



El impacto de los sistemas de recomendación en la propagación de la desinformación en redes sociales

Autor: Alfonso de Paz García

Tutor: Alejandro Bellogín Kouki

Índice

1. Introducción
 - i. Motivación
 - ii. Objetivos
2. Sistema desarrollado
 - i. Sistema de crawling
 - ii. Sistema de extracción de Twitter
 - iii. Sistema de generación de puntuaciones
 - iv. Sistema de generación de ficheros para la recomendación
3. Experimentos
 - i. Análisis del dataset final
 - ii. Descripción de los datos generados
 - iii. Recomendadores y métricas
 - iv. Resultados
4. Conclusiones y trabajo futuro

INTRODUCCIÓN

Motivación



**DESINFORMACIÓN
(MISINFORMATION)**

**RECOMMENDATION
SYSTEM**





INTRODUCCIÓN

Objetivos

Creación del dataset



- Proceso de búsqueda de noticias **reales** y **desinformativas**
- Extracción **automatizada** de las noticias.
- Extracción automatizada de información que **relacione** estas noticias **con usuarios**
- Procesar dicha información para obtener el conjunto de datos final

POLITIFACT

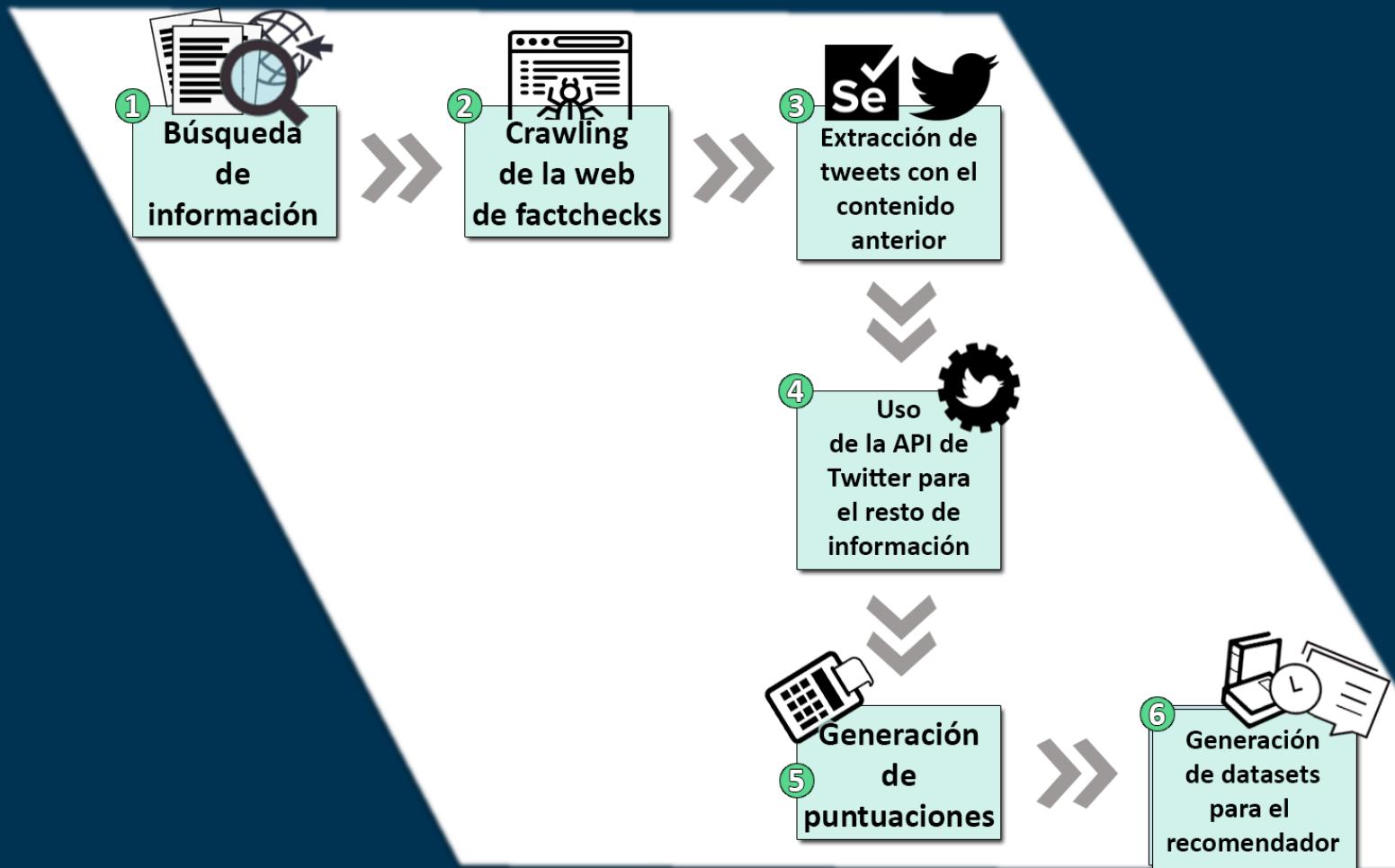


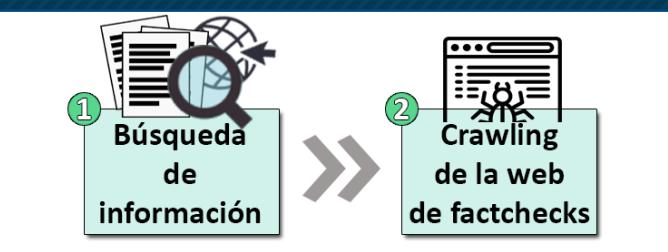
Comportamiento de los sistemas de recomendación

- Proceso de selección de los **recomendadores más representativos** en el área de estudio.
- Entrenar con distintos **subconjuntos del dataset** para probar distintas situaciones.
- Analizar los resultados con **métricas** orientadas a medir la cantidad **de desinformación**.



Sistema desarrollado





Sistema desarrollado

POLITIFACT

STATE EDITIONS

- California
- Florida
- Illinois
- Iowa
- Missouri
- New York
- North Carolina
- Penns
- Texas
- Virginia
- West V
- Vermont
- Wisconsin
- Michigan

ISSUES

- Online hoaxes
- Coronavirus
- Health Care
- Immigration
- Taxes
- Marijuana
- Environment

PEOPLE

- All People
- Donald Trump
- Mike Pence
- Joe Biden
- Kamala Harris
- Bernie Sanders
- Nancy Pelosi

Scorecard

True	Mostly True	Half True	Mostly False
3%	8%	8%	89%
38 Checks	37 Checks	38 Checks	84 Checks

For more details, see the Principles of Politifact

All Fact-checks for Coronavirus

- Mike Pence: Says since he and President Trump took office, the policies "led to an unemployment rate of 3.7% in Wisconsin and 61,000 Wisconsin jobs, including 15,000 manufacturing jobs."
- Facebook posts: Says Anthony Fauci was indicted "for crimes including lying and conspiring with the enemy and fraud."
- Says the "The Republicans are state Republicans, but it represents 17% (coronavirus) deaths in our state."
- Chris Karpinski: From 2018
- John Cornyn: From 2020
- Van Wauquiez: From 2020
- Viral Image: From 2020

Photo of Democratic senators huddling without masks is from 2018

IF YOUR TIME IS SHORT

The photo went viral during the COVID-19 pandemic. It's from September 2018.

See the sources for this fact-check

In the trial of Judge Amy Coney Barrett's Supreme Court confirmation hearings, Republican Sen. John Cornyn accused a group of Democratic colleagues on the Judiciary Committee of failing to wear masks.

Cornyn shared a [photo](#) on Oct. 14 that shows Sen. Sheldon Whitehouse, Richard Blumenthal and Cory Booker huddled together. The caption: "What's missing in this photo?" Cornyn's quote, retweet then added, "Mask."

The original tweet was posted about nine hours before Cornyn retweeted it.

What's missing in this photo? From 2018

Térmetro de Politifact

Verdaderas Falsas



1 MacIver Institute

2 Analysis bashing DHS COVID-19 calculations is built on errors, omissions

3

4

5

6

7

2551

Description - 2 items

XPath location

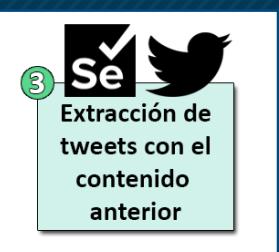
/site[1]/chapter[1]/linksection[1]

XML Outliner

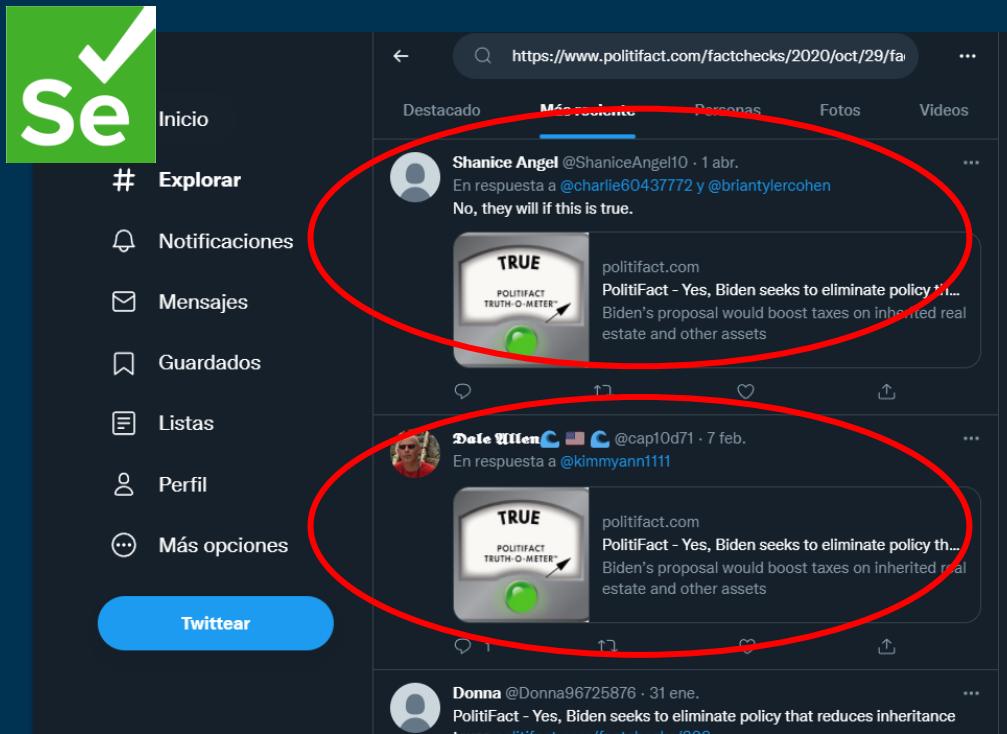
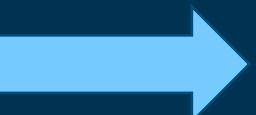
XML Validation and Well-F... /site[1]/chapter[1]/linksection[1]

Scrapy

Scrapy icon: A green circle with a white paintbrush inside.



