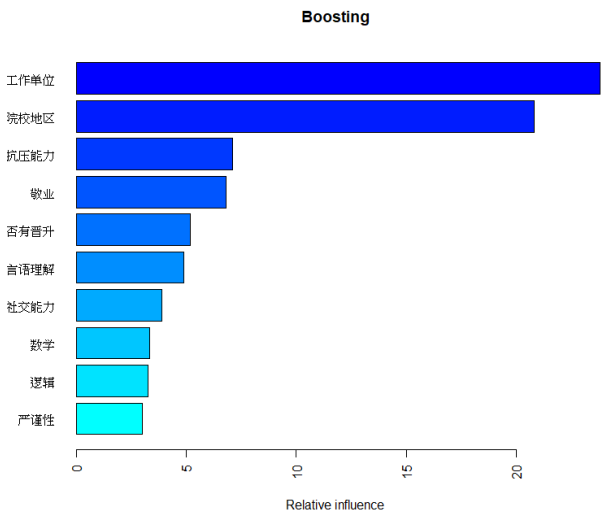
下圖為使用 boosting 算法跑出的預測結果。

```
> table(dta$是否1年内离职, bp$class)

       在职 离职
  在职 1217    0
  离职    0  242
```

下圖為 gbm 模型的 relative influence 及前十重要變數。

```
> print(importance)
                  Overall
性别             0.000000
工作单位       567.254896
工作单位类别    27.044328
入职年份         5.730876
学历             5.197385
毕业院校地区   493.736418
专业类别        51.940060
职称            40.428796
职业资格         0.000000
是否党员         4.741006
言语理解       130.998061
数学            82.900199
逻辑            92.000879
常识            22.081579
成就导向        14.402401
抗压能力       167.565596
灵活性          62.024621
影响性          66.113981
支配性          25.863762
外向性          22.425999
社交能力        81.651510
心理感受性      28.043930
创新            26.173300
敬业           146.505167
情绪稳定性      41.055088
严谨性          82.106316
完美主义倾向    47.220287
录用时岗位级别   6.456336
是否有晋升     129.950589
```



下圖為 Random Forest 跑出的預測結果。

```
> table(dta2$是否1年内离职, rp)
       rp
        在职 离职
  在职 1217    0
  离职    0  242
```

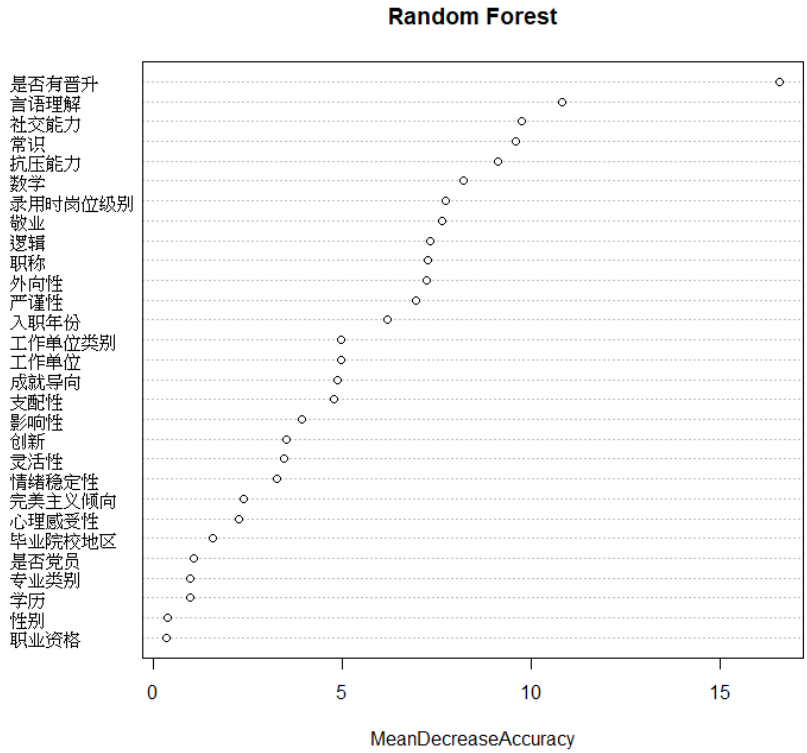
左圖為 MeanDecreaseAccuracy；右圖為 MeanDecreaseGini。

```
> print(importance(r, type = 1))
                 MeanDecreaseAccuracy
性别                        0.3931632
工作单位                    4.9710190
工作单位类别                4.9871153
入职年份                    6.1857156
学历                        0.9753027
毕业院校地区                1.5884117
专业类别                    0.9763469
职称                        7.2624575
职业资格                    0.3653418
是否党员                    1.0690326
言语理解                   10.8283225
数学                        8.2050712
逻辑                        7.3256053
常识                        9.6059001
成就导向                    4.8722095
抗压能力                    9.1158297
灵活性                      3.4789337
影响性                      3.9421469
支配性                      4.7747966
外向性                      7.2342872
社交能力                    9.7547769
心理感受性                  2.2783987
创新                        3.5382186
敬业                        7.6424926
情绪稳定性                  3.2909593
严谨性                      6.9502830
完美主义倾向                2.3848677
录用时岗位级别              7.7483977
是否有晋升                 16.5548521
```

