0)"
## devtools     "R (>= 3.0.2), usethis (>= 1.6.3)"
## diffobj      "R (>= 3.1.0)"
## digest       "R (>= 3.3.0)"
## dplyr        "R (>= 3.3.0)"
## DT           NA
## e1071        NA
## ellipsis     "R (>= 3.2)"
## evaluate     "R (>= 3.0.2)"
## fansi        "R (>= 3.1.0)"
## farver       NA
## fastmap      NA
## float        "R (>= 3.6.0), methods"
## forcats      "R (>= 3.2)"
## foreach      "R (>= 2.5.0)"
## fs           "R (>= 3.1)"
## gbm          "R (>= 2.9.0)"
## gdata        "R (>= 2.3.0)"
## generics     "R (>= 3.1)"
## gert         NA
## ggplot2      "R (>= 3.2)"
## gh           NA
## gitcreds     NA
## glue         "R (>= 3.2)"
## gower        NA
## gplots       "R (>= 3.0)"
## gtable       "R (>= 3.0)"
## gtools       "methods, stats, utils"
## haven        "R (>= 3.2)"
## highr        "R (>= 3.2.3)"
## hms          NA
## htmltools    "R (>= 2.14.1)"
## htmlwidgets  NA
##              Imports
## askpass      "sys (>= 2.1)"
## assertthat   "tools"
## backports    NA
## base64enc    NA
## BH           NA
## bitops       NA
## blob         "methods, rlang, vctrs (>= 0.2.1)"
## bookdown     "htmltools (>= 0.3.6), knitr (>= 1.22), rmarkdown (>= 2.4),\nxfun (>= 0.13), tinytex (>
## brew         NA
## brio         NA
## broom        "backports, dplyr (>= 1.0.0), ellipsis, generics (>= 0.0.2),\nglue, methods, purrr, rlan
## cachem       "rlang, fastmap"
## callr        "processx (>= 3.5.0), R6, utils"
## caret        "foreach, methods, plyr, ModelMetrics (>= 1.2.2.2), nlme,\nreshape2, stats, stats4, util
## caTools      "bitops"
## cellranger   "rematch, tibble"
## cli          "assertthat, glue, methods, utils"
## clipr        "utils"
## colorspace   "graphics, grDevices, stats"
## commonmark   NA
## covr         "digest, stats, utils, jsonlite, rex, httr, crayon, withr (>=\n1.0.2), yaml"
## cpp11        NA
## crayon       "grDevices, methods, utils"
## credentials  "openssl (>= 1.3), sys (>= 2.1), curl, jsonlite, askpass"
## crosstalk    "htmltools (>= 0.3.6), jsonlite, lazyeval, R6"
## curl         NA
## data.table   "methods"
## DBI          NA
## dbplyr       "assertthat (>= 0.2.0), blob (>= 1.2.0), DBI (>= 1.0.0), dplyr\n(>= 1.0.3), ellipsis, gl
## desc         "utils, R6, crayon, rprojroot"
## devtools     "callr (>= 3.4.4), cli (>= 2.0.2), covr (>= 3.5.1), desc (>=\n1.2.0), DT (>= 0.15), elli
## diffobj      "crayon (>= 1.3.2), tools, methods, utils, stats"
## digest       "utils"
## dplyr        "ellipsis, generics, glue (>= 1.3.2), lifecycle (>= 1.0.0),\nmagrittr (>= 1.5), methods
## DT           "htmltools (>= 0.3.6), htmlwidgets (>= 1.3), jsonlite (>=\n0.9.16), magrittr, crosstalk
## e1071        "graphics, grDevices, class, stats, methods, utils, proxy"
## ellipsis     "rlang (>= 0.3.0)"
## evaluate     "methods"
## fansi        NA
## farver       NA
## fastmap      NA
## float        "utils, tools"
## forcats      "ellipsis, magrittr, rlang, tibble"
## foreach      "codetools, utils, iterators"
## fs           "methods"
## gbm          "lattice, parallel, survival"
## gdata        "gtools, stats, methods, utils"
## generics     "methods"
## gert         "askpass, credentials (>= 1.2.1), openssl (>= 1.4.1),\nrstudioapi (>= 0.11), sys, zip (
## ggplot2      "digest, glue, grDevices, grid, gtable (>= 0.1.1), isoband,\nMASS, mgcv, rlang (>= 0.3.
