

La desinformación desde el análisis computacional del discurso

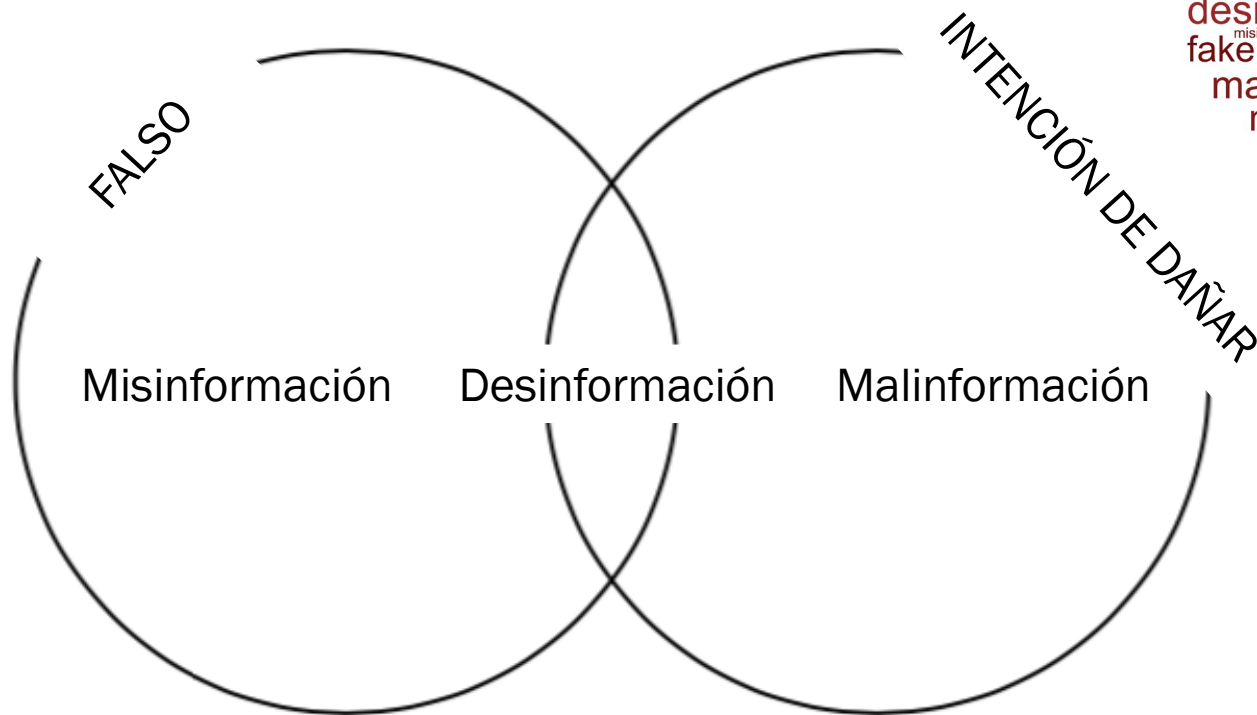
Maite Taboada
Simon Fraser University
Vancouver, Canadá
mtaboada@sfu.ca

Jornadas “La lingüística frente a la amenaza de la desinformación”
Universidad de Navarra
27 mayo 2024



SIMON FRASER UNIVERSITY
ENGAGING THE WORLD

Trastornos de la información



engaño truco sátira
suplantación propaganda news
falso
desinformación
fake misinformation estafa rumor
bullo
malinformación
montaje

El puzle de la desinformación

Inteligencia artificial y análisis del lenguaje

1

Educación

2

Parar la propagación



4

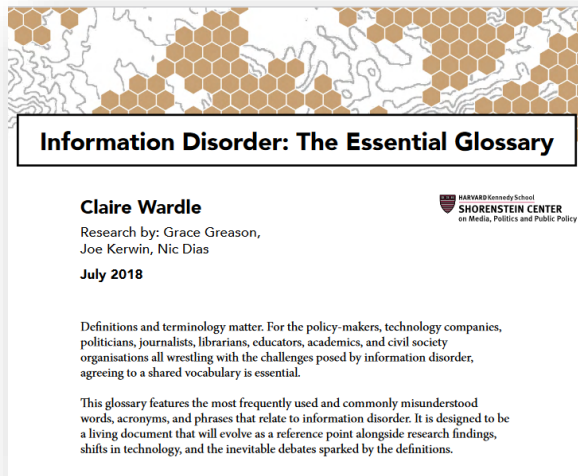
Verificación automática

3

Verificación experta de datos

Investigación dentro de las ciencias sociales

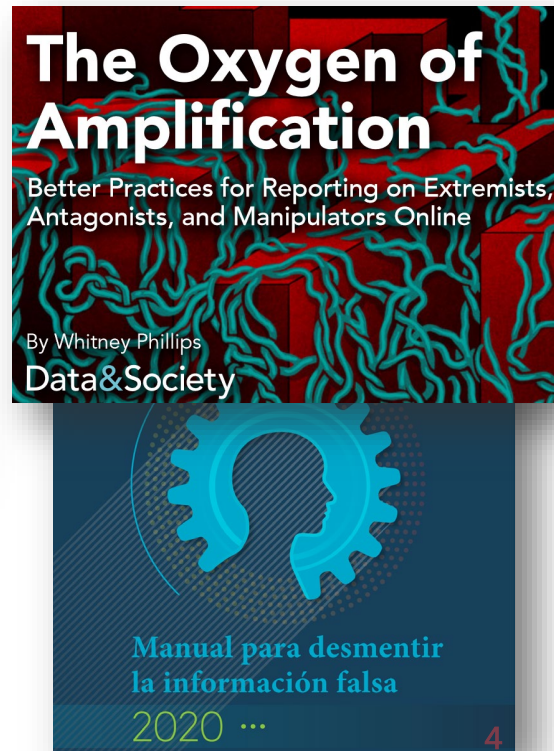
Definiciones



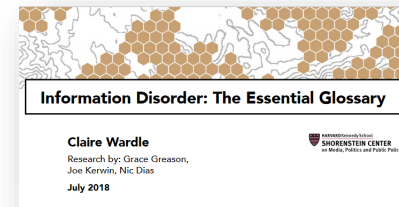
Propagación



Verificación y amplificación



Investigación: Definiciones



MALDITO BULO

- Bulo
 - Datos, hechos, imágenes
 - Contrastados que son falsos o manipulados
 - Sacados fuera de contexto
 - Satíricos, presentados como reales
- No hay pruebas
 - No se sabe si son ciertos

CHEQUEADO

- Inchequeable
- Verdadero
- Verdadero, pero...
- Discutible
- Apresurado
- Exagerado
- Engañoso
- Insostenible
- Falso

