

*Sentiment Analysis on anticovid-19 vaccination*

*Project of the Data Mining course 2020/2021*

*Edited by:*

*Castrignano Matteo*

*Dallatomasina Erica*

*Index*

1. *Introduction*

Our goal was to create an application that could analyze tweets to discover people's opinion on antivid-19 vaccines. It collects tweets via a scraper that will later be labeled in one of three classes (negative, positive, or neutral) depending on the content of the tweet itself. To do this it was necessary to train a classifier with previously labeled tweets and choosing a particular model. The whole thing was made in *Python* using the *SK-learn* library.

1. *Data collection*

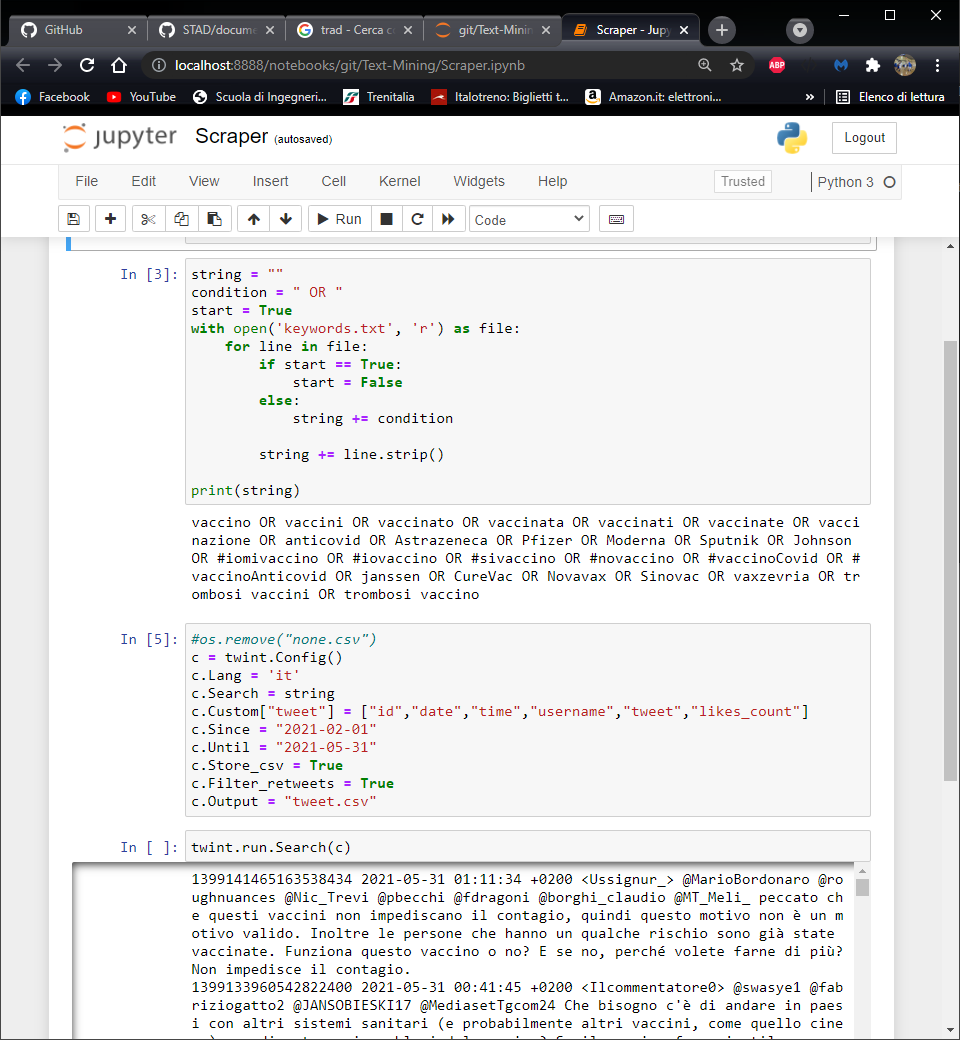
The raw tweets are scraped from Twitter using the *Twint* library. Using the functions offered from this library, we can set some options, but we have set only:

* The date, in order to scrape tweets posted only on a specific period.
* The language, in order to scrape only Italian tweet.
* Some keywords to scrape the tweets.

The scraped tweets are characterized by the following attributes (whose name is self-explanatory):

* id
* date
* time
* username
* tweet
* likes\_count

All information about *Twint* can be retrieve from GitHub repository: <https://github.com/twintproject/twint>



1. *Pre-processing*

In the pre-processing phase raw tweets are elaborated in order to clear them. In particular, we do the following steps:

* Remove all the links, emoji and images contained in the tweets
* Remove all the hashtag symbols (#) from the tweets. For example, if a tweet contains “#word” we transform it in “word”. We decided to maintain the word associated to the hashtag because it is often important to understand the stance of the tweet (for example there can be hashtags like #iomivaccino or #vaccinomorte that are indicative of the opinion about vaccines expressed in the tweet)
* Remove all the mentions to other users in the tweet
* Put the text of the tweets in lower case
* Remove all the punctuation
* Remove all the extra-spaces

***Before***

*"@valy\_s È una cosa normale. L'immunità di gregge è a senso unico. Se vaccini tutti non sconfiggerai mai la malattia. Quando lo capiranno, torneremo a vivere normalmente! Se vuoi fermare la pandemia, studiala e crea una medicina. Il vaccino serve a poco!"*

***After***

*e una cosa normale l immunita di gregge e a senso unico se vaccini tutti non sconfiggerai mai la malattia quando lo capiranno torneremo a vivere normalmente se vuoi fermare la pandemia studiala e crea una medicina il vaccino serve a poco*

1. *Supervised learning stage*
   1. *Tokenization, stop-word filtering and stemming*
   2. *Stem by relevant*
   3. *Classification and evaluation*
   4. *Cross-validation*
2. *Concept drift*