

---

# INSTITUTO TECNOLÓGICO SUPERIOR DE TEZIUTLÁN

---

## Tesis



“Desarrollo de un sistema de software basado en Transformers para la  
detección de lenguaje de odio en medios sociales en español”

PRESENTA:

**ANGEL OSWALDO VÁZQUEZ BENITO**

CON NÚMERO DE CONTROL

**21TE0021P**

PARA OBTENER EL GRADO ACADÉMICO DE:

**MAESTRÍA EN SISTEMAS COMPUTACIONALES**

CLAVE DEL PROGRAMA ACADÉMICO

**MPSCO-0127**

DIRECTOR (A) DE TESIS:

**Dr. Mario Andrés Paredes Valverde**

TEZIUTLÁN, PUEBLA, JULIO 2023

“La Juventud de hoy, Tecnología del Mañana”

## **AGRADECIMIENTOS**

Quiero expresar mi infinito agradecimiento a mi director y asesor de tesis el Dr. Mario Andrés Paredes Valverde quien en todo momento mostró apertura a resolver las dudas e inquietudes que durante el proceso fueron surgiendo, así como el tiempo y paciencia que dedicó a poder atenderme y reconocer su compromiso para con este trabajo.

Agradezco al Consejo Nacional de Humanidades Ciencia y Tecnología (CONAHCYT) por la oportunidad de crecer como profesionista y el apoyo al desarrollo de este proyecto que me deja mil aprendizajes de forma académica que seguramente se verán reflejados en un futuro.

También quiero reconocer el apoyo brindado por parte del sistema DIF municipal de Chignautla, Puebla que desde el primer momento creyó en este proyecto brindando asesoría psicológica para el enriquecimiento y corrección de los datos que son vitales en los resultados de este proyecto, especialmente a la Ing. María del Carmen Castro Abad presidenta de dicha organización por su atención, tiempo y disposición.

Finalmente agradezco a cada uno de los docentes que fueron parte de esta formación académica cada uno aportando su granito de arena en distintas áreas me llevo de cada uno grandes aprendizajes.

## **DEDICATORIA PERSONAL**

Para mi madre que siempre ha sido ejemplo de esfuerzo y perseverancia porque en esos momentos bajos ella me recuerda lo capaz que soy con su cariño, comprensión, paciencia y amor.

Para mis abuelos que me demuestran que tanto en el cielo como en la tierra siempre habrá alguien acompañándome en cada momento de mi vida y que sin ellos no sería posible llegar hasta este punto.

A mi padre y hermanos por su apoyo en todo momento, comprensión y por alentarme a continuar con mis metas.

A mis amigas Guadalupe Bandala y Anai Ortiz quienes han sido un sustento y apoyo sincero y que me han brindado su apoyo incondicional.

Al profesor Carlos Ramírez por su ánimo, consejos y apoyo en momentos de duda y que me acompañó en este viaje de tesis desde el momento uno.

# ÍNDICE GENERAL

RESUMEN .....	10
ABSTRACT .....	11
INTRODUCCIÓN.....	12
CAPÍTULO I GENERALIDADES DEL PROYECTO.....	13
Marco teórico.....	13
1.1.1 Redes sociales .....	13
1.1.2 Conductas de odio en redes sociales .....	14
1.1.3 Misoginia .....	15
1.1.4 Misoginia y su relación con las redes sociales.....	16
1.1.5 Web Scraping .....	17
1.1.6 Procesamiento del Lenguaje Natural (PLN) .....	18
1.1.7 Aprendizaje automático .....	18
1.1.8 Transformers .....	19
1.1.9 Metodología CRISP-DM.....	21
1.1.10 Python .....	25
Planteamiento del problema.....	27
Justificación .....	28
Hipótesis.....	29
Objetivo general.....	29
Objetivos específicos .....	30
Alcances y limitaciones .....	30
1.1.11 Alcances.....	30
1.1.12 Limitaciones .....	31

CAPÍTULO II	ESTADO DEL ARTE .....	32
	Trabajos relacionados.....	32
	Análisis comparativo de los trabajos relacionados.....	35
CAPÍTULO III	METODOLOGÍA Y DESARROLLO .....	39
	Solución propuesta.....	39
	3.1.1 Conjunto de datos.....	40
	3.1.2 Módulo de recolección de datos .....	40
	3.1.3 Modelo de detección de lenguaje misógino .....	40
	3.1.4 Servicio Web.....	40
	3.1.5 Aplicación Web .....	41
	Comprensión del negocio.....	41
	Comprensión de los datos.....	42
	3.1.6 Identificación de comentarios objetivos .....	42
	3.1.7 Propuesta de fuentes de extracción de datos .....	44
	3.1.8 Recopilación de datos.....	44
	3.1.9 Preparación de los datos .....	45
	3.1.10 Modelado y evaluación .....	52
	3.1.11 Despliegue .....	53
CAPÍTULO IV	RESULTADOS .....	61
CAPÍTULO V	CONCLUSIONES.....	72
	Conclusiones.....	72
	Recomendaciones .....	72
	5.1.1 Enriquecimiento del conjunto de datos .....	72

5.1.2	Análisis automático en publicaciones de redes sociales .....	73
REFERENCIAS .....		74
ANEXOS I .....		76
5.1.3	Maqueta de Aplicación Web .....	76
5.1.4	Extracto de conjunto de datos de entrada para el entrenamiento del modelo	77

## ÍNDICE DE FIGURAS

Figura 1 Redes Sociales Actuales.....	13
Figura 2 Interacción con redes sociales.....	<b>iError! Marcador no definido.</b>
Figura 3 Proceso de una red neuronal con transformers. ....	21
Figura 4 Ciclo de vida de la metodología CRISP-DM.....	22
Figura 5 Logo de Python.....	<b>iError! Marcador no definido.</b>
Figura 6 Arquitectura de solución propuesta.....	39
Figura 7 Tecnología empleada para el servicio WEB (Backend)	<b>iError! Marcador no definido.</b>
Figura 8 Tecnología empleada para la aplicación WEB (Frontend)	<b>iError! Marcador no definido.</b>
Figura 9 Ejemplo de comentario objetivo obtenido de Facebook.....	43
Figura 10 Ejemplo de comentario objetivo obtenido de Facebook.....	43
Figura 11 Ejemplo de comentario objetivo obtenido de Facebook.....	44
Figura 12 Logo de Twitter y Facebook. ....	<b>iError! Marcador no definido.</b>
Figura 13 Resultados de entrenamiento del modelo.....	52
Figura 14 Promedios de precisión y F1 del modelo obtenido.....	53
Figura 15 longitud de caracteres en los elementos del conjunto de datos.....	61
Figura 16 longitud de palabras en los elementos del conjunto de datos. ....	62
Figura 17 top 10 de palabras recurrentes en el conjunto de datos. ....	63
Figura 18 top 25 de palabras recurrentes en el conjunto de datos. ....	64
Figura 19 Nube de etiquetas de palabras recurrentes .....	64
Figura 20 Nube de etiquetas de palabras recurrentes 2.....	65
Figura 21 Interfaz principal de aplicación WEB .....	65

Figura 22 Analisis de texto en aplicación WEB 1 .....	66
Figura 23 Analisis de texto en aplicación WEB 2 .....	67
Figura 24 Analisis de texto en aplicación WEB 3 .....	67
Figura 25 Analisis de texto en aplicación WEB 4 .....	68
Figura 26 Analisis de texto en aplicación WEB 5 .....	68
Figura 27 Analisis de texto en aplicación WEB 6 .....	69
Figura 28 Analisis de texto en aplicación WEB 7 .....	69
Figura 29 Resultado del análisis y clasificación de texto .....	70
Figura 30 Representación de porcentaje sin contenido misógino .....	71



## ÍNDICE DE TABLAS

Tabla 1 Tabla comparativa del estado del arte .....	35
Tabla 2 Lista de personas identificadas para extracción de datos.....	42
Tabla 3 Ejemplos de textos extraídos.....	51
Tabla 4 Caso de clasificación de texto 1 .....	53
Tabla 5 Caso de clasificación de texto 2.....	54
Tabla 6 Caso de clasificación de texto 3.....	55
Tabla 7 Caso de clasificación de texto 4.....	56
Tabla 8 Caso de clasificación de texto 5.....	57
Tabla 9 Caso de clasificación de texto 6.....	59
Tabla 10 Caso de clasificación de texto 7.....	60

## **RESUMEN**

Las redes sociales se han convertido en un medio de comunicación global, sin embargo, a medida que el contenido en línea continúa creciendo, también lo hace la difusión del discurso de odio, el cual puede entenderse como un ataque directo a las personas en función de su raza, etnia, origen nacional, afiliación religiosa, orientación sexual, casta, sexo, género, identidad de género y enfermedad o discapacidad grave. Por otro lado, los avances recientes en técnicas de aprendizaje profundo han revolucionado muchas aplicaciones de procesamiento de lenguaje natural, tal es el caso de los Transformers, que son modelos de aprendizaje profundo que adoptan el mecanismo de auto atención, ponderando diferencialmente el significado de cada parte de los datos de entrada. Dicho esto, el objetivo de este proyecto es desarrollar un sistema de software basado en Transformers para la detección de lenguaje de odio en medios sociales en español, concretamente, fuentes de datos de México, con el objetivo de contribuir al soporte a la toma de decisiones por parte de instituciones gubernamentales de seguridad ciudadana y demás organizaciones civiles enfocadas en proporcionar a la ciudadanía una vida libre de violencia.

## **ABSTRACT**

Social networks have become a global means of communication, however, as online content continues to grow, so does the spread of hate speech, which can be understood as a direct attack on people based on different characteristics like ethnicity, national origin, religious affiliation, sexual orientation, caste, sex, gender, gender identity, and serious illness or disability. On the other hand, recent advances in deep learning techniques have revolutionized many natural language processing applications, such is the case of Transformers, which are deep learning models that adopt the self-attention mechanism, differentially weighting the meaning of each word. part of the input data. The objective of this project is to develop a software system based on Transformers for the detection of hate language in social media in Spanish, specifically, data sources from Mexico, with the aim of contributing to decision-making support. by government citizen security institutions and other civil organizations focused on providing citizens with a life free of violence.

## **INTRODUCCIÓN**

Con el paso del tiempo la comunicación humana se ha adaptado a la evolución de las tecnologías. Actualmente, existe un sinnúmero de redes que nos permiten establecer comunicación desde cualquier parte del mundo, sin embargo, el objetivo inicial de estas redes se ha ido distorsionando al encontrar contenido que no solo fomenta una sana comunicación sino que también se puede identificar la difusión de ideas, lemas, actitudes e incluso conductas a través de Internet que menosprecian, atacan y humillan a otras personas. Este tipo de conductas suele ser dirigida a grupos concretos por motivos de raza, género, sexualidad o religión.

Las mujeres forman parte de estos grupos vulnerables es por ello que este trabajo aborda este tipo de conductas con carácter misógino recolectando datos reales de redes sociales para que a partir de ellos sea posible realizar un análisis que permita la creación de un modelo con el fin de identificar lenguaje de odio en texto basado en el contenido de los datos recolectados todo ello apoyado de una metodología que se adapta a las necesidades de proyectos que requieren el procesamiento de datos en volumen.

# CAPÍTULO I GENERALIDADES DEL PROYECTO

## Marco teórico

### 1.1.1 Redes sociales

Las redes sociales son aquellos medios que a través del uso de las tecnologías de la información nos permiten comunicarnos de forma no presencial con cualquier persona, en cualquier país a través de cualquier idioma, y es que las redes sociales no conocen de fronteras. En la actualidad las redes sociales vigentes como se observa en la Figura 1 evolucionan y seguirán evolucionando con la finalidad de poder ofrecer más herramientas que permitan la comunicación lo más parecida posible a un contacto presencial o incluso ofreciendo opciones que de la forma mencionada no puede ser llevada a cabo.

Red social	Tema y dirección web
Facebook	Interés general <a href="https://www.facebook.com/">https://www.facebook.com/</a>
YouTube	Sitio donde subir, alojar y compartir videos gratis. <a href="https://www.youtube.com/">https://www.youtube.com/</a>
Google+	Red social de Google. Atrayente por sus comunidades, por el servicio de subir y compartir fotos y por la autoridad que aporta a los blogs y sitios de Internet. <a href="https://plus.google.com/">https://plus.google.com/</a>
Reddit	Comunidad donde compartir enlaces a sitios web, servicios, noticias de actualidad y todo tipo de contenido de interés. <a href="https://www.reddit.com/">https://www.reddit.com/</a>
Twitter	Publicación instantánea de noticias y todo tipo de acontecimientos que pueden convertirse en virales en la red. Admite imágenes y videos. <a href="https://twitter.com/">https://twitter.com/</a>
Instagram	Aplicación, comunidad y red social para subir, editar y compartir fotos. Solo se pueden subir desde dispositivos móviles. <a href="https://www.instagram.com/">https://www.instagram.com/</a>
LinkedIn	Currículos personales e información de negocios. De interés para profesionales y todo el que busque trabajo y oportunidades laborales de valor. <a href="https://www.linkedin.com/">https://www.linkedin.com/</a>
Imgur	Sitio de hospedaje y publicación de imágenes. <a href="https://imgur.com">https://imgur.com</a>
Tumblr	Comunidad de microbloging. Se publican y comparten post cortos, generalmente con imágenes. <a href="https://www.tumblr.com/">https://www.tumblr.com/</a>
Pinterest	Sitio donde subir, mostrar, compartir con amigos imágenes y videos, organizadas en colecciones y categorías. <a href="https://www.pinterest.com/">https://www.pinterest.com/</a>
Vimeo	Red social donde subir y guardar videos para compartir con cualquier usuario. <a href="https://vimeo.com/">https://vimeo.com/</a>
deviantART	Inmensa comunidad artistica donde sus miembros

*Figura 1 Redes Sociales Actuales*

Las redes sociales corresponden al término que empezó a emplearse por antropólogos y sociólogos ingleses a mediados de los años 50 del siglo XX, para caracterizar y estudiar el conjunto de relaciones humanas que tienen un impacto duradero en la vida de un individuo y ofrecen unos patrones y colaboraciones contingentes para orquestar las soluciones en la vida cotidiana (Wolfe, 2011).

Acceder a una red social no tiene ninguna complejidad, y cada vez son más las acciones que se pueden ejecutar a través de esta. Una red social garantiza una comunicación sin importar la ubicación o idioma de los participantes, pero actualmente nada garantiza que esta comunicación se ejerza de forma sana.

### **1.1.2 Conductas de odio en redes sociales**

Internet se ha posicionado como una herramienta a través de la cual los seres humanos han basado su vida diaria por extremista que parezca, pero en ciertos casos las actividades que forman parte del día a día no se realizarían sin este medio.

Su uso va desde temas políticos, empresariales, sociales a simplemente solo diversión. Pero como todo aquello que tiene ventajas también tiene desventajas y es que este medio ha abierto una brecha intangible de la que muy poco se habla y que tanto se sufre día a día: las conductas de odio. Si bien antes la violencia solo se reconocía por causar algún daño físico hoy se sabe que va más allá y que incluso existen distintos tipos de conductas que incitan a cualquier variante en las conductas de odio, que aún no eran erradicadas en su totalidad si habían disminuido, como la discriminación, sin embargo, han tomado realce por el uso de redes sociales.

En los últimos años se detecta un aumento significativo de las reacciones xenófobas, racistas e intolerantes en general, peligrosamente encauzadas a través de grupos y partidos políticos de ultraderecha con una influencia creciente en la vida pública. Esas actitudes intolerantes, basadas en fuertes prejuicios hacia otras identidades culturales y dirigidas a excluir y marginar al «diferente», que se agrupan

generalmente bajo el concepto de «comportamientos de odio» (Copello & Rodríguez, 2021).

### **1.1.3 Misoginia**

Está por obviedad, que el llamar xenofóbico u homofóbico hacen referencia al odio o la aversión que tiene una raza con otra o la intolerancia a la libertad sexual y de género, en cualquier caso, se trata de aversiones dicotómicas, sin embargo, la misoginia, es un término que no contrasta polos opuestos.

Para tener una idea o bagaje más amplio sobre este término con relación al odio y violencia generados y que también incluye acciones de mujeres contra mujeres, debemos empezar esbozando y deconstruyendo el término en primera instancia.

Como muchos autores coinciden, el objetivo de la misoginia es la deshumanización de la mujer y esto se basa en la creencia de que la mujer es inferior al hombre, lo cual ha sido justificado desde diferentes perspectivas; religiosas, biológicas, seudocientíficas e incluso políticas donde muchas de ellas, rompen una barrera generacional.

Proveniente del griego antiguo μισόγυνος, compuesto por μισέω (miseo), "detestar, odiar", y del griego γυνή (gyné), "mujer", misoginia es un término que tiene orígenes al menos desde el siglo IV a.C., hace referencia al odio, rechazo, aversión por las mujeres solo por el hecho de ser mujer. Solía ser representada en comedias y tragedias griegas de esos tiempos.

Actualmente desde la perspectiva de género la misoginia se produce cuando un hombre atenta contra la dignidad, corporeidad y salud mental de una mujer, o al menos siempre se ha tenido la tendencia a transpolar este término, sin embargo, la misoginia no es únicamente lo anterior descrito, es también cualquier conducta de odio contra las mujeres, en donde no necesariamente el autor de dichas conductas debe ser un varón como en el mayor de los casos se tiende a generalizar.

Rita Segato habla ampliamente de ello y propone un argumento que se desarrolla en dos textos, género y colinealidad, y el papel del estado y de las leyes frente a la estructura y las transformaciones de la violencia de género. Para Segato los términos «criollo», «prejuicio», «homofobia» y «misoginia» son sinónimos, esto tomando en cuenta que el mundo criollo es un mundo letal para las mujeres. Y eso tiene que ver con el frente colonial. “No soy la única en decir algo así”, afirma la autora argumentando que esto también ha sido dicho por investigadoras francesas para el África.

Si en el genocidio la construcción retórica del odio al otro conduce la acción de su eliminación, en el feminicidio la misoginia por detrás del acto es un sentimiento más próximo al de los cazadores por su trofeo: se parece al desprecio por su vida o a la convicción de que el único valor de esa vida radica en su disponibilidad para la apropiación. Es por ello por lo que se muestra convencida de que la víctima es el desecho del proceso, una pieza descartable, y de que condicionamientos y exigencias extremas para atravesar el umbral de la pertenencia al grupo de pares. (Segato, 2016)

#### **1.1.4 Misoginia y su relación con las redes sociales**

En la actualidad el término no ha tornado su sentido original y se lucha cada día para erradicarlo y mantener la integridad de las mujeres. Internet ha funcionado como un medio para comunicar al mundo lo que las mujeres viven día con día, lo que pasa en un hemisferio es conocido en el otro gracias a la difusión en medios digitales principalmente en redes sociales, pues por años los medios tradicionales no mostraban nada de la realidad.

Pero si bien, así como ha servido para exhibir los maltratos, las injusticias, la violencia en contra del género femenino también ha sido parteaguas para expresar las desigualdades, para denigrar y agredir a las mujeres en una práctica globalizada



sin importar las consecuencias que esto pueda tener, todo esto en muchas ocasiones bajo el anonimato.

