Topic: Predicting the survival of person based on the features Using decision trees

CODE:-

# to get the data from my google drive if not already downloaded,uncomment the following two lines

#train\_url <- "https://drive.google.com/open?id=0ByQqscFZfvzcWjFkVWxnY0tYbTg"

#train1 <- read.csv(train\_url)

#import the necessary libraries

library(readr)

library(dplyr)

library(party)

library(rpart)

library(rpart.plot)

library(ROCR)

library(ggplot2)

library(rattle)

library(rpart.plot)

library(RColorBrewer)

# set seed for constant output

set.seed(100)

#show first few entries of data

head(train)

#define the test split

test=tail(train)

#plot the number of males and females in each class

ggplot(train,aes(x=factor(Pclass),fill=factor(Sex)))+

geom\_bar(position="dodge")

ggplot(train,aes(x=factor(Pclass),fill=factor(Sex)))+

geom\_bar(position="dodge")+

facet\_grid(". ~ Survived")

posn.j <- position\_jitter(0.5, 0)

ggplot(train,aes(x=factor(Pclass),y=Age,col=factor(Sex)))+

geom\_jitter(size=3,alpha=0.5,position=posn.j)+

facet\_grid(". ~ Survived")

#define the decision tree model

fit <- rpart(Survived ~ Pclass + Sex + Age + SibSp + Parch + Fare + Embarked, data = train, method = "class")

#show the tree

fancyRpartPlot(fit)

#Predict the values for test data

my\_prediction <- predict(fit, test, type = "class")

# Make a dataframe to show the output

my\_solution <- data.frame(PassengerId = test$PassengerId, Survived = my\_prediction)

#Print the predicted output

my\_solution

SCREENSHOTS:







