Project on Credit card Fraud Detection

SWAPNIL SANJAY CHAVAN

10/17/2019

## Business Problem: Credit card fraud detection using supervise ML Model/Algorithm in R language

### Fraud Detection

#### Assumptions Made before the Poject:

Assumption# 1

Assumption# 2

#### Library used

library(randomForest)

## randomForest 4.6-14

## Type rfNews() to see new features/changes/bug fixes.

library(e1071)  
library(rpart)  
library(rpart.plot)  
library(caTools)

#### Read and Classify the Data

card<-read.csv("creditcard.csv")  
  
str(card)

#### Class Factor

card$Class<-factor(card$Class)

#### Run code on both ends set any value I put 1

set.seed(1)

#### Find Number of Rows

nrow(card)

## [1] 284807

#### Divide entire Dataset into Training and Testing Datasets

T= Training Datasets(70% Data)

F= Testing Datasets(30% Data)

split<-sample.split(card$Class,SplitRatio = 0.7)  
train<-subset(card,split==T)  
cv<-subset(card,split==F)  
nrow(train)  
nrow(cv)

#### Creat table and Find Fraudulant transaction

efficiency is greater than 99 percent

table(cv$Class)

##   
## 0 1   
## 85295 148

85265/(85265+148)

## [1] 0.9982672

#### Logistic regression:

Logistic regression because set transaction is in 0 and 1

Dependant variable= categorical variable / Dependant vaviable= Set of parameter

binomial= Dependant variable in 0 and 1

.= Consider all variable as independant variable except Class

Class= Dependant variable

glm=Logistic regression Model

glm.model<-glm(Class~.,data=train,family = "binomial")

#### Predicting glm.Model

glm.predict<-predict(glm.model,cv,type = "response")  
glm.predict  
nrow(cv)

#### Compare Class and Predicted glm model value to categorise fraudulant and Non-fraudulent transaction

if predicted value is greater than 0.5 it is fraudulent transaction

if predicted value is less than 0.5 it is non-fraudulent transaction

table produce gives confusion matrixa and accuracy is arount 99%

table(cv$Class,glm.predict>0.5)

##   
## FALSE TRUE  
## 0 85279 16  
## 1 69 79

#### Put and All predicated value in one file and Get working directory

save it as a “checkcredit.csv”

cv$Class  
glm.predict  
cv$predict<-glm.predict  
write.csv(cv,"checkcredit.csv")  
  
getwd()

#### Tree Model developenent for better visualisation

For better visualiasation of the concept to the world

Basically How Algorithm Works

rpart= Classification and Regression Tree

minbucket= min number of observation at each node is 20

tree.model<-rpart(Class~.,data = train,method = "class",minbucket=20)

#### Explanation of Tree

left side Yes

right side No

0= Non Fraudulant

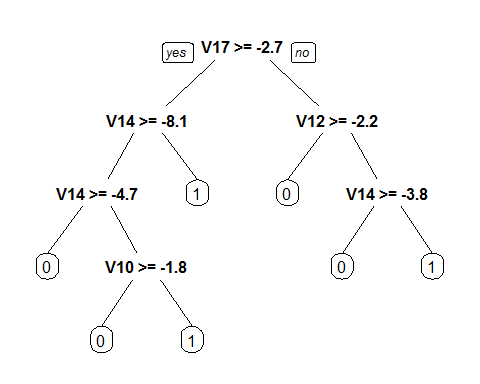
1= fraudulant

if V17 value satisfies first condition it goes to left and check next condition.

if not it goes to right and check the next condition.

According to condition satisfaction we can easily understand which transaction is fraudulant and which is not

prp(tree.model)



tree.predict<-predict(tree.model,cv,type="class")  
tree.predict  
table(cv$Class,tree.predict)

## Finish the Program, Happy Coading…….

### SWAPNIL SANJAY CHAVAN

### Github: sscswapnil

### <https://github.com/sscswapnil/Fraud-detection-with-R.git>