

# Group project : Predicting political orientation from Twitter contents

Natascia Caria, María Emilia Charnelli, Francesco Ferretto,  
Matteo Lavina, Andrea Sinigaglia,

November 1, 2019

## Abstract

## 1 Phase 1 - Collecting data

Collecting data by going through profile on Twitter randomly.

### 1.1 Strategy for collect users

We did an script in Python using the Tweepy library for accessing the Twitter API [1]. Also, we used the random functions of the Numpy library [1].

In order to collect users we first take an initial list of  $n$  users from Italy who meet certain conditions. Those users are added to the selected user list. Then from these users we start an iterative process. First we select users randomly from the selected user list and from each selected user we take  $k$  of their friends randomly, and if those new users selected (friends) meet certain conditions, they are added to the selected user list of users. This process is repeated several times until complete the 1000 users required for the analysis.

We choose the friends of every user instead of their followers because we consider that a user is following accounts that he chooses to follow because he knows or he likes those users.

We used the search method of Tweepy to get the initial list of users. That method returns a list of  $t$  tweets by geolocalization, in this case from Italy. From these list of tweets we get the unique users and we select the users who meet certain conditions to create the initial list of users. We could have used the search method for obtaining all the users for the analysis. But to ensure to obtain randomly users who aren't bots and who are useful for this task, we considered that is better to iterate over friends of an initial list.

For obtaining a random sample of users we used the function "choice" of the random package in the Numpy library. Also, to obtain different users in each iteration we used the function "choice" with a parameter that allows choosing  $k$  users in a list using no replacement.

For check that every user meet with certain conditions, we define a function to determine is a user is useful for our analysis:

- `min_n_tweets = 10`. Minimum number of tweets for consider a user as useful.
- `min_n_friends = 15`. Minimum number of friends for consider a user as useful.
- `tweet_language = 'it'`. The language of the last tweet should be italian for consider a user as useful.
- `min_date_activity_user = 1/6/2019`. The last activity of the user should be after the 1st of June.

## 1.2 Strategy to detect the party of the users collected

First of all, we considered a temporal window between June-September to analyse the tweets of every users collected with the script mentioned in the previous section.

## References

- [1] Tweepy Library. <https://www.tweepy.org/>. [Online; accessed 01-October-2019].