



THE MOOD OF TWITTERERS IN MEXICO

(EL ESTADO DE ÁNIMO DE LOS TUITEROS EN MÉXICO)



INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA

September, 2018



The three pillars of official statistics



The ~~three~~ four pillars of official statistics



BIG DATA



“Big data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it...”

Dan Ariely



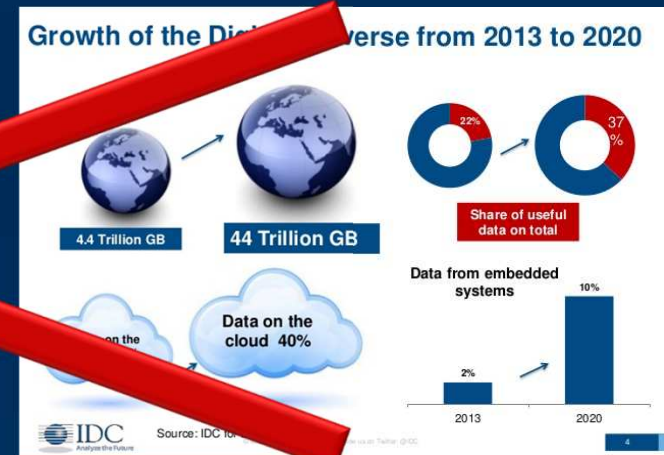


The Big Data definition evolves

Initially it was about...

- *Volume*
- *Velocity*
- *Variety*
- *Veracity*
- *Value*

the 3 V's



Instead...

Big Data is an approach to use and re-use the totality of a set of information in a diversity of possible purposes, normally different to those that originated the information set in the first place.



Convergence of two agendas



 Big data.

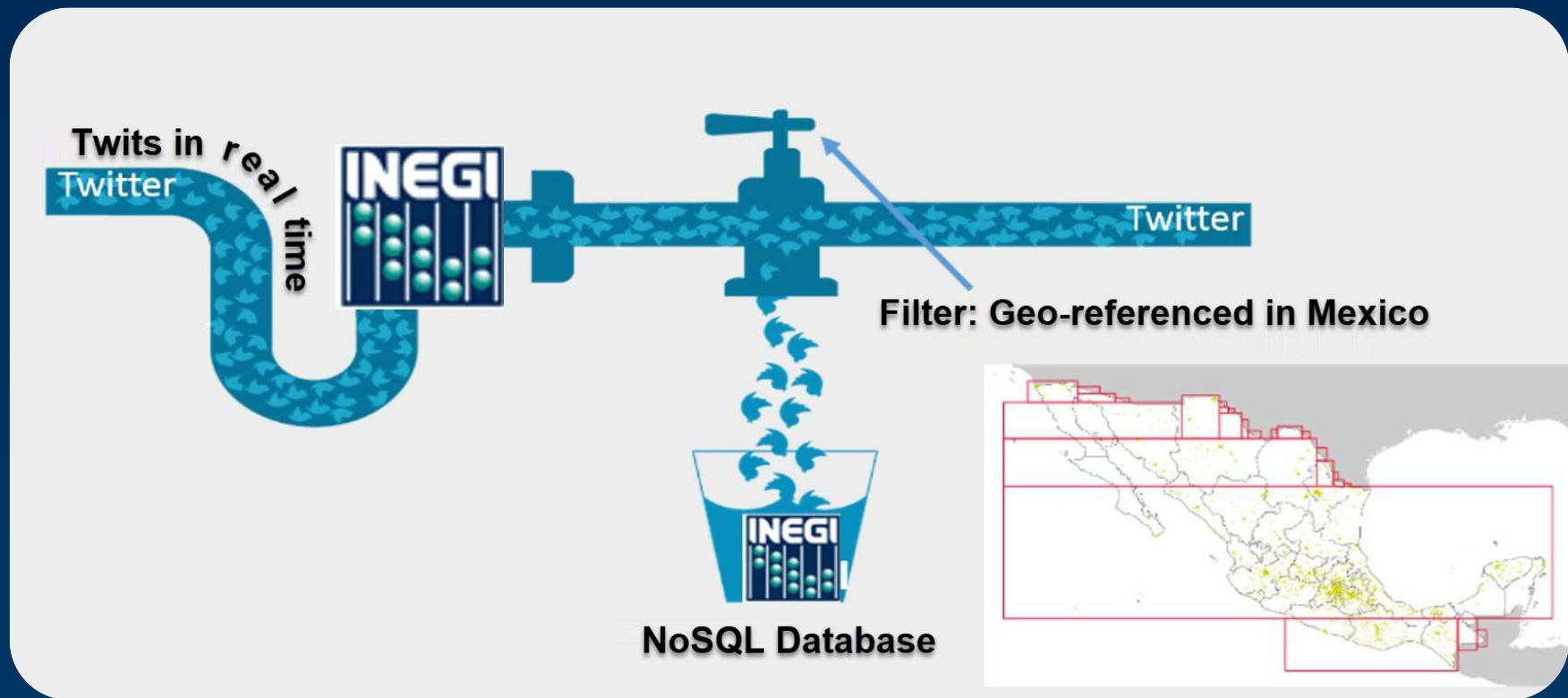
 Subjective Well Being
(Martin Seligman).



