Universidad Nacional Mayor de San Marcos
Universidad del Perú. Decana de América
Facultad de Ingeniería de Sistemas e Informática
Escuela Profesional de Ingeniería de Software

Macht: una aplicación basada en un modelo de análisis de sentimiento aplicado a la identificación de mensajes en español de testimonios de violencia de género en Twitter

TESIS

Para optar el Título Profesional de Ingeniero de Software

AUTOR

Ivonne Stephany SOLDEVILLA PACHECO

ASESOR

Dra. Nora Bertha LA SERNA PALOMINO

Lima, Perú

2022

Referencia bibliográfica

Soldevilla, I. (2022). *Macht: una aplicación basada en un modelo de análisis de sentimiento aplicado a la identificación de mensajes en español de testimonios de violencia de género en Twitter*. [Tesis de pregrado, Universidad Nacional Mayor de San Marcos, Facultad de Ingeniería de Sistemas e Informática, Escuela Profesional de Ingeniería de Software]. Repositorio institucional Cybertesis UNMSM.

Metadatos complementarios

Datos de autor	
Nombres y apellidos	Ivonne Stephany Soldevilla Pacheco
Tipo de documento de identidad	DNI
Número de documento de identidad	72861278
URL de ORCID	https://orcid.org/0000-0002-5561-8807
Datos de asesor	
Nombres y apellidos	Nora Bertha La Serna Palomino
Tipo de documento de identidad	DNI
Número de documento de identidad	07665297
URL de ORCID	https://orcid.org/0000-0002-4292-344X
Datos del jurado	
Presidente del jurado	
Nombres y apellidos	Luis Angel Guerra Grados
Tipo de documento	DNI
Número de documento de identidad	15644548
Miembro del jurado 1	
Nombres y apellidos	Rosa Menendez Mueras
Tipo de documento	DNI
Número de documento de identidad	10246770
Miembro del jurado 2	
Nombres y apellidos	Nora Bertha La Serna Palomino
Tipo de documento	DNI
Número de documento de identidad	07665297

Datos de investigación	
Línea de investigación	C.0.3.22. Ingeniería de Software
Grupo de investigación	No aplica
Agencia de financiamiento	Sin financiamiento
Ubicación geográfica de la investigación	Edificio: Universidad Nacional Mayor de San Marcos País: Perú Departamento: Lima Provincia: Lima Distrito: Cercado de Lima Latitud: -12.0545901 Longitud: -77.0833251
Año o rango de años en que se realizó la investigación	2019-2022
URL de disciplinas OCDE	Ingeniería de sistemas y comunicaciones https://purl.org/pe-repo/ocde/ford#2.02.04 Ciencias de la computación https://purl.org/pe-repo/ocde/ford#1.02.01 Informática y Ciencias de la Información https://purl.org/pe-repo/ocde/ford#1.02.00

Universidad Nacional Mayor de San Marcos

Universidad del Perú, DECANA DE AMÉRICA

Facultad de Ingeniería de Sistemas e Informática

Escuela Profesional de Ingeniería de Software

Acta de Sustentación Virtual de Tesis

Siendo las quince (15) horas del día 11 (once) del mes de noviembre de 2022, se reunieron en la sala virtual meet.google.com/efn-jfqz-xjq, presidido por el Mg. Luis Angel Guerra Grados, Mg. Rosa Menendez Mueras (Miembro) y la Dra. Nora Bertha La Serna Palomino (Miembro Asesor), para la sustentación virtual de la Tesis intitulada ***"MACHT: UNA APLICACIÓN BASADA EN UN MODELO DE ANÁLISIS DE SENTIMIENTO APLICADO A LA IDENTIFICACIÓN DE MENSAJES EN ESPAÑOL DE TESTIMONIOS DE VIOLENCIA DE GÉNERO EN TWITTER"****, por la Bachiller* ***Ivonne Stephany Soldevilla Pacheco,*** *para optar el Título Profesional de Ingeniero de Software.*

Acto seguido de la exposición de la Tesis, el Presidente invitó a la bachiller a dar respuesta a las preguntas establecidas por los Miembros del Jurado.

La bachiller en el curso de sus intervenciones demostró pleno dominio del tema, al responder con acierto y fluidez a las observaciones y preguntas formuladas por los señores miembros del Jurado.

Finalmente habiéndose efectuado la calificación correspondiente por los Miembros del Jurado, la bachiller obtuvo la nota de 19 (Diecinueve).

A continuación, el Presidente del Jurado, Mg. Luis Angel Guerra Grados, declara a la bachiller ***Ingeniero de Software****.*

Siendo las 15:55 horas, se levantó la sesión.

Mg. Luis Angel Guerra Grados
Presidente

Mg. Rosa Menendez Mueras
Miembro

Dra. Nora Bertha La Serna Palomino
Miembro Asesor

Universidad Nacional Mayor de San Marcos

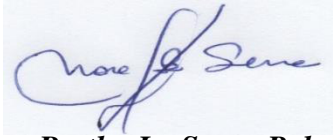
Universidad del Perú. Decana de América

Facultad de Ingeniería de Sistemas e Informática

Escuela Profesional de Ingeniería Software

INFORME DE EVALUACIÓN DE ORIGINALIDAD
Nº 015-EPISW-FISI-2022

1. *Autoridad Académica que emite el **I**nforme de Originalidad:*	*Directora de la Escuela Profesional de Ingeniería de Software*
2. *Apellidos y Nombres de la autoridad académica:*	*Dra. Nora Bertha La Serna Palomino*
3. *Operador del programa informático de similitudes:*	*Dra. Nora Bertha La Serna Palomino*
3. *Documento evaluado:*	*Tesis para Pregrado* *Título: "**Macht: Una aplicación basada en un modelo de análisis de sentimiento aplicado a la identificación de mensajes en español de testimonios de violencia de género en Twitter**"*
5. *Autores del documento:*	*Soldevilla Pacheco, Ivonne Stpehany*
6. Fecha de recepción de documento	*Recepción: 10/06/2022*
7. *Fecha de aplicación del programa detector de similitudes:*	*Revisión: 07/08/2022*
8. Software utilizado:	*Turnitin*
9. Configuración del programa detector de similitudes:	*Excluye textos entrecomillados: Sí* *Excluye biografías: Sí* *Excluye cadenas menores a 40 palabras: Sí* *Otro criterio (especificar): No*
10. Porcentaje de similitudes según programa detector de similitudes	*Diez por ciento (10%)*
11. Fuentes originales de las similitudes encontradas	*Se adjuntan en 02 (dos) fojas al presente informe*
12. Observaciones:	*Ninguna*
13. Calificación de originalidad *i. Documento cumple criterios de originalidad, sin observaciones* *ii. Documento cumple criterios de originalidad, con observaciones* *iii. Documento no cumple criterios de originalidad.*	*Documento cumple criterio de originalidad, sin observación*
14. *Fecha del Informe:*	*11/11/2022*



Dra. Nora Bertha La Serna Palomino

Directora (e) de la EPISW

AGRADECIMIENTOS

Dedico este trabajo a mi gata Asha la cual pacientemente espero por un nombre durante tres años, mismo tiempo que he tardado en terminar este proyecto.

Macht: Una aplicación basada en un modelo de análisis de sentimiento aplicado a la identificación de mensajes en español de testimonios de violencia de género en Twitter

RESUMEN

La violencia contra la mujer es un grave problema de salud pública y social. En el Perú, el número de casos que se eleva año tras año ha permitido que cada vez más sectores de la población perciban la importancia de resolver esta problemática. Por ello, el presente estudio busca construir una plataforma web capaz de clasificar mensajes en dos categorías: "La mujer pasó por un proceso violento" y "La mujer no pasó por un proceso violento", con la finalidad de realizar procesos de concientización más específicos que permitan fomentar la creación de espacios seguros. En estos espacios se buscaría que los testimonios de las víctimas sean escuchados, el brindar soporte emocional, enseñar a identificar signos de violencia en hogares y relaciones y brindar información acerca de las medidas tomadas contra la violencia a la mujer en el Perú.

La metodología aplicada considera la construcción de un conjunto de datos públicos con 1042 tweets en español etiquetados por 22 voluntarios. El modelo considera el proceso de ajuste a 3 modelos BERT pre-entrenados (SpanBERT, BETO, multilingualBERT), con los cuales se realizaron 2916 experimentos para encontrar el modelo con mejor desempeño, obteniendo un Área Bajo la Curva de 0.9349 y una precisión de 0.9043. La investigación aporta un nuevo dato público etiquetado en español, en 3 rangos de edad. Cualquier persona de cualquier parte del mundo podrá acceder a la aplicación y probar el rendimiento del modelo.

Macht: An application based on a sentiment analysis model to the identification of messages in Spanish with gender violence content on Twitter

ABSTRACT

Violence against women is a serious public and social health problem. In Peru, the number of cases rising year after year has allowed more and more segments of the population to perceive the importance of solving this problem. For this reason, the present study seeks to build a web platform capable of classifying messages into two categories: "The woman went through a violent process" and "The woman did not go through a violent process", in order to create specific awareness processes, that allow the possibility of promoting the creation of safe spaces. These spaces would seek to ensure that the testimonies of the victims are heard, to provide emotional support, to teach how to identify signs of violence in homes and relationships, and to provide information about the measures taken to mitigate violence against women in Peru.

The applied methodology considers the construction of a public dataset with 1042 tweets in Spanish tagged by 22 volunteers. The model considers the fine-tuning process to 3 pre-trained BERT models (SpanBERT, BETO, multilingualBERT), with which 2916 experiments were carried out to find the model with the best performance, obtaining an Area Under the Curve of 0.9349 and precision of 0.9043. The research provides new public data labeled in Spanish, in 3 age ranges. The resulting model can be accessed from anywhere by anyone in the world.

TABLA DE CONTENIDOS

Lista de Figuras

Lista de Tablas

Lista de Gráficos

CAPÍTULO 1: PLANTEAMIENTO DEL PROBLEMA

1.1. Antecedentes

Publicaciones realizadas por la OMS arrojan que al menos una de cada tres mujeres en el mundo (35%) han pasado por procesos de violencia física y/o sexual por parte de sus parejas o un tercero en un punto de su vida. (Organización Mundial de la Salud, 2017)

1.1.1. Violencia contra la mujer en el mundo

El informe *estimaciones mundiales y regionales de la violencia contra la mujer: prevalencia y efectos de la violencia conyugal y de la violencia no conyugal en la salud* publicado por la OMS (Organización Mundial de la Salud, 2013) y elaborado en conjunto con la Escuela de Higiene y Medicina Tropical de Londres y el Consejo Sudafricano de Investigaciones Médicas, reúne los datos 79 países, más dos territorios y realiza un examen sistemático sobre la prevalencia de violencia conyugal y no conyugal. Entre las conclusiones más destacadas se muestra que la prevalencia de violencia contra la mujer por parte de su pareja empieza con medidas muy altas a una de edad temprana de entre 15 y 19 años y alcanzan sus mayores porcentajes entre los 40 y 44. Ver Tabla 1.1.

RANGO DE EDAD	PREVALENCIA %
15-19	29.4
20-24	31.6
25-29	32.3
30-34	31.1
35-39	36.6
40-44	37.8
45-49	29.2
50-54	25.5
55-59	15.1
60-64	19.6
65-69	22.2

Tabla 1.1. Prevalencia de violencia contra la mujer por parte de la pareja según rango de edad (Organización Mundial de la Salud, 2013)

1.1.2. Impactos en la salud generados por la violencia contra la mujer

Entre los diferentes problemas de salud en los que puede decantar la violencia contra la mujer, encontramos enfermedades de transmisión sexual como: SIDA, Sífilis, Clamidia o Gonorrea. También abortos inducidos y diferentes problemas psicológicos como: Depresión, Alcoholismo, etc (Briere & Jordan, 2004).

El nivel de gravedad y prevalencia de estos varía de acuerdo con el contexto económico, religioso, cultural, etc. La OMS en conjunto con la OPS nos muestra un cuadro resumen de todos estos posibles impactos. Ver Tabla 1.2.

Físicas	Sexuales y Reproductivas
• Lesiones agudas como contusiones, cortes, laceraciones, heridas punzantes, quemaduras o mordeduras, así como huesos o dientes rotos. • Lesiones más graves que pueden provocar una discapacidad, como lesiones en la cabeza, los ojos, los oídos, el tórax o el abdomen. • Trastornos digestivos, problemas de salud a largo plazo o mala salud, incluido el síndrome de dolor crónico. • Muerte, por ejemplo, por suicidio o asociada con el SIDA.	• Interrupción del embarazo de manera segura o no. • Embarazo no deseado o no planeado. • Infecciones de transmisión sexual, dónde se incluye el VIH. • Complicaciones dentro de embarazo o aborto espontáneo. • Infecciones o hemorragias vaginales. • Infecciones en las vías urinarias, así como también en la zona pélvica crónica. • Desgarros entre la vagina y el recto o con la vejiga. • Relaciones sexuales con dolor. • Disfunción sexual.
Mentales	**Conductuales**
• Trastornos de los hábitos alimenticios y de sueño. • Trastornos de ansiedad y estrés. • Intentos de suicidio o autoagresiones graves. • Autoestima Baja. • Problemas de depresión.	• Uso desmedido de sustancias alucinógenas o alcohol. • Múltiples compañeros sexuales. • Elección nuevamente de parejas abusivas. • Porcentaje bajo de uso de condones u otros anticonceptivos.

Tabla 1.2. Problemas de salud producto de la violencia hacia la mujer (García-Moreno, Guedes, & Knerr, 2013)

Entre los más peligrosos efectos se encuentra la muerte. Se estima que de las 87,000 mujeres que fueron asesinadas globalmente en el 2017, más de la mitad (50,000) fueron matadas por sus parejas o miembros familiares. Esto quiere decir que 137 mujeres fueron asesinadas a diario por un miembro de su familia.

El estudio global de homicidio de la Oficina de Drogas y Crímenes de las Naciones Unidas del año 2019 (United Nations Office on Drugs and Crime, 2019) muestra, además, que un tercio (30,000) de los asesinatos a las mujeres en el 2017 fueron cometidos por su actual o expareja.

En este reporte, Asia ocupa el primero lugar de casos de feminicidios cometidos por parejas íntimas o miembros familiares por continente, sin embargo, si se toman las cifras de muertes por cada 100 000 habitantes de género femenino, África ocupa el primer lugar y América el segundo. Ver Figura 1.1.

Figura 1.1. Cantidad de feminicidios según continente por parte de parejas íntimas o miembros familiares (United Nations Office on Drugs and Crime, 2019)

1.1.3. Violencia contra la mujer en el Perú

El resumen estadístico publicado por el Programa Nacional para la Prevención y Erradicación de la Violencia contra las Mujeres e Integrantes del Grupo Familiar – AURORA muestra que solo en enero del año 2019 se registraron 14991 casos de violencia contra la mujer, violencia familiar y violencia sexual en los Centros de Emergencia Mujer (CEM) a nivel nacional, siendo 12575 los casos presentados por mujeres. Esto se puede visualizar en la Tabla N.º 1.3.

Mes	Total	Mujer	Hombre
Enero	**14,491**	12,575	1,916
Febrero	**12,941**	11,134	1,807
Marzo	**14,420**	12,433	1,987
Abril	**14,419**	12,380	2,039
Mayo	**15,259**	12,894	2,365
Junio	**14,804**	12,522	2,282
Julio	**15,334**	12,808	2,526
Agosto	**15,245**	12,954	2,291
Septiembre	**16,210**	13,881	2,329
Octubre	**16,289**	13,836	2,453
Noviembre	**16,240**	13,852	2,388
Diciembre	**16,233**	13,823	2,410
TOTAL	**181,885**	**155,092**	**26,793**
%	**100%**	**85%**	**15%**

Tabla 1.3. Casos atendidos por sexo según mes (Programa Naciona Aurora, 2019)

La cifra de casos reportados en los CEM con respecto al año 2018 presenta un aumento de 36% (Programa Naciona Aurora, 2019) esto podría deberse a la mayor difusión y crecimiento de los CEM a nivel Nacional, lo que ha logrado que una mayor cantidad de casos sean atendidos y que las cifras reales se estén transparentando.

Respecto a la variación porcentual, entre los años 2019 y 2020 (Ver Tabla N° 1.4). El origen de la estrepitosa caída (-37.1%) en el número de casos reportados a partir del mes marzo del presente año sería consecuencia del inicio de la cuarentena decretada en el país por el presidente Vizcarra a mediados de dicho mes, esto habría generado que muchas mujeres dejen de denunciar en los diferentes CEM, debido a que muchas veces es necesario realizar el proceso de manera presencial para una correcta atención y que solo los casos graves de violencia familiar o sexual atendidos mediante la Línea 100 son derivados a los CEM. Esta caída se corresponde con el aumento de llamadas a la Línea 100 que creció en un 97% respecto al año anterior (Ver Tabla N° 1.5).

Mes	2019	2020	
Ene	14,491	18,466	**27.4%**
Feb	12,941	17,181	**32.8%**
Mar	14,420	9,357	**-35.1%**
Abr	14,419	0	**-100.0%**
May	15,259	0	**-100.0%**
Jun	14,804	0	**-100.0%**
Jul	15,334	5,658	**-63.1%**
Ago	15,245	4,899	**-67.9%**
Set	16,210	7,582	**-53.2%**
Oct	16,289	17,539	**7.7%**
Nov	16,240	17,681	**8.9%**
Dic	16,233	16,132	**-0.6%**
Total	**181,885**	**114,495**	**-37.1%**

Tabla 1.4 Variación porcentual de los casos atendidos en los CEM del año 2020 en relación al año 2019 en cada mes (Programa Nacional Aurora, 2020)

La variación negativa más alta en casos atendidos (Ver Tabla N° 1.4) respecto al año anterior se obtuvo entre los meses de marzo y septiembre mientras que el aumento de llamadas a la línea 100 (Ver Tabla N° 1.5), más alta se dio entre los meses de mayo y octubre llegando a un pico de 190% en julio.

