Homework 6

Text Processing and Sentiment Analysis of Twitter Data

Ομάδα G

Ευαγγέλου Σωτήρης 2159

Καλαής Κωσταντίνος 2146

Χατζηευφραιμίδης Λευτέρης 2209

UNIVERSITY OF THESSALY

Why Twitter Data?

- Twitter is a social media application with million messages per day, including huge amounts of information about almost all type of industries, lawmakers, celebrities etc.
- It provides easy access for collecting and analysing real time data, so we can do any sort of research of these real life situations and understand real user opinions, complaints and suggestions.

From Data Collection to Sentiment Analysis

Data Collecting

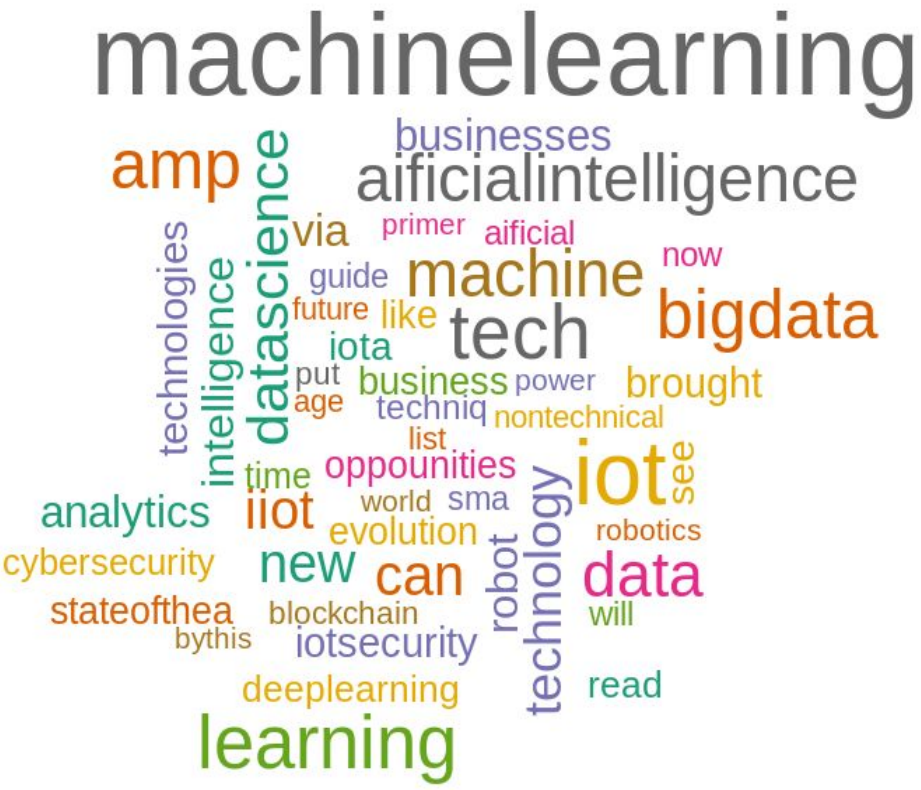
- Data in the form of raw tweets is retrieved by using "**twitteR**", an R package which provides access to the Twitter API. The API requires us to register a developer account with twitter and fill in parameters such as consumer_key, consumer_secret and access_token.

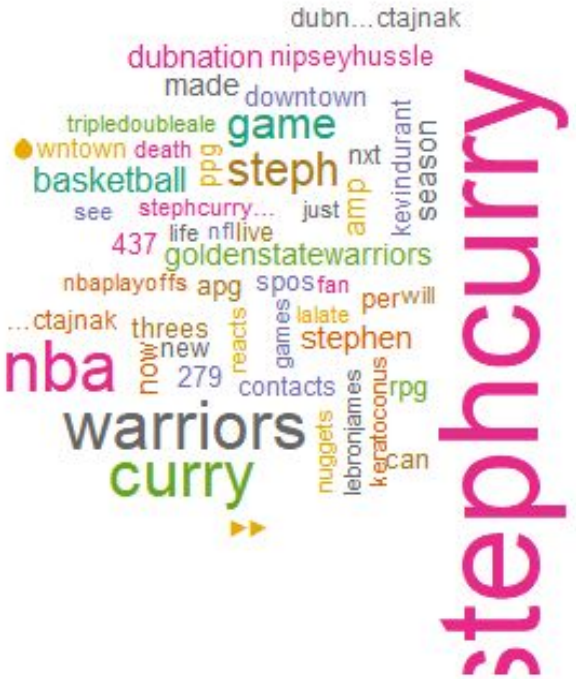
Data Pre-Processing

- Initially the accessed tweets are turned into dataframes to process.
- We pre-process the data and remove tabs, blank spaces, links etc.
- Lastly we remove **stop words** - a set of commonly used words in any language with no semantic meaning for our research - that deteriorate the analysis .

