install.packages("data.table")

library("data.table")

library("e1071")

#Reading the data

Data <- fread("finalData1.csv")

#Data Preparation

Data$VacantBuilding <- as.factor(Data$VacantBuilding)

Data$Num.Open.Violations <- as.factor(Data$Num.Open.Violations)

Data$AssessedValue <- as.factor(Data$AssessedValue)

Data$LandUse <- as.factor(Data$LandUse)

Data$YearBuilt <- as.factor(Data$YearBuilt)

#Categorizing owner occupied into buckets

Vector1 <- Data$Owner\_Occupied

Vector1 <- as.numeric(Vector1)

for(i in 1:14907)

{

if(Vector1[i]<0.25)

{

Vector1[i] <- "Very Low"

}

else if(Vector1[i]<0.5)

{

Vector1[i] <- "Low"

}

else if(Vector1[i]<0.75)

{

Vector1[i] <- "High"

}

else if(Vector1[i]>0.74)

{

Vector1[i] <- "Very High"

}

}

Owner\_occupied\_Cat <- as.factor(Vector1)

#Categorizing H1-3 into buckets

Vector2 <- Data$H\_.1.3.P

Vector2 <- as.numeric(Vector2)

for(i in 1:14907)

{

if(Vector2[i]<0.25)

{

Vector2[i] <- "Very Low"

}

else if(Vector2[i]<0.5)

{

Vector2[i] <- "Low"

}

else if(Vector2[i]<0.75)

{

Vector2[i] <- "High"

}

else if(Vector2[i]>0.74)

{

Vector2[i] <- "Very High"

}

}

H1\_3\_Cat <- as.factor(Vector2)

#Categorizing H4 into buckets

Vector3 <- Data$H.4..P

Vector3 <- as.numeric(Vector3)

for(i in 1:14907)

{

if(Vector3[i]<0.25)

{

Vector3[i] <- "Very Low"

}

else if(Vector3[i]<0.5)

{

Vector3[i] <- "Low"

}

else if(Vector3[i]<0.75)

{

Vector3[i] <- "High"

}

else if(Vector3[i]>0.74)

{

Vector3[i] <- "Very High"

}

}

H4\_Cat <- as.factor(Vector3)

#Categorizing Total taxes owed into buckets

Vector4 <- Data$Total.Taxes.Owed

Vector4 <- as.numeric(Vector4)

for(i in 1:14907)

{

if(Vector4[i]<2000)

{

Vector4[i] <- 0

}

else if(Vector4[i]<4000)

{

Vector4[i] <- 2000

}

else if(Vector4[i]<6000)

{

Vector4[i] <- 4000

}

else if(Vector4[i]<8000)

{

Vector4[i] <- 6000

}

else if(Vector4[i]<10000)

{

Vector4[i] <- 8000

}

else if(Vector4[i]<20000)

{

Vector4[i] <- 10000

}

else if(Vector4[i]<30000)

{

Vector4[i] <- 20000

}

else if(Vector4[i]<40000)

{

Vector4[i] <- 30000

}

else if(Vector4[i]<50000)

{

Vector4[i] <- 40000

}

else if(Vector4[i]<100000)

{

Vector4[i] <- 50000

}

else if(Vector4[i]>99999)

{

Vector4[i] <- 100000

}

}

Total\_taxes\_Cat <- as.factor(Vector4)

#Categorizing crimes into buckets

Vector5 <- Data$Total\_crimes

Vector5 <- as.numeric(Vector5)

Vector5[11759]

for(i in 1:14907)

{

if(Vector5[i]<6)

{

Vector5[i] <- 5

}

else if(Vector5[i]<16)

{

Vector5[i] <- 15

}

else if(Vector5[i]<30)

{

Vector5[i] <- 30

}

else if(Vector5[i]>29)

{

Vector5[i] <- 45

}

}

Total\_crimes\_Cat <- as.factor(Vector5)

# Combining the categorized fields with the dataframe

Data <- data.frame(Data, Total\_taxes\_Cat,Total\_crimes\_Cat, H1\_3\_Cat, H4\_Cat,Owner\_occupied\_Cat)

# Preparing the training data

data\_yes <- Data[Data$VacantBuilding== "Y",]

data\_no <- Data[Data$VacantBuilding == "N",]

#Randomizing the rows

data\_no <- data\_no[sample(nrow(data\_no)),]

data\_no\_equal <- data\_no[1:746,]

data\_final <- rbind(data\_yes,data\_no\_equal)

data\_final <- data\_final[sample(nrow(data\_final)),]

#Training and test data

data\_test <- data\_final[1:1268,]

data\_train <- data\_final[1269:1492,]

colnames((data\_test))

#Building the model

Naive\_Bayes\_Model=naiveBayes(VacantBuilding ~ Num.Open.Violations + Total\_crimes\_Cat , data= data\_train)

nb.pred <- predict(Naive\_Bayes\_Model, data\_test)

#Showing the confusion matrix

table(data\_test$VacantBuilding,nb.pred)