Mes	Años		Variación %
	2019	2020	
Ene	9,768	12,893	**32%**
Feb	10,054	13,753	**37%**
Mar	10,992	14,049	**28%**
Abr	10,274	16,037	**56%**
May	9,863	23,644	**140%**
Jun	10,039	24,072	**140%**
Jul	9,259	26,869	**190%**
Ago	9,212	24,990	**171%**
Set	9,624	24,744	**157%**
Oct	9,253	19,219	**108%**
Nov	9,993	17,948	**80%**
Dic	11,455	17,573	**53%**
Total	**119,786**	**235,791**	**97%**

Tabla 1.5. Variación porcentual de las consultas atendidas en la Linea100 (Programa Nacional Aurora, 2020)

Por otro lado, si bien el protocolo para la actuación de los CEM y comisarías cercanas o las comisarías de protección contra la violencia familiar especializadas ya se encuentra

creado en el marco de la Ley Nº30364[1] presentadas por la Dirección general contra la Violencia de Género ya se encuentra creado, (Dirección general contra la Violencia de Género, 2018) este no se cumple al pie de la letra y es desconocido por muchas de las victimas que se dirigen a las comisarías a presentar sus denuncias.

Tal como podemos ver en la Figura N.º 1.2 se involucra tanto a comisarías especializadas como no especializadas, la Dirección general contra la Violencia de Género en el protocolo de actuación conjunta de los Centros de Emergencia Mujer y comisarías o comisarias especializadas en materia de protección contra la violencia familiar de la policía nacional del Perú. Se especifica que todas las comisarías de la Policía Nacional del Perú, independientemente de la especialidad, están obligadas a recibir, registrar y tramitar de inmediato las denuncias verbales o escritas de actos de violencia (Dirección general contra la Violencia de Género, 2018). Sin embargo, esto en la práctica no se cumple, puesto que las comisarías no aceptan las denuncias y las mujeres al no conocer el protocolo desisten.

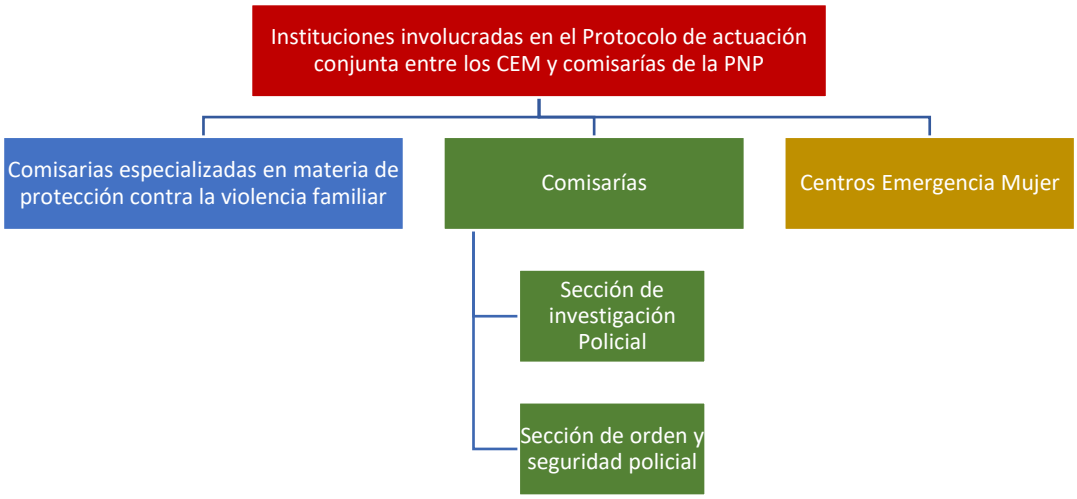


Figura 1.2. Instituciones involucradas en el Protocolo de actuación conjunta entre los CEM y comisarías de la PNP (Dirección general contra la Violencia de Género, 2018)

[1] Ley Nº 30364, se aprueba la Ley para prevenir, sancionar y erradicar la violencia contra las mujeres y los integrantes del grupo familiar, que tiene por objeto prevenir, erradicar y sancionar toda forma de violencia producida en el ámbito público o privado contra las mujeres por su condición de tales, y contra los integrantes del grupo familiar; en especial, cuando se encuentran en situación de vulnerabilidad, por la edad o situación física como las niñas, niños, adolescentes, personas adultas mayores y personas con discapacidad. (Ministerio de la Mujer y Poblaciones Vulnerables, s.f.)

1.1.4. Implicancias de los medios electrónicos en la Violencia a la Mujer.

El uso de medios electrónicos o redes sociales en los últimos años ha creado una nueva plataforma dónde muchas mujeres pueden compartir sus denuncias y experiencias, obteniendo en ciertos casos respaldo. Un ejemplo de esto es el movimiento tanto online como offline #NiUnaMenos que nació en Argentina pero que rápidamente se expandió a lo largo de todo Latinoamérica dónde miles de mujeres protestaron contra los casos de feminicidios.

Otro uso de estos medios es el de alerta o levantamiento de consciencia, ejemplo de ello es el hashtag #MiPrimerAcoso surgido en México como parte de una réplica de #PrimeiroAssedio en Brasil, dónde muchas mujeres contaron como fue el primer acoso sexual del que fueron víctimas.

Sin embargo, la tecnología también ha servido como medio para perpetuar la violencia. La Asociación para el Progreso de las Comunicaciones (APC) ha definido a la violencia de género en línea como "cualquier forma de violencia basada en género que se comete o se agrava, en parte o totalmente, por el uso de tecnologías de información y comunicación".

En el año 2018 Amnistía Internacional (International, 2018) como parte de una investigación, realizó una encuesta en 8 países: el resultado fue que el 23% de las mujeres encuestadas afirmaron haber experimentado abusos o acoso en internet.

Así también tal como señala el reporte de la situación de América Latina sobre la violencia de género ejercida por medios electrónicos (Organizaciones de América Latina, 2017), el Instituto de las Mujeres de la Ciudad de México, elaboró la investigación "Plan de acciones públicas para la visibilización y prevención de la violencia y el acoso sexual contra las mujeres en las redes sociales", donde se descubrió que la plataforma principal utilizada para promover campañas de odio contra mujeres y dónde se realizan difusiones de índole sexual es Twitter; por otro lado en Facebook es dónde se encontraron mayores ataques hacia mujeres que defienden sus derechos.

Lamentablemente en el caso del Perú, las cifras oficiales respecto a las distintas modalidades de violencia de género solo datan entre los años 2018 y mediados del 2021. En todo el año 2020 el 84% (616 casos) de víctimas fueron mujeres mientras que solo hasta el mes de agosto del año 2021 este mismo grupo representaba el 89% con 727 casos

(Observatorio Nacional de la Violencia contra las mujeres e integrantes del grupo familiar.)

1.1.5. Percepción de violencia hacia la mujer en el Perú.

El año 2019 el Instituto de Opinión Pública de la Pontificia Universidad Católica del Perú, por el encargo del Movimiento Manuela Ramos, realizó un sondeo de opinión ciudadana.

El principal objetivo fue "conocer las percepciones de la población respecto a la violencia contra las mujeres, en tres regiones del país: Ayacucho, Lima y Ucayali" (Perú, Pública, & Ramos, 2019, pág. 2)

En el tercer tópico respecto a la "Percepción del aumento de la violencia contra la mujer a nivel local", en las zonas tanto urbanas como rurales de las regiones de Ayacucho y Ucayali, más del 25% de los encuestados respondió que si considera que la violencia contra la mujer en su distrito aumentó (Ver Figura 1.3).

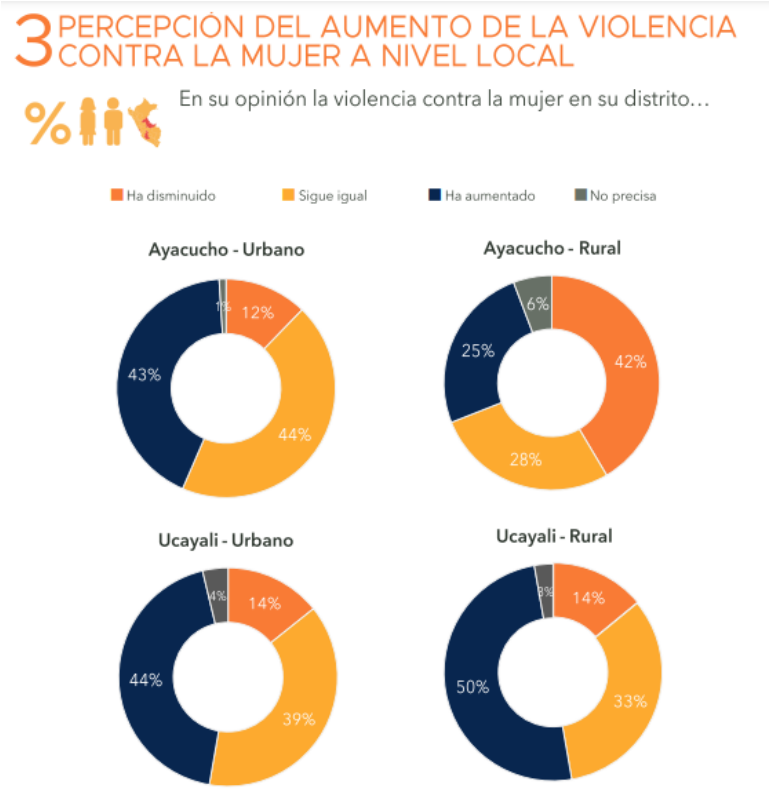


Figura 1.3. Percepción el aumento de la violencia contra las mujeres a nivel local en las regiones de Ayacucho y Ucayali, tanto a nivel urbano como rural (Perú, Pública, & Ramos, 2019)

Por otro lado, en Lima-Callao la percepción de aumento fue de un 30% frente a un 16% que cree que disminuyo (Ver Figura 1.4).

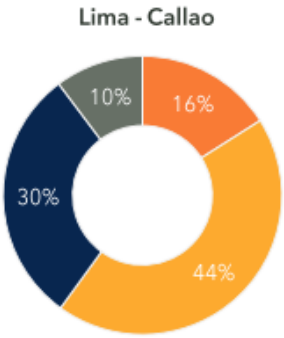


Figura 1.4 Percepción del amento de la violencia contra la mujer a nivel local en Lima - Callao (Perú, Pública, & Ramos, 2019)

Si ahondamos más respecto a las situaciones y/o hechos que implican violencia hacia la mujer tenemos que un 15% considera que el "No permitir que la pareja trabaje o estudie" no es violencia, a su vez también un 23% no considera como un hecho de violencia "Silbidos y piropos hacia las mujeres en la vía pública". (Ver figura 1.5)[1]

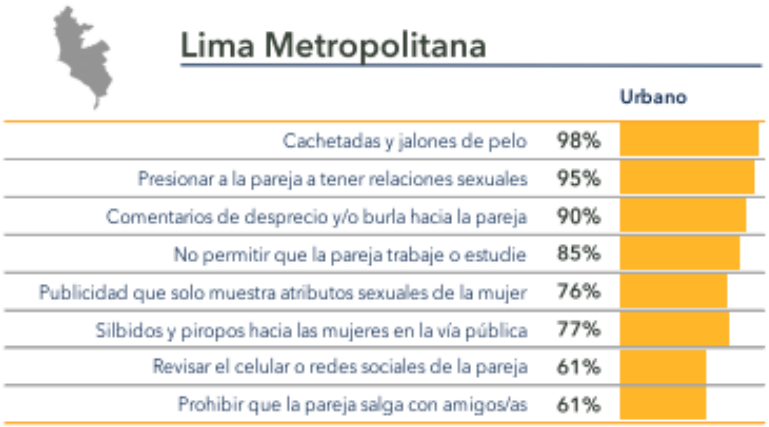


Figura 1.5. Percepción de situaciones que implican violencia contra la Mujer en Lima Metropolitana (Perú, Pública, & Ramos, 2019)

Finalmente, en cuanto a las actitudes frente a la violencia contra la mujer en une relación de pareja; más de un 60% en todas las regiones encuestadas tanto a nivel urbano como rural afirma en que están de acuerdo en que "los problemas de violencia entre una pareja son un asunto que solo deben resolver entre ambos", además un preocupante 20% de mujeres y 23% hombres en la zona rural de Ayacucho considera que "Hay ocasiones en las que las mujeres merecen ser golpeadas" (Ver Figura 1.6).

[1] El porcentaje de amarillo corresponde a los que consideraron que "Sí son hechos de violencia"

6. Estaría de acuerdo o en desacuerdo con las siguientes afirmaciones...

Porcentaje **"De acuerdo/Totalmente de acuerdo"**

	Lima		Ayacucho				Ucayali			
	Urbano		Urbano		Rural		Urbano		Rural	
	Hombre	Mujer	Hombre	Mujer	Hombre	Mujer	Hombre	Mujer	Hombre	Mujer
Los problemas de violencia entre una pareja son un asunto que solo deben resolver entre ambos	61.5	58.5	75.0	75.0	87.7	78.0	82.8	74.3	73.5	69.0
Sólo un hombre mentalmente enfermo es capaz de golpear a su pareja	72.0	71.0	60.0	56.5	52.8	57.1	74.7	76.2	71.0	69.5
Revisar el celular de la pareja es una forma de control	46.5	49.5	36.0	41.5	51.8	55.6	48.0	47.5	57.0	51.0
Los celos son una demostración de amor	11.5	4.0	19.5	9.0	26.7	10.7	23.2	12.4	25.0	18.5
Hay ocasiones en las que las mujeres merecen ser golpeadas	4.0	2.0	6.5	7.5	16.4	18.0	16.7	9.4	23.0	20.0
Una mujer debe tolerar que su pareja la golpee para mantener la familia unida	2.0	0.5	3.5	4.5	9.2	11.2	7.6	5.9	10.5	8.0

Figura 1.6. Actitudes frente a la violencia contra la mujer en une relación de pareja (Perú, Pública, & Ramos, 2019)

1.2. Problemas

1.2.1. Problema General

¿La construcción de una herramienta capaz de clasificar mensajes de redes sociales como Twitter en categorías “La mujer pasó por un proceso violento” y “La mujer no pasó por un proceso violento" serviría para visibilizar la violencia contra la mujer?

1.2.2. Problemas Específicos

Los problemas específicos de la presente tesis son los siguientes:

(a) ¿El diseño e implementación de un modelo de aprendizaje profundo permite clasificar testimonios sobre violencia en redes sociales como Twitter en las categorías de “La mujer pasó por un proceso violento” y “La mujer no pasó por un proceso violento”?

(b) ¿El diseño de una herramienta web que permita el uso del modelo, facilitará la clasificación y detección de testimonios de violencia?

(c) ¿La construcción de un dataset en español de 1042 mensajes de la red social de Twitter puede recoger la percepción sobre testimonios de violencia?

1.3. Objetivos

1.3.1. Objetivo General

Construir una herramienta capaz de clasificar mensajes de redes sociales como Twitter en categorías de “La mujer pasó por un proceso violento” y “La mujer no pasó por un proceso violento” que sirva para visibilizar la violencia contra la mujer..

1.3.2. Objetivos Específicos

Los objetivos específicos de la presente tesis son los siguientes:

(a) Diseñar e implementar un modelo de aprendizaje profundo que permita clasificar testimonios sobre violencia en redes sociales como Twitter en las categorías de “La mujer pasó por un proceso violento” y “La mujer no pasó por un proceso violento”.

(b) Diseñar una herramienta web que permita hacer el uso del modelo, de tal manera que facilite la clasificación y detección de testimonios de violencia.

(c) Construir un dataset en español de 1042 mensajes de la red social de Twitter, a partir de la traducción al idioma español del dataset propuesto por Schrading (Schrading, 2015), que recoja la percepción sobre testimonios de violencia.

1.4. Justificación

(a) El Perú es un país dónde la percepción de violencia hacia la mujer ha ido en aumento a través de los años. En el año 2016 con la encuesta nacional IOP PUCP “Roles y Violencia de Género” la percepción de que la violencia “Había aumentado mucho” en el país pasó de un 67.5% a un 76% entre los años 2012 y 2016 (Instituto de Opinión Pública de la Pontificia Universidad Católica del Perú, 2016, pág. 22).

(b) A partir de los resultados obtenidos de acuerdo con los grupos de edad planteados en la presente investigación se puede realizar un proceso de concientización más específico. Esto para fomentar la creación de espacios seguros dónde puedan ser escuchados testimonios de víctimas, se brinde soporte emocional, se enseñe a identificar signos de violencia en hogares y relaciones. Así como también se brinde información acerca de medidas tomadas contra la violencia a la mujer en el Perú.

(c) El dataset construido representa un gran aporte que puede ser utilizado para posteriores estudios que aborden la temática de violencia o percepción de

violencia hacia las mujeres en el Perú debido a que cuenta con diferentes grupos de edad y género.

(d) A partir de esta herramienta se puede realizar un análisis de la situación actual de percepción de violencia en el Perú en la red social de Twitter con el uso de la API que proporciona.

1.5. Alcance

El producto a obtener con la presente tesis es un aplicativo web capaz de realizar la clasificación de testimonios recogidos de la red social Twitter con la finalidad de dividirlos en las categorías “La mujer pasó por un proceso violento” y “La mujer no pasó por un proceso violento”.

1.6. Organización de la tesis

La presente tesis se encuentra organizada en 6 capítulos, los cuales se mencionan a continuación.

El primer capítulo presenta los antecedentes a partir de los cuales nace el planteamiento del problema y se plantean los objetivos a cumplir.

En el segundo capítulo se aborda el marco teórico, dónde se muestran conceptos relativos al problema, así como también los conceptos y definiciones de las tecnologías a utilizar.

En el tercer capítulo, se realiza una revisión exhaustiva de trabajos de investigación relacionados con la clasificación de textos, análisis de sentimientos y predicción mediante Aprendizaje Profundo.

