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

banking.df <- banking.df[,c(1:9,96)]

head(banking.df)

train.index <- sample(c(1:dim(banking.df)[1]), dim(banking.df)[1]\*0.5)

train.df <- banking.df[train.index, ]

valid.df <- banking.df[-train.index, ]

Regression:

logit.reg <- glm(Bankrupt. ~ ., data = train.df, family = "binomial"); summary(logit.reg)

A screenshot of a computer

Description automatically generated

# Using logistic regression we are predicting the probabilities of Y=1 with the validation dataset

logit.reg.pred <- predict(logit.reg, valid.df, type = "response")

# first 5 actual and predicted records

data.frame(actual = valid.df$Bankrupt.[1:5], predicted = logit.reg.pred[1:5])A screenshot of a computer

Description automatically generated

#0.5 above as 1 and 0.5 below as 0

logit.reg.pred.classes <- ifelse(logit.reg.pred > 0.5, 1, 0)

confusionMatrix(as.factor(logit.reg.pred.classes), as.factor(valid.df$Bankrupt.))

A screenshot of a computer

Description automatically generated