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

INTRODUCCIÓN

Motivación

DESINFORMACIÓN (MISINFORMATION)

RECOMMENDATION SYSTEM

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

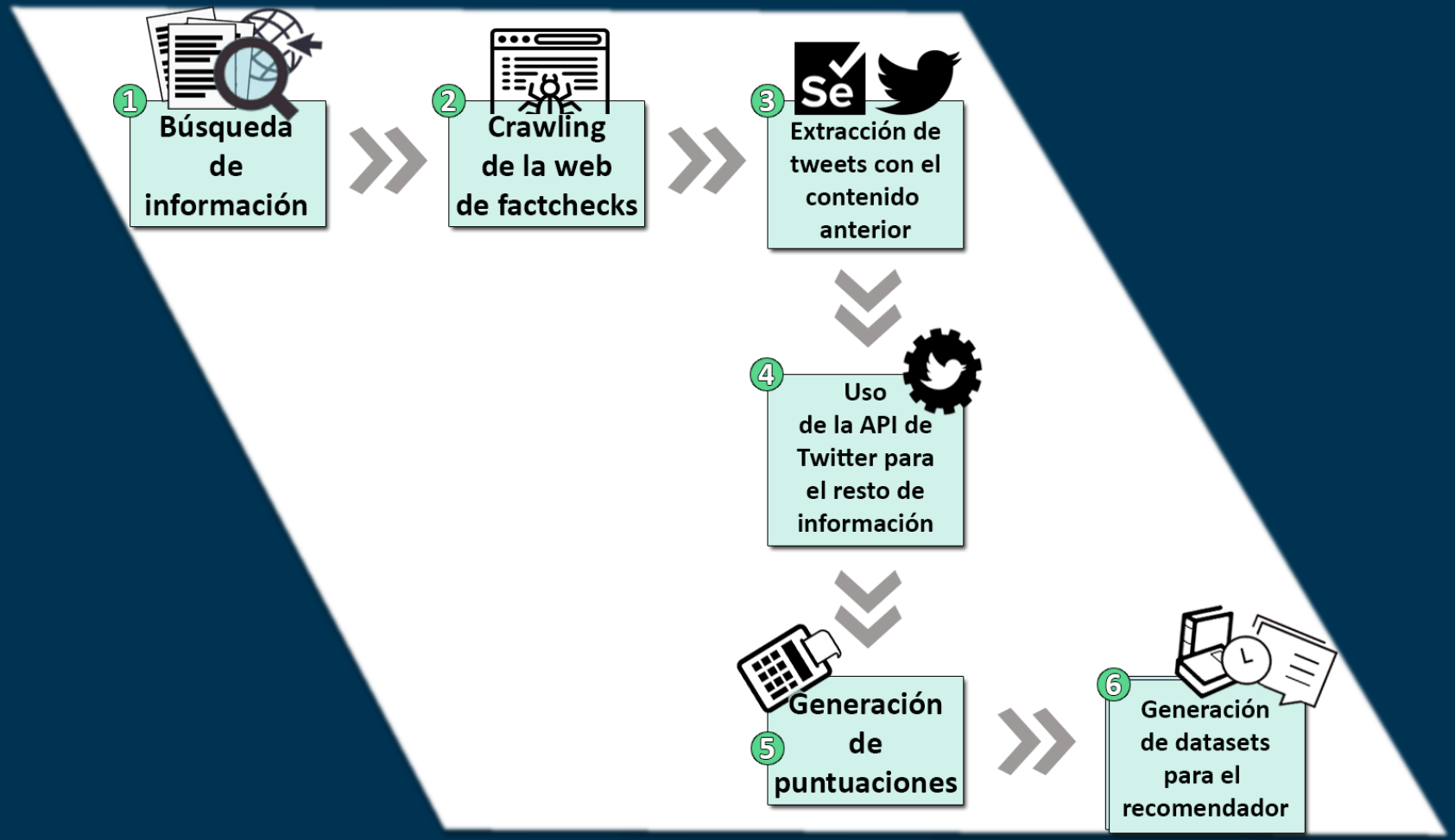
INTRODUCCIÓN

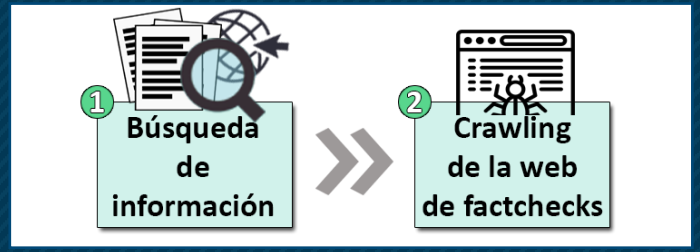
Objetivos

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**.