La misoginia hasta hace algunos años pasaba desapercibida en chistes, bromas, comentarios, etc., pues resultaban ser "sutiles" para la sociedad. No se dimensionaba el impacto que estos pudieran tener sobre el grupo afectado pero que con el uso de redes sociales se han visto expuestos. Actualmente el feminismo ha hecho frente a esta situación, protestando a través del ciberactivismo y es que esto se ha vuelto una necesidad puesto que en redes sociales el éxito de los *influencers* se basa en un relato despreocupado que se acompaña de un humor ofensivo, generalmente sexista cuando no misógino, basado en la burla, especialmente contra las mujeres, los personajes femeninos de videojuegos, e incluso sus propias seguidoras. Con todo, la crítica a estos contenidos y las acusaciones directas de machismo revierten, pese a todo, incrementando el éxito de estos creadores de contenidos, ídolos para los adolescentes y jóvenes de hoy en día (Cerva D., 2020).

### **1.1.5 Web Scraping**

A lo largo del tiempo se han desarrollado nuevas herramientas dentro de la investigación, muchas de ellas permiten la recolección automatizada de datos de la web con base en determinados patrones identificados por el analista de datos. Dentro de estas herramientas ha cobrado mucha popularidad el Web Scraping, la cual consiste en recolectar información en sitios de internet que a pesar de ser pública, no cuenta con una estructura clara ya que la información que muestra se encuentra dispersa o esta de forma masiva.

De forma clara una página web está compuesta por elementos (títulos, párrafos, imágenes, enlaces, estilos, tablas, entre otros) los cuales se identifican dentro de la estructura de la página por medio de etiquetas. Dichos elementos se definen tanto con una etiqueta de inicio y una de cierre y entre estas se encuentra el contenido en la mayoría de los casos texto que puede ser el objetivo de utilizar el web Scraping

“las posibilidades son amplias; dependen del problema particular y del ingenio que el analista de datos aplique para encontrar una solución” (López, 2018).

### **1.1.6 Procesamiento del Lenguaje Natural (PLN)**

La interacción entre la humanidad y las máquinas tecnológicas también se ha visto en la necesidad de poder establecer lazos de comunicación, pero para poder llevarla a cabo es importante que no solo el humano logre interpretar lo que observa en una computadora si no que la computadora también sea capaz de poder interpretar la forma de comunicación humana.

Por Procesamiento de Lenguaje Natural (PLN, denominado también NLP por sus siglas en inglés) se entiende la habilidad de la máquina para procesar la información comunicada, no simplemente las letras o los sonidos del lenguaje (Gelbukh, 2010).

### **1.1.7 Aprendizaje automático**

El aprendizaje automático o Machine Learning es una disciplina del campo de la Inteligencia Artificial que funciona a través de algoritmos, los cuales instruyen a los ordenadores el identificar patrones en datos masivos y elaborar predicciones (análisis predictivo). Este aprendizaje dota a las computadoras de características que les permite realizar funciones o actividades de forma autónoma, es decir, sin necesidad de ser programados.

Pese a que el término se utilizó por primera vez en 1959, es en los últimos años que ha ganado relevancia debido al aumento de la tecnología y de la capacidad computacional además del descubrimiento y explotación de los datos. Las técnicas de aprendizaje automático son una parte fundamental del Big Data.

#### ***Algoritmos de aprendizaje automático***

Aprendizaje supervisado: Este tipo de algoritmos cuentan con un aprendizaje previo basado en un sistema de etiquetas asociadas a unos datos que les permiten tomar

decisiones o hacer predicciones. Un claro ejemplo es un detector de spam que etiqueta un e-mail de forma binaria (como spam o no) dependiendo de los patrones que ha aprendido en el historial de correos.

### ***Aprendizaje no supervisado***

A diferencia del tipo anterior este tipo de algoritmos no cuentan con un conocimiento previo. Se enfrentan a los datos con el objetivo de poder encontrar patrones que permitan organizarlos. Por ejemplo, en el campo del marketing se utilizan para extraer patrones de datos masivos que provienen de las redes sociales y crear campañas de publicidad altamente segmentadas las cuales a partir de lo anterior pueden identificar preferencias en usuarios dando así un uso óptimo a los datos que ocupa como objeto de estudio.

### ***Aprendizaje por refuerzo***

El objetivo de este es que un algoritmo aprenda a partir de la propia experiencia. Esto es, que sea capaz de tomar la mejor decisión posible ante diferentes situaciones de acuerdo con un proceso de prueba y error en el que se recompensan las decisiones correctas. En la actualidad está siendo aplicado para posibilitar el reconocimiento facial, hacer diagnósticos médicos o clasificar secuencias de ADN (Iberdrola, 2022).

## **1.1.8 Transformers**

Actualmente existen algunas tecnologías que nos proporcionan herramientas o programas para poder procesar el lenguaje natural. Con el paso del tiempo estas herramientas cambian y también dan paso al surgimiento de algunas nuevas tal es el caso de los Transformers.

Hasta hace apenas unos años atrás las Redes Recurrentes dominaban el área para el procesamiento del lenguaje, pues el lenguaje humano es precisamente una secuencia de palabras. Y estas redes se especializan en procesar secuencias. Pero estas redes recurrentes presentaban un gran inconveniente que se relacionaba con tener una memoria de corto plazo.

Por ejemplo, para la generación de texto la red podrá producir frases relativamente cortas y bastante coherentes. Pero cuando la secuencia generada es extensa, la red no está en capacidad de tener como referencia el texto que apareció al inicio, y así el texto generado no será del todo coherente.

Los Transformers nacieron inicialmente como una alternativa al problema de la traducción de texto de un idioma a otro. Pero debido a que estas redes procesan de forma paralela en comparación con las Redes Recurrentes en donde se procesan uno a uno, las reglas del juego cambian para la comprensión del lenguaje natural (NLU, por sus siglas en inglés Natural Language Understanding), un subconjunto del procesamiento del lenguaje natural (NLP), que se ha convertido en uno de los pilares de la Inteligencia Artificial en una economía digital global (Rothman, 2021).

En este tipo de redes denominadas Transformers la totalidad de la secuencia de entrada es procesada en paralelo por la red, a diferencia de las Redes Recurrentes en donde se procesan uno a uno (es decir de forma serial) los elementos de la secuencia.

En dicho proceso representado en la Figura 3 esta secuencia es inicialmente convertida en una representación numérica usando un embedding, posteriormente se añade una codificación de posición y los vectores resultantes ingresan a la etapa de codificación, que se encarga de extraer la información más relevante de la secuencia para finalmente ser conectada al decodificador para así mostrar la salida o resultado.

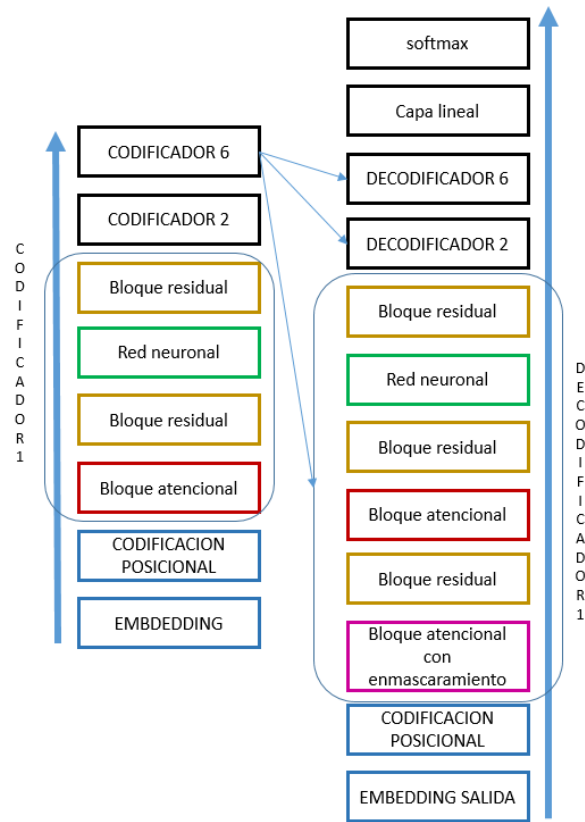


Figura 2 Proceso de una red neuronal con transformers.

Un ejemplo de esto es el transformer de transferencia de texto a texto del método PYTHON, que está capacitado para traducir entre todos los pares de combinaciones de características del método PYTHON: un modelo único que puede predecir métodos completos a partir de cadenas de texto en lenguaje natural (cadenas de documentación) y resumir código en cadenas de documentación de cualquier estilo común (Colin, 2020).

### 1.1.9 Metodología CRISP-DM

Esta metodología puede ser considerada como reglamentaria cuando hablar de proyectos que van enfocados a la extracción y procesamiento de datos se refiere.

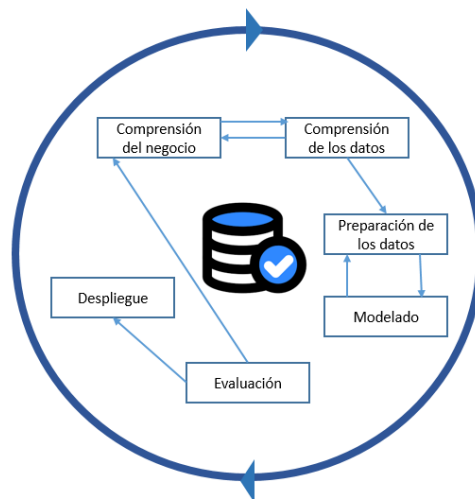
Durante 20 años la metodología CRISP-DM ha sido fuente de inspiración de otros estándares como SEMMA de SAS o ASUM-DM de IBM, así como ha dado lugar a

múltiples variantes que amplían o particularizan CRISP-DM a una industria o tipo de proyecto.

Usar metodologías como CRISP-DM en proyectos que involucran el análisis de datos asegura calidad en los datos con los que se trabaja, lo cual conlleva a obtener resultados acertados al aportar un marco conceptual que ayuda en diferentes aspectos como coordinar las actividades marcando un proceso a seguir, ayuda a gestionar expectativas ya que permite predecir pasos a futuro.

### ***Fases de la metodología CRISP-DM***

La metodología CRISP-DM se conceptualiza en 6 fases que se muestran en la Figura 4.



*Figura 3 Ciclo de vida de la metodología CRISP-DM.*

### ***Comprensión del negocio***

En esta fase, la comprensión de los objetivos y exigencias que abarca el proyecto es el punto primordial. Después, ese entendimiento de objetivos es transportado a definir un problema de minería de datos y en establecer un plan adecuado para el alcance de los objetivos planteados. Es importante no establecer objetivos

inalcanzables y marcar criterios de evaluación tan realistas como sea posible todo ello considerando y previendo la cantidad de datos que se marcaran como objetivo.

### ***Comprensión de los datos***

Es dentro de la comprensión de los datos que es necesario realizar la recolección inicial de estos y comenzar a familiarizarse con ellos a través de la exploración de estos. En esta exploración es necesario identificar la distribución de los atributos clave, en las relaciones entre subconjuntos pequeños de los atributos o en las propiedades que existen dentro del conjunto total de datos.

También es necesario realizar una caracterización general de los datos recolectados y esta actividad puede apoyarse de cuestionamientos como:

- ¿Qué formato tienen los datos?
- ¿Cuál es el número de registros?
- ¿Cuántos campos tiene un registro?
- ¿Las características son suficientes para satisfacer los requerimientos?

### ***Preparación de los datos***

La fase de preparación de datos comprende las actividades que permiten construir el conjunto de datos final, algunas de las actividades en esta etapa son:

#### *Selección de los datos:*

Se determina que datos se emplearán en el análisis usando los criterios que se definen para catalogar su relevancia y calidad.

#### *Limpieza de datos:*

Esta actividad es de suma importancia porque de ella dependerá la calidad de los datos que se usarán como valores de entrada en el proceso de análisis y modelado.

#### *Construcción de nuevas variables:*

A partir de los datos originalmente capturados, se generan atributos derivados, nuevos registros o valores transformados de atributos existentes, en función de los requerimientos para preparar la entrada a las herramientas de modelado.

#### Integración de datos:

Se realizan agregaciones que resumen información contenida en varios registros.

#### Integración de los datos y su formato:

Estas transformaciones se refieren a modificaciones sintácticas que se hacen sobre los datos, estas alteraciones no deben modificar el significado, pero son requeridas para la etapa de modelado.

### **Modelado**

Es dentro de esta fase que se aplican técnicas de modelado que también comprenden herramientas de minería de datos. Estas actividades utilizan como parámetros de entrada los datos que han sido tratados en la fase anterior (preparación de los datos) debido a ello es posible que se tenga que regresar a la etapa anterior y realizar ajustes o calibrar los datos para obtener resultados óptimos.

Es por ello por lo que esta fase se compone de actividades como:

- Selección de la técnica de modelado.
- Diseño de prueba.
- Construcción del modelo.

### **Evaluación**

En esta etapa se realiza la evaluación de los modelos obtenidos con el fin de poder determinar si responden a las necesidades planteadas en las etapas anteriores.

En el momento que se identifica que se ha obtenido un modelo que presenta un nivel mayor de calidad desde la perspectiva del análisis de datos es necesario revisar



los pasos seguidos para tener la confianza de que el modelo cumple con su tarea y poder pasar a la etapa de despliegue.

### ***Despliegue***

Durante esta etapa se realiza la explotación de los modelos dentro del centro productivo. Esto no es necesariamente el final del flujo puesto que al no cumplir con los estándares necesarios es posible que se requiera rehacer el modelo, para esta etapa el monitoreo es vital (Smartup, 2019.)

#### **1.1.10 Python**

Python tuvo sus inicios a principios de los años 90, cuando a Guido Van Rossum, quien fungía como trabajador del Centrum Wiskunde & Informatica (CWI), un centro de investigación ubicado en Holanda tuvo la iniciativa de desarrollar un nuevo lenguaje retomando principios del lenguaje 'ABC' el cual también fue desarrollado por Guido en colaboración con sus compañeros. El principal objetivo era crear un lenguaje de programación cuyo aprendizaje para el usuario en cuanto a escritura y entendimiento fuera muy fácil de aprender.

Actualmente Python es un lenguaje de programación de alto nivel para desarrollar diferentes tipos de aplicaciones, además de ser un lenguaje multiplataforma de código abierto que permite desarrollar software sin limitación alguna. Con el paso del tiempo, Python ha tomado un papel importante en diferentes proyectos debido a la gran gama de opciones y posibilidades que presenta ya que impulsa facilidades en ámbitos como inteligencia artificial, big data, aprendizaje automático y ciencia de datos, además de otras ramas que se encuentran en tendencia.

Python es un lenguaje de programación sumamente poderoso ofreciendo así una infinidad de herramientas que pueden ser de apoyo para diferentes proyectos adecuándose a sus objetivos.

### ***Analítica de datos y big data***

Estos dos términos son temas de interés actual en el ámbito tecnológico por lo cual Python agrega simplicidad y una opción de bibliotecas de datos que ayudan al procesamiento de datos es por ello por lo que Python resulta ideal para el análisis y gestión de datos en gran volumen y en tiempo real.

Actualmente es una herramienta recurrentemente utilizada por diferentes empresas y aunque en ocasiones no es de forma directa es Python quien está detrás de muchas herramientas de análisis.

### ***Minería de datos***

Para la minería de datos Python puede dar resultados muy útiles a través de la limpieza y organización de datos usando algoritmos de aprendizaje automático para simplificar este tipo de procesos todo esto con el objetivo de predecir futuras tendencias.

### ***Ciencia de datos***

Python está abarcando más territorio incluso desbancando a MATLAB el cual es un lenguaje utilizado por científicos para el procesamiento de datos en volumen, esto debido a la sencillez y gran potencia que tiene para trabajar con un gran número de datos ofreciendo diferentes bibliotecas que ayudan a la realización de este tipo de tareas.

### ***Inteligencia artificial***

La inteligencia artificial (IA) ha sido una rama de la ciencia que ha tenido una gran evolución en los últimos años. Siendo un lenguaje robusto, pero al mismo tiempo sencillo de escribir lo hace el candidato perfecto para encajar con la IA. Y es que poder plasmar ideas que resultan complejas en un fragmento de código pequeño ha

hecho que sea uno de los lenguajes que están directamente asociados con el impulso de la IA.

## **Planteamiento del problema**

Si bien la comunicación entre humanos es algo que forma parte de sus necesidades básicas al igual que todo ha tenido que evolucionar según las necesidades de la sociedad, las tecnologías de la información también han sido adaptadas para este ámbito. A mediados del siglo XX los medios de comunicación apoyados por tecnología han ido evolucionado de forma sorprendente, cada vez ofrecen más y más herramientas que permiten la comunicación no presencial a través de diversas formas tal es el caso de las redes sociales.

En México, en 2020, se estimó una población de 84.1 millones de usuarios de internet, que representan 72% de la población de seis años o más. Las principales actividades que realizan los usuarios de Internet en 2020 son comunicarse (93.8%), buscar información (91.0%) y acceder a redes sociales (89.0%) (INEGI, 2021).

Las conductas de odio a través de las redes sociales se caracterizan por tocar temas que atacan diferentes vulnerabilidades en la sociedad o en la integridad de una persona de forma directa, muchas de estas conductas impactan directamente en temas como la misoginia.

Al analizar la situación de temas misóginos alrededor del mundo se encuentra que, en Europa, una de cada 10 mujeres entrevistadas reporta haber recibido acoso sexual en línea a partir de los 15 años siendo el contenido principal ofensas u ofertas de índole sexual. Igualmente, además de que en algunos casos también contienen amenazas de agresión física o sexual por lo que prácticamente la mitad de las mujeres violentadas temen por su seguridad.

En un contexto más cercano, un estudio realizado por el Centro de Investigación Pew en Estados Unidos permitió identificar mientras los hombres reciben amenazas o insultos por medios electrónicos, son las mujeres quienes reciben contenido con

acoso de género o contenido sexual lo cual da como resultado que el 21% de las mujeres entre 18 y 29 años entrevistadas reportó haber sido acosada sexualmente en línea, más del doble que los hombres de la misma edad (9%) (GeoINT, 2022.)

Según cifras del INEGI se estima que en México la población usuaria de internet que fue víctima de ciberacoso aumentó de 21% en 2020 a 21.7 % en 2021. Para el año 2021 el 22.8% de mujeres que usan internet manifestaron haber sido víctimas de ciberacoso lo cual tiene mayor recurrencia en los estados de Michoacán, Guerrero y Oaxaca (MOCIBA, 2021).

El rango de usuarios en redes sociales hace imposible saber con exactitud cuántos usuarios las utilizan tomando en cuenta que existe más de una red social y que existen cifras alarmantes que corresponden a perfiles no fidedignos. Este tipo de situaciones han fomentado la alteración del objetivo principal de la comunicación en ellas. Es decir, poder escribir un comentario con actitudes de odio es algo posible de realizar, pese a que muchas de las redes sociales actuales han desarrollado mecanismos para evitar este tipo de situaciones ninguno es tan potente como la mente humana y a veces basta con crear una nueva palabra o modificar una ya existente con diferentes caracteres para que este pase por desapercibida.

