

Exploring the political pulse of a country using data Science tools

Miguel G. Folgado

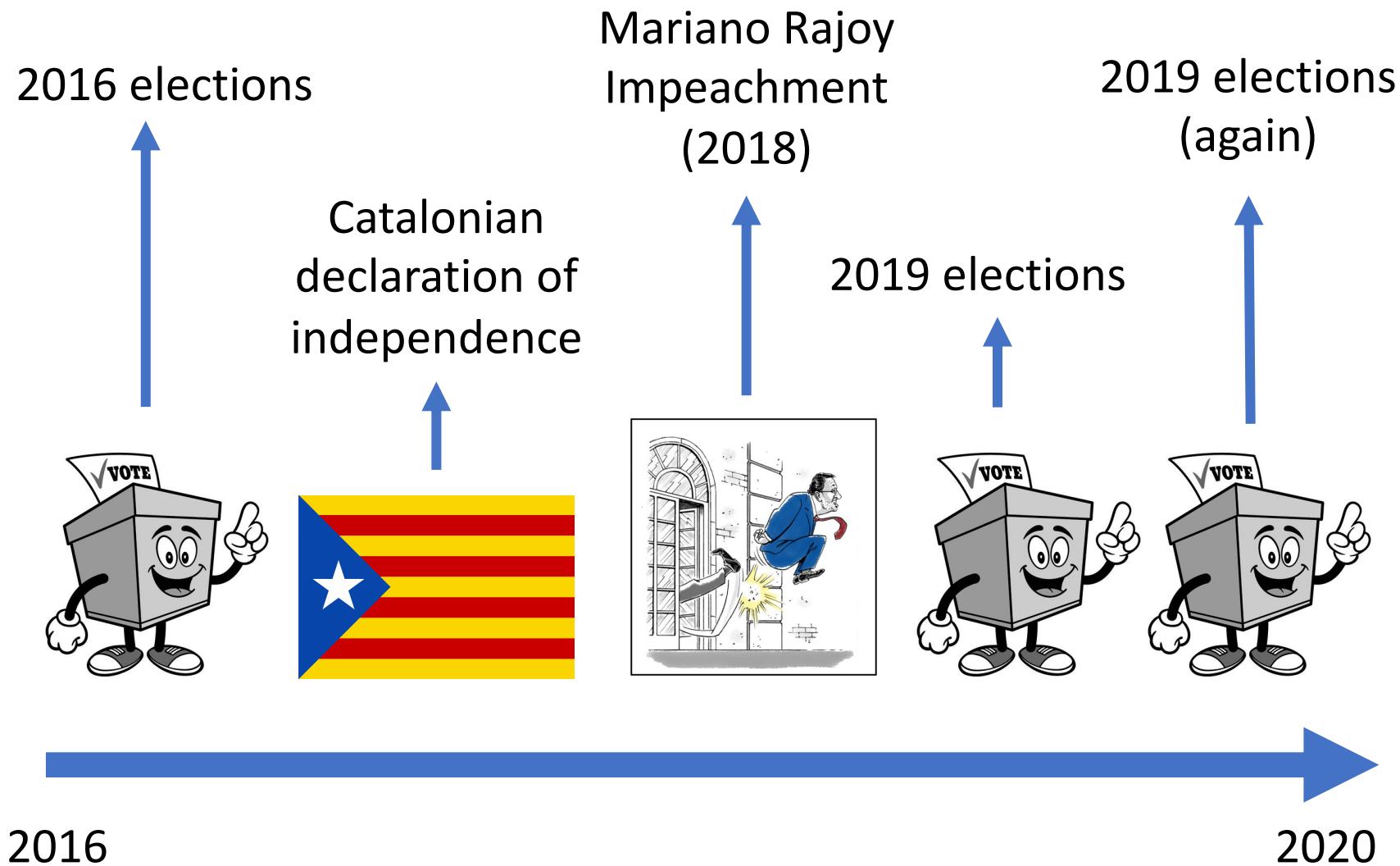
Based on:

Miguel Folgado and Verónica Sanz

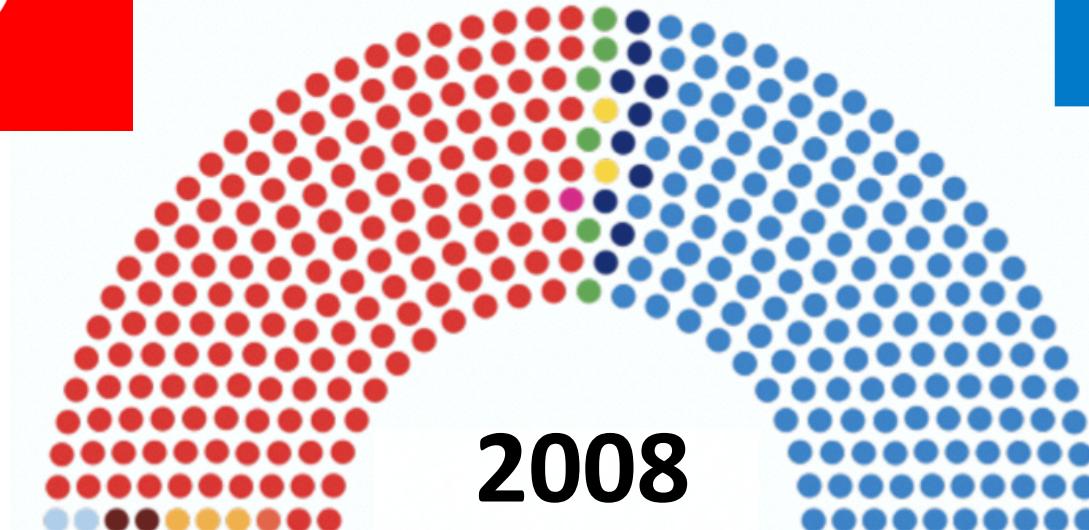
arXiv:2011.10264
[10.31235/osf.io/u73d9](https://doi.org/10.31235/osf.io/u73d9)