Investigación: Propagación

RESEARCH

SOCIAL SCIENCE

The spread of true and false news online

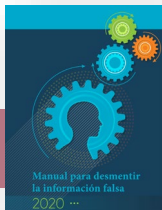
Soroush Vosoughi,¹ Deb Roy,¹ Sinan Aral^{2*}

LAS NOTICIAS FALSAS SON MÁS VIRALES

- Las noticias falsas y la desinformación
- También las noticias satíricas
 - Se propagan más rápidamente
 - Llegan a más gente

Investigación: Verificación y amplificación

VERIFICACIÓN



- Herramienta importante
 - Precisa de expertas
 - Política, economía, salud, cambio climático...
- Cuidado con el método
 - Para los desmentidos
 - Métodos basados en psicología cognitiva
 - El bocadillo de la verdad



AMPLIFICACIÓN



- Cuidado con amplificar las noticias
- A veces la verificación o la discusión en los medios simplemente difunde la noticia falsa más
- Manipulación: globos sonda
- Psicología cognitiva
 - Nos acordamos de la noticia
 - ¡No si de era falsa o no!

Un problema social



NO ES INDIVIDUAL

- No lo podemos resolver como individuos
- Necesitamos de las instituciones
 - Educación
 - Regulación de los medios y las plataformas digitales

NO ES IGUAL PARA TODAS

- La desinformación afecta desproporcionadamente a ciertos colectivos
 - Mujeres
 - Inmigrantes
 - Grupos minoritarios
 - Periodistas
 - Activistas

El puzle de la desinformación

Inteligencia artificial y análisis del lenguaje

1

Educación

2

Parar la propagación



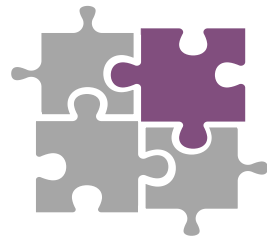
4

Verificación automática

3

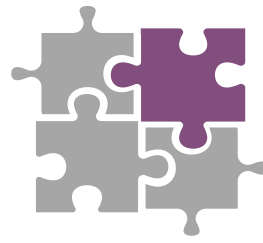
Verificación experta de datos

4. INTELIGENCIA ARTIFICIAL Y ANÁLISIS DEL LENGUAJE

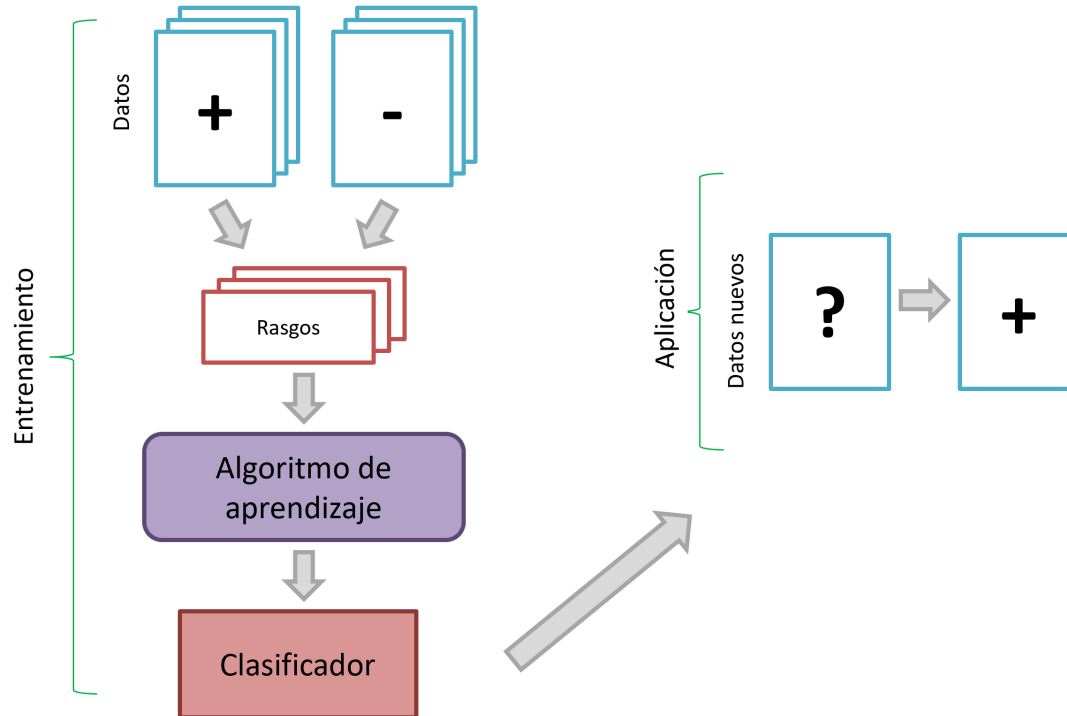


- Hipótesis
 - El lenguaje de las noticias falsas es diferente
 - Hay rasgos que las identifican
 - Independientemente de la distribución, el sesgo, etc.
- Propuesta
 - Podemos desarrollar un sistema que detecte las noticias falsas usando rasgos lingüísticos
- Técnicas
 - Inteligencia artificial, aprendizaje automático
 - Procesamiento del lenguaje natural

INTELIGENCIA ARTIFICIAL Y APRENDIZAJE AUTOMÁTICO

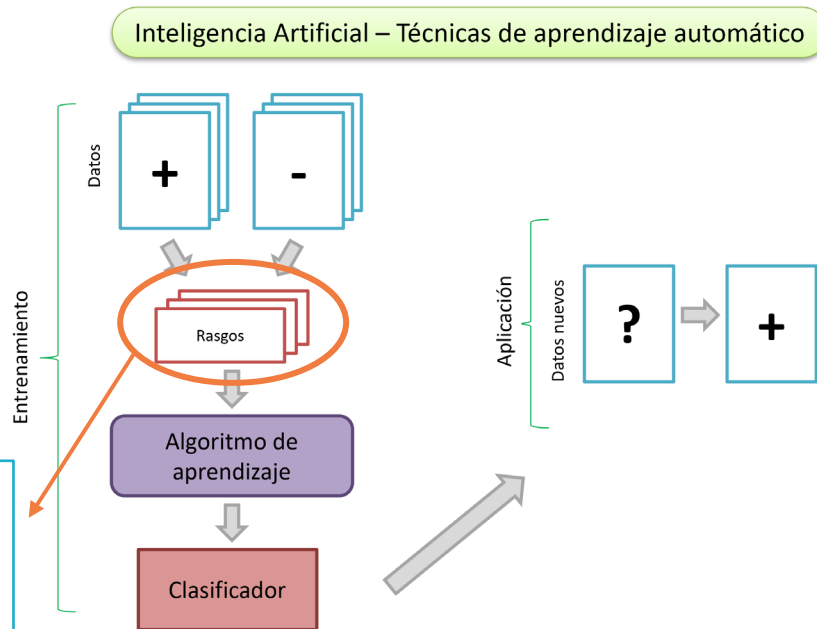


Inteligencia Artificial – Técnicas de aprendizaje automático



PROCESAMIENTO DEL LENGUAJE NATURAL

- Campo que solapa la informática, la inteligencia artificial y la lingüística
- Extracción de información lingüística para diferentes aplicaciones
 - Reconocimiento de entidades
 - Búsqueda y recuperación de información
 - Resumen automático
 - Traducción automática
 - Análisis de la opinion
 - ...