## **Justificación**

Con el paso del tiempo y la evolución de la tecnología, las redes sociales han facilitado la comunicación entre personas respondiendo así a la necesidad de la comunicación. Actualmente es mucho más fácil comunicarnos, podemos hablar y ver a una persona.

En ocasiones el contenido en redes sociales también conlleva cosas que la gente no es capaz de decir cara a cara. El anonimato y cierta sensación de seguridad al saber que a través de un perfil falso es fácil llevar a cabo mayor violencia verbal. “En México 9 millones de mujeres mayores de 12 años han sufrido acoso cibernético; las adolescentes y jóvenes las más expuestas” (Gov.Mx, 2022.)

Este proyecto tiene como uno de sus objetivos aportar a diferentes aspectos en primer lugar socialmente a la erradicación de conductas de odio reflejadas en contenido público de redes sociales como Facebook, enfocándose principalmente a contenido misógino que, en base a diferentes fuentes, en México es un problema real y que afecta a una gran parte de mujeres además es un problema que da pie a severas consecuencias

También en el ámbito tecnológico con modelos predictivos asociados a los resultados obtenidos que servirán como base para el desarrollo de herramientas computacionales o bien para investigaciones posteriores dentro de este ámbito que aporten al avance de la detección de conductas de odio y temas adyacentes.

El retomar temas como conductas de odio a la mujer en medios sociales es de suma importancia puesto que al tomarlo de diferentes perspectivas como la red social a analizar y el idioma de origen del contenido objetivo se contribuye a entender mejor y ampliar el panorama de un tema que afecta a una gran parte de la población y que persiste a pesar de los cambios o medidas que se han aplicado a través de trabajos relacionados previamente.

## **Hipótesis**

Se estima que mediante la construcción de un conjunto de datos de más de 1700 elementos provenientes de redes sociales es posible entrenar un modelo capaz de identificar conductas misóginas en un texto con una precisión por encima del 75%.

## **Objetivo general**

- Construir y validar un conjunto de modelos predictivos basados en la tecnología Transformers para la detección automática de lenguaje misógino a partir de información proveniente de redes sociales en español.

## **Objetivos específicos**

- Analizar el estado del arte referente a detección de odio en medios sociales.
- Analizar y seleccionar fuentes de medios sociales en el contexto de lenguaje misógino.
- Recopilar, pre procesar y analizar un conjunto de datos provenientes de las fuentes de datos seleccionadas.
- Construir modelos predictivos basados en la tecnología Transformers para la detección automática de lenguaje misógino.
- Configurar y validar el grupo de modelos de análisis predictivo desarrollado con base en un conjunto de datos de lenguaje misógino distinto al recopilado y procesado inicialmente.
- Diseñar y desarrollar una aplicación web que integre los modelos validados para al menos un caso de estudio.

## **Alcances y limitaciones**

### **1.1.11 Alcances**

- La obtención y construcción de un conjunto de datos de por lo menos 1700 elementos de forma balanceada que permita el correcto entrenamiento del modelo final.
- El enriquecimiento de un conjunto de datos a través de la extracción de datos de por lo menos más de una red social.
- La corrección de los datos a través de la revisión de un experto en el tema garantizando así un mejor resultado y un modelo capaz de detectar conductas misóginas en cualquier texto.
- La creación de una aplicación web que permita la explotación del modelo y probar los resultados de una forma práctica.

### **1.1.12 Limitaciones**

- El cambio de políticas en redes sociales que pueden llegar a afectar o frenar la obtención de datos en ellas, tal es el caso de Twitter que actualmente se encuentra reduciendo la cantidad de tweets que se pueden recuperar a través de su API de forma gratuita, presentando costos muy altos si es que se necesitan a grandes volúmenes.
- La cantidad de datos que se encuentran en idiomas ajenos al español o algunos otros que se encuentran con algunas palabras en portugués o inglés esto puede afectar directamente al modelo.
- El adoptar nuevas palabras como ofensivas o dar un nuevo sentido a una oración con propósitos misóginos es difícil de manejar es por ello por lo que los resultados y el modelo obtenido es aplicable para su estudio en un contexto actual.
- Para poder obtener un conjunto de datos que sea certero y con un buen rumbo es necesario el acompañamiento de especialistas en el tema lo cual depende de la disposición de algunas organizaciones dentro de la región.

## **CAPÍTULO II ESTADO DEL ARTE**

### **Trabajos relacionados**

A continuación, se anexan algunos datos de proyectos o artículos de investigación que se relacionan con el tema principal de este trabajo, tomando como punto central algún tema aunado a conductas que fomentan el odio.

El apoyo de estos trabajos relacionados aporta a una rama específica de desarrollo y justificación de este proyecto.

En (Tanase M., 2020) se puede encontrar un esfuerzo por poder hacer el reconocimiento binario de conductas agresivas sobre tuits escritos en español. Esto aplicando procesamiento de lenguaje natural basadas en arquitecturas transformer.

Nos presenta un enfoque de la tarea compartida de detección automatizada de agresividad en muestras tomadas de redes sociales en español del 2020. Se experimentó con el ajuste de modelos que se basan en Transformer que han sido pre entrenados, soluciones que lograron resultados positivos en múltiples Tareas de PNL. Determinan que el método que proponen se puede aplicar con éxito utilizando modelos pre entrenados tanto en español como otros idiomas. Además, incluyeron varias combinaciones de conjuntos de datos anotados para varias tareas relacionadas (es decir, incitación al odio, lenguaje ofensivo y acoso) para poder realizar ajustes que apunten a una mejor precisión. Descubrimos que el rendimiento de los modelos se puede mejorar agregando más datos, incluso si están etiquetados para una tarea ligeramente diferente.

En (Plaza-Del-Arco F., 2020) se encuentra un trabajo centrado en poder encontrar de forma específica la detección de comentarios contra mujeres e inmigrantes en redes sociales centrándose en textos en español. Para lograr sus objetivos se aplicaron diferentes técnicas de aprendizaje como NB, SVM, regresión logística (LR), árbol de decisión (DT) y un clasificador de votación por conjuntos. Todo esto



utilizando una biblioteca de aprendizaje automático de software libre para la programación de Python.

Este artículo describe tres enfoques diferentes para clasificar expresiones de odio contra las mujeres y los inmigrantes en los tuits en español: un enfoque basado en el léxico, un enfoque de aprendizaje automático supervisado y un enfoque de aprendizaje profundo, la mayor parte de su investigación trata sobre documentos en inglés, ya que argumentan que los estudios y recursos para el estudio de otros idiomas era insuficiente, como el español. Antes de ejecutar sus experimentos, se encontraron con la necesidad de preprocesar el conjunto de datos.

En (Aluru S., 2020) proponen el primer análisis a gran escala del odio multilingüe mediante el análisis del rendimiento de los modelos de aprendizaje profundo en 16 conjuntos de datos de 9 idiomas diferentes también se realizaron experimentos en diversas condiciones (bajos y altos recursos, entornos monolingües y multilingües) para una variedad de idiomas. A partir de ello logran identificar que formas de entrenamiento pueden ser más precisos.

Después de su proceso se logró construir un catálogo indicando qué modelo es efectivo para un idioma en particular dependiendo de la extensión de los datos disponibles.

En (Mozafari M., 2020) trabajaron sobre palabras o lenguaje que algunos consideran profano, vulgar u ofensivo este estudio revela que los conjuntos de datos de referencia para el discurso de odio y las tareas de identificación de lenguaje abusivo contienen rarezas que causan una alta preferencia por los clasificadores para clasificar algunos simples en clases específicas.

Estas rarezas identificadas se asocian principalmente con una alta correlación entre algunos datos específicos de un conjunto de entrenamiento y una clase negativa específica. El empleo de un enfoque de evaluación de dominios cruzados, utilizando clasificadores entrenados en estos conjuntos de datos, demostraron algunos sesgos en estos clasificadores. Por lo tanto, también utilizando un clasificador basado en

BERT pre entrenado. Su trabajo se considera una extensión de trabajos previos relacionados al tema en el que proponen una transferencia y enfoque de aprendizaje para la identificación del discurso de odio en las redes sociales en línea mediante el empleo de una combinación del modelo pre entrenado no supervisado BERT y el nuevo ajuste supervisado.

En (Calderón C., 2020) analizan el rechazo verbal al extranjero como potencial detector de discurso de odio a través de dos análisis de contenido de tuits en español recogidos con la API de Twitter.

Analizaron el rechazo verbal al extranjero como potencial detector de discurso de odio a través de dos análisis de contenido de tuits en español obtenidos con la API de Twitter, uno de ellos de forma automática con técnicas de extracción y procesamiento de big data. En ambos casos el rechazo hacia los migrantes fue significativamente mayor que hacia los refugiados, como se había observado ya en contextos internacionales.

Se basa en el análisis de contenido y en la clasificación automatizada de textos basada en el aprendizaje automático supervisado (supervised machine learning). Esta técnica se aplicó a mensajes de Twitter, de manera que cada tuit conformaba una unidad de análisis.

En (Fragoso P., 2019) nos muestran como resultado de la colaboración del Centro de Investigación en Ciencias de Información Geoespacial (Centro Geo) con el Centro de Investigación y Estudios Superiores en Antropología Social (CIESAS) a través de un análisis logran la creación de una tipología de las expresiones que agreden y discriminan a las mujeres en Twitter y que por su popularidad pasan desapercibidas o son interpretadas como poco nocivas.

Dentro de su investigación destacan el empleo de metodologías transdisciplinarias para integrar el desarrollo de algoritmos que ayuden a la identificación de lenguaje agresivo en la plataforma digital.

Dentro de sus objetivos se encuentra la construcción de modelos predictivos a través de la categorización de una base de datos que contiene miles de tuits catalogados con contenido misógino y no misógino que permitan identificar otras expresiones similares (Fragoso P., 2019).

## **Análisis comparativo de los trabajos relacionados.**

En la siguiente sección (Tabla 1) se encuentra el cuadro comparativo de los trabajos descritos dentro del estado del arte, en él se puede observar el contraste de tecnologías y resultados, así como el aspecto de cada uno de ellos que resultan de interés.

*Tabla 1 Tabla comparativa del estado del arte*

<b>Autor</b>	<b>Título del artículo</b>	<b>Tecnologías</b>	<b>Resultados</b>	<b>Interés en artículo</b>
<b>(Tanase M, 2020)</b>	Detecting Aggressiveness in Mexican Spanish Social Media Content by Fine-Tuning Transformer-Based Models.	Modelos basados en transformer.	Se determinó que el rendimiento de los modelos se puede mejorar agregando más datos, incluso si están etiquetados para una tarea ligeramente diferente.	Su procesamiento basado en modelos que implementan tecnología transformer.

<p><b>Plaza-Del-Arco F., 2020)</b></p>	<p>Detecting Misogyny and Xenophobia in Spanish Tweets Using Language Technologies</p>	<p>Librerías de machine learning compatibles con Python.</p>	<p>Determinaron que es necesario trabajar antes de lograr un sistema preciso.  Y a pesar de que los textos objetivos fueron en español los criterios de clasificación fueron en inglés.</p>	<p>La forma de preprocesar el conjunto de datos y la depuración de contenido que debe ser modificado para cumplir las características y hacer un aporte eficiente.</p>
<p><b>(Aluru S., 2020)</b></p>	<p>Deep Learning Models for Multilingual Hate Speech Detection*</p>	<p>Modelos:</p> <ul style="list-style-type: none"> <li>• CNN-GRU</li> <li>• BERT</li> <li>• MBERT</li> </ul>	<p>Construcción de un catálogo indicando qué modelo es efectivo para un idioma en particular dependiendo de la extensión de los datos disponibles</p>	<p>La forma en que abordan sus objetivos ocupando diversos idiomas en diferentes condiciones además de ser comparados bajo diferentes modelos de procesamiento y entrenamiento.</p>

<b>(Mozafari M., 2020)</b>	Hate speech detection and racial bias mitigation in social media based on BERT model	Modelo de lenguaje llamado BERT (Representaciones de codificador bidireccional Transformers)	Este estudio revela que los conjuntos de datos de referencia para las tareas de identificación de discurso de odio y lenguaje abusivo contienen rarezas que causan una alta preferencia por los clasificadores.	La detección de sesgo y su propuesta de un mecanismo de alivio del sesgo para disminuir el impacto de las rarezas en los datos de entrenamiento
<b>(Calderón C., 2020)</b>	Rechazo y discurso de odio en Twitter: análisis de contenido de los tuits sobre migrantes y refugiados en español.	Supervised machine learning.	Lograron obtener resultados traducidos a cifras: En un 16,7% de los tuits se expresó aceptación hacia el colectivo, mientras que un 38,3% de los tuits fueron mensajes neutros.	La red social que abordan Twitter y que al igual que en muchos más casos ocupan a esta red social como primer objetivo para la extracción y análisis de datos.

<b>(Fragoso P., 2019 )</b>	Misoginia en pocas palabras: Identificación y análisis de violencia escrita contra las mujeres en textos cortos de Twitter.	Modelos predictivos Machine learning.	En este trabajo se identificó y analizó violencia escrita con contenido misógino en redes sociales, y establecer correlaciones entre este fenómeno y la presencia de delitos violentos contra las mujeres y niñas en territorio mexicano	El entorno de estudio en territorio mexicano y su relación con consecuencias directas como el feminicidio a partir de este tipo de conductas.
----------------------------	---	---------------------------------------	--	---

Si bien existen trabajos que ocupan el aprendizaje automático y distintas herramientas de inteligencia artificial, los Transformers han llegado a revolucionar la precisión de modelos predictivos, y pese a que si han sido aplicados en trabajos enfocados en conductas de odio y redes sociales actualmente la misoginia en contenido en español es un tema que se encuentra en etapas muy tempranas en cuanto a modelos que implementan Transformers. Además, en el presente trabajo se realiza un dataset que resulta del conjunto de tweets pero que también toma en cuenta contenido en Facebook añadiendo con el fin de enriquecer los datos de entrada y obtener resultados favorables.

## CAPÍTULO III METODOLOGÍA Y DESARROLLO

### Solución propuesta

A continuación, en la figura 6 se presenta la arquitectura del proyecto compuesto de 5 elementos principales descritos en los siguientes apartados.

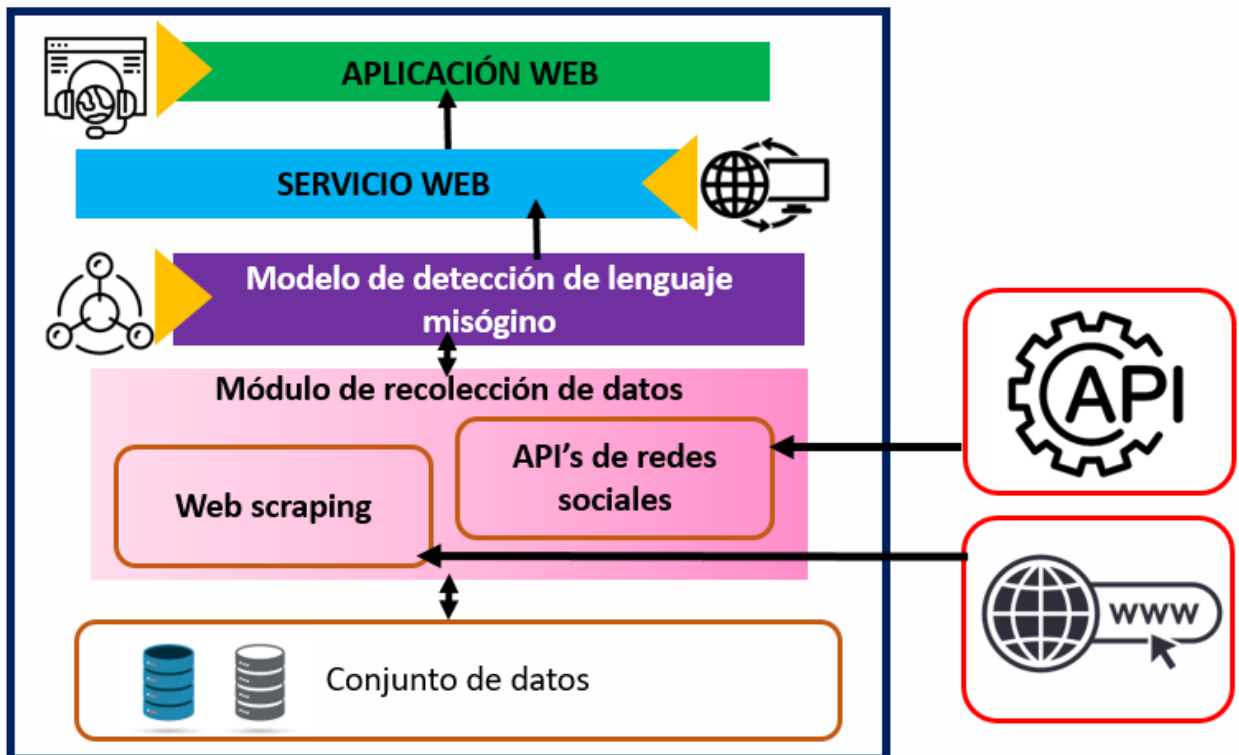


Figura 4 Arquitectura de solución propuesta.

La solución propuesta se basa en la extracción de información a través de herramientas de web scraping todo esto regido en un proceso que conlleva el seguimiento de una metodología para poder obtener resultados positivos. Con el fin de que a través de técnicas de procesamiento de datos apoyada por el lenguaje de programación Python sea posible establecer un modelo que responda a la detección de lenguaje de odio en medios sociales, una vez obtenido el modelo será expuesto a través de un servicio web para poder consumirlo y usarlo en la evaluación de valores de entrada para observar y contrastar el contenido.

### **3.1.1 Conjunto de datos**

Entender la importancia del conjunto de datos implica saber que son el pilar de este proyecto puesto que a partir de ellos es que se va a realizar el entrenamiento del modelo que será capaz de categorizar el contenido.

Para este conjunto de datos es importante considerar como meta alrededor de 1700 registros de texto (Comentarios y/o tweets) que a través de algunos pasos de depuración conformen materia suficiente para un enriquecimiento al modelo y su acercamiento con lo óptimo.

### **3.1.2 Módulo de recolección de datos**

Para el módulo de recolección de datos se realizó la extracción de comentarios públicos en Facebook a través de técnicas de web scraping o bien con el API publica de Twitter y las llaves que proporciona para obtener tweets en volumen que puedan ser tratados, analizados y clasificados con contenido misógino o no según corresponda.

### **3.1.3 Modelo de detección de lenguaje misógino**

En esta etapa se desarrolló el entrenamiento de un modelo que a través de las librerías de Transformers en Python las cuales comprenden el uso de una red neuronal que aprende contexto y, por lo tanto, significado mediante el seguimiento de relaciones en datos secuenciales como las palabras de una oración.