Since February 2014



Collecting tweets



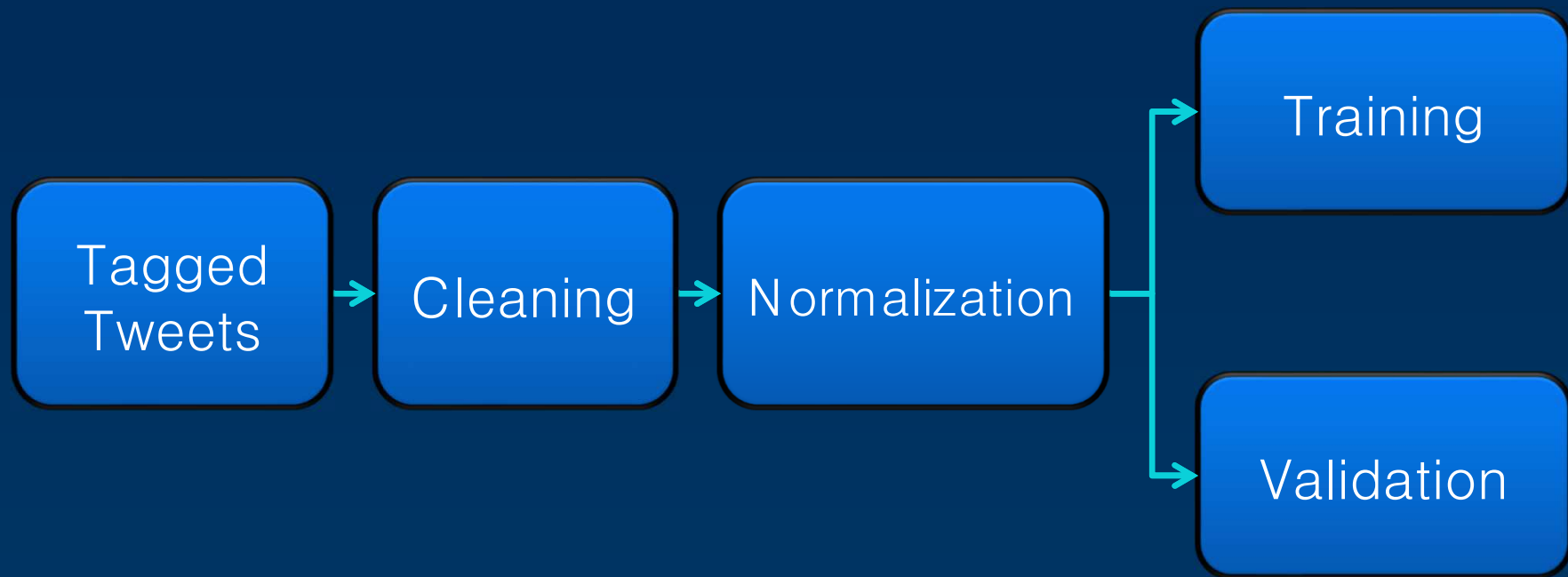
NoSQL Database

NoSQL Database

300 million tweets



Automatic analysis and classification



Cleaning of the tagged set

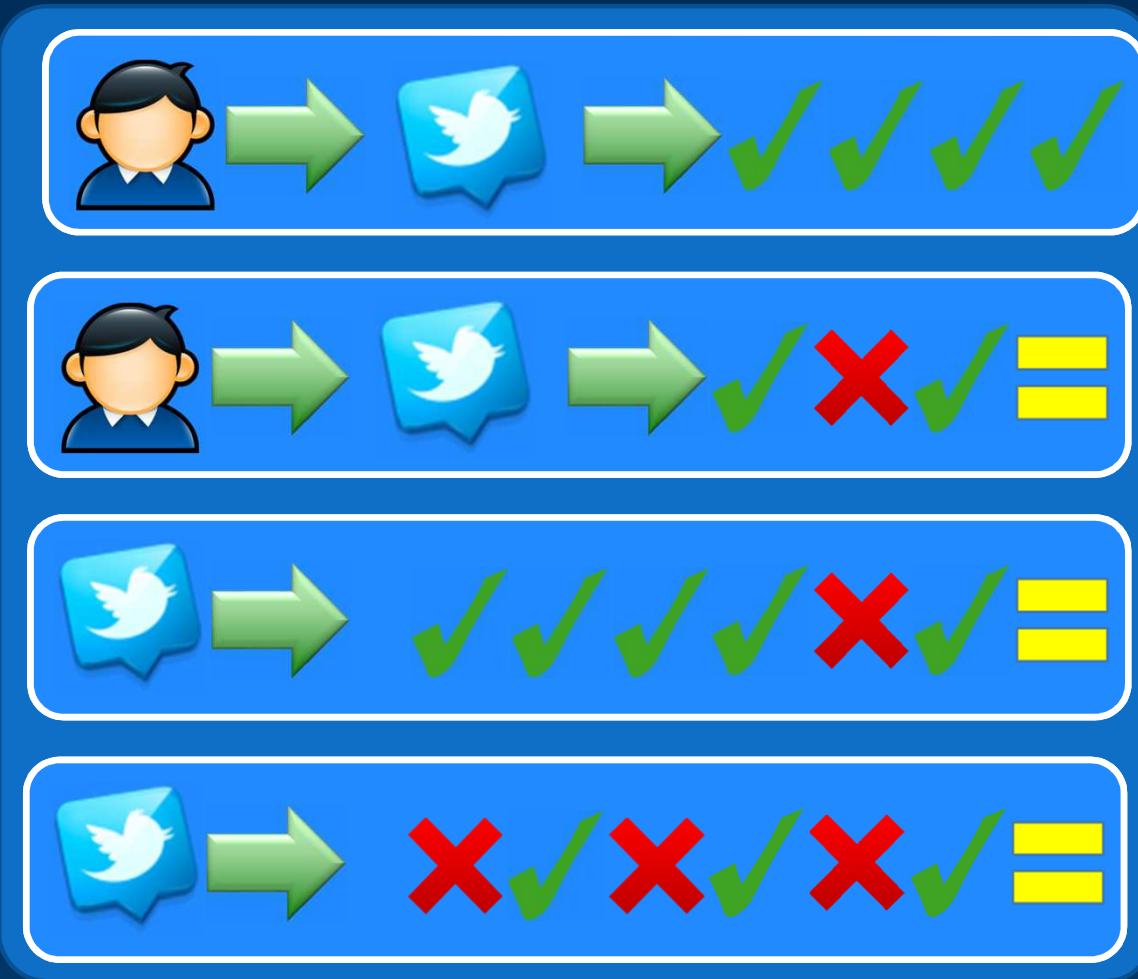
Cleaning Entropy

Tagged

Tweets



Cleaning of the tagged set (cleaning)