En el cuarto capítulo, se realiza la definición del modelo de aprendizaje profundo propuesto y la metodología a seguir para su implementación, así como también se detalla la herramienta que permitirá la clasificación y predicción.

En el quinto capítulo, se detalla la implementación de la solución y el entrenamiento de los modelos de aprendizaje profundo que fueron propuestos.

Finalmente, el capítulo 6 recoge las conclusiones obtenidas durante la presente investigación, así como también los posibles trabajos futuros a realizar.

1.7. Propuesta

Se propone el desarrollo de una herramienta de apoyo para la clasificación de textos basado en un modelo BERT, el cual deberá ser capaz de clasificar mensajes testimoniales o no testimoniales de violencia hacia la mujer a partir de un texto ingresado de forma manual o el ID de un tweet. El esquema de la herramienta planteada se presenta en la Figura 1.7.

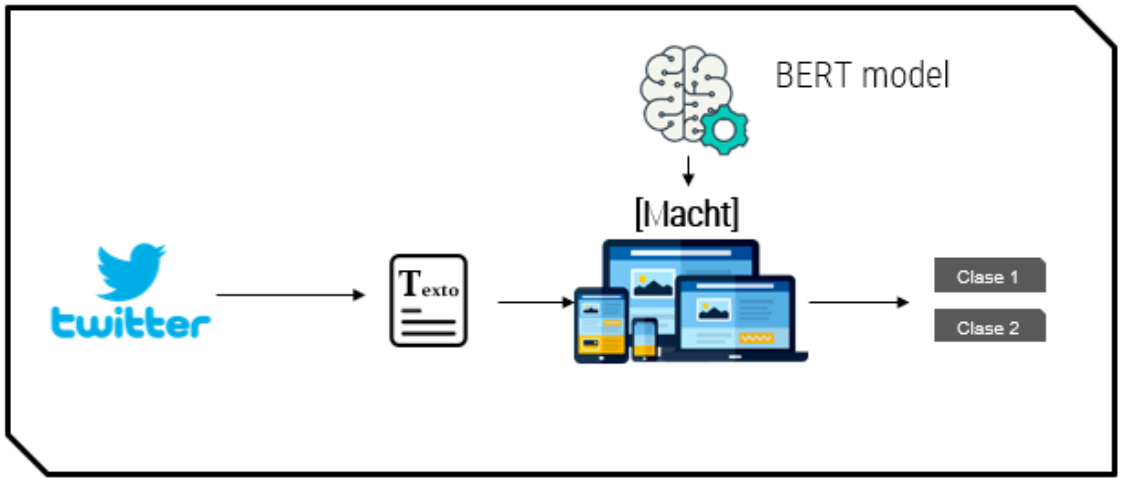


Figura 1.7. Esquema general de la propuesta

CAPÍTULO 2: MARCO TEÓRICO

En este capítulo se abordan conceptos teóricos relacionados al problema planteado en la primera sección de la presente tesis. Es así como se definen términos cómo feminismo, violencia a la mujer, feminicidio, entre otros. Se hace hincapié especialmente en la información de estudios que buscan responder la pregunta de porque se origina esta problemática en el Perú, cuántas modalidades de esta existen y porque en muchos casos las mujeres no buscan ayuda a pesar de ser víctimas de dicha violencia.

Por consiguiente, en la segunda sección del capítulo se hace una revisión de los conceptos y tecnologías a usar en la construcción de la solución. La tecnología revisada más importante es el modelo de BERT, del cual se basa la propuesta planteada en el capítulo previo.

2.1. Conceptos complementarios relativos al problema

2.1.1. Feminismo

La definición que le asigna la Real Academia Española es ***"Principio de igualdad de derechos de la mujer y el hombre".*** Respecto a las diferentes olas del feminismo, a cada una de ellas se le ha asignado una definición diferente.

2.1.2. Violencia a la Mujer

La definición de violencia a la mujer de acuerdo a las Naciones Unidas, es escrita como ***"todo acto de violencia de género que resulte, o pueda tener como resultado un daño físico, sexual o psicológico para la mujer, inclusive las amenazas de tales actos, la coacción o la privación arbitraria de libertad, tanto si se producen en la vida pública como en la privada".***

2.1.3. Feminicidio

La definición que le asigna la Real Academia Española es *"**Asesinato de una mujer a manos de un hombre por machismo o misoginia**"*. Esto se corresponde con la definición de Rusell (Russell, 1992) "***se trata del asesinato de mujeres por hombres motivados por el odio, el desprecio, el placer o la suposición de propiedad sobre las mujeres***"

2.1.4. Modalidades de violencia basada en género en el Perú.

La violencia basada en género como menciona el Ministerio de la Mujer y Poblaciones Vulnerables (Ministerio de la Mujer y Poblaciones Vulnerables, 2015) tiene una naturaleza constante, múltiple y generalizada, además se encuentra presente en diversos espacios de la vida social. Johan Galtung, propone para esto un modelo triangular de la violencia. En el cual observamos cuatro tipos de violencia.

1. Violencia Directa: Esta primera es la más visible, y para Magallón [1]en el caso de las mujeres se ejerce contra sus derechos de sobrevivencia, de identidad, de bienestar y de libertad, a través del feminicidio, el maltrato, el desprecio, el acoso, la alienación identitaria proveniente de los modelos hegemónicos de feminidad, la ciudadanía de segunda categoría y la sistemática negación de derechos y de opciones y elecciones de vida para las mujeres.
2. Violencia Estructural: De Magallon también conocemos que la violencia estructural se vincula a lo económico. Algunas de las expresiones más evidentes son; el trato desigual para el acceso a la vivienda u otros tipos de propiedad, salarios disparejos a pesar de igualdad de experiencia y capacidad, acceso diferenciado a posiciones de toma de decisiones y poder, trabajo sexual asociado a mujeres y feminización de la pobreza.
3. Violencia Cultural: Cumple la función de legitimar a las otras dos formas de violencia mencionadas previamente. Aquí se muestra como se ha encasillado a la mujer en contraposición al hombre en los espacios de cuidado familiar y no en un mundo de producción, racionalidad, creación y transformación de cultura a los que estos segundos se asocian. Debido a esto, a los hombres se les reconoce además capacidades para participar en espacios de política, arte y ciencia dónde se suele ostentar más reconocimiento social y prestigio.
4. Violencia simbólica: La diferencia entre los sexos tiene la naturaleza de una institución que delimita lo subjetivo de las estructuras mentales, así como el objetivo de las estructuras sociales, al punto de que la superioridad masculina no es necesaria de justificar, sino que se expresa y refuerza en todos los niveles de forma permanente mediante los discursos y costumbres.

[1] Véase MAGALLÓN PORTOLÉS, CARMEN (2005). Epistemología y violencia. Aproximación a una visión integral sobre la violencia hacia las mujeres.

El triángulo planteado por Johan Galtung, lo podemos observarlo en la figura 2.1.

Figura 2.1. Triángulo de Galtung (Ministerio de la Mujer y Poblaciones Vulnerables, 2015)

En el Marco Conceptual para las políticas públicas y acción del estado (Ministerio de la Mujer y Poblaciones Vulnerables, 2015) se hace mención a la existencia de distintas modalidades de violencia basada en género que se dan a lo largo de nuestra región. Un grupo de estas tienen un proceso de intervención más favorable, debido a que se encuentran legisladas y cuentan con registros sistemáticos. Adicionalmente presentan un análisis de las modalidades más comunes de violencia y de las que se tiene información sobre su nivel de prevalencia. Véase figura 2.2.

Figura 2.2 Modalidades de Violencia presentes en el Perú (Ministerio de la Mujer y Poblaciones Vulnerables, 2015)

2.1.5. ¿Por qué las mujeres víctimas de violencia de pareja en el Perú no buscan ayuda?

Si bien con la creación de los CEM las tasas de denuncias aumentaron considerablemente, aún existe una gran cantidad de mujeres que no buscan ayuda de ningún tipo de institución pública o privada. Esto se da por diversos motivos: Vergüenza, miedo a represalias, etc.

Jhon Ortega parte del grupo de GRADE[1]en su estudio refiere que, en el último año, de cada 3 mujeres víctimas de violencia severa, 1 de ellas no buscó ayuda en ninguna institución ni comento a amigos ni familiares, por otro lado, el porcentaje que no llega a denunciar a su agresor es de más del 60 %. En el caso de las áreas rurales la situación empeora: ya que la tasa de no denuncias llega a sobrepasar el 70%. Esta tasa de no denuncias (Ver Figura 2.3), considera a las víctimas que no acudieron a reportar la situación de violencia a las distintas instituciones del Estado, dónde se incluyen: Defensorías del Niño y Adolescente (DNA), comisarías, Ministerio de la Mujer y Poblaciones Vulnerables y Defensoría del Pueblo. ((GRADE), 2019)

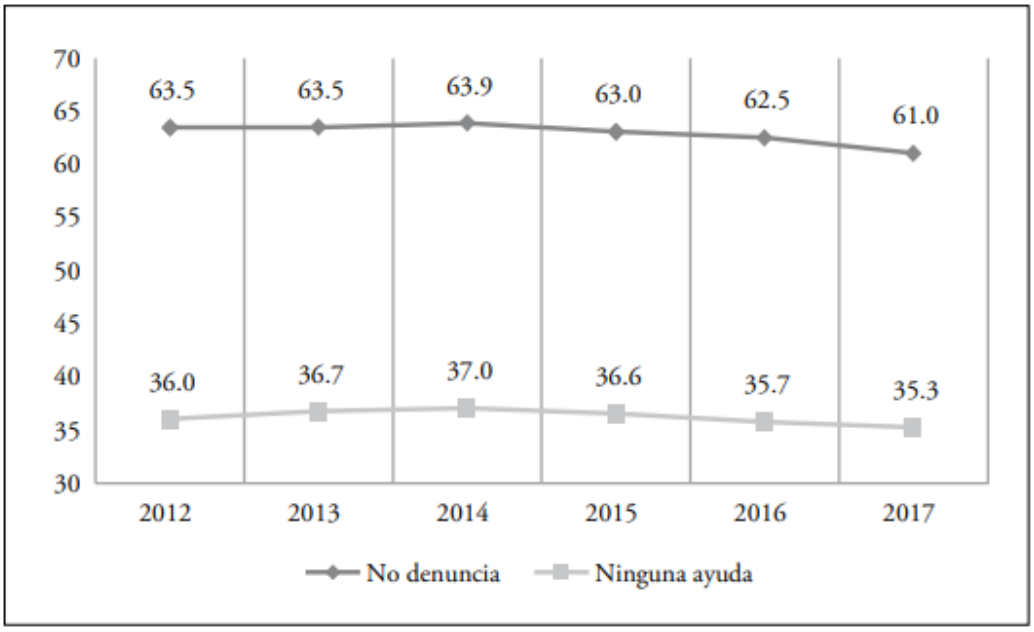


Figura 2.3 Porcentaje de víctimas de violencia física severa que buscan ningún tipo de ayuda ni realizan denuncia ((GRADE), 2019)

[1] Grupo de Análisis para del Desarrollo

En cuanto a las razones de no denuncia. Durante último año de la muestra, la respuesta con más alto porcentaje fue que "vergüenza" (25%), el segundo lugar lo ocupó el "miedo a represalias" con un 18 % mientras que las razones "no era necesario "y "**no sabía dónde ir"** bordean el 15%. Sin embargo, es importante resaltar, que a pesar de que las agresiones son graves, parte de la muestra aún presenta un "miedo a la separación" (6 %) y "miedo a causarle problemas"(11 %) ((GRADE), 2019). Esto lo podemos observar en la Tabla 2.1.

	2012	2013	2014	2015	2016	2017
Vergüenza	26.3	26.0	25.9	25.2	25.5	25.0
Miedo a represalias	17.8	16.9	17.2	17.8	18.1	18.4
No era necesario	15.7	16.1	16.3	16.3	16.2	15.9
No sabía a dónde ir	15.5	16.4	16.5	16.4	15.8	15.4
Miedo a causarle problemas	9.1	9.2	9.4	9.9	10.0	10.5
Miedo a la separación	6.9	7.0	6.3	6.1	6.2	6.2
Otros	8.7	8.4	8.5	8.3	8.3	8.7
Observaciones	**5491**	**5207**	**5068**	**5127**	**5096**	**5069**

Nota: Otros considera «no sirve de nada», «es parte de la vida» y «ella tuvo la culpa». Las tasas incluyen los tres años previos a cada encuesta.
Fuente: Endes 2009-2017.

Tabla 2.1 Razones principales para no buscar ayuda en instituciones (%) ((GRADE), 2019)

Otra alerta que nos dan estos resultados es que el proceso de 5 años en el que se realizó el estudio no se vio un cambio o descenso considerable en las cifras obtenidas. Por ello, el estudio plantea dos preguntas: **¿qué factores determinan que una víctima busque ayuda?** y **¿qué factores influyen en que una víctima denuncie ante la Policía, busque apoyo familiar o amigos, o recurra a ambas opciones?** Para dar respuesta a ambas interrogantes, se empleó información de la Endes[1] sobre mujeres que sufrieron violencia física o sexual de parte de su pareja o compañero. Se analizaron estos datos a la luz de la teoría propuesta por Liang, Goodman, Tummala-narra y Weintraub, quiénes plantean que la búsqueda de ayuda pasa por tres etapas: la asimilación de la violencia como un problema, decisión de buscar ayuda y decisión del tipo de ayuda a la cual recurrir. Este modelo podemos verlo en la Figura N° 2.4.

[1] La Encuesta Demográfica y de Salud Familiar

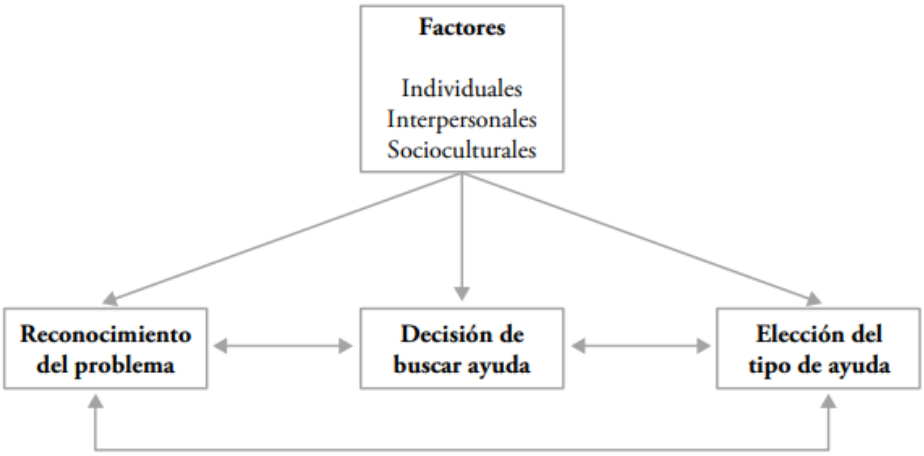


Figura 2.4 Modelo de Búsqueda de Ayuda[1]

1. Reconocimiento del problema: Es considerado el primer y más importante paso dentro del proceso de búsqueda de ayuda. En esta etapa la víctima, según menciona Prochaska y otros (1992), suele negar la existencia de la violencia comparando su caso con otros más graves para de esa manera restarle importancia. Es así como a medida que la violencia se va incrementado tanto en frecuencia como severidad, la aceptación comienza a darse. Todo este proceso puede verse afectado por factores propios de cada individuo como por ejemplo nivel educativo, económico, etc.
2. Decisión de buscar ayuda: En el momento en el que la mujer que ha sido agredida logra identificar las señales de las agresiones, los tipos de respuesta que se pueden presentar son diversos. Estas pueden ser pasivas, como esperar que su pareja cambie, o en contraposición, buscar eliminar la violencia por completo con ayuda especializada. La búsqueda de ayuda, sin embargo, dependiendo de la que se necesite, requiere que la víctima sea capaz de reconocer que por sí misma es incapaz de darle fin a la violencia. **Lamentablemente, hay condiciones sociales que reducen la libertad de las mujeres para poder realizar el proceso de búsqueda de ayuda especializada, con lo cual se ven empujadas a optar por los propios medios con los que cuentan para eliminarla. Dichas condiciones pueden ser el aislamiento social y cultura, la inmigración, bajo poder adquisitivo, comisarías lejanas o de difícil acceso.** ((GRADE), 2019)

[1] Tomado de Liang y otros (2005).

3. Elección del tipo de ayuda: La elección del tipo de ayuda o soporte está fuertemente vinculado con el tipo de violencia del que está siendo víctima. Por ejemplo, en los casos en que la agredida considere que se encuentra enfrentando un problema de violencia psicológica leve, es probable que el tipo de ayuda que busque se encuentre dentro de las fuentes informales como pueden ser amigas. Sin embargo, si la violencia es severa, la víctima necesita además de sanción para el agresor, un apoyo psicológico y un acompañamiento para prevenir nuevos ataques. ((GRADE), 2019)

Finalmente, toda esta información es resumida en la Tabla N° 2.2 que muestra la tasa de reportes violencia física de acuerdo con el tipo de ayuda a la que la víctima acudió.

	2009	2010	2011	2012	2013	2014	2015	2016	2017
Informal	28.3	30.4	29.6	28.3	28.7	29.3	32.1	33.0	32.5
Padres	16.5	18.0	17.3	15.6	16.4	17.1	19.4	18.8	19.5
Hermanos	9.1	10.3	10.8	9.5	10.4	11.3	11.0	11.4	11.2
Suegros	5.0	5.5	5.1	5.3	5.4	5.7	5.8	6.3	6.3
Amigos	2.2	2.3	1.9	2.3	2.3	2.0	2.5	2.6	1.9
Formal	23.7	26.4	26.0	26.9	23.9	25.2	26.2	26.3	28.2
Denuncia 1/	20.9	23.1	23.0	24.6	21.7	23.1	23.5	23.7	25.4
Demuna	3.2	4.3	3.9	3.1	3.4	3.0	3.5	3.1	3.7
CEM	1.2	1.4	1.0	1.6	1.0	1.2	1.1	1.2	1.5
Solo informal	18.3	18.2	17.4	16.9	17.7	17.8	19.2	20.4	19.4
Solo formal	13.7	14.2	13.8	15.5	13.0	13.7	13.3	13.6	15.1
Informal y formal	10.0	12.2	12.2	11.4	11.0	11.5	12.9	12.6	13.1
Alguna ayuda	42.0	44.5	43.4	43.8	41.6	43.0	45.4	46.6	47.6
Obs.	5594	5103	5184	5171	4950	4789	7155	6662	6438

Nota: La tasa incluye los tres años previos a cada encuesta.
1/ Considera comisarías, juzgados y Fiscalía.