Cada URL se busca en Twitter de forma automatizada usando Selenium



Sistema desarrollado



Limitaciones de la API de Twitter

- Tweets de hasta 7 días de antigüedad
- 100 tweets por petición
- 180 peticiones por minuto



Con el Identificador del tweet nos saltamos esta restriccción

4



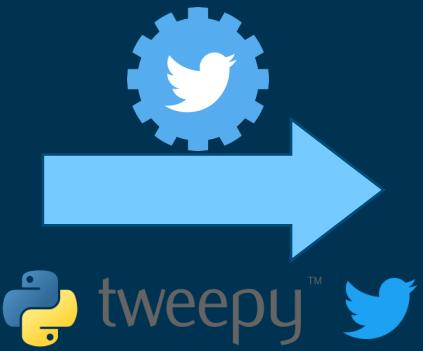
Uso
de la API de
Twitter para
el resto de
información

Sistema desarrollado

Cada tweet extraído con Selenium
se busca en la API

```
https://twitter.com/ShaniceAngel10/status/13774066016198942
https://twitter.com/cap10d71/status/1358274594365448193
https://twitter.com/Donna96725876/status/13559832637104087
https://twitter.com/ShaniceAngel10/status/13774066016198942
https://twitter.com/BigLasagnaMI/status/1355889904635543557
https://twitter.com/ShaniceAngel10/status/13774066016198942
https://twitter.com/Donna96725876/status/13559832637104087
https://twitter.com/BigLasagnaMI/status/1355889904635543557
https://twitter.com/ShaniceAngel10/status/13774066016198942
https://twitter.com/Donna96725876/status/13559832637104087
https://twitter.com/ShaniceAngel10/status/13774066016198942
https://twitter.com/ShaniceAngel10/status/13774066016198942
https://twitter.com/ShaniceAngel10/status/13774066016198942
https://twitter.com/ShaniceAngel10/status/13774066016198942
```

TweetIDs

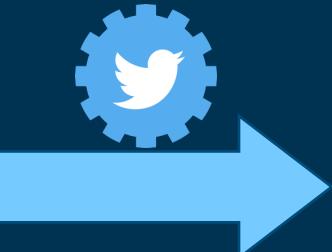


```
{
  "id_str": "1358147955052539904",
  "user_id": "8953122",
  "user_screen_name": "PolitiFact",
  "text": "President Biden hasn't canceled student loans. He has proposed a plan to do so, but it hasn't been finalized. The White House says it's 'under review.'",
  "created_at": {
    "year": 2021,
    "month": 2,
    "day": 6,
    "time_h": 20,
    "time_m": 18,
    "time_s": 0
  },
  "retweet_count": 11,
  "favorite_count": 47,
  "in_reply_to_status_id_str": null,
  "in_reply_to_user_id_str": null,
  "entities": {
    "hashtags": [],
    "user_mentions": [],
    "urls": [
      {
        "url": "https://t.co/KcMaKPRX29",
        "expanded_url": "http://bit.ly/2MSf4lp",
        "display_url": "bit.ly/2MSf4lp",
        "indices": [
          50,
          73
        ]
      }
    ]
  }
},
{
  "id_str": "1357741528022618112",
  "user_id": "8953122",
  "user_screen_name": "PolitiFact"
}
```

Sistema desarrollado



Para cada usuario se extraen
150 tweets de su timeline



Tweet de @Ejemplo

```
{  
  "id_str": "1358147955052539904",  
  "user_id": "8953122",  
  "user_screen_name": "PolitiFact",  
  "text": "President Biden hasn't canceled student lo  
  "created_at": {  
    "year": 2021,  
    "month": 2,  
    "day": 6,  
    "time_h": 20,  
    "time_m": 18,  
    "time_s": 0  
  },  
  "retweet_count": 11,  
  "favorite_count": 47,  
  "in_reply_to_status_id_str": null,  
  "in_reply_to_user_id_str": null,  
  "entities": {  
    "hashtags": [],  
    "user_mentions": [],  
    "urls": [  
      {  
        "url": "https://t.co/KcMaKPRX29",  
        "expanded_url": "http://bit.ly/2MSf4lp",  
        "display_url": "bit.ly/2MSf4lp",  
        "indices": [  
          50,  
          73  
        ]  
      }  
    ]  
  },  
  {  
    "id_str": "1357741528022618112",  
    "user_id": "8953122",  
    "user_screen_name": "PolitiFact" |
```

X150 tweets



Sistema desarrollado

Dataset final de puntuaciones

```
[  
  {  
    "user": "85273142",  
    "rate": 1,  
    "tweet_id_rated": "1377406601619894272",  
    "tweet_id": "1377406601619894272",  
    "tweet_in_reply": null,  
    "claim": "https://www.politifact.com/factchecks/2020/",  
    "claim_in_reply": null  
  },  
  {  
    "user": "2655912266",  
    "rate": -1,  
    "tweet_id_rated": "1358274594365448193",  
    "tweet_id": "1358271156126552064",  
    "tweet_in_reply": "1358274594365448193",  
    "claim": null,  
    "claim_in_reply": "https://www.politifact.com/factche  
  },  
  {  
    "user": "1316970577198870531",  
    "rate": 1,  
    "tweet_id_rated": "1355983263710408706",  
    "tweet_id": "1355983263710408706",  
    "tweet_in_reply": null,  
    "claim": "https://www.politifact.com/factchecks/2020/",  
    "claim_in_reply": null  
  },  
]
```

- **User**
- **Rate (0,1,-1)**
- **tweet_id_rated**
- **tweet_id**
- **tweet_in_reply**
- **claim**
- **claim_in_reply**

Sistema desarrollado



(Tweet Padre)

JSON generado para ese tweet

```
{  
  "user": "85273142",  
  "rate": 1,  
  "tweet_id_rated": "1377406601619894272",  
  "tweet_id": "1377406601619894272",  
  "tweet_in_reply": null,  
  "claim": "https://www.politifact.com/factchecks/2020/oct/29/facebook-posts/yes-biden-",  
  "claim_in_reply": null  
},
```

Claim



Sistema desarrollado

Tweet puntuado (el del padre)

Charlie "Charlie Pfohl" Pfohl
@charlie60437772

En respuesta a @ShaniceAngel10 y @briantylercohen

Your taxes will not change then

Traducir Tweet

1:15 a. m. · 1 abr. 2021 · Twitter for iPhone

(Claim del hijo)

TRUE

POLITIFACT TRUTH-O-METER

politifact.com

PolitiFact - Yes, Biden seeks to eliminate policy that reduce... Biden's proposal would boost taxes on inherited real estate and other assets

(Tweet ID del hijo)

1358274594365448193

(Tweet ID del padre)

1377399121565282306



JSON generado para ese tweet

```
{  
  "user": "2655912266",  
  "rate": -1,  
  "tweet_id_rated": "1377406601619894272", (Tweet ID del hijo)  
  "tweet_id": "1377399121565282306", (Tweet ID del padre)  
  "tweet_in_reply": "1377406601619894272", (Tweet ID del hijo)  
  "claim": null,  
  "claim_in_reply": "https://www.politifact.com/factchecks/2020/oct/29/facebook-posts/yes-biden-"}  
,
```



Sistema desarrollado

Tweet neutral del timeline del usuario



Shanice Angel
@ShaniceAngel10

The fact that this needs to even be said is proof of how broken this country is.

Traducir Tweet

4:35 a. m. · 4 ago. 2021 · Twitter Web App

...

Reply Retweet Like Share



JSON generado para ese tweet

```
{  
  "user": "2655912266",  
  "rate": 0,  
  "tweet_id_rated": null,  
  "tweet_id": "1422747972295708675",  
  "tweet_in_reply": null,  
  "claim": null,  
  "claim_in_reply": null  
},
```



Sistema desarrollado

Hilo de un mismo autor

 **Shanice Angel**
@ShaniceAngel10 ...

The fact that this needs to even be said is proof of how broken this country is.

[Traducir Tweet](#)

4:35 a. m. · 4 ago. 2021 · Twitter Web App

 **Shanice Angel**
@ShaniceAngel10 ...

En respuesta a [@charlie60437772](#) y [@briantylercohen](#)

No, they will if this is true.

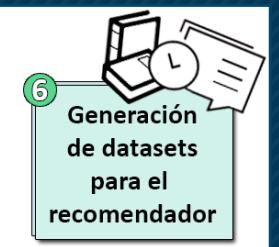
[Traducir Tweet](#)



politifact.com

PolitiFact - Yes, Biden seeks to eliminate policy that reduce...

Biden's proposal would boost taxes on inherited real estate and other assets



Sistema desarrollado

Dataset final

```
{  
  "user": "8953122",  
  "rate": 1,  
  "tweet_id_rated": "1357741528022618112",  
  "tweet_id": "1357741528022618112",  
  "tweet_in_reply": null,  
  "claim": "https://www.politifact.com/factchecks/2021/feb/04/facebook-posts/joe",  
  "claim_in_reply": null  
},  
{  
  "user": "902268741844692992",  
  "rate": 1,  
  "tweet_id_rated": "1357533915486056450",  
  "tweet_id": "1357533915486056450",  
  "tweet_in_reply": null,  
  "claim": "https://www.politifact.com/factchecks/2021/feb/04/facebook-posts/joe",  
  "claim_in_reply": null  
},  
{  
  "user": "364593753",  
  "rate": 1,  
  "tweet_id_rated": "1357382676655833089",  
  "tweet_id": "1357382676655833089",  
  "tweet_in_reply": null,  
  "claim": "https://www.politifact.com/factchecks/2021/feb/04/facebook-posts/joe",  
  "claim_in_reply": null  
},  
{  
  "user": "4924987672",  
  "rate": 1,  
  "tweet_id_rated": "1357359343990149120",  
  "tweet_id": "1357359343990149120",  
  "tweet_in_reply": null,  
  "claim": "https://www.politifact.com/factchecks/2021/feb/04/facebook-posts/joe",  
  "claim_in_reply": null  
},  
{  
  "user": "738935939808927744",  
  "rate": 1,  
  "tweet_id_rated": "1351698393723236352",  
  "tweet_id": "1351698393723236352",  
  "tweet_in_reply": null,  
  "claim": "https://www.politifact.com/factchecks/2020/oct/07/facebook-posts/no-",  
  "claim_in_reply": null  
},  
{  
  "user": "395192293",  
  "rate": 1,
```



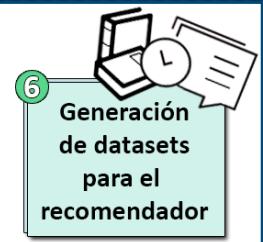
Datasets generados aleatoriamente para la recomendación

UserID	ItemID	Rate
1344356576786866176	512__b	0.521
93215780	546__g	5.279
121415	UserID	ItemID
	1344356576786866176	512__b
	93215780	546__g
	12141545	UserID
		ItemID
		Rate
	1344356576786866176	512__b
93819	93215780	546__g
	1214154537314832391	564__b
	23452847	443__b
115265	9381972	2814186879
70269		801__g
		415196079
		626__g
		1480259144
		129__g
	11526555	938197264010444800
88836	7026935	44175599
115123		632__g
		14091197
		40__g
	115265558344635648	874__b
	8883659	7026935011109273
	11512393	17919393
		6__b
		22555372
		140__g
	888365903930417154	56__g
	1151239321074184192	208__b
		4520484798
		208__g
		161142392
		72__b
	902267	3519064697
120474		616__b
	52320662	616__b
	9022687	2835451658
		504__b
	342833326	854__b
	12047485	197371392
		845__b
	20644656	854__g
	902268741844692992	696__g
	1204748512275120128	338__g
		28448196
		895__g
	747153392	882__g
		154697628
		286__g
	1307768174	845__b
		18584875
		845__b
	4924987672	969__g

Formato JSON

Formato TSV

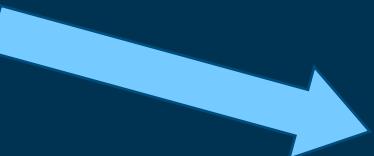
Sistema desarrollado



Dataset final

```
{
  "user": "8953122",
  "rate": 1,
  "tweet_id": "1157741520022618112",
  "tweet_id_r": "1157741520022618112",
  "tweet_in_reply": null,
  "claim": "https://www.politifact.com/factchecks/2021/feb/04/facebook-posts/joe",
  "claim_in_reply": null
},
{
  "user": "98026804188402992",
  "rate": 1,
  "tweet_id": "1157741520022618112",
  "tweet_id_r": "1157741520022618112",
  "tweet_in_reply": null,
  "claim": "https://www.politifact.com/factchecks/2021/feb/04/facebook-posts/joe",
  "claim_in_reply": null
},
{
  "user": "36459375",
  "rate": 1,
  "tweet_id": "1157741520022618112",
  "tweet_id_r": "1157741520022618112",
  "tweet_in_reply": null,
  "claim": "https://www.politifact.com/factchecks/2021/feb/04/facebook-posts/joe",
  "claim_in_reply": null
},
{
  "user": "45949872",
  "rate": 1,
  "tweet_id": "1157741520022618112",
  "tweet_id_r": "1157741520022618112",
  "tweet_in_reply": null,
  "claim": "https://www.politifact.com/factchecks/2021/feb/04/facebook-posts/joe",
  "claim_in_reply": null
},
{
  "user": "73893593988927744",
  "rate": 1,
  "tweet_id": "1157741520022618112",
  "tweet_id_r": "1157741520022618112",
  "tweet_in_reply": null,
  "claim": "https://www.politifact.com/factchecks/2021/feb/04/facebook-posts/joe",
  "claim_in_reply": null
},
{
  "user": "395392293",
  "rate": 1,
}
```