Spain is different

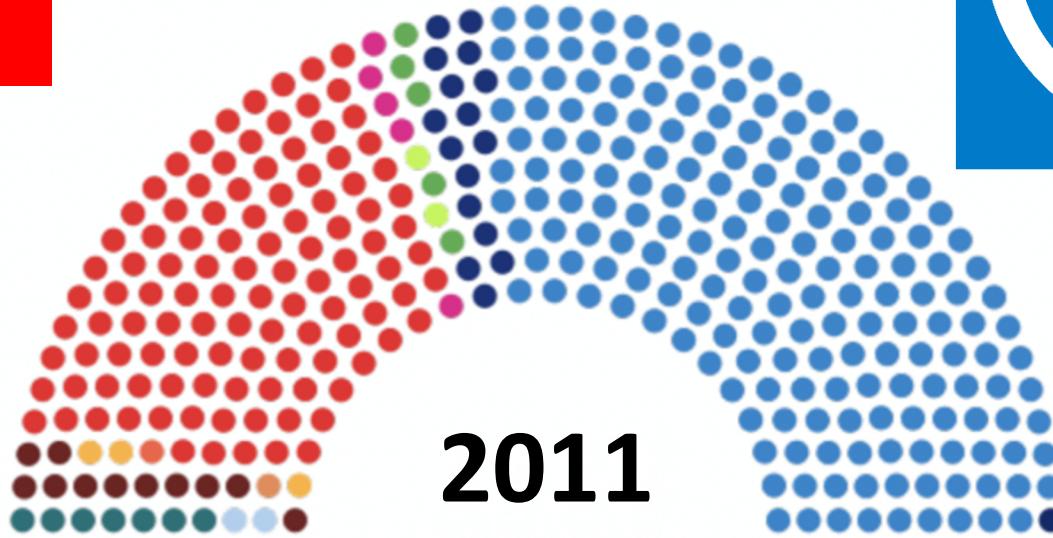


Evolution of the Spanish parliament in the last years



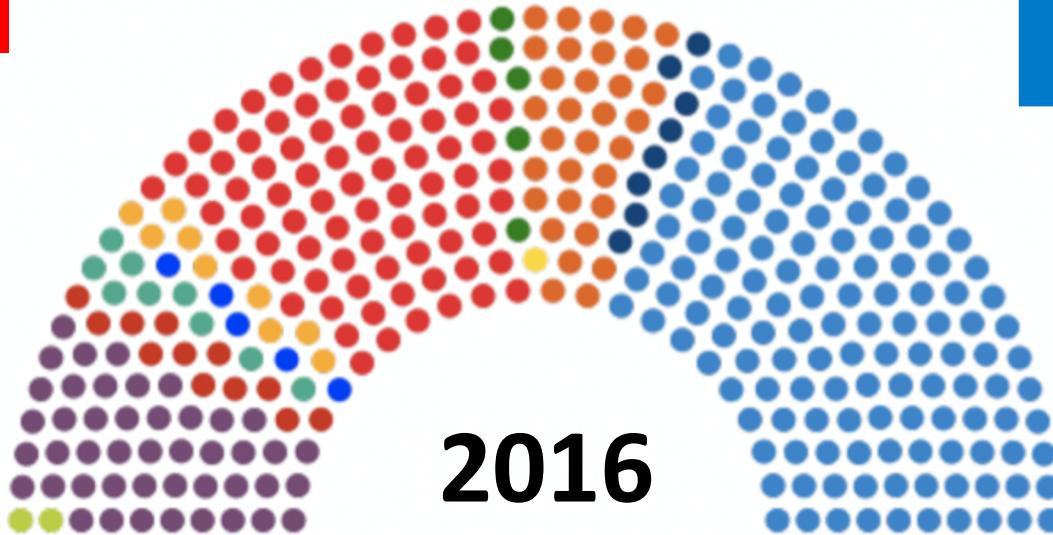
Highly Bi-partisan

Evolution of the Spanish parliament in the last years

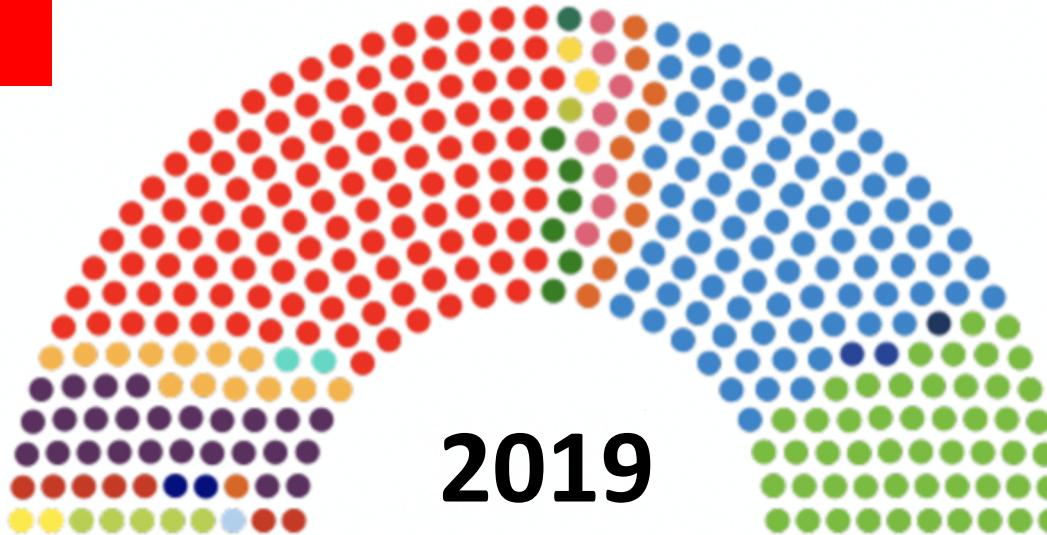


Highly Bi-partisan

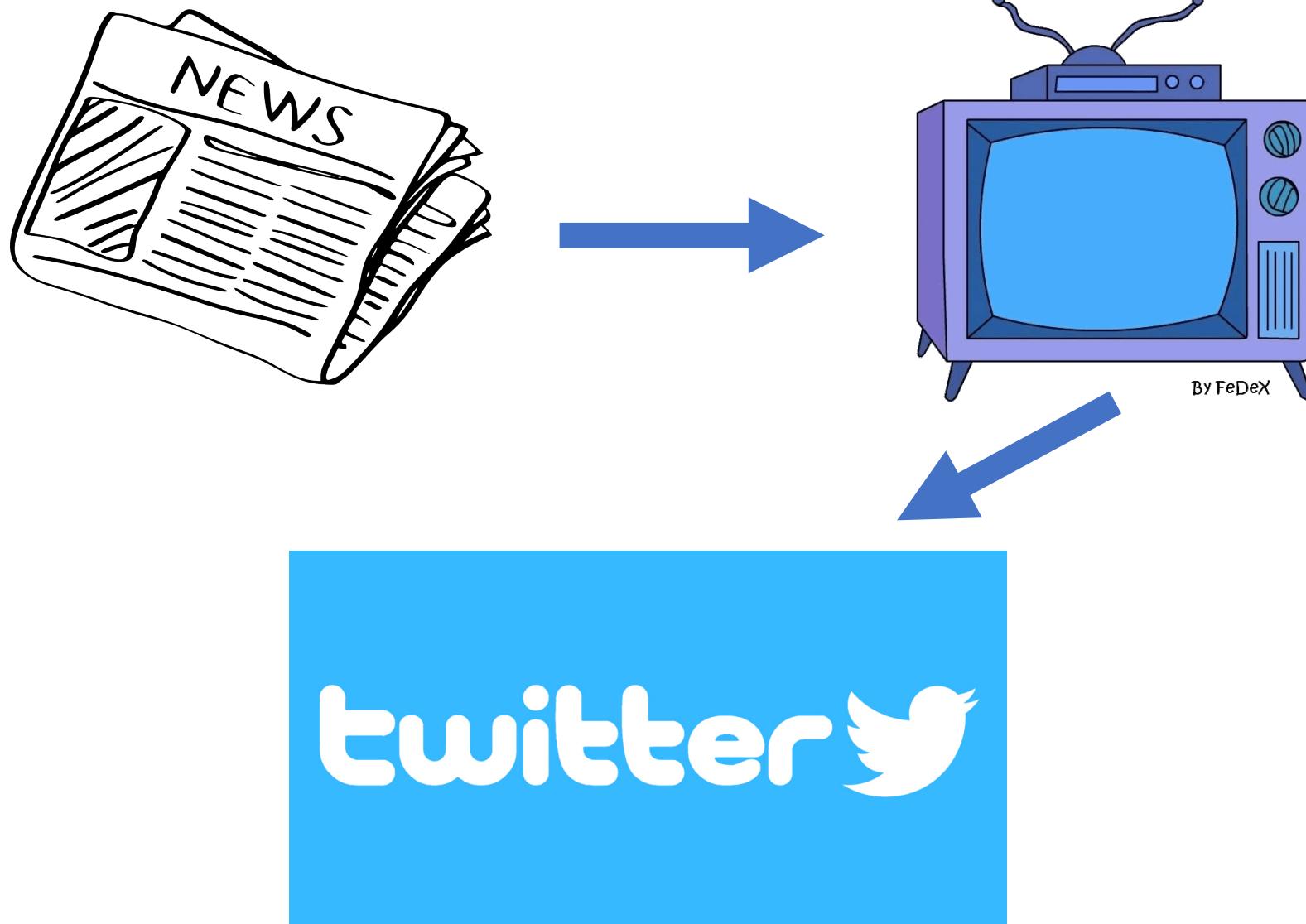
Evolution of the Spanish parliament in the last years



Evolution of the Spanish parliament in the last years



