**Water Quality Analysis in R**

**Code:**

library(ggplot2)

library(dplyr)

library(tidyverse)

library(tidyr)

library(caret)

library(Hmisc)

data<-read.csv("water\_potability.csv")

head(data)

data<-data[rowSums(is.na(data)) == 0,]

counts<-table(data$Potability)

barplot(counts, main="Distribution of Unsafe and Safe Water",xlab="Potability",col=counts)

ggplot(data,aes(x=ph))+geom\_histogram(aes(fill=factor(Potability)),position ="identity")+labs(fill="Potability",y="Count")

ggplot(data,aes(x=Hardness))+geom\_histogram(aes(fill=factor(Potability)),position ="identity",binwidth=5)+labs(fill="Potability",y="Count")

ggplot(data,aes(x=Solids))+geom\_histogram(aes(fill=factor(Potability)),position ="identity")+labs(fill="Potability",y="Count")

ggplot(data,aes(x=Chloramines))+geom\_histogram(aes(fill=factor(Potability)),position ="identity")+labs(fill="Potability",y="Count")

ggplot(data,aes(x=Sulfate))+geom\_histogram(aes(fill=factor(Potability)),position ="identity")+labs(fill="Potability",y="Count")

ggplot(data,aes(x=Conductivity))+geom\_histogram(aes(fill=factor(Potability)),position ="identity")+labs(fill="Potability",y="Count")

ggplot(data,aes(x=Organic\_carbon))+geom\_histogram(aes(fill=factor(Potability)),position ="identity")+labs(fill="Potability",y="Count")

ggplot(data,aes(x=Trihalomethanes))+geom\_histogram(aes(fill=factor(Potability)),position ="identity")+labs(fill="Potability",y="Count")

ggplot(data,aes(x=Turbidity))+geom\_histogram(aes(fill=factor(Potability)),position ="identity")+labs(fill="Potability",y="Count")

mat<-data.matrix(data,rownames.force = NA)

correlation<-rcorr(mat)

correlation

allmodels<-paste(names(getModelInfo()), collapse=', ')

allmodels

modelLookup("rf")

modelLookup("earth")

model = train(Potability ~ ., data = data, method='rf')

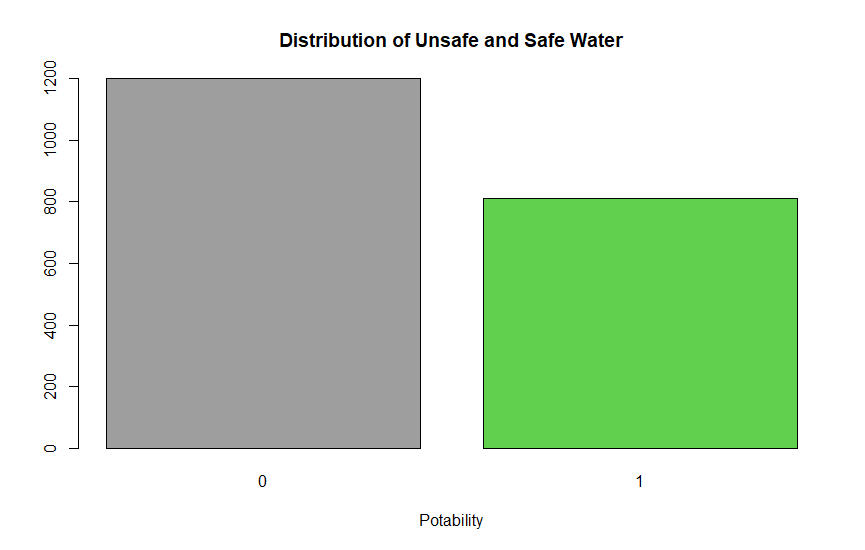
model

plot(model)

predicted <- predict(model,data)