El modelo obtenido será de vital importancia puesto que de él depende la correcta clasificación del contenido que será expuesto a través del sistema para ser catalogado como misógino o no.

### **3.1.4 Servicio Web**

Para este módulo del proyecto se requiere el uso de tecnología que permita la creación de un servicio web capaz de poder emplear el modelo entrenado con Transformers para que pueda recibir texto desde cualquier aplicación de cliente y



sea capaz de regresar un valor que permita identificar si se encuentra contenido misógino en el valor de entrada.

Para la realización del servicio Web se empleó el marco de trabajo Django que por su compatibilidad con el lenguaje Python y su simpleza es una excelente opción para la creación de servicios con una arquitectura bastante digerible y potente.

### **3.1.5 Aplicación Web**

Para esta aplicación Web se utilizó como marco de trabajo de apoyo a Angular que, al ser una plataforma de desarrollo construida sobre TypeScript y basado en componentes cumple perfectamente con el objetivo de desarrollar aplicaciones de una sola página a través de componentes ordenados y bien identificados haciendo que el frontend sea presentable y potente (En los anexos "5.1.3 Maqueta de Aplicación Web" se puede consultar la estructura propuesta).

### **Comprensión del negocio**

Debido a la flexibilidad de la metodología CRISP-DM se determina que las etapas que propone se ajustan con los objetivos de este proyecto los cuales proponen la construcción y validación de un conjunto de modelos predictivos basados en la tecnología Transformers que permitan detectar de forma automática lenguaje misógino a partir de información proveniente de redes sociales en español.

La importancia del manejo de información que será empleada para la construcción de modelos resulta fundamental, es por ello por lo que las etapas 2 y 3 de esta metodología juegan un papel importante para obtener resultados favorables en las etapas posteriores.

Dentro del objetivo para los datos se contempló un total de 2000 registros que ayuden a entrenar un modelo capaz de identificar contenido misógino a partir de comentarios en redes sociales que se encuentren en idioma español.

## Comprensión de los datos

### 3.1.6 Identificación de comentarios objetivos

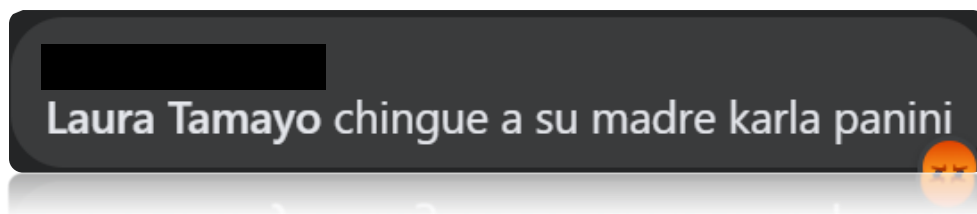
Dentro de estas actividades se realizó la comprensión de requisitos que debe tener un comentario que ayude a la correcta categorización de contenido misógino y la búsqueda de comentarios objetivos en redes sociales a través del ingreso a publicación de mujeres que se encuentran como tendencia en los últimos meses para ello es importante crear una lista con algunos nombres identificados como opción para posible texto objetivo, durante algunas revisiones a primera instancia se logran identificar los nombres enlistados en la Tabla 2.

*Tabla 2 Lista de personas identificadas para extracción de datos*

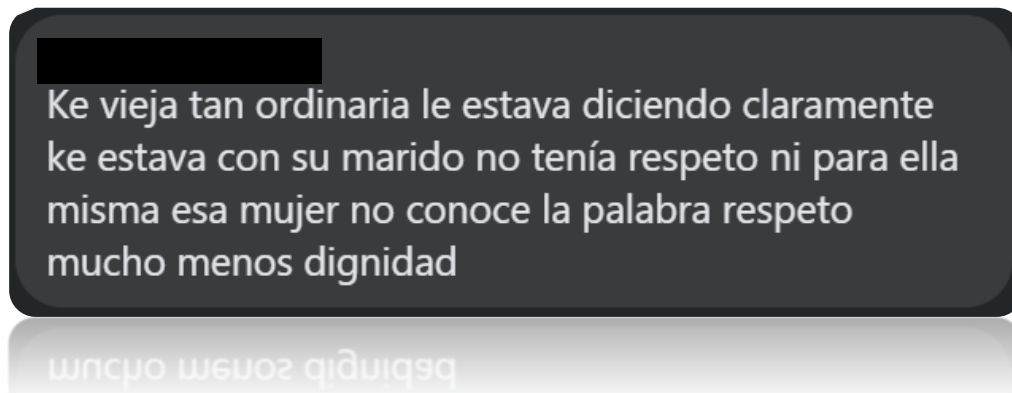
<b>Nombre</b>
Yalitza Aparicio
Tattis Beauty
Halle Bailey
Mon Laferte
Gloria Trevi
Bárbara de Regil
Karely Ruíz
Danna Paola
Rebeca Mendiola
Belinda Peregríni

Alondra Castro
Sol León
Adela Micha
Kimberly Loaiza

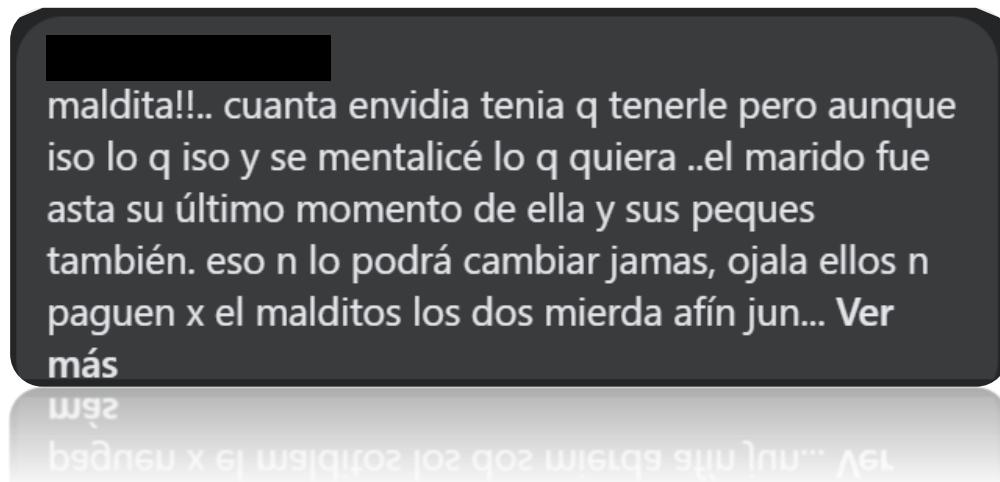
En las Figura 9, Figura 10 y Figura 11 se muestran comentarios que resultan atractivos como objetivo pues contienen cierto nivel de agresión o palabras identificadas o relacionadas directamente con gestos misóginos, estos son ejemplos de publicaciones referentes a mujeres enlistadas anteriormente que cumplen alguna característica para denominarlo como ofensivo.



*Figura 5 Ejemplo de comentario objetivo obtenido de Facebook.*



*Figura 6 Ejemplo de comentario objetivo obtenido de Facebook.*



*Figura 7 Ejemplo de comentario objetivo obtenido de Facebook.*

### **3.1.7 Propuesta de fuentes de extracción de datos**

En esta fase se determinó utilizar a Twitter y Facebook como medios para la obtención de datos debido a que dentro de su contenido es fácil poder identificar comentarios que puedan enriquecer al proyecto y que debido a que es contenido publico pueden emplearse técnicas de extracción de datos para poder comenzar con la recopilación de datos.

### **3.1.8 Recopilación de datos**

#### ***Extracción en Facebook***

Para la extracción de datos en Facebook se emplea un Web Scraper con código en Python que a través de las Url's proporcionadas realiza la extracción de los comentarios en su primer nivel de jerarquía de la publicación concentrándolos en este caso en un archivo. json para la limpieza y tratamiento de los datos obtenidos.

#### ***Extracción en Twitter***

Twitter cuenta con un conjunto de funciones programáticas que permite a los desarrolladores interactuar con esta plataforma y su contenido, en este caso se empleó su API que a través de llaves proporcionadas a través de la solicitud a Twitter de forma directa para su uso en este proyecto es posible la creación de una herramienta que permita concentrar comentarios que contengan una palabra o frase ingresada por el usuario obteniendo como producto un archivo json que permita la manipulación de estos.

### 3.1.9 Preparación de los datos

#### *Análisis de los datos*

Una vez realizadas las actividades propuestas en la etapa anterior se logró obtener datos que a su vez pueden ser subdivididos de dos formas: por una parte, texto que contiene palabras estratégicas directamente relacionadas con contenido misógino en español mientras del otro lado encontramos texto extraído como comentarios en publicaciones de Facebook.

La Tabla 3 presenta el número de registros para Twitter, mientras que la Tabla 4 muestra el número de registros para Facebook.

*Tabla 3 Número de comentarios obtenidos en Twitter*

<b>Twitter</b>	
<b>Palabra/Frase Buscada:</b>	<b># de Tweets con Coincidencia.</b>
Facilota	125
Golfa	1685
Maldita criada	12
Maldita zorra	288

Mantenida	7395
Piruja	1177
Putita	19535
Vieja interesada	25
Vieja inútil	758
Vieja pendeja	787
Puta	0

*Tabla 4 Número de comentarios obtenidos en Facebook*

<b>FACEBOOK</b>	
<b>Publicación referente a:</b>	<b># de Comentarios:</b>
Belinda Peregrin	120
Kenia Os	22
Liz Vega	151
Miriam Carballo	1200
Karely Ruiz	31
Yalitza Aparicio	32

Alondra Castro	21
Sol León	23
Paulina Rubio	24
Kimberly Loaiza	20
Adela Micha	23

A continuación, se dan a conocer algunas actividades llevadas a cabo para la depuración de los datos puesto que de los datos en crudo se logran identificar algunas inconsistencias como:

- Textos o comentarios repetidos.
- Textos que incluyen oraciones o contenido en idiomas ajenos al español.
- Textos u oraciones que se quedan truncados o terminan con `...'`
- Registros que cuentan con poco contenido semántico.

### ***Eliminar comentarios repetidos***

De un total de 33,454 comentarios obtenidos de ambas redes sociales gran contenido de este se encontraba repetido debido a que eran comentarios compartidos, replicas o re-tweets debido a ello se opta por eliminar la mayor cantidad posible de este tipo de datos basura.

Del registro original se extrae el comentario sin identificador para una comparación directa de cadena por cadena ignorando los comentarios iguales.

En el siguiente código se presenta un método que recorre a todos los elementos eliminando de forma directa textos que ya se encuentran en el nuevo conjunto de datos que se construye.

```

#Angel Oswaldo Vázquez Benito
#abril 2023
#Fragmento de código dedicado a eliminar registros repetidos de forma
directa
import json
import os
import pandas as pd
directorio = os.listdir('datosCrudos')

# Opening JSON file
for archivo in directorio:
    print(archivo)
    f = open('carpetaPrueba/'+archivo, encoding="utf8")
    data = json.load(f)
    result = []
    arrErrores = []
    for item in data:
        try:
            if item not in result:
                result.append(item)
        except:
            arrErrores.append(item)
    print(archivo)
    out_file = open('datosSinRepetir/'+archivo, "w")
    json.dump(result, out_file, indent=6)
    out_file.close()

```

### ***Validación de longitud de cadena y caracteres especiales***

Del resultado del paso anterior se detecta que existen comentarios que, a pesar de ser ofensivos, la longitud de la cadena es demasiado pequeño para contribuir de forma significativa al aprendizaje del modelo, es por ello que se somete a un proceso para eliminar los registros cuya longitud sea menor a 50 caracteres además de eliminar los comentarios que comiencen con 'RT' o terminen con '...' puesto que este tipo de comentarios representan re tweets cuyo contenido es repetido en "n" ocasiones y que el comentario original ya forma parte del conjunto de datos.

En el siguiente código se evalúa a cada elemento comparando la longitud de la cadena directamente si es mayor a 50 de no cumplir con esta característica no se toma en cuenta para el conjunto de datos.

```

#Angel Oswaldo Vázquez Benito
#abril 2023

```



```

#Fragmento de código dedicado a eliminar registros que contengan texto
con menos de 50 caracteres de longitud y que no terminen en puntos
suspensivos
import json
import os
import pandas as pd

directorio = os.listdir('datosSinRepetir')

# Opening JSON file
for archivo in directorio:
    print(archivo)
    f = open('datosSinRepetir/'+archivo, encoding="utf8")
    #print(f)
    data = json.load(f)
    result = []
    arrErrores = []
    for item in data:
        try:
            text = item['Text']
            if len(text) > 50 and text[-3:] != '...' and text[-1:] !=
'...':
                print(text[-3:])
                result.append(item)
        except:
            arrErrores.append(item)
    print(archivo)
    out_file = open('datosDepurados/'+archivo, "w")
    json.dump(result, out_file, indent=6)
    out_file.close()

```

### ***Rectificar idioma y concentración de datos en español***

En el caso de algunas búsquedas se detecta la presencia de comentarios que, a pesar de contener la frase objetivo, se encuentran en idiomas en donde dicha palabra representa un cognado es por ello por lo que se requiere de la limpieza de idioma puesto que este trabajo se centra en el análisis de contenido en español. Dicho proceso se realiza utilizando una librería de Python para la etiqueta de idioma detectado en el texto del comentario y la supresión de aquellos que no son reconocidos como español.

Dentro del siguiente fragmento de código se presenta un método que se apoya de una librería capaz de identificar el idioma del texto que se dé como entrada, iterando

cada elemento del conjunto de datos se etiqueta el idioma detectado y de no ser español se elimina.

```
#Angel Oswaldo Vázquez Benito
#abril 2023
#Fragmento de código dedicado a identificar el idioma del texto y
eliminar registros que no se encuentren en español
import os
import json
import pandas as pd
from langdetect import detect, detect_langs

def language_detection(texto, metodo = "single"):

    if(metodo.lower() != "single"):
        result = detect_langs(texto)
    else:
        result = detect(texto)
    return result

directorio = os.listdir(' datosDepurados ')
for archivo in directorio:
    nombreArchivo = (archivo.rsplit('.', 1)[0])
    f = open(' datosDepurados /'+archivo, encoding="utf8")
    data = json.load(f)
    arrResult = []
    arrErrores = []
    arrEspañol = []
    for i in data:
        if 'Text' in i:
            try:
                language = language_detection(i['Text'], "single")
                i['Language'] = language
                arrResult.append(i)
                if language == 'es':
                    arrEspañol.append(i)
            except:
                arrErrores.append(i)

#Guardar archivo de items con idioma
out_file = open('datosConIdioma/'+archivo, "w")
json.dump(arrResult, out_file, indent=6)
out_file.close()

#Guardar archivo de items con idioma solo Español
out_file = open('datosEnEspañol/'+archivo, "w")
json.dump(arrEspañol, out_file, indent=6)
out_file.close()

pd.DataFrame(arrErrores).to_excel(nombreArchivo+"ErrorIdioma.xlsx")
```

### ***Etiqueta de datos***

Una vez pre procesados los datos, se procede a asignar una etiqueta a cada uno de ellos, es decir, se asigna el valor de 0 para aquellos que no presentan contenido misógino y el valor de 1 en donde se identifica contenido que si lo es. Esto se realiza ordenando los archivos por número de caracteres es decir tomando como prioridad los comentarios más grandes.

Es importante mencionar la contribución, apoyo y revisión durante este proceso de especialistas en el área de psicología por parte del Instituto Tecnológico Superior de Teziutlán y del sistema D.I.F. del municipio de Chignautla, Puebla y fue en esta etapa en donde se eliminaron y adicionaron comentarios que pudieran enriquecer más el valor semántico de términos misóginos para obtener una mejor versión del conjunto de datos.

En la Tabla 5 se aprecian algunos ejemplos de texto extraído con su etiqueta de valor asignada de forma manual durante esta etapa.

*Tabla 5 Ejemplos de textos extraídos*

<b><i>Texto</i></b>	<b>Valor De Etiqueta</b>
<i>@lucia_flaquitaN: Ya han pasado 8 días, desde que taladre profundamente mi culo con un pepino □ y aun traigo un poco inflamado mi ano. No cabe duda que soy una puta.</i>	1
<i>Zorra ya estoy muerto, perdon si t menti adiós piruja.</i>	1

<i>@desgraciatdel96 Eso sí, Eiza González me puede dar un beso cuando quiera.</i>	0
<i>@earperspain @villaneve_spain @OhanaAFW @croquetavegana_ JAJAJAJA na, fuera de bromas. A mi si me gusta mucho 😍 y luego Eiza González tb</i>	0
<i>@folkloss Adiós a mi sueño de ver a Eiza González como Evelyn :( me encanta como actriz y aparte es hermosa</i>	0

### 3.1.10 Modelado y evaluación

Una vez obtenido el conjunto de datos final (Fragmento de elementos puede ser consultado en los Anexos "5.1.4 Extracto de conjunto de datos de entrada para el entrenamiento del modelo") se procedió con el entrenamiento para la obtención del modelo. A continuación, se presentan los valores obtenidos como se muestra en la Figura 13: Durante las tres iteraciones se logra apreciar una pérdida en el entrenamiento y validación que ronda en cifras bastante cercanas se refleja en la exactitud reflejada que como se observa en los tres casos es superior a un 75%.

Epoch	Training Loss	Validation Loss	Accuracy
1	0.513200	0.960224	0.802083
2	0.645100	0.826237	0.786458
3	0.414200	0.904688	0.802083

*Figura 8 Resultados de entrenamiento del modelo*

En la Figura 14 se observa que la precisión promedio como resultado de las dos diferentes clasificaciones posibles es de 81%, por otro lado el recall nos muestra un

promedio de 75% de esta forma la métrica F1 arroja un promedio del 76% aproximadamente haciendo de este modelo algo estable.

```
In [33]: # Previous
print(classification_report(test['Value'], predicciones))
```

	precision	recall	f1-score	support
0	0.87	0.57	0.69	79
1	0.76	0.94	0.84	113
accuracy			0.79	192
macro avg	0.81	0.75	0.76	192
weighted avg	0.80	0.79	0.78	192

*Figura 9 Promedios de precisión y F1 del modelo obtenido*

### 3.1.11 Despliegue

Una vez obtenido el modelo con las características previamente descritas se procede a su exposición mediante un servicio web que es capaz de recibir un texto como entrada y devolviendo un 0 para el caso de no tener contenido misógino y un 1 cuando si lo hay, se procede a la observación de textos reales extraídos de Twitter y Facebook.

En la Tabla 6 se observa la primera evaluación de texto en donde el resultado obtenido coincide con la salida que se esperaba.

*Tabla 6 Caso de clasificación de texto 1*

<b>EVALUACIÓN DE TEXTO</b>	
<b>Texto Por Evaluar:</b>	<i>"@FCBarcelona_es eso piensa la zorra de tu madre de ti, pobrecita que le salió tan retrasado el niño."</i>

<b>Valor Esperado:</b>	1 (Contenido misógino)
<b>Resultado:</b>	
	
<b>Observaciones:</b>	El valor obtenido coincide con el valor esperado.