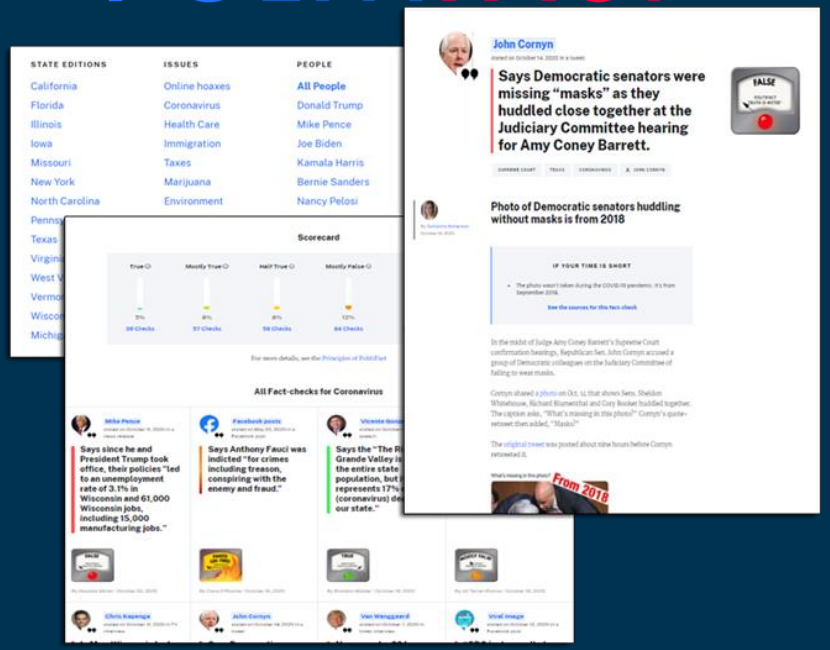
Sistema desarrollado

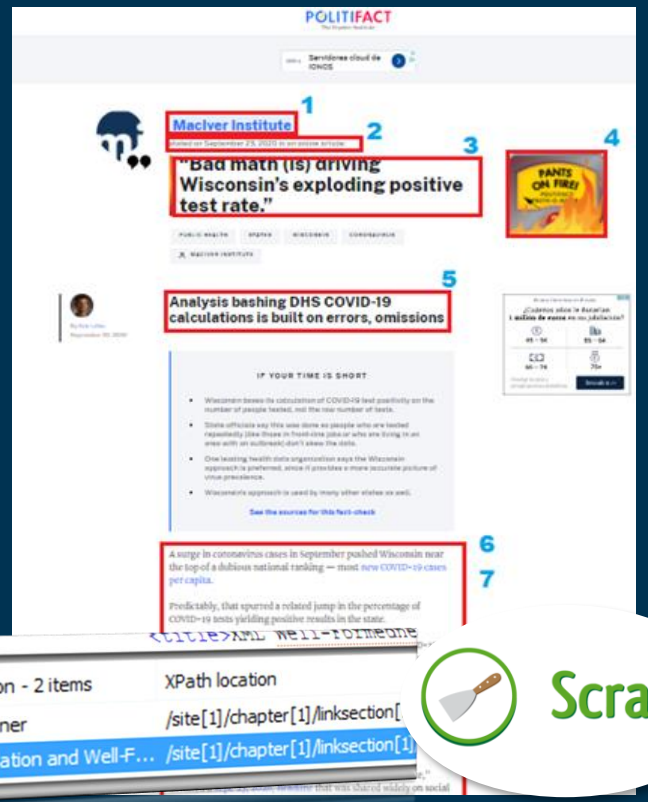




Sistema desarrollado

POLITIFACT





Térmometro de Politifact

Verdaderas

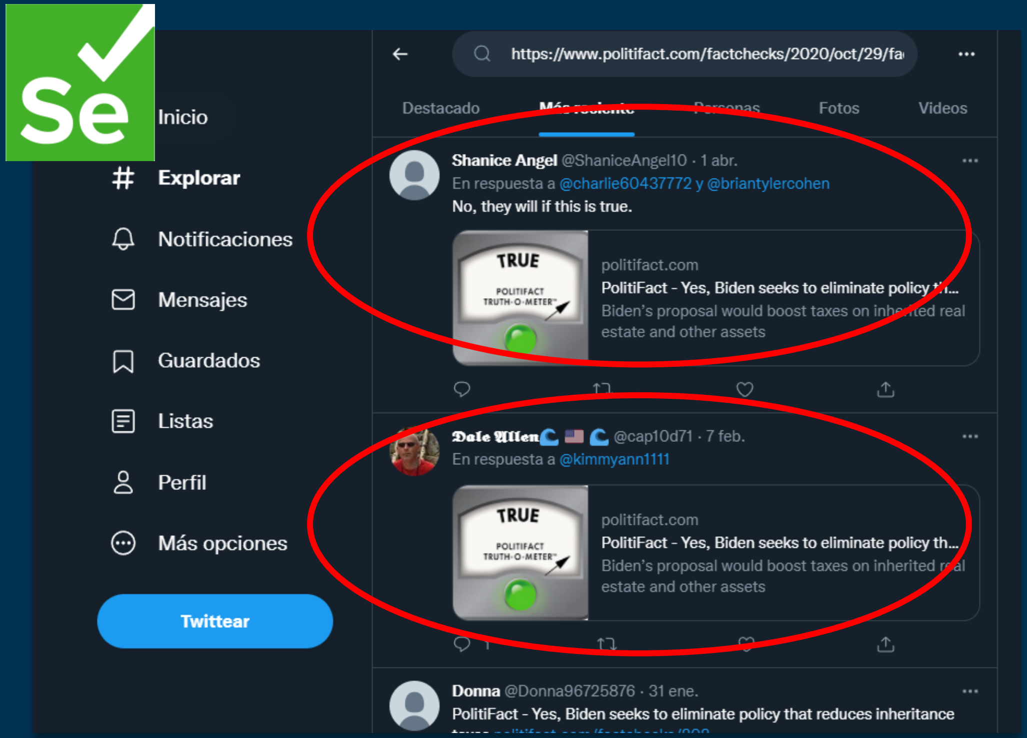
Falsas

Sistema desarrollado

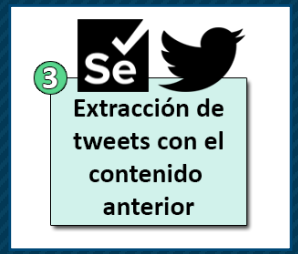
3 Extracción de tweets con el contenido anterior

Cada URL se busca en Twitter de forma automatizada usando Selenium

https://www.politifact.com/factchecks/2020/oct/29/facebook-po
https://www.politifact.com/factchecks/2020/oct/29/facebook-po
https://www.politifact.com/factchecks/2020/oct/29/facebook-po
https://www.politifact.com/factchecks/2020/oct/29/facebook-po
https://www.politifact.com/factchecks/2020/oct/29/facebook-po
https://www.politifact.com/factchecks/2020/oct/29/facebook-po
https://www.politifact.com/factchecks/2020/oct/29/facebook-po
https://www.politifact.com/factchecks/2020/oct/29/facebook-po
https://www.politifact.com/factchecks/2020/oct/29/facebook-po
https://www.politifact.com/factchecks/2020/oct/29/facebook-po
https://www.politifact.com/factchecks/2020/oct/29/facebook-po
https://www.politifact.com/factchecks/2020/oct/29/facebook-po
https://www.politifact.com/factchecks/2020/oct/29/facebook-po
https://www.politifact.com/factchecks/2020/oct/29/facebook-po



Duración alrededor de 4h para buscar unas 850 publicaciones



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

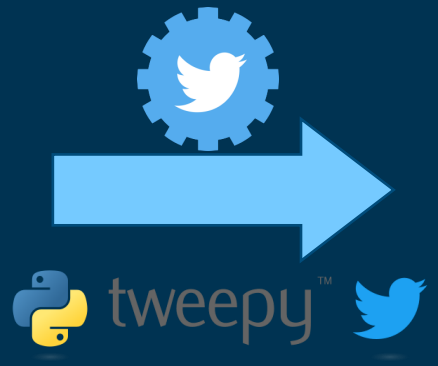
Sistema desarrollado

4 Uso de la API de Twitter para el resto de información

Cada tweet extraído con Selenium se busca en la API

https://twitter.com/ShaniceAngel10/status/1377406601619894
https://twitter.com/cap10d71/status/1358274594365448193
https://twitter.com/Donna96725876/status/13559832637104087
https://twitter.com/ShaniceAngel10/status/1377406601619894
https://twitter.com/BigLasagnaMI/status/1355889904635543557
https://twitter.com/ShaniceAngel10/status/1377406601619894
https://twitter.com/Donna96725876/status/13559832637104087
https://twitter.com/BigLasagnaMI/status/1355889904635543557
https://twitter.com/ShaniceAngel10/status/1377406601619894
https://twitter.com/Donna96725876/status/13559832637104087
https://twitter.com/ShaniceAngel10/status/1377406601619894
https://twitter.com/ShaniceAngel10/status/1377406601619894
https://twitter.com/ShaniceAngel10/status/1377406601619894
https://twitter.com/ShaniceAngel10/status/1377406601619894

TweetIDs




```
{
  "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"
```

4 Uso de la API de Twitter para el resto de información

Sistema desarrollado

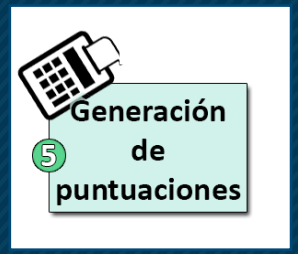
Para cada usuario se extraen 150 tweets de su timeline