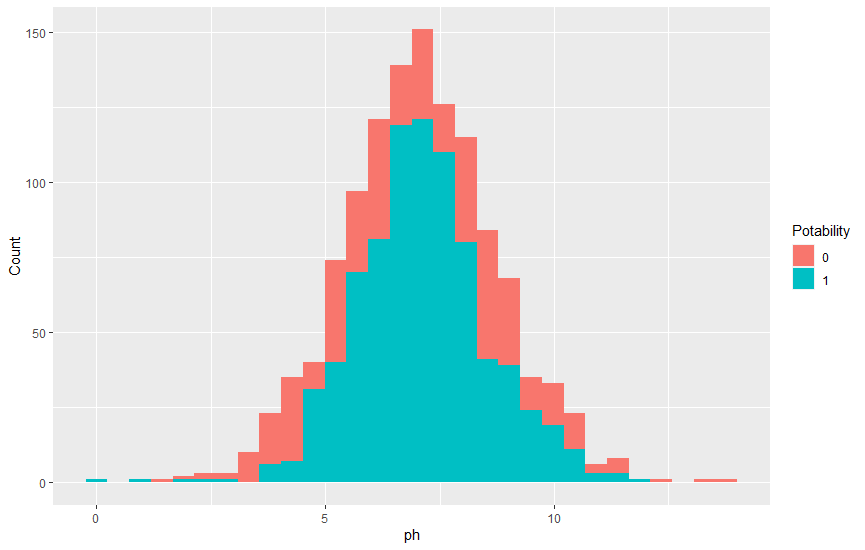
predicted

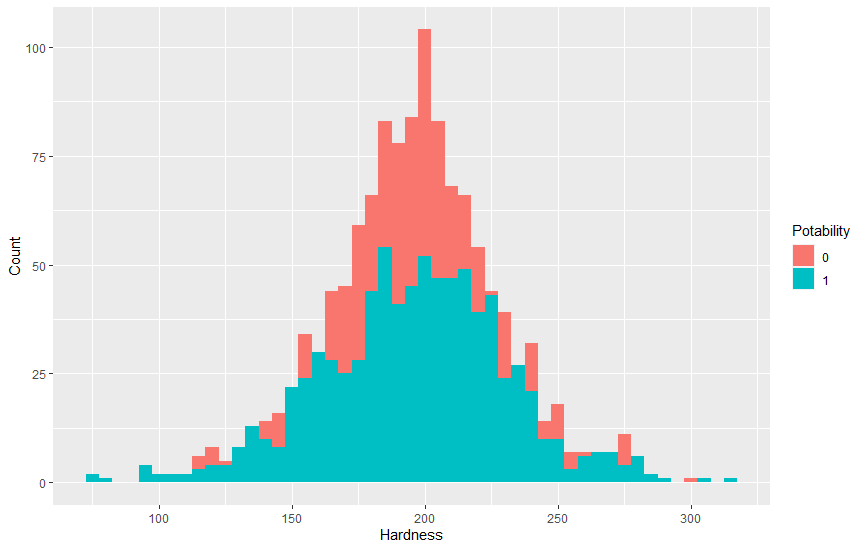
**Screen Shots of Results:**

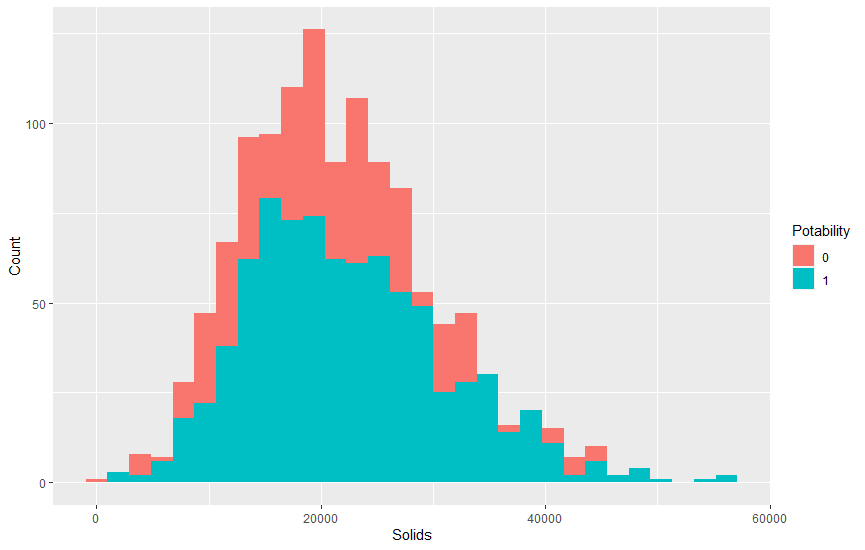
****

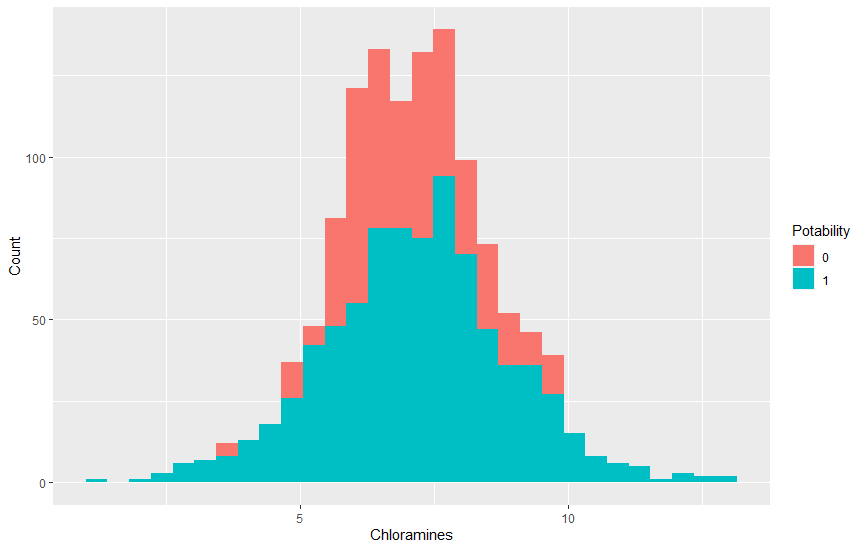
40 % of samples collected are not potable while 60% of records in data set are potable.