Political Rings



Data set

The main problem in data analysis is obtaining **data sets**

Official method: Open a twitter developer account and use the official api tools



Problem: Only the last **3200** tweets per account

Maybe this quantity would be huge for the most part of the population, but not for politicians!!

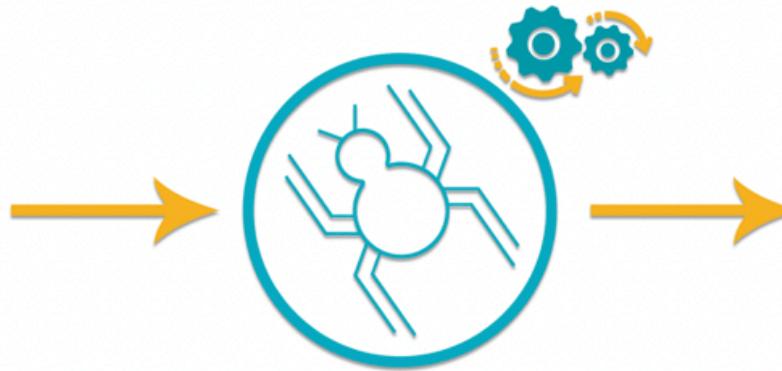
Data set

Alternative methods?

What is Web Scraping?