@Ejemplo

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

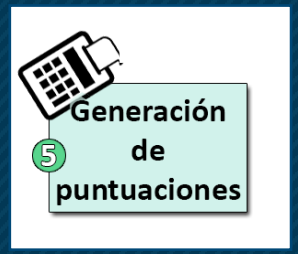


Sistema desarrollado

Dataset final de puntuaciones

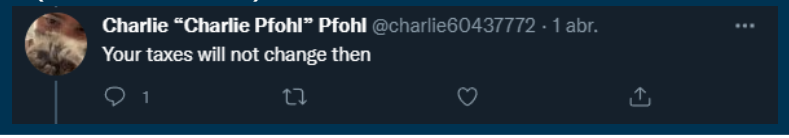

```
[
  {
    "user": "85273142",
    "rate": 1,
    "tweet_id_rated": "1377406601619894272",
    "tweet_id": "1377406601619894272",
    "tweet_in_reply": null,
    "claim": "https://www.politifact.com/factchecks/2020/
    "claim_in_reply": null
  },
  {
    "user": "2655912266",
    "rate": -1,
   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

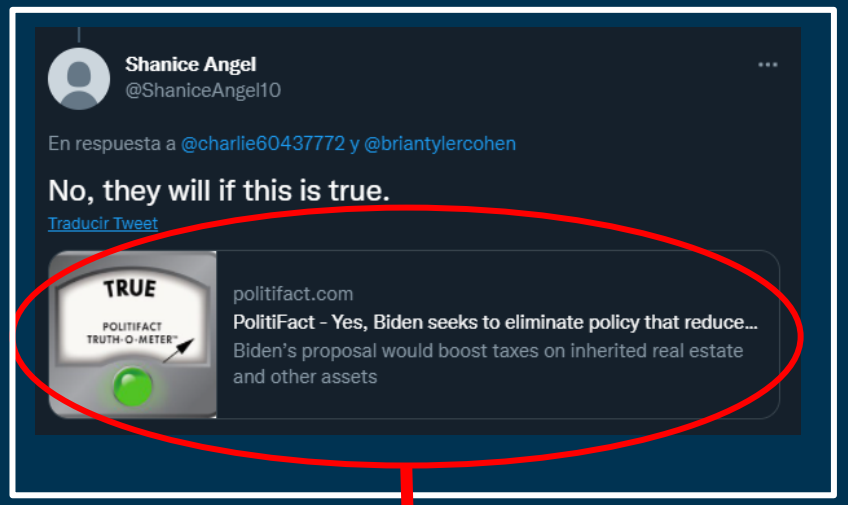

- User
- Rate (0,1,-1)
- tweed_id_rated
- tweet_id
- tweet_in_replay
- claim
- claim_in_replay



Sistema desarrollado

(Tweet Padre)

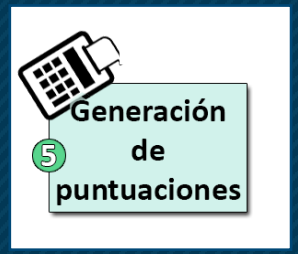




Claim

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
},
```



Sistema desarrollado

Tweet puntuado (el del padre)

(Claim del hijo)

JSON generado para ese tweet

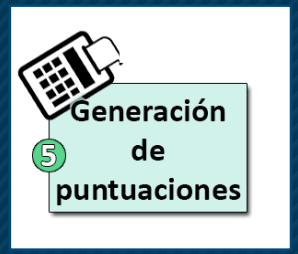
```
{
  "user": "2655912266",
  "rate": -1,
  "tweet_id_rated": "1377406601619894272", (Tweet ID del hijo)
  "tweet_id": "1377399121565282306",(Tweet ID del padre)
  "tweet_in_reply": "1377406601619894272", (Tweet ID del hijo)
  "claim": null,
  "claim_in_reply": "https://www.politifact.com/factchecks/2020/oct/29/facebook-posts/yes-biden-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

(Tweet ID del hijo)

1358274594365448193

(Tweet ID del padre)

1377399121565282306

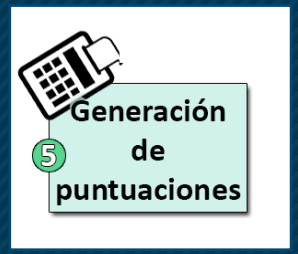


Sistema desarrollado

Tweet neutral del timeline del usuario

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

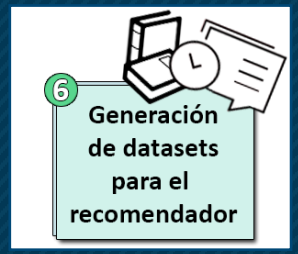
Shanice Angel
@ShaniceAngel10

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,
```


Formato JSON

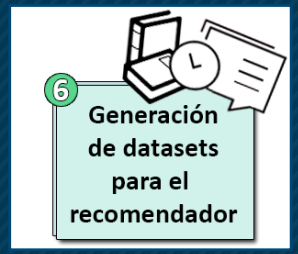
Datasets generados aleatoriamente para la recomendación

UserID	ItemID	Rate
1344356576786866176	512__b	0.521
93215780	546__g	5.279

UserID	ItemID	Rate
1344356576786866176	512__b	0.521
93215780	546__g	5.279

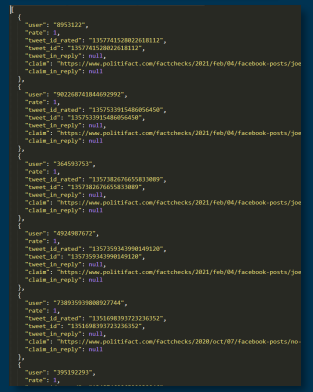
UserID	ItemID	Rate
1344356576786866176	512__b	0.521
93215780	546__g	5.279
1214154537314832391	564__b	1.196
23452847	443__b	1.051
2814186879	801__g	5.055
415196079	626__g	5.432
1480259144	129__g	4.823
938197264010444800	564__b	1.182
44175599	632__g	4.667
14091197	40__g	4.823
1152655583444635648	874__b	0.966
702693550111092737	574__g	4.621
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
197371392	845__b	0.66
20644656	854__g	5.421
902268741844692992	696__g	5.047
1204748512275120128	338__g	5.012
28448196	895__g	5.219
747153392	882__g	4.787
154697628	286__g	5.225
1307768174	845__b	1.236
18584875	845__b	1.223
4924987672	969__g	4.619

Formato TSV



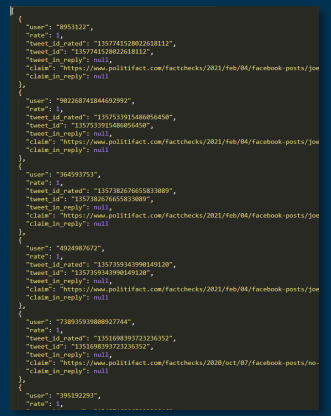
Sistema desarrollado

Dataset final



(proporción de desinformación)

Dataset final



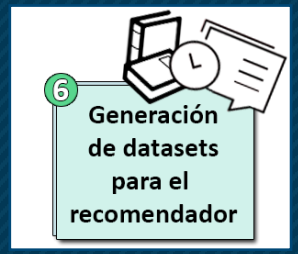
(proporción de neutralidad)

(con los tweets neutrales obtenido de los timelines)

Dataset generado aleatoriamente para la recomendación

UserID	ItemID	Rate
13443565767868661 76	512__b	0.521
93215780	546__g	5.279
12141545373148323 91	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
11512393210741841 92	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)



Sistema desarrollado

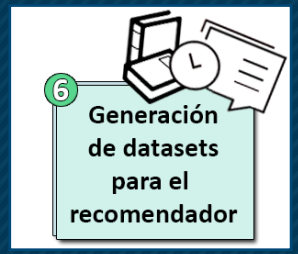
Dataset final

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

(con los tweets neutrales obtenido de los timelines)