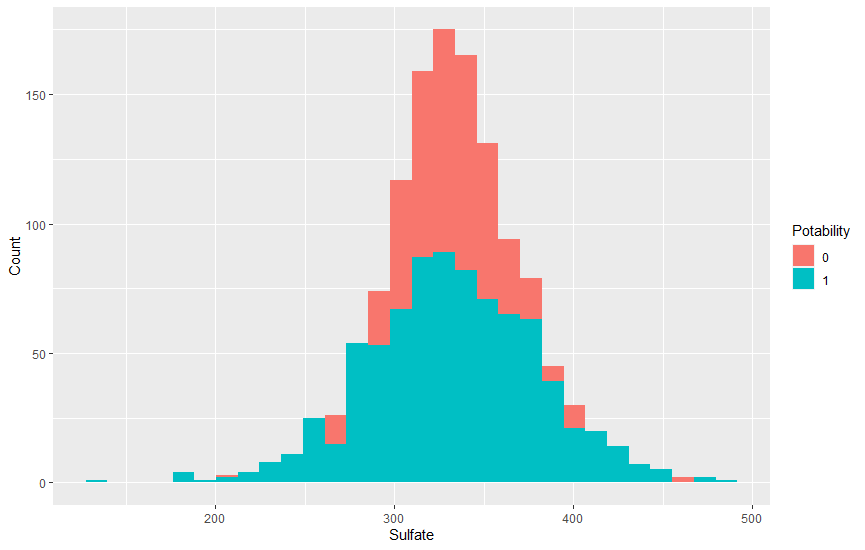
**Following 9 Graphs Represents Count of Records of drinkanle and undrinkable water based on attributes of ph, hardness, chloramine, solids, sulfate, conductivity, trihalomethane, organic carbon, turbidity.**

