Summary

Stepan

13 10 2020

setwd('C:/Users/Stepan/Desktop/6 курс/Машинное обучение/DZ4')  
d1 <- read.csv2('DATASET\_test.csv.', header = TRUE, encoding = 'UNICOD')  
#d1 <- read.csv2('')  
d2 <- read.csv2('DATASET\_summmary.csv', header = TRUE, encoding = 'UNICOD')  
d2 <- d2[-1]  
d <- dplyr::bind\_cols(d1, d2)  
d <- d[,-1]  
d$CLIENT\_TOGETHER\_INCOME <- as.factor(d$CLIENT\_TOGETHER\_INCOME)  
head(d)

## LOAN\_PRODUCT\_TYPE LOAN\_AMOUNT BRANCH\_REGION  
## 1 Cпоживчий кредит без довiдки про доходи 16816.14 Харків  
## 2 Cпоживчий кредит без довiдки про доходи 11210.76 Харків  
## 3 Cпоживчий кредит без довiдки про доходи 11210.76 Полтавська область  
## 4 Cпоживчий кредит без довiдки про доходи 16816.14 Запорізька область  
## 5 Cпоживчий кредит без довiдки про доходи 16816.14 Полтавська область  
## 6 Cпоживчий кредит без довiдки про доходи 10373.44 Черкаська область  
## CLIENT\_GENDER CLIENT\_FAMILYSTATUS CLIENT\_COUNTDEPENDENTS  
## 1 Чоловік Одружений (-на) 0  
## 2 Чоловік Розлучений (-на) 0  
## 3 Жінка Неодружений(-на) 0  
## 4 Чоловік Розлучений (-на) 0  
## 5 Жінка Неодружений(-на) 0  
## 6 Чоловік Одружений (-на) 1  
## CLIENT\_TOTALEXPERIENCE CLIENT\_LASTEXPERIENCE CLIENT\_EDUCATION  
## 1 30 2 Середня - технічна  
## 2 21 7 Середня - технічна  
## 3 28 22 Середня - технічна  
## 4 23 6 Вища освіта  
## 5 17 2 Середня - технічна  
## 6 20 14 Вища освіта  
## CLIENT\_ACTIVITYTYPE CLIENT\_TOGETHER\_INCOME LOAN\_OUTSTANDINGLOANSCOUNT  
## 1 торгівля та послуги 6000 2  
## 2 інша 4000 2  
## 3 інша 5200 2  
## 4 інша 10000 6  
## 5 торгівля та послуги 5800 6  
## 6 інша 9900 4  
## LOAN\_EXISTINGCUSTOMERFLAG LOAN\_OVERDUE\_EXIST\_FLAG SCR\_WORKPLACECONFIRMED  
## 1 Так Ні Не визначено  
## 2 Так Так Не визначено  
## 3 Ні Так Не визначено  
## 4 Ні Так Не визначено  
## 5 Ні Так Не визначено  
## 6 Ні Ні Не визначено  
## EMPLOYMENTTYPE DECADE LOAN\_TERM ZODIAC ZODIAC\_CHINA AGE  
## 1 Офіційно працевлаштований третья декада 1095 лев Тигр 54  
## 2 Офіційно працевлаштований первая декада 1092 скорпион Бык 43  
## 3 Офіційно працевлаштований вторая декада 1095 рак Тигр 54  
## 4 Офіційно працевлаштований вторая декада 1094 телец Петух 48  
## 5 Офіційно працевлаштований вторая декада 1094 скорпион Свинья 45  
## 6 Офіційно працевлаштований вторая декада 364 овен Тигр 43  
## CHANGE\_WORK  
## 1 була зміна місця роботи  
## 2 була зміна місця роботи  
## 3 була зміна місця роботи  
## 4 була зміна місця роботи  
## 5 була зміна місця роботи  
## 6 була зміна місця роботи  
## MATCH\_OF\_THE\_REGISTRATION\_PLACE\_WITH\_THE\_ACTUAL\_RESIDENCE  
## 1 ИСТИНА  
## 2 ИСТИНА  
## 3 ИСТИНА  
## 4 ИСТИНА  
## 5 ЛОЖЬ  
## 6 ИСТИНА  
## FAMILY\_PHONE REAL\_ESTATE CAR  
## 1 не вказано телефон рідних має власну нерухомість не має власного  
## 2 не вказано телефон рідних не має власної нерухомості не має власного  
## 3 не вказано телефон рідних не має власної нерухомості не має власного  
## 4 не вказано телефон рідних має власну нерухомість не має власного  
## 5 вказано телефон рідних має власну нерухомість не має власного  
## 6 не вказано телефон рідних має власну нерухомість не має власного  
## INFORMATION\_IN\_MGB INFORMATION\_IN\_BKI STRABIS  
## 1 відсутня наявна не має ознаки страбісу  
## 2 відсутня наявна є ознака страбісу  
## 3 відсутня наявна не має ознаки страбісу  
## 4 відсутня наявна не має ознаки страбісу  
## 5 відсутня наявна не має ознаки страбісу  
## 6 відсутня наявна не має ознаки страбісу  
## PERCENT\_IN\_THE\_LOAN\_AMOUNT DELAY p\_sr p\_mr p\_pr  
## 1 2.80 нет прострочки 14897.72 16354.910 14419.56  
## 2 2.80 нет прострочки 12666.07 13736.473 13013.55  
## 3 2.16 нет прострочки 14005.06 11064.337 13797.27  
## 4 1.68 есть прострочка 19361.02 15610.066 18608.73  
## 5 2.90 есть прострочка 14674.55 19378.418 14256.73  
## 6 1.05 нет прострочки 19249.43 9929.325 18484.02  
## p\_ft p\_rf nn  
## 1 11719.648 12518.611 -0.131584948  
## 2 11719.648 8410.632 -0.290397688  
## 3 11719.648 10865.871 -0.198134642  
## 4 11254.198 17571.975 0.007642399  
## 5 11719.648 11392.028 -0.337598037  
## 6 5430.537 19955.960 -0.457279574