Proporción de neutralidad:

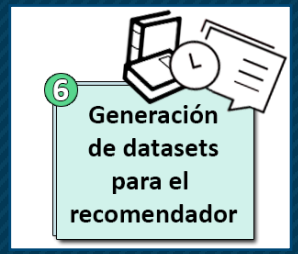
EJ: 150 tweets por 6025 users
Un 25% de neutralidad →

37 tweets x 6025 = ~230K elem

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

→ (Neutralidad)



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
```

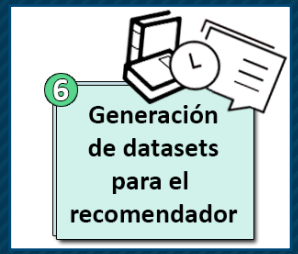
Archivo de mapeo de las claims

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

Claims etiquetadas:

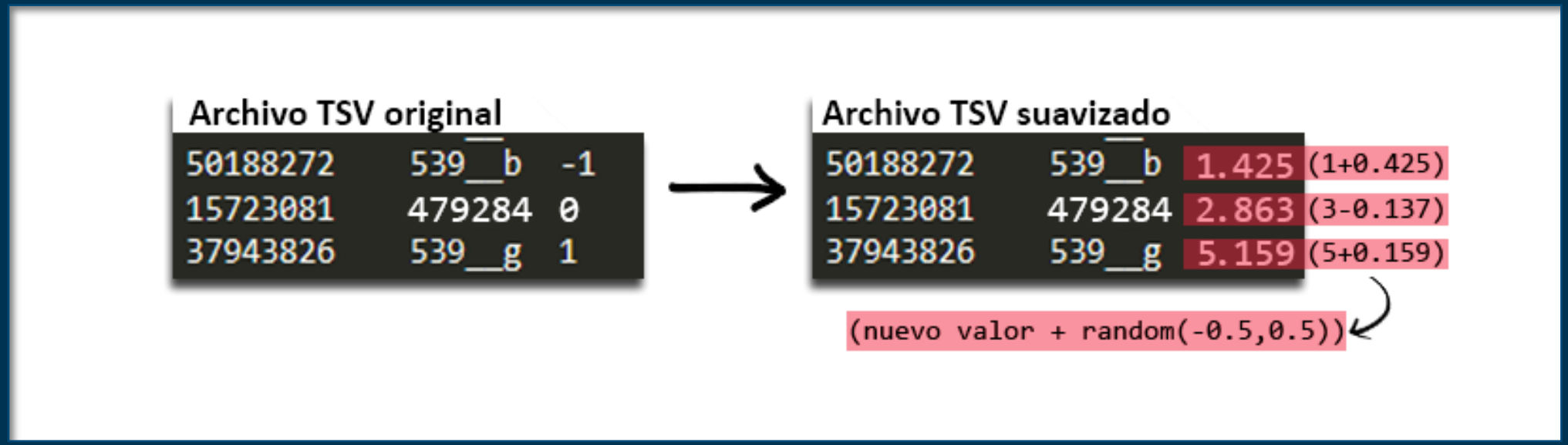
- __b → El usuarios está relacionado con dicha claim por **desinformar**.