Website Pages,
Unstructured Data



Web Scraping/
Data Extraction



Structured Data

We use the Python package **GetOldTweets3** to obtain the data set

Data set



Data set: 138203 Tweets from 53 Twitter accounts

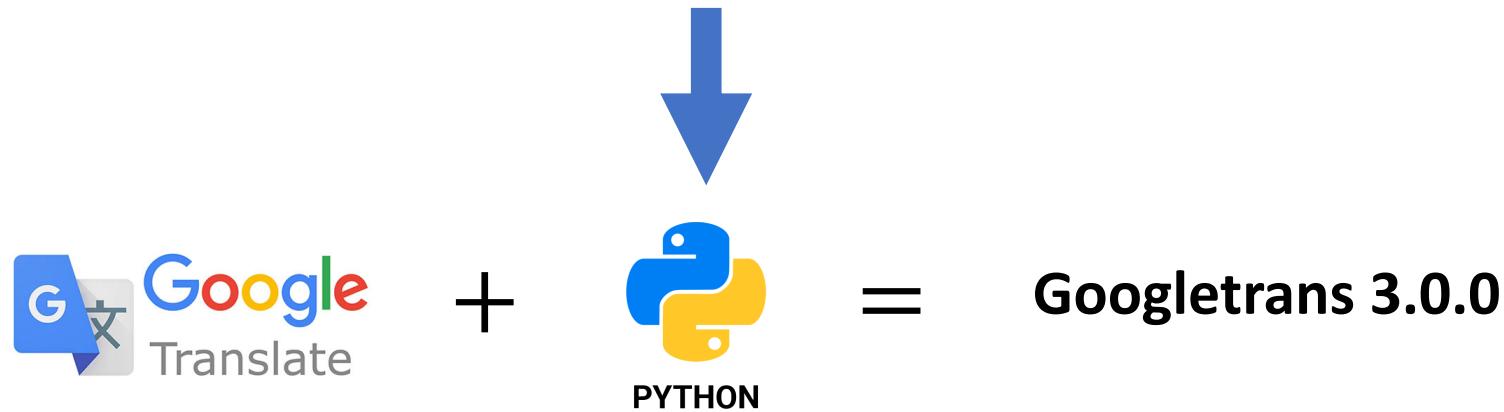
January 2016 – June 2020

Preprocessing data

@sanchezcastejon 2018/03/26



PP y Cs vuelven a demostrarnos que son dos caras de la misma moneda: la desigualdad y la precariedad. Malas noticias para la clase trabajadora.



PP and Cs once again show us that they are two sides of the same coin: inequality and precariousness. Bad news for the working class.

Preprocessing data

@sanchezcastejon 2018/03/26



PP and Cs once again show us that they are two sides of the same coin:
inequality and precariousness. Bad news for the working class.

Cleaned Tweets



PP Cs show us two sides coin inequality precariousness. Bad news working
class

Preprocessing data

@sanchezcastejon 2018/03/26



PP and Cs once again show us that they are two sides of the same coin:
inequality and precariousness. Bad news for the working class.

Cleaned Tweets



PP Cs show us two sides coin inequality precariousness. Bad news working
class

Lemmanization



PP Cs show us two side coin inequaly precariousness Bad news work class

Preprocessing data

@sanchezcastejon 2018/03/26



PP and Cs once again show us that they are two sides of the same coin:
inequality and precariousness. Bad news for the working class.

Cleaned Tweets



PP Cs show us two sides coin inequality precariousness. Bad news working
class

Lemmanization



PP Cs show us two side coin inequaly precariousness Bad news work class

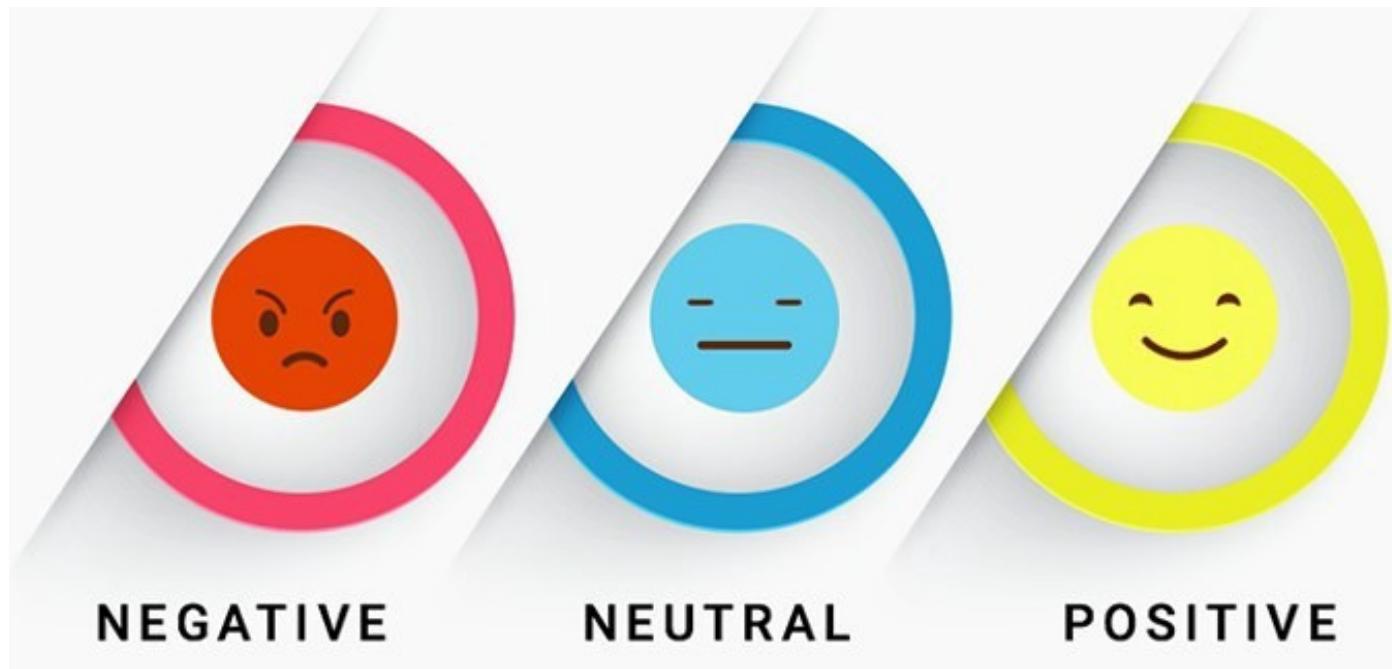
Tokennization



(PP, Cs, show, us, two, side, coin, inequaly, precariousness, Bad, news, work,
class)

Sentimental analysis

Sub-field of Natural Language Processing (NLP) that tries to identify and extract opinions within a given text