## gh           "cli (>= 2.0.1), gitcreds, httr (>= 1.2), ini, jsonlite"
## gitcreds     NA
## glue         "methods"
## gower        NA
## gplots       "gtools, stats, caTools, KernSmooth"
## gtable       "grid"
## gtools       NA
## haven        "forcats (>= 0.2.0), hms, methods, Rcpp (>= 0.11.4), readr (>=\n0.1.0), rlang (>= 0.4.0
## highr        NA
## hms          "ellipsis, lifecycle, methods, pkgconfig, rlang, vctrs (>=\n0.2.1)"
## htmltools    "utils, digest, grDevices, base64enc, rlang"
## htmlwidgets  "grDevices, htmltools (>= 0.3), jsonlite (>= 0.9.16), yaml"
##              LinkingTo
## askpass      NA
## assertthat   NA
## backports    NA
## base64enc    NA
## BH           NA
## bitops       NA
## blob         NA
## bookdown     NA
## brew         NA
## brio         NA
## broom        NA
## cachem       NA
## callr        NA
## caret        NA
## caTools      NA
## cellranger   NA
## cli          NA
## clipr        NA
## colorspace   NA
## commonmark   NA
## covr         NA
## cpp11        NA
## crayon       NA
## credentials  NA
## crosstalk    NA
## curl         NA
## data.table   NA
## DBI          NA
## dbplyr       NA
## desc         NA
## devtools     NA
## diffobj      NA
## digest       NA
## dplyr        NA
## DT           NA
## e1071        NA
## ellipsis     NA
## evaluate     NA
## fansi        NA
## farver       NA
## fastmap      NA
## float        NA
## forcats      NA
## foreach      NA
## fs           NA
## gbm          NA
## gdata        NA
## generics     NA
## gert         NA
## ggplot2      NA
## gh           NA
## gitcreds     NA
## glue         NA
## gower        NA
## gplots       NA
## gtable       NA
## gtools       NA
## haven        "Rcpp"
## highr        NA
## hms          NA
## htmltools    NA
## htmlwidgets  NA
##              Suggests
## askpass      "testthat"
## assertthat   "testthat, covr"
## backports    NA
## base64enc    NA
## BH           NA
## bitops       NA
## blob         "covr, crayon, pillar (>= 1.2.1), testthat"
## bookdown     "htmlwidgets, rstudioapi, miniUI, rsconnect (>= 0.4.3), servr\n(>= 0.13), shiny, testit
## brew         NA
## brio         "testthat (>= 2.1.0), covr"
## broom        "AER, akima, AUC, bbmle, betareg, biglm, binGroup, boot,\nbtergm, car, caret, cluster,
## cachem       "testthat"
## callr        "cli, covr, crayon, fansi, pingr, ps, rprojroot, spelling,\ntestthat, tibble, withr (>=
## caret        "BradleyTerry2, e1071, earth (>= 2.2-3), fastICA, gam (>=\n1.15), ipred, kernlab, knitr
## caTools      "MASS, rpart"
## cellranger   "covr, testthat (>= 1.0.0), knitr, rmarkdown"
## cli          "callr, covr, grDevices, htmlwidgets, knitr, mockery, ps (>=\n1.3.4.9000), rmarkdown, r
## clipr        "covr, knitr, rmarkdown, rstudioapi (>= 0.5), testthat (>=\n2.0.0)"
## colorspace   "datasets, utils, KernSmooth, MASS, kernlab, mvtnorm, vcd,\ntcltk, shiny, shinyjs, ggpl
## commonmark   "curl, testthat, xml2"
## covr         "R6, curl, knitr, rmarkdown, htmltools, DT (>= 0.2), testthat,\nrlang, rstudioapi (>= 0
## cpp11        "bench, brio, callr, cli, covr, decor, desc, ggplot2, glue,\nknitr, lobstr, mockery, pr
## crayon       "mockery, rstudioapi, testthat, withr"
## credentials  "testthat, knitr, rmarkdown"
## crosstalk    "shiny, ggplot2, testthat (>= 2.1.0)"
## curl         "spelling, testthat (>= 1.0.0), knitr, jsonlite, rmarkdown,\nmagrittr, httpuv (>= 1.4.4
## data.table   "bit64 (>= 4.0.0), bit (>= 4.0.4), curl, R.utils, xts,\nnanotime, zoo (>= 1.8-1), yaml,
## DBI          "blob, covr, dbplyr, dplyr, glue, hms, knitr, magrittr,\nrmarkdown, rprojroot, RMariaDB
## dbplyr       "bit64, covr, knitr, Lahman, nycflights13, odbc, RMariaDB (>=\n1.0.2), rmarkdown, RPost
## desc         "covr, testthat, whoami, withr, spelling, gh"
## devtools     "BiocManager (>= 1.30.10), curl (>= 4.3), digest (>= 0.6.25),\nfoghorn (>= 1.3.1), gmai
## diffobj      "knitr, rmarkdown"
## digest       "tinytest, knitr, rmarkdown, minidown"
## dplyr        "bench, broom, callr, covr, DBI, dbplyr (>= 1.4.3), knitr,\nLahman, lobstr, microbenchma
## DT           "knitr (>= 1.8), rmarkdown, shiny (>= 1.2.0), testit"
## e1071        "cluster, mlbench, nnet, randomForest, rpart, SparseM, xtable,\nMatrix, MASS, slam"
## ellipsis     "covr, testthat"
## evaluate     "testthat, lattice, ggplot2"
## fansi        "unitizer, knitr, rmarkdown"
## farver       "testthat (>= 2.1.0), covr"
## fastmap      "testthat (>= 2.1.1)"
## float        NA
## forcats      "covr, dplyr, ggplot2, knitr, readr, rmarkdown, testthat"
## foreach      "randomForest, doMC, doParallel, testthat, knitr, rmarkdown"
## fs           "testthat, covr, pillar (>= 1.0.0), tibble (>= 1.1.0), crayon,\nrmarkdown, knitr, withr
## gbm          "covr, gridExtra, knitr, pdp, RUnit, splines, tinytest, vip,\nviridis"
## gdata        "RUnit"
## generics     "covr, pkgload, testthat, tibble"
## gert         "spelling, knitr, rmarkdown, testthat"
## ggplot2      "covr, dplyr, ggplot2movies, hexbin, Hmisc, knitr, lattice,\nmapproj, maps, maptools, mu
## gh           "covr, knitr, rmarkdown, rprojroot, spelling, testthat (>=\n2.3.2), withr"
## gitcreds     "codetools, testthat, knitr, mockery, oskeyring, rmarkdown,\nwithr"
## glue         "testthat, covr, magrittr, crayon, knitr, rmarkdown, DBI,\nRSQLite, R.utils, forcats, m
## gower        "tinytest (>= 0.9.3),"
## gplots       "grid, MASS, knitr"
## gtable       "covr, testthat, knitr, rmarkdown, ggplot2, profvis"
## gtools       NA
## haven        "covr, fs, knitr, rmarkdown, testthat, pillar (>= 1.4.0), cli,\ncrayon"
## highr        "knitr, testit"
## hms          "crayon, lubridate, pillar (>= 1.1.0), testthat (>= 3.0.0)"
## htmltools    "markdown, testthat, withr, Cairo, ragg, shiny"
## htmlwidgets  "knitr (>= 1.8), rmarkdown, testthat"
##              Enhances
## askpass      NA
## assertthat   NA
## backports    NA
## base64enc    "png"
## BH           NA
## bitops       NA
## blob         NA
## bookdown     NA
## brew         NA
## brio         NA
## broom        NA
## cachem       NA
## callr        NA
## caret        NA
## caTools      NA
## cellranger   NA
## cli          NA
## clipr        NA
## colorspace   NA
## commonmark   NA
## covr         NA
## cpp11        NA
## crayon       NA
## credentials  NA
## crosstalk    NA
## curl         NA
## data.table   NA
## DBI          NA
## dbplyr       NA
## desc         NA
## devtools     NA
## diffobj      NA
## digest       NA
## dplyr        NA
## DT           NA
## e1071        NA
## ellipsis     NA
## evaluate     NA
## fansi        NA
## farver       NA
## fastmap      NA
## float        NA
## forcats      NA
## foreach      NA
## fs           NA
## gbm          NA
## gdata        NA
## generics     NA
## gert         NA
## ggplot2      "sp"
## gh           NA
## gitcreds     NA
## glue         NA
## gower        NA
## gplots       NA
## gtable       NA
## gtools       NA
## haven        NA
## highr        NA
## hms          NA
## htmltools    "knitr"
## htmlwidgets  "shiny (>= 1.1)"
##              License                                  License_is_FOSS License_restricts_use OS_type
## askpass      "MIT + file LICENSE"                     NA              NA                    NA
## assertthat   "GPL-3"                                  NA              NA                    NA
## backports    "GPL-2 | GPL-3"                          NA              NA                    NA
## base64enc    "GPL-2 | GPL-3"                          NA              NA                    NA
## BH           "BSL-1.0"                                NA              NA                    NA
## bitops       "GPL (>= 2)"                             NA              NA                    NA
## blob         "GPL-3"                                  NA              NA                    NA
## bookdown     "GPL-3"                                  NA              NA                    NA
## brew         "GPL-2"                                  NA              NA                    NA
## brio         "MIT + file LICENSE"                     NA              NA                    NA
## broom        "MIT + file LICENSE"                     NA              