Normalization

Words correction

(dictionary, statistical & heuristic)

rojooooooooo jajajajajaja
↓ ↓
rojo ja

Codification

(diacritic symbols)

caña perdón amig@s
↓ ↓ ↓
cana perdon amigs

Lematization

(FreeLing)

rompí llegó subió
↓ ↓ ↓
romper llegar subir

Polarity of Emoticons

(polarity tag)

😊 😐 😡
↓ ↓ ↓
+1 0 -1

Q-Grams

(q=4)

romper perder amigs
↓ ↓ ↓
romp perd amig

Filtered

(without stop words)

adverbs adjectives
interjections nouns
verbs hashtags

Set of tagged tweets



- 54 131 tweets manually tagged by 9 330 people

- Classification system:

<https://cienciadedatos.inegi.org.mx/pioanalysis/>





Estar enamorada es como ir en un Ferrari a 240 kms/h. Se siente CHINGON pero sabes que en cualquier momento viene el putazo (:

nivel 0

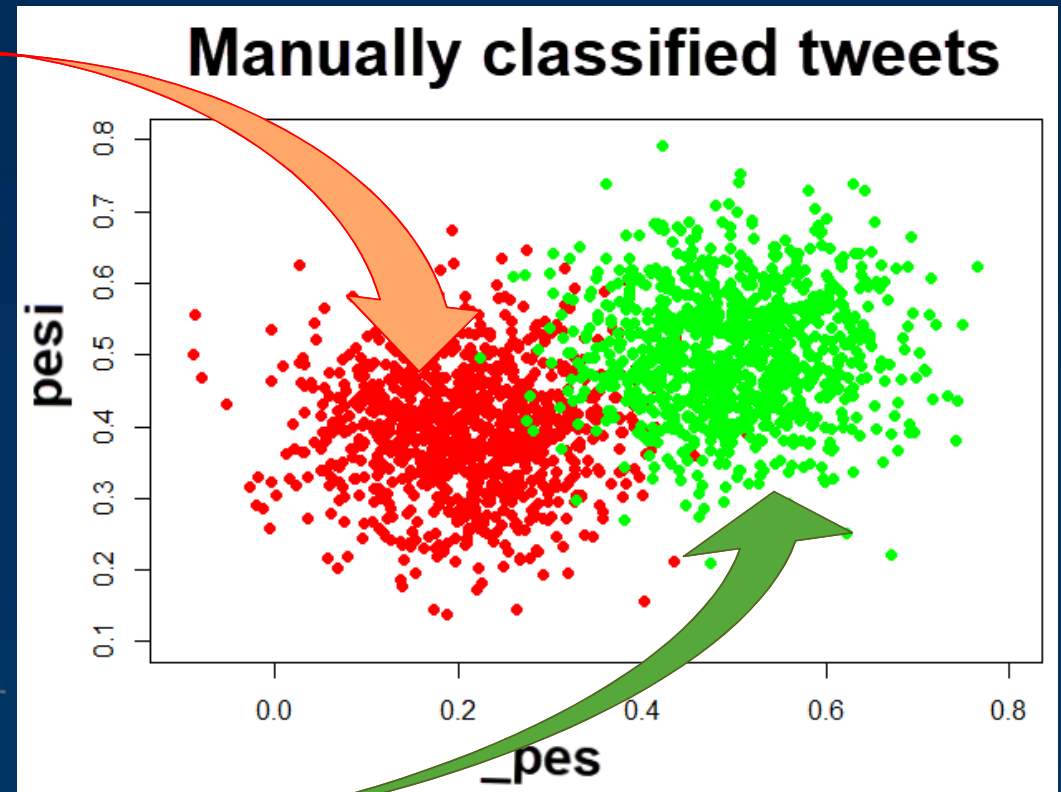
¿El tuitero se sentía?



