Code in python shell:

import numpy as np

import pandas as pd

import time

import nltk

import matplotlib.pyplot as plt

import matplotlib.pyplot as plt

from nltk.corpus import stopwords

from sklearn.model\_selection import train\_test\_split

from sklearn import naive\_bayes

from sklearn.metrics import roc\_auc\_score

from sklearn.feature\_extraction.text import TfidfVectorizer

df=pd.read\_csv(r"C:\Users\user\Desktop\mini project\youtube\yt.csv")

df.head()

stopset=set(stopwords.words("english"))

vectorizer=TfidfVectorizer(use\_idf=True, lowercase=True, strip\_accents='ascii',stop\_words=stopset)

y=df.LikeCount

x=vectorizer.fit\_transform(df.Comment)

y.shape

x.shape

x\_train,x\_test,y\_train,y\_test=train\_test\_split(x,y,test\_size=0.2)

clf=naive\_bayes.MultinomialNB()

clf.fit(x\_train,y\_train)

roc\_auc\_score(y\_test,clf.predict\_proba(x\_test)[:,1])

plt.rcdefaults()

object = ('positive', 'negative')

y\_pos = np.arange(len(object))

performance = [0.75,0.82]

plt.bar(y\_pos,performance, align='center',alpha=0.4)

plt.xticks(y\_pos,object)

plt.ylabel('accuracy')

plt.title('youtube reviews')

plt.show()

print('success')code in R studio:

library(SocialMediaLab)

# Google developer API key

apikey <- "AIzaSyC1BgCMWNVP3yynFMP9FRO7miNktCohyhk"

key <- AuthenticateWithYoutubeAPI(apikey)

# Collect data using Youtube video IDs

video <- c('5eDqRysaico', 'dJclNIN-TPo')

ytdata <- CollectDataYoutube(video, key, writeToFile = FALSE)

str(ytdata)

write.csv(ytdata, file='C:/Users/user/Desktop/result/yt9.csv', row.names = F)

# Read Youtube data file

data <- read.csv(file.choose(), header = T)

str(data)

data <- data[data$ReplyToAnotherUser != FALSE,]

y <- data.frame(data$User, data$ReplyToAnotherUser)

## Create user network

library(igraph)

net <- graph.data.frame(y, directed = T)

net <- simplify(net)

V(net)

E(net)

V(net)$label <- V(net)$name

V(net)$degree <- degree(net)

# Histogram of node degree

hist(V(net)$degree,

col = 'green',

main = 'Histogram of the Node Degree',

ylab = 'Frequency',

xlab = 'Degree of vertices')

# Network diagram

plot(net,

vertex.size =0.2\*V(net)$degree,

edge.arrow.size = 0.1,

vertex.label.cex = 0.01\*V(net)$degree)

## sentiment analysis

library(naivebayes)

library(syuzhet)

library(plyr)

library(ggplot2)

# Read data file

data <- read.csv(file.choose(), header = T)

str(data)

comments <- iconv(data$Comment, to = 'utf-8')

# Obtain sentiment scores

s <- get\_nrc\_sentiment(comments)

head(s)

s$neutral <- ifelse(s$negative+s$positive==0, 1, 0)

head(s)

comments[218]

comments[2]

# Bar plot

barplot(100\*colSums(s)/sum(s),

las = 2,

col = rainbow(10),

ylab = 'Percentage',

main = 'Sentiment Scores foryoutube reviews')