Para la Tabla 7 se puede observar que durante el segundo caso el valor también coincide con lo que se esperaba como respuesta del servicio web.

*Tabla 7 Caso de clasificación de texto 2*

<b>EVALUACIÓN DE TEXTO</b>	
<b>Texto Por Evaluar:</b>	<p><i>"Nuestro país y sus pinches leyes que para algunos privilegiados hasta las sentencias se reducen,por eso más personas sin valores,normas hacen lo que se les pega la gana.Bravo"</i></p>

<b>Valor Esperado:</b>	0 (Sin contenido misógino)
<b>Resultado:</b>	
 <p>The screenshot shows a REST client interface. At the top, there is a method dropdown set to 'GET' and a URL: 'http://127.0.0.1:8000/polls/pruebaRapida?texto=Nuestro país y sus pinches leyes que para algunos pri ...'. A 'Send' button is to the right. Below the URL bar, there are tabs for 'Params', 'Authorization', 'Headers (9)', 'Body', 'Pre-request Script', 'Tests', and 'Settings'. The 'Params' tab is active, showing a table with one query parameter: 'texto' with the value 'Nuestro país y sus pinches leyes que para algunos privilegiados hasta las sent...'. Below the table, there are tabs for 'Body', 'Cookies', 'Headers (9)', and 'Test Results'. The 'Body' tab is active, showing a status bar with '200 OK', '61 ms', and '296 B'. There are also buttons for 'Pretty', 'Raw', 'Preview', 'Visualize', and 'HTML'.</p>	
<b>Observaciones:</b>	El valor obtenido coincide con el valor esperado.

En el caso del tercer tweet comparado también existe una coincidencia tanto con el valor esperado y el valor arrojado dando positivo a contenido misógino, tal como se aprecia en la Tabla 8.

Tabla 8 Caso de clasificación de texto 3

<b>EVALUACIÓN DE TEXTO</b>	
<b>Texto Por Evaluar:</b>	<p><i>"@MilanoRMM: Muy linda ayer combatiendo el machismo esperancita, pero el viern es entras gratis en la discoteca ¿no? Maldita criada"</i></p>

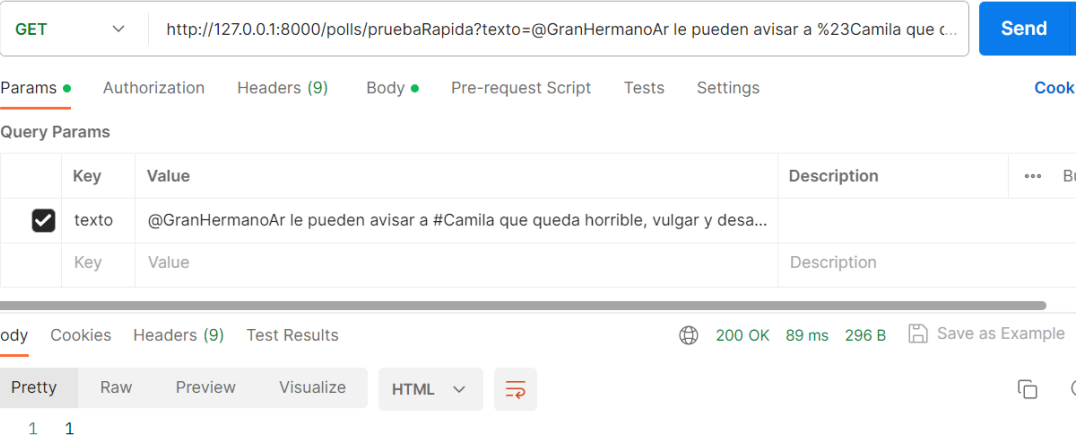
<b>Valor Esperado:</b>	1 (Contenido misógino)
<b>Resultado:</b>	
	
<b>Observaciones:</b>	El valor obtenido coincide con el valor esperado.

En el caso de la Tabla 9 número se espera un valor positivo de contenido misógino que justamente coincide con el valor que arroja al ser evaluado por el modelo.

Tabla 9 Caso de clasificación de texto 4

<b>EVALUACIÓN DE TEXTO</b>	
<b>Texto Por Evaluar:</b>	<p><i>"@GranHermanoAr le pueden avisar a #Camila que queda horrible, vulgar y desagradable que este todo el día mostrando el coyote??</i></p>

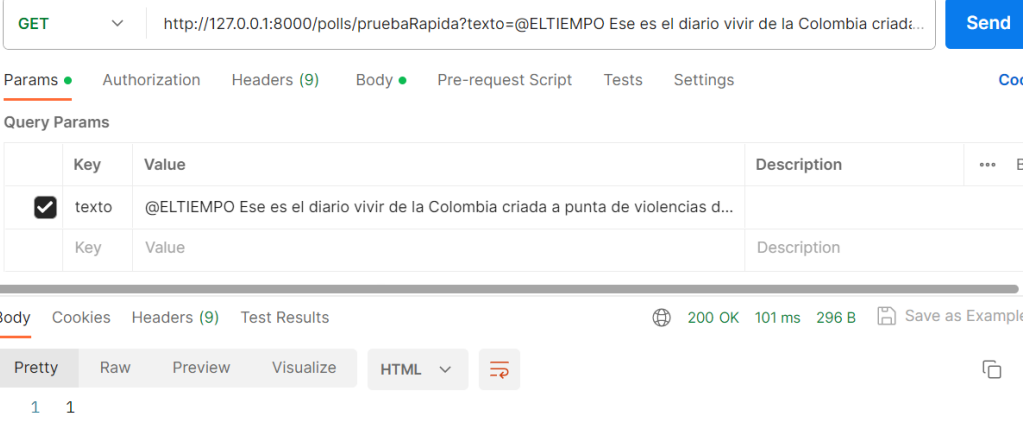


	#GranHernano2023 #GranHernanoArgentina"
<b>Valor Esperado:</b>	1 (Contenido misógino)
<b>Resultado:</b>	
 <p>The screenshot shows a REST client interface. At the top, a GET request is defined with the URL: http://127.0.0.1:8000/polls/pruebaRapida?texto=@GranHernanoAr le pueden avisar a %23Camila que c... The 'Send' button is highlighted in blue. Below the URL bar, there are tabs for Params, Authorization, Headers (9), Body, Pre-request Script, Tests, and Settings. The 'Params' tab is active, showing a table with one query parameter: 'texto' with the value '@GranHernanoAr le pueden avisar a #Camila que queda horrible, vulgar y desa...'. Below the table, there are tabs for Body, Cookies, Headers (9), and Test Results. The 'Body' tab is active, showing a status of 200 OK, a response time of 89 ms, and a size of 296 B. There are also options to 'Save as Example' and view the response in 'Pretty', 'Raw', 'Preview', or 'Visualize' format. The response content is currently empty.</p>	
<b>Observaciones:</b>	El valor obtenido coincide con el valor esperado.

Para el caso de la Tabla 10 se observa que pese a que el valor esperado es negativo para contenido misógino el valor obtenido no coincide esta vez arrojando un 1 como resultado.

*Tabla 10 Caso de clasificación de texto 5*

<b>EVALUACIÓN DE TEXTO</b>
----------------------------

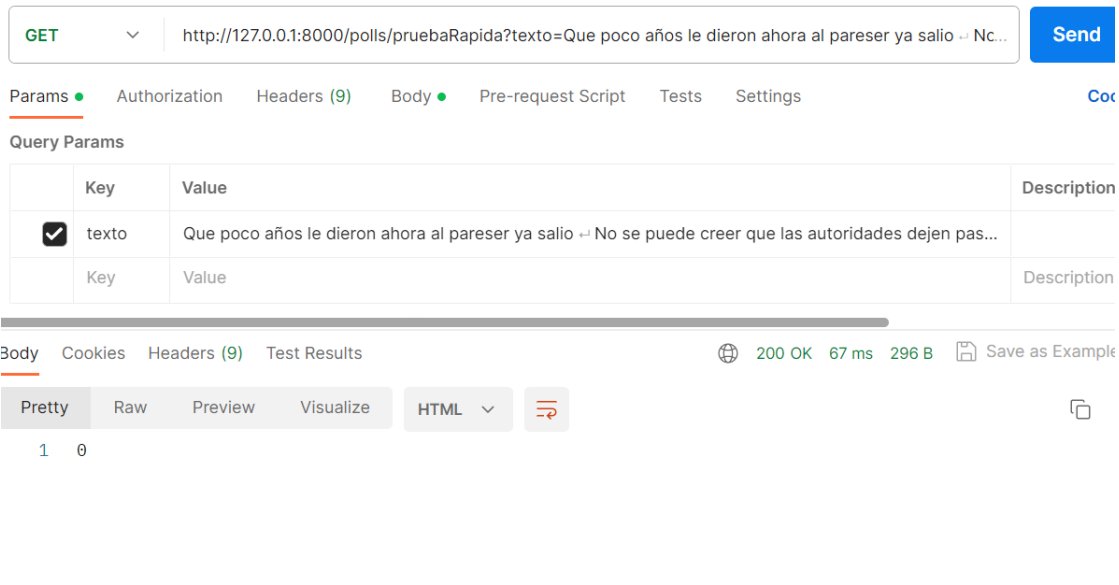
<p><b>Texto Por Evaluar:</b></p>	<p><i>"@ELTIEMPO Ese es el diario vivir de la Colombia criada a punta de violencias de toda clase desde hace muchos años con personajes que ya saben quiénes Fueron porque ahora hay otro actor más grave : La Maldita corrupción rampante e impunidad increíble en todos los ámbitos del país."</i></p>									
<p><b>Valor Esperado:</b></p>	<p>0 (Sin contenido misógino)</p>									
<p><b>Resultado:</b></p>										
 <p>The screenshot shows a REST client interface with the following details:</p> <ul style="list-style-type: none"> <li>Method: GET</li> <li>URL: http://127.0.0.1:8000/polls/pruebaRapida?texto=@ELTIEMPO Ese es el diario vivir de la Colombia criada...</li> <li>Params: Authorization, Headers (9), Body (selected), Pre-request Script, Tests, Settings</li> <li>Query Params table: <table border="1"> <thead> <tr> <th>Key</th> <th>Value</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>texto</td> <td>@ELTIEMPO Ese es el diario vivir de la Colombia criada a punta de violencias d...</td> <td></td> </tr> <tr> <td>Key</td> <td>Value</td> <td>Description</td> </tr> </tbody> </table> </li> <li>Response: 200 OK, 101 ms, 296 B</li> <li>Body tab selected, showing raw response content.</li> </ul>		Key	Value	Description	texto	@ELTIEMPO Ese es el diario vivir de la Colombia criada a punta de violencias d...		Key	Value	Description
Key	Value	Description								
texto	@ELTIEMPO Ese es el diario vivir de la Colombia criada a punta de violencias d...									
Key	Value	Description								
<p><b>Observaciones:</b></p>	<p>El valor esperado No coincide con el valor obtenido.</p>									

En los próximos dos casos correspondientes a la Tabla 11 y Tabla 12 el valor de salida si coincide con el valor esperado viendo de esta forma como se comporta el modelo con datos de entrada completamente nuevos.

Tabla 11 Caso de clasificación de texto 6

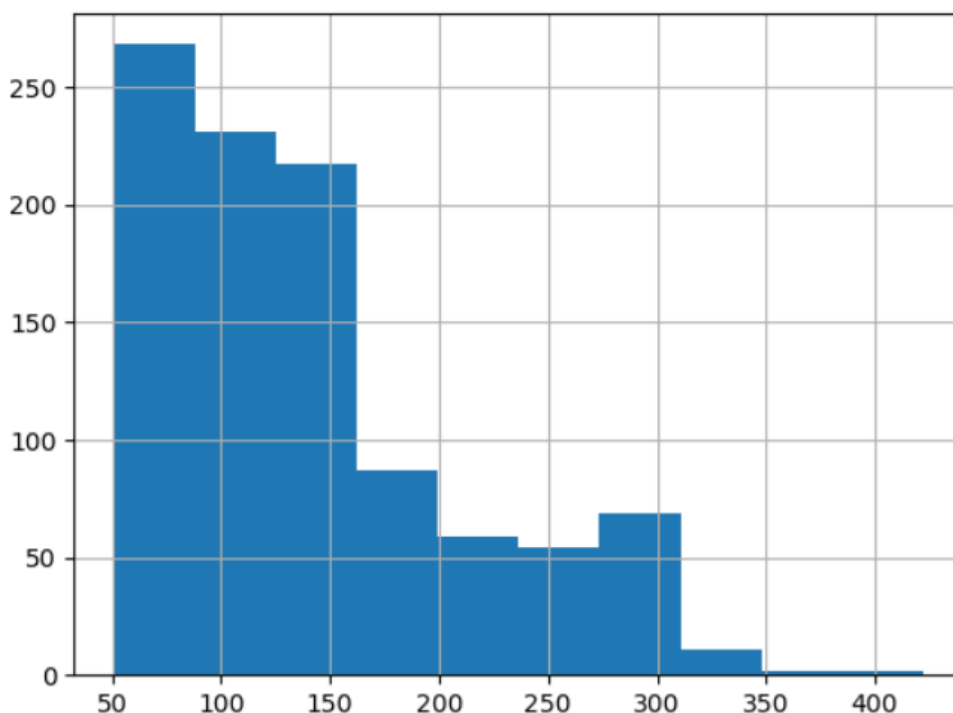
<b>EVALUACIÓN DE TEXTO</b>	
<b>Texto Por Evaluar:</b>	<i>"@rominacongusto A mí que chucha que puedas viajar piruja, por qué las ganas de querer compararnos entre mujeres. Suficiente tenemos con los hombres quitándonos espacios para que ahora nos toque lidiar con babosas cómo esta"</i>
<b>Valor Esperado:</b>	1 (Contenido misógino)
<b>Resultado:</b>	
 <p>The screenshot shows a REST client interface. At the top, a GET request is defined with the URL <code>http://127.0.0.1:8000/polls/pruebaRapida?texto=@rominacongusto A mí que chucha que puedas viajar ...</code> and a 'Send' button. Below the URL bar, tabs for 'Params', 'Authorization', 'Headers (9)', 'Body', 'Pre-request Script', 'Tests', and 'Settings' are visible. The 'Query Params' section contains a table with one entry: 'texto' with the value '@rominacongusto A mí que chucha que puedas viajar piruja, por qué las ganas ...'. The 'Body' tab is selected, showing a status of '200 OK', '92 ms', and '296 B'. At the bottom, there are tabs for 'Pretty', 'Raw', 'Preview', 'Visualize', and 'HTML', with 'Pretty' selected.</p>	
<b>Observaciones:</b>	El valor obtenido coincide con el valor esperado.

Tabla 12 Caso de clasificación de texto 7

<b>EVALUACIÓN DE TEXTO</b>										
<b>Texto Por Evaluar:</b>	<p><i>"Que poco años le dieron ahora al parecer ya salio</i></p> <p><i>No se puede creer que las autoridades dejen pasar por alto una violación</i></p> <p><i>Por eso estamos como estamos"</i></p>									
<b>Valor Esperado:</b>	0 (Sin contenido misógino)									
<b>Resultado:</b>										
 <p>The screenshot shows a REST client interface with the following details:</p> <ul style="list-style-type: none"> <li><b>Method:</b> GET</li> <li><b>URL:</b> http://127.0.0.1:8000/polls/pruebaRapida?texto=Que poco años le dieron ahora al parecer ya salio ↵ Nc...</li> <li><b>Params:</b> Authorization, Headers (9), Body, Pre-request Script, Tests, Settings</li> <li><b>Query Params:</b> <table border="1"> <thead> <tr> <th>Key</th> <th>Value</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td><input checked="" type="checkbox"/> texto</td> <td>Que poco años le dieron ahora al parecer ya salio ↵ No se puede creer que las autoridades dejen pas...</td> <td></td> </tr> <tr> <td>Key</td> <td>Value</td> <td>Description</td> </tr> </tbody> </table> </li> <li><b>Response:</b> 200 OK, 67 ms, 296 B</li> <li><b>Body:</b> Pretty, Raw, Preview, Visualize, HTML</li> <li><b>Content:</b> 1 0</li> </ul>		Key	Value	Description	<input checked="" type="checkbox"/> texto	Que poco años le dieron ahora al parecer ya salio ↵ No se puede creer que las autoridades dejen pas...		Key	Value	Description
Key	Value	Description								
<input checked="" type="checkbox"/> texto	Que poco años le dieron ahora al parecer ya salio ↵ No se puede creer que las autoridades dejen pas...									
Key	Value	Description								
<b>Observaciones:</b>	El valor obtenido coincide con el valor esperado.									

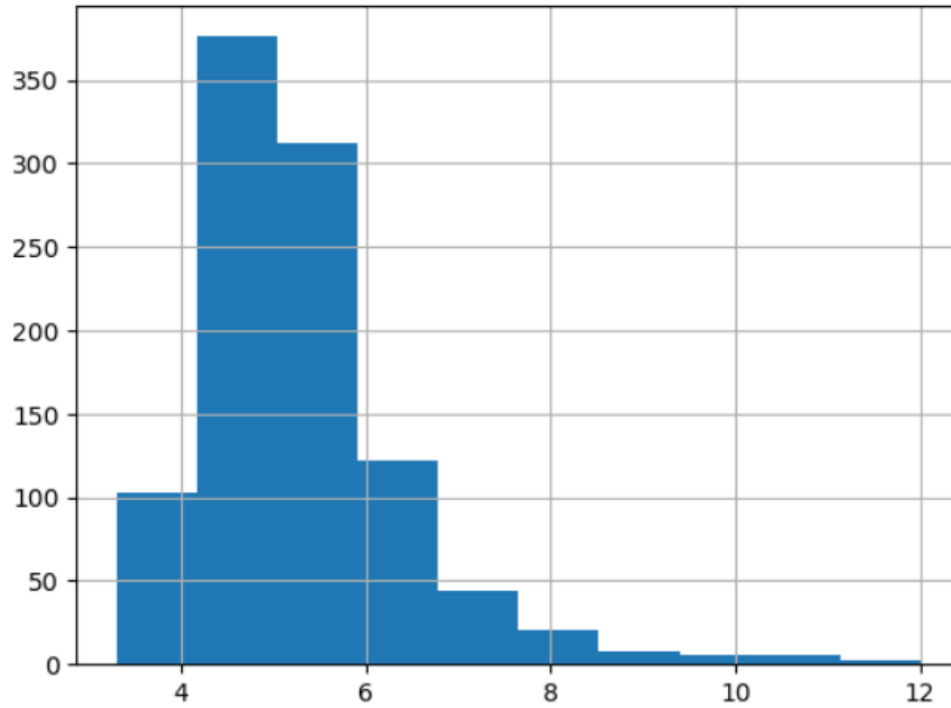
## CAPÍTULO IV RESULTADOS

A continuación, se presenta el análisis descriptivo del conjunto de datos el cual como se menciona en secciones anteriores es de suma importancia pues con base en este se lleva a cabo el entrenamiento del modelo. En la Figura 15 se puede ver representada la longitud de caracteres en cada elemento del conjunto observando así que se componen de textos cuyo rango de caracteres se encuentra entre los 50 y los 450 viendo mayor concurrencia en los primeros 150 caracteres.