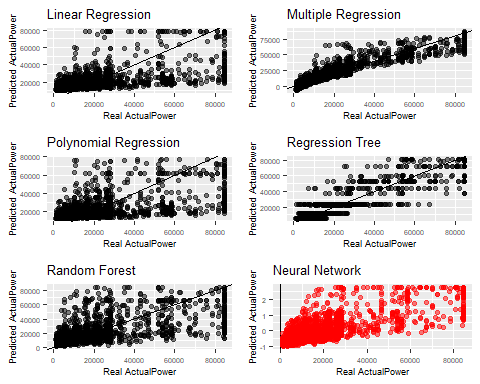
#install.packages('gridExtra')  
library(gridExtra)

## Warning: package 'gridExtra' was built under R version 3.6.3

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.6.3

g\_sr <- ggplot(d, aes(x=LOAN\_AMOUNT, y=p\_sr)) +   
 geom\_abline(intercept=0, slope=1) +  
 geom\_point(alpha=0.5) + labs(title="Linear Regression", x="Real ActualPower", y="Predicted ActualPower") +   
 theme(plot.title=element\_text(size=10), axis.title.x=element\_text(size=7), axis.title.y=element\_text(size=7), axis.text.x=element\_text(size=5), axis.text.y=element\_text(size=5)) + theme(legend.position="none")  
  
g\_mr <- ggplot(d, aes(x=LOAN\_AMOUNT, y=p\_mr)) +   
 geom\_abline(intercept=0, slope=1) +  
 geom\_point(alpha=0.5) + labs(title="Multiple Regression", x="Real ActualPower", y="Predicted ActualPower") +   
 theme(plot.title=element\_text(size=10), axis.title.x=element\_text(size=7), axis.title.y=element\_text(size=7), axis.text.x=element\_text(size=5), axis.text.y=element\_text(size=5)) + theme(legend.position="none")  
  
g\_pr <- ggplot(d, aes(x=LOAN\_AMOUNT, y=p\_pr)) +   
 geom\_abline(intercept=0, slope=1) +  
 geom\_point(alpha=0.5) + labs(title="Polynomial Regression", x="Real ActualPower", y="Predicted ActualPower") +   
 theme(plot.title=element\_text(size=10), axis.title.x=element\_text(size=7), axis.title.y=element\_text(size=7), axis.text.x=element\_text(size=5), axis.text.y=element\_text(size=5)) + theme(legend.position="none")   
  