(proporción de desinformación)



Dataset final

```
{
  "user": "8953122",
  "rate": 1,
  "tweet_id": "1157741520022618112",
  "tweet_id_r": "1157741520022618112",
  "tweet_in_reply": null,
  "claim": "https://www.politifact.com/factchecks/2021/feb/04/facebook-posts/joe",
  "claim_in_reply": null
},
{
  "user": "98026804188402992",
  "rate": 1,
  "tweet_id": "1157741520022618112",
  "tweet_id_r": "1157741520022618112",
  "tweet_in_reply": null,
  "claim": "https://www.politifact.com/factchecks/2021/feb/04/facebook-posts/joe",
  "claim_in_reply": null
},
{
  "user": "36459375",
  "rate": 1,
  "tweet_id": "1157741520022618112",
  "tweet_id_r": "1157741520022618112",
  "tweet_in_reply": null,
  "claim": "https://www.politifact.com/factchecks/2021/feb/04/facebook-posts/joe",
  "claim_in_reply": null
},
{
  "user": "45949872",
  "rate": 1,
  "tweet_id": "1157741520022618112",
  "tweet_id_r": "1157741520022618112",
  "tweet_in_reply": null,
  "claim": "https://www.politifact.com/factchecks/2021/feb/04/facebook-posts/joe",
  "claim_in_reply": null
},
{
  "user": "73893593988927744",
  "rate": 1,
  "tweet_id": "1157741520022618112",
  "tweet_id_r": "1157741520022618112",
  "tweet_in_reply": null,
  "claim": "https://www.politifact.com/factchecks/2021/feb/04/facebook-posts/joe",
  "claim_in_reply": null
},
{
  "user": "45949872",
  "rate": 1,
  "tweet_id": "1157741520022618112",
  "tweet_id_r": "1157741520022618112",
  "tweet_in_reply": null,
  "claim": "https://www.politifact.com/factchecks/2021/feb/04/facebook-posts/joe",
  "claim_in_reply": null
},
{
  "user": "73893593988927744",
  "rate": 1,
  "tweet_id": "1157741520022618112",
  "tweet_id_r": "1157741520022618112",
  "tweet_in_reply": null,
  "claim": "https://www.politifact.com/factchecks/2021/feb/04/facebook-posts/joe",
  "claim_in_reply": null
},
{
  "user": "395392293",
  "rate": 1,
}
```

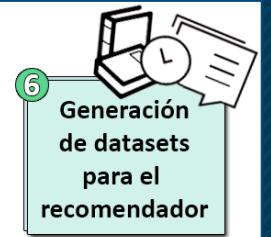
(proporción de neutralidad)

(con los tweets neutrales obtenido de los timelines)

Dataset generado aleatoriamente para la recomendación

UserID	ItemID	Rate
1344356576786866176	512_b	0.521
	93215780	5.279
1214154537314832391	564_b	1.196
	23452847	1.051
	2814186879	5.055
	415196079	5.432
	1480259144	4.823
938197264010444800	564_b	1.182
	44175599	3.463
	14091197	2.987
1152655583444635648	1357533915486056450	3.345
702693550111092737	435753324448605678	2.767
	17919393	1.392
	22555372	5.481
888365903930417154	56_g	5.069
1151239321074184192	208_b	0.697
	4520484798	4.991
	161142392	1.387
	3519064697	616_b
	52320662	1.182
	616_b	1.161
2835451658	504_b	1.39
342833326	854_b	1.282

→ (Neutralidad)



Sistema desarrollado

Dataset final

```
{
  "user": "49011122",
  "rate": 1,
  "tweet_id": "135745102002181127",
  "tweet_id_r": "135745102002181127",
  "tweet_in_reply": null,
  "class_in_reply": "www.politifact.com/factchecks/2021/feb/04/facebook-posts/no"
  "class_in_reply_r": null
},
{
  "user": "8902871846992992",
  "rate": 1,
  "tweet_id": "135733915486564548",
  "tweet_id_r": "135733915486564548",
  "tweet_in_reply": null,
  "class_in_reply": "www.politifact.com/factchecks/2021/feb/04/facebook-posts/no"
  "class_in_reply_r": null
},
{
  "user": "25659315",
  "rate": 1,
  "tweet_id": "135745102002181127",
  "tweet_id_r": "135745102002181127",
  "tweet_in_reply": null,
  "class_in_reply": "www.politifact.com/factchecks/2021/feb/04/facebook-posts/no"
  "class_in_reply_r": null
},
{
  "user": "490489877",
  "rate": 1,
  "tweet_id": "135733915486564548",
  "tweet_id_r": "135733915486564548",
  "tweet_in_reply": null,
  "class_in_reply": "www.politifact.com/factchecks/2021/feb/04/facebook-posts/no"
  "class_in_reply_r": null
},
{
  "user": "72093590880897746",
  "rate": 1,
  "tweet_id": "135733915486564548",
  "tweet_id_r": "135733915486564548",
  "tweet_in_reply": null,
  "class_in_reply": "www.politifact.com/factchecks/2021/feb/04/facebook-posts/no"
  "class_in_reply_r": null
},
{
  "user": "39593203",
  "rate": 1,
  "tweet_id": "135733915486564548",
  "tweet_id_r": "135733915486564548",
  "tweet_in_reply": null,
  "class_in_reply": "www.politifact.com/factchecks/2020/oct/07/facebook-posts/no"
  "class_in_reply_r": null
}
}
```

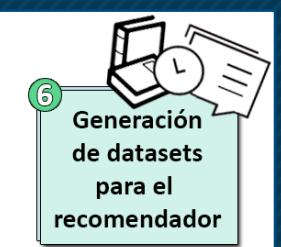
Proporción de desinformación:

- En función de las tuplas
- En función de los usuarios



Dataset generado aleatoriamente para la recomendación

UserID	ItemID	Rate
1344356576786866176	512_b	0.521
93215780	546_g	5.279
1214154537314832391	564_b	1.196
23452847	443_b	1.051
2814186879	801_g	5.055
415196079	626_g	5.432
1480259144	129_g	4.823
938197264010444800	564_b	1.182
44175599	234757654486057654	3.463
14091197	126753391548612265	2.987
1152655583444635648	1357533915486056450	3.345
702693550111092737	435753324448605678	2.767
17919393	6_b	1.392
22555372	140_g	5.481
888365903930417154	56_g	5.069
1151239321074184192	208_b	0.697
4520484798	208_g	4.991
161142392	72_b	1.387
3519064697	616_b	1.182
52320662	616_b	1.161
2835451658	504_b	1.39
342833326	854_b	1.282



Sistema desarrollado

Dataset final

```
{
  "user": "49011122",
  "rate": 1,
  "tweet_id": "1357445103802181127",
  "tweet_id_rate": "1357445103802181127",
  "tweet_in_reply": null,
  "class_in_reply": "www.politifact.com/factchecks/2021/feb/04/facebook-posts/no_claim_in_reply"
},
{
  "user": "8902818166992992",
  "rate": 1,
  "tweet_id": "1357439515480564548",
  "tweet_id_rate": "1357439515480564548",
  "tweet_in_reply": null,
  "class_in_reply": "www.politifact.com/factchecks/2021/feb/04/facebook-posts/no_claim_in_reply"
},
{
  "user": "25659315",
  "rate": 1,
  "tweet_id": "1357439515480564548",
  "tweet_id_rate": "1357439515480564548",
  "tweet_in_reply": null,
  "class_in_reply": "www.politifact.com/factchecks/2021/feb/04/facebook-posts/no_claim_in_reply"
},
{
  "user": "459489577",
  "rate": 1,
  "tweet_id": "1357439515480564548",
  "tweet_id_rate": "1357439515480564548",
  "tweet_in_reply": null,
  "class_in_reply": "www.politifact.com/factchecks/2021/feb/04/facebook-posts/no_claim_in_reply"
},
{
  "user": "728935919808937747",
  "rate": 1,
  "tweet_id": "1351089372258352",
  "tweet_id_rate": "1351089372258352",
  "tweet_in_reply": null,
  "class_in_reply": "www.politifact.com/factchecks/2020/oct/07/facebook-posts/no_claim_in_reply"
},
{
  "user": "3955923207",
  "rate": 1,
  "...
```

Proporción de neutralidad:

**EJ: 150 tweets por 6025 users
Un 25% de neutralidad →**

37 tweets x 6025 = ~230K elem

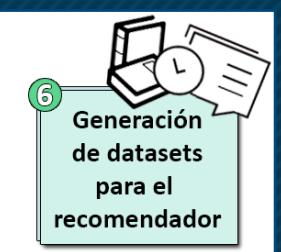


(con los tweets neutrales obtenido de los timelines)

Dataset generado aleatoriamente para la recomendación

UserID	ItemID	Rate
1344356576786866176	512_b	0.521
93215780	546_g	5.279
1214154537314832391	564_b	1.196
23452847	443_b	1.051
2814186879	801_g	5.055
415196079	626_g	5.432
1480259144	129_g	4.823
938197264010444800	564_b	1.182
44175599	234757654486057654	3.463
14091197	126753391548612265	2.987
115265558344635648	1357533915486056450	3.345
702693550111092737	435753324448605678	2.767
17919393	6_b	1.392
22555372	140_g	5.481
888365903930417154	56_g	5.069
1151239321074184192	208_b	0.697
4520484798	208_g	4.991
161142392	72_b	1.387
3519064697	616_b	1.182
52320662	616_b	1.161
2835451658	504_b	1.39
342833326	854_b	1.282

→ (Neutralidad)



6
Generación
de datasets
para el
recomendador

Sistema desarrollado

Dataset final

```
{  
  "user": "50188272",  
  "rate": -1,  
  "tweet_id_rated": "1337140529826435072",  
  "tweet_id": "1327007193661575168",  
  "tweet_in_reply": null,  
  "claim": null,  
  "claim_in_reply": "https://www.politifact.com/factchecks/2020/nov/  
},  
{  
  "user": "15723081",  
  "rate": -1,  
  "tweet_id_rated": "1326910757393608704",  
  "tweet_id": "1326631254545018884",  
  "tweet_in_reply": null,  
  "claim": "https://www.politifact.com/factchecks/2020/nov/10/insta",  
  "claim_in_reply": null,  
},  
{  
  "user": "37943826",  
  "rate": 1,  
  "tweet_id_rated": "1326355666475347970",  
  "tweet_id": "1326355666475347970",  
  "tweet_in_reply": "1326631254545018884",  
  "claim": "https://www.politifact.com/factchecks/2020/nov/10/insta",  
  "claim_in_reply": null  
},
```



(Claims mapeadas)

Archivo TSV generado

50188272	539	__b	-1
15723081	539	__b	-1
37943826	539	__g	1

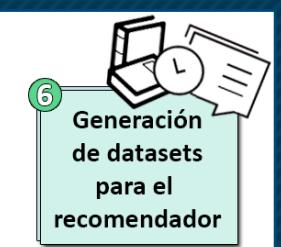
Archivo de mapeo de las claims

538	<https://www.politifact.com/
539	<https://www.politifact.com/
540	<https://www.politifact.com/

Claims etiquetadas:

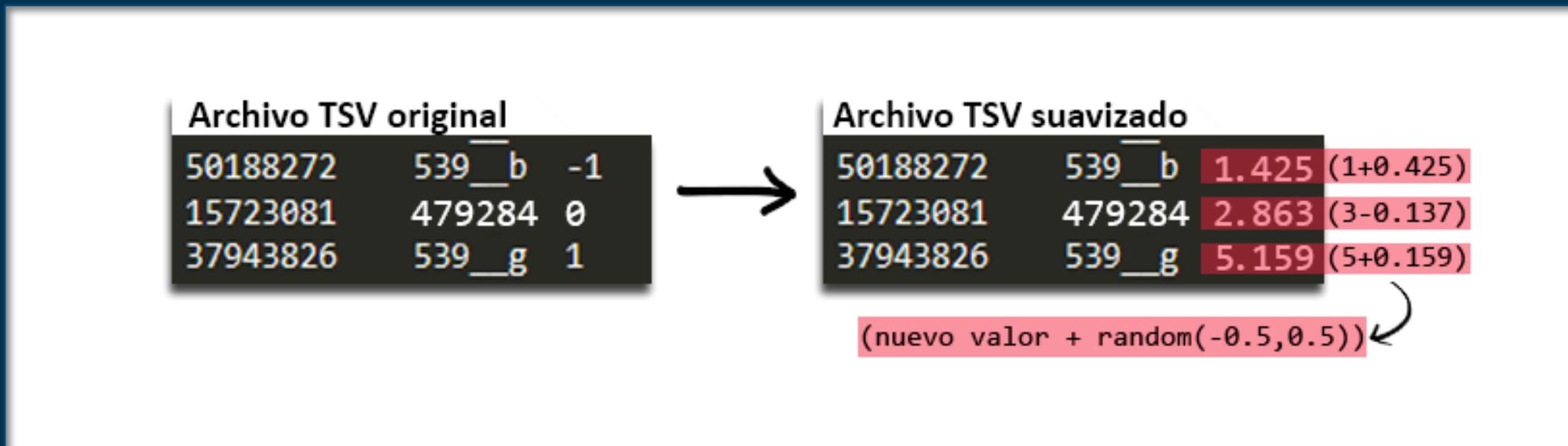
- __b → El usuarios está relacionado con dicha claim por **desinformar**.
- __g → El usuarios está relacionado con dicha claim por usarla para **verificar**.

Tener los claims etiquetados nos permite saber la cantidad de desinformación recomendada a un usuario



Sistema desarrollado

Suavizado de los datos



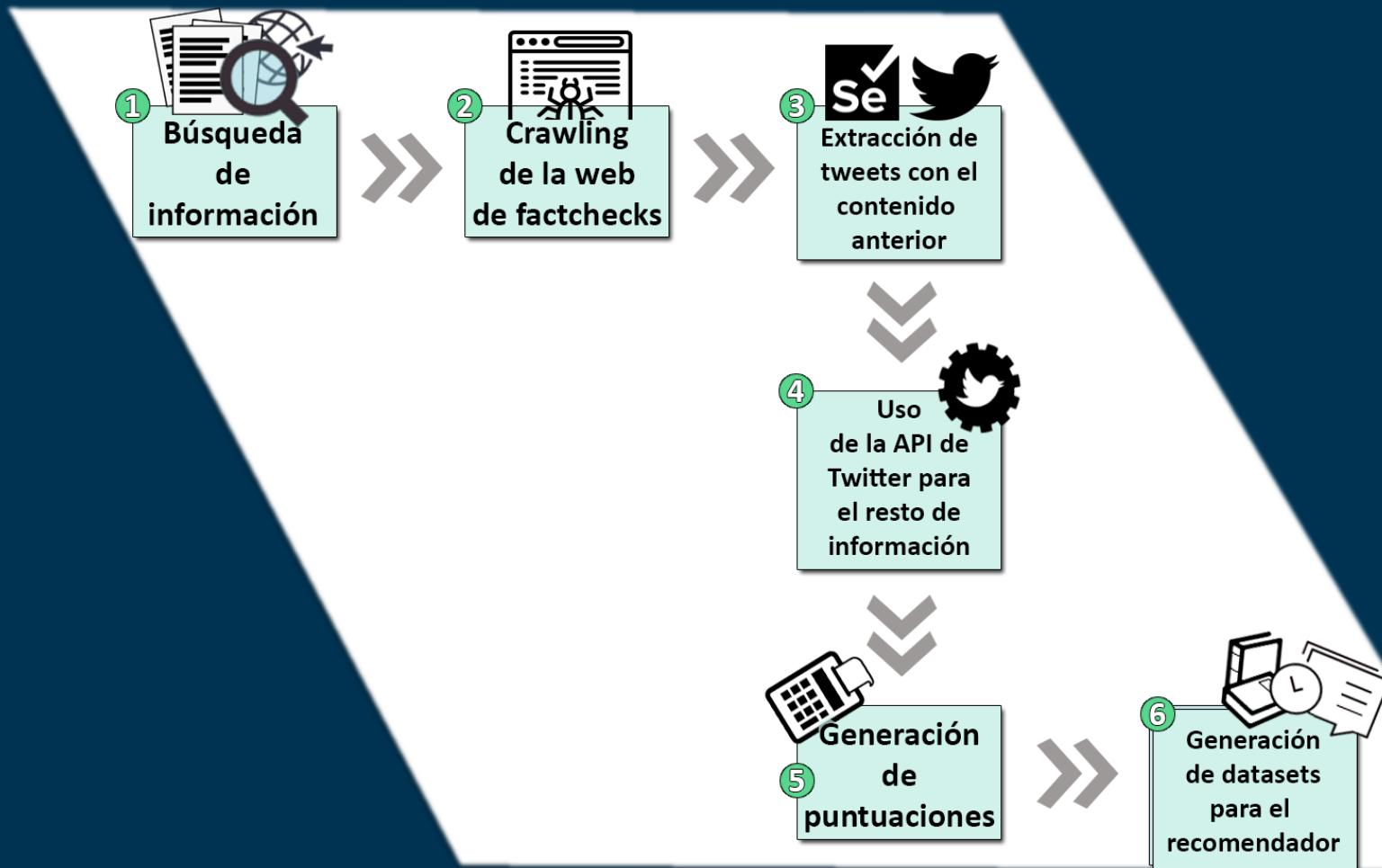
Objetivos para mejorar la recomendación:

- Evitar tener valores negativos
- Darle variedad a los datos

Pasos:

- -1 → 1
- 0 → 3
- 1 → 5
- Ruido random de [-0.5 – 0.5]

Sistema desarrollado



Experimentos

Análisis del dataset final



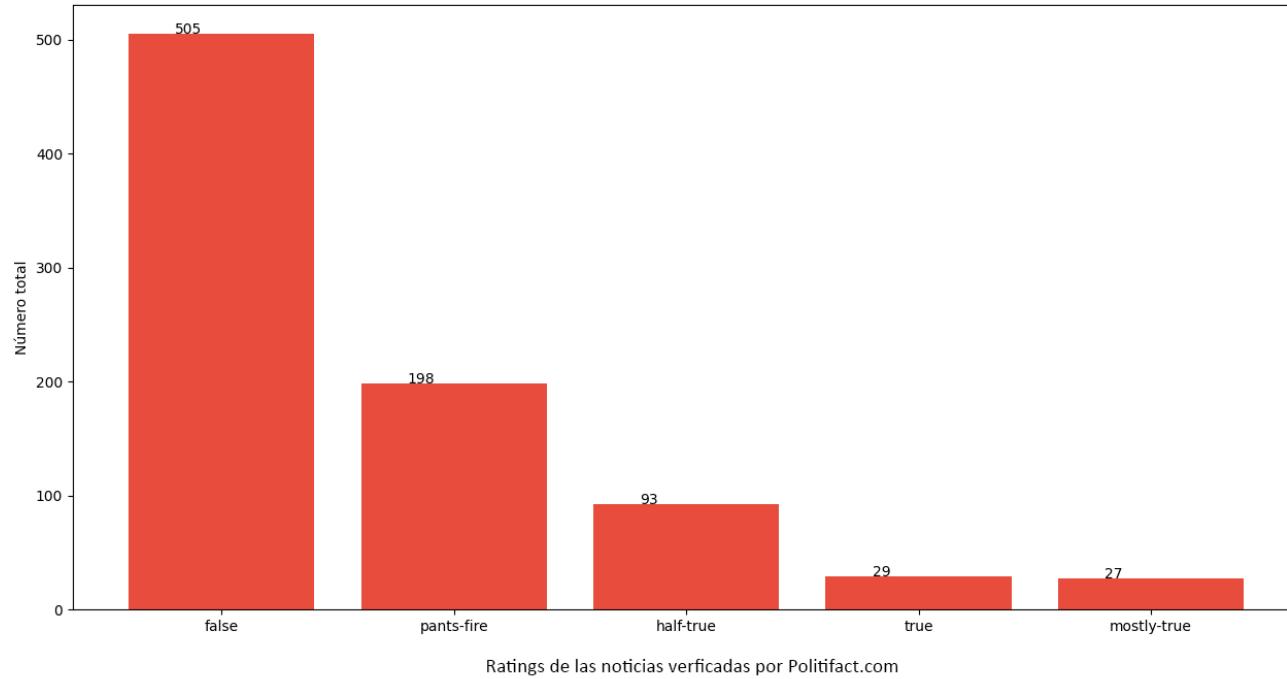
- Entre enero de 2020 y febrero de 2021
- Actualidad de Estados Unidos



Experimentos

Análisis del dataset final

Politifact



Falsos : false y pants-fire

Verdaderos : half-true, true y mostly-true

Politifact (elementos verificados)

703 elementos como **falsos**

149 elementos como **verdaderos**

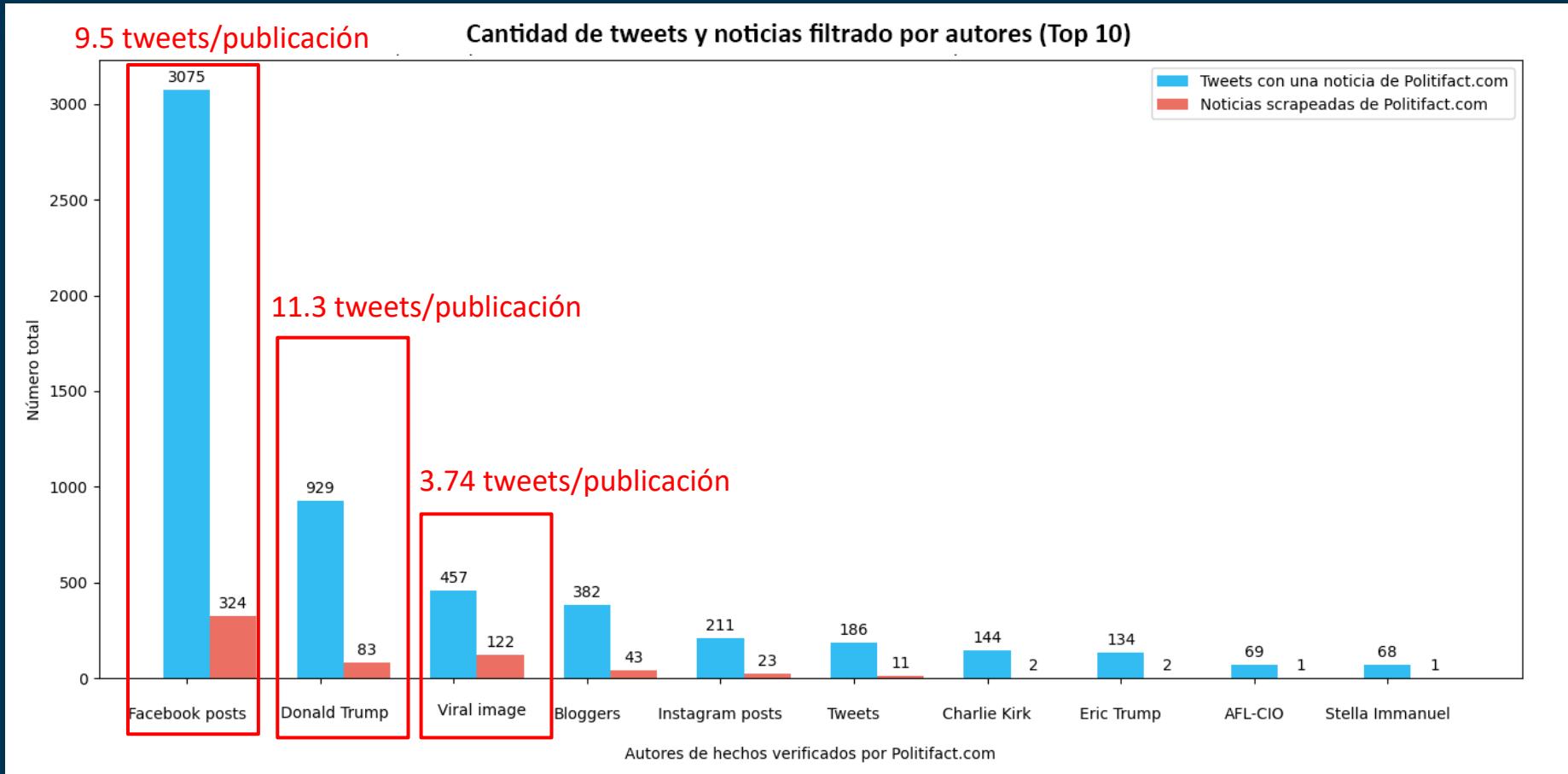
Twitter (tweets que comparten estas publicaciones de Politifact)

