CODE SNIPPETS for Data COLLECTION, MODELLING and ANALYSIS

1. Twitter Data Pull.R :
2. #loading library
4. library(rtweet)
6. #twitter authentication
7. twitter\_token <- create\_token(
8. app = "socialNM",
9. consumer\_key = "XmV6T42a2RembSjcDvMwN4fWk",
10. consumer\_secret = "LSyo6zcm9s17zv6fexhKsDZJKDiUeDnnCBIT5ZqiOthOAi9afm")
12. #data pull for multiple keywords
13. digitalization <- search\_tweets(q= "#digitalization", n=1000,
14. geocode = "20.593684,78.962880,1000mi")
16. digitalization1 <- search\_tweets(q= "#digitalindia", n=1000,
17. geocode = "20.593684,78.962880,1000mi")
19. makeinIndia <- search\_tweets(q= "#makeinindia", n=1000,
20. geocode = "20.593684,78.962880,1000mi")
22. pmkvy <- search\_tweets(q= "#PradhanMantriKaushalVikasYojana OR #pmkvy", n=1000,
23. geocode = "20.593684,78.962880,1000mi")
25. skillindia <- search\_tweets(q= "#skillindia OR #PMKK", n=2000,
26. geocode = "20.593684,78.962880,1000mi")
28. betibachao <- search\_tweets(q= "#betibachaobetipadhao", n=1000,
29. geocode = "20.593684,78.962880,1000mi")
31. education <- search\_tweets(q= "narendramodi+education", n=1000,
32. geocode = "20.593684,78.962880,1000mi")
34. youth <- search\_tweets(q= "narendramodi+youth", n=1000,
35. geocode = "20.593684,78.962880,1000mi")

38. employment <- search\_tweets(q= "narendramodi+employment", n=1000,
39. geocode = "20.593684,78.962880,1000mi")
41. sheet1 = rbind(digitalization,digitalization1)
43. # labelling data with topic
44. sheet1$Topic= "digitalization"
45. betibachao$Topic= "betibachao"
46. education$Topic= "education"
47. employment$Topic= "employment"
48. makeinIndia$Topic= "makeinIndia"
49. pmkvy$Topic= "pmkvy"
50. skillindia$Topic= "skillindia"
51. youth$Topic= "youth"
53. #combining all the dataframes and writing to CSV
54. final = rbind(digitalization,betibachao,education,employment,makeinIndia,pmkvy,youth,skillindia)
55. write.csv(final, "combined data.csv")
56. Sentiment An.R
57. #load libraries
58. **if** (!require(sentimentr)) {install.packages("Sentiment An.R")} # ensure java is up to date!
59. library(sentimentr)
60. library(dplyr)
61. library(tidytext)
62. library(tidyverse)
64. #read the data to be analysed
65. final = read.csv("C:/CBAcontent/Practicum/data collection/twitter data/combined data.csv")
66. textdf = data\_frame(text = as.character(final$text))
68. #get sentiment for each of the word token
69. new\_senti = as.vector(final$text) %>% get\_sentences() %>% sentiment()
71. #get average sentiment per tweet
72. senti\_per\_tweet = new\_senti  %>% select(element\_id,sentiment) %>%group\_by(element\_id) %>% summarise\_all(funs(mean))
73. write.csv(senti\_per\_tweet, "senti\_per\_tweet.csv")

3) Trends Data Pull

1. #load required libraries
2. library(gtrendsR)
3. library(maps)
4. library(ggplot2)
5. library(lettercase)
6. library(viridis)
7. library(pals)
8. library(scico)
9. library(ggrepel)
10. library(tidyverse)
12. #fetch data for required terms
13. term1 <- gtrends("digital india", time = "all", gprop = "web", geo = c("IN"))
14. term2 <- gtrends("digitalization", time = "all", gprop = "web", geo = c("IN"))
15. term3 <- gtrends("make in india", time = "all", gprop = "web", geo = c("IN"))
16. term4 <- gtrends("pmkvy", time = "all", gprop = "web", geo = c("IN"))
17. term5 <- gtrends("skill india", time = "all", gprop = "web", geo = c("IN"))
18. term6 <- gtrends("Pradhan Mantri Kaushal Vikas", time = "all", gprop = "web", geo = c("IN"))
19. term7 <- gtrends("beti bachao beti padhao", time = "all", gprop = "web", geo = c("IN"))
20. term8 <- gtrends("connect to PM", time = "all", gprop = "web", geo = c("IN"))
21. term9 <- gtrends("narendra modi", time = "all", gprop = "web", geo = c("IN"))
22. term10 <- gtrends("youth empowerment", time = "all", gprop = "web", geo = c("IN"))
23. #combine the data for required KPIs
24. interest\_over\_time = rbind(term1$interest\_over\_time,term2$interest\_over\_time,term3$interest\_over\_time,term4$interest\_over\_time,term5$interest\_over\_time,term6$interest\_over\_time,term7$interest\_over\_time,term8$interest\_over\_time,term9$interest\_over\_time,term10$interest\_over\_time)
25. related\_queries = rbind(term1$related\_queries,term2$related\_queries,term3$related\_queries,term4$related\_queries,term5$related\_queries,term6$related\_queries,term7$related\_queries,term8$related\_queries,term9$related\_queries,term10$related\_queries)
26. interest\_by\_region = rbind(term1$interest\_by\_region,term2$interest\_by\_region,term3$interest\_by\_region,term4$interest\_by\_region,term5$interest\_by\_region,term6$interest\_by\_region,term7$interest\_by\_region,term8$interest\_by\_region,term9$interest\_by\_region,term10$interest\_by\_region)
28. #write data to CSV
29. write.csv(interest\_over\_time, "interest\_over\_time.csv", row.names = FALSE, sep = ",")
30. write.csv(related\_queries, "related\_queries.csv",row.names = FALSE, sep = ",")
31. write.csv(interest\_by\_region, "interest\_by\_region.csv",row.names = FALSE, sep = ",")

4) News API data

1. **import** requests
2. **import** pandas as pd
4. data = requests.get("https://newsapi.org/v2/everything?apiKey=69daccc4fd664a7fb1944c5a17b8c29a&q=Digitalization").json()
6. articles = data["articles"]
8. #source-id , source-name, author, title, description, content
9. df = pd.DataFrame(columns = ["source-id" , "source-name", "author", "title", "description", "content"])
10. count = 0
11. **for** article **in** articles:
12. arr = []
13. arr.append(article["source"]["id"])
14. arr.append(article["source"]["name"])
15. arr.append(article["author"])
16. arr.append(article["title"])
17. arr.append(article["description"])
18. arr.append(article["content"])
19. df.loc[count] = arr
20. count += 1
22. df.head()
24. df.to\_csv("output\_api.csv")

Topic Analysis

1. source("https://raw.githubusercontent.com/ruchib19/NLP-using-UDPipe/master/dependency-nlp-an.R")
2. runGitHub("NLP-using-UDPipe","ruchib19")
4. source("https://raw.githubusercontent.com/sudhir-voleti/text-topic-analysis-shinyapp/master/dependency-text-topic-analysis-shinyapp.R")
5. runGitHub("text-topic-analysis-shinyapp","sudhir-voleti")