PROJECT MIDWAY REPORT

TAIWANESE BANKRUPTCY PREDICTION

Bankruptcy or business failure can have a negative impact both on the enterprise itself and the global economy. Business practitioners, investors, governments, and academic researchers have long studied ways to identify the potential risk of business failure to reduce the economic loss caused by bankruptcy. In short, bankruptcy prediction is very important for many related financial institutions. In general, the aim is to predict the likelihood that a firm may go bankrupt. Financial institutions need effective prediction models to make appropriate lending decisions.

**A. Pre-processing:** We have to pre-processing the data for the analysis phase

*#Loading the data*

banking.df <- read.csv("data.csv")

head(banking.df)

**Missing Values:**

#We can find the missing values using the below command

any(is.na(banking.df))

No value is missing in the dataframe

**Data partition for validation:**

set.seed(1, sample.kind = "Rounding")

validation\_index <- createDataPartition(y = banking.df$Bankrupt, times = 1, p = 0.2, list = FALSE)

validation <- banking.df[validation\_index,]

banking.df <- banking.df[-validation\_index,]

**Identifying & Removing Predictors:**

nzv <- nearZeroVar(df)

nzv[!nzv %**in**% 1] *#to avoid delete the variable of interest*

## Variable Format Changes:

## banking.df$Bankrupt <- factor(df$Bankrupt, labels = c("non\_bankruptcy", "bankruptcy"))

## validation$Bankrupt <- factor(validation$Bankrupt, labels = c("non\_bankruptcy", "bankruptcy"))

## Train and test sets:

To train the models, test, and optimize we are using 80% of the data for training the model and 20% for testing

train\_index <- createDataPartition(y = banking.df$Bankrupt, times = 1, p = 0.8, list = FALSE)

train\_set <- banking.df[train\_index,]

test\_set <- banking.df[-train\_index,]

**B. Analysis**

Simple analysis

summary(banking.df[,1:12])

Bankrupt. ROA.C..before.interest.and.depreciation.before.interest

Min. :0.000 Min. :0.024

1st Qu.:0.000 1st Qu.:0.476

Median :0.000 Median :0.503

Mean :0.032 Mean :0.506

3rd Qu.:0.000 3rd Qu.:0.536

Max. :1.000 Max. :1.000

ROA.A..before.interest.and...after.tax ROA.B..before.interest.and.depreciation.after.tax

Min. :0.000 Min. :0.034

1st Qu.:0.536 1st Qu.:0.527

Median :0.560 Median :0.552

Mean :0.559 Mean :0.554

3rd Qu.:0.590 3rd Qu.:0.585

Max. :0.985 Max. :0.958

Operating.Gross.Margin Realized.Sales.Gross.Margin Operating.Profit.Rate Pre.tax.net.Interest.Rate

Min. :0.000 Min. :0.000 Min. :0.000 Min. :0.000

1st Qu.:0.601 1st Qu.:0.601 1st Qu.:0.999 1st Qu.:0.797

Median :0.606 Median :0.606 Median :0.999 Median :0.797

Mean :0.608 Mean :0.608 Mean :0.999 Mean :0.797

3rd Qu.:0.614 3rd Qu.:0.614 3rd Qu.:0.999 3rd Qu.:0.798

Max. :1.000 Max. :1.000 Max. :1.000 Max. :0.851

After.tax.net.Interest.Rate Non.industry.income.and.expenditure.revenue

Min. :0.000 Min. :0.000

1st Qu.:0.809 1st Qu.:0.303

Median :0.809 Median :0.304

Mean :0.809 Mean :0.304

3rd Qu.:0.809 3rd Qu.:0.304

Max. :0.864 Max. :1.000

Continuous.interest.rate..after.tax. Operating.Expense.Rate

Min. :0.000 Min. :0.00e+00

1st Qu.:0.782 1st Qu.:0.00e+00

Median :0.782 Median :0.00e+00

Mean :0.781 Mean :1.98e+09

3rd Qu.:0.782 3rd Qu.:4.07e+09

Max. :0.829 Max. :9.99e+09

## Distributions and Relationships

This graph shows the high prevalence in the data, only 3% of the companies in the data have a bankruptcy and the other 97% have no bankruptcy

## 