5792 tweets con claims puntuadas como **falsas**

415 tweets con claims puntuadas como **verdaderas**

Experimentos

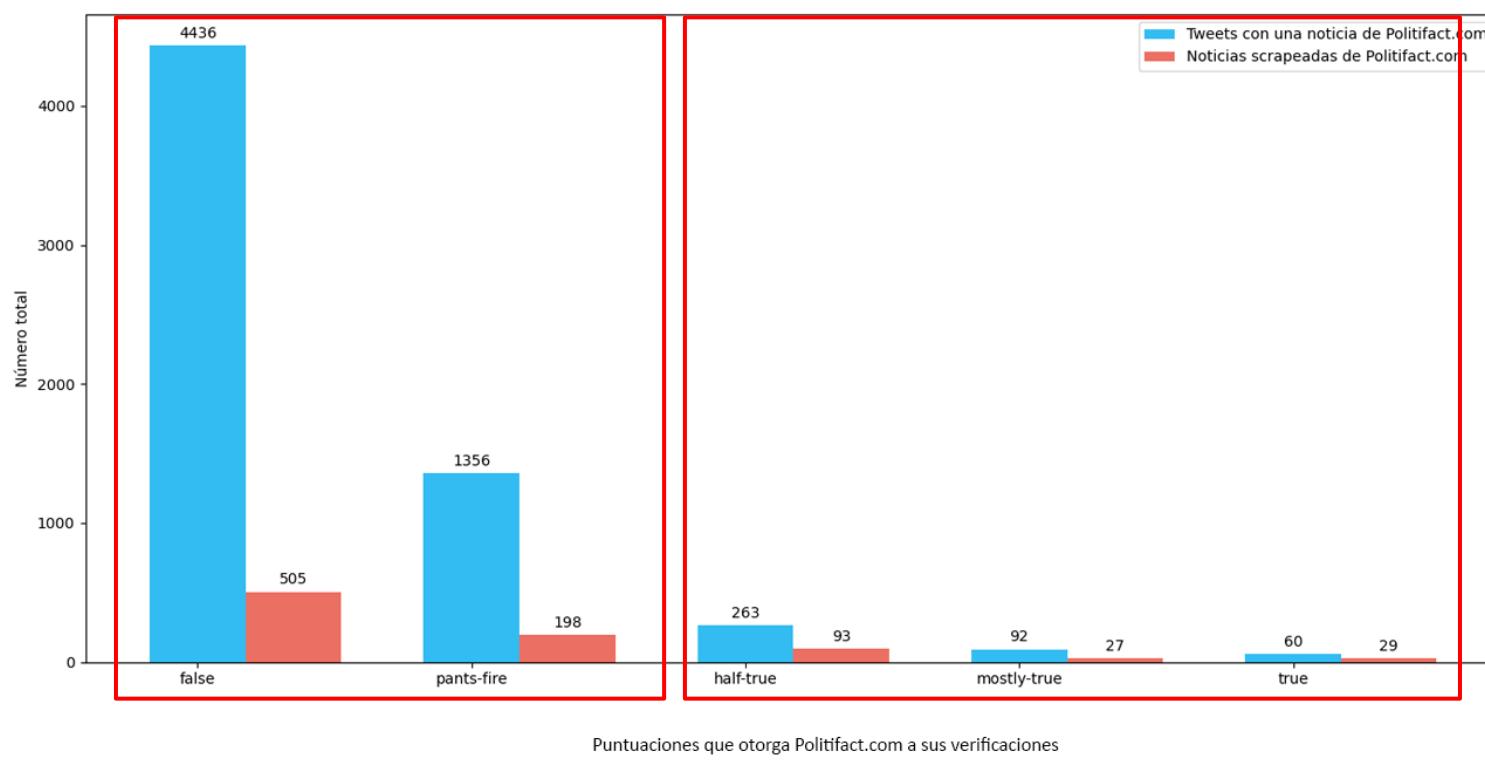
Análisis del dataset final



Una mayor proporción indica un mayor interés por compartir dichas verificaciones en la red social

¿Qué se comparte más?

Cantidad de tweets y noticias filtrado por puntuaciones (Top 10)



Interacción en twitter

8.24 tweets/publicación con puntuación **falsa**

2.78 tweets/publicación con puntuación **verdadera**

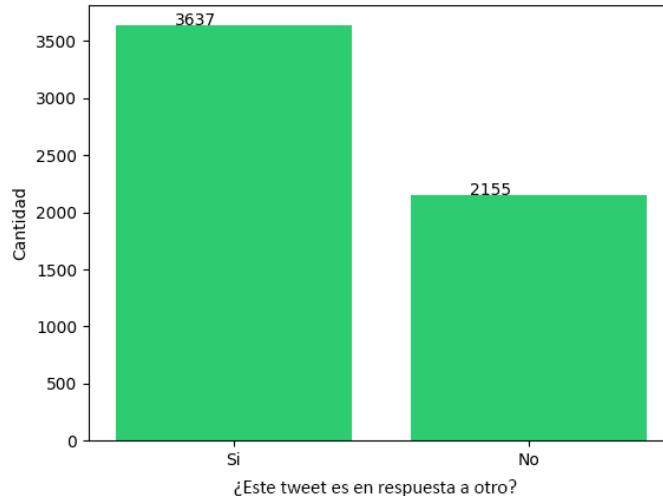
x4!!

Experimentos

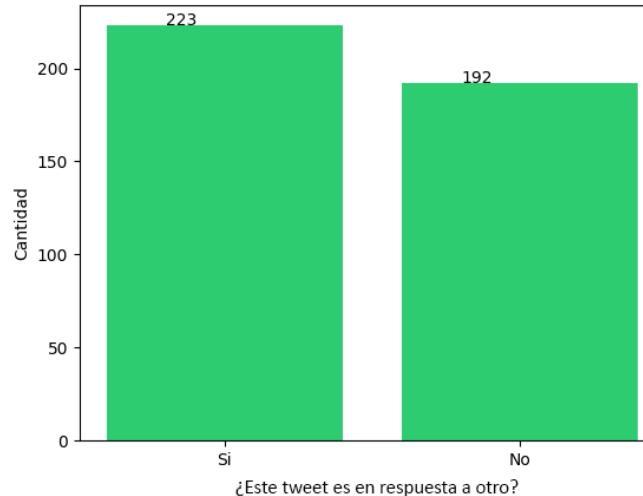
Análisis del dataset final

¿Cuándo se responde más?

Tweets en respuesta a otro
(filtrado por elementos de Politifact falsos)



Tweets en respuesta a otro
(filtrado por elementos de Politifact verdaderos)



Es más fácil desinformar desmintiendo una noticia verdadera que inventando algo



51% son tweets en respuesta

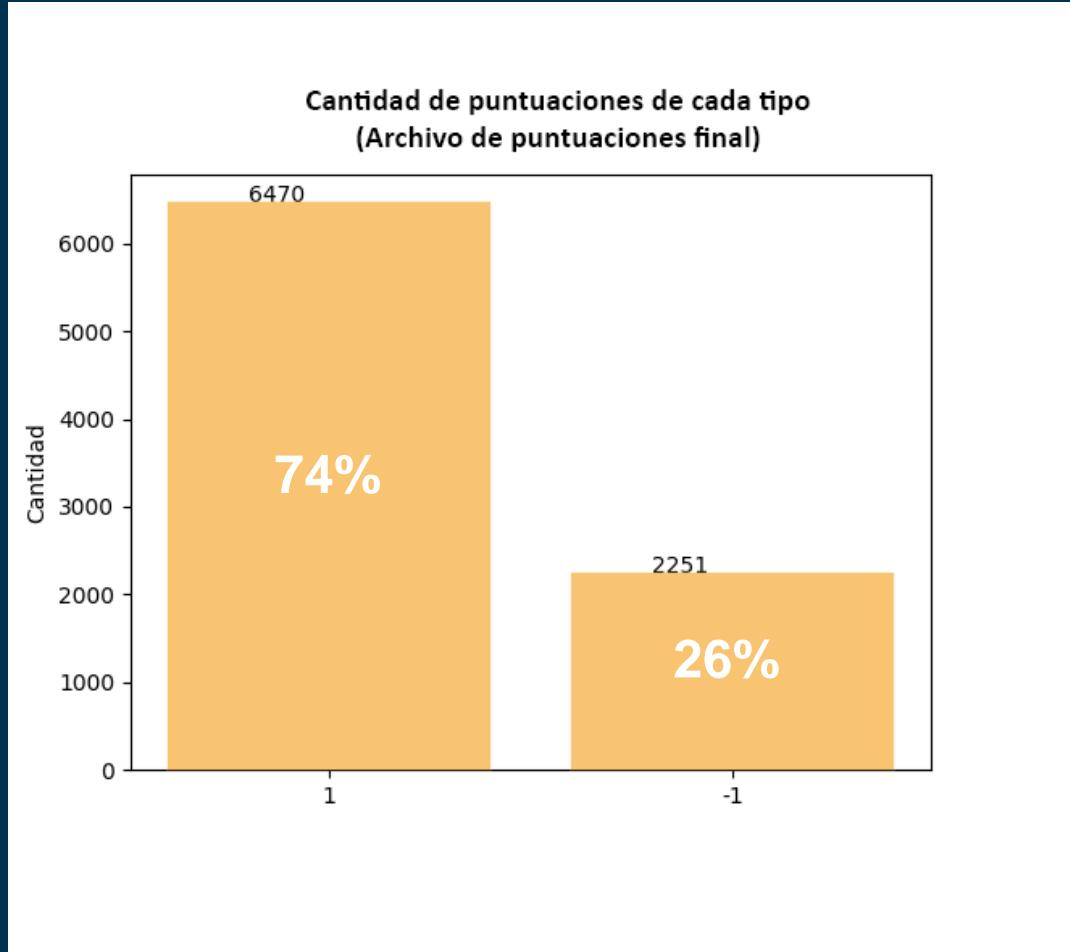


62% son tweets en respuesta

Experimentos

¿Cómo está repartido el dataset final?