Tabla 2.2 Porcentaje de reporte de víctimas de violencia física (no severa y severa) por tipo de fuente de ayuda y año ((GRADE), 2019)

Se observa que los medios informales y de los que difícilmente se puede tener información son a los que más se acude, puesto que la víctima se siente más segura y con menos miedo de ser juzgada por ser el ámbito familiar.

2.2. Conceptos y Definiciones de la(s) Tecnología(s) a usar en la construcción de solución.

2.2.1. Procesamiento de lenguaje natural (NLP)

El procesamiento del lenguaje natural busca entender el lenguaje humano con el propósito de realizar diferentes tareas. Esto, mediante el uso de computadores que se encarguen de realizar dicho proceso.

Es un campo de carácter interdisciplinario, que combina ciencia computacional y cognitiva, lingüística computacional e inteligencia artificial. Desde la óptica de la ingeniería, NLP busca la manera desarrollar nuevas aplicaciones que faciliten la interacción computadoras y el lenguaje humano. (Deng & Liu, 2018)

2.2.1.1. Aplicaciones de NLP

Las típicas aplicaciones del procesamiento del lenguaje natural sol: Entendimiento del lenguaje hablado, reconocimiento de voz, sistemas de diálogo, análisis léxico, análisis de sentimientos, generador de lenguaje natural. (Deng & Liu, 2018)

2.2.2. BERT

Es un modelo de representación de lenguaje, que significa Bidirectional Encoder Representations from Transformers (BERT). Está diseñado para entrenar representaciones bidireccionales profundas a partir de texto sin etiquetar, mediante el condicionamiento conjunto de comparación –de un contexto izquierdo y derecho– en todas las capas. Lo que hace diferente a BERT de otros modelos de representación contextual como Elmo, ULMFit y OpenAI GTP, es que se trata de la primera representación de lenguaje profundamente bidireccional - no supervisado, previamente entrenado, utilizando solo un corpus de texto plano (Devlin, Chang, Lee, & Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, 2018).

Adicional a ello, BERT contiene un codificador con 12 bloques transformadores, 12 cabezales de auto-atención y un tamaño oculto de 768. Toma la entrada de una secuencia de no más de 512 tokens y emite la representación de ésta. La secuencia, tiene uno o dos segmentos. El primer token de la secuencia es siempre [CLS], que contiene la clasificación especial incrustada. Otro token especial es [SEP] que se utiliza para separar segmentos. En cuanto a las tareas de clasificación de texto, BERT toma el estado oculto final “h” del primer token [CLS] como la representación de toda la secuencia y agrega un clasificador softmax simple en la parte superior de BERT para predecir la probabilidad de etiqueta (Sun, Qiu, Xu, & Huang, 2019) .

2.3. Resumen

La violencia basada en género se encuentra compuesta de diferentes tipos de violencia; siendo las del tipo directa y estructural, las más sencillas de reconocer en el ámbito social y económico. El tener definidos además diferentes modalidades de violencia, se facilita la obtención de información y medición de su nivel de prevalencia.

En los resultados del estudio de ((GRADE), 2019), se muestra que el mayor impedimento para no realizar una denuncia en caso de ser víctima de una de las modalidades de violencia es la "vergüenza" a la que la víctima siente que se sentirá expuesta al denunciar; esto adicional al "miedo a represalias" demuestra además que no sienten que el gobierno sea capaz de mantener su seguridad y protegerlas durante todo este proceso.

Complementario a estas cifras, basados en el modelo de búsqueda de ayuda que plantea las 3 fases que se deben atravesar. Vemos que el paso primordial es el reconocimiento del problema, el cual tiene una estrecha relación con la necesidad de la visibilizarían del problema.

En cuanto a la tecnología seleccionada, de acuerdo con la información revisada, podemos ver que procesamiento del lenguaje natural, tiene al análisis de sentimientos entre sus aplicaciones. Es así que BERT, como modelo previamente entrenado, bidireccional y no supervisado es ideal para la clasificación de mensajes, práctica que se encuentra dentro del análisis de sentimientos.

En el siguiente capitulo, se realizará una revisión más profunda acerca de las técnicas y modelos de clasificación de mensajes que contengan alguna modalidad de violencia hacia la mujer, así como también diferentes herramientas que existen para medir el desempeño de dichos modelos de clasificación.

CAPÍTULO 3: ESTADO DEL ARTE

En el presente capítulo se presenta una revisión de la literatura acerca del procesamiento del lenguaje natural utilizando diferentes técnicas de inteligencia artificial. En esta revisión se identifican técnicas y herramientas que nos permitan clasificar mensajes.

3.1. Revisión Sistemática de la Literatura

La metodología propuesta, se realizar a partir del procedimiento propuesto por Kitchenham y Charters (Kitchenham & Charters, 2007), el cual plantea 3 etapas.

- Planificación: En esta etapa las preguntas de investigación son planteadas, las cuales incluyen los criterios para la inclusión y exclusión de artículos definiendo así el protocolo de revisión.
- Revisión: En esta etapa se ejecuta el protocolo de revisión, seleccionando así los artículos que pasen exitosamente dichos filtros.
- Resultados: Presentación de los resultados obtenidos.

3.2. Planificación de la Revisión

Para desarrollar la revisión de la literatura, siguiendo con las etapas de la metodología propuesta, se llevaron a cabo las siguientes preguntas.

> **Q1:** ¿Qué técnicas de clasificación de mensajes con contenido de violencia en redes sociales existen?
>
> **Q2:** ¿Qué técnicas de identificación de mensajes que contengan algún tipo de violencia hacia la mujer en las redes sociales existen?
>
> **Q3:** ¿Qué modelos de clasificación de mensajes que contengan algún tipo de violencia hacia la mujer existen?
>
> **Q4:** ¿Qué tipos de herramientas existen para medir el desempeño de modelos de clasificación de textos?

La búsqueda se realizó considerando las bases de datos de: IEEE Xplore, Scopuss y Wos. Los documentos seleccionados corresponden a cadenas de búsqueda específicas detalladas en la Tabla 3.1 y que fueron publicadas entre los años 2016 y 2021.

Fuente	Cadena de búsqueda
Scopus	(domestic AND violence) (text classification) OR (identify messages)) AND PUBYEAR > 2015 AND PUBYEAR < 2022
IEEE Xplore	(domestic violence) AND (machine learning) OR (text classification)
Wos	(Natural Language Processing) AND (Social Media) and (Domestic Violence) OR (Domestic Abuse)

Tabla 3.1Cadenas de búsqueda utilizadas en las bases de datos

Criterios de inclusión	Criterios de exclusión
• Modelos, métodos y técnicas para la clasificación de mensajes. • Artículos que abordan la identificación y clasificación de mensajes con contenido violento hacia las mujeres en redes sociales u otras plataformas web. • Responden una o más preguntas de la investigación planteada. • Artículos que se encuentren entre los años 2016 y 2022.	• Artículos que no aborden la clasificación. • Actas, carteles, tesis, talleres y libros.

Tabla 3.2. Criterios de inclusión y exclusión

3.3. Realización de la Revisión

Prosiguiendo con la búsqueda, esta se realizó de acuerdo con las cadenas de búsqueda de manera sistemática. Resultado de ello se encontraron 47 artículos, posterior a ello se aplicaron los criterios de inclusión y exclusión, quedando seleccionados 13. Finalmente se realizó un análisis de la introducción, resumen y conclusión para medir la capacidad de respuesta a las preguntas de investigación de cada artículo seleccionado.

El proceso de desarrollo de esta revisión se puede observar en la Figura 3.1, de igual forma la cantidad de artículos encontrados en cada motor de búsqueda se describen en la Tabla 3.3.

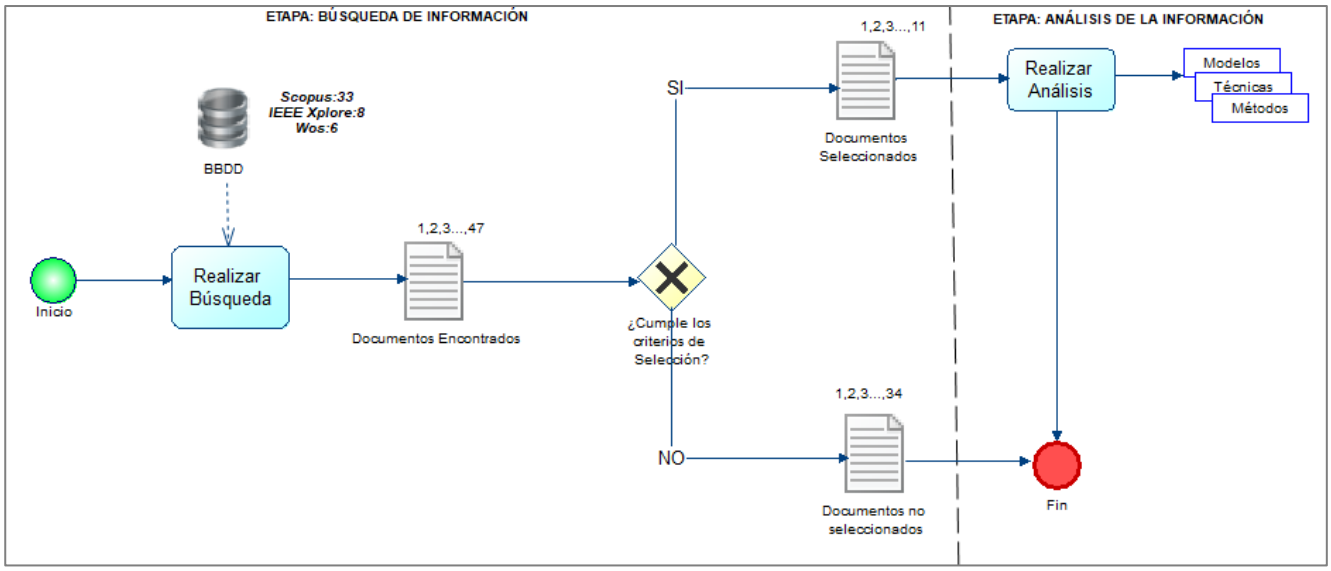


Figura 3.1. Proceso de revisión de la literatura

Fuente	Estudios potencialmente elegibles	Estudios seleccionados
Scopus	33	4
IEEE Xplore	8	7
Wos	6	2
TOTAL	**47**	**13**

Tabla 3.3Cantidad de documentos por motor de búsqueda

3.4. Resultados

3.4.1. Documentos seleccionados

Luego de ejecutar la planificación y revisión se muestran los documentos obtenidos.

Véase Tabla 3.4.

ID	Título	Autor	Revista
A01	Combating the challenges of social media hate speech in a polarized society. A Twitter ego lexalytics approach	(Udanor & Anyanwu, 2019)	Data Technologies and Applications
A02	Detecting misogyny in Spanish tweets. An approach based on	(García-Díaz, Cánovas-	Elsevier

	linguistics features and word embeddings	García, Colomo-Palacios, & Valencia-García, 2020)	
A03	Metadata Based Multi-Labelling of YouTube Videos	(Agarwal, Gupta, Singh, & Saxena, 2017)	IEEE
A04	Deep Learning for Multi-Class Identification From Domestic Violence Online Posts	(SUBRAMANI, y otros, 2019)	IEEE Access
A05	Are They Our Brothers? Analysis and Detection of Religious Hate Speech in the Arabic Twittersphere	(Albadiy, Kurdiz, & Mishra, 2018)	IEEE
A06	Predicting Cyberbullying on Social Media in the Big Data Era Using Machine Learning Algorithms: Review of Literature and Open Challenges	(AL-GARADI, y otros, 2019)	IEEE Access
A07	Sentiment Analysis in Turkish with Deep Learning	(Demirci, Keskin, & Dogan, 2019)	IEEE
A08	Analyzing Abusive Text Messages to Detect Digital Dating Abuse	(Roy, McClendon, & Hodges, 2018)	IEEE

A09	Domestic Violence Crisis Identification From Facebook Posts Based on Deep Learning	(SUBRAMANI, WANG, & VU, Domestic Violence Crisis Identification From Facebook Posts Based on Deep Learning, 54075-54085)	IEEE Access
A10	Automatic classification of social media reports on violent incidents in South Africa using machine learning	(Kotzé, Senekal, & Daelemans, 2020)	South African Journal of Science
A11	Arabic Text Classification using Feature-Reduction Techniques for Detecting Violence on Social Media	(ALSaif & Alotaibi, 2019)	(IJACSA) International Journal of Advanced Computer Science and Applications
A12	CyberBERT: BERT for cyberbullying identification	(Sayanta & Sriparna, 2020)	Multimedia Systems - Springer
A13	Cyberbullying Detection using Pre-Trained BERT Model	(Yadav, Kumar, & Chauhan)	IEEE

Tabla 3.4 Documentos seleccionados para el estudio

3.5. Análisis

En esta sección se responden a las preguntas de investigación que fueron propuestas en la planificación de la revisión.

3.5.1. ¿Qué técnicas de clasificación de mensajes con contenido de violencia en redes sociales existen?

Las técnicas de clasificación de mensajes se refieren al conjunto de prácticas y herramientas utilizadas, para realizar un etiquetado de lo que se considera violento o no dentro de una red social. Todo esto extraído de la literatura, véase Tabla 3.5.

Id	Técnica	Descripción	Fuente
T1	Clasificación automática	Clasificación automática de mensajes mediante programas, diccionario de datos, etc.	(Udanor & Anyanwu, 2019)
T2	Clasificación Manual Asistida	Clasificación manual de mensajes usualmente realizada con la asistencia de personas que no figuran como autores del artículo.	(García-Díaz, Cánovas-García, Colomo-Palacios, & Valencia-García, 2020) (Yadav, Kumar, & Chauhan)
T3	Clasificación Manual en más de 2 categorías.	Clasificación manual realizada por los autores en más de 2 categorías.	(Agarwal, Gupta, Singh, & Saxena, 2017) (Sayanta & Sriparna, 2020)
T4	Clasificación Manual en más de 2 categorías con asistencia.	Clasificación manual de mensajes en más de 2 categorías usualmente realizada con la asistencia de personas que no figuran como autores del artículo.	(SUBRAMANI, y otros, 2019) (Roy, McClendon, & Hodges, 2018) (SUBRAMANI, WANG, & VU, Domestic Violence Crisis Identification From Facebook Posts Based on Deep Learning, 54075-54085)
T5	Clasificación personalizada por grupo especializado	Clasificación de mensajes por parte de expertos contratados ajenos a la investigación.	(Albadiy, Kurdiz, & Mishra, 2018)
T6	Clasificación Manual en 2 categorías	Clasificación de mensajes realizada por los autores en 2 categorías.	(Demirci, Keskin, & Dogan, 2019)

Tabla 3.5 Técnicas de clasificación de mensajes con contenido de violencia en redes sociales

3.5.2. ¿Qué técnicas de identificación de mensajes que contengan algún tipo de violencia hacia la mujer en las redes sociales existen?

Se realizó una revisión en la literatura acerca del proceso de obtención de data, así como también el proceso de etiquetado de una forma más específica para los casos de estudio que incluyeran mensajes violentos a la mujer en redes sociales. Ver Tabla 3.6.

Id	Técnica	Descripción Dataset	Técnica Específica	Fuente
TVM1	Clasificación Manual Asistida	Conjunto de 3 dataset: - Violencia contra mujeres relevantes: Abarca comentarios hacia mujeres con relevancia o allegada política como Greta Thunberg o políticas españolas. - Españoles europeos vs Latinoamericanos: Se tomo en cuenta la latitud para saber el lugar de procedencia del usuario del tweet. Este proceso se llevó a cabo dado que de acuerdo con el contexto de determinada región el significado de una palabra puede variar. - Descredito, dominación, acoso sexual, y estereotipos: Se seleccionaron tweets dónde se discutían tópicos relacionados a la violencia basada en género.	- Si el tweet contenía ciertos tratos misóginos como objetivación, acoso sexual, dominación, etc. Era clasificado como Misógino. - Si la cuenta pertenece a una web de noticias fue catalogado como no Misógino. Para clasificar un mensaje como misógino, es necesario que más de dos de los anotadores lo califique como tal y que ninguno lo califique como no misógino.	(García-Díaz, Cánovas-García, Colomo-Palacios, & Valencia-García, 2020)
TVM2	Clasificación Manual en más de 2 categorías con asistencia.	Dataset compuesto por post de Facebook con términos de búsqueda “Violencia Doméstica “y “Abuso Doméstico”	Fue categorizado en 4: - Historia Personal. - Recaudación de fondos - Empatía - General El grado de acuerdo entre los dos estudiantes y el profesional psiquiátrico en el tema de violencia doméstica fue de 0.81. En caso de incertidumbre se etiqueto de acuerdo al consejo del experto.	(SUBRAMANI, y otros, 2019)
TVM3	Clasificación Manual en más de 2 categorías	Dataset compuesto por post en la página web " athinline.org” cercanamente alineados a diferentes escenarios de abuso de pareja.	Se reclutó un grupo de 44 estudiantes, los cuáles tomaron un cuestionario para establecer una línea base. Adicionalmente se les hizo alcance de 5	(Roy, McClendon, & Hodges, 2018)

	con asistencia.		historias de violencia en diferentes escenarios. Se les pidió que se hicieran pasar por la pareja romántica abusiva para la obtención de los mensajes.	
TVM4	Clasificación Manual en más de 2 categorías con asistencia	Dataset compuesto por post de Facebook con términos de búsqueda "Violencia Doméstica "y "Abuso Doméstico"	Se clasifico en las categorías "Crítico" (Necesidad de apoyo y soporte emocional) y " No crítico". El grupo de personas que realizó la revisión de forma individual. Era obligatorio que la etiqueta a colocar fuera por decisión unánime.	(SUBRAMANI, WANG, & VU, Domestic Violence Crisis Identification From Facebook Posts Based on Deep Learning, 54075-54085)
TVM5	Clasificación Manual en más de 2 categorías.	Dataset compuesto por tweets entre los años 2017 y 2018 de las regiones de Riyadh, Makkah, regiones del este y Asir.	Se clasificó en 3 categorías: Violencia física, violencia psicológica y no violencia.	(ALSaif & Alotaibi, 2019)

Tabla 3.6 Técnicas de identificación de mensajes que contengan algún tipo de violencia hacia la mujer en las redes sociales

3.5.3. ¿Qué modelos de clasificación de mensajes que contengan algún tipo de violencia hacia la mujer existen?