Nuestro proyecto



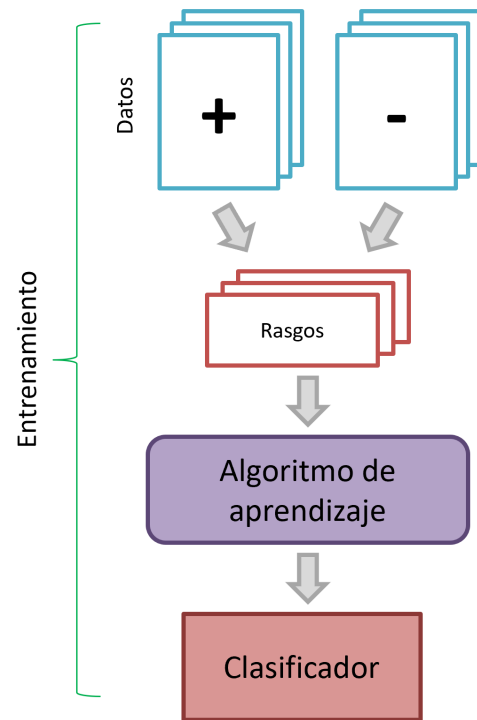
- Hipótesis
 - El lenguaje de las noticias falsas es diferente del lenguaje de las noticias contrastadas
- Métodos
 - Aprendizaje automático, PLN
 - Lingüística de corpus
 - Análisis multidimensional (Biber 1995)
 - Análisis de la variación lingüística en el texto



Fatemeh Torabi Asr

El aprendizaje automático necesita datos


- Los métodos actuales de aprendizaje automático utilizan grandes cantidades de datos
 - Aprendizaje supervisado
 - Input: corpus con etiquetas (spam/no_spam)
 - Output: un clasificador que diferencia esos dos tipos de textos en el corpus



Los datos


- La desinformación y las noticias falsas son un problema
- Son un problema de big data
- Los estamos solucionando con ‘small data’

Big Data and quality data for fake news and misinformation detection

Fatemeh Torabi Asr and Maite Taboada 

Abstract

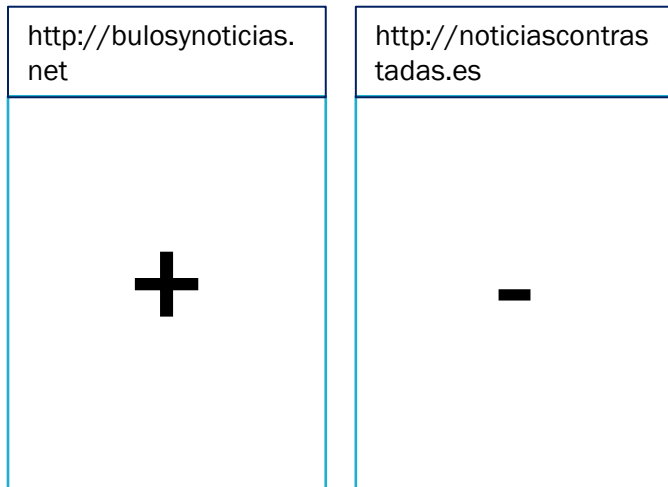
Fake news has become an important topic of research in a variety of disciplines including linguistics and computer science. In this paper, we explain how the problem is approached from the perspective of natural language processing, with the goal of building a system to automatically detect misinformation in news. The main challenge in this line of research is collecting quality data, i.e., instances of fake and real news articles on a balanced distribution of topics. We review available datasets and introduce the MisInfoText repository as a contribution of our lab to the community. We make available the full text of the news articles, together with veracity labels previously assigned based on manual assessment of the articles' truth content. We also perform a topic modelling experiment to elaborate on the gaps and sources of imbalance in currently available datasets to guide future efforts. We appeal to the community to collect more data and to make it available for research purposes.

Big Data & Society
January–June 2019: 1–14
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2053951719843310
journals.sagepub.com/home/bds




¿De dónde vienen los datos?

- Necesitamos dos tipos de datos
 - Ejemplos positivos (spam, noticias falsas)
 - Ejemplos negativos (no_spam, noticias contrastadas)
- Problemas
- Ruido
 - bulosynoticias.net igual reproduce noticias de medios fiables
- Sesgo
 - A mí me parece que bulosynoticias.net publica noticias falsas
- Necesitamos
 - Artículos contrastados, uno a uno, por expertas