NA                    NA
## cachem       "MIT + file LICENSE"                     NA              NA                    NA
## callr        "MIT + file LICENSE"                     NA              NA                    NA
## caret        "GPL (>= 2)"                             NA              NA                    NA
## caTools      "GPL-3"                                  NA              NA                    NA
## cellranger   "MIT + file LICENSE"                     NA              NA                    NA
## cli          "MIT + file LICENSE"                     NA              NA                    NA
## clipr        "GPL-3"                                  NA              NA                    NA
## colorspace   "BSD_3_clause + file LICENSE"            NA              NA                    NA
## commonmark   "BSD_2_clause + file LICENSE"            NA              NA                    NA
## covr         "GPL-3"                                  NA              NA                    NA
## cpp11        "MIT + file LICENSE"                     NA              NA                    NA
## crayon       "MIT + file LICENSE"                     NA              NA                    NA
## credentials  "MIT + file LICENSE"                     NA              NA                    NA
## crosstalk    "MIT + file LICENSE"                     NA              NA                    NA
## curl         "MIT + file LICENSE"                     NA              NA                    NA
## data.table   "MPL-2.0 | file LICENSE"                 NA              NA                    NA
## DBI          "LGPL (>= 2.1)"                          NA              NA                    NA
## dbplyr       "MIT + file LICENSE"                     NA              NA                    NA
## desc         "MIT + file LICENSE"                     NA              NA                    NA
## devtools     "GPL (>= 2)"                             NA              NA                    NA
## diffobj      "GPL (>= 2)"                             NA              NA                    NA
## digest       "GPL (>= 2)"                             NA              NA                    NA
## dplyr        "MIT + file LICENSE"                     NA              NA                    NA
## DT           "GPL-3 | file LICENSE"                   NA              NA                    NA
## e1071        "GPL-2 | GPL-3"                          NA              NA                    NA
## ellipsis     "GPL-3"                                  NA              NA                    NA
## evaluate     "MIT + file LICENSE"                     NA              NA                    NA
## fansi        "GPL (>= 2)"                             NA              NA                    NA
## farver       "MIT + file LICENSE"                     NA              NA                    NA
## fastmap      "MIT + file LICENSE"                     NA              NA                    NA
## float        "BSD 2-clause License + file LICENSE"    NA              NA                    NA
## forcats      "MIT + file LICENSE"                     NA              NA                    NA
## foreach      "Apache License (== 2.0)"                NA              NA                    NA
## fs           "GPL-3"                                  NA              NA                    NA
## gbm          "GPL (>= 2) | file LICENSE"              NA              NA                    NA
## gdata        "GPL-2"                                  NA              NA                    NA
## generics     "MIT + file LICENSE"                     NA              NA                    NA
## gert         "MIT + file LICENSE"                     NA              NA                    NA
## ggplot2      "MIT + file LICENSE"                     NA              NA                    