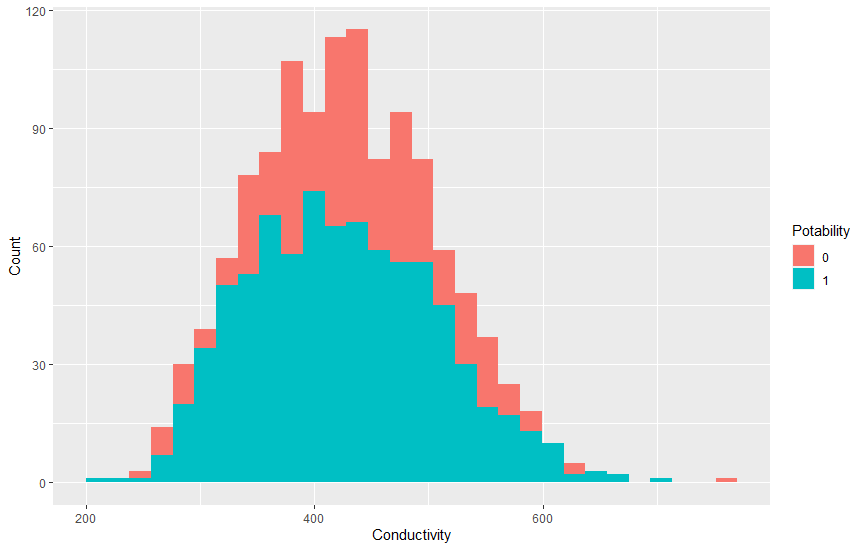
****

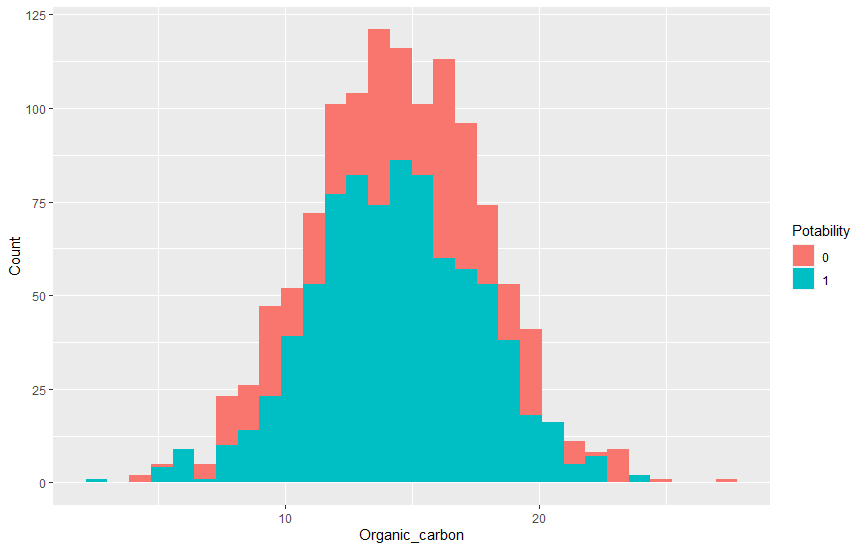
****

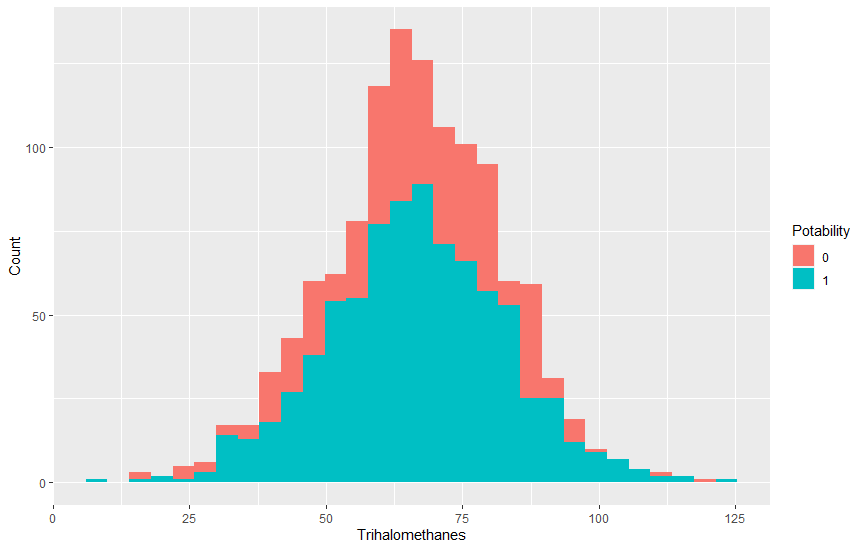
****

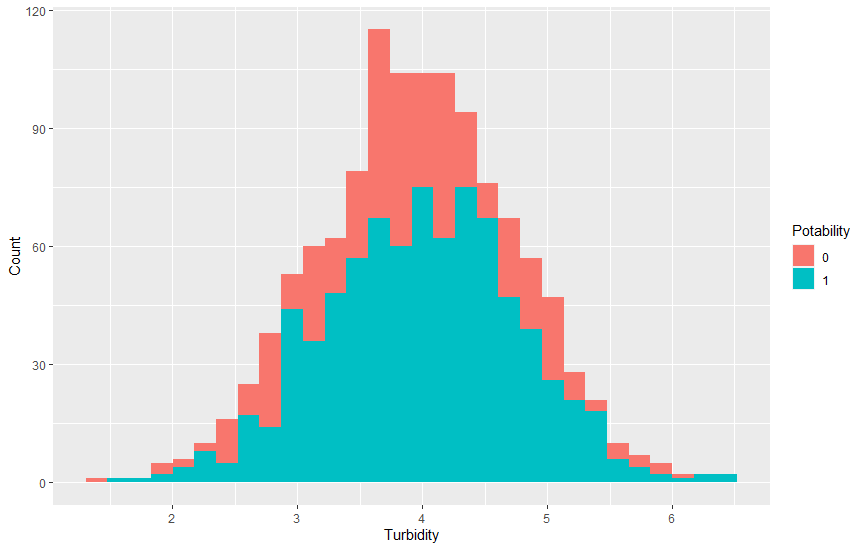
****

****

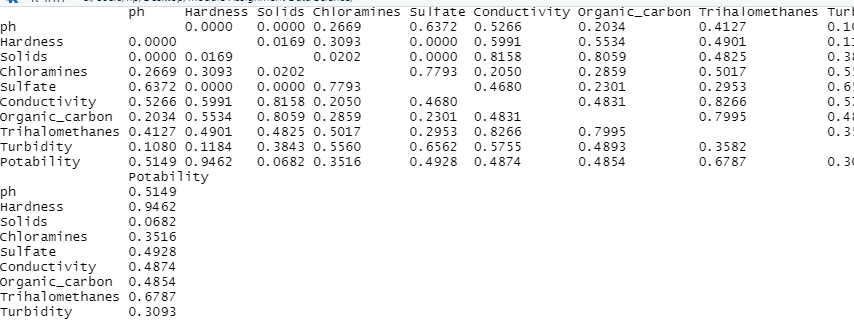