Machine learning algorithm SVM



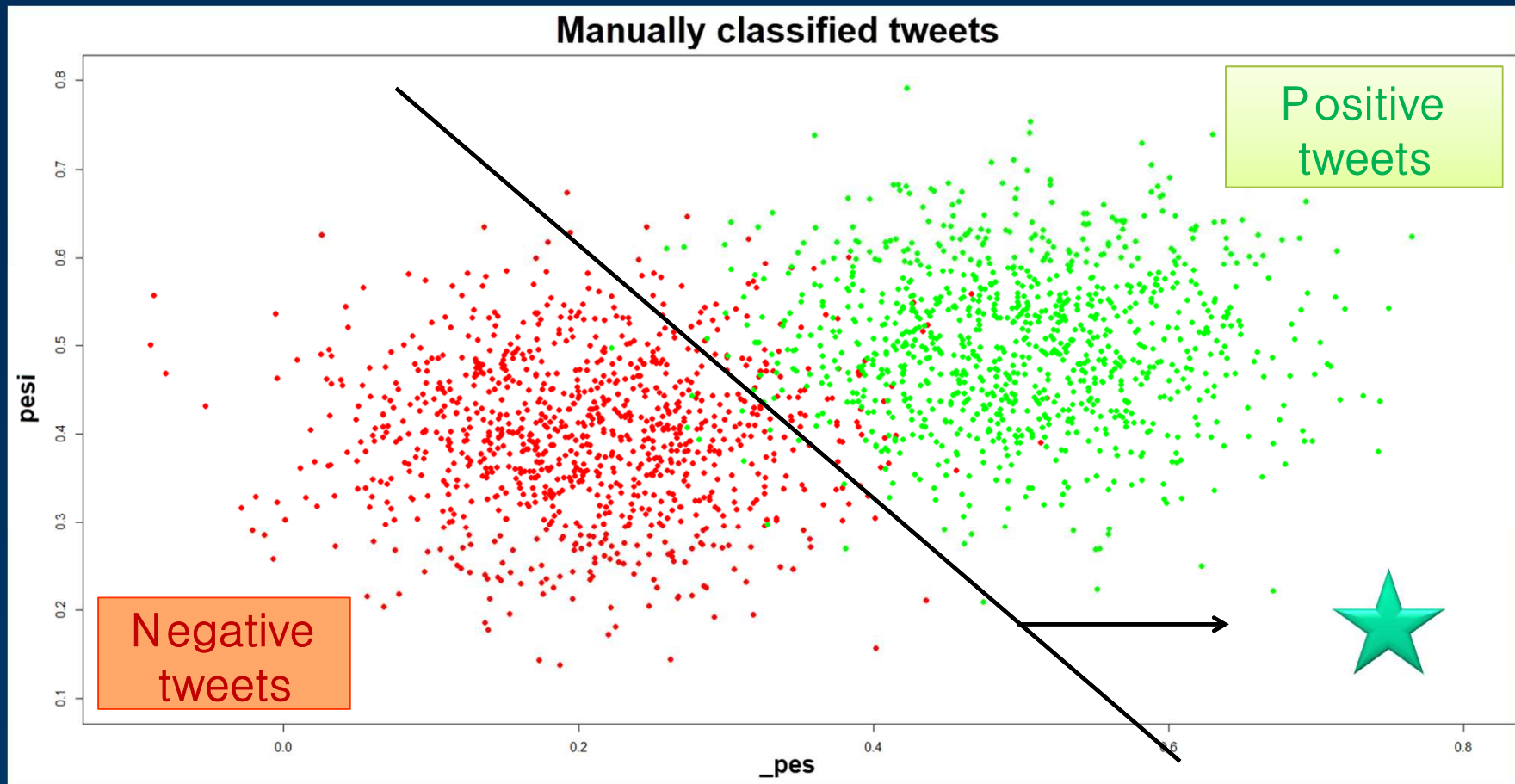
	tag	_pes	pesi
Tweet 1	NEGATIVE	0.22	0.48
Tweet 2	NEGATIVE	0.12	0.55
Tweet 3	NEGATIVE	0.25	0.39
...
Tweet n	POSITIVE	0.6	0.35



