



TECNOLÓGICO NACIONAL DE MÉXICO

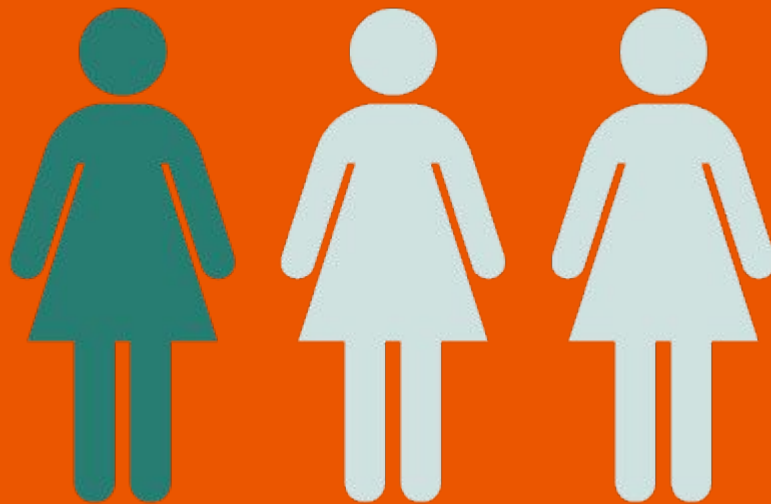
Detección de misoginia en redes sociales de habla hispana

Ángel Rodríguez Ulloa
Asesor de Tesis: Dr. Arnoldo Diaz-Ramirez

Viernes 17 de Febrero de 2023



Introducción



1 de cada 3 mujeres mayores a 15 años ha sido víctima de violencia física y/o sexual.



Violence Contra la Mujer (VCM)

- La mayoría de la VCM es cometida por una pareja íntima.
- **38%** de todos los asesinatos de mujeres son cometidos por una pareja íntima.
- Menos del **40%** de las mujeres que han sufrido violencia buscan ayuda.
- Solo $\frac{3}{4}$ de los países mundialmente tienen leyes para proteger a las víctimas de abuso doméstico.



Consecuencias de VCM

VCM tiene consecuencias de la salud:

- **42%** de las víctimas presentan lesiones.
- **16%** más probable tener un bebé con bajo peso.
- Estrés postraumático, depresión, ansiedad, abuso de sustancias.
- Inhabilidad para trabajar y aislamiento.
- Muerte.



VCM en Latinoamérica

- México reportó **854** feminicidios en el año 2022.
- **98%** de los feminicidios en latinoamérica no son penalizados.
- **81%** de los casos globales de VCM ocurren en latinoamérica.



https://drive.google.com/file/d/1jvGGra31Q361fOuNChetkBu0pva_MGxF/view

<https://reliefweb.int/report/brazil/double-pandemic-gender-based-violence-latin-america-and-early-experience-women-during>

<https://www.gob.mx/conavim/articulos/que-es-el-feminicidio-y-como-identificarlo?idiom=es>



Misoginia

- Misoginia significa odio hacia la mujer
- Es un problema sistémico, **NO** aislado.
- La misoginia es una causa principal de la VCM.



Misoginia en Medios Sociales


La misoginia en medios sociales tiene muchas caras:

- Lenguaje de odio que fomenta inequidades de género.
- Acoso
- Venganza porno
- Amenazas de violencia física o sexual
- Chistes sexistas



Misoginia en Medios Sociales

- $\frac{2}{3}$ de las mujeres reporteras han sido víctimas de amenazas, acoso sexual y otras formas de abuso en medios sociales.
- **38%** de las mujeres en EEUU y el **25%** de las mujeres en México han sido víctimas de acoso en línea.
- Más del **60%** de las víctimas de acoso sienten tristeza, molestia u otra inquietud por acoso en línea.



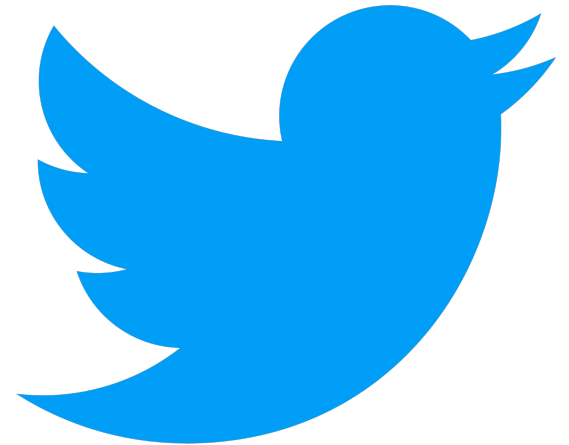
La misoginia en medios sociales es un indicador de VCM en la vida real.