****

****

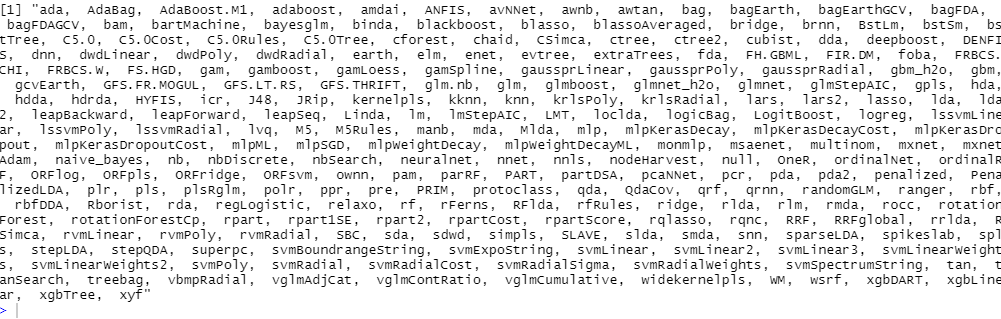
****

****

**Correlation matrix of all parameters:**

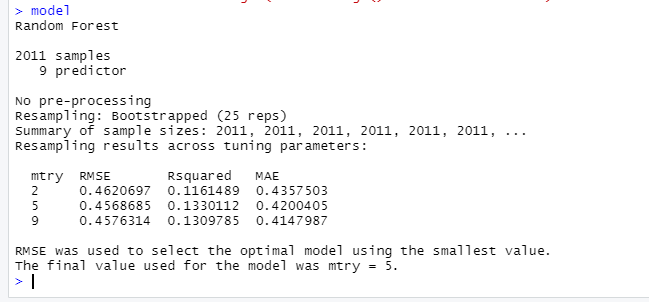
****

**Names of all Machine Learning algorithms in caret package:**

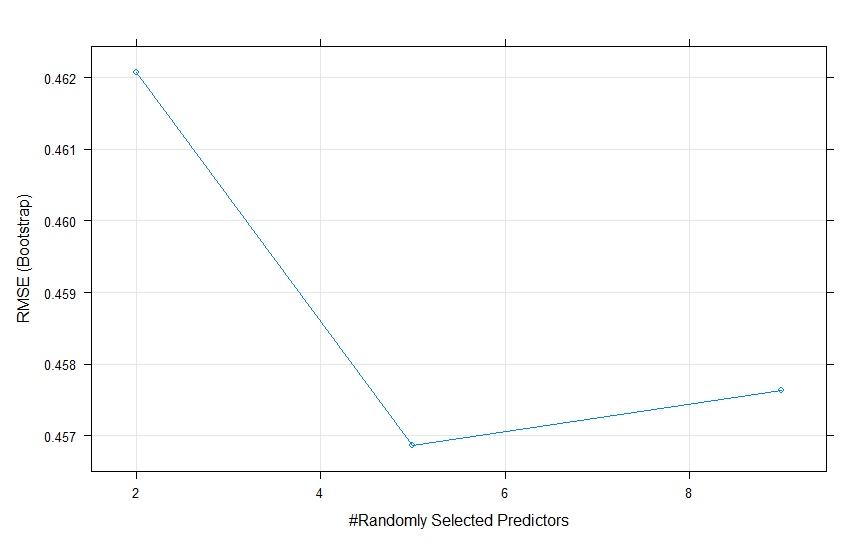