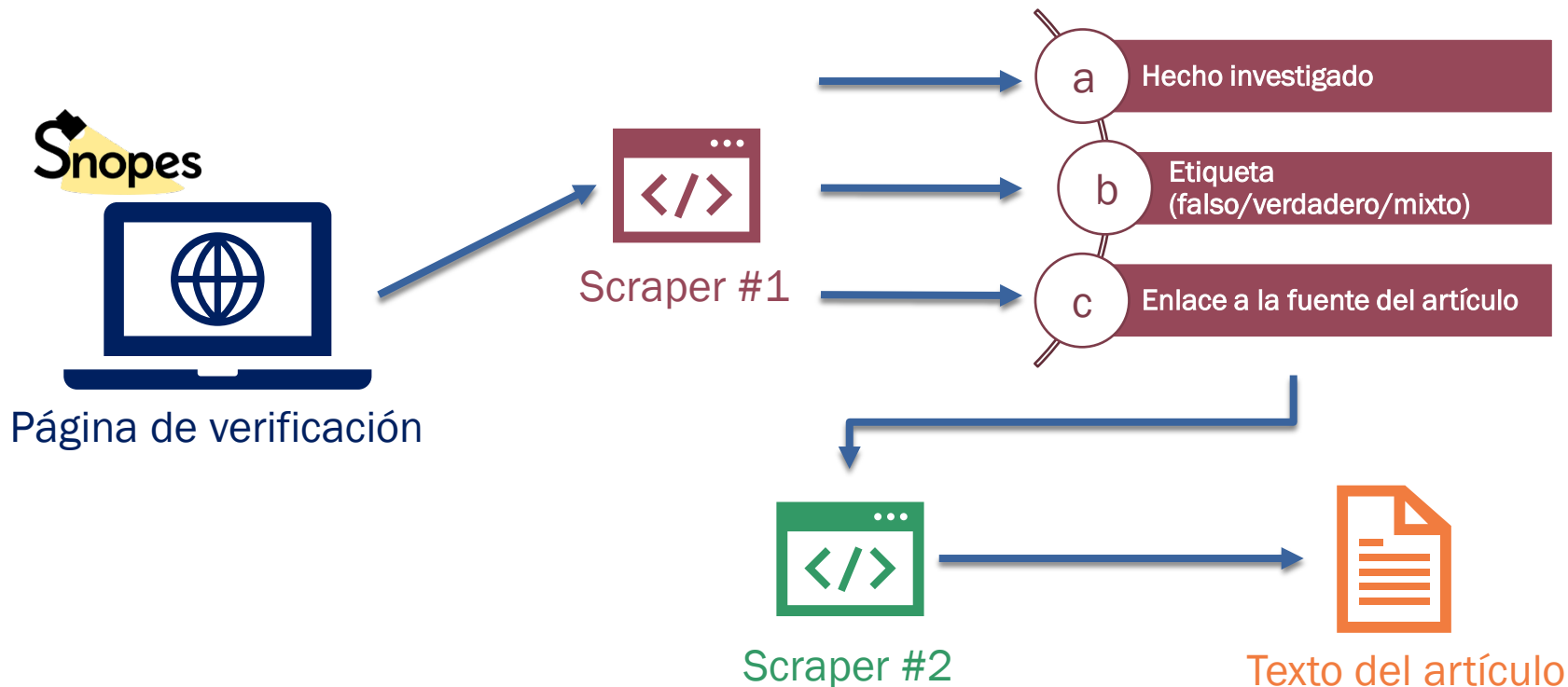


Artículos de verificadores de datos

- Fuentes
 - BuzzFeed – lista de las top noticias falsas del año
 - Artículos en Snopes, Politifact y Emergent
 - Cada uno, etiquetado individualmente
 - Falso/auténtico



Proceso de recopilación de datos



Did a Woman Open Fire at McDonald's for Forgetting Bacon on Her Burgers?

Adds new meaning to the definition of hangry.

By Madison Dapceвич

Published 6 June 2021



Image via [Christoph Schmidt/picture alliance via Getty Images](#)

Claim

A Michigan woman was sentenced to up to seven years in prison for opening fire at a McDonald's restaurant after workers forgot to put bacon on her burgers.

Rating



True

About this rating ⓘ

hecho investigado

etiqueta

¿enlace a la fuente?

¿enlace a la fuente?

¿enlace a la fuente?

Origin

A Michigan woman was sentenced to three to seven years in prison after opening fire in 2014 at a McDonald's restaurant where workers twice forgot to put bacon on her burgers during one visit.

Shaneka Torres, who was 30 at the time of her 2015 sentencing, reportedly became angry after the burger she ordered came without the requested bacon. She was offered a free meal by the restaurant, but bacon was left out of the second order, as well. That's when she fired a shot through the drive-thru window, according to [The Associated Press](#). No one was injured.

The case went viral again in May 2021 when TikTok user Tanner Lane [shared](#) the story to the social media platform:

Following Torres' arrest, [Time reported](#) that the woman was arrested at her home about 30 minutes after the 3 a.m. shooting on Feb. 10, 2014. Her defense attorney argued that Torres discharged the weapon by accident, and that that the incident was not correlated with the burger mishap.

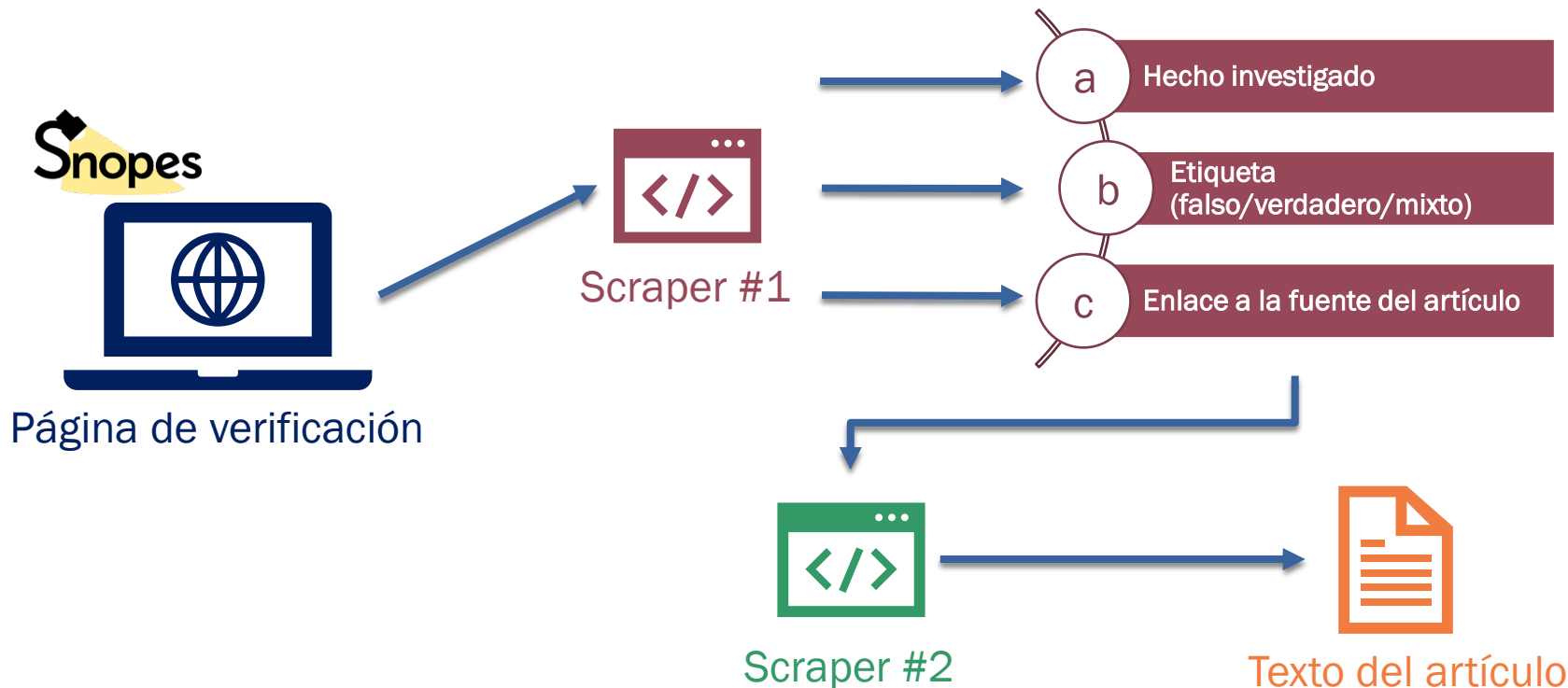
The jury in her trial reportedly deliberated for one hour before finding her guilty of carrying a concealed weapon, discharging a firearm into a building, and felony use of a firearm. The news website Michigan Live reported that Torres was also ordered to stay away from all McDonald's restaurants, but that it was at the discretion of the Michigan Department of Corrections to decide if the ban should continue after her prison release.