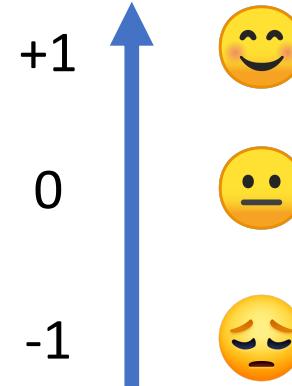
Sentimental analysis

Sub-field of Natural Language Processing (NLP) that tries to identify and extract opinions within a given text

Python Library



vaderSentiment



Hate	-0.57	Table	0.0	Good	0.44
Lazy	-0.36	Fruit	0.0	Beauty	0.59
Sad	-0.48	Chair	0.0	Love	0.64

Sentimental analysis

@sanchezcastejon 2018/03/26



PP and Cs once again show us that they are two sides of the same coin:
inequality and precariousness. Bad news for the working class.

pre-processing data

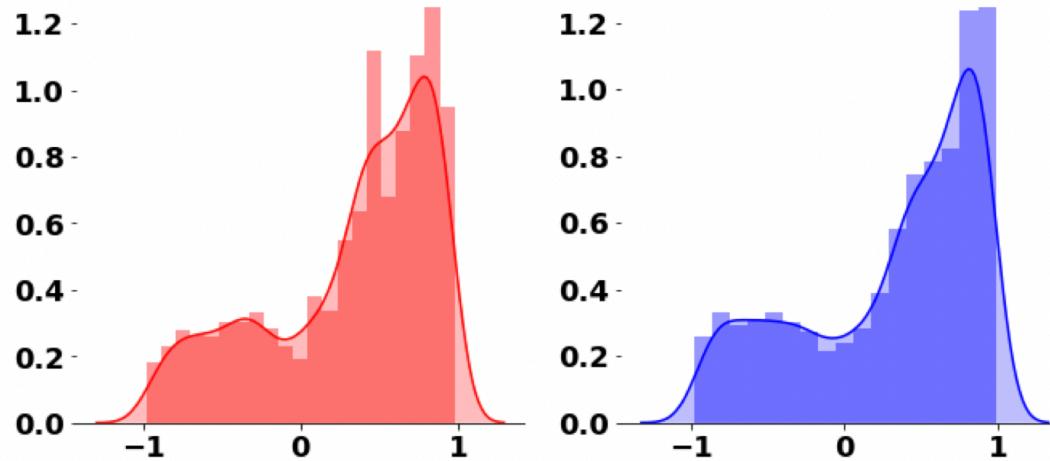


(PP, Cs, show, us, two, side, coin, inequaly, precariousness, Bad, news, work,
class)