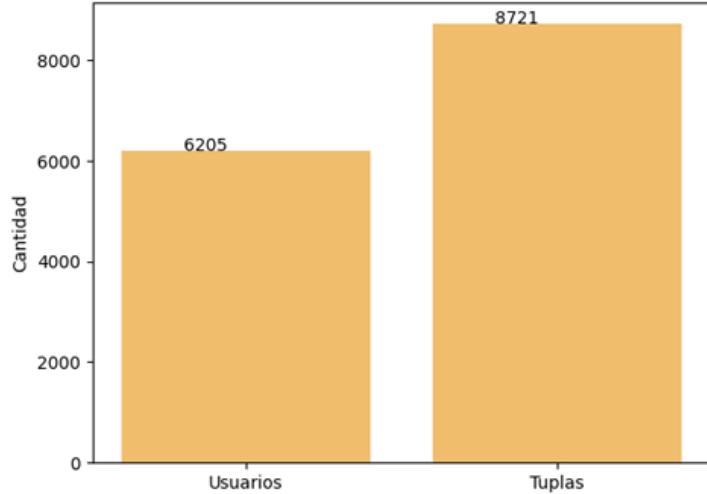
Análisis del dataset final



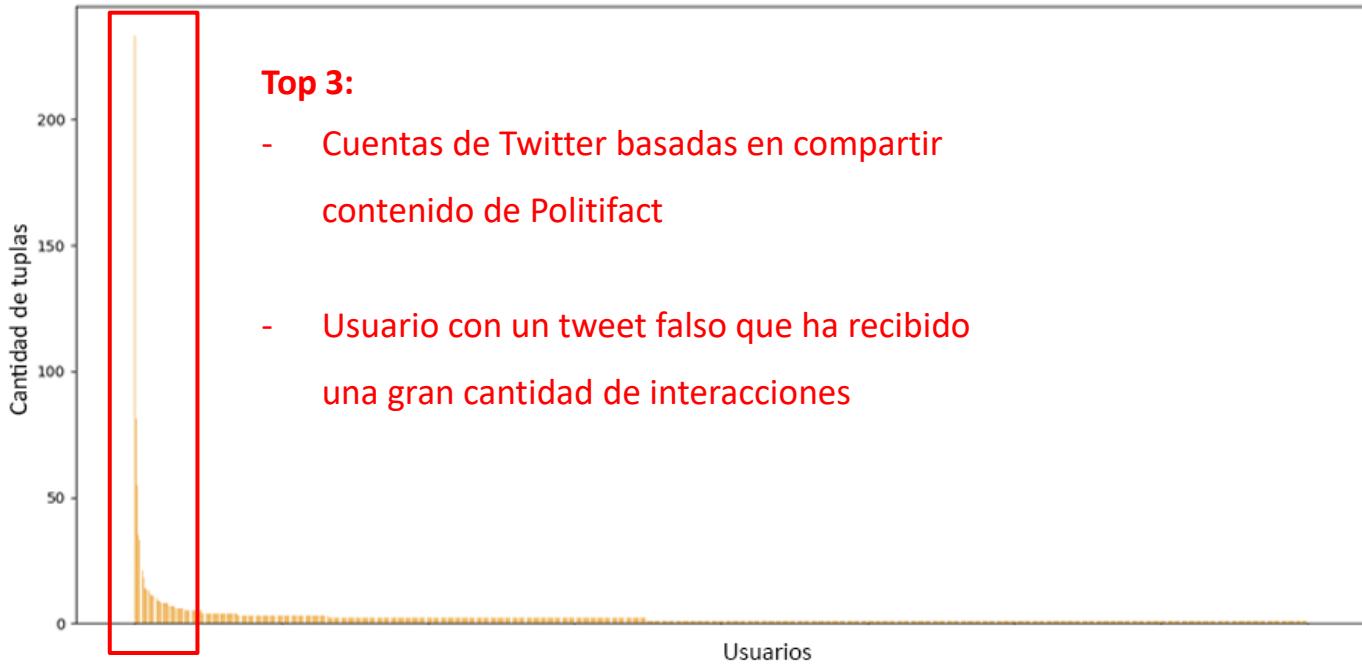
```
{  
  "user": "50188272",  
  "rate": -1,  
  "tweet_id_rated": "1337140529826435072",  
  "tweet_id": "1327007193661575168",  
  "tweet_in_reply": null,  
  "claim": null,  
  "claim_in_reply": "https://www.politifact.com/factchecks/2017/jan/10/trump/trumps-claims-about-muslims-are-false/",  
},  
{  
  "user": "15723081",  
  "rate": -1,  
  "tweet_id_rated": "1326910757393608704",  
  "tweet_id": "1326631254545018884",  
  "tweet_in_reply": null,  
  "claim": "https://www.politifact.com/factchecks/2017/jan/10/trump/trumps-claims-about-muslims-are-false/",  
  "claim_in_reply": null,  
},  
{  
  "user": "37943826",  
  "rate": 1,  
  "tweet_id_rated": "1326355666475347970",  
  "tweet_id": "1326355666475347970",  
  "tweet_in_reply": "1326631254545018884",  
  "claim": "https://www.politifact.com/factchecks/2017/jan/10/trump/trumps-claims-about-muslims-are-false/",  
  "claim_in_reply": null  
},
```

¿Cuántos elementos hay por cada usuarios en el dataset final?

Diferencia entre usuarios y tuplas en el conjunto de datos de puntuaciones



Distribución de tuplas por usuario en el conjunto de datos de puntuaciones



Media de 1.4 tuplas/usuarios

Mediana de 1 tuplas/usuarios

Esta distribución dificulta la recomendación, lo que se intenta solucionar con el suavizado de los datos

Experimentos

Descripción de los datos generados para recomendar

Dataset final

Proporción de desinformación:

- En función de las tuplas
 - En función de los usuarios

10%
30%
60%
90%

Proporción de neutralidad

0%
25%
50%

Datasets generados aleatoriamente para la recomendación

UserID	ItemID	Rate
1344356576786866176	512_b	0.521
93215780	546_g	5.279
12141545373		
28	1344356576786866176	512_b
4	93215780	546_g
14	12141545373	
2	1344356576786866176	512_b
9381972640	93215780	546_g
281		
41	1214154537314832391	564_b
148		
28452847	2814186879	443_b
7026935501	93819726401	801_g
4		
415196079	4180259144	626_g
1		
1480259144	938197264010444800	129_g
8883659039	115265558344	564_b
11512393210	115265558344	1.182
210	938197264010444800	44175599
1		632_g
45	115265558344	40_g
1	904101917	4.823
35	115265558344635648	874_b
2	702693550111092737	574_g
39		4.621
8883659039	17919393	6_b
115123932107		1.392
28	22555372	140_g
452		5.481
1	88836590393041754	56_g
16	115123932107147418192	208_b
351		0.697
9022687418	4520484798	208_g
12047485122	1614124392	72_b
283		1.387
34	3519064697	616_b
7	53230662	616_b
19		1.161
3	2835451658	504_b
2		1.39
13	342833326	854_b
90226874184	197371392	845_b
12047485122		0.66
45	20644655	854_b
2	902268741844692992	696_g
74		5.047
34	1204748512275120128	338_g
15		5.012
19	28448196	895_g
130		5.219
1	747153392	882_g
492		4.787
2	154697628	286_g
74	1307768174	845_b
15		5.225
130	18584875	845_b
1		1.236
492	4924987677	964_g
2		4.619

24 ficheros usados para recomendar y obtener resultados

Experimentos

Recomendadores y métricas

Recomendadores

No personalizados

- **Random**
- **MostPop**

FC basado en vecinos próximos

- **ItemKNN**
- **UserKNN**

FC basado en fact. Matrices

- **BPRMF**
- **FunkSVD**
- **SVD++**

FC basado en redes neuronales

- **MultiVAE**

Métricas

• **Recuento de desinformación (RD)**

• **Diferencia de ratio de desinformación (DRD)**

• **Contador de recomendaciones (CR)**

• **Contador de recomendaciones promedio (CRP)**

(Pensadas para ItemKNN y UserKNN)



Experimentos

Resultados

Ratio	Recm	RD@5	RD@10	RD@20	DRD	% CR	CRP@5	CRP@10	CRP@20
0.1	Random	1.44	2.10	3.67	-0.05	100%	5	10	20
0.1	MostPop	1.00	1.00	1.00	0.09	100%	5	10	20
0.1	ItemKNN	1.10	1.59	1.62	0.04	68.2%	3.92	6.77	10.55
0.1	UserKNN	1.07	1.47	1.52	0.05	54.1%	4.09	7.19	12.85
0.1	FunkSVD	1.50	2.73	5.73	-0.16	100%	5	10	20
0.1	SVDpp	1.39	2.18	4.12	-0.08	100%	5	10	20
0.1	BPRMF	1.00	1.06	1.78	0.03	100%	5	10	20
0.1	MultiVAE	1.02	1.15	1.68	0.05	100%	5	10	20
0.3	Random	1.78	3.04	5.97	0.05	100%	5	10	20
0.3	MostPop	1.92	3.9	7.86	-0.002	100%	5	10	20
0.3	ItemKNN	1.62	3.95	5.64	0.08	52.8%	3.95	6.78	11.31
0.3	UserKNN	1.58	4.28	6.35	0.07	39.3%	3.95	6.94	11.72
0.3	FunkSVD	1.80	3.28	7.59	-0.052	100%	5	10	20
0.3	SVDpp	2.22	4.14	8.30	-0.06	100%	5	10	20
0.3	BPRMF	1.43	3.82	7.84	0.03	100%	5	10	20
0.3	MultiVAE	2.10	3.67	6.58	0.03	100%	5	10	20
0.6	Random	2.43	4.61	9.16	0.18	100%	5	10	20
0.6	MostPop	4.05	6.77	15.69	-0.04	100%	5	10	20
0.6	ItemKNN	2.44	4.93	10.14	0.14	45.9%	3.81	6.04	9.66
0.6	UserKNN	2.02	4.36	9.12	0.19	30.4%	3.55	5.30	7.47
0.6	FunkSVD	3.40	7.05	12.57	0.03	100%	5	10	20
0.6	SVDpp	3.26	6.28	12.23	0.03	100%	5	10	20
0.6	BPRMF	3.84	7.32	14.49	-0.06	100%	5	10	20
0.6	MultiVAE	3.40	6.36	11.88	0.06	100%	5	10	20
0.9	Random	3.58	7.10	14.18	0.20	100%	5	10	20
0.9	MostPop	5.00	9.99	19.01	-0.05	100%	5	10	20
0.9	ItemKNN	4.54	8.33	16.14	0.10	56.1%	3.61	6.08	9.28
0.9	UserKNN	4.41	7.92	15.38	0.14	30.8%	3.35	5.10	5.53
0.9	FunkSVD	4.33	8.41	16.93	0.05	100%	5	10	20
0.9	SVDpp	3.87	7.52	14.58	0.19	100%	5	10	20
0.9	BPRMF	4.98	9.69	19.05	-0.04	100%	5	10	20
0.9	MultiVAE	4.90	9.73	19.16	-0.04	100%	5	10	20

Media de 1.4 tuplas/usuarios
Mediana de 1 tuplas/usuarios

Ratio	Recm	RD@5	RD@10	RD@20	DRD	% CR	CRP@5	CRP@10	CRP@20
0.1	Random	1.33	1.80	2.97	-0.04	100%	5	10	20
0.1	MostPop	1.00	1.00	1.00	0.06	100%	5	10	20
0.1	ItemKNN	2.34	3.84	4.73	-0.01	76.7%	4.32	8.00	14.45
0.1	UserKNN	2.73	3.59	4.90	-0.02	60.8%	4.23	7.965	14.81
0.1	FunkSVD	1.01	1.32	2.97	-0.05	100%	5	10	20
0.1	SVDpp	1.47	2.15	3.95	-0.09	100%	5	10	20
0.1	BPRMF	1.00	1.03	1.44	0.03	100%	5	10	20
0.1	MultiVAE	1.03	1.07	1.39	0.04	100%	5	10	20
0.3	Random	1.65	2.64	5.01	0.05	100%	5	10	20
0.3	MostPop	1.00	2.93	5.88	0.06	100%	5	10	20
0.3	ItemKNN	3.27	3.484	5.84	0.08	71.1%	4.3	7.83	13.58
0.3	UserKNN	2.93	3.45	6.15	0.07	51.4%	3.82	6.43	10.10
0.3	FunkSVD	1.48	3.04	6.69	-0.06	100%	5	10	20
0.3	SVDpp	1.49	2.54	4.76	0.06	100%	5	10	20
0.3	BPRMF	1.14	2.60	5.72	0.01	100%	5	10	20
0.3	MultiVAE	1.36	2.33	4.94	0.05	100%	5	10	20
0.6	Random	1.89	3.33	6.56	0.26	100%	5	10	20
0.6	MostPop	3.83	6.78	11.73	-0.03	100%	5	10	20
0.6	ItemKNN	2.59	3.78	8.24	0.21	61.7%	3.40	5.53	8.14
0.6	UserKNN	2.40	4.81	7.52	0.27	44.9%	3.17	4.96	7.76
0.6	FunkSVD	3.08	6.27	11.80	0.03	100%	5	10	20
0.6	SVDpp	2.48	4.83	9.39	0.13	100%	5	10	20
0.6	BPRMF	3.62	7.03	13.30	-0.05	100%	5	10	20
0.6	MultiVAE	3.22	6.08	11.31	0.04	100%	5	10	20
0.9	Random	2.62	5.10	10.26	0.38	100%	5	10	20
0.9	MostPop	4.99	9.02	19.02	-0.03	100%	5	10	20
0.9	ItemKNN	3.52	6.59	13.53	0.23	67.8%	3.70	6.22	9.29
0.9	UserKNN	3.51	6.08	12.45	0.30	39.66%	3.32	5.38	7.83
0.9	FunkSVD	4.35	8.76	17.56	0.02	100%	5	10	20
0.9	SVDpp	3.36	6.57	12.38	0.28	100%	5	10	20
0.9	BPRMF	4.87	9.07	18.64	-0.03	100%	5	10	20
0.9	MultiVAE	4.67	9.28	18.55	-0.03	100%	5	10	20