En total se encontraron 12 modelos de clasificación de mensajes. Se específico para la mayoría de ellos la precisión y exactitud obtenida. Destacándose de la literatura la investigación realizada por S. Subramani et al. (SUBRAMANI, y otros, 2019) Dónde se logró una precisión del 94.40 para clasificar post de Facebook con las etiquetas "critico" o "no crítico" Véase Tabla 3.7.

ID	Clasificación	Resultados	Fuente
CVM1	Bosques Aleatorios	Accuracy: 82.092% Accuracy:84.40% Accuracy: 89.24%	(García-Díaz, Cánovas-García, Colomo-Palacios, & Valencia-García, 2020) (SUBRAMANI, y otros, 2019) (SUBRAMANI, WANG, & VU, Domestic Violence Crisis Identification From Facebook Posts Based on Deep Learning, 54075-54085)

CVM2	Arboles de decisión	Accuracy:82.77% Precision: 83.66 % Accuracy:89.83% Precision: 89.80 %	(SUBRAMANI, y otros, 2019) (SUBRAMANI, WANG, & VU, Domestic Violence Crisis Identification From Facebook Posts Based on Deep Learning, 54075-54085)
CVM3	Linear Support Vector Machines	Accuracy: 84.480% AUC: 0.94%	(García-Díaz, Cánovas-García, Colomo-Palacios, & Valencia-García, 2020) (Roy, McClendon, & Hodges, 2018)
CVM4	Máquina de soporte vectorial	Accuracy:90.81% Precision: 91.00 % Accuracy:92.20% Precision: 92.90 %	(SUBRAMANI, y otros, 2019) (SUBRAMANI, WANG, & VU, Domestic Violence Crisis Identification From Facebook Posts Based on Deep Learning, 54075-54085)
CVM5	Redes Neuronales Convolucionales (CNN)	Accuracy:90.93% Precision: 91.33 % Accuracy:93.82% Precision: 93.90 %	(SUBRAMANI, y otros, 2019) (SUBRAMANI, WANG, & VU, Domestic Violence Crisis Identification From Facebook Posts Based on Deep Learning, 54075-54085)
CVM6	Redes Neurales Recurrentes (RNN)	Accuracy:67.65% Precision: 69.33 % Accuracy:85.72% Precision: 86.00 %	(SUBRAMANI, y otros, 2019) (SUBRAMANI, WANG, & VU, Domestic Violence Crisis Identification From Facebook Posts Based on Deep Learning, 54075-54085)
CVM7	LSTM	Accuracy:90.99% Precision: 91.00 % Accuracy:94.02% Precision: 94.10 %	(SUBRAMANI, y otros, 2019) (SUBRAMANI, WANG, & VU, Domestic Violence Crisis Identification From Facebook Posts Based on Deep Learning, 54075-54085)
CVM8	GRU	Accuracy:91.78% Precision: 91.66 %	(SUBRAMANI, y otros, 2019) (SUBRAMANI, WANG, & VU,

		Accuracy:94.26% Precision: 94.50 %	Domestic Violence Crisis Identification From Facebook Posts Based on Deep Learning, 54075-54085)
CVM9	BLSTM	Accuracy:91.29% Precision: 91.66 % **Accuracy:** 94.16% **Precision:** 94.40 %	(SUBRAMANI, y otros, 2019) (SUBRAMANI, WANG, & VU, Domestic Violence Crisis Identification From Facebook Posts Based on Deep Learning, 54075-54085)
CVM10	Regresión Logística	Accuracy:90.45% Precision: 91.00 % Accuracy:90.74% Precision: 90.80 %	(SUBRAMANI, y otros, 2019) (SUBRAMANI, WANG, & VU, Domestic Violence Crisis Identification From Facebook Posts Based on Deep Learning, 54075-54085)
CVM11	Naive Bayes	AUC: -- Accuracy:91.66% Precision: 91.80 %	(Roy, McClendon, & Hodges, 2018) (SUBRAMANI, WANG, & VU, Domestic Violence Crisis Identification From Facebook Posts Based on Deep Learning, 54075-54085)
CVM12	Sequential Minimal Optimization (SMO)	Accuracy:84.886 %	(García-Díaz, Cánovas-García, Colomo-Palacios, & Valencia-García, 2020)

Tabla 3.7 Modelos de clasificación de mensajes que contengan algún tipo de violencia hacia la mujer

3.5.4. ¿Qué tipos de herramientas existen para medir el desempeño de modelos de clasificación de textos?

De la revisión de los artículos se encontraron 2 herramientas para la medición del desempeño, siendo la más utilizada la Matriz de Confusión. Véase Tabla 3.8.

ID	Clasificación	Detalle	Fuente
H1	Matriz de Confusión	Herramienta que permite la visualización del desempeño de un algoritmo.	1. (García-Díaz, Cánovas-García, Colomo-Palacios, & Valencia-García, 2020) 2. (Agarwal, Gupta, Singh, & Saxena, 2017) 3. (SUBRAMANI, y otros, 2019) 4. (Albadiy, Kurdiz, & Mishra, 2018) 5. (Demirci, Keskin, & Dogan, 2019) 6. (Roy, McClendon, & Hodges, 2018) 7. (SUBRAMANI, WANG, & VU, Domestic Violence Crisis Identification From Facebook Posts Based on Deep Learning, 54075-54085) 8. (Kotzé, Senekal, & Daelemans, 2020) 9. (ALSaif & Alotaibi, 2019) 10. (Sayanta & Sriparna, 2020) 11. (Yadav, Kumar, & Chauhan)
H2	Curva Roc	Gráfico que muestra el rendimiento de un modelo de clasificación, representa la sensibilidad frente a la especificidad.	1. (Albadiy, Kurdiz, & Mishra, 2018) 2. (Roy, McClendon, & Hodges, 2018)

Tabla 3.8. Herramientas para medir el desempeño de modelos de clasificación de textos

3.6. Resumen

El resultado de la revisión sistemática de la literatura nos dio un total de 47 posibles estudios elegibles, siendo finalmente seleccionados 13 de las fuentes de Scopus, IEEE Xplore y Wos; todos estos dentro del rango 2016-2022. Respecto a la construcción de los dataset con clasificación, la técnica más utilizada, presente en tres artículos, fue la clasificación manual de mensajes en más de 2 categorías esto con ayuda y/o asistencia de especialistas en el área de estudio y que no figuran como autores del artículo.

Finalmente, en cuanto a las herramientas utilizadas para la medición del desempeño de los modelos, la matriz de confusión fue la que más se empleó, en contraste con la curva roc que se utilizó de manera complementaria en solo dos de los artículos.

En el siguiente capítulo, se realizará la justificación del modelo seleccionado partiendo de la revisión ejecutada en el presente capítulo. Es así como veremos la razón de la selección del modelo y su comparativa con otros modelos expuestos previamente.

CAPÍTULO 4: MODELO DE APRENDIZAJE PROFUNDO – APORTE TEÓRICO

En el presente capítulo justificamos el modelo seleccionado, explicamos el procedimiento utilizado para la generación del dataset y las técnicas que se emplearán para entrenar y medir el rendimiento del modelo.

4.1. Selección y justificación del modelo

La justificación del modelo parte un análisis realizado durante todo el capítulo tres respecto a las diferentes metodologías y modelos que se pueden aplicar para realizar la clasificación de textos.

Los 2 últimos artículos seleccionados en el capítulo 3 nos muestran investigaciones dónde se realiza una comparativa de los resultados obtenidos entre modelos tradicionales y BERT. El primer artículo aborda la identificación de post relacionados al Cyberbullying en tres diferentes datasets (Sayanta & Sriparna, 2020). Por otro lado, en el segundo trabajo se aborda la misma temática para solo dos dataset (Yadav, Kumar, & Chauhan).

Este proceso fue reforzado por una serie de experimentos con distintos dataset para proporcionar una comparación entre BERT y modelos tradicionales de Machine Learning para la clasificación de textos (Gonzales-Carbajal & Garrido-Merchán, 2020). De la misma forma un estudio orientado a la clasificación de textos relacionados a discursos de odio realizó una comparativa similar (Plaza-del-Arco, Molina-González, Ureña-López, & Martín-Valdivia, 2021).

Los resultados obtenidos muestran que, en seis de los ocho experimentos recopilados, BERT obtiene los mejores indicadores de desempeño con valores superiores a 0.90. Véase Tabla 4.1.

PROCESAMIENTO DE LENGUAJE NATURAL PARA CLASIFICACIÓN DE TEXTOS								
MODELOS	**EXPERIMENTOS**							
	IMDB	**Portuguese news**	**Chinese hotel reviews**	**Cyberbullying - Twitter**	**Cyberbullying - Wikipedia**	**Cyberbullying - Formspring**	**Spanish HS datset (HaterNet)**	**Spanish HS datset (HaterEval)**
	%Accuracy	%Accuracy	%Accuracy	F1 -score	F1 -score	F1 -score	%Precision	%Precision
Logistic Regression	0.8949	—	—	0.81	0.75	0.72	74.33	61.51
SVM	—	—	—	0.8	0.77	0.75	73.65	61.27
CNN	—	—	—	0.93	0.81	0.91	74.2	75.1
BilSTM				0.93	0.87	0.91	78.7	75.36
LSTM	—	—	—	0.85 + (RNN)	0.61 + (RNN)	0.88 + (RNN)	76.63	75.48
Predictor (auto_ml)	—	0.848	0.7399	—	—	—	—	—
BERT	0.9387	0.9093	0.9381	0.94	0.91	0.92	74.84	75.26

Tabla 4.1 Comparativa de modelos para el NPL mediante experimentos

4.2. Modelo de aprendizaje profundo propuesto

En la figura 4.1. se presenta el esquema general del modelo propuesto para lograr la clasificación de mensajes con contenido de violencia de género. Así mismo en la figura 4.2 se muestra la arquitectura de la solución.

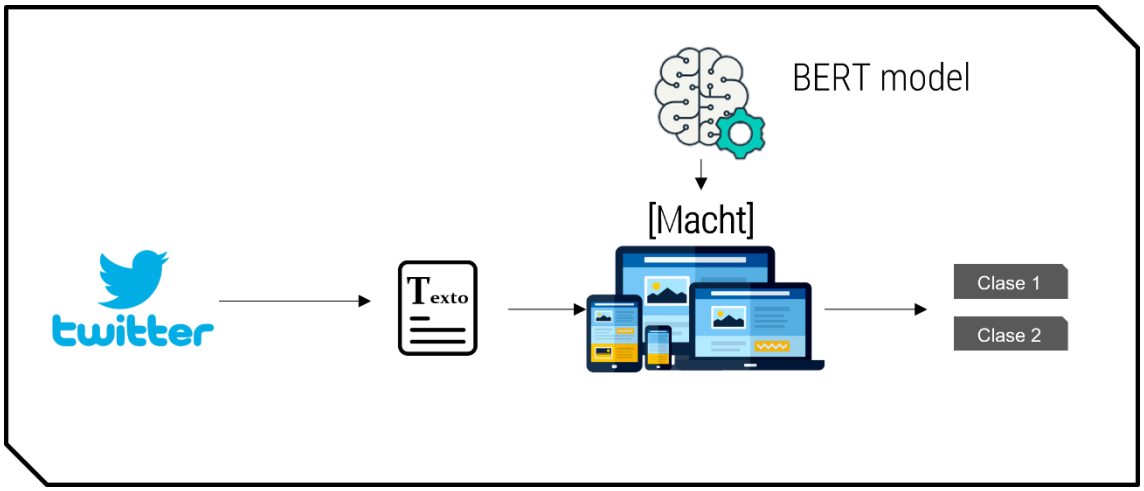


Figura 4.1. Esquema general de la propuesta

La arquitectura del artefacto está alojada en Amazon Web Services, el cual se puede observar 2 instancias de Amazon EC2, una primera instancia que aloja el front-end desarrollado en Vue.js, con la que el usuario será capaz de interactuar con el modelo. Por

otro lado, en la segunda instancia se aloja el back-end que está desarrollado en flask, como también el modelo de BERT que fue escrito en pythorch.

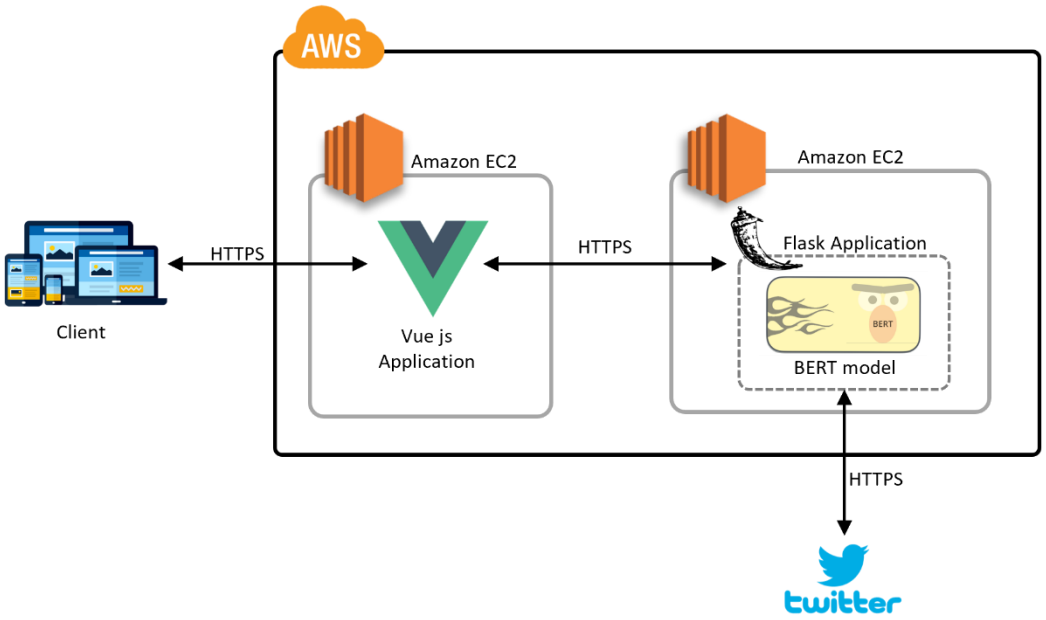


Figura 4.2. Arquitectura de la propuesta

4.2.1. Modelos de BERT

En este trabajo se utilizó el modelo preentrenado BERT al cual se le realizó el proceso de fine-tunning para la tarea de clasificación de textos con contenido violento o no violento hacia la mujer, se utilizó su versión entrenada *BETO-uncased* al cual se le modificaron las últimas capas agregando un modelo secuencial compuesto por una capa Linear de entrada, una capa Relu y una capa lineal de salida de 2 nodos (Véase Figura 4.3).

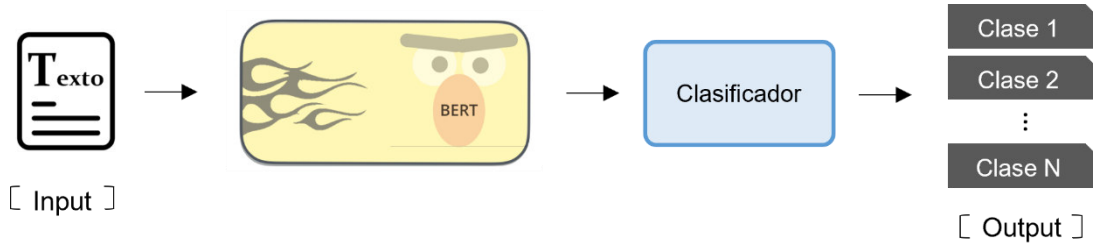


Figura 4.3. Esquema general del modelo

4.2.2. Interpretación y evaluación

En esta fase se evaluarán los resultados de los modelos de forma analítica haciendo el uso de la matriz de confusión (Risch, Stoll, Ziegele, & Krestel, 2019) mostrada en la Tabla 4.2.

Matriz de Confusión		Predicción	
		NV	V
Real	NV	VN	FP
	V	FN	VP

Tabla 4.2. Matriz de confusión para evaluar los modelos de DL

NV: Texto etiquetado como No violencia

V: Texto etiquetado como Violencia

VN: Número de textos clasificados por el modelo como "La mujer no pasó por un proceso violento" que en realidad los son.

FP: Número de textos clasificados por el modelo como "La mujer pasó por un proceso violento", pero que en realidad son "La mujer no pasó por un proceso violento"

FN: Número de textos clasificados por el modelo como "La mujer no pasó por un proceso violento", pero que en realidad son "La mujer pasó por un proceso violento".

VP: Número de textos clasificados por el modelo como "La mujer pasó por un proceso violento" que en realidad lo son.

Con los indicadores de desempeño:

- $Sensibilidad = \frac{\mathrm{VP}}{\mathrm{VP+FN}}\ x\ 100\%$
- $Especificidad = \frac{\mathrm{VN}}{\mathrm{VN+FP}}\ x\ 100\%$
- $Precisión = \frac{\mathrm{VP+VN}}{\mathrm{VN+FP\ +VP+FN}}\ x\ 100\%$

Sensibilidad: Indica el porcentaje de textos etiquetados como "La mujer pasó por un proceso violento" que realmente pertenecen a dicha categoría.