NA
## gh           "MIT + file LICENSE"                     NA              NA                    NA
## gitcreds     "MIT + file LICENSE"                     NA              NA                    NA
## glue         "MIT + file LICENSE"                     NA              NA                    NA
## gower        "GPL-3"                                  NA              NA                    NA
## gplots       "GPL-2"                                  NA              NA                    NA
## gtable       "GPL-2"                                  NA              NA                    NA
## gtools       "GPL-2"                                  NA              NA                    NA
## haven        "MIT + file LICENSE"                     NA              NA                    NA
## highr        "GPL"                                    NA              NA                    NA
## hms          "MIT + file LICENSE"                     NA              NA                    NA
## htmltools    "GPL (>= 2)"                             NA              NA                    NA
## htmlwidgets  "MIT + file LICENSE"                     NA              NA                    NA
##              NeedsCompilation Built
## askpass      "yes"            "4.0.4"
## assertthat   "no"             "4.0.4"
## backports    "yes"            "4.0.3"
## base64enc    "yes"            "4.0.3"
## BH           "no"             "4.0.3"
## bitops       "yes"            "4.0.3"
## blob         "no"             "4.0.4"
## bookdown     "no"             "4.0.4"
## brew         NA               "4.0.3"
## brio         "yes"            "4.0.4"
## broom        "no"             "4.0.4"
## cachem       "yes"            "4.0.4"
## callr        "no"             "4.0.4"
## caret        "yes"            "4.0.4"
## caTools      "yes"            "4.0.4"
## cellranger   "no"             "4.0.4"
## cli          "no"             "4.0.4"
## clipr        "no"             "4.0.4"
## colorspace   "yes"            "4.0.4"
## commonmark   "yes"            "4.0.4"
## covr         "yes"            "4.0.4"
## cpp11        "no"             "4.0.4"
## crayon       "no"             "4.0.4"
## credentials  "no"             "4.0.4"
## crosstalk    "no"             "4.0.4"
## curl         "yes"            "4.0.4"
## data.table   "yes"            "4.0.4"
## DBI          "no"             "4.0.4"
## dbplyr       "no"             "4.0.4"
## desc         "no"             "4.0.4"
## devtools     "no"             "4.0.4"
## diffobj      "yes"            "4.0.4"
## digest       "yes"            "4.0.4"
## dplyr        "yes"            "4.0.4"
## DT           "no"             "4.0.4"
## e1071        "yes"            "4.0.4"
## ellipsis     "yes"            "4.0.4"
## evaluate     "no"             "4.0.4"
## fansi        "yes"            "4.0.4"
## farver       "yes"            "4.0.4"
## fastmap      "yes"            "4.0.4"
## float        "yes"            "4.0.3"
## forcats      "no"             "4.0.4"
## foreach      "no"             "4.0.4"
## fs           "yes"            "4.0.4"
## gbm          "yes"            "4.0.4"
## gdata        "no"             "4.0.4"
## generics     "no"             "4.0.4"
## gert         "yes"            "4.0.4"
## ggplot2      "no"             "4.0.4"
## gh           "no"             "4.0.4"
## gitcreds     "no"             "4.0.4"
## glue         "yes"            "4.0.4"
## gower        "yes"            "4.0.3"
## gplots       "no"             "4.0.4"
## gtable       "no"             "4.0.4"
## gtools       "yes"            "4.0.3"
## haven        "yes"            "4.0.4"
## highr        "no"             "4.0.4"
## hms          "no"             "4.0.4"
## htmltools    "yes"             "4.0.4"
## htmlwidgets  "no"              "4.0.4"
##  [ reached getOption("max.print") -- omitted 125 rows ]
```