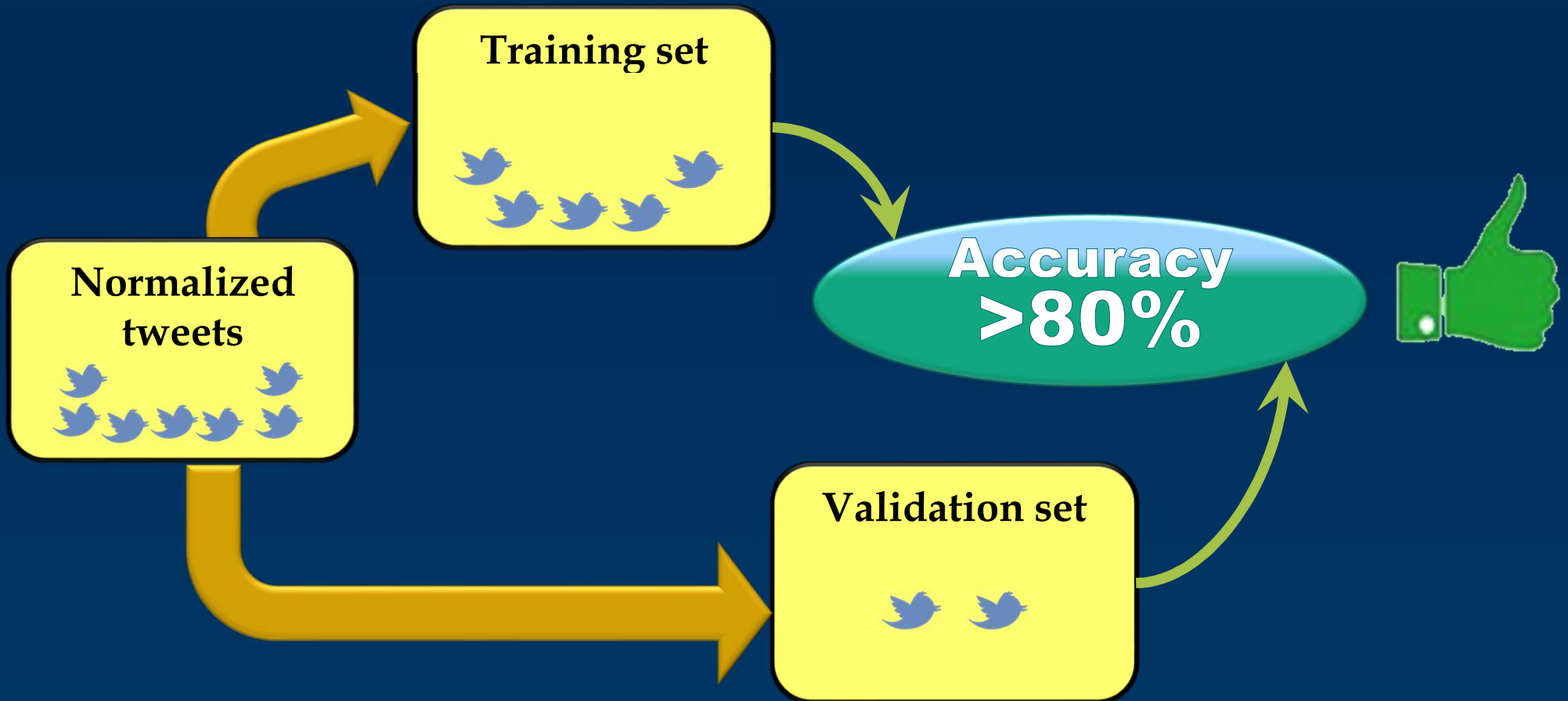
PDOD4 Manually classified tweets

PAEZ DOMINGUEZ OLINCA DESSIREE, 07/09/2018

Training the SVM algorithm



Evaluating innovative classifiers



Automatically classifying tweets



Unclassified tweets

Normalization
Vector representation
Classification



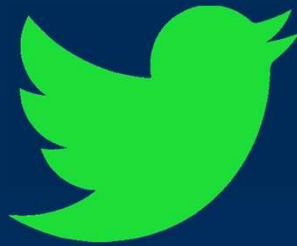
Hundred of millions tagged tweets



Positivity quotient

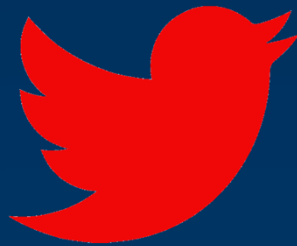


POSITIVES

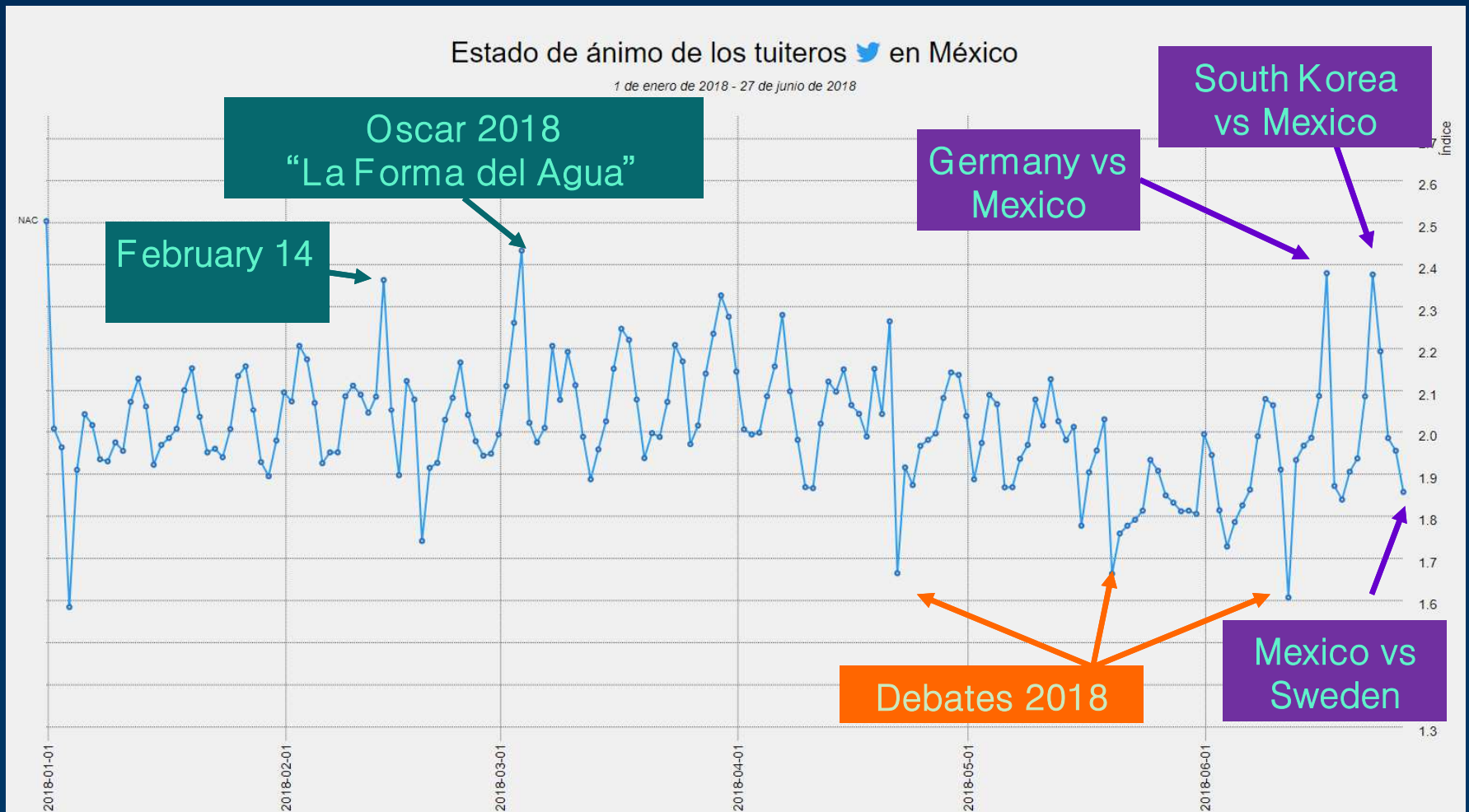


Positivity
quotient

NEGATIVES



The mood of tweeters in Mexico (showing 2018)



Link:



 <http://www.inegi.org.mx/>

 <http://www.beta.inegi.org.mx/app/animotuitero/#/app/multiline>



Other INEGI projects with Twitter:



- 🐦 Domestic tourism.
- 🐦 Mental health.
- 🐦 Mobility in Mexico City.
- 🐦 New agglomerations.
- 🐦 Consumer confidence.
- 🐦 Insecurity.