Proporción de desinformación calculada **según todas las tuplas** y proporción de neutralidad del 0%

Proporción de desinformación calculada **según los usuarios** y proporción de neutralidad del 0%

Experimentos

Resultados

Proporción de desinformación calculada **según todas las tuplas** y proporción de neutralidad del 0%

Ratio	Recm	RD@5	RD@10	RD@20	DRD	% CR	CRP@5	CRP@10	CRP@20
0.1	Random	1.44	2.10	3.67	-0.05	100 %	5	10	20
0.1	MostPop	1.00	1.00	1.00	0.09	100 %	5	10	20
0.1	ItemKNN	1.10	1.59	1.62	0.04	68.2 %	3.92	6.77	10.55
0.1	UserKNN	1.07	1.47	1.52	0.05	54.1 %	4.09	7.19	12.85
0.1	FunkSVD	1.50	2.73	5.73	-0.16	100 %	5	10	20
0.1	SVDpp	1.39	2.18	4.12	-0.08	100 %	5	10	20
0.1	BPRMF	1.00	1.06	1.78	0.03	100 %	5	10	20
0.1	MultiVAE	1.02	1.15	1.68	-0.05	100 %	5	10	20

x3

Ratio	Recm	RD@5	RD@10	RD@20	DRD	% CR	CRP@5	CRP@10	CRP@20
0.1	Random	1.33	1.80	2.97	-0.04	100 %	5	10	20
0.1	MostPop	1.00	1.00	1.00	0.06	100 %	5	10	20
0.1	ItemKNN	2.34	3.84	4.73	-0.01	76.7 %	4.32	8.00	14.45
0.1	UserKNN	2.73	3.59	4.90	-0.02	60.8 %	4.23	7.965	14.81
0.1	FunkSVD	1.01	1.32	2.97	-0.05	100 %	5	10	20
0.1	SVDpp	1.47	2.15	3.95	-0.09	100 %	5	10	20
0.1	BPRMF	1.00	1.03	1.44	0.03	100 %	5	10	20
0.1	MultiVAE	1.03	1.07	1.39	0.04	100 %	5	10	20

Proporción de desinformación calculada **según los usuarios** y proporción de neutralidad del 0%

Experimentos

Resultados

Proporción de desinformación calculada **según todas las tuplas** y proporción de neutralidad del 0%

0.9	Random	3.58	7.10	14.18	0.20	100 %	5	10	20
0.9	MostPop	5.00	9.99	19.01	-0.05	100 %	5	10	20
0.9	ItemKNN	4.54	8.33	16.14	0.10	56.1 %	3.61	6.08	9.28
0.9	UserKNN	4.41	7.92	15.38	0.14	30.8 %	3.35	5.10	5.53
0.9	FunkSVD	4.33	8.41	16.93	0.05	100 %	5	10	20
0.9	SVDpp	3.87	7.52	14.58	0.19	100 %	5	10	20
0.9	BPRMF	4.98	9.69	19.05	-0.04	100 %	5	10	20
0.9	MultiVAE	4.90	9.73	19.16	-0.04	100 %	5	10	20

0.9	Random	2.62	5.10	10.26	0.38	100 %	5	10	20
0.9	MostPop	4.99	9.02	19.02	-0.03	100 %	5	10	20
0.9	ItemKNN	3.52	6.59	13.53	0.23	67.8 %	3.70	6.22	9.29
0.9	UserKNN	3.51	6.08	12.45	0.30	39.66 %	3.32	5.38	7.83
0.9	FunkSVD	4.35	8.76	17.56	0.02	100 %	5	10	20
0.9	SVDpp	3.36	6.57	12.38	0.28	100 %	5	10	20
0.9	BPRMF	4.87	9.07	18.64	-0.03	100 %	5	10	20
0.9	MultiVAE	4.67	9.28	18.55	-0.03	100 %	5	10	20

Proporción de desinformación calculada **según los usuarios** y proporción de neutralidad del 0%

Resultados

Proporción de desinformación calculada **según todas las tuplas** y proporción de neutralidad del 0%

0.9	Random	3.58	7.10	14.18	0.20	100 %	5	10	20
0.9	MostPop	5.00	9.99	19.01	-0.05	100 %	5	10	20
0.9	ItemKNN	4.54	8.33	16.14	0.10	56.1 %	3.61	6.08	9.28
0.9	UserKNN	4.41	7.92	15.38	0.14	30.8 %	3.35	5.10	5.53
0.9	FunkSVD	4.33	8.41	16.93	0.05	100 %	5	10	20
0.9	SVDpp	3.87	7.52	14.58	0.19	100 %	5	10	20
0.9	BPRMF	4.98	9.69	19.05	-0.04	100 %	5	10	20
0.9	MultiVAE	4.90	9.73	19.16	-0.04	100 %	5	10	20

Algoritmos
no personalizados

0.9	Random	2.62	5.10	10.26	0.38	100 %	5	10	20
0.9	MostPop	4.99	9.02	19.02	-0.03	100 %	5	10	20
0.9	ItemKNN	3.52	6.59	13.53	0.23	67.8 %	3.70	6.22	9.29
0.9	UserKNN	3.51	6.08	12.45	0.30	39.66 %	3.32	5.38	7.83
0.9	FunkSVD	4.35	8.76	17.56	0.02	100 %	5	10	20
0.9	SVDpp	3.36	6.57	12.38	0.28	100 %	5	10	20
0.9	BPRMF	4.87	9.07	18.64	-0.03	100 %	5	10	20
0.9	MultiVAE	4.67	9.28	18.55	-0.03	100 %	5	10	20

Proporción de desinformación calculada **según los usuarios** y proporción de neutralidad del 0%

Resultados

Proporción de desinformación calculada **según todas las tuplas** y proporción de neutralidad del 0%

Ratio	Recm	RD@5	RD@10	RD@20	DRD	% CR	CRP@5	CRP@10	CRP@20
0.9	Random	3.58	7.10	14.18	0.20	100 %	5	10	20
0.9	MostPop	5.00	9.99	19.01	-0.05	100 %	5	10	20
0.9	ItemKNN	4.54	8.33	16.14	0.10	56.1 %	3.61	6.08	9.28
0.9	UserKNN	4.41	7.92	15.38	0.14	30.8 %	3.35	5.10	5.53
0.9	FunkSVD	4.33	8.41	16.93	0.05	100 %	5	10	20
0.9	SVDpp	3.87	7.52	14.58	0.19	100 %	5	10	20
0.9	BPRMF	4.98	9.69	19.05	-0.04	100 %	5	10	20
0.9	MultiVAE	4.90	9.73	19.16	-0.04	100 %	5	10	20

Algoritmos FC
basados en vecinos
próximos

0.9	Random	2.62	5.10	10.26	0.38	100 %	5	10	20
0.9	MostPop	4.99	9.02	19.02	-0.03	100 %	5	10	20
0.9	ItemKNN	3.52	6.59	13.53	0.23	67.8 %	3.70	6.22	9.29
0.9	UserKNN	3.51	6.08	12.45	0.30	39.66 %	3.32	5.38	7.83
0.9	FunkSVD	4.35	8.76	17.56	0.02	100 %	5	10	20
0.9	SVDpp	3.36	6.57	12.38	0.28	100 %	5	10	20
0.9	BPRMF	4.87	9.07	18.64	-0.03	100 %	5	10	20
0.9	MultiVAE	4.67	9.28	18.55	-0.03	100 %	5	10	20

Proporción de desinformación calculada **según los usuarios** y proporción de neutralidad del 0%

Resultados

Ratios bajos (<0.5)

Ratio	Recm	RD@5	RD@10	RD@20	DRD	% CR	CRP@5	CRP@10	CRP@20
0.1	Random	1.44	2.10	3.67	-0.05	100 %	5	10	20
0.1	MostPop	1.00	1.00	1.00	0.09	100 %	5	10	20
0.1	ItemKNN	1.10	1.59	1.62	0.04	68.2 %	3.92	6.77	10.55
0.1	UserKNN	1.07	1.47	1.52	0.05	54.1 %	4.09	7.19	12.85
0.1	FunkSVD	1.50	2.73	5.73	-0.16	100 %	5	10	20
0.1	SVDpp	1.39	2.18	4.12	-0.08	100 %	5	10	20
0.1	BPRMF	1.00	1.06	1.78	0.03	100 %	5	10	20
0.1	MultiVAE	1.02	1.15	1.68	0.05	100 %	5	10	20

Algoritmos FC
basados en factorización
de matrices

Ratios altos (>0.5)

0.9	Random	3.58	7.10	14.18	0.20	100 %	5	10	20
0.9	MostPop	5.00	9.99	19.01	-0.05	100 %	5	10	20
0.9	ItemKNN	4.54	8.33	16.14	0.10	56.1 %	3.61	6.08	9.28
0.9	UserKNN	4.41	7.92	15.38	0.14	30.8 %	3.35	5.10	5.53
0.9	FunkSVD	4.33	8.41	16.93	0.05	100 %	5	10	20
0.9	SVDpp	3.87	7.52	14.58	0.19	100 %	5	10	20
0.9	BPRMF	4.98	9.69	19.05	-0.04	100 %	5	10	20
0.9	MultiVAE	4.90	9.73	19.16	-0.04	100 %	5	10	20

(A partir de ahora solo mostraremos la proporción de desinformación calculada **según todas las tuplas**)

Resultados

Ratios bajos (<0.5)

Ratio	Recm	RD@5	RD@10	RD@20	DRD	% CR	CRP@5	CRP@10	CRP@20
0.1	Random	1.44	2.10	3.67	-0.05	100 %	5	10	20
0.1	MostPop	1.00	1.00	1.00	0.09	100 %	5	10	20
0.1	ItemKNN	1.10	1.59	1.62	0.04	68.2 %	3.92	6.77	10.55
0.1	UserKNN	1.07	1.47	1.52	0.05	54.1 %	4.09	7.19	12.85
0.1	FunkSVD	1.50	2.73	5.73	-0.16	100 %	5	10	20
0.1	SVDpp	1.39	2.18	4.12	-0.08	100 %	5	10	20
0.1	BPRMF	1.00	1.06	1.78	0.03	100 %	5	10	20
0.1	MultiVAE	1.02	1.15	1.68	0.05	100 %	5	10	20

Algoritmos FC
basados en redes
neuronales

Ratios altos (>0.5)

0.9	Random	3.58	7.10	14.18	0.20	100 %	5	10	20
0.9	MostPop	5.00	9.99	19.01	-0.05	100 %	5	10	20
0.9	ItemKNN	4.54	8.33	16.14	0.10	56.1 %	3.61	6.08	9.28
0.9	UserKNN	4.41	7.92	15.38	0.14	30.8 %	3.35	5.10	5.53
0.9	FunkSVD	4.33	8.41	16.93	0.05	100 %	5	10	20
0.9	SVDpp	3.87	7.52	14.58	0.19	100 %	5	10	20
0.9	BPRMF	4.98	9.69	19.05	-0.04	100 %	5	10	20
0.9	MultiVAE	4.90	9.73	19.16	-0.04	100 %	5	10	20