g\_ft <- ggplot(d, aes(x=LOAN\_AMOUNT, y=p\_ft)) +   
 geom\_abline(intercept=0, slope=1) +  
 geom\_point(alpha=0.5) + labs(title="Regression Tree", x="Real ActualPower", y="Predicted ActualPower") +   
 theme(plot.title=element\_text(size=10), axis.title.x=element\_text(size=7), axis.title.y=element\_text(size=7), axis.text.x=element\_text(size=5), axis.text.y=element\_text(size=5)) + theme(legend.position="none")  
  
g\_rf <- ggplot(d, aes(x=LOAN\_AMOUNT, y=p\_rf)) +   
 geom\_abline(intercept=0, slope=1) +  
 geom\_point(alpha=0.5) + labs(title="Random Forest", x="Real ActualPower", y="Predicted ActualPower") +   
 theme(plot.title=element\_text(size=10), axis.title.x=element\_text(size=7), axis.title.y=element\_text(size=7), axis.text.x=element\_text(size=5), axis.text.y=element\_text(size=5)) + theme(legend.position="none")  
  
g\_nn <- ggplot(d, aes(x=LOAN\_AMOUNT, y=nn)) +   
 geom\_abline(intercept=0, slope=1) +  
 geom\_point(color = 'red', alpha=0.5) + labs(title="Neural Network", x="Real ActualPower", y="Predicted ActualPower") +   
 theme(plot.title=element\_text(size=10), axis.title.x=element\_text(size=7), axis.title.y=element\_text(size=7), axis.text.x=element\_text(size=5), axis.text.y=element\_text(size=5)) + theme(legend.position="none")  
  
gridExtra::grid.arrange(g\_sr,g\_mr,g\_pr,g\_ft,g\_rf,g\_nn,ncol=2)



#Calc prediction error and visualize it

sr <- mean ((d$LOAN\_AMOUNT - d$p\_sr) ^ 2)  
mr <- mean ((d$LOAN\_AMOUNT - d$p\_mr) ^ 2)  
pr <- mean ((d$LOAN\_AMOUNT - d$p\_pr) ^ 2)  
ft <- mean ((d$LOAN\_AMOUNT - d$p\_ft) ^ 2)  
rf <- mean ((d$LOAN\_AMOUNT - d$p\_rf) ^ 2)  
nn <- mean ((d$LOAN\_AMOUNT - d$nn) ^ 2)  
mse <- data.frame(sr,mr,pr,ft,rf,nn)  
head(mse)

## sr mr pr ft rf nn  
## 1 199990348 45119952 198427351 51351540 198435544 735129248

#install.packages('reshape')  
library(reshape)

## Warning: package 'reshape' was built under R version 3.6.3

mse1 <- reshape::melt.data.frame(mse)

## Using as id variables

## Using as id variables  
head(mse1)

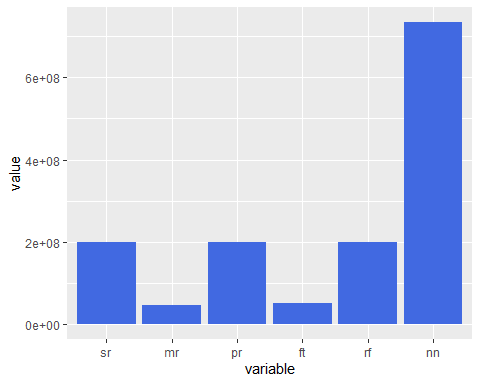
## variable value  
## 1 sr 199990348  
## 2 mr 45119952  
## 3 pr 198427351  
## 4 ft 51351540  
## 5 rf 198435544  
## 6 nn 735129248

b1 <- ggplot(mse1, aes(x=variable, y=value)) +  
 geom\_bar(stat="summary", fun.y="mean", fill = 'royalblue')

## Warning: Ignoring unknown parameters: fun.y

b1

## No summary function supplied, defaulting to `mean\_se()`



**# Висновок: Найменша похибка в багатофакторній моделі.**

#Save results

ggsave("plot.jpg", plot=b1 + theme\_classic(), width=20, height=15,  
units="cm", dpi=600)

## No summary function supplied, defaulting to `mean\_se()`

**#Висновок: результати моделювання збережені у файлі.**