!Thank you!





Conociendo México

01 800 111 46 34

www.inegi.org.mx

atencion.usuarios@inegi.org.mx



INEGI Informa



@INEGI_INFORMA



**INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA**



Help us to

Mood

Leads people wanting to help to another page
state and temporality

2016 a 24 de enero de 2018

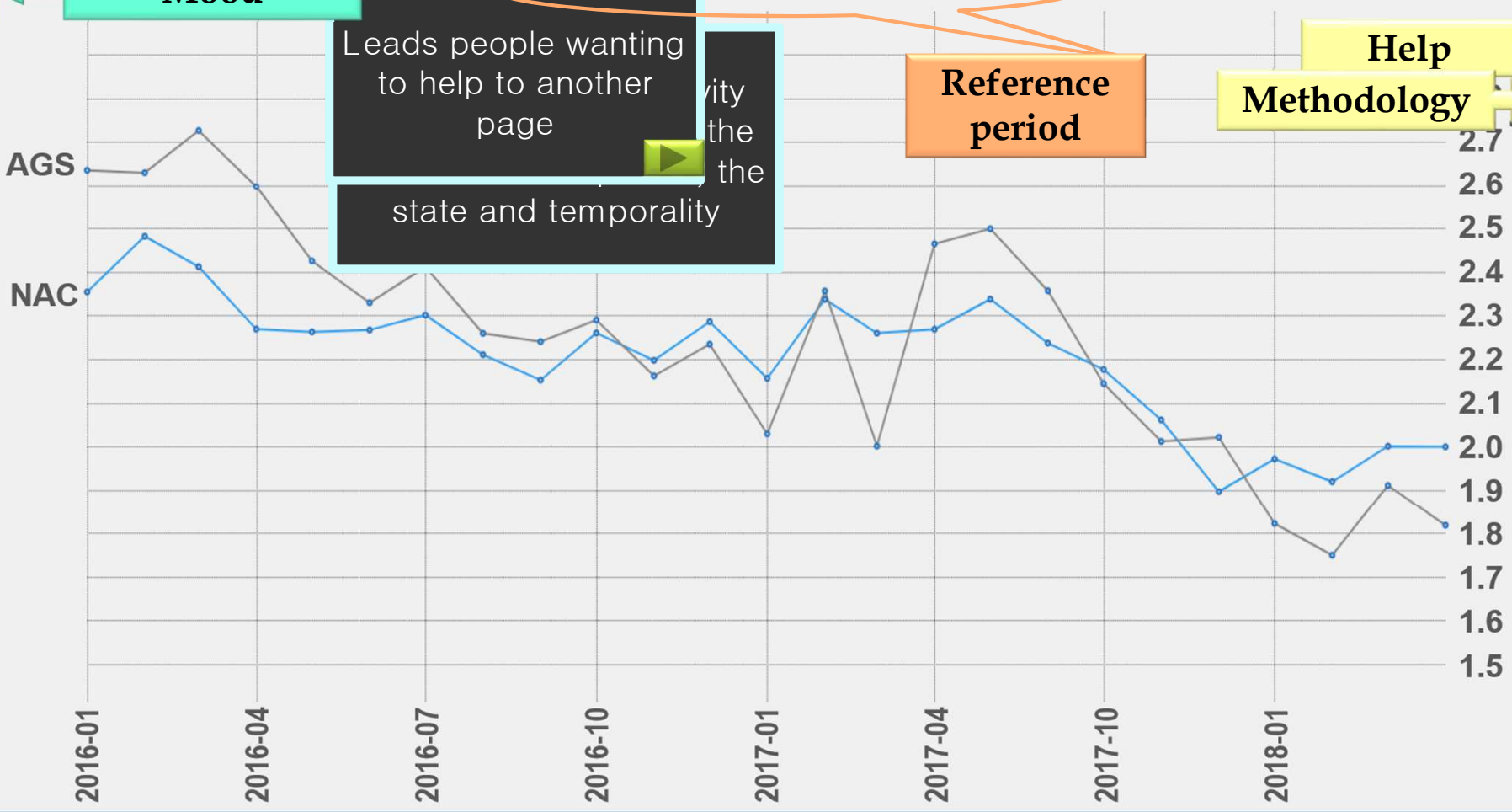
Reference period

Help Methodology

Help



Estado de ánimo de los tuiteros en México



Estado de ánimo de los tuiteros en M

1 de enero de 2016 a 24 de enero de 2018

Entidad federativa

Nuevo León

Nuevo León

Nacional (NAC)

Shows periods for selection

Filtros por fecha

Filtro previo seleccionado: 1 de enero de 2016 - 24 de enero de 2018

Calendar Selection of states

Shows the temporality of the indicator

Shows, at the upper right corner, the National level and a selecting bar for the state of interest