- Áreas con mayores porcentajes de tweets misóginos también tienen mayores porcentajes de violencia doméstica y familiar.
- Areas con mayores porcentajes de tweets misóginos tienen mayor porcentaje de reportes de violencia sexual.



Twitter

- **#7** en medios sociales más importantes mundialmente
- Más de **500 millones** de tweets publicados al día
- **30%** de la audiencia de Twitter son mujeres



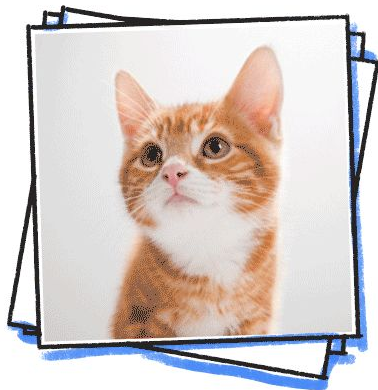
**En los últimos 5 años ha
crecido el interés en crear
modelos de aprendizaje
máquina para detectar tweets
misóginos en varios lenguajes.**



Aprendizaje Máquina (*Machine Learning*)

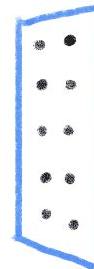
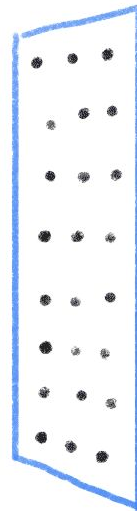
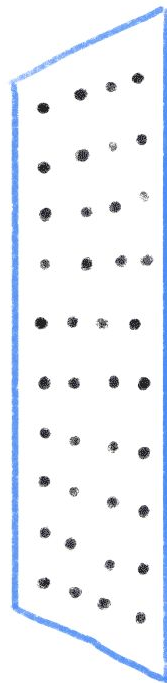
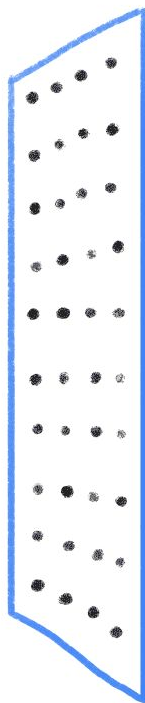
- Algoritmos matemáticos que le “enseñan” a una computadora a completar tareas sin usar métodos basados en reglas explícitas.
- Los modelos se entrenan al exponerlos a una gran cantidad de datos pertinentes a la tarea que buscas completar.

CAT



(LAELED)
PHOTOS

DOG



OUTPUT



Procesamiento de Lenguaje Natural

- Tecnologías que se usan para ayudar a las computadoras a “entender” el lenguaje humano.
- E.j., “Hey Google”, Siri, o Alexa

