My Code:

setwd("/Users/rushikeshkhankar/Desktop/R/1555052405\_datasets")

getwd()

library(e1071)

library(caret)

library(readxl)

library(plyr)

library(caTools)

library(randomForest)

AttDS <- read\_excel("Desktop/R/1555052405\_datasets/Attribute DataSet.xlsx")

Dress\_Sales <- read\_excel("Desktop/R/1555052405\_datasets/Dress Sales.xlsx")

View(AttDS)

View(Dress\_Sales)

#remove dress id column

AttDS\_ = AttDS[2:14]

Dress\_Sales\_ = Dress\_Sales[2:24]

View(Dress\_Sales)

#check the unique values for each columns

#lapply(attribset[2:14], unique)

#values checking

#Style

AttDS\_$Style[AttDS\_$Style == 'sexy'] = 'Sexy'

#Price

AttDS\_$Price[AttDS\_$Price == 'low'] = 'Low'

AttDS\_$Price[AttDS\_$Price == 'high'] = 'High'

#Size

AttDS\_$Size[AttDS\_$Size == 's'] = 'S'

AttDS\_$Size[AttDS\_$Size == 'small'] = 'S'

#Season

AttDS\_$Season[AttDS\_$Season == 'spring'] = 'Spring'

AttDS\_$Season[AttDS\_$Season == 'summer'] = 'Summer'

AttDS\_$Season[AttDS\_$Season == 'Automn'] = 'Autumn'

AttDS\_$Season[AttDS\_$Season == 'winter'] = 'Winter'

#NeckLine

AttDS\_$NeckLine[AttDS\_$NeckLine == 'sweetheart'] = 'Sweetheart'

#SleeveLenght

AttDS\_$SleeveLength[AttDS\_$SleeveLength == 'sleevless'] = 'sleeveless'

AttDS\_$SleeveLength[AttDS\_$SleeveLength == 'sleeevless'] = 'sleeveless'

AttDS\_$SleeveLength[AttDS\_$SleeveLength == 'threequater'] = 'threequarter'

AttDS\_$SleeveLength[AttDS\_$SleeveLength == 'thressqatar'] = 'threequarter'

AttDS\_$SleeveLength[AttDS\_$SleeveLength == 'urndowncollor'] = 'turndowncollor'

#Decoration

AttDS\_$Decoration[AttDS\_$Decoration == 'embroidary'] = 'embroidery'

AttDS\_$Decoration[AttDS\_$Decoration == 'sequined'] = 'sequins'

AttDS\_$Decoration[AttDS\_$Decoration == 'ruched'] = 'ruche'

AttDS\_$Decoration[AttDS\_$Decoration == 'none'] = 'null'

#Pattern Type

AttDS\_$`Pattern Type`[AttDS\_$`Pattern Type` == 'none'] = 'null'

AttDS\_$`Pattern Type`[AttDS\_$`Pattern Type` == 'leapord'] = 'leopard'

#factoring

AttDS\_$Style = factor(AttDS\_$Style,

levels = c( 'bohemian', 'Brief','Casual','cute', 'fashion', 'Flare','Novelty','OL','party', 'Sexy','vintage', 'work'),

labels = c(0,1,2,3,4,5,6,7,8,9,10,11))

AttDS\_$Price = factor(AttDS\_$Price,

levels = c('Low','Medium', 'Average','High','very-high'),

labels = c(0,1,2,3,4))

AttDS\_$Size = factor(AttDS\_$Size,

levels = c('free', 'L' ,'M','S' ,'XL'),

labels = c(0,1,2,3,4))

AttDS\_$Season = factor(AttDS\_$Season,

levels = c('Autumn', 'Spring', 'Summer', 'Winter'),

labels = c(0,1,2,3))

AttDS\_$NeckLine = factor(AttDS\_$NeckLine,

levels = c("o-neck","v-neck","boat-neck","peterpan-collor","ruffled","turndowncollor","slash-neck","mandarin-collor","open", "sqare-collor","Sweetheart", "Scoop","halter","backless","bowneck","NULL" ),

labels = c(0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15))

AttDS\_$SleeveLength = factor(AttDS\_$SleeveLength,

levels = c("sleevless","Petal", "full","butterfly" ,"short","threequarter","halfsleeve","cap-sleeves","turndowncollor","capsleeves","half","NULL" ),

labels = c(0,1,2,3,4,5,6,7,8,9,10,11))

AttDS\_$waiseline = factor(AttDS\_$waiseline,

levels = c("empire","natural","null","princess","dropped" ),

labels = c(0,1,2,3,4))

AttDS\_$Material = factor(AttDS\_$Material,

levels = c("null","microfiber","polyster","silk","chiffonfabric","cotton","nylon","other","milksilk","linen","rayon","lycra","mix","acrylic","spandex","lace","modal","cashmere","viscos","knitting","sill","wool","model","shiffon" ),

labels = c(0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23))

AttDS\_$FabricType = factor(AttDS\_$FabricType,

levels = c("chiffon","null","broadcloth","jersey","other","batik","satin","flannel","worsted","woolen","poplin","dobby","knitted","tulle","organza","lace","Corduroy","terry" ),

labels = c(0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17))