Daily, Weekly, Monthly, Quarterly or Annual Indicator

Últimos 30 días

Últimos 60 días

Últimos 15 días

Últimos 90 días

Cerrar

enero de 2016 - 24 de enero de 2018

1.5

2016-01

2016-04

2016-07

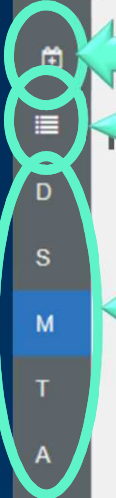
2016-10

2017-01

2017-04

2017-10

2018-01

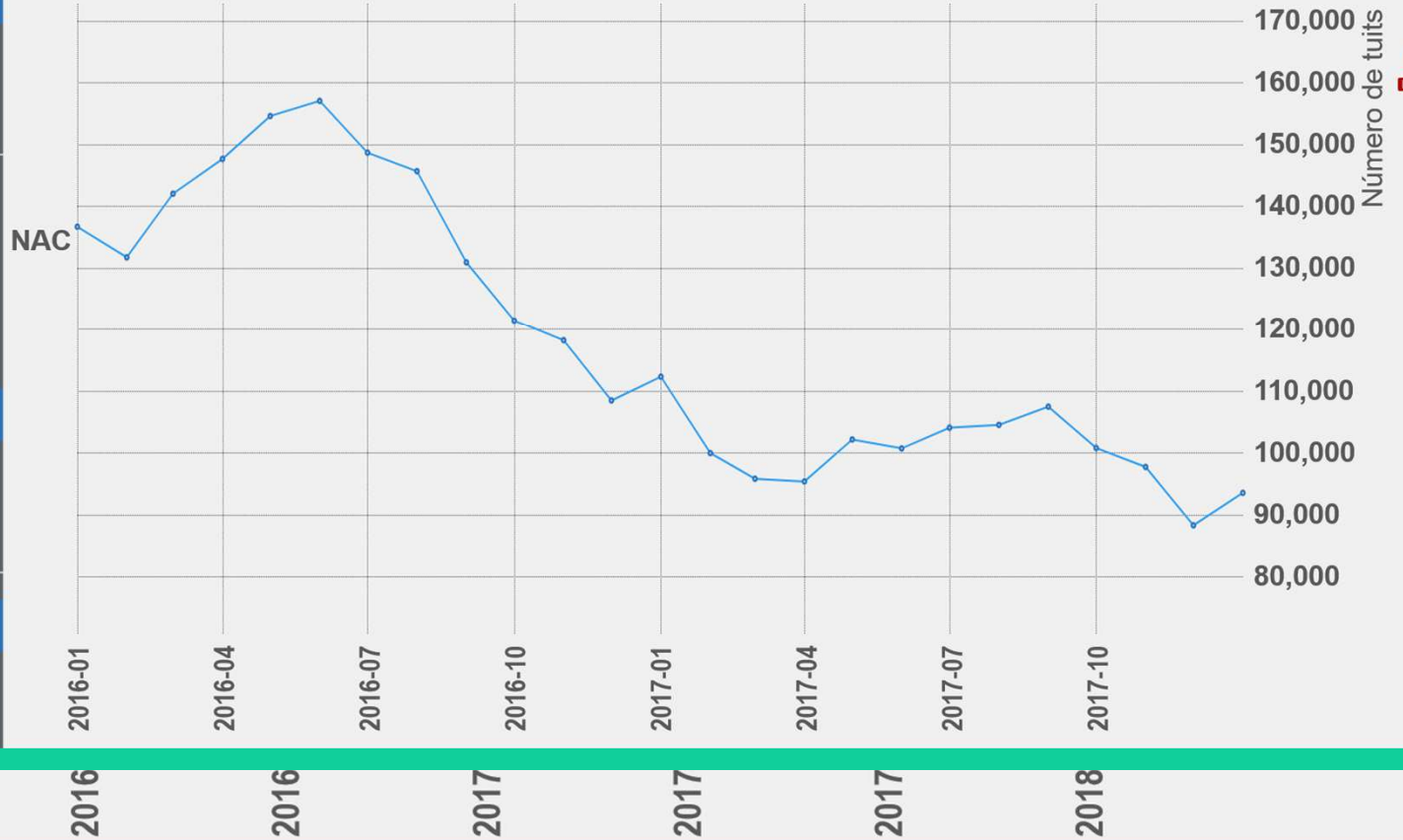


Estado de ánimo de los tuiteros en México

1 de enero de 2016 a 24 de enero de 2018

Promedio diario de recolección de tuits en México

1 de enero de 2016 a 24 de enero de 2018



Gathering

Shows the number of tweets gathered

PDOD23

Slide 27

PDOD23 Daily average of tweets gathered in Mexico
PAEZ DOMINGUEZ OLINCA DESSIREE, 07/09/2018

Mapa del estado de ánimo de los tuiteros en México

1 de enero de 2016 a 24 de enero de 2018



Map

Shows, on the map, the states coloured according to the positivity quotient

PDOD24
PDOD25

Shows the tweets of all people in the state or the country

All

Residents

Visitors

Shows the tweets of people visiting the state

Índice = Positivos (👍) / Negativos (👎)

Slide 28

PDOD24 Map of the mood of tweeterers in Mexico
PAEZ DOMINGUEZ OLINCA DESSIREE, 07/09/2018

PDOD25 From January 1st 2016 to January 24th 2018
PAEZ DOMINGUEZ OLINCA DESSIREE, 07/09/2018