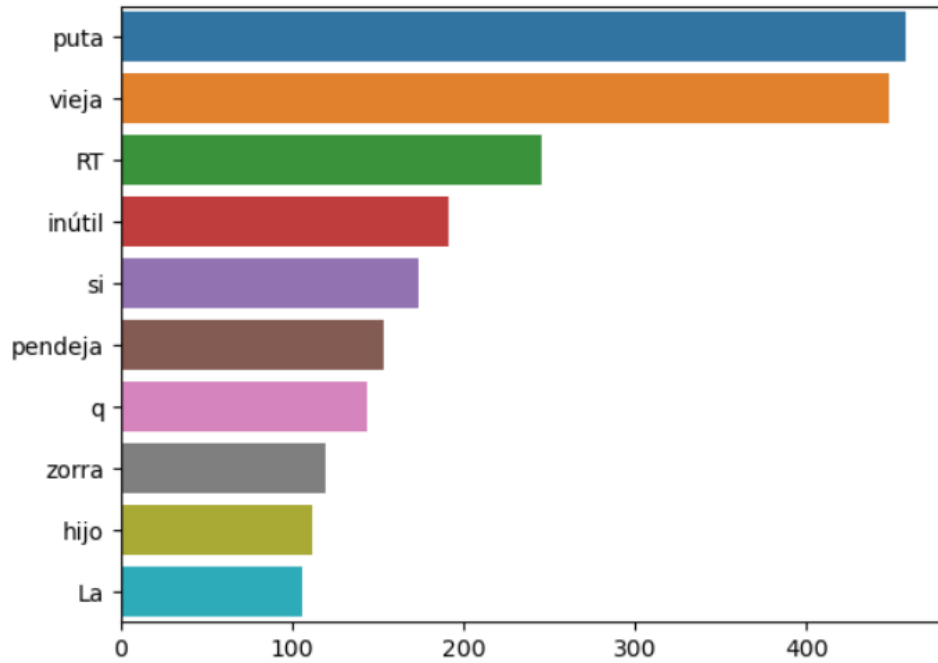
*Figura 10 longitud de caracteres en los elementos del conjunto de datos.*

En la Figura 16 se encuentra representada la longitud de las palabras que componen a cada elemento notando que la longitud promedio de las palabras oscila entre 3 y 12, siendo 5 la longitud más común presentando una frecuencia por encima de 350.



*Figura 11 longitud de palabras en los elementos del conjunto de datos.*

En la Figura 17 se destacan las 10 palabras más frecuentes en general en el contenido del conjunto de datos siendo la palabra 'puta' la más encontrada con una frecuencia superior a 400. También hacen presencia palabras como 'vieja', 'inútil', 'pendeja' y 'zorra' teniendo presencia en un buen número de elementos.



*Figura 12 top 10 de palabras recurrentes en el conjunto de datos.*

Para el siguiente caso (Figura 18) se observan las 25 palabras con más presencia en el conjunto de datos en donde se anexan 'vieja', 'putita', 'madre', 'mierda'.

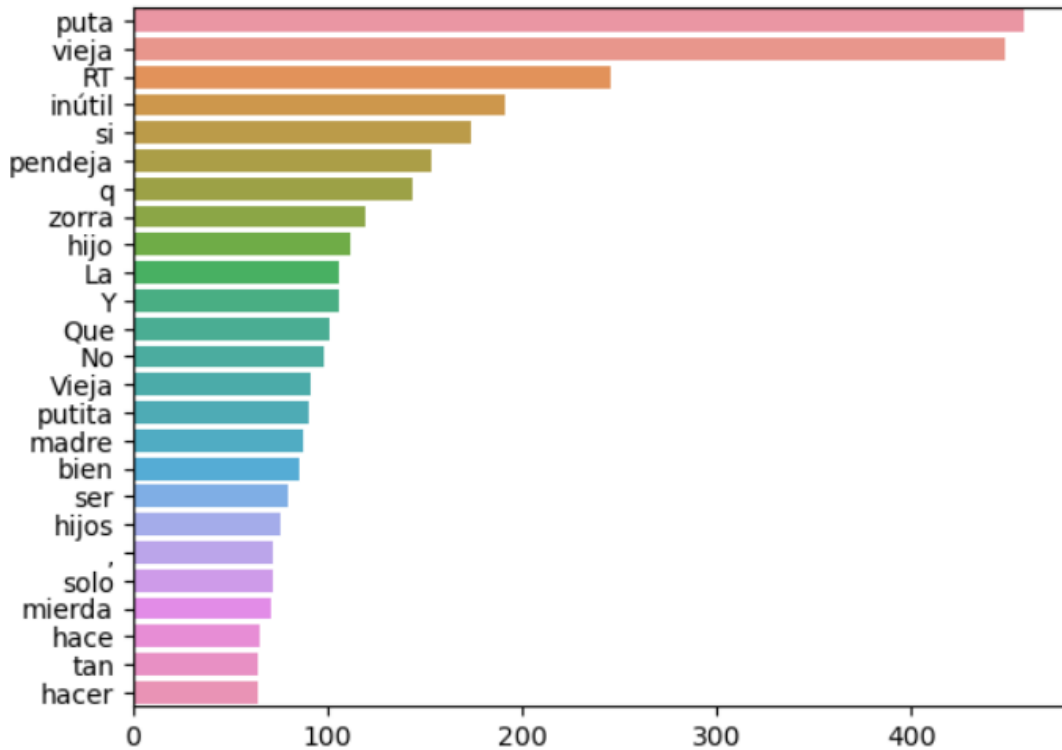


Figura 13 top 25 de palabras recurrentes en el conjunto de datos.

En la Figura 19 y Figura 20 se presentan las nubes de palabras que coinciden con las palabras con más frecuencia en los elementos del conjunto de datos.

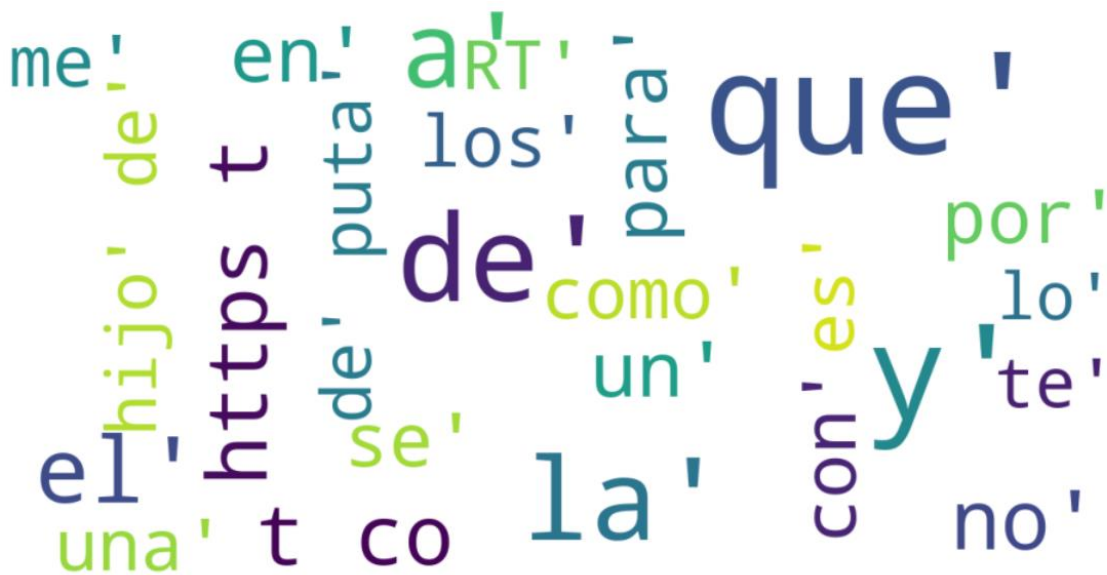
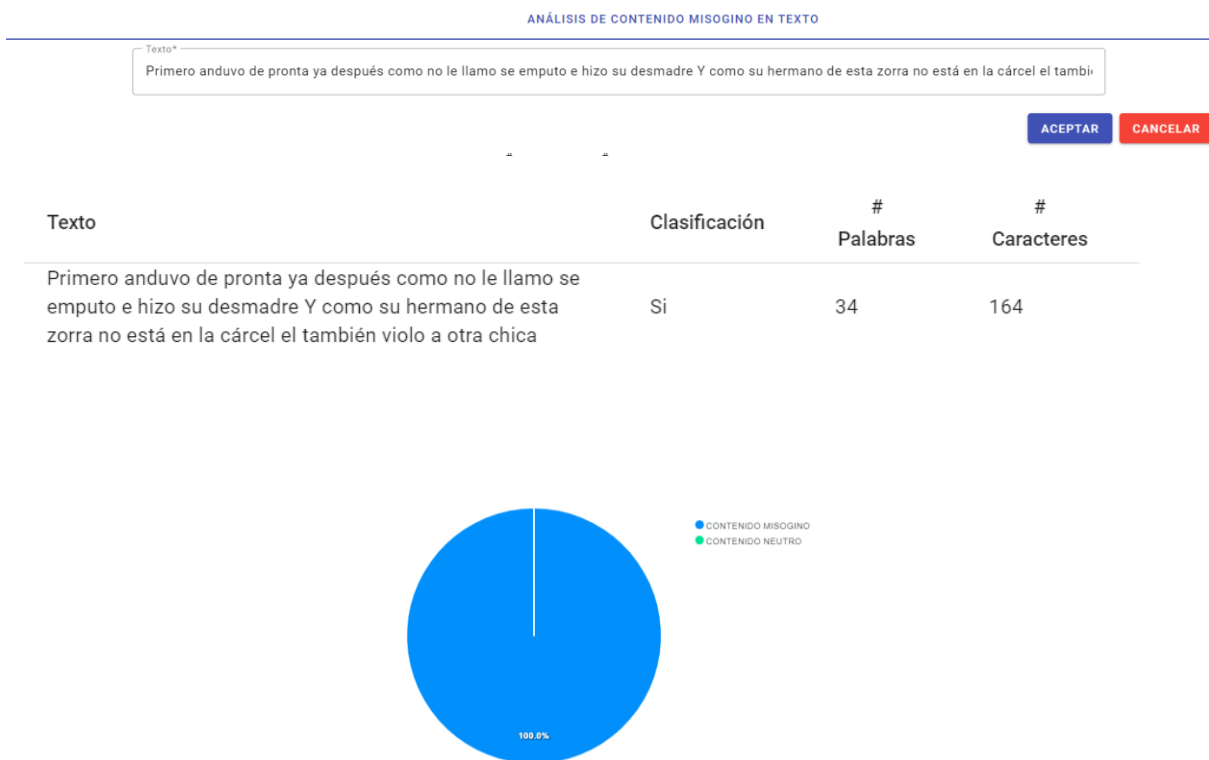


Figura 14 Nube de etiquetas de palabras recurrentes





presenta 34 palabras compuestas a su vez de 164 caracteres elevando la gráfica a un 100% del lado de contenido misógino.



*Figura 17 Análisis de texto en aplicación WEB 1*

Para el segundo comentario (Figura 23) se obtiene un resultado sin contenido misógino en el texto compuesto por 31 palabras y 168 caracteres balanceando la gráfica a partes iguales.

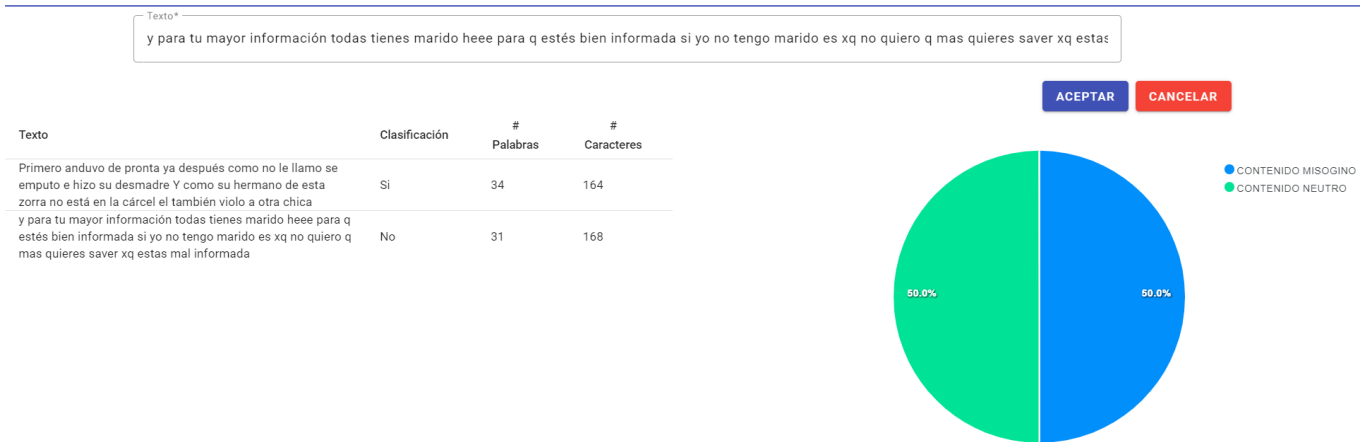


Figura 18 Análisis de texto en aplicación WEB 2

Para el tercer comentario (Figura 24) ingresado se obtiene contenido misógino al texto compuesto por 21 palabras y 137 caracteres.



Figura 19 Análisis de texto en aplicación WEB 3

El caso número 4 representado a continuación (Figura 25) arroja también detección de lenguaje misógino en el texto compuesto por 10 palabras y 55 caracteres.



Figura 20 Análisis de texto en aplicación WEB 4

Comentario número 5 (Figura 26) con 18 palabras 96 caracteres resultando en positivo a contenido misógino.

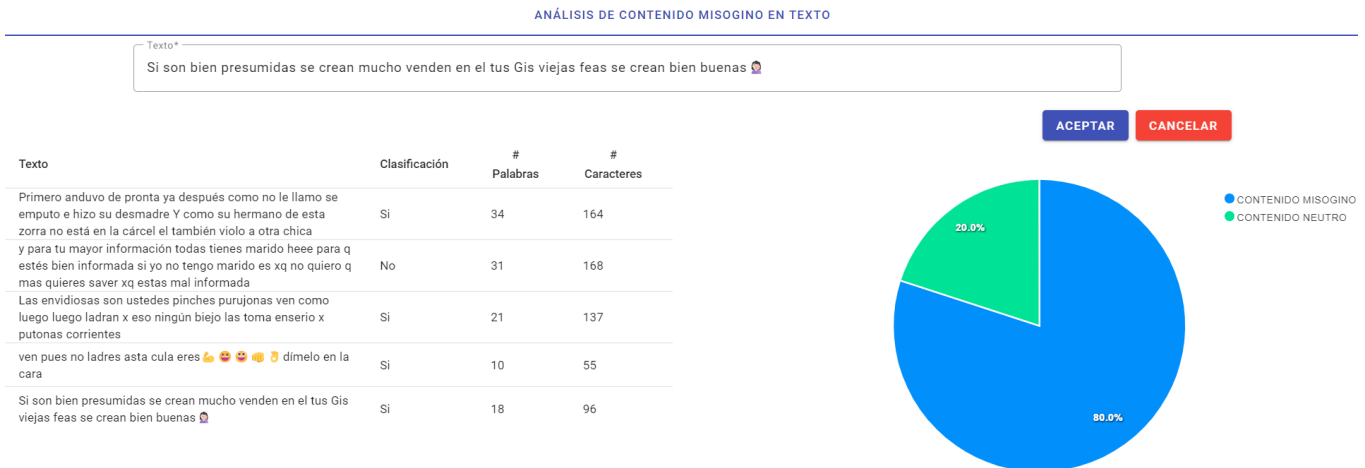


Figura 21 Análisis de texto en aplicación WEB 5

Comentario número 6 (Figura 27) con 11 palabras 71 caracteres resultando en negativo a contenido misógino.



Figura 22 Análisis de texto en aplicación WEB 6

Se repite el proceso durante 9 veces (Figura 28) adicionando así 4 comentarios más con las siguientes características:

- Comentario 6: 11 palabras 71 caracteres sin contenido misógino encontrado.
- Comentario 7: 32 palabras 162 caracteres sin contenido misógino encontrado.
- Comentario 8: 77 palabras 430 caracteres sin contenido misógino encontrado.
- Comentario 9: 32 palabras 182 caracteres sin contenido misógino encontrado.

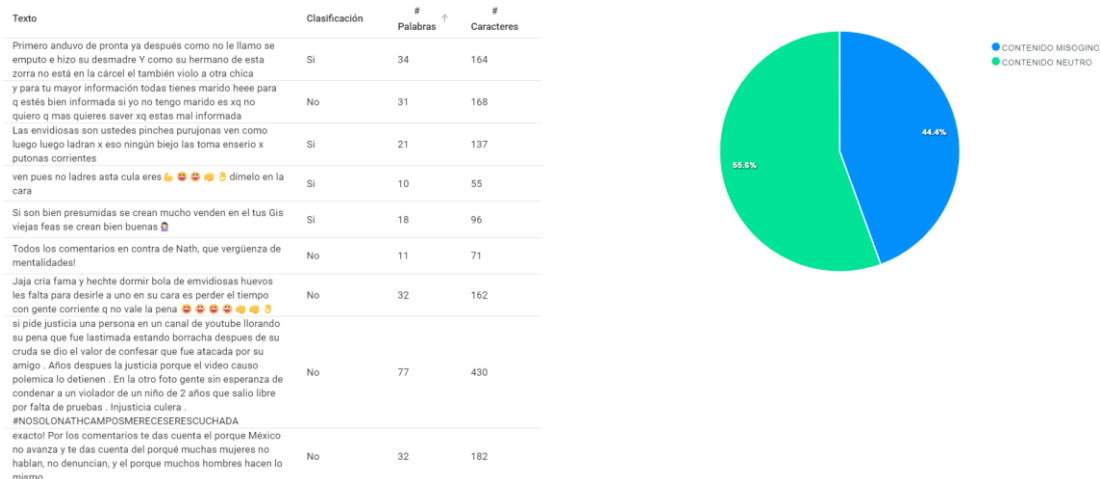


Figura 23 Análisis de texto en aplicación WEB 7

Al final se obtiene la evaluación para los 9 comentarios ingresados de los cuales suman un total de 266 palabras como se observa en la Figura 29 con 1465 lo cual nos da un promedio de 29 palabras y 163 caracteres por comentario.

