**Twitter Sentiment Analysis of Popular programming language Python**

Purpose

The purpose of this project is to apply learnings about Sentiment Analysis from IT518-Foundations of Data Science Course. Focuses on analyzing and understanding sentiments on popular programming language ‘Python’. In this study, data is collected and analyzed from recent twitter data related to the “python” to provide insights on how and where python is utilized in Data Science field.

Sentiment Analysis

Sentiment analysis is contextual mining of text which identifies and extracts information in source material and assisting a business to understand the social sentiment of their brand, product or service while monitoring online conversations.

**Steps 1**: To have access to the Twitter API, login to Twitter Developer website and create an application. After registering, create an access token and grab your application’s Consumer Key, Consumer Secret, Access token and Access token secret from Keys and Access Tokens tab.

**Step 2.** To Build corpus,

I will first use some packages in R. These are twitter, plyr, stringr and ggplot2. You can install these packages by the following commands:

A screenshot of a cell phone

Description automatically generated

Step 3: After twitter auth, we can fetch most recent tweets related to any keyword. I have used #python as python is mostly used for data analysis.

p\_tweets <- searchTwitter('python', n=1500, lang="en")

p\_tweets <- twListToDF(p\_tweets)

This command will get 1500 tweets related to Python. The function “searchTwitter” is used to download recent tweets. Now we need to convert this list of 1500 tweets into the data frame.

Step 4: Build a corpus and clean the corpus by removing punctuation, repeating words, links, numbers, and white spaces

p\_text <- p\_tweets$text

p\_tweets <- tolower(ag\_text)

p\_tweets <- gsub("rt", "", p\_tweets)

p\_tweets <- gsub("@\\w+", "", p\_tweets)

p\_tweets <- gsub("[[:punct:]]", "", p\_tweets)

p\_tweets <- gsub("http\\w+", "", p\_tweets)

p\_tweets <- gsub("[ |\t]{2,}", "", p\_tweets)

p\_tweets <- gsub("^ ", "", p\_tweets)

p\_tweets <- gsub(" $", "", p\_tweets)

p\_tweets <- gsub("and", "", p\_tweets)

p\_tweets <- gsub("the", "", p\_tweets)

p\_tweets <- gsub("[^\x01-\x7F]", "", p\_tweets)

p\_bd\_corpus = Corpus(VectorSource(p\_tweets))

# Inspect the corpus

inspect(p\_bd\_corpus[1:10])

p\_bd\_corpus\_bd\_clean <- tm\_map(p\_bd\_corpus, tolower)

p\_bd\_corpus\_bd\_clean <- tm\_map(p\_bd\_corpus, removePunctuation)

p\_bd\_corpus\_bd\_clean <- tm\_map(p\_bd\_corpus, removeNumbers)

p\_bd\_corpus\_bd\_clean <- tm\_map(p\_bd\_corpus, removeWords, stopwords("en"))

p\_bd\_corpus\_bd\_clean <- tm\_map(p\_bd\_corpus, stripWhitespace)

p\_bd\_corpus\_bd\_clean <- iconv(p\_bd\_corpus,to="utf-8-mac")

p\_bd\_corpus\_bd\_clean <- tm\_map(p\_bd\_corpus, removeWords, c("python", "Python"))

Step 5: Visualizing the tweets using Word Cloud

A black and silver text on a screen

Description automatically generated

A picture containing drawing

Description automatically generated

According to this wordcloud and Bar plot, we can see that Python is the most used term in the tweets followed by Bigdata, datascience, IOT, tenserflow, analytics which shows that while tweeting about python the person also connect the term python with the words like gdata, datascience, IOT, tenserflow.

References

# Shashank Gupta, Sentiment Analysis: Concept, Analysis and Applications <https://towardsdatascience.com/sentiment-analysis-concept-analysis-and-applications-6c94d6f58c17>, 2018