Germany VS Mexico

June 17th 2018



Germany

0

-

1



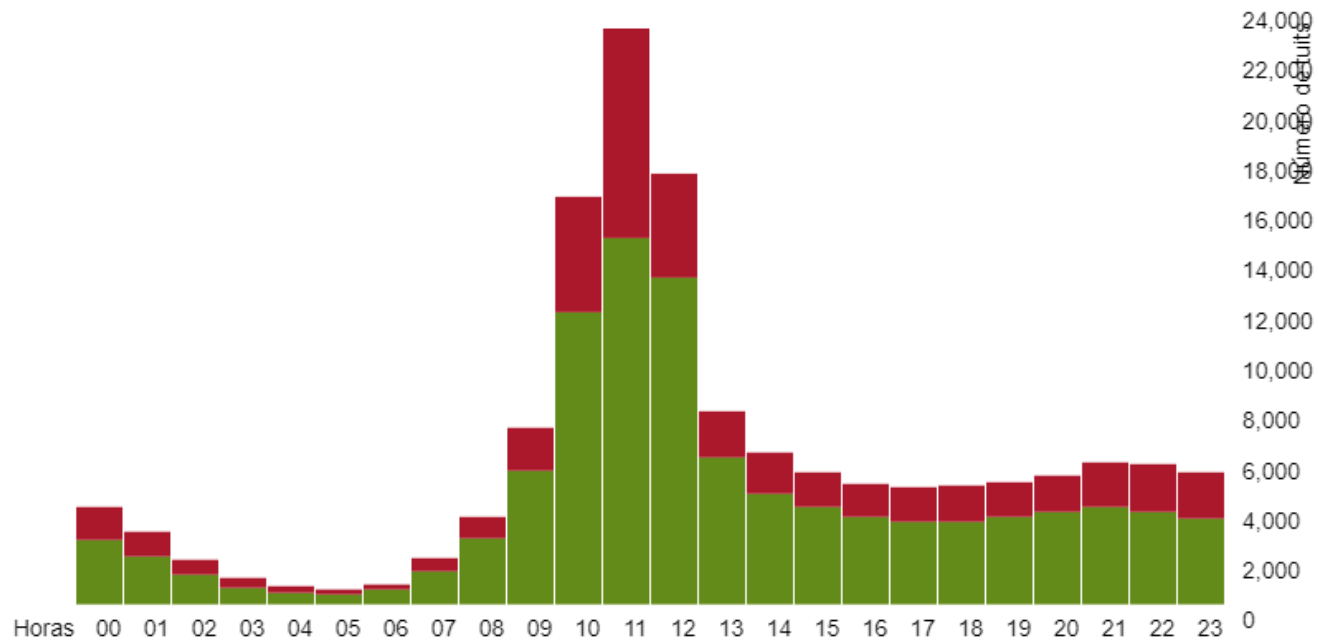
Mexico

Group Stage - Group F - Matchday 1 of 3



Classification by hour of the day

June 17th 2018

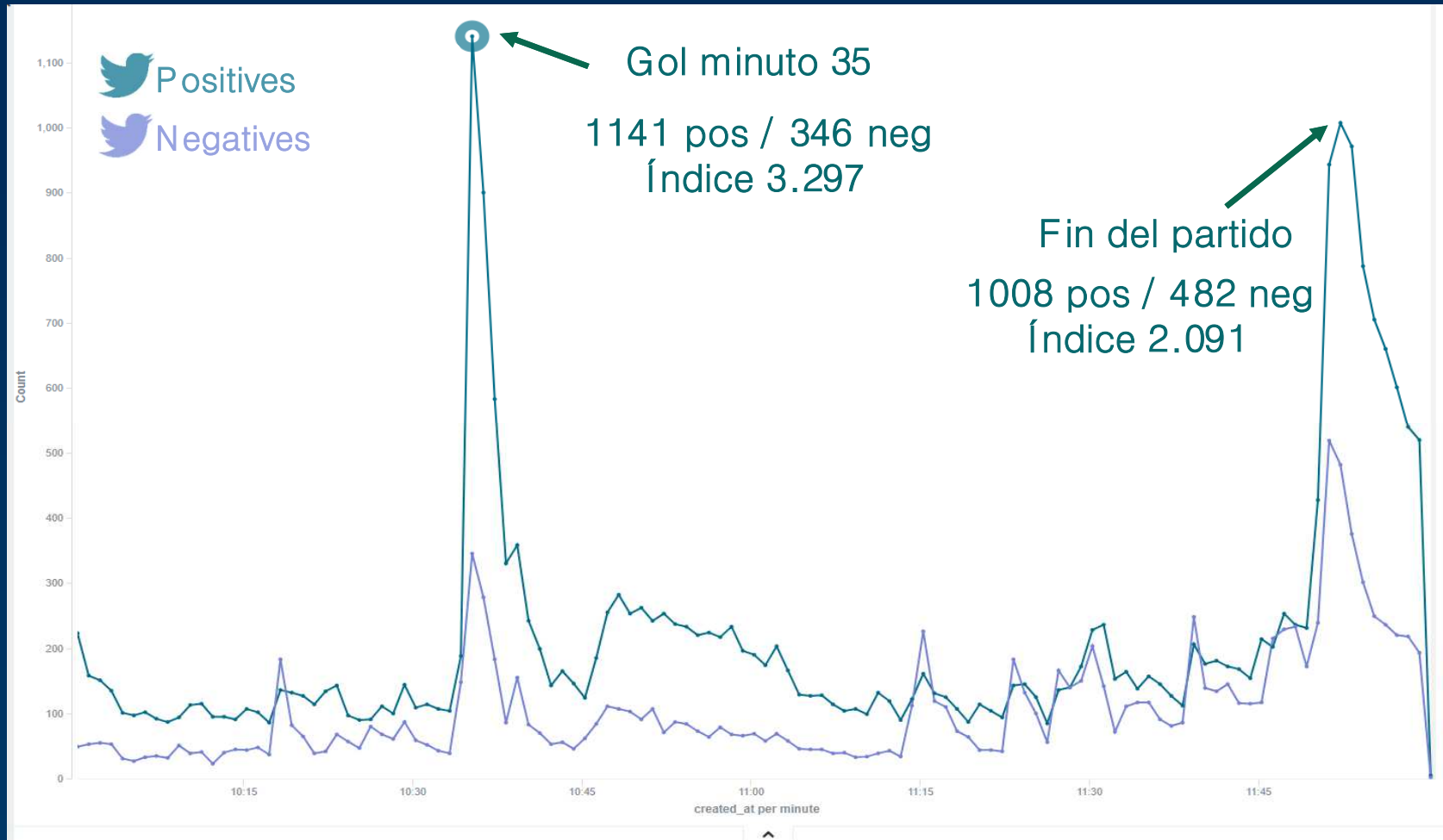


Clasificación

#Hashtags

Noticias

Gathering by minute during the match Germany vs Mexico



Word clouds of all tweets in the minute of each goal Germany vs Mexico