****

**C:\Users\hp\Desktop\Water_Quality _Analysis\WaterQuality\Capture3.PNG**

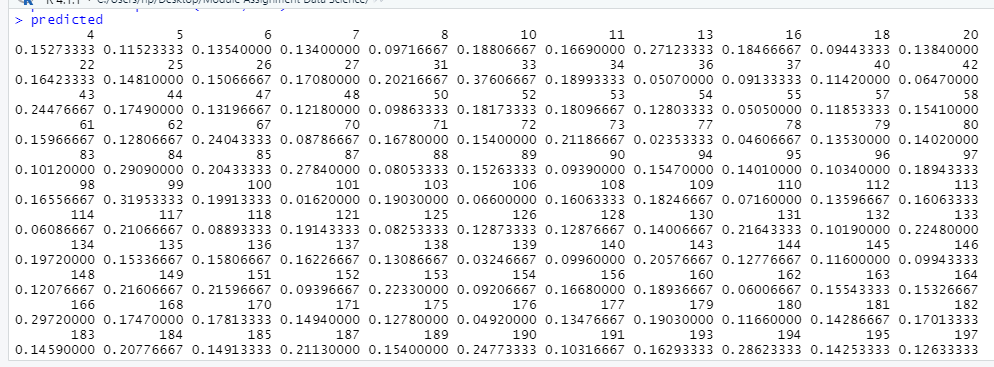
**Creating the training model for our data:**

****

**Plotting the model:**

****

**Predicting from the model:**

****