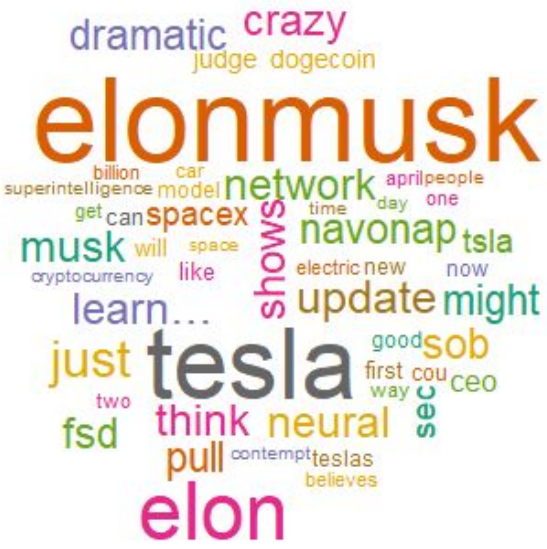
Word Clouds

Word clouds are unique diagrams used in sentiment analysis to show the most used words associated with a certain word, and hashtag in our situation.


machinelearning
businesses
amp
aificialintelligence
datascience
via
primer
aificial
technologies
intelligence
guide
machine
now
bigdata
future
like
tech
iota
put
business
power
brought
age
techniq
nontechnical
list
oppounities
iot
see
time
iiot
world
sma
analytics
evolution
robotics
technology
cybersecurity
new
can
robot
data
stateofthea
blockchain
will
bythis
iotsecurity
deeplearning
read
learning


dubn...ctajnak
dubnation
nipseyhussle
made
downtown
tripledoubleale
game
wntown
death
ppg
steph
nxt
basketball
kevindurant
season
see
stephcurry...
just
amp
437
life
nfllive
goldenstatewarriors
nbaplayoffs
apg
spos
fan
perwill
...ctajnak
threes
lalate
games
reacts
stephen
nba
now
new
279
contacts
rpg
warriors
nuggets
lebronjames
keratoconus
can
curry
stephcurry


archives…
defacedrediscovered
sitemovie
workshops
handson
def…
days
delighted
submit
year
unfounately
good
pictures
de…
smoke
defconpaies
2019
hacking
1995
first
hack
plz
talk
atalk
village27
schedules
early
announce
outbreak
one
especially
ever
pic
talks
gonna
live
inform
like
packet
ceasars
anyone
pay
friends
defcon
cigarettes
announcement
hackers
hack…


dramatic
crazy
judge
dogecoin
elonmusk
billion
car
superintelligence
model
network
aprilpeople
get
can
spacex
time
day
one
shows
navonap
musk
will
space
tsla
cryptocurrency
like
electric
new
now
learn…
update
might
just
tesla
good
sob
first
cou
ceo
way
two
sec
fsd
think
neural
pull
contempt
teslas
believes
elon

Sentiment Analysis

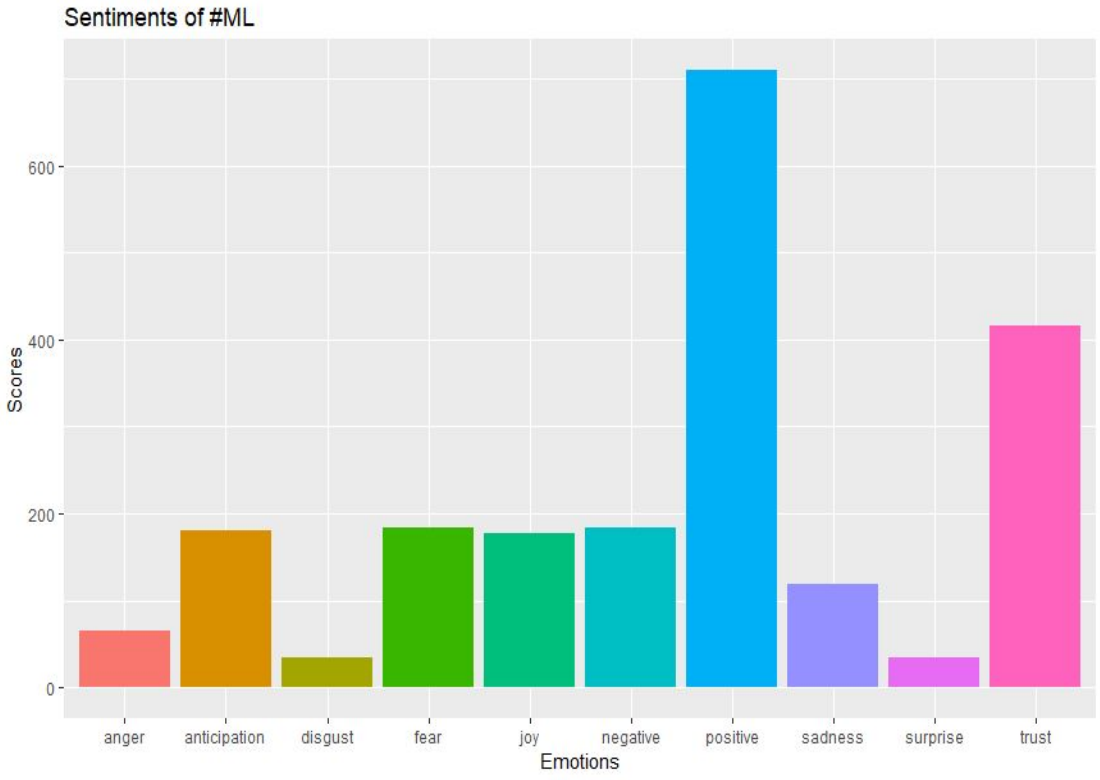
- Sentiment analysis is a type of data mining that measures the inclination of people’s opinions through natural language processing and text analysis, which are used to extract and analyze subjective information from the Web - mostly social media and similar sources. The analyzed data quantifies the general public's sentiments or reactions toward certain products, people or ideas.

