

Answer:

X1\_pred\_train<- predict(zip\_Model, zipTrain)

> X1\_actual\_train = zipTrain$X1

> X1\_train <- sqrt(mean((X1\_actual\_train - X1\_pred\_train)^2)) #RMSE Train

> print(X1\_train)

[1] 0.1575174

>

> X1\_predict<- predict(zip\_Model, zipTest)

> X1\_actual = zipTest$X1

> X1\_test <- sqrt(mean((X1\_actual - X1\_predict)^2)) #RMSE Test

> print(X1\_test)

[1] 0.3894424

>

>

> # Part 2: K-nearest neighbor

>

> library(class)

>

> knn\_1\_train<-knn(train=zipTrain[,2:257],test=zipTrain[,2:257],cl = as.factor(zipTrain$X1),k=1) # Knn model for K=1

> knn\_1\_pred <- as.numeric(as.character(knn\_1\_train))

> knn\_act = zipTrain$X1

> RMSE\_1\_train <- sqrt(mean((knn\_act - knn\_1\_pred)^2)) #RSME

> print(RMSE\_1\_train)

[1] 0

>

> knn\_1<-knn(train=zipTrain[,2:257],test=zipTest[,2:257],cl = as.factor(zipTrain$X1),k=1) # Knn model for K=1

> knn\_1\_predict <- as.numeric(as.character(knn\_1))

> knn\_actual = zipTest$X1

> RMSE\_1\_test <- sqrt(mean((knn\_actual - knn\_1\_predict)^2)) #RSME

> print(RMSE\_1\_test)

[1] 0.1572427

>

> knn\_3\_train<-knn(train=zipTrain[,2:257],test=zipTrain[,2:257],cl = as.factor(zipTrain$X1),k=3) # Knn model for K=3

> knn\_3\_pred <- as.numeric(as.character(knn\_3\_train))

> RMSE\_3\_train <- sqrt(mean((knn\_act - knn\_3\_pred)^2)) #RSME

> print(RMSE\_3\_train)

[1] 0.07099012

>

> knn\_3<-knn(train=zipTrain[,2:257],test=zipTest[,2:257],cl = as.factor(zipTrain$X1),k=3) # Knn model for K=3

> knn\_3\_predict <- as.numeric(as.character(knn\_3))

> RMSE\_3\_test<-sqrt(mean((knn\_actual - knn\_3\_predict)^2)) #RSME

> print(RMSE\_3\_test)

[1] 0.1738384

>

> knn\_5\_train<-knn(train=zipTrain[,2:257],test=zipTrain[,2:257],cl = as.factor(zipTrain$X1),k=5) # Knn model for K=5

> knn\_5\_pred <- as.numeric(as.character(knn\_5\_train))

> RMSE\_5\_train <- sqrt(mean((knn\_act - knn\_5\_pred)^2)) #RSME

> print(RMSE\_5\_train)

[1] 0.07589163

>

> knn\_5<-knn(train=zipTrain[,2:257],test=zipTest[,2:257],cl = as.factor(zipTrain$X1),k=5) # Knn model for K=5

> knn\_5\_predict <- as.numeric(as.character(knn\_5))

> RMSE\_5\_test<-sqrt(mean((knn\_actual - knn\_5\_predict)^2)) #RSME

> print(RMSE\_5\_test)

[1] 0.1738384

>

> knn\_7\_train<-knn(train=zipTrain[,2:257],test=zipTrain[,2:257],cl = as.factor(zipTrain$X1),k=7) # Knn model for K=7

> knn\_7\_pred <- as.numeric(as.character(knn\_7\_train))

> RMSE\_7\_train <- sqrt(mean((knn\_act - knn\_7\_pred)^2)) #RSME

> print(RMSE\_7\_train)

[1] 0.08049523

>

> knn\_7<-knn(train=zipTrain[,2:257],test=zipTest[,2:257],cl = as.factor(zipTrain$X1),k=7) # Knn model for K=7

> knn\_7\_predict <- as.numeric(as.character(knn\_7))

> RMSE\_7\_test<-sqrt(mean((knn\_actual - knn\_7\_predict)^2)) #RSME

> print(RMSE\_7\_test)

[1] 0.1815683

>

> knn\_15\_train<-knn(train=zipTrain[,2:257],test=zipTrain[,2:257],cl = as.factor(zipTrain$X1),k=15) # Knn model for K=15

> knn\_15\_pred <- as.numeric(as.character(knn\_15\_train))

> RMSE\_15\_train <- sqrt(mean((knn\_act - knn\_15\_pred)^2)) #RSME

> print(RMSE\_15\_train)

[1] 0.09674322

>

> knn\_15<-knn(train=zipTrain[,2:257],test=zipTest[,2:257],cl = as.factor(zipTrain$X1),k=15) # Knn model for K=15

> knn\_15\_predict <- as.numeric(as.character(knn\_15))

> RMSE\_15\_test<-sqrt(mean((knn\_actual - knn\_15\_predict)^2)) #RSME

> print(RMSE\_15\_test)

[1] 0.1961161