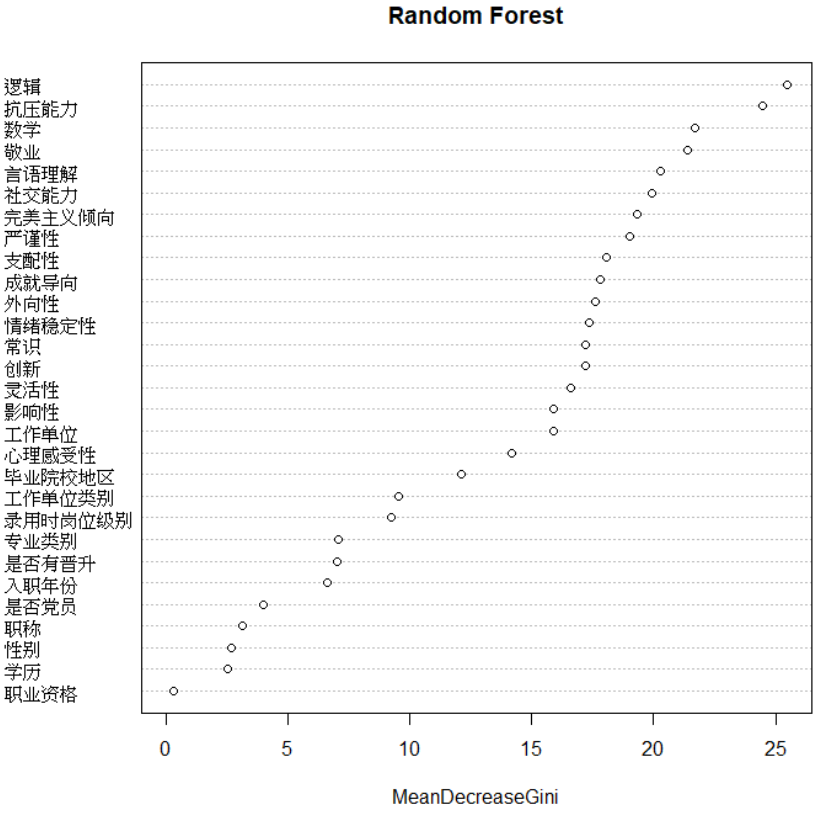
```
> print(importance(r, type = 2))
                 MeanDecreaseGini
性别                    2.7003264
工作单位               15.8723336
工作单位类别            9.5272192
入职年份                6.6307938
学历                    2.5122017
毕业院校地区           12.1255405
专业类别                7.0593838
职称                    3.1515274
职业资格                0.3224313
是否党员                4.0169366
言语理解               20.2793350
数学                   21.7048946
逻辑                   25.4746362
常识                   17.2217428
成就导向               17.8311600
抗压能力               24.4598955
灵活性                 16.6085496
影响性                 15.8991854
支配性                 18.0778324
外向性                 17.6099065
社交能力               19.9386810
心理感受性             14.1690061
创新                   17.2190657
敬业                   21.3938448
情绪稳定性             17.3652738
严谨性                 19.0097206
完美主义倾向           19.3341606
录用时岗位级别          9.2191841
是否有晋升              7.0369760
```