Especificidad: Es el porcentaje de textos clasificados como "La mujer no pasó por un proceso violento" que realmente pertenecen a dicha categoría.

Precisión: Indica el porcentaje de textos **correctamente clasificados**, por el modelo, en las categorías “La mujer pasó por un proceso violento” y “La mujer no pasó por un proceso violento”.

4.3. Metodología para el desarrollo

Una investigación en el campo de las metodologías ágiles las presenta dividiéndolas en cuatro categorías: introducción y adaptación, factores humanos y sociales, la percepción de los métodos agiles y estudios comparativos. Los resultados indican que la introducción de métodos ágiles a proyectos de software de tamaño reducido genera grandes beneficios.

Existen diversas metodologías ágiles para el desarrollo de software. Mencionaremos las tres principales que se abordan en el artículo de Balaguera (Amaya Balaguera, 2015).

1. Mobile – D.
2. HMD (Hybrid Methodology Design)
3. Mobile Development Process Spiral

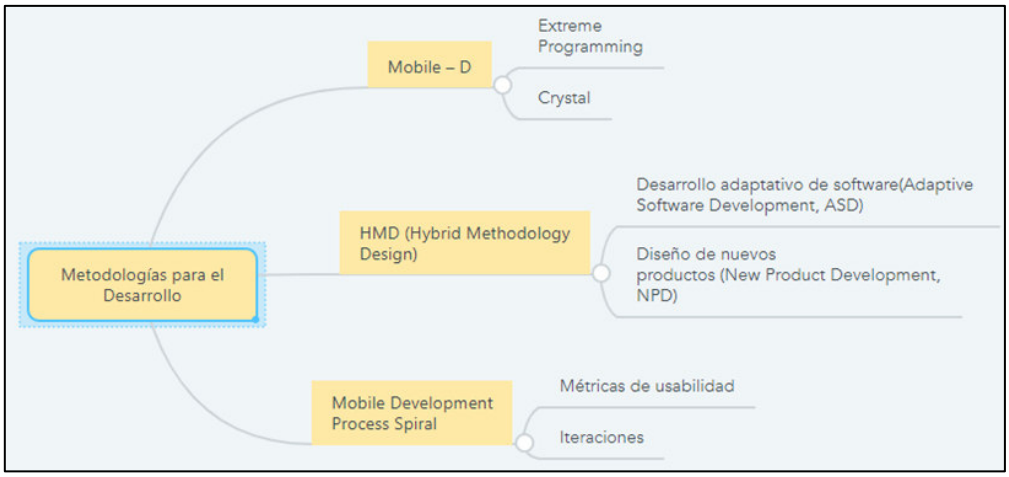


Figura 4.4 Metodologías de Desarrollo (Amaya Balaguera, 2015)

En la Tabla 4.3. se muestra los criterios definidos para realizar el benchmarking basado en Oliver Pérez, de las metodologías mencionadas previamente, esto con la finalidad de elegir la mejor para el trabajo de investigación, cada criterio posee un peso ponderado de acuerdo con la importancia asignada

Criterio de Evaluación	Peso Ponderado
Independencias de tecnologías	0.25
Documentación Estricta	0.25
Enfoque en los procesos	0.125
Enfoque en las personas	0.125
Resultados Rápidos	0.0625
Manejo de Tiempo	0.0625
Iterativo	0.0625
Respuesta a los cambios	0.0625
TOTAL	**1**

Tabla 4.3 Criterios de Evaluación

La distribución de los pesos se efectuó de acuerdo con la importancia asignada a cada criterio para la solución propuesta.

4.3.1. Puntajes por criterio

Se debe tener en cuenta que se ha establecido un rango de puntajes:

- 0 (mínimo)
- 2 (medio)
- 4 (alta).

Independencia de tecnologías

Se evalúa el grado en que la metodología de desarrollo es independiente de la tecnología a utilizar para construir la solución. En la tabla 4.4 se muestran los puntajes respectivos:

Grado de Accesibilidad	Puntaje
Muy Independiente	4
Independiente	2
No independiente / no especifica	0

Tabla 4.4. Puntaje del Criterio Independencia de tecnologías

Documentación Estricta

Se evalúa si la metodología cuenta con una documentación estricta. Véase Tabla 4.5.

Análisis de Información	Puntaje
Posee documentación	4
No posee / No específica	0

Tabla 4.5. Puntaje del Criterio de Documentación Estricta

Enfoque en los procesos

Mide si la metodología se encuentra enfocada en los procesos. Véase Tabla 4.6.

Grado de Interoperabilidad	Puntaje
Altamente enfocado	4
Medianamente enfocado	2
No enfocado / no especifica	0

Tabla 4.6. Puntaje del Criterio de Enfoque en los procesos

Enfoque en las personas.

Mide si la metodología se encuentra enfocada en las personas. Véase Tabla 4.7.

Grado de concientización	Puntaje
Altamente enfocado	4
Medianamente enfocado	2
No enfocado / no especifica	0

Tabla 4.7 Puntaje del Criterio de Enfoque en las personas

Resultados rápidos

Se evalúa la capacidad que posee la metodología para proveer métodos y técnicas que ayuden a obtener resultados de manera rápida. Véase Tabla 4.8.

Grado de Escalabilidad	Puntaje
Resultados rápidos	4
Resultados medianamente rápidos	2
Lentos / no especifica	0

Tabla 4.8. Puntaje del Criterio de Resultados rápidos

Manejo de Tiempo

Se califica si la metodología ofrece técnicas y herramientas para un adecuado y eficiente manejo de tiempo. Véase Tabla 4.9.

Generación de reportes	Puntaje
Ofrece manejo de tiempo	4
No posee / no especifica	0

Tabla 4.9. Puntaje del Criterio de Manejo de Tiempo

Iterativo

Califica si la metodología es iterativa. Véase Tabla 4.10.

Diseño	Puntaje
Iterativo	4
No iterativo / no especifica	0

Tabla 4.10 Puntaje del Criterio Iterativo

Respuesta a los cambios

Califica si la metodología permite responder de manera rápida y oportuna a nuevos o cambios de requerimientos. Véase Tabla 4.11.

Diseño	Puntaje
Alta capacidad de respuesta	4
Media capacidad de respuesta	3
Baja capacidad de respuesta	2
Nula capacidad de respuesta / no especifica	0

Tabla 4.11. Puntaje del Criterio de Respuesta a los cambios

En la Figura 4.5 se muestran los puntajes que obtuvieron las metodologías definidas previamente. Como se observa, la metodología Mobile Development Process Spiral escogida previamente fue la que obtuvo el mayor puntaje.

BENCHMARKING METODOLOGÍAS DE DESARROLLO							
CRITERIOS COMPARATIVOS	PESOS	SOLUCIONES					
		Mobile-D		HMD (Hybrid Methodology Design)		Mobile Development Process Spiral (ELECCIÓN)	
		CALIFICACIÓN	RESULTADO	CALIFICACIÓN	RESULTADO	CALIFICACIÓN	RESULTADO
Independencias de tecnologías	0.25	0	0	0	0	2	0.5
Documentación Estricta	0.25	2	0.5	0	0	2	0.5
Más enfocado en los procesos	0.125	2	0.25	4	0.5	2	0.25
Más enfocado a las personas	0.125	4	0.5	0	0	4	0.5
Resultados Rápidos	0.0625	4	0.25	0	0	4	0.25
Manejo de Tiempo	0.0625	4	0.25	2	0.125	4	0.25
Iterativo	0.0625	4	0.25	4	0.25	4	0.25
Respuesta a los cambios	0.0625	4	0.25	0	0	4	0.25
TOTAL	1	24	2.25	10	0.875	26	2.75

Figura 4.5 Benchmarking Metodologías de Desarrollo

CAPÍTULO 5: IMPLEMENTACIÓN Y VALIDACIÓN DE LOS MODELOS DE BERT – APORTE PRÁCTICO

A continuación, se presentará la implementación y validación de los 3 modelos propuestos mediante un data set que se construyó para los fines de este trabajo.

5.1. Diseño de la Solución

5.1.1. Dataset

Uno de los objetivos de la presente tesis es la construcción de un dataset en español de más de 1000 mensajes que recoja la percepción sobre testimonios de violencia. Para ello se realizó la traducción de parte del dataset de Schrading (Schrading, 2015) y su posterior división en encuestas mediante la herramienta de formularios de Google. Para etiquetado se contó con la participación de 22 personas, 8 de 15-19 años (4 mujeres y 4 hombres), 7 de 20 a 24 años (4 mujeres y 3 hombres) y 7 de 25 a 29 años (4 mujeres y 3 hombres). Para contar con el compromiso y la veracidad en los resultados se recompensó a los participantes con S/.10 por cada encuesta llenada; siendo el pago total por participante S/.100 y una sumatoria de S/.2200 nuevos soles.

5.1.1.1. Tweets a clasificar

Se presenta el proceso realizado para realizar la recopilación de los mensajes a clasificar en el formulario. (Véase Figura 5.1)

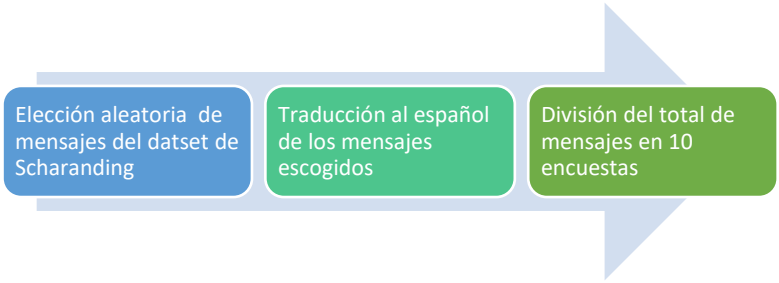


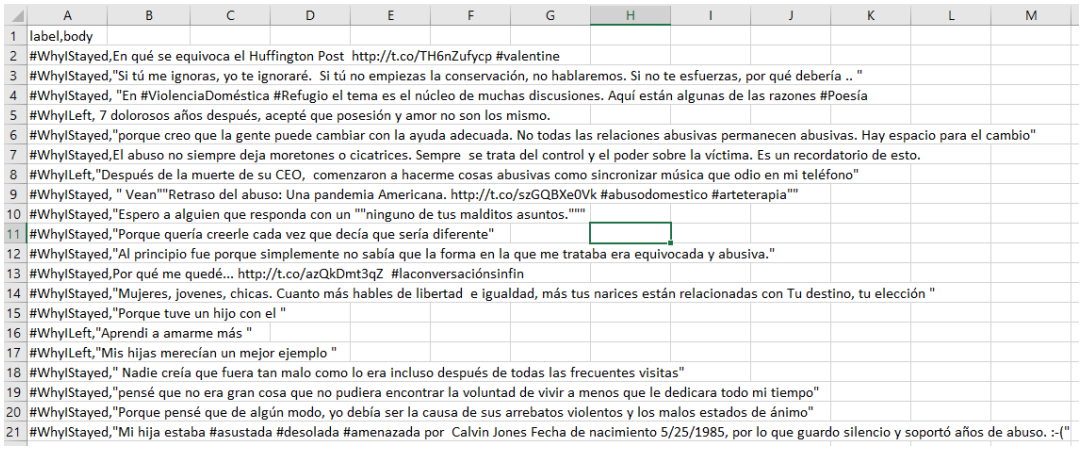
Figura 5.1 Proceso de recopilación de mensajes

5.1.1.2. Elección aleatoria de mensajes

A partir del dataset de Schrading (Schrading, 2015), se escogió de forma aleatoria en grupos de 1050 diferentes mensajes, que después de realizar una primera revisión inicial quedaron en un total de 1042.

5.1.1.3. Traducción al español

Como segundo paso se realizó la traducción al español. Dicha información se guardó en un documento Excel manteniendo el formato del dataset original.

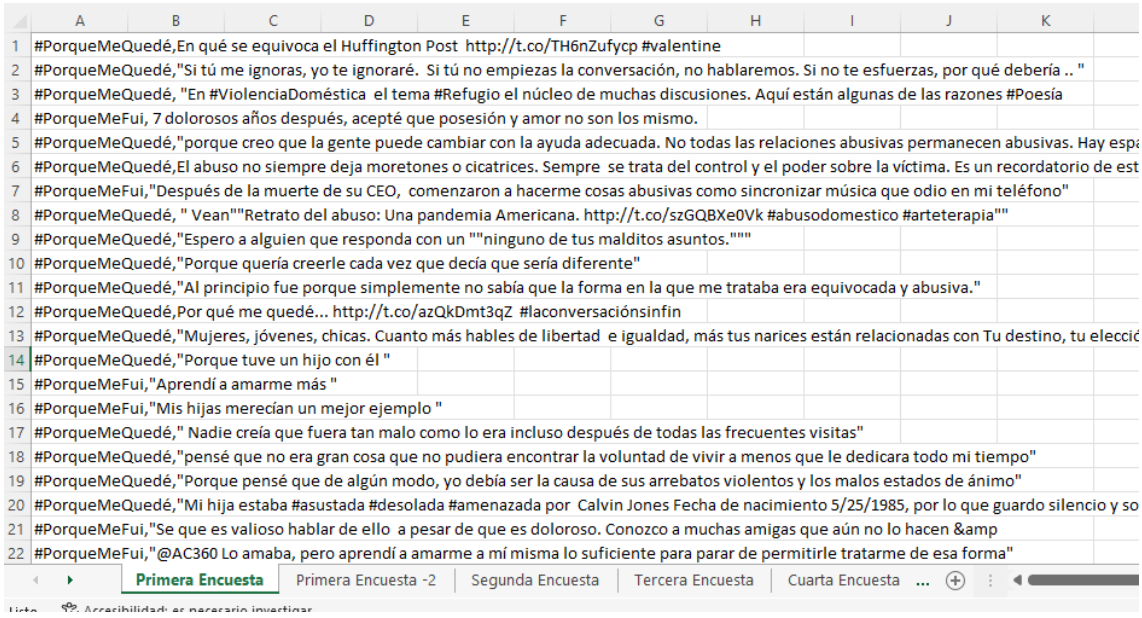


	A	B	C	D	E	F	G	H	I	J	K	L	M
1	label,body												
2	#WhyIStayed,En qué se equivoca el Huffington Post http://t.co/TH6nZufycp #valentine												
3	#WhyIStayed,"Si tú me ignoras, yo te ignoraré. Si tú no empiezas la conservación, no hablaremos. Si no te esfuerzas, por qué debería .. "												
4	#WhyIStayed, "En #ViolenciaDoméstica #Refugio el tema es el núcleo de muchas discusiones. Aquí están algunas de las razones #Poesía												
5	#WhyILeft, 7 dolorosos años después, acepté que posesión y amor no son los mismo.												
6	#WhyIStayed,"porque creo que la gente puede cambiar con la ayuda adecuada. No todas las relaciones abusivas permanecen abusivas. Hay espacio para el cambio"												
7	#WhyIStayed,El abuso no siempre deja moretones o cicatrices. Sempre se trata del control y el poder sobre la víctima. Es un recordatorio de esto.												
8	#WhyILeft,"Después de la muerte de su CEO, comenzaron a hacerme cosas abusivas como sincronizar música que odio en mi teléfono"												
9	#WhyIStayed, " Vean""Retraso del abuso: Una pandemia Americana. http://t.co/szGQBXe0Vk #abusodomestico #arteterapia""												
10	#WhyIStayed,"Espero a alguien que responda con un ""ninguno de tus malditos asuntos.""""												
11	#WhyIStayed,"Porque quería creerle cada vez que decía que sería diferente"												
12	#WhyIStayed,"Al principio fue porque simplemente no sabía que la forma en la que me trataba era equivocada y abusiva."												
13	#WhyIStayed,Por qué me quedé... http://t.co/azQkDmt3qZ #laconversaciónsinfin												
14	#WhyIStayed,"Mujeres, jovenes, chicas. Cuanto más hables de libertad e igualdad, más tus narices están relacionadas con Tu destino, tu elección "												
15	#WhyIStayed,"Porque tuve un hijo con el "												
16	#WhyILeft,"Aprendi a amarme más "												
17	#WhyILeft,"Mis hijas merecían un mejor ejemplo "												
18	#WhyIStayed," Nadie creía que fuera tan malo como lo era incluso después de todas las frecuentes visitas"												
19	#WhyIStayed,"pensé que no era gran cosa que no pudiera encontrar la voluntad de vivir a menos que le dedicara todo mi tiempo"												
20	#WhyIStayed,"Porque pensé que de algún modo, yo debía ser la causa de sus arrebatos violentos y los malos estados de ánimo"												
21	#WhyIStayed,"Mi hija estaba #asustada #desolada #amenazada por Calvin Jones Fecha de nacimiento 5/25/1985, por lo que guardo silencio y soportó años de abuso. :-("												

Figura 5.2 Mensajes Traducidos

5.1.1.4. División de mensajes para encuestas

Finalmente, como último paso debido a la gran cantidad de mensajes se realizó la división de los mensajes en un grupo de 10 para los formularios a utilizar.



