Meghana_Nadig_Assignment2

Assignment 2 Meghana B Nadig

Introduction to Data Mining / Machine Learning (DA 5030) NUID : 001236144

1.

```
m <- mean(USArrests$Assault)
sd <- sd(USArrests$Assault)
View(USArrests)
USArrests$z <- (USArrests$Assault - m) / sd
View(z)
USArrests$z <- abs(USArrests$z)
Filter_USArrests <- USArrests[USArrests$z > 1.5,]
View(Filter_USArrests)

#Hence the state which is a outlier is Florida, Maryland, North Carolina,
North Dakota.
```

2.

```
Mur <- USArrests$Murder
Ass <- USArrests$Assaults
var <- USArrests[,-3:-6]
View(var)
cor(var,method = "pearson")
install.packages("ggplot2")
library("ggplot2")
ggplot(USArrests,aes(Murder,Assault))+geom_line()
ggplot(USArrests,aes(USArrests$Murder, USArrests$Assault))+geom_line()


# There is a high positive correlation between the two of 0.8
```

3.

```
DataAssginment2 <- read.table("C:/Users/Meghana
Nadig/Desktop/Assignment_2.csv",sep = ",",header = TRUE)
View(DataAssginment2)
```

#Calculating Moving Avg tail(DataAssginment2) n <- nrow(DataAssginment2) View(n) last_3 <- DataAssginment2[n:(n-2),2] View(last_3) sma <- mean(last_3)

#Calculating Weighted Avg w <- c(4,1,1) View(w) sw <- w*last_3 View(sw) F <- sum(sw)/sum(w) wma <- F View(F)

#Calculating Exponential Smoothning DataAssginment2Ft < −0DataAssginment2E <- 0 DataAssginment2$Ft[1] <- DataAssginment2[1,2] for(i in 2:nrow(DataAssginment2)) { DataAssginment2$Ft[i] <- DataAssginment2Ft[i − 1] + 0.2 ∗ DataAssginment2E[i-1] DataAssginment2E[i] < −DataAssginment2[i, 2] − DataAssginment2Ft[i] } View(DataAssginment2)
n <- nrow(DataAssginment2) res <- DataAssginment2Ft[n] + 0.2 ∗ DataAssginment2E[n] es <- res print(res)

#Line Regression Forescating library(ggplot2) ggplot(DataAssginment2,aes(DataAssginment2Year, DataAssginment2Subscribers))+geom_line() model <- lm(DataAssginment2Subscribers DataAssginment2Year) print(model) var <- -3.666e+10 + 1.828e+07*2017 lrt <- var print(var)

The Forecast for the year 2017 is 210760000.

4.

#SMA DataAssginment2Ft < −0DataAssginment2E <- 0 DataAssginment2Ft[1] < −DataAssginment2Subscribers[1] DataAssginment2Ft[2] < −mean(DataAssginment2Subscribers[1]) DataAssginment2Ft[3] < −mean(DataAssginment2Subscribers[1:2])

DataAssginment2E[1] < −0DataAssginment2E[2] <- 0 DataAssginment2E[3] < −DataAssginment2Subscribers[3] - DataAssginment2$Ft[3] for (i in 4:nrow(DataAssginment2)) { last_3 <- DataAssginment2[(i-1):(i-3),2] DataAssginment2$Ft[i] <- mean(last_3) DataAssginment2E[i] < −DataAssginment2Subscribers[i] - DataAssginment2$Ft[i]

}

```
View(DataAssginment2)
result <- DataAssginment2$E^2
mean(result)
  #1.17563e+15 <- value for SMA


#WMA
DataAssginment2$Ft <- 0
DataAssginment2$E <- 0
w <- c(4,1,1)
DataAssginment2$Ft[1] <- DataAssginment2$Subscribers[1]
DataAssginment2$Ft[2] <- (4*DataAssginment2$Subscribers[1])/4
```

```
DataAssginment2$Ft[3] <-
(DataAssginment2$Subscribers[1])+(4*DataAssginment2$Subscribers[2])/5

DataAssginment2$E[1] <- 0
DataAssginment2$E[2] <- DataAssginment2$Subscribers[2] -
DataAssginment2$Ft[2]
DataAssginment2$E[3] <- DataAssginment2$Subscribers[3] -
DataAssginment2$Ft[3]
for (i in 4:nrow(DataAssginment2))
{
  last_3 <- DataAssginment2[(i-1):(i-3),2]
  sw <- w*last_3
  DataAssginment2$Ft[i] <- sum(sw)/sum(w)
  DataAssginment2$E[i] <- DataAssginment2$Subscribers[i] -
DataAssginment2$Ft[i]

}

View(DataAssginment2)
result <- DataAssginment2$E^2
mean(result)
  #7.091546e+14 <- value for WMA



#Exponential Smoothning
DataAssginment2$Ft <- 0
DataAssginment2$E <- 0
DataAssginment2$Ft[1] <-  DataAssginment2[1,2]
for(i in 2:nrow(DataAssginment2))
{
  DataAssginment2$Ft[i] <- DataAssginment2$Ft[i-1]+0.2*DataAssginment2$E[i-1]
  DataAssginment2$E[i] <- DataAssginment2[i,2] - DataAssginment2$Ft[i]
}
View(DataAssginment2)
result <- DataAssginment2$E^2
mean(result)
  # 3.166964e+15 <- value for Exponential smoothning

#Linear Regression trend Line
DataAssginment2$Ft <- 0
DataAssginment2$E <- 0
for(i in 1:nrow(DataAssginment2))
{
  model <- lm(DataAssginment2$Subscribers[i] ~ DataAssginment2$Year[i])
  var <- -3.666e+10 + 1.828e+07*DataAssginment2$Year[i]
  DataAssginment2$Ft <- var
  DataAssginment2$E[i] <- DataAssginment2[i,2] - DataAssginment2$Ft[i]
}
```

```
View(DataAssginment2)
result <- DataAssginment2$E^2
mean(result)
  # 1.773756e+14 <- value for linear regression trend line
```

5. Simple Moving average model has the smallest MSE.

6. wt <- c(3,2,1) res <- c(lrt,es,wma) avg <- sum(wt*res)/sum(wt) View(avg)