AttDS\_$Decoration = factor(AttDS\_$Decoration,

levels = c("ruffles","null","embroidary","bow","lace","beading","sashes","hollowout","pockets","sequined" ,"applique","button","Tiered","rivet","feathers","flowers","pearls","pleat","crystal","ruched","draped","tassel","plain","cascading" ),

labels = c(0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23))

AttDS\_$`Pattern Type` = factor(AttDS\_$`Pattern Type`,

levels = c("animal","print","dot","solid","null","patchwork","striped","geometric","plaid","leopard","floral","character","splice","leapord","none" ),

labels = c(0,1,2,3,4,5,6,7,8,9,10,11,12,13,14))

AttDS\_$Recommendation = sapply(AttDS\_$Recommendation, factor)

#count of missing values in attribset \_ dataset

colSums(is.na(AttDS\_))

#create the function

getmode <- function(v) {

uniqv <- unique(v)

uniqv[which.max(tabulate(match(v, uniqv)))]

}

#Fill missing values with mode

AttDS\_$Price[is.na(AttDS\_$Price) == TRUE] <- getmode(AttDS\_$Price)

AttDS\_$Season[is.na(AttDS\_$Season) == TRUE] <- getmode(AttDS\_$Season)

AttDS\_$NeckLine[is.na(AttDS\_$NeckLine) == TRUE] <- getmode(AttDS\_$NeckLine)

AttDS\_$waiseline[is.na(AttDS\_$waiseline) == TRUE] <- getmode(AttDS\_$waiseline)

AttDS\_$Material[is.na(AttDS\_$Material) == TRUE] <- getmode(AttDS\_$Material)

AttDS\_$FabricType[is.na(AttDS\_$FabricType) == TRUE] <- getmode(AttDS\_$FabricType)

AttDS\_$Decoration[is.na(AttDS\_$Decoration) == TRUE] <- getmode(AttDS\_$Decoration)

AttDS\_$`Pattern Type`[is.na(AttDS\_$`Pattern Type`) == TRUE] <- getmode(AttDS\_$`Pattern Type`)

AttDS\_data <- data.frame(AttDS\_)

str(AttDS\_)

#update columns name in Dres\_sale\_ dataset

Dress\_Sales\_ = rename(Dress\_Sales\_, c('41314'='2/9/2013'))

Dress\_Sales\_ = rename(Dress\_Sales\_, c('41373'='4/9/2013'))

Dress\_Sales\_ = rename(Dress\_Sales\_, c('41434'='6/9/2013'))

Dress\_Sales\_ = rename(Dress\_Sales\_, c('41495'='8/9/2013'))

Dress\_Sales\_ = rename(Dress\_Sales\_, c('41556'='10/9/2013'))

Dress\_Sales\_ = rename(Dress\_Sales\_, c('41617'='12/9/2013'))

Dress\_Sales\_ = rename(Dress\_Sales\_, c('41315'='2/10/2013'))

Dress\_Sales\_ = rename(Dress\_Sales\_, c('41374'='4/10/2013'))

Dress\_Sales\_ = rename(Dress\_Sales\_, c('41435'='6/10/2013'))

Dress\_Sales\_ = rename(Dress\_Sales\_, c('40400'='8/10/2013'))

Dress\_Sales\_ = rename(Dress\_Sales\_, c('41557'='10/10/2013'))

Dress\_Sales\_ = rename(Dress\_Sales\_, c('41618'='12/10/2013'))

#Covert all variable types to numeric

Dress\_Sales\_ <- as.data.frame(apply(Dress\_Sales\_, 2, as.numeric))

#Mean row

Dress\_Sales\_ = as.matrix(Dress\_Sales\_)

k <- which(is.na(Dress\_Sales\_), arr.ind = TRUE)

Dress\_Sales\_[k] <- rowMeans(Dress\_Sales\_, na.rm = TRUE)[k[,1]]

Dress\_Sales\_ = as.data.frame(Dress\_Sales\_)

#sum all values on row on (total sales)

Dress\_Sales\_$total\_sales = rowSums(Dress\_Sales\_)

head(Dress\_Sales\_)

#merge data

merged\_data <- data.frame(AttDS\_, Dress\_Sales\_)

merged\_data

str(merged\_data)

#splitting dataset

set.seed(100)

spl = sample.split(merged\_data$Recommendation, SplitRatio = 0.7)

train = subset(merged\_data, spl==TRUE)

test = subset(merged\_data, spl==FALSE)

print(dim(train)) ; print(dim(test))

#Naive Bayes Model

#build model

Naive\_Model = naiveBayes(Recommendation ~., data = train)

#Create Confusion Matrix

confusionMatrix(train$Recommendation, predict(Naive\_Model, train), positive = '1')

#predict test set

Naive\_Predict = predict(Naive\_Model, test)

table(Naive\_Predict, test$Recommendation)

#regression

regress\_Sales = lm(formula = total\_sales ~ Style+Season+Material+Price, data = train)

summary(regress\_Sales)

plot(regress\_Sales)

#Regression (total sales n ratings)

regress\_Rating = lm(formula = total\_sales ~ Rating, data = train)

summary(regress\_Rating)

plot(regress\_Rating, pch = 16, col = "blue")