	A	B	C	D	E	F	G	H	I	J	K
1	#PorqueMeQuedé,En qué se equivoca el Huffington Post http://t.co/TH6nZufycp #valentine										
2	#PorqueMeQuedé,"Si tú me ignoras, yo te ignoraré. Si tú no empiezas la conversación, no hablaremos. Si no te esfuerzas, por qué debería .. "										
3	#PorqueMeQuedé, "En #ViolenciaDoméstica el tema #Refugio el núcleo de muchas discusiones. Aquí están algunas de las razones #Poesía										
4	#PorqueMeFui, 7 dolorosos años después, acepté que posesión y amor no son los mismo.										
5	#PorqueMeQuedé,"porque creo que la gente puede cambiar con la ayuda adecuada. No todas las relaciones abusivas permanecen abusivas. Hay espa										
6	#PorqueMeQuedé,El abuso no siempre deja moretones o cicatrices. Sempre se trata del control y el poder sobre la víctima. Es un recordatorio de est										
7	#PorqueMeFui,"Después de la muerte de su CEO, comenzaron a hacerme cosas abusivas como sincronizar música que odio en mi teléfono"										
8	#PorqueMeQuedé, " Vean""Retrato del abuso: Una pandemia Americana. http://t.co/szGQBXe0Vk #abusodomestico #arteterapia""										
9	#PorqueMeQuedé,"Espero a alguien que responda con un ""ninguno de tus malditos asuntos.""""										
10	#PorqueMeQuedé,"Porque quería creerle cada vez que decía que sería diferente"										
11	#PorqueMeQuedé,"Al principio fue porque simplemente no sabía que la forma en la que me trataba era equivocada y abusiva."										
12	#PorqueMeQuedé,Por qué me quedé... http://t.co/azQkDmt3qZ #laconversaciónsinfin										
13	#PorqueMeQuedé,"Mujeres, jóvenes, chicas. Cuanto más hables de libertad e igualdad, más tus narices están relacionadas con Tu destino, tu elecció										
14	#PorqueMeQuedé,"Porque tuve un hijo con él "										
15	#PorqueMeFui,"Aprendí a amarme más "										
16	#PorqueMeFui,"Mis hijas merecían un mejor ejemplo "										
17	#PorqueMeQuedé," Nadie creía que fuera tan malo como lo era incluso después de todas las frecuentes visitas"										
18	#PorqueMeQuedé,"pensé que no era gran cosa que no pudiera encontrar la voluntad de vivir a menos que le dedicara todo mi tiempo"										
19	#PorqueMeQuedé,"Porque pensé que de algún modo, yo debía ser la causa de sus arrebatos violentos y los malos estados de ánimo"										
20	#PorqueMeQuedé,"Mi hija estaba #asustada #desolada #amenazada por Calvin Jones Fecha de nacimiento 5/25/1985, por lo que guardo silencio y so										
21	#PorqueMeFui,"Se que es valioso hablar de ello a pesar de que es doloroso. Conozco a muchas amigas que aún no lo hacen &amp										
22	#PorqueMeFui,"@AC360 Lo amaba, pero aprendí a amarme a mí misma lo suficiente para parar de permitirle tratarme de esa forma"										

Primera Encuesta | Primera Encuesta -2 | Segunda Encuesta | Tercera Encuesta | Cuarta Encuesta ...

Figura 5.3 División de mensajes en 10 grupos

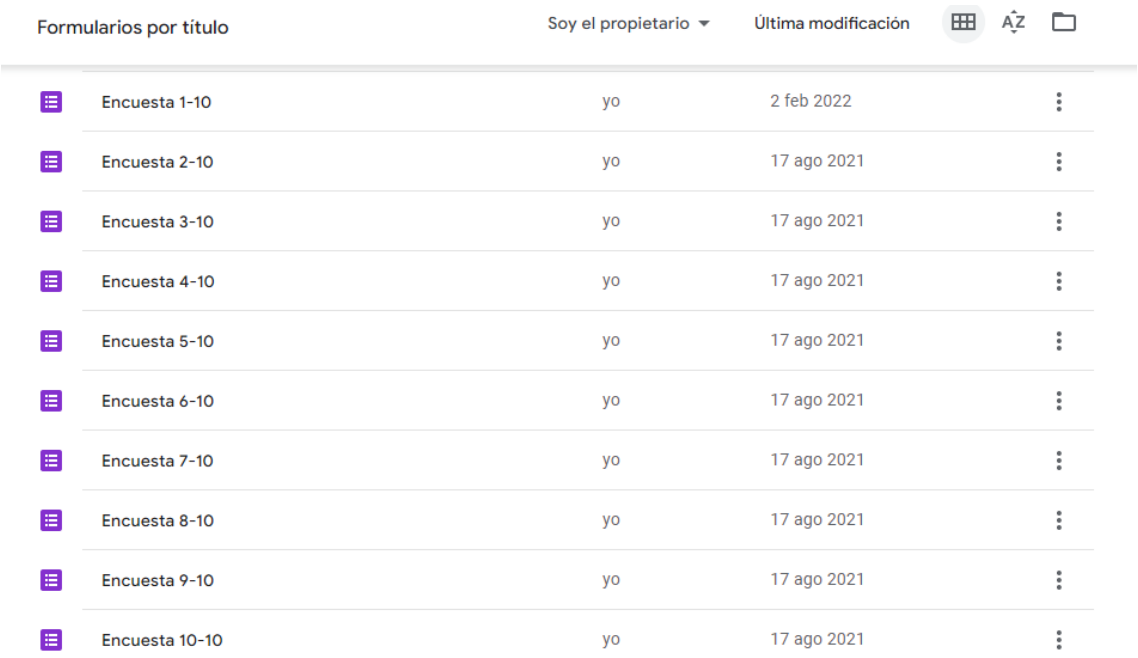


Figura 5.4 Encuestas creadas

5.1.2. Características de formulario

A continuación, se presentan las características del formulario utilizado para la recopilación de información.

5.1.2.1. Introducción

Se refiere a la parte inicial de la encuesta la cual contiene un resumen del motivo del estudio, así como también las indicaciones del proceso a realizar. Véase Figura 5.5.

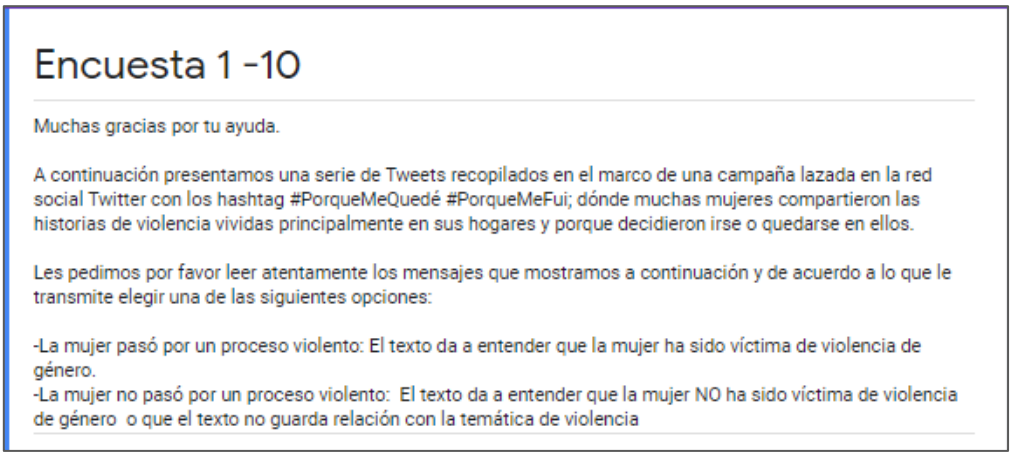


Figura 5.5 Texto Introductorio de formulario de encuesta

5.1.2.2. Grupo de edad

Primera pregunta del formulario de carácter seleccionable. Se realizó una división en 6 rangos. Véase Figura 5.6.

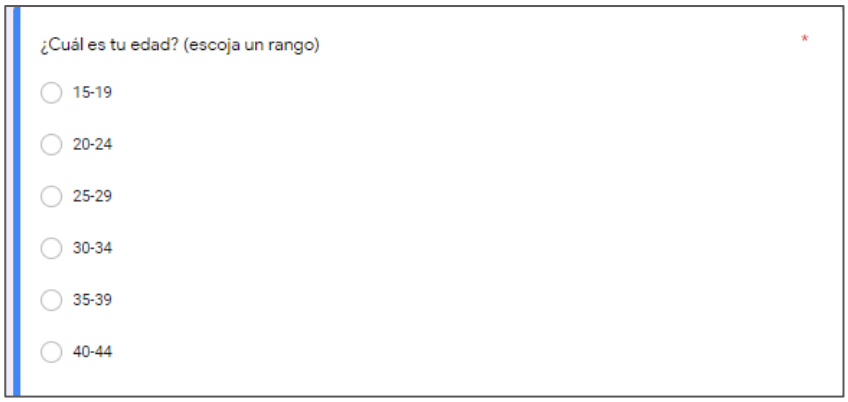


Figura 5.6 Rangos de edad en formulario

Para la elección de este rango, se tuvo en cuenta el informe de la INEI (Instituto Nacional de Estadística e Informática - Encuesta Nacional de Hogares, 2021) el cual señala que el mayor uso de internet (por encima del 75%) se da entre los 12 y 40 años, alcanzando el punto más alto entre los 19 y 24 con un valor de 89.6.

Adicionalmente, se consideró el estudio publicado por la OMS en conjunto con la Escuela de Higiene y Medicina Tropical de Londres y el Consejo Sudafricano de Investigaciones Médicas (Organización Mundial de la Salud, 2013), la cual evidencia que la problemática de prevalencia de la violencia hacia la mujer empieza en edades tempranas (15-19) y alcanza su pico más alto entre los 40 y 44.

5.1.2.3. Género

Segunda pregunta del formulario de carácter seleccionable. Estaba conformada por cinco opciones. Véase Figura 5.7.

Figura 5.7 Opciones de Género en formulario

La importancia de esta pregunta radica en la diferencia de resultados que se puede obtener en cuánto a la percepción, dentro de un mismo rango de edad, dependiendo del género que tenga el sujeto.

5.1.2.4. Clasificación de Tweets recopilados

Es la última sección del formulario, la cual se presenta en forma de cuadricula de varias opciones, es dónde se realizaba la clasificación en las categorías “La mujer pasó por un proceso violento” y “La mujer no pasó por un proceso violento”. Véase Figura 5.8.

Figura 5.8 Clasificación de tweets

5.2. Implementación de la solución

Se utilizaron 3 modelos preentrenados de BERT (SpanBERT (Joshi, et al., 2020), BETO (Cañete, José, et al., 2020) y Multilingual BERT (Google, 2019)) a los cuales se le realizó el proceso de fine-tunning para la tarea de clasificación de mensajes en las categorías “La mujer pasó por un proceso violento” y “La mujer no pasó por un proceso violento”, se modificaron las últimas capas agregando un modelo secuencial compuesto por una

capa Linear de entrada, una capa Relu[1] y una capa lineal de salida de 2 nodos. Para cada modelo se plantearon los escenarios de la tabla 5.1

Escenario	Condiciones de entrenamiento	Tipo de balanceo	Train-test size (%)
1	Se entrena los modelos con los tweets originales.	Sin balanceo Random Oversampling Random Undersampling	90-10 80-20 70-30
2	Entrenamiento de los modelos aplicando las condiciones de limpieza para cada tweet.		

Tabla 5.1 Definición de los escenarios de entrenamiento

Los tipos de balanceo utilizados para estos entrenamientos son:

Sin balanceo: Se entrena el modelo con la cantidad de mensajes del dataset original.

Random OverSampling: Al dataset original se agrega nuevos mensajes de la categoría con la representación más baja. Para nuestro estudio, estos serían mensajes etiquetados como "La mujer no pasó por un proceso violento".

Random UnderSampling: Al dataset original se le reducen mensajes de la categoría con la representación más alta. Para nuestro estudio, estos serían mensajes etiquetados como "La mujer pasó por un proceso violento".

Además, para el entrenamiento de los modelos se contempla un conjunto de experimentos, que parten del planteamiento de escenarios y se ejecutan de manera secuencial siguiendo el esquema de la figura 5.9. Estos experimentos contemplan la calibración de parámetros y el entrenamiento secuencial de cada uno de los 3 modelos pre-entrenados de BERT.

[1] Rectificador lineal unitario.

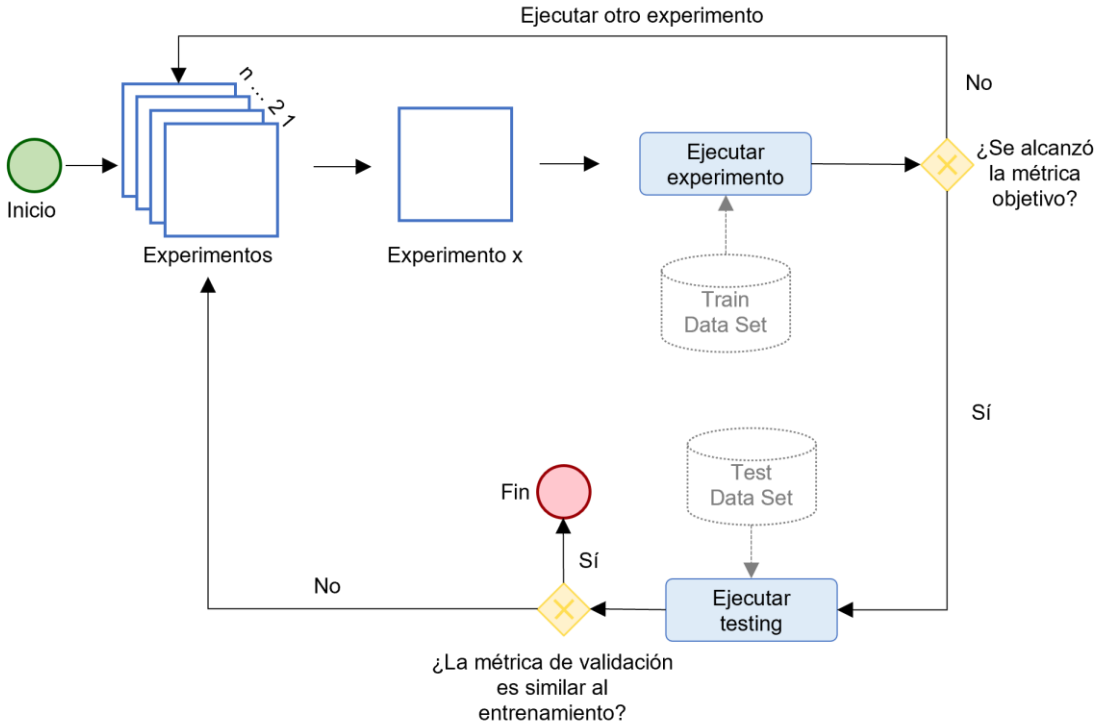


Figura 5.9 Clasificación de tweets

5.3. Ambiente de entrenamiento y validación

5.3.1. Ambiente de entrenamiento

- RAM: 25 GB
- CPU Xeon 2.3 GHz
- GPU Nvidia Tesla V100 16GB
- Disco duro: 100 GB
- Python 3.6
- Pytorch 1.7

5.3.2. Ambiente de validación

- RAM: 8 GB
- CPU Core i3 2.3 GHz
- Disco duro: 100 GB
- Python 3.6
- Pytorch 1.7

5.4. Instancias de pruebas

Se tiene un total de 1042 tweets que fue expuesto en el capítulo 5.1. Las etiquetas finales para los mensajes, se decidieron por mayoría es decir si eran 8 encuestados se necesitaba del acuerdo de al menos 5 con el label “La mujer paso por un proceso de violencia” para colocar dicho valor. El detalle de las etiquetas finales obtenidas se muestra en la tabla 5.2.

Etiqueta	Detalle	#Registros		
		15-19 años	20-24 años	25-29 años
0	La mujer no pasó por un proceso de violencia	162	221	207
1	La mujer pasó por un proceso de violencia	880	821	835

Tabla 5.2 Detalles de las instancias de pruebas

Para ver los atributos del dataset puede acceder al siguiente enlace:

https://github.com/NahumFGz/BERT-Backend-Flask

5.5. Preparación de datos

Como se definió en la sección 5.2 para los datos de entrenamiento y validación de cada se escenario se definieron 3 particiones de entrenamiento y validación el detalle se observa en la tabla 5.3. y para los escenarios 2 se consideran los criterios de limpieza definidos en la tabla 5.4.

id	%TrainSet	%TestSet	# Registros Train	# Registros Test
1	70	30	727	315
2	80	20	831	211
3	90	10	935	107

Tabla 5.3 Definición de los escenarios de entrenamiento

N°	Criterio de limpieza
1	Eliminar los nombres de los usuarios etiquetados en el mensaje.
2	Hacer un Split para los hashtags (ej. “#EjemploDeSplit” pasa a ser “Ejemplo De Split”)
3	Eliminar los hashtags de los mensajes
4	Eliminar los saltos de línea.
5	Eliminar emoticones.

Tabla 5.4 Definición de los escenarios de entrenamiento

5.6. Entrenamiento de los modelos de BERT

Se presentan los experimentos para la validación de los 3 modelos preentrenados de BERT en cada uno de los escenarios.

5.6.1. Definir experimentos

Se definen un conjunto de experimentos con el objetivo de obtener los modelos con el mejor rendimiento. El detalle de los parámetros de cada experimento del primer, segundo se puede observar en la tabla 5.5. donde se muestran los parámetros que se utilizaron para generar los experimentos.

Modelos	Dataset	Tipo de balanceo	Test size	Lr	Num epochs
SpanBERT	15 – 19 años	Sin balanceo	10% 20% 30%	2e-5 3e-5 5e-5	5 6 7 8 9 10
		RandomOverSampling			
		RandomUnderSampling			
	20 – 24 años	Sin balanceo			
		RandomOverSampling			
		RandomUnderSampling			
	25 – 29 años	Sin balanceo			
		RandomOverSampling			
		RandomUnderSampling			
BETO	15 – 19 años	Sin balanceo	10% 20% 30%	2e-5 3e-5 5e-5	5 6 7 8 9 10
		RandomOverSampling			
		RandomUnderSampling			
	20 – 24 años	Sin balanceo			
		RandomOverSampling			
		RandomUnderSampling			
	25 – 29 años	Sin balanceo			
		RandomOverSampling			
		RandomUnderSampling			
multilingual	15 – 19 años	Sin balanceo	10% 20% 30%	2e-5 3e-5 5e-5	5 6 7 8 9 10
		RandomOverSampling			
		RandomUnderSampling			
	20 – 24 años	Sin balanceo			
		RandomOverSampling			
		RandomUnderSampling			
	25 – 29 años	Sin balanceo			
		RandomOverSampling			
		RandomUnderSampling			

Tabla 5.5 Parámetros de entrenamiento del primer escenario y segundo escenario, al combinar todas las columnas obtenemos 2916 experimentos

5.6.2. Ejecutar experimento

Los experimentos se ejecutan según el planteamiento de la figura 5.9 y los parámetros definidos en el punto anterior.