- We have used the inbuilt sentiment analyzer in R, which uses the NRC sentiment dictionary to calculate the presence of eight different emotions and their corresponding valence in a text.

- The graphs contains a list of tweets in real time along with their sentiment score and we can see below some of these instances.
- These kind of graphs are used a lot in the recent years by businesses, politicians, celebrities etc. in order to control the masses and even manipulate them in some cituations.

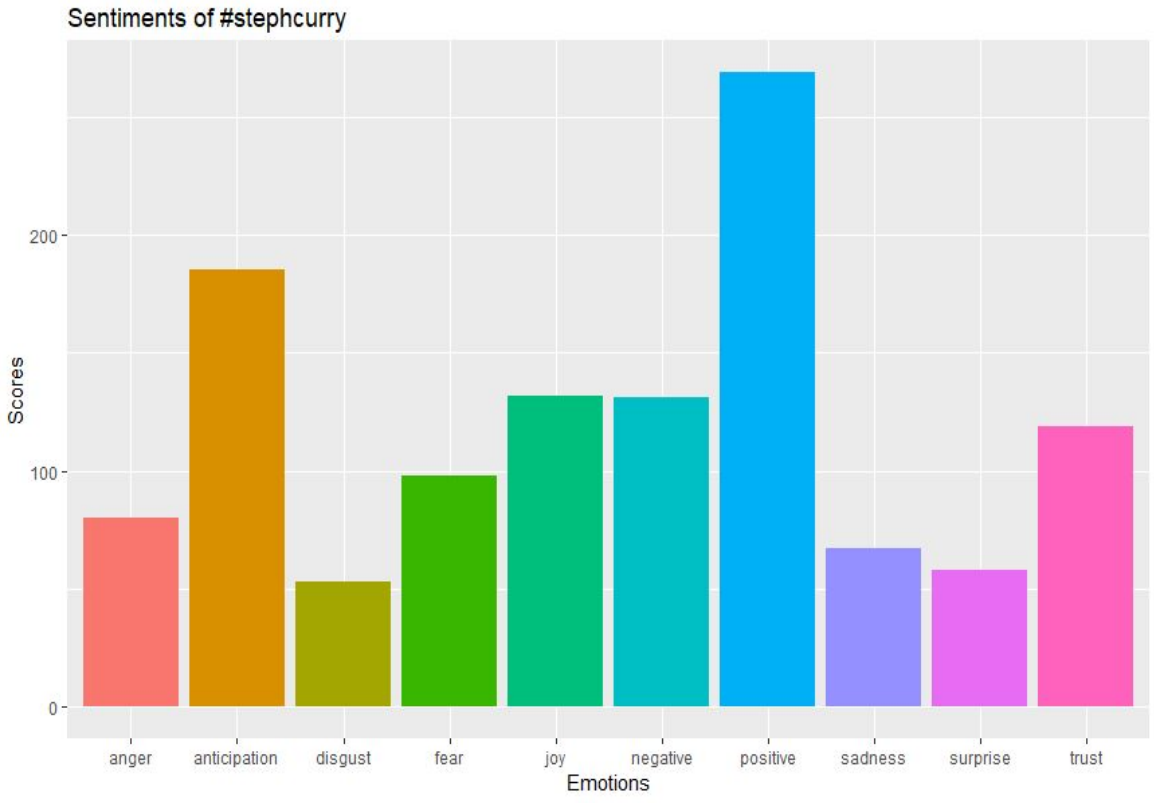
#ML sentiment analysis graph

From the following plot of sentiment analysis on tweets with #ML, it's observable that Machine Learning has a positive impact on the consumers and is extremely trusted by them. There is also a good level of anticipation maybe on its abilities of making predictions, processes' optimization and facing easier future work.