Nuestro modelo para la detección de lenguaje misógino en tweets en español



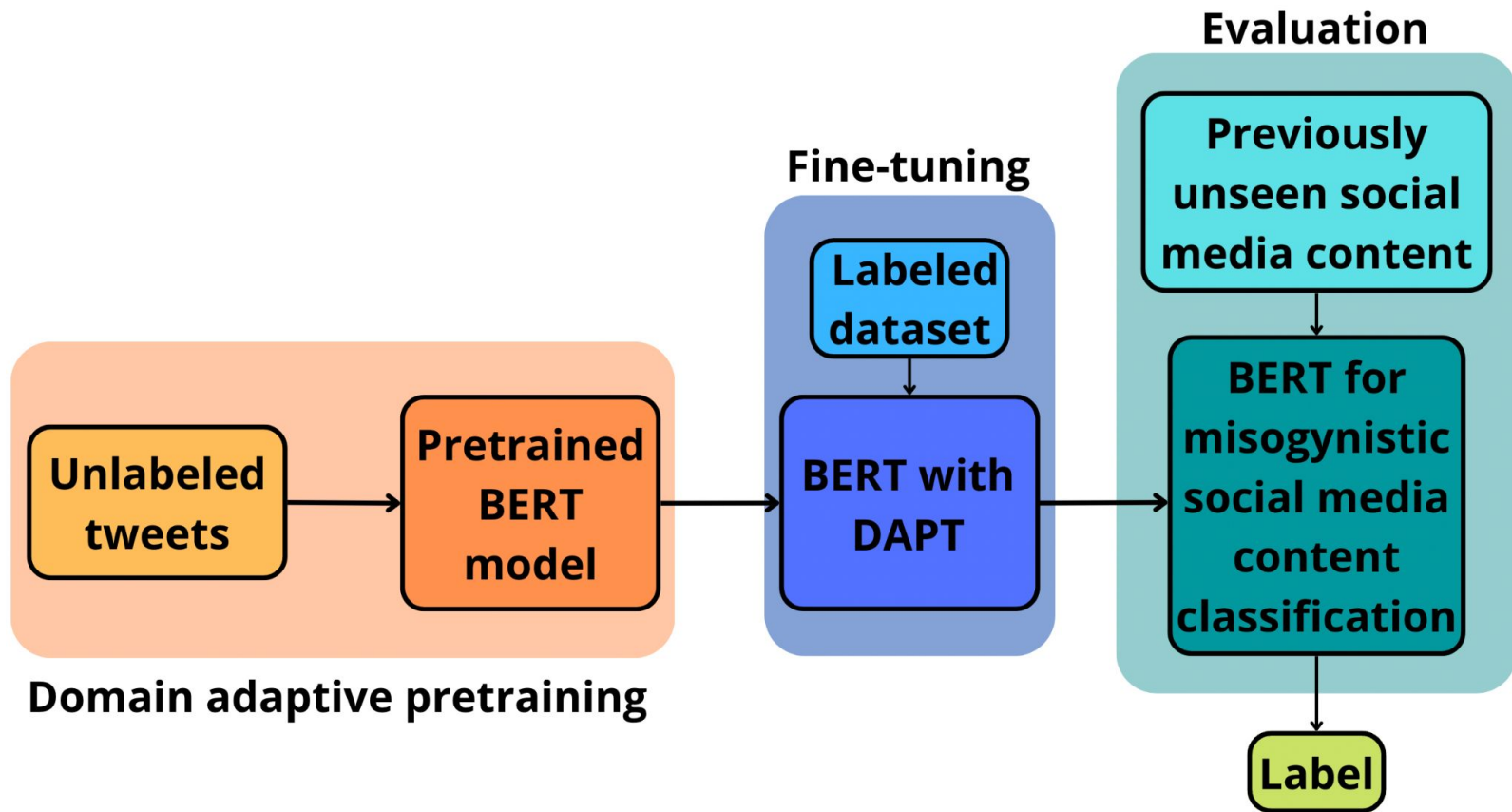
Nuestro Modelo

- Arquitectura existente de aprendizaje máquina llamada **BETO**, pre-entrenada para entender el idioma Español
- Entrenada por nosotros dos veces:
 - 5 millones de tweets en español no etiquetadas
 - 3 mil tweets etiquetadas como misóginas o no misóginas



Datos de entrenamiento

- 5.5 millones de tweets en Español.
 - Los tweets contienen profanidades populares.
- Conjunto de datos IberEval
 - 3 mil tweets etiquetadas como misóginas o no misóginas
 - Este conjunto de datos es popular, se ha usado en casi todos los proyectos de investigación para la detección de misoginia desde 2018.



Arquitectura para detección de contenido de Twitter misógino en Español.

Resultados Iniciales



Resultados

Model	Acc.	Pr.	Re.	F-1
García et al. [23]	83.30	N/A	N/A	83.30
Baseline	79.90	73.10	85.04	78.62
BETO _{profane}	84.60	79.64	86.70	83.02
BETO _{covid}	79.90	73.10	85.04	78.62

Table 5: Performance of our proposed models for misogyny identification on the IberEval dataset.

Descubrimiento de Sesgo



Sesgo

Un sistema tiene sesgo cuando las dos condiciones son verdad:

- Los resultados discriminan hacia un grupo.
- La discriminación es sistémica.

Mislabeled Text

@ClaudiaLopez como no apreciar un escote profundo un elegante vestido o una minifalda encantadora esto no es acoso hace parte de la magia mujer hombre si una persona se siente acosada al primer instante debe manifestarlo no hacer un juego y luego un escán

@N/A el acoso en la mujer es tan solo cuando el hombre no le gusta, si el hombre le gustara, no fuera acoso!!!

@bbcmundo lamentablemente ahora a todo le quieren llamar acoso, eso provocará radicalismo y falta de atención a casos reales que requerirán intervención penal, no caigamos en el juego de feminazis

*@N/A @elespectador se puede llamar acoso cuando algunas mujeres usan minifaldas, escotes, descaderados, transparencias etc, para provocarnos ?
mirar n buen trasero o un buen escote es acoso??*

Table 6: Misogynistic tweets that were labelled as non-misogynistic by BETO_{profane}.