Texto	Clasificación	# Palabras	# Caracteres
Primero anduvo de pronta ya después como no le llamo se emputo e hizo su desmadre Y como su hermano de esta zorra no está en la cárcel el también violo a otra chica	Si	34	164
y para tu mayor información todas tienes marido heee para q estés bien informada si yo no tengo marido es xq no quiero q mas quieras saver xq estas mal informada	No	31	168
Las envidiosas son ustedes pinches purujonas ven como luego luego ladran x eso ningún biejo las toma enserio x putonas corrientes	Si	21	137
ven pues no ladres asta cula eres 🙄😏👊🔥 dímelo en la cara	Si	10	55
Si son bien presumidas se crean mucho venden en el tus Gis viejas feas se crean bien buenas 🙄	Si	18	96
Todos los comentarios en contra de Nath, que vergüenza de mentalidades!	No	11	71
Jaja cría fama y hechte dormir bola de emvidiosas huevos les falta para desirle a uno en su cara es perder el tiempo con gente corriente q no vale la pena 🙄😏👊🔥	No	32	162
si pide justicia una persona en un canal de youtube llorando su pena que fue lastimada estando borracha despues de su cruda se dio el valor de confesar que fue atacada por su amigo . Años despues la justicia porque el video causo polemica lo detienen . En la otro foto gente sin esperanza de condenar a un violador de un niño de 2 años que salio libre por falta de pruebas . Injusticia culera . #NOSOLONATHCAMPOSMERECESERESCUCHADA	No	77	430
exacto! Por los comentarios te das cuenta el porque México no avanza y te das cuenta del porqué muchas mujeres no hablan, no denuncian, y el porque muchos hombres hacen lo mismo	No	32	182

*Figura 24 Resultado del análisis y clasificación de texto*

La Figura 30 representa que de los 9 comentarios sometidos a evaluación 5 de ellos fueron identificados como texto sin contenido misógino lo cual representa al 55.6% de los comentarios.

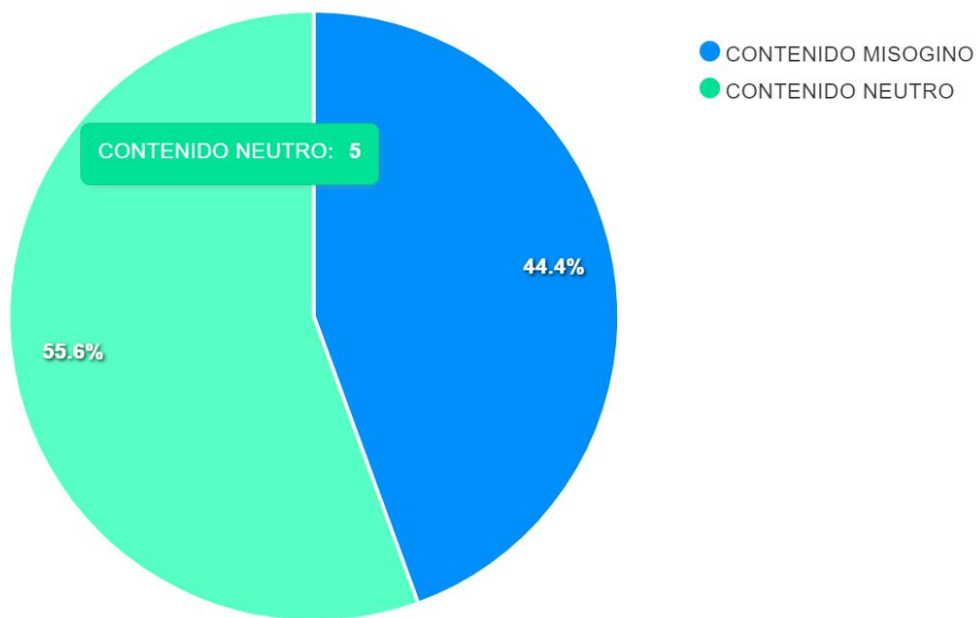


Figura 25 Representación de porcentaje sin contenido misógino

Mientras que los 4 comentarios restantes fueron catalogados con contenido misógino representando así al 44.4% de los comentarios totales como se ve en la Figura 31.

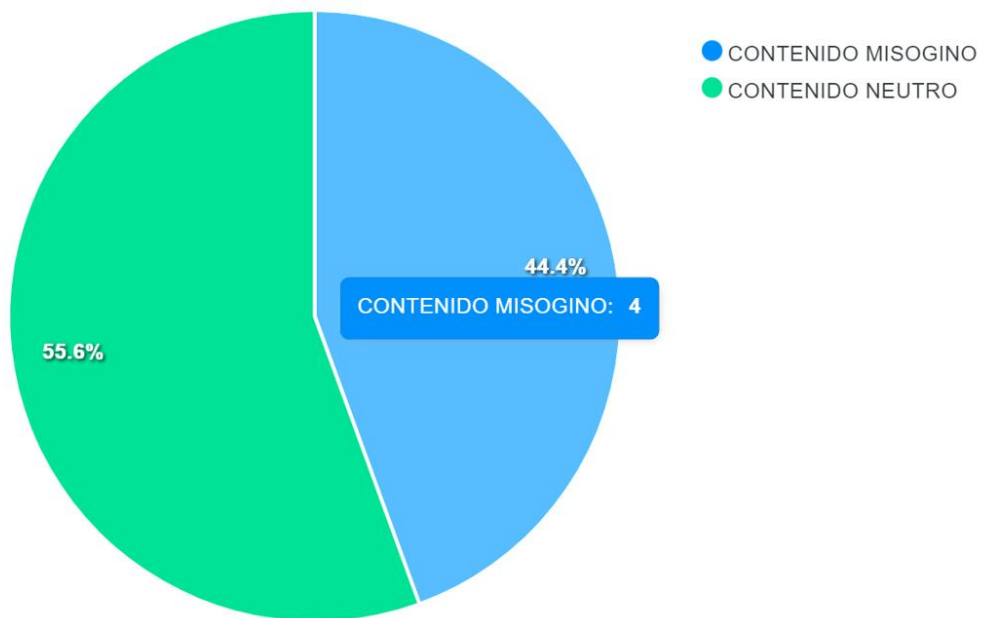


Figura 31 Representación de porcentaje con contenido misógino

# **CAPÍTULO V CONCLUSIONES**

## **Conclusiones**

Es importante mencionar que un trabajo de esta naturaleza debe contar con una selección de datos bien fundamentada es decir, entre más cuidadosa sea la selección y clasificación de los datos que usan durante el entrenamiento del modelo mejores resultados se obtienen en un entorno de despliegue real.

El acompañamiento de un especialista en el área también es fundamental sobre todo en el caso de análisis de lenguaje natural ya que en la etapa de etiquetar a cada uno de los elementos llegan a surgir dudas que de no ser resueltas podrían afectar al continuar con el proceso.

Como se puede observar en las etapas de prueba y despliegue se obtienen resultados favorables al someter a clasificación datos reales obtenidos de Facebook para lo cual en contraste de su clasificación a través del modelo se puede determinar que en ambos casos solo existe un comentario cuya evaluación resulta incorrecta dando así un 85% de efectividad en las pruebas ejecutadas en la etapa uno (exposición del modelo en servicio WEB) y un 88% para la etapa en pruebas a través de la aplicación WEB final.

## **Recomendaciones**

Si bien los resultados obtenidos fueron favorables y bien permiten la detección de contenido misógino en comentarios de Facebook, así como en tweets obtenidos de Twitter se notan algunas características que podrían ser clave para la obtención de resultados más precisos, a continuación, se describe cada una de estas.

### **5.1.1 Enriquecimiento del conjunto de datos**

Como se puede observar el modelo si es capaz de identificar el contenido objetivo en texto que es ingresado como entrada, pero de aquellos que formaron parte del margen de error se observa que su corrección podría influir en el hecho de poder



incluir dentro del conjunto de datos elementos que a pesar de contener palabras de connotación positiva llevan implícito un mensaje despectivo al respecto como se puede observar en la siguiente oración.

*"No cabe duda que eres una princesa hermosa pero una pijuja en la cama"*

Este tipo de oraciones generalmente muestran una mala clasificación por que son asociadas directamente con las palabras claves o que forman parte de la descripción del sujeto, en este caso "princesa, hermosa y puta" pasan a ser los adjetivos calificativos, pero con el hecho de tener un halago entre ellas su clasificación no logra detectar la parte despectiva de forma óptima.

### **5.1.2 Análisis automático en publicaciones de redes sociales**

La herramienta web creada permite la interacción de forma directa con el modelo y en los ejemplos pasados las pruebas se realizaron ocupando contenido extraído principalmente de Facebook de esta forma se iba viendo el resultado por comentario ingresado de forma manual la idea de esta mejora propone la integración de un módulo que a través del enlace web de una publicación permita la extracción de todos los comentarios que contenga sometiendo a cada uno de ellos a su clasificación y de esta forma tener un panorama más general sobre una publicación en concreto.

## REFERENCIAS

- INEGI. (2021). *ENCUESTA NACIONAL SOBRE DISPONIBILIDAD Y USO DE TECNOLOGÍAS DE LA INFORMACIÓN EN LOS HOGARES, 2020*.
- Pulido, M. B., Soto, Á. D., Lozano, F. M., Quintero Peña, W., & Grancolombiano, P. (2021). Redes sociales y relaciones digitales, una comunicación que supera el cara a cara. *Revista Internacional de Pedagogía e Innovación Educativa*, 1(1), 123–148. <https://doi.org/10.51660/RIPIE.V1I1.29>
- Chávez, C. F. (2018). *CIUDAD EDUCATIVA CIENCIAS DE LA INFORMACIÓN ¿SON EFICIENTES LAS REDES SOCIALES?* <https://canalbiblos.blogspot.com/2013/01/tipologia-de-las-redes-sociales.html>
- Gelbukh, A. (n.d.). Procesamiento de lenguaje natural y sus aplicaciones Related papers. *Artículo Invitado. Komputer Sapiens*.
- Lorenzo Copello, P., & Daunis Rodriguez, A. (2021). *Odio, prejuicios y derechos humanos*. Comares.
- Rothman, D. (2021). *Transformers for Natural Language Processing: Build innovative deep neural ... - Denis Rothman - Google Libros*. <https://books.google.es/books?hl=es&lr=&id=Cr0YEAAAQBAJ&oi=fnd&pg=PP1&dq=transformers+python&ots=a9p2XoZh31&sig=EIVTNqj1n0mwrS07fY-NVh8E5sQ#v=onepage&q=transformers%20python&f=false>
- Aluru, S. S., Mathew, B., Saha, P., & Mukherjee, A. (2020). *Deep Learning Models for Multilingual Hate Speech Detection*. <http://arxiv.org/abs/2004.06465>
- Calderón, C. A., Blanco-Herrero, D., & Apolo, M. B. V. (2020). Rejection and hate speech in Twitter: Content analysis of tweets about migrants and refugees in Spanish. *Revista Española de Investigaciones Sociológicas*, 172, 21–39. <https://doi.org/10.5477/cis/reis.172.21>
- Plaza-Del-Arco, F. M., Molina-González, M. D., Ureña-López, L. A., & Martín-Valdivia, M. T. (2020). Detecting Misogyny and Xenophobia in Spanish Tweets Using Language Technologies. *ACM Transactions on Internet Technology*, 20(2). <https://doi.org/10.1145/3369869>
- Tanase, M.-A., Zaharia, G.-E., Cercel, D.-C., & Dascalu, M. (2020). *Detecting Aggressiveness in Mexican Spanish Social Media Content by Fine-Tuning Transformer-Based Models*. [https://www.facebook.com/communitystandards/hate\\_speech](https://www.facebook.com/communitystandards/hate_speech).
- Watanabe, H., Bouazizi, M., & Ohtsuki, T. (2018). Hate Speech on Twitter: A Pragmatic Approach to Collect Hateful and Offensive Expressions and Perform Hate Speech Detection. *IEEE Access*, 6, 13825–13835. <https://doi.org/10.1109/ACCESS.2018.2806394>
- Bianchi, F. D. (2001). *Introducción a MATLAB*.
- Cerva Cerna, D., & Cerva Cerna, D. (2020). La protesta feminista en México. La misoginia en el discurso institucional y en las redes sociodigitales. *Revista Mexicana de Ciencias Políticas y Sociales*, 65(240), 177–205. <https://doi.org/10.22201/FCPYS.2448492XE.2020.240.76434>

- *Crisp-DM: los 6 pasos del proceso de Data Mining - Blog Smartup.* (n.d.). Retrieved December 1, 2022, from <https://blog.smartup.es/crisp-dm-6-pasos-proceso-data-mining/>
- *GeoINT - Research website.* (n.d.). Retrieved November 23, 2022, from <http://geoint.mx/site/publicacion/id/127.html>
- INEGI. (n.d.). *POBLACIÓN QUE UTILIZÓ INTERNET (Porcentaje).*
- Lopez, J. (n.d.). *Web scraping.* Retrieved November 23, 2022, from [www.programacion.net](http://www.programacion.net)
- *“Machine Learning”: definición, tipos y aplicaciones prácticas - Iberdrola.* (n.d.). Retrieved November 24, 2022, from <https://www.iberdrola.com/innovacion/machine-learning-aprendizaje-automatico>
- *Mujeres de 18 a 24 años tienen 27% más de probabilidad de ser víctimas de violencia en línea: Informe del MESECVI | Instituto Nacional de las Mujeres | Gobierno | gob.mx.* (n.d.). Retrieved December 1, 2022, from <https://www.gob.mx/inmujeres/articulos/la-cirberviolencia-y-el-ciberacoso-es-una-realidad-que-debe-ser-visibilizada-con-datos>
- Sergio, J., Efraín, S., & Farfán, M. (2014). *El arte de programar en R: un lenguaje para la estadística.* [www.imta.gob.mx](http://www.imta.gob.mx).

**ANEXOS I**

**5.1.3 Maqueta de Aplicación Web**

The wireframe consists of a rounded rectangular container. At the top, there are two horizontal input fields. The first is labeled 'TITULO' and the second is labeled 'VALOR DE ENTRADA'. Below these fields is a table with three columns: 'TEXTO', 'VALOR', and '#'. The table has three empty rows. To the right of the table are two buttons: a green 'ACEPTAR' button and a red 'CANCELAR' button. Further right is a pie chart divided into three segments: a blue segment labeled '1', a yellow segment labeled '2', and a green segment labeled '3'. To the right of the pie chart are three checkboxes, each followed by the text 'ITEM 1', 'ITEM 2', and 'ITEM 3' respectively.

TEXTO	VALOR	#

ITEM 1

ITEM 2

ITEM 3

## 5.1.4 Extracto de conjunto de datos de entrada para el entrenamiento del modelo

```
[
  {
    "Text": "@alex141176 Siliconesca putanesca guilota facilota jajaja",
    "Language": "es",
    "Value": 1
  },
  {
    "Text": "@Ary_vrey Quiere saber si todavía eres bien facilota y efectivamente por eso le aburres o todavía no envías nudes jajajaja",
    "Language": "es",
    "Value": 1
  },
  {
    "Text": "@CarLon_2020 No se, yo dudo que el adjetivo sea misogino. Más bien, suena a dos cosas: él se la cogio o él sabe que ella es facilota. Pensaria más bien que ofendio a los amigos \\/ padrinos de la doña y por eso lo sacan.",
    "Language": "es",
    "Value": 1
  },
  {
    "Text": "@coronel0204 Por ser tremenda golfa prostituta facilota perra hija de la perra mayor",
    "Language": "es",
    "Value": 1
  },
  {
    "Text": "@epigmenioibarra La botarga Xóchitl facilota da pena ajena! Ridícula y mentirosa! □□□",
    "Language": "es",
    "Value": 1
  },
  {
    "Text": "@Ernesto14836468 @DavidPramo2 Ve con la.chaira cherry,.ella te puede ayudar con lo que pides Esa.cherry chaira es una.facilota en todos los aspectos□□□",
    "Language": "es",
    "Value": 1
  },
  {
    "Text": "@EsdeProfugos En ese momento debió el de irse o al menos decir algo, de victima queda en patético asi ... ella queda como una facilota por dejársela barata jajaja",
    "Language": "es",
    "Value": 1
  },
  {
    "Text": "@FabiolaMDF Estas totalmente equivocada la facilota y mala esposa es la 'niña' traicionera ☹️👉👉👉👉 https:\\\\t.co\\YJNqBDwEd2",
    "Language": "es",
    "Value": 1
  },
],
```

```

{
  "Text": "@FelipeCalderon @SChecoPerez Miles de niños no podrán decir eso jamás 'Feliz Día del niño' gracias a tí y a tú maldita y absurda guerra, todo por legitimar una presidencia espuria. La vida te cobrará todo el daño que le hicieron a México tu y tú miserable familia junto con la de tú infiel y facilota mujer.",
  "Language": "es",
  "Value": 1
},
{
  "Text": "@HaffelyGalarza3 Pobre niña, que culpa tiene de tener una madre tan facilota, y quien sabe quien sea el padre",
  "Language": "es",
  "Value": 1
},
{
  "Text": "@InstaYeri @yerimua Igual no vales verga lo unico que sabes hacer es enseñar el culo y las tetas y andar de puta dando la panocha pinche facilota que por dinero te vendes hija de perra",
  "Language": "es",
  "Value": 1
},
{
  "Text": "@jenarovillamil @canalcatorcemx @MxPlusTV Hay chulis mejor vivorea de Bety la facilota de palacio con Marcelo, todavía está tibio el muerto y no respeta!!!!",
  "Language": "es",
  "Value": 1
},
{
  "Text": "@Luisfecalderonz Ay pobrecito de ti pinche moles. Después de todo el daño que hizo tu padrastro y la facilota de tu mamá, todavía quisieras haber hecho tus tranzas. Está cabron ser ustedes!!!! https://t.co/GdEKMv0Tzv",
  "Language": "es",
  "Value": 1
},
{
  "Text": "@Mr_Malvado ¿Ya no extrañas a Ochoa? Qué facilota eres 😊",
  "Language": "es",
  "Value": 1
},
{
  "Text": "No mi amor, tu no eres una bandida, eres una facilota",
  "Language": "es",
  "Value": 1
},
{
  "Text": "Nunca faltan los desesperados que se ligan lo que sea en cualquier lugar. Nunca falta la facilota en busca de atención. https://t.co/N1SBurrurx",
  "Language": "es",
  "Value": 1
},
{

```

```

    "Text": "Pobre Acnédres tendrá que volver a las chaquetas, se va su
fundita □ y con lo facilota que es no lo va a estar esperando afuera □
#ElElegido",
    "Language": "es",
    "Value": 1
  },
  {
    "Text": "Usar falda por el calor no me hace una fácilota, y tampoco
menos hombre",
    "Language": "es",
    "Value": 1
  },
  {
    "Text": "Yo creo q si paso, pero nmms se me fue una pregunta bn
facilota ☺",
    "Language": "es",
    "Value": 1
  },
  {
    "Text": " \/\Bien dice eso y siente más y más como esa concha facilota
le está masajeando cada cm de pija suya. ¿Acaso la muy trola quiere cagar
por la concha mestizos bastardos y quizás mal formados?, Que sucia madre
será en el futuro no muy lejano, por eso mismo lo eligió como su macho.
https:\\\/t.co\/9kgCgO3fC6",
    "Language": "es",
    "Value": 1
  },
  {
    "Text": "(Mi hermana suele decir que si yo hubiera sido una tía tendría
cinco o seis hijos, cada uno de un padre diferente, sería la más
borracha, inútil y golfa del barrio. Es asombroso lo bien que me
conoce.)",
    "Language": "es",
    "Value": 1
  },
  {
    "Text": "@abrahamendieta @Mzavalagc No eres más hijo de puta porque la
golfa que te malparió no es más alta.",
    "Language": "es",
    "Value": 1
  },
  {
    "Text": "@aguakatetwitter Diría que de hecho la tendencia es a más fea
más golfa",
    "Language": "es",
    "Value": 1
  },
  {
    "Text": "@AlbertoCaliu Pues nada, que no pase hambre ese culito, mmmm
te pillo en un callejón con 4 amigos y te haremos correrte como una
golfa. ¿No es lo que quieres? ¿No es lo que defiendes? SUBNORMAL.",
    "Language": "es",
    "Value": 1
  },
  {