下圖為 MeanDecreaseAccuracy 的視覺化圖表。

110305075 繆孟珊



下圖為 MeanDecreaseGini 的視覺化圖表。



此處附上完整版 R code。

```
library(tidyverse)

library(data.table)

library(ggplot2)

library(ALSM)

library(adabag)

library(gbm)

library(ada)

library(caret)

#匯入資料

dta <- as.data.frame(fread("C:/Users/ingri/OneDrive/桌面/政大/HRDA/畢業生數據 2_.csv", sep=","))


dta <- dta[,3:32]

dta[["是否 1 年内离职"]] <- factor(dta[["是否 1 年内离职"]])

colnames(dta)

levels(dta[["是否 1 年内离职"]]) <- list(在职  = 0,  离职  = 1)

set.seed(4410)


#boosting

dta$性别  <- factor(dta$性别, levels = c("男", "女"))

dta$学历  <- factor(dta$学历, levels = c("大专","本科", "硕士", "博士"))

dta$工作单位  <- factor(dta$工作单位, levels = c("A 分公司", "B 分公司", "C 分公司", "D 分公司", "E 分公司","F 分公司","G 分公司","H 分公司","I 分公司","J 分公司","K 分公司","L 分公司","M 分公司","N 分公司","O 分公司","P 分公司","Q 分公司","R
```

```
分公司","S 分公司","T 分公司"))

dta$职业资格 <- factor(dta$职业资格, levels = c("无","软体设计师","助理人力资源管理师","全国计算机信息高新技术考试合格证","会计从业资格","电工证","仓库保管工","三级企业人力资源管理师","CCNA 证书"))

dta$毕业院校地区 <- factor(dta$毕业院校地区, levels = c("上海","山东","山西","广东","广西","云南","内蒙古","天津","北京","四川","宁夏","甘肃","吉林","安徽","江西","江苏","辽宁","武汉","河北","河南","贵州","重庆","陕西","浙江","海南","湖北","湖南","黑龙江","新疆","境外","福建"))

dta$专业类别 <- factor(dta$专业类别, levels = c("工学","文学","农学","体育","医学","其他","法学","哲学","教育学","理学","管理学"))

dta$职称 <- factor(dta$职称, levels = c("工程师","无","助理工程师","助理会计师","助理经济师"))


dta$工作单位类别 <- factor(dta$工作单位类别, levels = c("A 类", "B 类","C 类","D 类","E 类","F 类","G 类"))

dta$是否党员 <- factor(dta$是否党员, levels = c("是", "否"))


#by boosting

b <- boosting(是否 1 年内离职~., dta)

bp <- predict(b, dta)

table(dta$是否 1 年内离职, bp$class)

importanceplot(b, top = 5)


#by gbm_model

dta$是否 1 年内离职 <- ifelse(dta$是否 1 年内离职 == "在职", 0, 1)
```

```
gbm_model <- gbm(是否 1 年内离职 ~ .,
                 data = dta,
                 distribution = "bernoulli",
                 n.trees = 500,
                 interaction.depth = 3,
                 shrinkage = 0.01,
                 cv.folds = 5)
n_trees <- length(gbm_model$trees)
# Calculate feature importance using permutation-based method
importance <- varImp(gbm_model, scale = FALSE, numTrees = n_trees)

print(importance)

summary(
  gbm_model, # gbm object
  cBars = 10, # the number of bars to draw. length(object$var.names)
  plotit = TRUE, # an indicator as to whether the plot is generated.defult TRUE.
  method = relative.influence, # The function used to compute the relative influence. 亦可使用 permutation.test.gbm
  las = 2,
  main = "Boosting"
)
```

110305075 繆孟珊

```
#Random forest

library(randomForest)

set.seed(4410)

dta2 <- read.csv("C:/Users/ingri/OneDrive/桌面/政大/HRDA/畢業生數據 2_.csv")

dta2 <- dta2[,3:32]

dta2[["是否 1 年內离职"]] <- factor(dta2[["是否 1 年內离职"]])

levels(dta2[["是否 1 年內离职"]]) <- list(在职  = 0,  离职  = 1)

set.seed(101010)

r <- randomForest(是否 1 年內离职~., data = dta2, proximity = TRUE, importance
= TRUE, na.rm = TRUE)

rp <- predict(r, dta2)

table(dta2$是否 1 年內离职, rp)


# Relative Influence

var_importance <- importance(r)

print(var_importance)


print(importance(r, type = 1))

#視覺化圖表

varImpPlot(r, sort = TRUE, type = 1,main = "Random Forest")

varImpPlot(r, type = 2,main = "Random Forest")
```