Bias

- Analizamos el conjunto de datos de IberEval y descubrimos **469** tweets mal etiquetados (**14%** de los datos).
- Había un tweet que ni estaba en Español: *Hoxe meu abuelo vaime ensenar a usar a escopeta e estou superemocionada a mĩ na nai non lle fai puta gracia pero me da iw´a.*

Label and number missed	Example
Derailing (18)	@N/A <i>desnudas provocais al igual que un hombre, simplemente porque es un instinto humano.</i>
Stereotype and objectification (10)	<i>-¿porque los hombres se casan con las mujeres? -porque las cabras no saben fregar los platos.</i>
Dominance (24)	@N/A <i>un RT para difundir SEMEN! Movimiento machista internacional. Únete como activista!!</i>
Sexual harassment and threats of violence (6)	@N/A <i>Comemela mientras te meto los dedos</i>
Discredit (22)	@IrantzuVarela <i>A mi lo que digas tú me importa una puta mierda la verdad.</i>

Table 8: Misogynistic tweets wrongfully labeled as non-misogynistic on the IberEval training dataset.

Label and number missed	Example
Aimed at men (157)	<i>@LopezDoriga Cállate puta!</i>
Feminist discourse (66)	<i>@N/A @El_Pais Porque no tiene la menor empatía con víctimas de acoso ni de violación y porque defiende un derecho autoinventado del hombre de hacer con las mujeres lo que le dé la gana.</i>
Female dogs (28)	<i>Porque ladra tanto mi perra? La puta madre cállate un poco</i>
Clearly sarcastic (89)	<i>feliz año zorra, te amo imbécil :(@N/A</i>
Other (49)	<i>Otra puta vez enseñando prioridades.</i>

Table 7: Non-misogynistic tweets wrongfully labeled as misogynistic on the IberEval training dataset.



El conjunto de datos IberEval tiene un sesgo misógino

- Los tweets que fueron mal etiquetados como no misóginos reflejan las creencias misóginas de la sociedad.
- Los tweets que fueron mal etiquetados como misóginos reflejan una falta de atención o descuido hacia el tema de VCM.



Conclusiones

- Las personas que etiquetaron el conjunto de datos IberEval no tenían el entendimiento necesario sobre la misoginia para etiquetarlo.
- Hay que asegurarnos que nuestros datos de entrenamiento no discriminen a las poblaciones que queremos ayudar.
- La creación de modelos para detección de misoginia puede mejorar con un enfoque interdisciplinario.

Resultados de nuestro modelo después de remover el sesgo



Remover el sesgo

- Arreglamos los tweets mal etiquetados
- Quitamos el tweet que no estaba en español



Resultados finales

Model	Acc.	Pr.	Re.	F-1
García et al. [23]	83.30	N/A	N/A	83.30
BETO _{profane}	84.60	79.64	86.70	83.02
Debiased BETO _{profane}	84.35	84.64	83.93	84.28

Table 10: Performance comparison of BETO_{profane} before and after our IberEval dataset debiasing.

Conclusion



Conclusión

- Creamos un modelo para la detección de contenido misógino de habla hispana en Twitter
- Nuestro modelo supera el estado de arte.
- Descubrimos un sesgo misógino en el conjunto de datos IberEval. Este conjunto de datos es el que más se utiliza para entrenar modelos de detección de misoginia.



Conclusión

- Continuamos brindando atención a un problema grave: La violencia contra la mujer.
- Hay que pensar de manera crítica sobre cómo tratamos a las mujeres a nuestro alrededor.
- Necesitamos hacer cambios en nuestro comportamiento para asegurarnos de que las mujeres tengan espacios seguros, respetuosos y equitativos.
- **La violencia contra la mujer es prevenible pero tenemos que hacer el trabajo.**

Mis contribuciones como estudiante de maestría: 2020-2023

Domain-Adaptive Pretraining on a BERT Model for the Automatic Detection of Misogynistic Tweets in Spanish

Dalia A. Rodríguez¹, Julia Díaz-Escobar¹, Arnoldo Díaz-Ramírez^{1*} and Leonardo Trujillo²

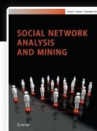
^{1*}Department of Computer Systems, Tecnológico Nacional de México/IT Mexicali, Av Tecnológico s/n, Mexicali, 21376, Baja California, México.

²Department of Electrics and Electronics, Tecnológico Nacional de México/IT Tijuana, Blvd. Industrial s/n, Tijuana, 22430, Baja California, México.

