Machine Learning-02 Logistic Regression

Created by H. M. Samadhi Chathuranga Rathnayake

```r
setwd("D:\\Workshops\\R Programming for Data Science Workshop\\Part 04 - Machine Learning\\Datasets")
df=read.csv("loan.CSV")
head(df)

df=df[complete.cases(df),] #Removing missing values
summary(df)

nrow(df)

#Factorize the categorical variables
str(df)

df$Gender=factor(df$Gender)
df$Married=factor(df$Married)
df$Education=factor(df$Education)
df$Self_Employed=factor(df$Self_Employed)
df$Property_Area=factor(df$Property_Area)
df$Loan_Status=factor(df$Loan_Status)

str(df)

#Fitting the logistic regression model
fit=glm(Loan_Status~.,data=df,family = "binomial")
summary(fit)

#Testing the prediction accuracy
set.seed(7777)
trainID=sample(1:nrow(df),0.8*nrow(df))
train=df[trainID,]
test=df[-trainID,]


fit_train=glm(Loan_Status~.,data=train,family = "binomial")
summary(fit_train)

y_pred_probs=predict(fit_train,test,type="response")
y_actual=test$Loan_Status

y_pred=ifelse(y_pred_probs>=0.5,"Y","N")

conf=table(y_actual,y_pred)

accuracy=sum(diag(conf))/length(y_pred)*100
accuracy
```