**DATA MINING (SENTIMENT ANALYSIS)**

#data mining

#install.packages("tm")

#install.packages("wordcloud")

library(tm)

library(wordcloud)

getwd()

setwd("C:\\Users\\10ani\\Desktop\\DS\_classes")

comments=read.csv("comments.csv")

myCorpus=Corpus(VectorSource(comments$Comments))

class(myCorpus)

myCorpus[[1]][1] #convert to lower case

myCorpus[[1]][2]

strwrap(myCorpus)

myCorpus=tm\_map(myCorpus,content\_transformer(tolower))

myCorpus=tm\_map(myCorpus,removeNumbers) #remove numbers

myCorpus=tm\_map(myCorpus,removeWords,stopwords("english")) #remove stopwords

myCorpus=tm\_map(myCorpus,removePunctuation) #remove punctuations

myCorpus=tm\_map(myCorpus,stemDocument) #carry out steming

myCorpus=tm\_map(myCorpus,removeWords,c("get","took","gave")) #remove selected stopwords

myCorpus=tm\_map(myCorpus,stripWhitespace) #remove white spaces

head(myCorpus,13) #it is not a dataframe

tdm=TermDocumentMatrix(myCorpus)

mat=as.matrix(tdm)

v=sort(rowSums(mat),decreasing = TRUE)

d=data.frame(word=names(v),freq=v)

wordcloud(d$word,d$freq)

#skip random order

wordcloud(d$word,d$freq,random.order = FALSE,rot.per = 0.3)

#control location

wordcloud(d$word,d$freq,random.order = FALSE,rot.per = 0.3,scale=c(4,0.5),max.words = 15,colors = brewer.pal(8,"Dark2"))

title (main="WORD CLOUD")

#SENTIMENT ANALYSIS

getwd()

setwd("C:\\Users\\10ani\\Desktop\\DS\_classes")

comments=read.csv("comments.csv")

library(tm)

comments1=comments

neg\_words=read.csv("neg\_words.csv")

pos\_words=read.csv("pos\_words.csv")

#neg\_words=neg\_words("llloocaattiioonn",what="character")---to run and read data set

#install.packages("stringr")

library(stringr)

#install.packages("plyr")

library(plyr)

#REMOVE PUNCTUATIONS,NUMBERS(gsub=changing one words with other)

comments1$Comments =gsub("[[:punct:]]"," ",comments1$Comments)

comments1$Comments= gsub("[[:digit:]]"," ",comments1$Comments)

comments1$Comments= gsub("[^[:alnum:]]"," ",comments1$Comments)

comments1$Comments= gsub("xxxx"," ",comments1$Comments)

comments1$Comments= tolower(comments1$Comments)

for(i in 1:nrow(comments1)){

comments1$NN[i]=sum(!is.na(match(unlist(strsplit(comments1$Comments[i]," ")),neg\_words$X2.faced)))

}

#comments1$neg\_words=sum(!is.na(match(unlist(strsplit(comments1$Comments[i]," ")),neg\_words$X2.faced)))

#sum(!is.na(match(unlist(strsplit(comments1$Comments[1]," ")),"call")))

for(i in 1:nrow(comments1)){

comments1$PP[i]=sum(!is.na(match(unlist(strsplit(comments1$Comments[i]," ")),pos\_words$a.)))

}

comments1$pos\_words=NULL

comments1$Polarity=ifelse(comments1$PP>comments1$NN,"pos",ifelse(comments1$NN>comments1$PP,"neg","neutral"))

---------------------------

# CHECK WORD IN DETAILS

---------------------------

install.packages("syuzhet")

library(syuzhet)

getwd()

setwd("C:\\Users\\10ani\\Desktop\\DS\_classes")

apple=read.csv("apple.csv")

library(tm)

tweet=iconv(apple$text)

sentiment=get\_nrc\_sentiment(tweet)

head(sentiment)

get\_nrc\_sentiment("merry")

get\_nrc\_sentiment("sudden")

get\_nrc\_sentiment("i am a player")

get\_nrc\_sentiment("he is the best")

colSums(sentiment)

barplot(colSums(sentiment),main = "sentiment analysis",ylab = "count",xlab = "emotion",col=rainbow(30))