*Corresponding author(s). E-mail(s): adiaz@itmexicali.edu.mx;
Contributing authors: drodriguez@itmexicali.edu.mx;
jdiaz@itmexicali.edu.mx; leonardo.trujillo.ttl@gmail.com;

Abstract

Violence against women is a major social issue. One out of every three women worldwide have been subjected to physical or sexual violence. The pervasiveness of violence against women in the physical world and



Social Network Analysis and Mining

[Editorial board](#) [Aims & scope](#) [Journal updates](#)

Social Network Analysis and Mining (SNAM) is a multidisciplinary journal serving researchers and practitioners in academia and industry. It is the main venue for a wide range of researchers and readers from computer science, network science, social sciences, mathematical sciences, medical and biological sciences, financial, management and political sciences. We solicit experimental and theoretical work on social network analysis and mining using a wide range of techniques from social sciences, mathematics, statistics, physics, network science and computer science. — [show all](#)

Received June 21, 2021, accepted August 2, 2021, date of publication August 9, 2021, date of current version August 24, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3103459

A Systematic Review of Computer Science Solutions for Addressing Violence Against Women and Children

DALIA ANDREA RODRÍGUEZ¹, **ARNOLDO DÍAZ-RAMÍREZ¹**, (Member, IEEE),
JESÚS ELÍAS MIRANDA-VEGA¹, **LEONARDO TRUJILLO²**, AND **PEDRO MEJÍA-ALVAREZ³**


¹Department of Computer Systems, Tecnológico Nacional de México/IT Mexicali, Mexicali, Baja California 21376, Mexico

²Department of Electrics and Electronics, Tecnológico Nacional de México/IT Tijuana, Tijuana, Baja California 22430, Mexico

³Cinvestav Guadalajara, Zapopan, Jalisco 45017, Mexico

Corresponding author: Arnoldo Díaz-Ramírez (adiaz@itmexicali.edu.mx)

This work was supported in part by the Tecnológico Nacional de México and in part by the National Council of Science and Technology (CONACyT).



Detection and prevention of gender violence with computer science tools

Detección y prevención de violencia de género con herramientas de las ciencias computacionales

Dalia Andrea Rodríguez
Arnoldo Díaz-Ramírez
Rafael Iván Ayala Figueroa

Tecnológico Nacional de México / IT de Mexicali
November 17, 2021

A Systematic Review of Computer Science Solutions for Addressing Violence Against Women and Children

Dalia Andrea Rodríguez
Arnoldo Díaz-Ramírez
Rafael Iván Ayala Figueroa
Leonardo Trujillo
Juan P. García Vázquez

Neo 2021
September 10, 2021



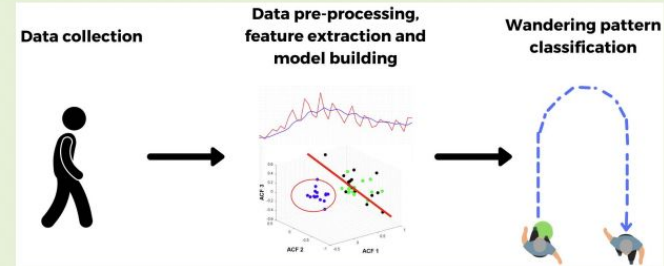
Time Series Data Processing for Classifying Wandering Patterns in People With Dementia

Arnoldo Díaz-Ramírez¹, Senior Member, IEEE, Jesús E. Miranda-Vega², Daniel Ramos-Rivera³,
Dalia Andrea Rodríguez⁴, Wendy Flores-Fuentes⁵, and Oleg Sergiyenko⁶

Abstract—As the population ages, dementia has become one of the main health issues worldwide affecting the elderly. It is a disease related to the damage of the brain cells, causing memory loss, impairing of written and spoken communication skills, difficulty in performing previously routine tasks, as well as personality and mood changes. There is no cure for dementia, but if diagnosed correctly, providing the proper treatment and support allows for a better quality of life for those affected by this disease. People with dementia tend to wander, and a relationship between the wandering pattern and the level of dementia has been established. In this paper, two-

time series techniques, the autocorrelation function and the partial autocorrelation function used in conjunction with the machine learning algorithms, including linear discriminant analysis, multivariate Gaussian, adaptive boost, and k-nearest neighbors, were evaluated to classify wandering patterns in people affected by dementia. The main contribution of this work is the use of time-series data techniques and machine learning algorithms to classify wandering patterns. The use of smoothing filters and time series feature extraction techniques, used in combination with ML algorithms, showed a very good performance, with an accuracy greater than 90%.

Index Terms—Dementia, wandering patterns, time series data processing, machine learning.



¡Gracias! :-)