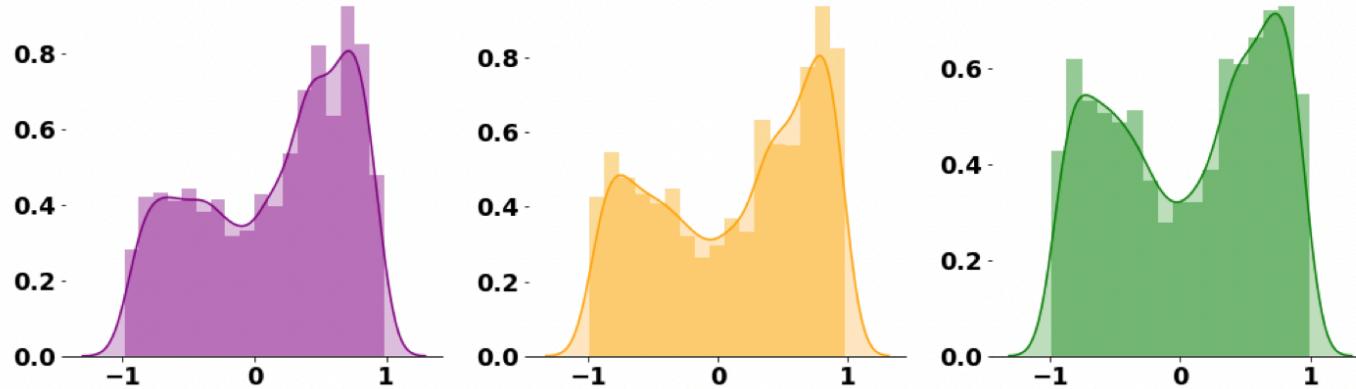
Total sentimental score: **-0.54**

Sentimental analysis

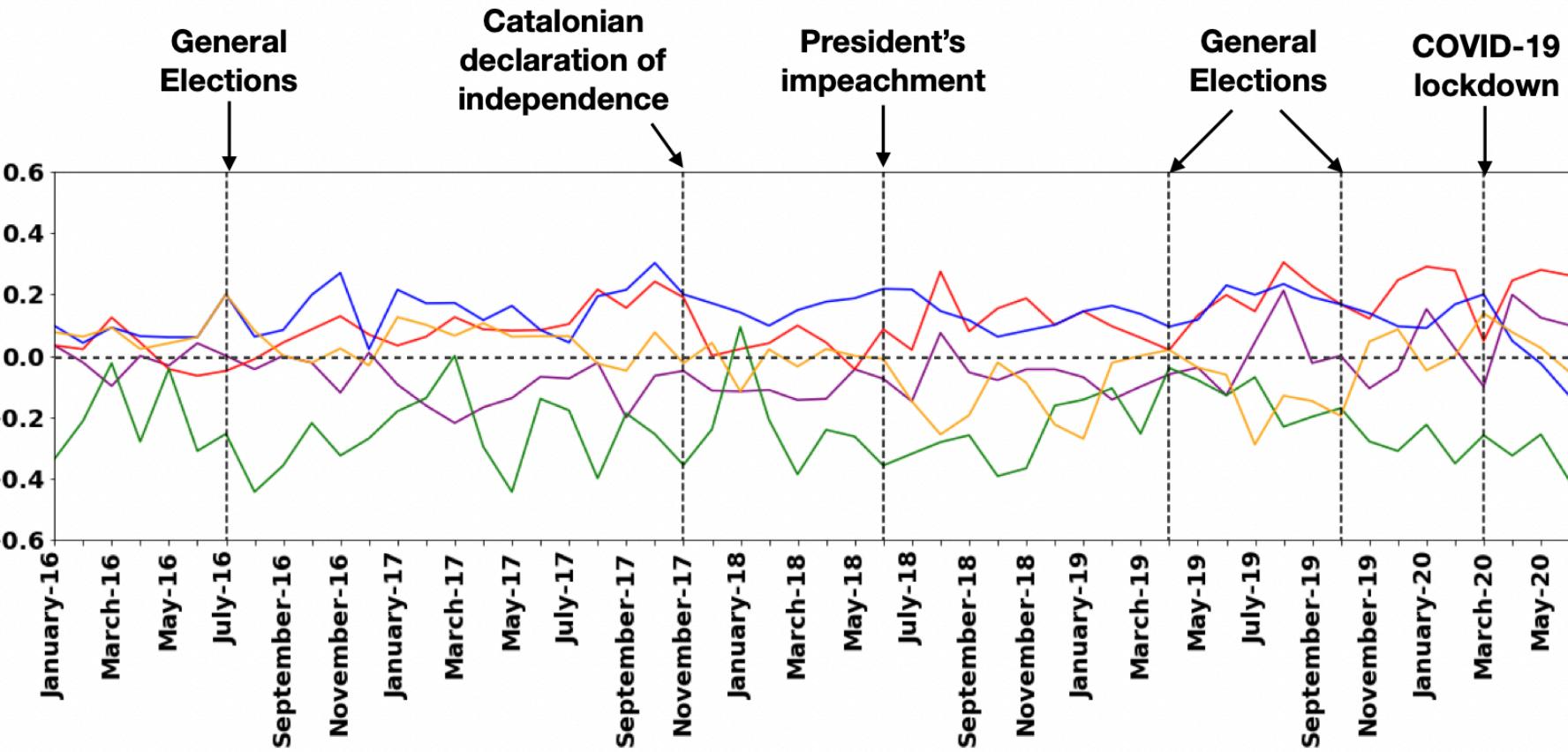
Classical Parties



New Parties



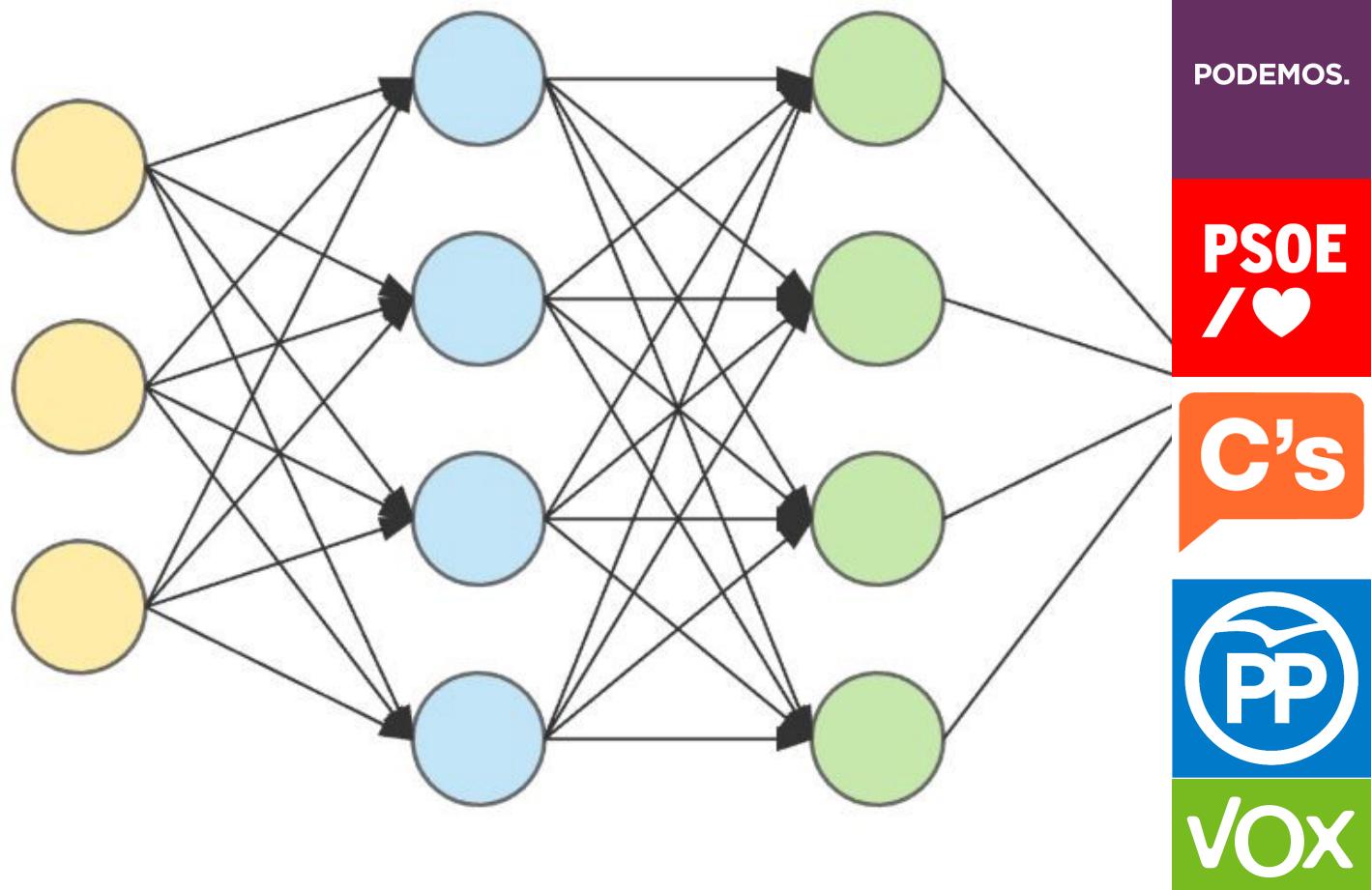
Time evolution of sentiments



Bag of Words (BOW)



Is it possible to predict the political party?