Snopes submitted a request to Kent County Circuit Court, the jurisdiction that convicted Torres, for more information surrounding the details of her arrest and subsequent sentencing. We did not hear back in time for publication but will update the article if we do.

By Madison Dapceвич

Published 6 June 2021

Proceso de recopilación de datos



Origin

A Michigan woman was sentenced to three to seven years in prison after opening fire in 2014 at a McDonald's restaurant where workers twice forgot to put bacon on her burgers during one visit.

Shaneka Torres, who was 30 at the time of her 2015 sentencing, reportedly became angry after the burger she ordered came without the requested bacon. She was offered a free meal by the restaurant, but bacon was left out of the second order, as well. That's when she fired a shot through the drive-thru window, according to [The Associated Press](#). No one was injured.

The case went viral again in May 2021 when TikTok user Tanner Lane [shared](#) the story to the social media platform:

Woman gets 3-7-years for shooting over bacon-less burger

April 21, 2015



Click to copy

RELATED TOPICS

[Shootings](#)

[Lifestyle](#)

[Business](#)

[Crime](#)

[Food and drink](#)

[Violent crime](#)

[General news](#)

GRAND RAPIDS, Mich. (AP) — A Michigan woman has been sentenced to three to seven years in prison for opening fire at a McDonald's restaurant after workers twice failed to put bacon on her burgers.

Authorities say 30-year-old Shaneka Torres became angry in February 2014 when the first burger she ordered at the restaurant's drive-up station was missing bacon. She was offered a free meal, but bacon also wasn't added to a second burger order.

Police say she fired a shot through the restaurant. No one was injured.

A Kent County Circuit Court jury convicted her March 25. Judge Paul Sullivan on Tuesday sentenced her to one to five years for shooting at a building, plus two years for possessing a firearm during a felony.

Torres says she's sorry "but it's over and done with."

<https://apnews.com/article/64f5dd3baa2942548a7eb210806c5d56>

Fase 1 (Snopes/Politifact)

Fase 2 (páginas externas)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	politifact_url_phase1	fact_tag_phase1	article_title_phase1	article_claim_phase1	article_claim_phase1	article_publication_date_phase1	article_researcher_by_phase1	article_edited_by_phase1	article_categories_phase1	original_url_phase1	page_is_first_publication_phase1	error_phase2	original_article_text_phase2	article_title_phase2	publish_date_phase2	author_phase2
2	http://www.politifact.com	Half-True	Georgia gas tax	"Contrary to what E. Frank	Wednesday, April	Hunt	Jim Tharpe	Gas Prices, Taxes	http://www.gasprices.com	TRUE	No Error	By Clay G. Co	Georgia Gas	2011-08-13		
3	http://www.politifact.com	TRUE	Health insurance	If you "alread	Barack Obama	Wednesday, April	Angie Drobni	Bill Adair	Health Care	http://www.frwed.com	TRUE	No Error	Congressional	Browse Congressional Bills		
4	http://www.politifact.com	TRUE	Analyst says	"Switzerland	Matt Miller	Wednesday, April	Louis Jacobson	Bill Adair	Health Care	http://www.politifact.com	TRUE	No Error	A dangerous	Why the Pub	2005-09-06	Matt Miller
5	http://www.politifact.com	Pants on Fire	Gov. Jan Brewer	"Our law enforcement	Jan Brewer	Wednesday, April	Robert Farley	Martha M. H.	Immigration	http://www.politifact.com	TRUE	No Error	In an exclusive	Brewer wants	2006-06-24	Brahm Resnik
6	http://www.politifact.com	Mostly False	Obama says	John Boehner	Barack Obama	Wednesday, April	Angie Drobni	Martha M. H.	Economy, Education	http://www.politifact.com	TRUE	No Error	2:11 P.M. CD	Remarks by the President	2006-09-05	
7	http://www.politifact.com	FALSE	Congressional	"Close to 30%	Daniel M. Claitor	Wednesday, April	Cary Spivak	Greg Borowski	Immigration	http://www.politifact.com	TRUE	No Error	Greetings, Be	Mielke For Congress		
8	http://www.politifact.com	Mostly True	Hillary Clinton	Says Donald Trump	Hillary Clinton	Wednesday, April	Jon Greenberg	Katie Sander	Foreign Policy	http://www.politifact.com	TRUE	No Error	Who is better	Commander-in-Chief	2012-09-07	Years Ago
9	http://www.politifact.com	Mostly True	Do girls make	Young women	Justin Trudeau	Wednesday, April	Joseph Cariz	Aaron Sharov	Public Health	https://africa.com	TRUE	No Error	Despite signs	Are 74% of African girls aged 15 to 24 HIV+?		
10	http://www.politifact.com	FALSE	Trump repeats	"I was totally	Donald Trump	Wednesday, April	Lauren Carro	Katie Sander	Iraq	http://www.politifact.com	TRUE	No Error	Donald Trump	Trump Respects	2012-02-20	
11	http://www.politifact.com	TRUE	Sen. Rob Portman	Says the federal	Rob Portman	Wednesday, April	Tom Feran	Robert Higgs	Energy	http://www.politifact.com	TRUE	No Error	The national	Toledo can lead	2007-08-27	
12	http://www.politifact.com	Half-True	Did Republican	"Republicans	SEIU on	Wednesday, April	Angie Drobni	Martha M. H.	Corporations	http://thomas.com	TRUE	No Error	Array (=>	2 C.H.R.1586 - 111th Congress	Rangel	
13	http://www.politifact.com	Half-True	Sid Miller: Po	Says a poll sh	Sid Miller	Wednesday, April	W. Gardner	John Bridges	History, Polls	http://www.politifact.com	TRUE	No Error	As Austin off	Efforts underway in large	Johnathan Silver	-
14	http://www.politifact.com	Half-True	Jeff Denham	Says Barack C	Jeff Denham	Wednesday, April	Chris Nichols	Gregory Favre	Homeland Security	http://www.politifact.com	TRUE	No Error	Denham: Pro	Denham: Pro	2013-09-04	
15	http://www.politifact.com	Mostly False	In Milwaukee	"In 1986, Pre	Donald Trump	Wednesday, April	Tom Kertsch	Greg Borowski	Corporations	http://www.politifact.com	TRUE	No Error	CLOSE Presid	Trump: We n	2013-09-02	Donald Trump
16	http://www.politifact.com	Mostly True	Diana DeGett	Women "recl	Diana DeGett	Wednesday, April	Louis Jacobson	Angie Drobni	Jobs, Women	http://thomas.com	TRUE	No Error	Senate Comm	S.181 - 111th Congress (20	Mikulski	
17	http://www.politifact.com	Pants on Fire	Wasserman	Says she was	Debbie Wasserman Schultz	Wednesday, April	Amy Sherman	Angie Drobni	Israel	http://m.washingtonpost.com	TRUE	No Error	Its the mess	Jewish Demo	2008-09-02	Brendan Smialowski
18	http://www.politifact.com	TRUE	Chris Christie	Says Bruce S	Chris Christie	Wednesday, April	Bill Wichert	Caryn Shinski	Pop Culture	http://www.politifact.com	TRUE	Not a news a	Published on Sep 5, 2012	Late Night with Jimmy Fallon. Se		

