**Week -7**

**48.** **Case study on Text Mining**

# 1. Case Study on Text Mining

install.packages("tm")

install.packages("wordcloud")

install.packages("RColorBrewer")

data <- Corpus(VectorSource(c("Text mining is fun", "R is great for text analysis", "Data Science is amazing")))

data

data\_clean <- tm\_map(data, content\_transformer(tolower))

data\_clean

dtm <- TermDocumentMatrix(data\_clean)

dtm

m <- as.matrix(dtm)

m

word\_freqs <- sort(rowSums(m),decreasing=TRUE)

word\_freqs

x=wordcloud(names(word\_freqs), word\_freqs,colors=brewer.pal(8, "Dark2"))

x

**49. Visualize data using any plotting framework**

plot(1:10, rnorm(10), type = "b", col = "blue", main = "Basic R Plot")

**50. Study and implementation of Data Visualization with ggplot2**

df <- data.frame(x = 1:10, y = rnorm(10))

ggplot(df, aes(x, y)) + geom\_point() + geom\_line() + ggtitle("ggplot2 Visualization")

**51. Load and Clean Data in R**

**data <- data.frame(ID = c(1, 2, 3, 4, NA), Name = c("A", "B", "C", "D", "E"))**

**data <- na.omit(data) # Removing NA values**

**data**

**52. Identify and Remove Duplicate Records using R**

duplicated\_data <- data.frame(ID = c(1, 2, 2, 3, 4), Value = c(10, 20, 20, 30, 40))

unique\_data <- duplicated\_data[!duplicated(duplicated\_data), ]

unique\_data

**53. Filter Rows Based on Multiple Conditions using R.**

duplicated\_data <- data.frame(ID = c(1, 2, 2, 3, 4), Value = c(10, 20, 20, 30, 40))

unique\_data <- duplicated\_data[!duplicated(duplicated\_data), ]

unique\_data

filtered\_data <- subset(unique\_data, ID > 1 & Value < 40)

filtered\_data

**54. Group Data and Calculate Aggregate.**

install.packages("dplyr")

aggr\_data <- duplicated\_data %>% group\_by(ID) %>% summarise(Total = sum(Value))

aggr\_data

**55. Merge Multiple Datasets using R**

data1 <- data.frame(ID = 1:3, Value1 = c(10, 20, 30))

data2 <- data.frame(ID = 2:4, Value2 = c(15, 25, 35))

merged\_data <- merge(data1, data2, by = "ID", all = TRUE)

merged\_data

**Assessment-7**

**Network Analysis:**

install.packages("igraph")

g <- graph(edges = c(1,2, 2,3, 3,4, 4,1, 2,4), directed = FALSE)

plot(g, vertex.size=degree(g) \* 5, vertex.color = "lightblue", main = "Network Graph")