- __g → El usuarios está relacionado con dicha claim por usarla para **verificar**.

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



Sistema desarrollado

Suavizado de los datos



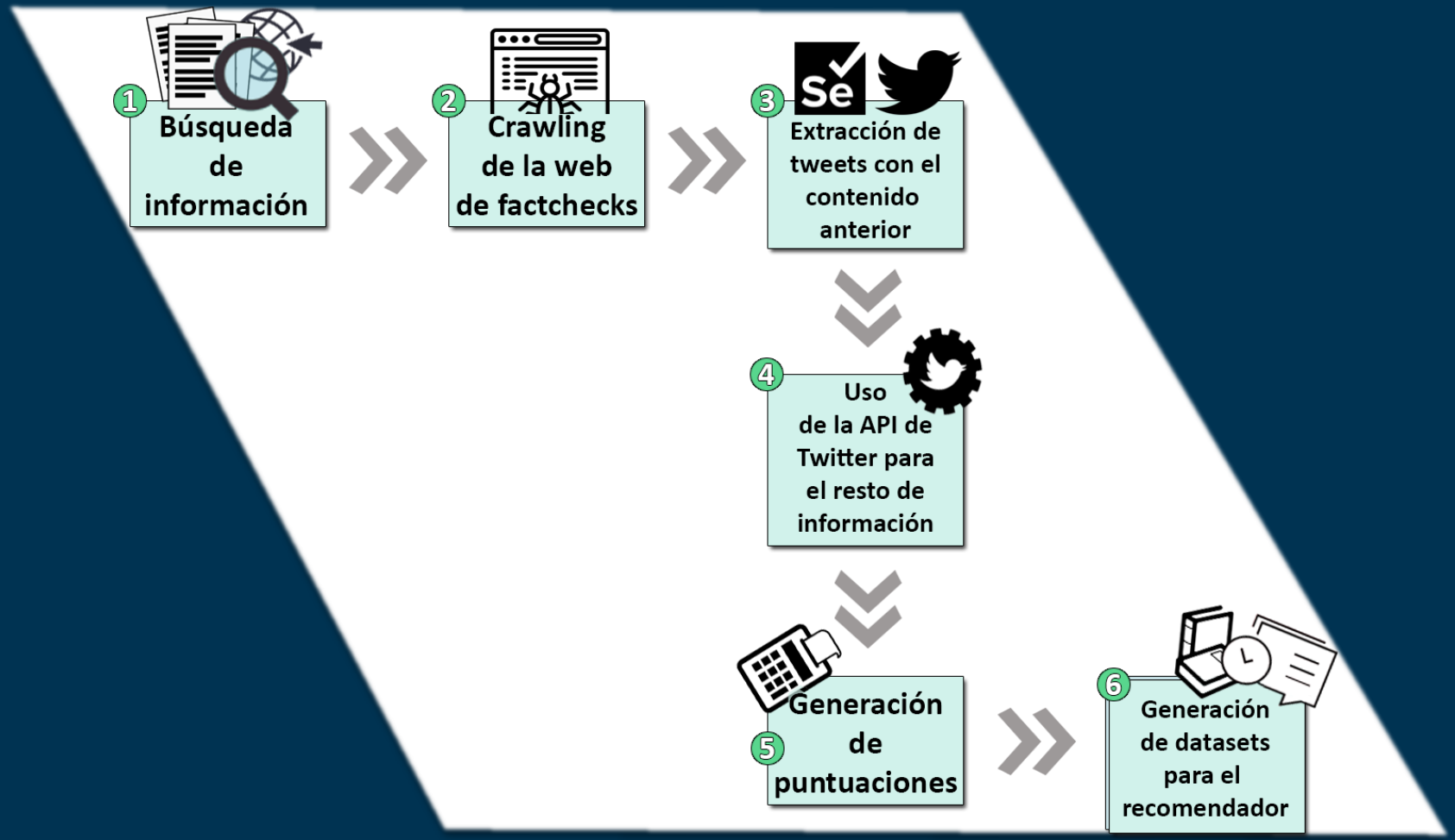
Objetivos para mejorar la recomendación:

- 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



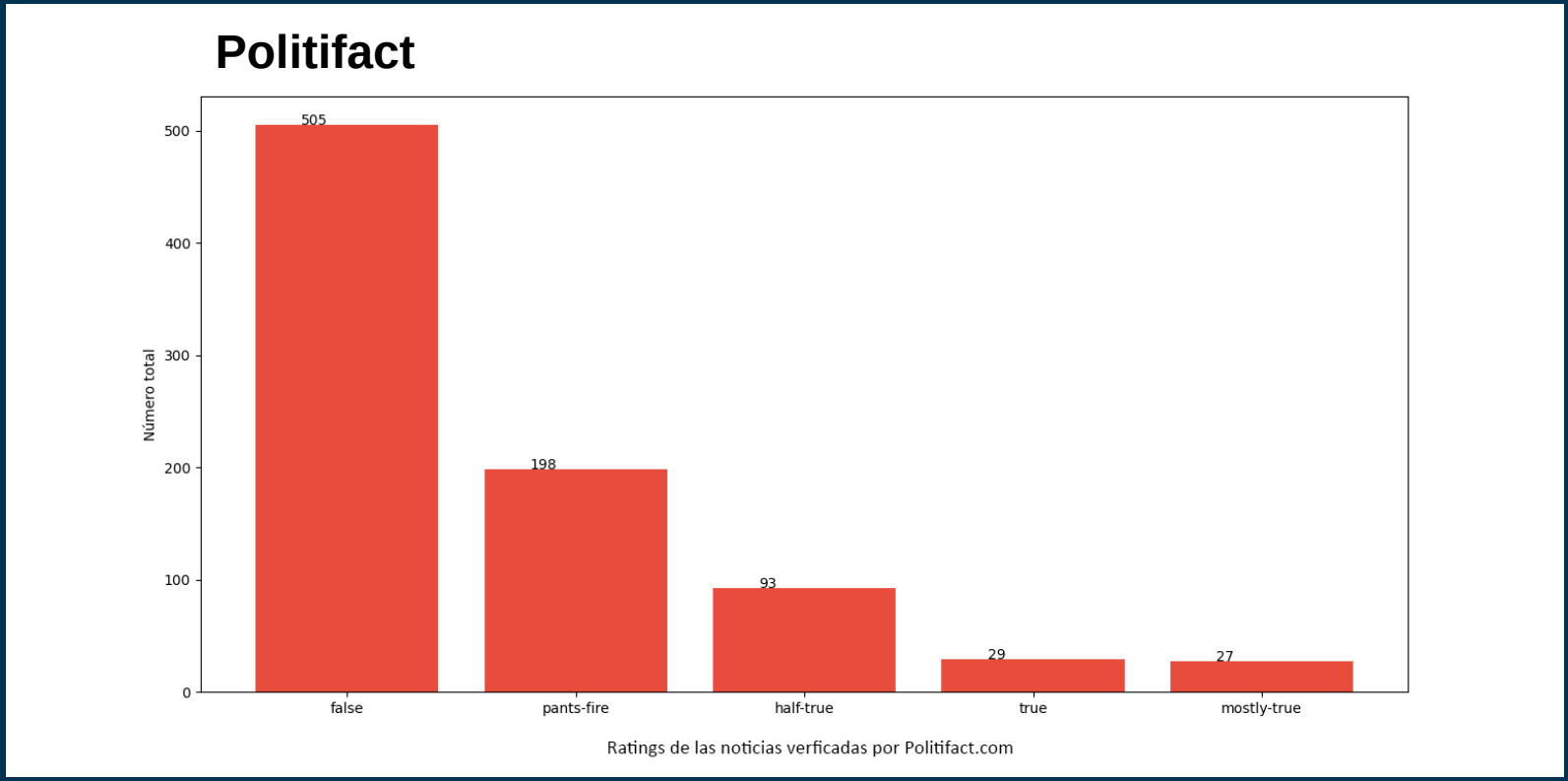
Experimentos

Análisis del dataset final

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

Experimentos

Análisis del dataset final



Falsos : false y pants-fire
Verdaderos : half-true, true y mostly-true

Politifact (elementos verificados)