Verificación manual

- Para cada enlace que pensamos contiene el artículo:
- Fase 1
 - El artículo del verificador y el enlace que estamos contemplando
- Fase 2
 - La página externa
 - ¿Existe todavía?
 - ¿Se refiere al evento investigado?
 - ¿Hemos extraído el texto correctamente?

Crowdsourcing

- Después de esta fase, teníamos 2,000 artículos
 - Validación manual
 - Comprobar el enlace
 - Comprobar que el enlace se refiere a la noticia en Snopes
- Mediante crowdsourcing

Corpus

Source	False	True	Small balanced sample	Large mixed sample
Snopes Silver	1,585	259	$2 * 259$	$2 * 1,300$
Buzzfeed USE	64	1,090	$2 * 64$	
Total	1,649	1,349	646	2,600

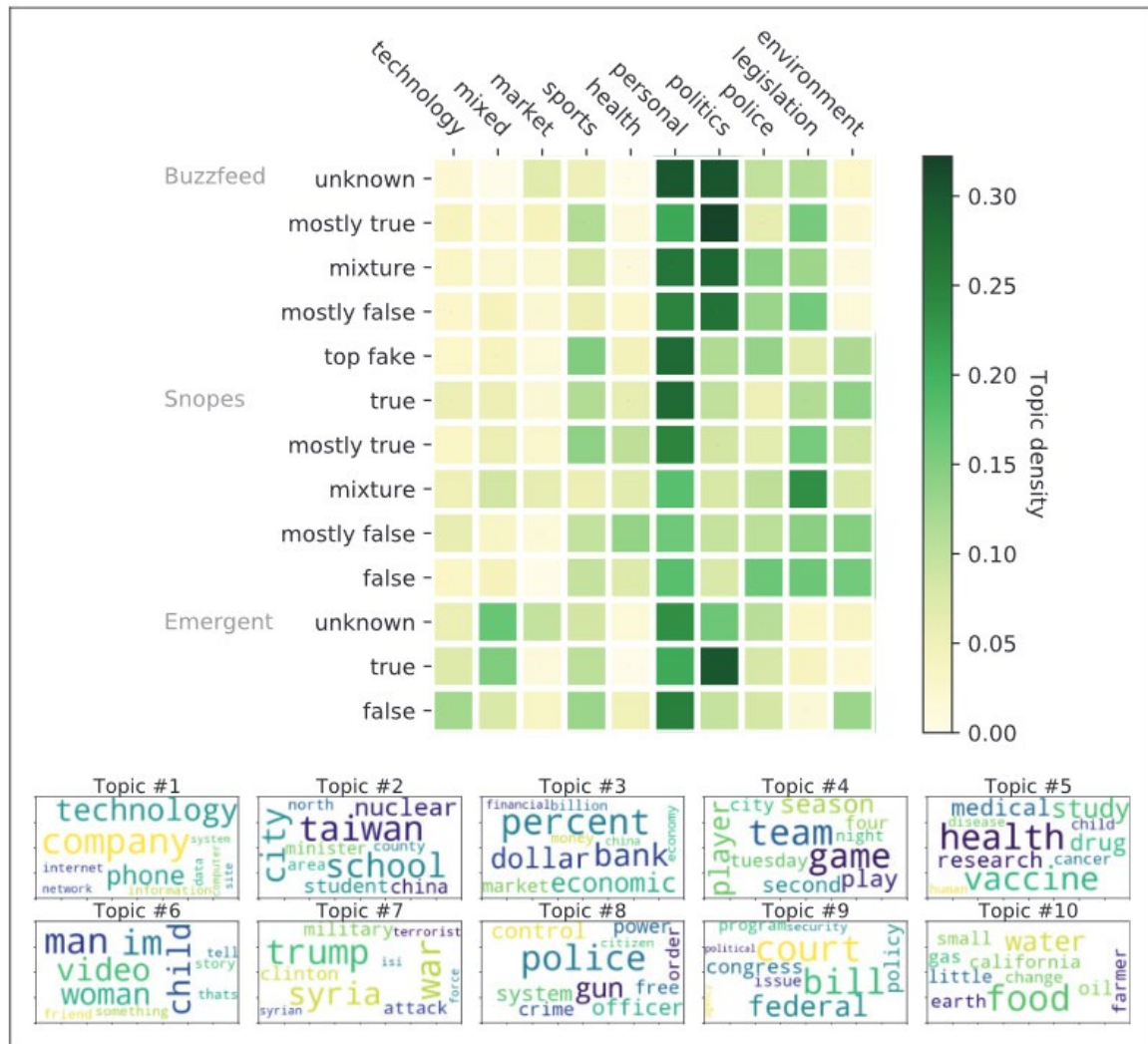
Snopes Silver, validado con crowdsourcing