Preprocessing data

@sanchezcastejon 2018/03/26



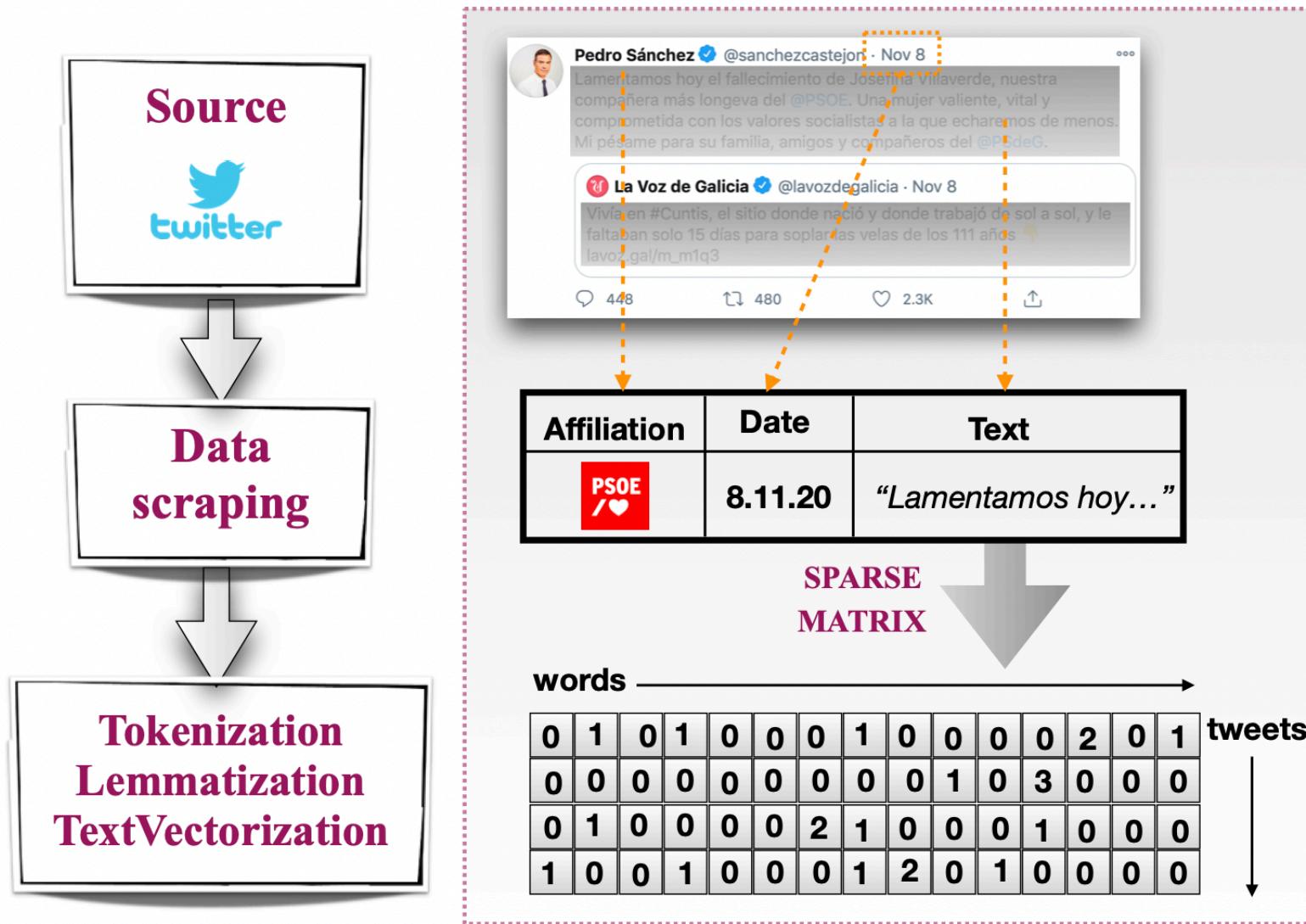
(PP, Cs, show, us, two, side, coin, inequaly, precariousness, Bad, news, work, class)