703 elementos como **falsos**

149 elementos como **verdaderos**

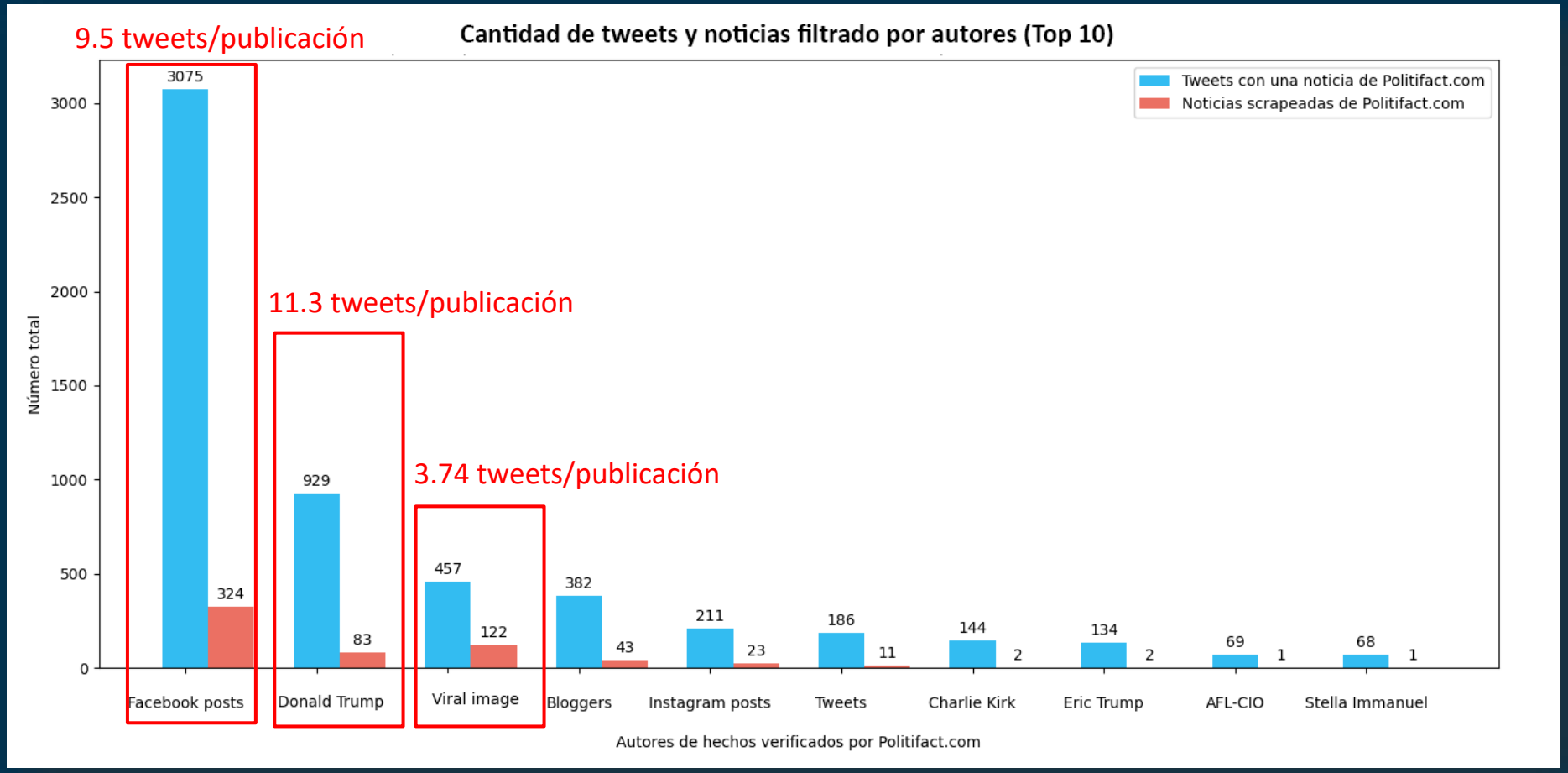
Twitter (tweets que comparten estas publicaciones de Politifact)

5792 tweets con claims puntuadas como **falsas**

415 tweets con claims puntuadas como **verdaderas**

Experimentos

Análisis del dataset final

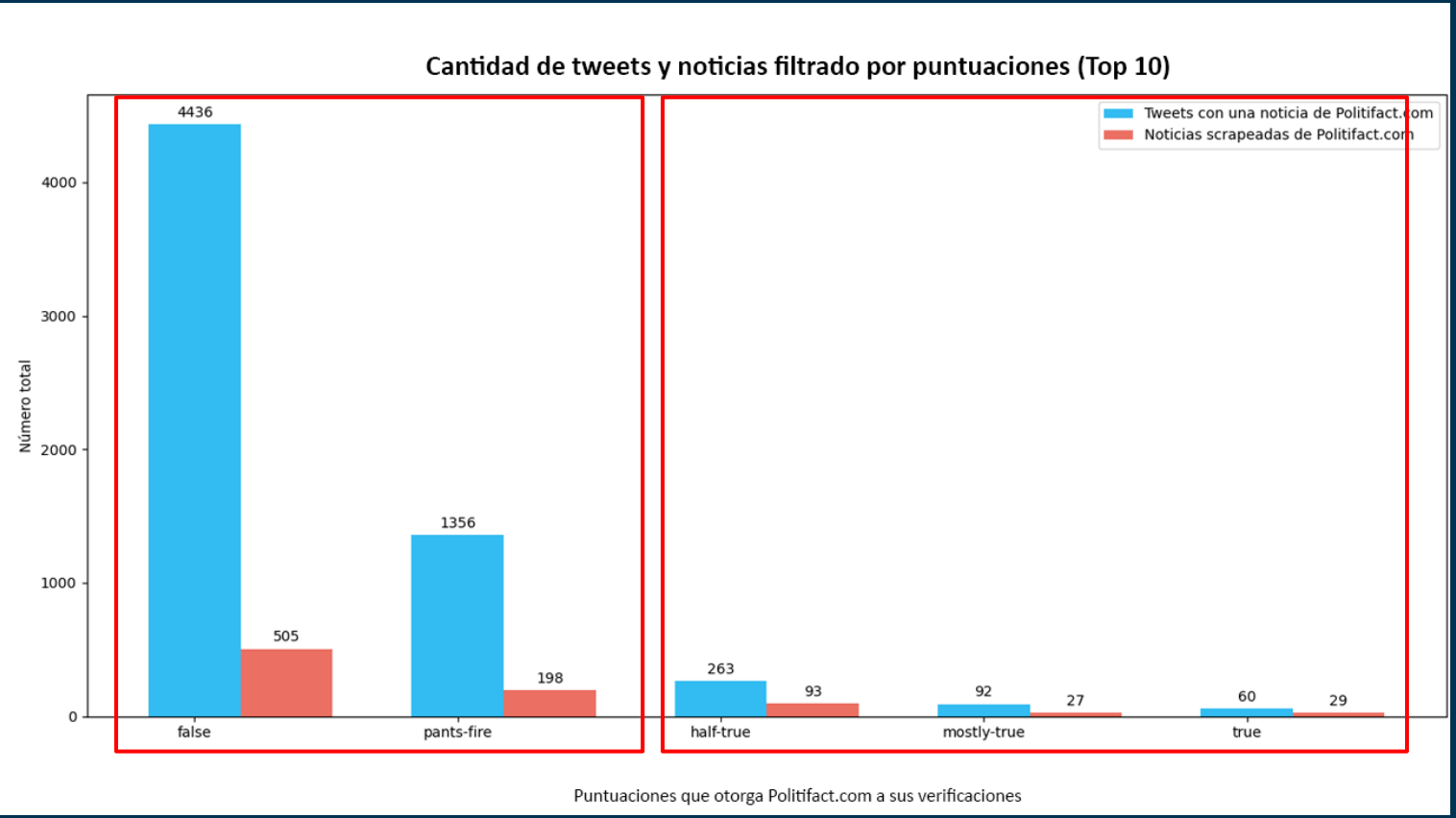


Una mayor proporción indica un mayor interés por compartir dichas verificaciones en la red social

Experimentos

Análisis del dataset final

¿Qué se comparte más?



Interacción en twitter

8.24 tweets/publicación con puntuación **falsa**