Buzzfeed, noticias falsas de las elecciones en EE.UU

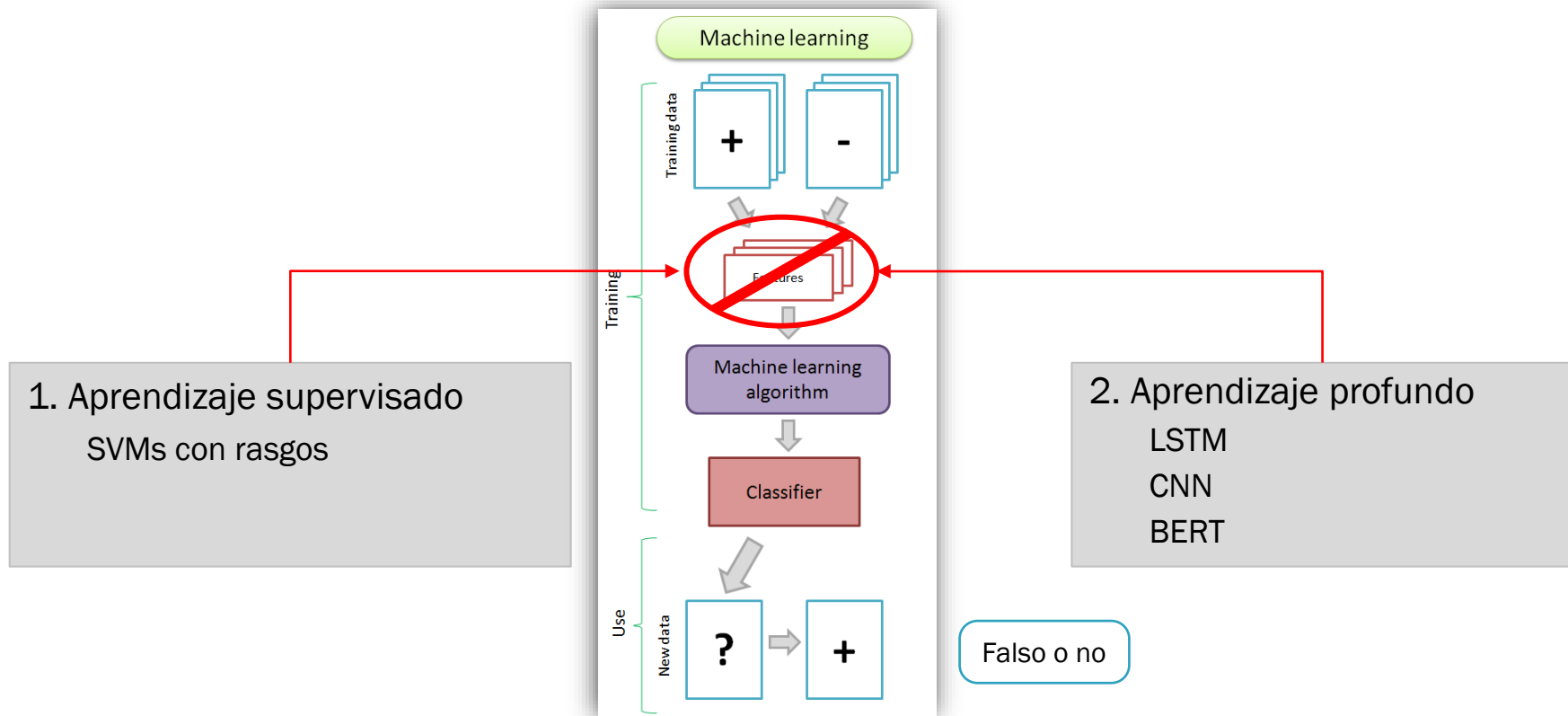
Cuando los juntamos, tenemos una mezcla de temas y personalidades

Usado como corpus de entrenamiento

Mezcla de temas y personalidades



¡Tenemos datos!



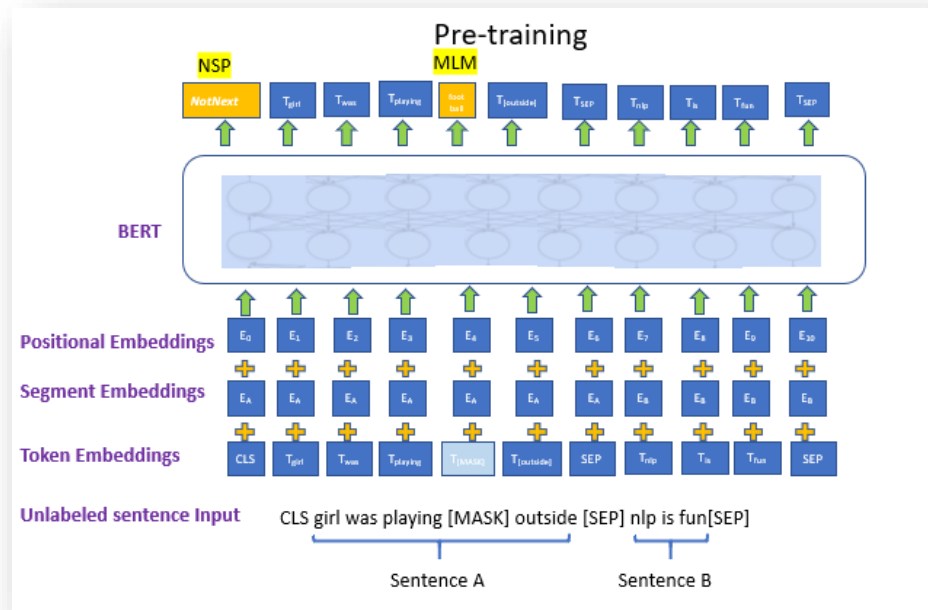
Métodos de clasificación – 1

- 1. SVMs – Máquinas de soporte vectorial
- Rasgos:
 - N-grams (TF-IDF)
 - Semánticos
 - LIWC, subjectividad
 - Sintácticos
 - POS – partes del discurso
 - Sustantivos, verbos, adjetivos, etc.
 - Nivel de lectura
 - Primaria/educación media/superior



Métodos de clasificación – 2

- 2. Aprendizaje profundo
- ULMFiT
 - Howard & Ruder 2018
 - Modelo de lenguaje pre-entrenado
 - Basado en vectores de palabras
- BERT
 - Modelo puntero de Google
 - También pre-entrenado



Corpus de entrenamiento y corpus de test

ENTRENAMIENTO

- Rashkin et al. (2017)
 - Propaganda/bulo/satírico/fiable
- Snopes Silver
 - Propio, mediante crowdsourcing
 - Corpus grande, temas mixtos
 - Corpus pequeño, menos mezcla

TEST

- Snopes Gold
 - Propio
 - Verificación del corpus de crowdsourcing
- BuzzFeed Top
 - Top 33 – noticias falsas
- Pérez Celebrity
 - Pérez-Rosas et al. (2017)
 - Noticias falsas sobre celebridades