TextVectorization

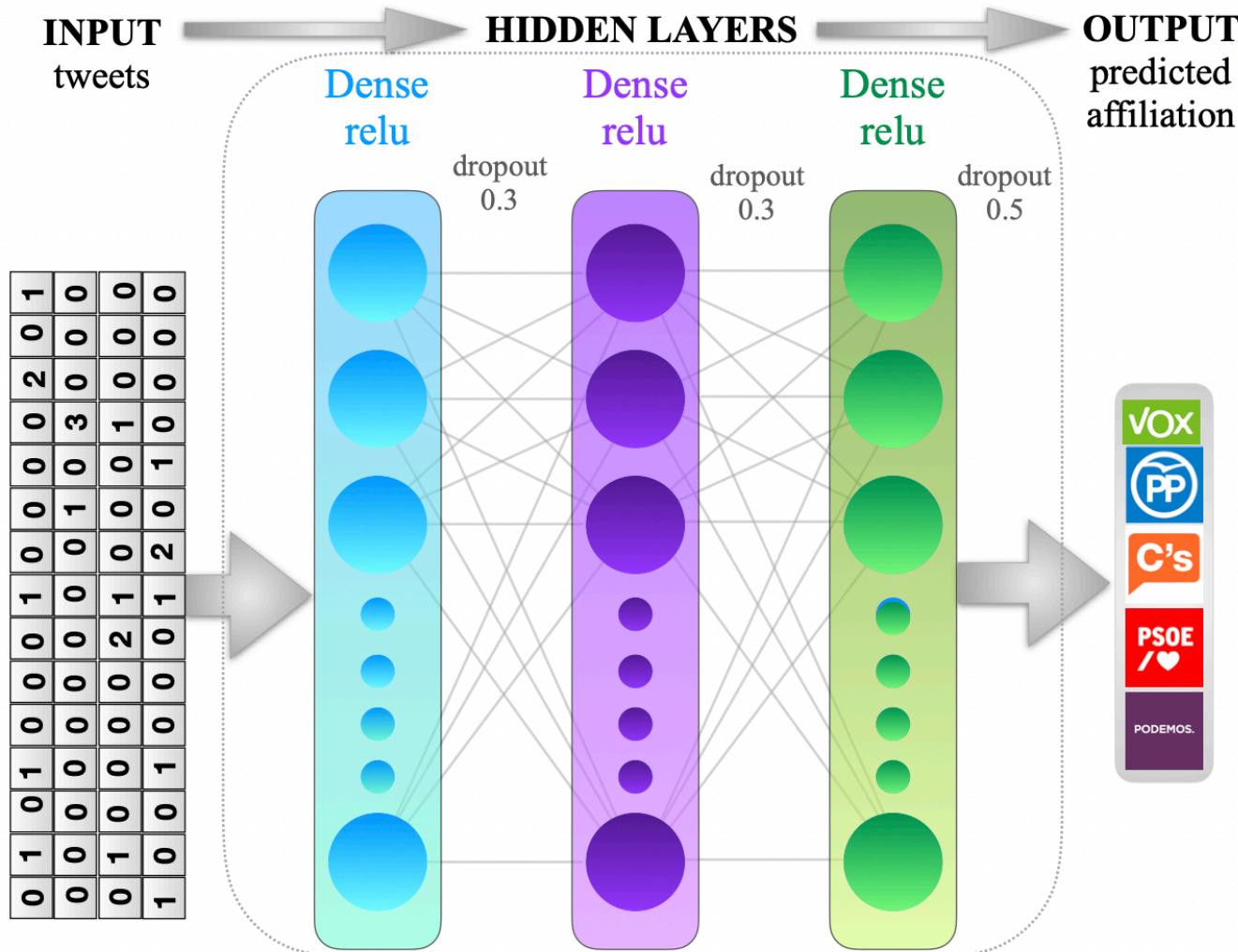


PP	hello	good	us	two	tv	coin	work	table	side
1	0	1	1	1	0	1	1	0	1

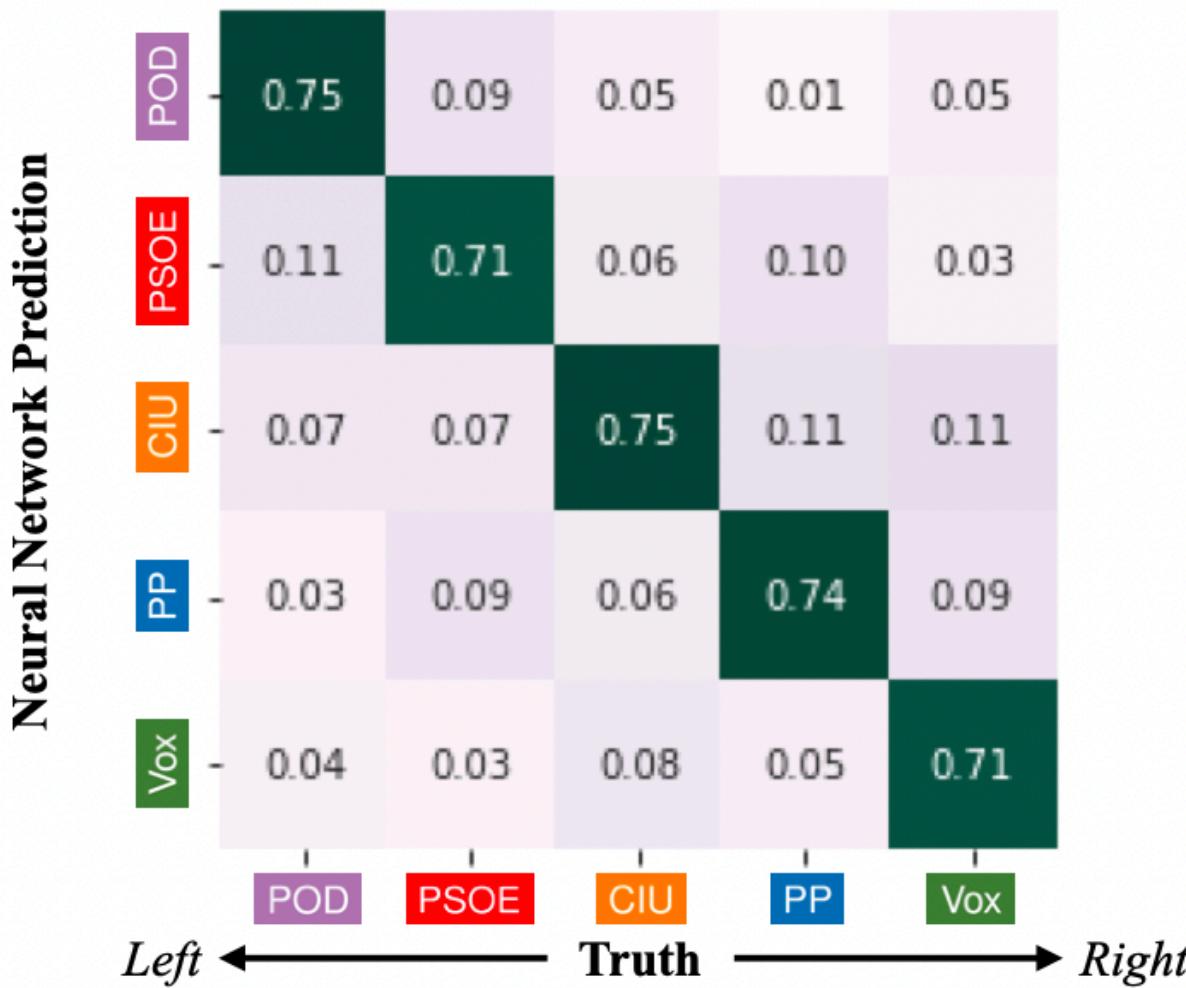
Preprocessing data



Neural networks to predict the political party



Neural networks to predict the political party



Neural networks to predict the political party

Neural Network Prediction