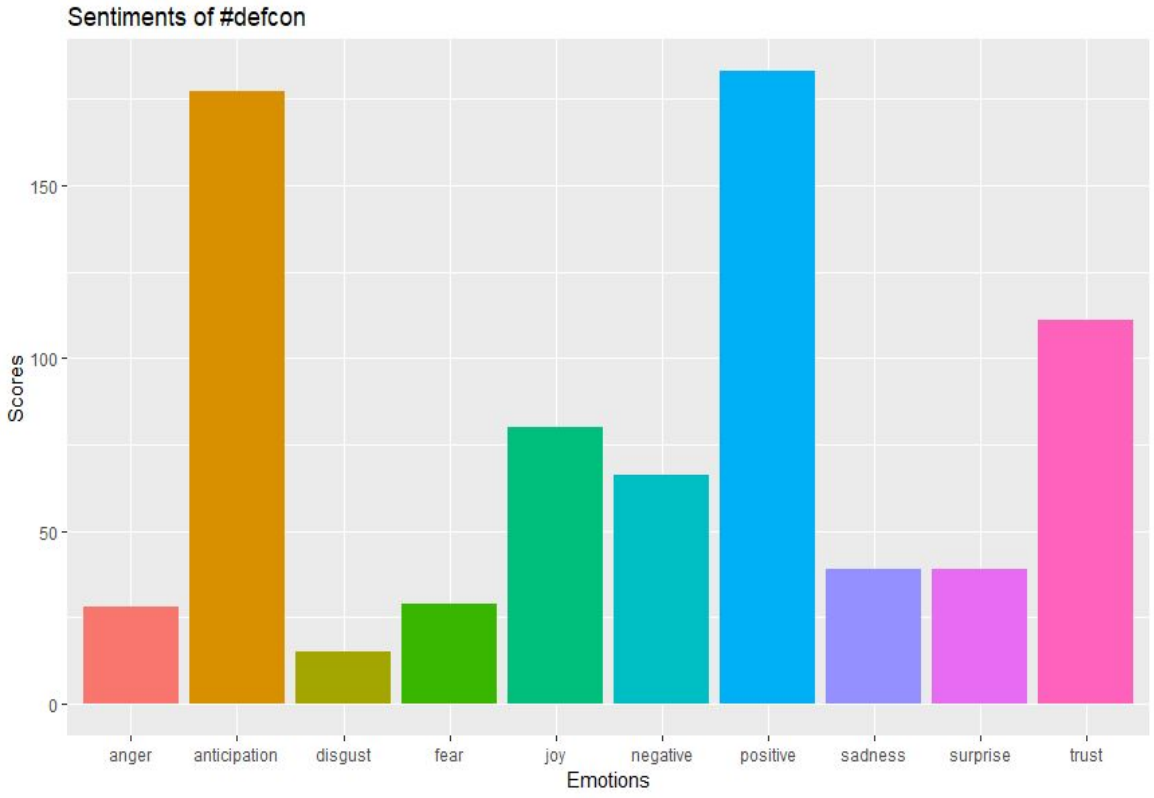
#stephcurry sentiment analysis graph

From the following plot of sentiment analysis on tweets with #stephcurry, it's observable that as a basketball player and icon he has made a positive impact on every sports fan. As a gifted scorer he is trusted by the fans, and they always anticipate his next big game. But, as with every good player, he also has a lot of people who conceive him negatively (other teams' fans) and fear that his next 40-point game will be against their favourite team.



#defcon sentiment analysis graph

Defcon is the biggest cybersecurity conference, which is held annually in Las Vegas, Nevada. It holds the biggest Capture The Flag contest and attracts thousands of people every year. From the graph we can already see the positiveness and anticipation for this event, and of course the joy and trust the people have on this conference. We can also see that the bar for negativeness is not the smallest, and that is because cybersecurity is by itself a special bound between good and bad.



#elonmusk sentiment analysis graph

Elon Musk is the real-life Tony Stark of our age. A great mind , CEO of Tesla and other great companies and a billionaire who is also not afraid to spend his money, have fun and even make fun of himself as a normal person. He is a diverse personality that is conceived differently by every person and we can also see that from the graph. We can see that there are a lot of positive thoughts about him and closely enough negative thoughts as well. A lot of people are anticipating his next moves with joy and trust , and a lot of people fear his next moves (because of his power) or even look at his moves with sadness or anger.