Resultados

Table 4. Performance of different text classification models in various data scenarios measured by F1-score on training data as well as three test datasets

Data scenario	Model	Train	BuzzTop	SnopesGold	PerezCeleb
Rashkin mapped (8,000 train items)	SVM N-gram	98	94	54	56
	SVM N-gram+Sem	97	94	52	56
	SVM N-gram+Sem+Syn	96	91	53	59
	SVM All features	94	88	55	56
Big mixed (2,600 train items)	SVM N-gram	86	97	54	51
	SVM N-gram+Sem	86	97	54	51
	SVM N-gram+Sem+Syn	86	97	54	51
	SVM All features	83	85	50	52
	ULMFiT	83	95	51	50
Small balanced (646 train items)	SVM N-gram	96	88	67	64
	SVM N-gram+Sem	95	88	70	64
	SVM N-gram+Sem+Syn	86	88	70	65
	SVM All features	94	97	59	58
	ULMFiT	75	73	61	58

- Sesgo
 - Artículos sobre el mismo tema
 - La distribución por tema es importante
 - Detectar todo como bulo
- Los SMVs hacen un trabajo adecuado
- Los métodos de aprendizaje profundo necesitan muchos más datos

Rasgos – ¿cuáles son significativos?

- Semánticos
 - Palabras negativas
 - Adverbios
 - Utilidad del análisis de la opinion
(Alonso, Vilares, Gómez-Rodríguez y Vilares 2021)
- LIWC
 - Contenido sexual
 - Pronombres
 - Adverbios
- Syntácticos
 - Verbos
 - Partículas
 - Adverbios
 - Pronombres (*they*=fake; *I*=true)
- Rasgos del texto
 - Bulos = más cortos
 - Menos apóstrofes
 - *do not* frente a *don't*

Método: Ratio y valor p en un análisis de correlación

Rasgos lingüísticos de los bulos

1 PREGUNTAS RETÓRICAS

- ¿Quién va a pagar por esto?
- Muchos españoles se preguntan, ¿qué pasa con la inmigración?

2 PALABRAS NEGATIVAS

- una situación escalofriante
- severamente perturbador
- una burla de la gente decente

3 EXAGERACIONES

- Muchos adverbios y adjetivos
- alucinante testimonio de un ciudadano que asiste estupefacto al diálogo entre un trabajador de La Caixa y un inmigrante

4 LENGUAJE DE OTREDAD/ IDENTIDAD

- Nosotros frente a 'ellos'
- esa gente
- estos

5 MEZCLA DE ESTILO

- Mezcla de lenguaje formal e informal

6 LENGUAJE NO ESTÁNDAR

- Errores gramaticales, faltas de ortografía
- El caso de los adolescentes de Macedonia

OUR BATTLE AGAINST FAKE NEWS

The aim of the fake news project is to help news readers to identify bias and misinformation in news articles in a quick and reliable fashion. We have collected news articles with veracity labels from fact-checking websites and used them to train text classification systems to detect fake from real news. You can paste a piece of text and examine its similarity to our collection of true vs. false news articles, or to news from four different types/genre. Enjoy testing!

FAKE NEWS DETECTION



Please select a method:

- ☒ Fact Checking
☐ Genre Classification

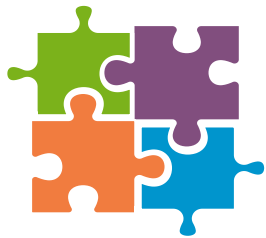
Please enter your text:

Clinton won the 2016 US election.

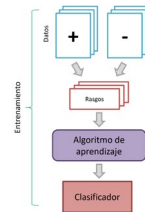
Submit

Refresh

<http://fakenews.research.sfu.ca/>



En resumen



VARIAS PIEZAS DEL PUZLE

- Educación
- Propagación
- Verificación
- Tecnología

- Un problema social
- La tecnología y la IA como soluciones a problemas sociales

LA APORTACIÓN DEL PLN

- Necesitamos datos de calidad, anotados por expertas
 - Nueva Ley de Servicios Digitales
 - Exige a las plataformas compartir datos
- El aprendizaje profundo no siempre funciona
- En situaciones de 'small data', métodos clásicos
 - Con rasgos lingüísticos

Maite Taboada
Simon Fraser University
mtaboada@sfu.ca

<http://www.sfu.ca/~mtaboada/>
<http://www.sfu.ca/discourse-lab>



Bibliografía

- Alonso, M. A., Vilares, D., Gómez-Rodríguez, C., y Vilares, J. (2021). Sentiment Analysis for Fake News Detection. *Electronics*, 10(11), 1348. <https://www.mdpi.com/2079-9292/10/11/1348>
- Asr, F. T., Mokhtari, M., y Taboada, M. (2023). Misinformation detection in news text: Automatic methods and data limitations. In S. M. Maci, M. Demata, M. McGlashan, & P. Seargeant (Eds.), *The Routledge Handbook of Discourse and Disinformation* (pp. 79-102). Routledge.
- Asr, F. T. y Taboada, M. (2018). The data challenge in misinformation detection: Source reputation vs. content veracity. *Proceedings of the 1st Workshop on Fact Extraction and Verification (FEVER), Conference on Empirical Methods in Natural Language Processing* (pp. 10-15). Brussels.
- Asr, F. T. y Taboada, M. (2019). Big data and quality data for fake news and misinformation detection. *Big Data & Society*, (January-June), 1-4.
- Biber, D. (1995). *Dimensions of Register Variation: A Cross-Linguistic Comparison*. Cambridge University Press.
- Cook, J. y Lewandowsky, S. (2011). *The Debunking Handbook*. <http://sks.to/debunk>
- Lewandowski, L. J., Coddington, R. S., Kleinmann, A. E., y Tucker, K. L. (2003). Assessment of reading rate in postsecondary students. *Journal of Psychoeducational Assessment*, 21(2), 134-144.
- Lewandowsky, S. y al., e. (2020). *The Debunking Handbook*. <http://sks.to/debunk>
- Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., y Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, 13(3), 106-131.
- Phillips, W. (2018). The oxygen of amplification. *Data & Society*, 22, 1-128.
- Taboada, M. (2016). Sentiment analysis: An overview from linguistics. *Annual Review of Linguistics*, 2, 325-347.
- Vosoughi, S., Roy, D., y Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146-1151.
- Wardle, C. (2018). *Information disorder: The essential glossary* [Report]. https://firstdraftnews.org/wp-content/uploads/2018/07/infoDisorder_glossary.pdf