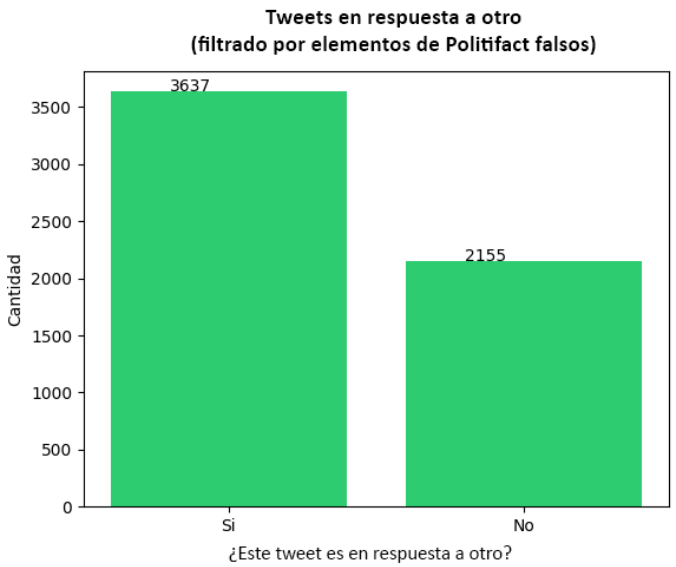
2.78 tweets/publicación con puntuación **verdadera**

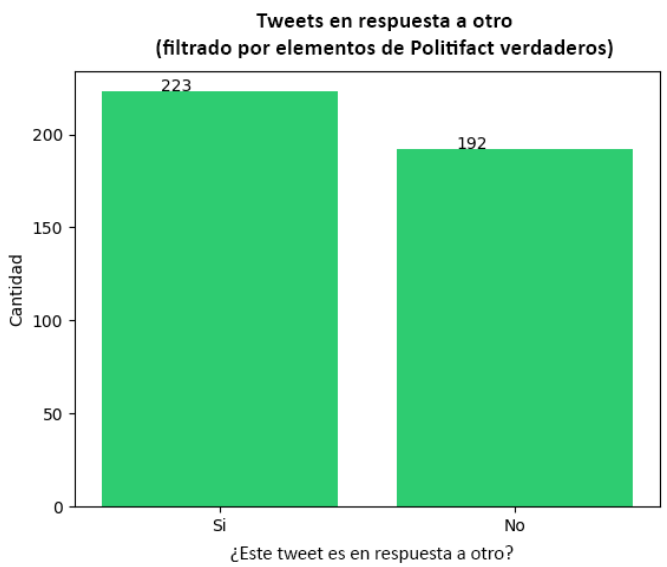
x4!!

Experimentos

Análisis del dataset final

¿Cuándo se responde más?





51% son tweets en respuesta

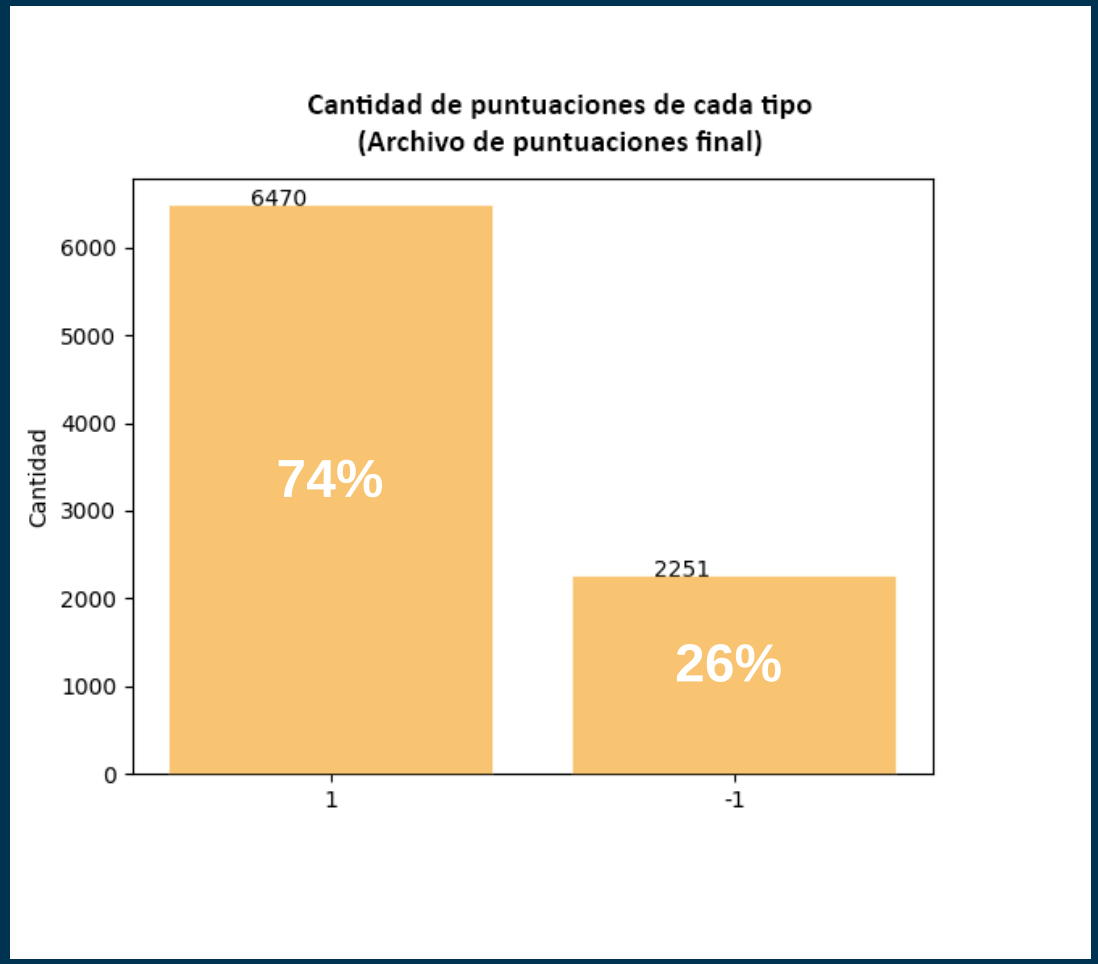
62% son tweets en respuesta

Es más fácil desinformar desmintiendo una noticia verdadera que inventando algo

Experimentos

Análisis del dataset final

¿Cómo está repartido el dataset final?

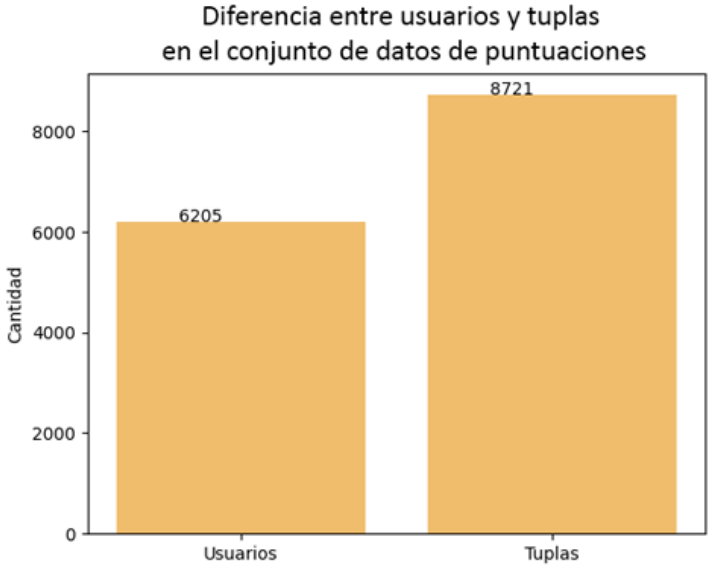


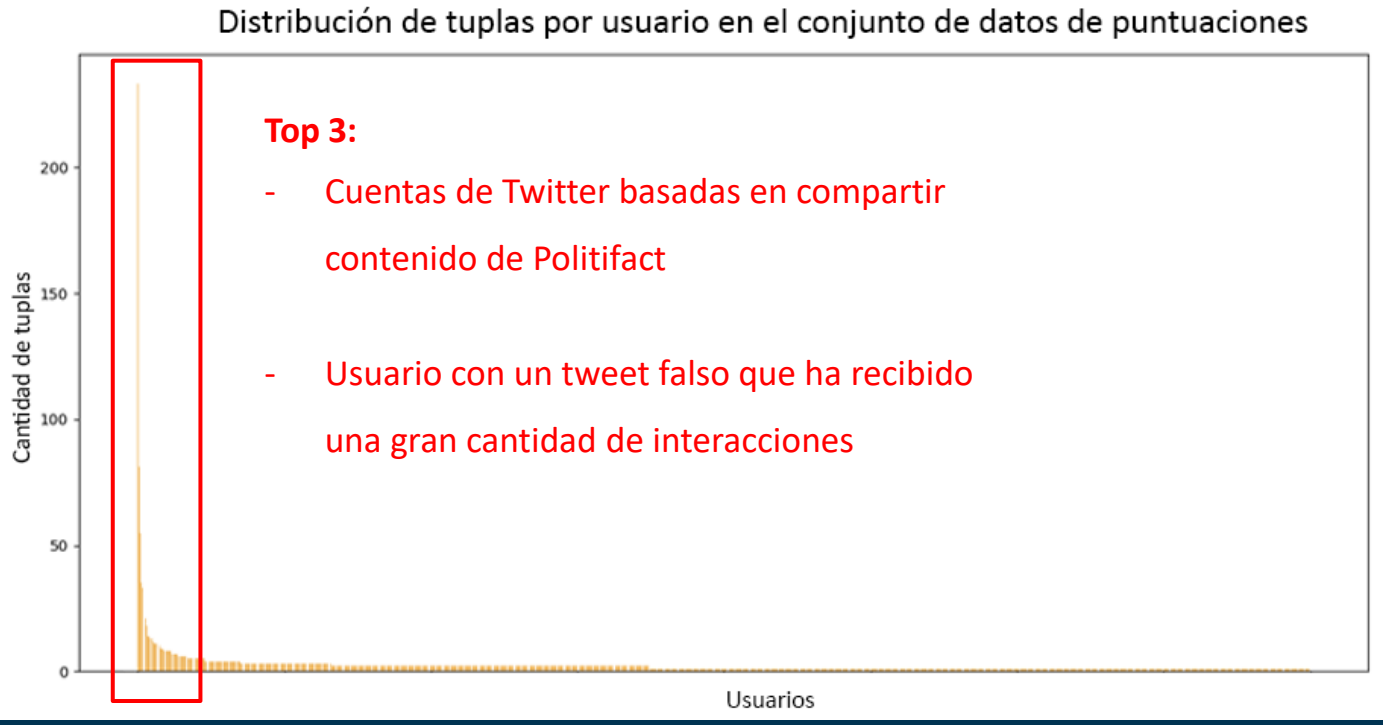