(A partir de ahora solo mostraremos la proporción de
desinformación calculada **según todas las tuplas**)

Resultados

Impacto de la neutralidad en la recomendación

Ratio	Recm	0	0.25	0.5
0.3	Random	3.04	1.00	1.00
0.3	MostPop	3.90	3.05	3.05
0.3	ItemKNN	3.95	ND	ND
0.3	UserKNN	4.28	2.00	0.00
0.3	FunkSVD	3.28	4.92	6.74
0.3	SVDpp	4.14	1.00	ND
0.3	BPRMF	3.82	6.07	8.95
0.3	MultiVAE	3.67	1.24	2.21
0.9	Random	7.10	1.00	1.00
0.9	MostPop	9.99	9.99	9.00
0.9	ItemKNN	8.33	ND	ND
0.9	UserKNN	7.92	2.19	4.18
0.9	FunkSVD	8.41	8.83	9.37
0.9	SVDpp	7.52	1.00	ND
0.9	BPRMF	9.69	10.0	10.0
0.9	MultiVAE	9.73	2.20	1.83

(A partir de ahora solo mostraremos la proporción de desinformación calculada **según todas las tuplas**)

Resultados

Impacto de la neutralidad en la recomendación

Métrica RD@10

(Neutralidad)

Ratio	Recm	0	0.25	0.5
0.3	Random	3.04	1.00	1.00
0.3	MostPop	3.90	3.05	3.05
0.3	ItemKNN	3.95	ND	ND
0.3	UserKNN	4.28	2.00	0.00
0.3	FunkSVD	3.28	4.92	6.74
0.3	SVDpp	4.14	1.00	ND
0.3	BPRMF	3.82	6.07	8.95
0.3	MultiVAE	3.67	1.24	2.21
0.9	Random	7.10	1.00	1.00
0.9	MostPop	9.99	9.99	9.00
0.9	ItemKNN	8.33	ND	ND
0.9	UserKNN	7.92	2.19	4.18
0.9	FunkSVD	8.41	8.83	9.37
0.9	SVDpp	7.52	1.00	ND
0.9	BPRMF	9.69	10.0	10.0
0.9	MultiVAE	9.73	2.20	1.83

(A partir de ahora solo mostraremos la proporción de desinformación calculada **según todas las tuplas**)

Resultados

Impacto de la neutralidad en la recomendación

Métrica RD@10

(Neutralidad)

Ratio	Recm	0	0.25	0.5
0.3	Random	3.04	1.00	1.00
0.3	MostPop	3.90	3.05	3.05
0.3	ItemKNN	3.95	ND	ND
0.3	UserKNN	4.28	2.00	0.00
0.3	FunkSVD	3.28	4.92	6.74
0.3	SVDpp	4.14	1.00	ND
0.3	BPRMF	3.82	6.07	8.95
0.3	MultiVAE	3.67	1.24	2.21
0.9	Random	7.10	1.00	1.00
0.9	MostPop	9.99	9.99	9.00
0.9	ItemKNN	8.33	ND	ND
0.9	UserKNN	7.92	2.19	4.18
0.9	FunkSVD	8.41	8.83	9.37
0.9	SVDpp	7.52	1.00	ND
0.9	BPRMF	9.69	10.0	10.0
0.9	MultiVAE	9.73	2.20	1.83

(A partir de ahora solo mostraremos la proporción de desinformación calculada **según todas las tuplas**)

Resultados

Impacto de la neutralidad en la recomendación

Métrica RD@10

Ratio	Recm	(Neutralidad)		
		0	0.25	0.5
0.3	Random	3.04	1.00	1.00
0.3	MostPop	3.90	3.05	3.05
0.3	ItemKNN	3.95	ND	ND
0.3	UserKNN	4.28	2.00	0.00
0.3	FunkSVD	3.28	4.92	6.74
0.3	SVDpp	4.14	1.00	ND
0.3	BPRMF	3.82	6.07	8.95
0.3	MultiVAE	3.67	1.24	2.21
0.9	Random	7.10	1.00	1.00
0.9	MostPop	9.99	9.99	9.00
0.9	ItemKNN	8.33	ND	ND
0.9	UserKNN	7.92	2.19	4.18
0.9	FunkSVD	8.41	8.83	9.37
0.9	SVDpp	7.52	1.00	ND
0.9	BPRMF	9.69	10.0	10.0
0.9	MultiVAE	9.73	2.20	1.83

¿Son los algoritmos basados en factorización de matrices los peores?

x2!!

Resultados

Impacto de la neutralidad en la recomendación

Métrica RD@10

Ratio	Recm	(Neutralidad)		
		0	0.25	0.5
0.3	Random	3.04	1.00	1.00
0.3	MostPop	3.90	3.05	3.05
0.3	ItemKNN	3.95	ND	ND
0.3	UserKNN	4.28	2.00	0.00
0.3	FunkSVD	3.28	4.92	6.74
0.3	SVDpp	4.14	1.00	ND
0.3	BPRMF	3.82	6.07	8.95
0.3	MultiVAE	3.67	1.24	2.21
0.9	Random	7.10	1.00	1.00
0.9	MostPop	9.99	9.99	9.00
0.9	ItemKNN	8.33	ND	ND
0.9	UserKNN	7.92	2.19	4.18
0.9	FunkSVD	8.41	8.83	9.37
0.9	SVDpp	7.52	1.00	ND
0.9	BPRMF	9.69	10.0	10.0
0.9	MultiVAE	9.73	2.20	1.83

¿Son los algoritmos basados en factorización de matrices los peores?

Resultados

Impacto de la neutralidad en la recomendación

Top peores algoritmos (de mayor a menor desinformación)

- MostPopular y BPMRF
- FunkSVD
- ItemKNN, UserKNN, MultiVAE
- **Random* y SVD++****

Ratio	Recm	(Neutralidad)		
		0	0.25	0.5
0.3	Random	3.04	1.00	1.00
0.3	MostPop	3.90	3.05	3.05
0.3	ItemKNN	3.95	ND	ND
0.3	UserKNN	4.28	2.00	0.00
0.3	FunkSVD	3.28	4.92	6.74
0.3	SVDpp	4.14	1.00	ND
0.3	BPRMF	3.82	6.07	8.95
0.3	MultiVAE	3.67	1.24	2.21
<hr/>				
0.9	Random	7.10	1.00	1.00
0.9	MostPop	9.99	9.99	9.00
0.9	ItemKNN	8.33	ND	ND
0.9	UserKNN	7.92	2.19	4.18
0.9	FunkSVD	8.41	8.83	9.37
0.9	SVDpp	7.52	1.00	ND
0.9	BPRMF	9.69	10.0	10.0
0.9	MultiVAE	9.73	2.20	1.83

* Buen método pero tiene el inconveniente de no ofrecer recomendaciones de calidad

** SVD++ nos falta conocer la neutralidad del 50%

Con mucha neutralidad, UserKNN y MultiVAE son las mejores opciones.

CONCLUSIONES Y TRABAJO FUTURO

Conclusiones

- Los datos extraídos son sencillos y naturales.
- Se ha demostrado que los tweets desinformativos son más polémicos y generan muchas interacciones.
- ItemKNN y UserKNN son los que más destacan con ratios altos pero con inconvenientes.
- No existe un algoritmo de recomendación superior a otro en cualquier situación

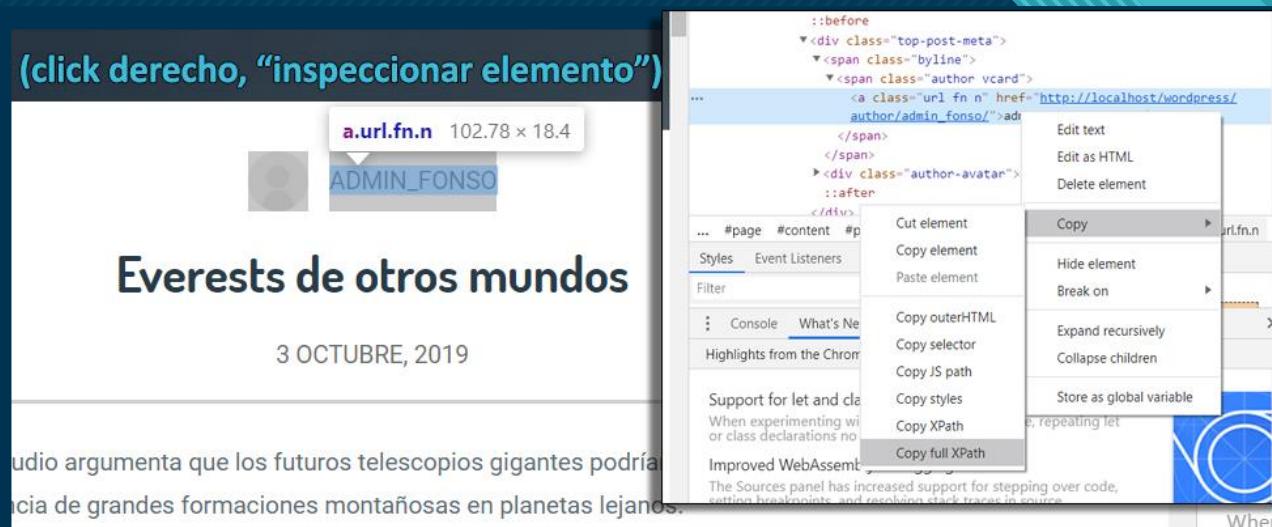
CONCLUSIONES Y TRABAJO FUTURO

Trabajo Futuro

- Reducción de las limitaciones a la hora de extraer los datos
- Añadir más variedad aumentando los portales verificadores
- Mejorar las recomendaciones (mejorando a nivel computacional y añadiendo mas variedad)
- Aplicar técnicas de procesamiento natural y análisis de redes sociales a los elementos desinformativos

Gracias

Explicación de XPATH



```
{  
  "allowed_domain_to_crawl" : "localhost",  
  "start_url_to_crawl" : "http://localhost/wordpress/",  
  
  "rest_api_host" : "http://127.0.0.1:8000/",  
  "rest_api_create_post_method" : "/api/posts/",  
  "rest_api_remove_post_method" : "/api/posts/remove/",  
  "rest_api_create_index" : "/api/index/create",  
  
  "author_full_xpath": "/html/body/div[1]/div[2]/div/main/article/header/div[1]/span/span/a",  
  "category_full_path": "/html/body/div[1]/div[2]/div/main/article/footer/span[1]/a",  
  "time_full_xpath": "/html/body/div[1]/div[2]/div/main/article/header/div[2]/span/a/time[1]"  
}
```

Archivo de configuración del crawler, **conf.json**

¿Cuáles son los claims más populares en Twitter?

Análisis del dataset final

Top 10	Claim	Puntuación	Tema
Claim_1	How President Barack Obama handled the outbreak of H1N1	False	Trump y COVID-19
Claim_2	Mask box label is legitimate, but people are misinterpreting it	False	COVID-19
Claim_3	The Obama administration "didn't do anything about" swine flu.	False	Trump y COVID-19
Claim_4	Says Nancy Pelosi was "caught trying to..."	False	COVID-19
Claim_5	"Wisconsin has more votes than people who..."	Pants in fire	Fraude electoral
Claim_6	"Nancy Pelosi may have just committed a violation of 18 U.S.C. § 2071..."	Pants in fire	Política
Claim_7	"We inherited a broken test" for COVID-19.	Pants in fire	Trump y COVID-19
Claim_8	Early morning election results from Michigan and Wisconsin show voter fraud.	False	Fraude electoral
Claim_9	Says photo of beaten woman is Aracely Henriquez...	False	George Floyd
Claim_10	Voters in Maricopa County, Ariz., were forced to...	False	Fraude electoral