```

```

    "Text": "@ALBERTOLUCA16 @MariettoPonce Otra vez..? No si bien dicen q
    esa doñita es una GOLFA.",
    "Language": "es",
    "Value": 1
  },
  {
    "Text": "@AlphariusFella @atomicjamoke @smorris837 @AlternatNews Cuando
    fue a Malvinas el gobernador Vernet encontró una foca que se parecía a tu
    madre, pero menos golfa, claro",
    "Language": "es",
    "Value": 1
  },
  {
    "Text": "@anadeaustria_ Esa es la típica golfa que quiere ser
    mantenida. Que la mantenga un tonto",
    "Language": "es",
    "Value": 1
  },
  {
    "Text": "@AnahRuiz1 Sin información no hay opinión posible y la
    información no la tienes. Anahí guapi, siempre ha habido viejas del
    visillo q han extraído sus conclusiones de la vecina golfa q iba con
    pantalones ajustados. Una pena que tú ya lo seas a tu edad y no esperes a
    estar aburrida.",
    "Language": "es",
    "Value": 1
  },
  {
    "Text": "@andyperbonie @carlosTROOP @ArturoVilla_ Solo una golfa con un
    culo tan grande y plano como una carpeta puede hacer ese comentario.",
    "Language": "es",
    "Value": 1
  },
  {
    "Text": "@AngMariscal4T La mayor golfa del chayote, sus servicios al
    mejor postor. Afortunadamente no nació en México.
    https://t.co/at9TCNIUsw",
    "Language": "es",
    "Value": 1
  },
  {
    "Text": "@ANIABELLO_R Quién le cree a una golfa con hambre??",
    "Language": "es",
    "Value": 1
  },
  {
    "Text": "@ANIABELLO_R Una golfa en busca de plata fácil, pruebas hay ?
    No , solo el testimonio de una vieja loca",
    "Language": "es",
    "Value": 1
  },
  {
    "Text": "@ArmandoDurn19 @guruchuirer Pues Beatriz, es igual a lo que
    dices cuando la esposa de obrador estaba casi moribunda la zopilota se le
    metió a López Obrador, como le llamarías a una mujer así,
    piruja, zorra, golfa, o ¿cómo?.",
    "Language": "es",

```



```

    "Value": 1
  },
  {
    "Text": "@ArturoVilla_ @Santi_ABASCAL Secretaria de estado? Esa es una
golfa criminal.",
    "Language": "es",
    "Value": 1
  },
  {
    "Text": "Felicidades sabemos q eres más Reyna q esa que compró la
corona pero abusada nunca te dejes de nadie",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Mimi aunque no tengas la corona tú eres la Reyna creo que la
donación que diste a esos niños es mejor que enriquecer a personas que
dicen que aran grandes cosas en Veracruz que hasta el sol de hoy no se a
visto nada de eso ☺",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Hermosa siempre!",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Así es, aunque la otra por sus astucia quedó, la del mejor
porte siempre me dije, eres tú Mimi, ánimo y no te interesen los haters",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Hermosa mujer... no sé la historia de ninguna de ellas pero
era MI FAVORITA...!",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Veracruz tenía una hermosa, educada, con porte, elegancia esa
Míriam Carballo",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Que Bonita muchacha. Ojalá la inscriban para mexicana
universal",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Queeee estas bien hermosa ... Pero también no te pases
jajaja",
    "Language": "es",
    "Value": 0
  },
  },

```

```

{
  "Text": "Regiaaaaaa ;;; me encanta tu porte y cuerpo super trabajado
☐",
  "Language": "es",
  "Value": 0
},
{
  "Text": "Es lo que merecíamos tener...mucho elegancia",
  "Language": "es",
  "Value": 0
},
{
  "Text": "hermosa te luciste espectacular .....",
  "Language": "es",
  "Value": 0
},
{
  "Text": "La verdad la verdad tu m pareces súper linda mimi . Se vienen
más cosas en tu vida . A seguir adelante",
  "Language": "es",
  "Value": 0
},
{
  "Text": "Muy hermosa Mimi no hagas caso a comentarios mediocres",
  "Language": "es",
  "Value": 0
},
{
  "Text": "eres la mas hermosa y lo seras en el carnaval jugaste limpio y
inteligente ♥",
  "Language": "es",
  "Value": 0
},
{
  "Text": "Felicidades mimi eres una reina, que se te resbalen los
comentarios ardididos de las ☐☐☐ para empezar cuando se a visto que sea
más bonito un chaneque que una jirafa?",
  "Language": "es",
  "Value": 0
},
{
  "Text": "Hay un futuro muy hermoso para ti",
  "Language": "es",
  "Value": 0
},
{
  "Text": "Felizidades ala princesa miimmii ya quisiera yo ser famosa
como tu",
  "Language": "es",
  "Value": 0
},
{
  "Text": "Cuando critiquen a Mimi ojalaaaa Y se vean a un espejo Y estén
mucho más linda que ella porque si no solo quedarán así ☐ ridiculaaas",
  "Language": "es",
  "Value": 0
}

```

```

},
{
  "Text": "!! Rebe cuando vas hacer un video como un update si se
resolvió lo de los bancos !! Imagínate creo que no abriría una cuenta por
aya 😊",
  "Language": "es",
  "Value": 0
},
{
  "Text": "Rebe se respeta tu decisión si no lo quieres decir es un no. Y
ya!\nPero hay personas tan.... Mal intencionadas",
  "Language": "es",
  "Value": 0
},
{
  "Text": "Es que el poderosisisimo bodorrio fue de impacto ♥",
  "Language": "es",
  "Value": 0
},
{
  "Text": "Desde que llegue a veracruz nunca he podido encontrarte en
plazas o en la calle, trosteeeee 😊",
  "Language": "es",
  "Value": 0
},
{
  "Text": "Ay no gente que no tiene nada que hacer tu tranqui estuvo
hermosa la boda",
  "Language": "es",
  "Value": 0
},
{
  "Text": "Yo esperando que salgan los tik Toks jajajajajaja",
  "Language": "es",
  "Value": 0
},
{
  "Text": "Ay no sin nombres no me gustan los chismes ☐",
  "Language": "es",
  "Value": 0
},
{
  "Text": "De que murieron los quemados de ardor envidia de la gente no
se ve tan mal",
  "Language": "es",
  "Value": 0
},
{
  "Text": "Orgullosa de ser kenini",
  "Language": "es",
  "Value": 0
},
{
  "Text": "Kenia lo está logrando ",
  "Language": "es",
  "Value": 0
}

```

```

},
{
  "Text": "Que llegue al primer lugar banda si se puedeeee",
  "Language": "es",
  "Value": 0
},
{
  "Text": "Me encantan sus canciones",
  "Language": "es",
  "Value": 0
},
{
  "Text": "Se convirtió en una reyna ",
  "Language": "es",
  "Value": 0
},
{
  "Text": "Ay no que hermosaaaaa♡♡♡",
  "Language": "es",
  "Value": 0
},
{
  "Text": "Mi Patrona ",
  "Language": "es",
  "Value": 0
},
{
  "Text": "ue llegue al primer lugar banda si se puedeeee",
  "Language": "es",
  "Value": 0
},
{
  "Text": "Me encantan sus canciones",
  "Language": "es",
  "Value": 0
},
{
  "Text": "kenia lo lograste!!!!!! ",
  "Language": "es",
  "Value": 0
},
{
  "Text": "la cantante y muy exitosa por cierto",
  "Language": "es",
  "Value": 0
},
{
  "Text": "La mejor de lo mejor ",
  "Language": "es",
  "Value": 0
},
{
  "Text": "Pero ni si quiera sabe entonar. Como quiere ser cantante ? Lo
que es una vloguer porque para eso no se necesita talento.",
  "Language": "es",
  "Value": 0
},

```

```

{
  "Text": "Si te quedo muy bonito. Como siempre",
  "Language": "es",
  "Value": 0
},
{
  "Text": "Te me haces una niña con cara linda",
  "Language": "es",
  "Value": 0
},
{
  "Text": "La más bella de todas ",
  "Language": "es",
  "Value": 0
},
{
  "Text": "Hermosa como siempre ",
  "Language": "es",
  "Value": 0
},
{
  "Text": "Eso que belleza nos les haga caso a esos comentarios te
maquillas hermoso ",
  "Language": "es",
  "Value": 0
},
{
  "Text": "Es muy bonita ",
  "Language": "es",
  "Value": 0
},
{
  "Text": "Soy tu fan y me encanta verte",
  "Language": "es",
  "Value": 0
},
{
  "Text": "Osea si xD pero no era manera de decirlo jaja",
  "Language": "es",
  "Value": 0
},
{
  "Text": "Hay Adela como que la .....",
  "Language": "es",
  "Value": 0
},
{
  "Text": "Toda la buena vibra para Adela Micha",
  "Language": "es",
  "Value": 0
},
{
  "Text": "Esta una diosaaaa! Bellísima, cada vez se supera más!",
  "Language": "es",
  "Value": 0
},
{

```

```

    "Text": "Está que me muero de lo perfecta ",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Yo quisiera tener el cuerpo de Kimberly Loaiza ",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Hermosa con i sin cirugía",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Ajjajajajaj y quien la critica por su cuerpo ? Si tiene un
cuerpaso que estupides ",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Wtf quien la critica, yo quiero ser ella",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Kim, no me gusta su trabajo, no me gusta su personalidad, no
consumo su contenido. Pero se ve hermosa",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Esta toda perfecta esta mujer ",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Se ve súper linda, aparte que es una bella persona siempre
hace donaciones a niños con necesidades, en Venezuela ayudo aún chico que
estaba construyendo una casita para unos niños en condición de calle y
ella nunca publicó sobre eso..! Dejemos aun lado la envida , el odio y la
burla debemos comenzar a cambiar esa mentalidad cada vez estamos peor
como sociedad",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Era muy muy bonita en verdad no necesitaba nada",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Que hermosa es, fan de Kim ",
    "Language": "es",
    "Value": 0
  },
  {

```

```

    "Text": "La neta quedo chida",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Ella se puede hacer lo quiera es su dinero",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Quienes te conocemos sabemos lo que vales. Opinan qué hacen aquí hablando si tanto les molesta están al pendiente de ella.",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "La más hermosa ",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Te veías muy bonita. Si no ganaste ntp, pensaste bien en que ese dinero te serviría para alguna inversión o algo y no te fuiste como gorda en tobogán metiendo un dineral que quizás luego pudieras arrepentírte... ",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Tu mi reina vete a concursos más grandes como mis universo , tú lo vales eres hermosa y inteligente.",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Muchas felicidades Mimmi, siempre será una Reina muchas bendiciones y que vengan nuevos éxitos",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "La Reyna para mí eres tú...por tu porte, belleza y calidad humana, lamentablemente los estándares de lo que es una Reyna han bajado mucho... saludos desde San Andrés Tuxtla.",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Que bonita Mimi sigue adelante...",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Bellisimaaa",
    "Language": "es",
    "Value": 0
  },
  },

```

```

{
  "Text": "Hermosa con o sin corona tu sigues siendo la misma Bella e
inteligente puedes llegar muy lejos mientras que otras solo asta que sus
seguidores se aburran y quedará en nada como antes",
  "Language": "es",
  "Value": 0
},
{
  "Text": "Bellísimo chamacona ",
  "Language": "es",
  "Value": 0
},
{
  "Text": "Primero dios ella sera nuestra reyna vamos mimi",
  "Language": "es",
  "Value": 0
},
{
  "Text": "Mimi debió haber ganado es las más bonita",
  "Language": "es",
  "Value": 0
},
{
  "Text": "Tu cuerpo,porte y belleza son para participar en verdaderos
certámenes y no en las que gana la que tiene más dinero... Hermosa te
luces con todo",
  "Language": "es",
  "Value": 0
},
{
  "Text": "que hermosa Mirian para mí tu eres una reina desde el primer
dia",
  "Language": "es",
  "Value": 0
},
{
  "Text": "Literal !! Si el carnaval se ganará por belleza ....Uff dejás
tiradas a todas ...ya eres una ganadora MUCHAS FELICIDADES ..",
  "Language": "es",
  "Value": 0
},
{
  "Text": "@DiosUneSusHijos: Di amén porque VIENE ALGO GRANDE PARA TÍ DE
PARTE DE DIOS. MAÑANA UNA HERMOSA NOTICIA TE DESPERTARÁ.",
  "Language": "es",
  "Value": 0
},
{
  "Text": "Como tú, ninguna; ni siquiera las rosas; como tú no hay otra
igual",
  "Language": "es",
  "Value": 0
},
{
  "Text": "Busqué una rosa para regalarte y por mucho que traté, fue
imposible encontrar una más bella que tú.",
  "Language": "es",

```



```

    "Value": 0
  },
  {
    "Text": "No hay flor como tú, imposible. Hay flores bonitas y hermosas,
pero como tú, ninguna tan bella",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Eres tú de esas mujeres maravillosas que conquistan sin faldas
cortas, sin grandes escotes. Solo con una sonrisa, tu mirada y tu forma
de ser.",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Ni las rosas contigo se comparan. Eres linda, encantadora,
eres única",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Me conquistaste con tu forma de ser y te adoro, por ser única
y original, por ser mujer.",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Tú eres maravillosa, así tal como eres. No tienes nada que
simular, nada que inventar para gustarme.",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Eres de esas mujeres que hipnotizan con su forma de ser. A mí
por supuesto, me tienes embrujado.",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Cada vez que te veo me dejas sin palabras, sin respiración y
sin aliento. Cada vez que me hablas, me enloquece y me fascina el sonido
de tu voz. Y si además me sonríes, entonces no se explicarte, todo lo que
por ti yo siento",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Si pudieras ganar dinero con los \"me gustas\" que te tengo,
te volvieras millonaria",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Tienes todo lo que me gusta, todo lo me hace falta y todo lo
que he soñado",
    "Language": "es",

```

```

    "Value": 0
  },
  {
    "Text": "Es tan lindo saber que tú existes y que me tocó vivir en este tiempo y este espacio para poder conocerte",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Como tú ninguna, para hacer otra como tú, hay que hacerla a la medida...",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Antes de volverte a mirar necesito comprarme un seguro de vida, no vaya a ser que me mate tu sonrisa...",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Eres de esa clase de personas, por las cuales a las estrellas se les piden deseos",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Solo con mirarte sé que nunca voy a sentir lo mismo por alguien más",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Eres la única persona a la que cuando veo, se me olvida el mundo.",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Eres ese mensaje que siempre me saca una sonrisa...",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Me miro en el espejo y hasta mi imagen está pensando en ti...",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Si solo verte y hablarte me enloquece, no quiero saber que me pasará cuándo pueda besarte.",
    "Language": "es",
    "Value": 0
  },
  {

```

```

    "Text": " Dices que soy un loco por quererte así. Eso no es nada,
locura es lo que tiene mi corazón por tenerte a ti!",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Por ti soy capaz de hacer cualquier locura, cualquier cosa,
menos dejarte de querer.",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Si te hace falta algún TE QUIERO, algún TE ADORO, sin reglas,
condiciones o algún loco enamorado que viva para ti, entonces piensa en
mí. Vivo esperando por ti.",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "No serás la chica más bonita del mundo, pero si eres la chica
que más bonito ha hecho mi mundo",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Eres eso a lo que todos le llaman mujer, a lo que yo
simplemente le llamo: bendición divina.",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "No se cómo lo nuestro empezó, pero sí sé que NO QUIERO que
termine nunca.",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "No miento si te digo que eres lo mejor que me ha pasado, lo
mejor que me pasa y lo mejor que quiero que me siga pasando.",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Cualquier persona puede hacerme feliz haciendo algo especial,
pero solo alguien especial puede hacerme feliz sin hacer nada... como lo
haces tú.",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Aunque tenga todo el dinero del mundo, si no tengo tu amor me
declaro en quiebra.",
    "Language": "es",
    "Value": 0
  },
  {

```

```

    "Text": "Al final me enamoré de ti y ahora vivir sin ti, es
imposible.",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": " Poco a poco caí en tu trampa y ahora me tienes encadenado a
ti y lo que es peor, es que no quiero salir.",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Te apareciste en mi vida de casualidad y ahora eres la causa
de mi felicidad.",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Por locuras como tú, hay locos como yo..!!",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "11 de cada 10 veces que te pienso, sonrío.",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Te conozco de memoria, porque no sales de mi mente...",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Contigo he tenido la historia de amor más bonita del mundo y
tú ni siquiera te has enterado...",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Creo que fuiste mi amor en otra vida, porque nada más verte mi
corazón comienza a latir.",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Cuando digo que busco un amor, no hablo de cualquier amor,
sino del tuyo..!!",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Te deseo de todo corazón lo mejor y lo mejor que te puede
pasar... soy yo.",
    "Language": "es",
    "Value": 0
  },
  {

```

```

    "Text": "Disculpa si me comporto como un imbécil, pero es que tus ojos
hipnotizan y vuelven loco.",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Si el propósito de la vida es el amor, entonces mi propósito
eres tú...",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": " Cómo no actuar como un imbécil, cuando tus ojos lo único que
hacen es hipnotizarme.",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Me gustas para que seas toda mi vida ¿Qué más puedo decirte
para que entiendas mi amor por ti?",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "No te prometo nada extraordinario, solo vivir por siempre de
ti enamorado.",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Todo comenzó con una mirada y una sonrisa, después vino un
\"me gustas\", seguido de un \"te quiero\", hasta llegar a un \"te amo\"
y a un \"no puedo vivir sin ti\"...",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Hasta las rosas siente celos, cuando pasas de mi mano.",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Tienes eso que todos los hombres gustan, que todos los hombres
buscan, el encanto irresistible de ser una mujer de verdad. Lo tienes
tú.",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Mujer eres rica y poderosa...",
    "Language": "es",
    "Value": 0
  },
  {
    "Text": "Con una sonrisa como la tuya, no necesitas ningún cosmético.",
    "Language": "es",
    "Value": 0
  }

```

```
},
{
  "Text": "Con una mirada como la tuya, no te hace falta una pistola, ni
un cañón.",
  "Language": "es",
  "Value": 0
},
{
  "Text": "Cuando estoy contigo siento que lo imposible se hace posible,
cundo te vas todo se hace difícil.",
  "Language": "es",
  "Value": 0
},
{
  "Text": "Primero eres un amigo, después una ilusión y ahora eres lo más
grande que llevo en el corazón.",
  "Language": "es",
  "Value": 0
}
]
```