```
{
  "user": "50188272",
  "rate": -1,
  "tweet_id_rated": "1337140529826435072",
  "tweet_id": "1327007193661575168",
  "tweet_in_reply": null,
  "claim": null,
  "claim_in_reply": "https://www.politifact.com/fac
},
{
  "user": "15723081",
  "rate": -1,
  "tweet_id_rated": "1326910757393608704",
  "tweet_id": "1326631254545018884",
  "tweet_in_reply": null,
  "claim": "https://www.politifact.com/factchecks/2
  "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/2
  "claim_in_reply": null
},
```

Experimentos

Análisis del dataset final

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





Top 3:

- Cuentas de Twitter basadas en compartir contenido de Politifact
- Usuario con un tweet falso que ha recibido una gran cantidad de interacciones

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

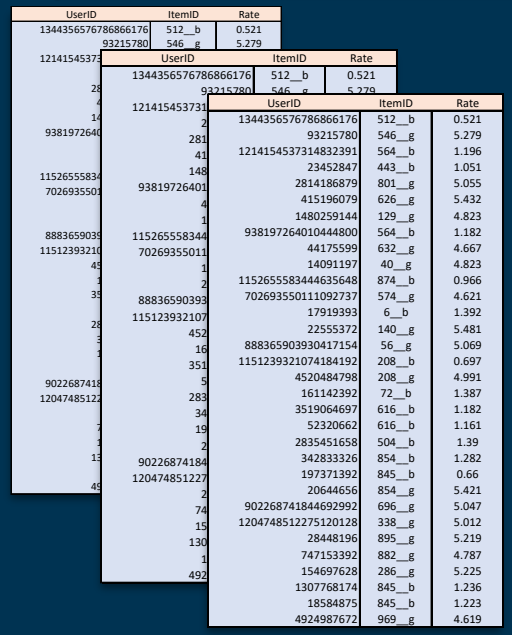
10%
30%
60%
90%

Proporción de neutralidad

0%
25%
50%

Debido a la alta dispersión y cantidad de elementos se requiere de mucha potencia computacional

Datasets generados aleatoriamente para la recomendación



UserID	ItemID	Rate
1344356576786866176	512__b	0.521
93215780	546__g	5.279
12141545373		
28		
4		
14		
9381972640		
11526555834		
7026935501		
4		
1		
8883659039		
11512393210		
45		
1		
35		
28		
7		
1		
9022687418		
12047485122		
7		
1		
13		
45		

UserID	ItemID	Rate
1344356576786866176	512__b	0.521
93215780	546__g	5.279
12141545373		
2		
281		
41		
148		
93819726401		
4		
1		
115265558344		
70269355011		
1		
2		
88836590393		
115123932107		
452		
16		
351		
5		
283		
34		
19		
2		
90226874184		
120474851227		
2		
74		
15		
130		
1		
492		

UserID	ItemID	Rate
1344356576786866176	512__b	0.521
93215780	546__g	5.279
12141545373148323 91	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	632__g	4.667
14091197	40__g	4.823
115265558344463564 8	874__b	0.966
702693550111092737	574__g	4.621
17919393	6__b	1.392
22555372	140__g	5.481
888365903930417154	56__g	5.069
115123932107418419 2	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
197371392	845__b	0.66
20644656	854__g	5.421
902268741844692992	696__g	5.047
120474851227512012 8	338__g	5.012
28448196	895__g	5.219
747153392	882__g	4.787
154697628	286__g	5.225
1307768174	845__b	1.236
18584875	845__b	1.223
4924987672	969__g	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					

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%

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

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%

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
0.9	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%

Experimentos

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)**

Experimentos

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)**

Experimentos

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)**

Experimentos

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)**

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)**

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

x2!!

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

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

Experimentos

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

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

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

Experimentos

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++****

(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

*** 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

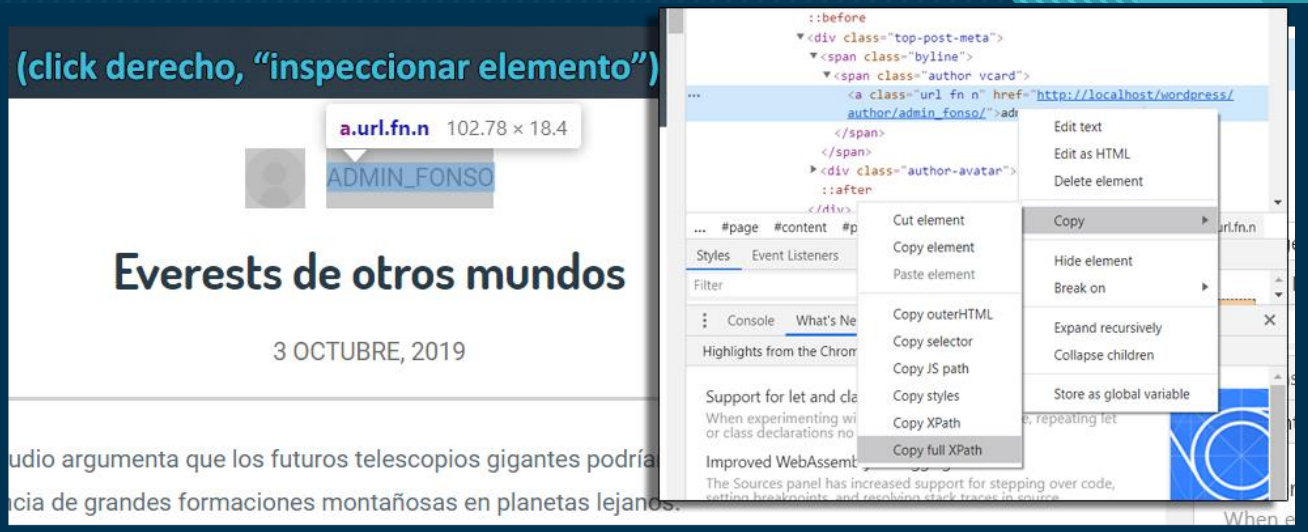
CONCLUSIONES Y TRABAJO FUTURO

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**

Experimentos

Análisis del dataset final

¿Cuáles son los claims más populares en Twitter?

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 Floydd
Claim_10	Voters in Maricopa County, Ariz., were forced to...	False	Fraude electoral