5.7. Resultados y Análisis

Debido a la gran cantidad de experimentos, en este punto se mostrarán los mejores resultados por cada escenario (Véase tablas 5.6 y 5.7), para ver el detalle de los resultados de cada experimento ver ANEXO A.

5.7.1. Escenario 1

De acuerdo con lo definido previamente, en el primer escenario se entraron los modelos con los tweets originales. De los experimentos realizados, se seleccionó el mejor resultado por modelo y TestSize.

Para los 3 tamaños, el mejor resultado se obtuvo con el modelo BETO. Resaltando entre ellos el experimento que se realizó sin balanceo y con un TestSize del 10%, el cual obtuvo un AUC de 0.9805. Cabe destacar que los mejores resultados se obtuvieron con el rango de edad 25-29.

Modelo	Balanceo	TestSize	Epochs	Learning rate	Acc	Sen	Esp	AUC
SpanBERT	RandomUnderSampling	10%	10	3e-05	0.9143	0.8462	0.9367	0.9464
BETO	**Unbalanced**	**10%**	**6**	**5e-05**	**0.9429**	**0.8846**	**0.962**	**0.9805**
multilingual	RandomUnderSampling	10%	9	5e-05	0.9333	0.9231	0.9367	0.9615
SpanBERT	RandomUnderSampling	20%	9	2e-05	0.8325	0.8333	0.8323	0.8701
BETO	RandomUnderSampling	**20%**	**6**	**5e-05**	**0.8852**	**0.8605**	**0.8916**	**0.9337**
multilingual	RandomUnderSampling	20%	6	5e-05	0.8469	0.8095	0.8563	0.9038
SpanBERT	RandomUnderSampling	30%	7	3e-05	0.8594	0.8182	0.8704	0.8963
BETO	RandomUnderSampling	**30%**	**10**	**3e-05**	**0.8754**	**0.8788**	**0.8745**	**0.9308**
multilingual	RandomUnderSampling	30%	7	5e-05	0.853	0.8788	0.8462	0.918

Tabla 5.6. Resultados de experimento en el primer Escenario.

5.7.2. Escenario 2

En el segundo escenario se entraron los modelos con los tweets con las condiciones de limpieza aplicadas. De los experimentos realizados, se seleccionó el mejor resultado por modelo y TestSize.

El mejor resultado para los tamaños de prueba de 10% y 30% se obtuvo con el modelo BETO, mientras que para el tamaño de 20% el modelo fue el multilingual. El mejor resultado se obtuvo con el experimento que se realizó con un balanceo RandomUnderSampling y con un TestSize del 10%, el cual obtuvo un AUC de 0.9834.

Este resultado, también de los rangos 25-29 supera al obtenido en el escenario 1, convirtiéndose así en el mejor obtenido de todos los experimentos y para ambos escenarios.

Modelo	Balanceo	TestSize	Epochs	Learning rate	Acc	Sen	Esp	AUC
SpanBERT	RandomUnderSampling	10%	9	5e-05	0.8667	0.9615	0.8354	0.9533
BETO	**RandomUnderSampling**	**10%**	**8**	**2e-05**	**0.9238**	**0.9615**	**0.9114**	**0.9834**
multilingual	RandomUnderSampling	10%	10	3e-05	0.9143	0.8846	0.9241	0.9304
SpanBERT	RandomUnderSampling	20%	7	5e-05	0.8469	0.8333	0.8503	0.9009
BETO	RandomUnderSampling	20%	5	5e-05	0.8612	0.9286	0.8443	0.939
multilingual	**RandomUnderSampling**	**20%**	**8**	**5e-05**	**0.8756**	**0.9286**	**0.8623**	**0.9304**
SpanBERT	RandomUnderSampling	30%	8	5e-05	0.885	0.8485	0.8947	0.9265
BETO	**RandomUnderSampling**	**30%**	**7**	**3e-05**	**0.8946**	**0.8939**	**0.8947**	**0.9367**
multilingual	RandomUnderSampling	30%	5	5e-05	0.8498	0.9091	0.834	0.9228

Tabla 5.7 Resultados de experimento en el segundo Escenario.

5.7.3. Análisis comparativo de modelos

Se realizará un análisis comparativo de los tres modelos utilizados, de acuerdo con su desempeño alcanzado en los indicadores Sensibilidad, Especificidad, Precisión y AUC. Se debe tener en cuenta que para los tres primeros indicadores se ordenaron los experimentos de acuerdo con los valores obtenidos y no por orden de ejecución de experimentos.

5.7.3.1. Sensibilidad

Según lo definido previamente, el indicador sensibilidad nos señala el porcentaje de textos etiquetados como “La mujer pasó por un proceso violento” que realmente pertenecen a dicha categoría. Teniendo en cuenta que el dataset es desbalanceado respecto a esta etiqueta, es normal encontrar registros que alcancen el valor 1.0 en este indicador. Esto, en contraste con otros indicadores puede llevar a un resultado no favorable.

Es así que en el gráfico 5.1 observamos que los valores más altos para el indicador "Sensibilidad" fueron obtenidos con el modelo multilingual, en contraste los resultados más bajos casi llegando al valor de 0.1 fueron obtenidos por el modelo SpanBERT.

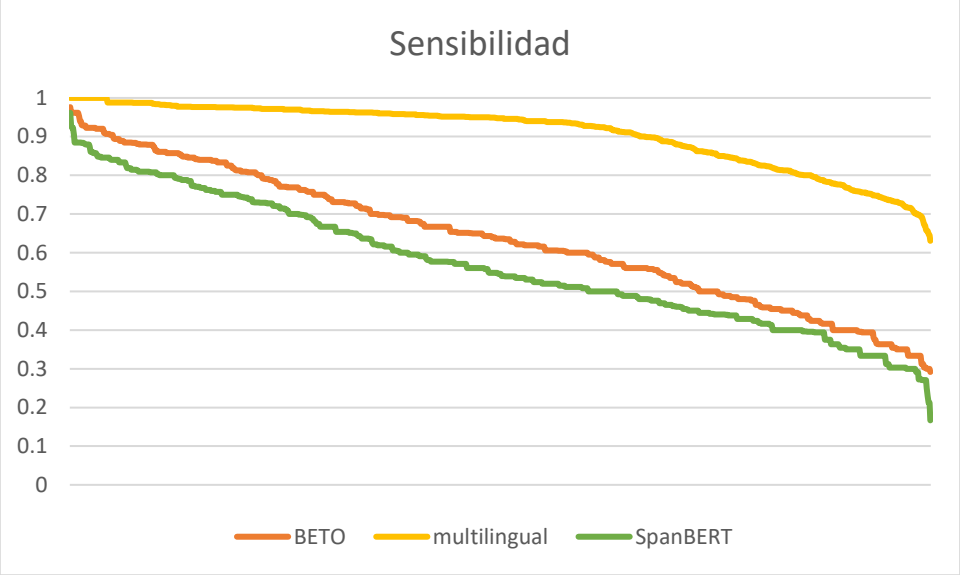


Gráfico 5.1 Análisis comparativo de modelos respecto a su desempeño para el indicador Sensibilidad

5.7.3.2. Especificidad

A diferencia del indicador de Sensibilidad, los valores obtenidos para el indicador de Especificidad fueron similares para todos los modelos (Véase gráfico 5.2), destacando levemente con valores más altos el modelo SpanBERT. Observamos también, que los resultados no llegan a desplomarse a números tan bajos como para el indicador de Sensibilidad, sino que se mantienen con valores superiores a 0.6.

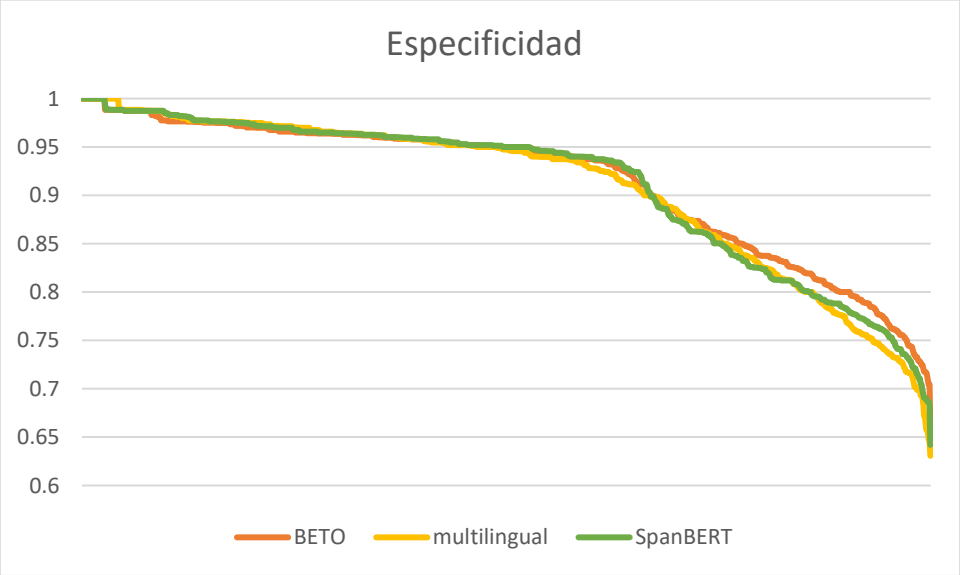


Gráfico 5.2 Análisis comparativo de modelos respecto a su desempeño para el indicador Especificidad

5.7.3.3. Precisión

Es uno de los indicadores más importantes, ya que es el porcentaje de textos correctamente clasificados por el modelo en ambas categorías. Observamos que, de forma similar al indicador de Especificidad, los valores más bajos son superior a 0.65. Sin embargo, ninguno de los valores obtiene un valor de 1.0.

El modelo con los mejores valores más altos es BETO, lo cual se condice con los resultados presentados en las secciones previas. Respecto a los modelos multilingual y SpanBert el gráfico nos muestra que obtuvieron resultados similares pero menores a los de BETO. (Véase Gráfico 5.3)

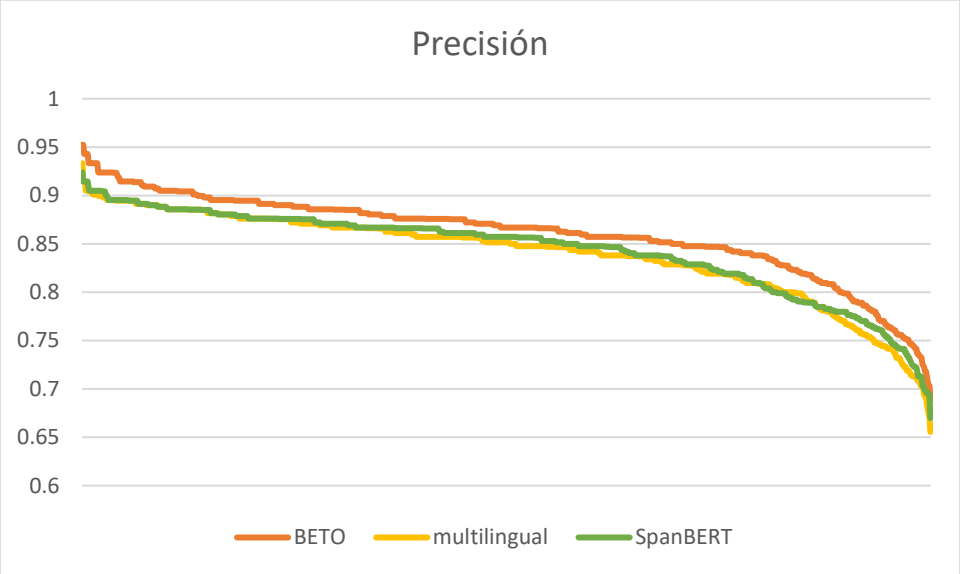


Gráfico 5.3. Análisis comparativo de modelos respecto a su desempeño para el indicador Precisión

5.7.3.4. AUC

En ambos escenarios, tanto con los tweets originales como los procesados por métodos de limpieza, los mejores resultados se obtuvieron con el modelo BETO (Véase gráfico 5.4). Estos resultados se condicen con lo presentado en las secciones 5.7.1. y 5.7.2. dónde de los 6 mejores resultados obtenidos, 5 de ellos se alcanzaron con el modelo “BETO”. Por otro lado, vemos que los experimentos con los modelos SpanBert y multilingual obtuvieron resultados similares.

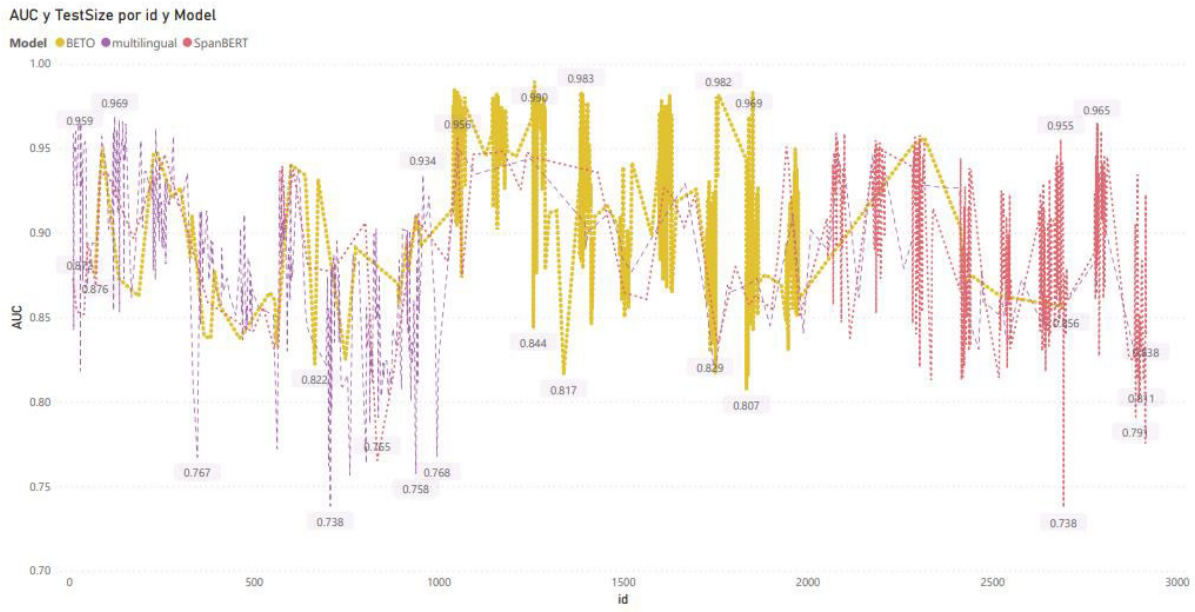


Gráfico 5.4 Análisis comparativo de modelos respecto a su desempeño para el indicador AUC

5.8. Plataforma web

Como parte de los objetivos, se diseñó una herramienta web utiliza el modelo seleccionado, de tal manera que el proceso de clasificación de un mensaje con contenido de violencia hacia la mujer es más sencillo e intuitivo. Se adjuntan capturas de la plataforma web construida con los resultados arrojados.

La figura 5.10 muestra la pantalla de inicio del sistema web. En esta se encuentra un primer mensaje dónde se nos indica el objetivo que tiene la plataforma, así como también que podemos realizar una prueba dando clic al botón “Nuevo”.

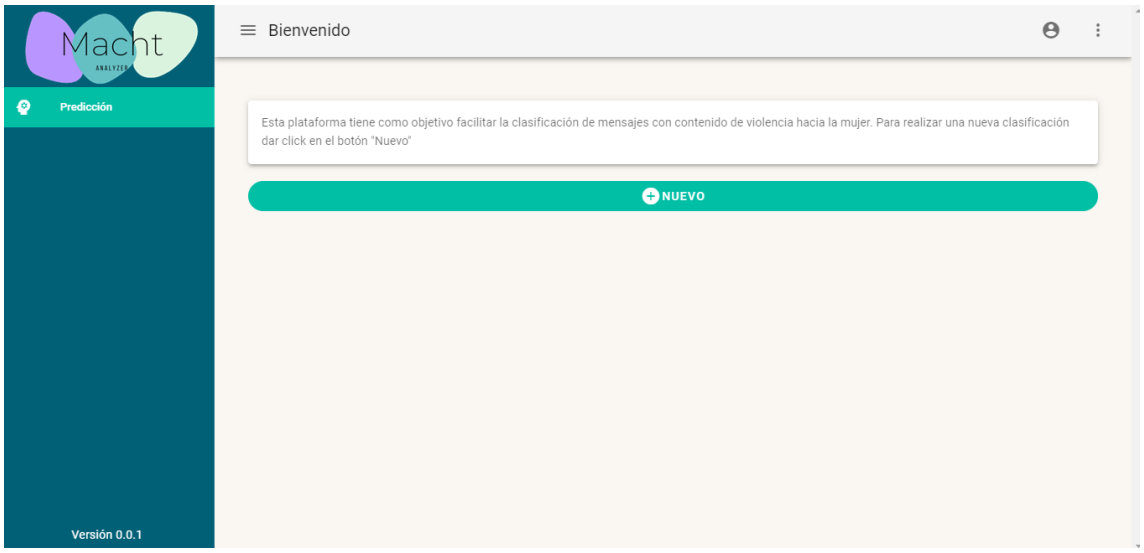


Figura 5.10. Pantalla de inicio del sistema

La siguiente pantalla se mostrará una vez se haya dado clic en el botón “Nuevo”. En esta, se muestra un mensaje de guía el cual nos indica que dependiendo del tipo de entrada (Texto o ID Tweet) a seleccionar se completarán diferentes tipos de campos (Véase Figura 5.11).

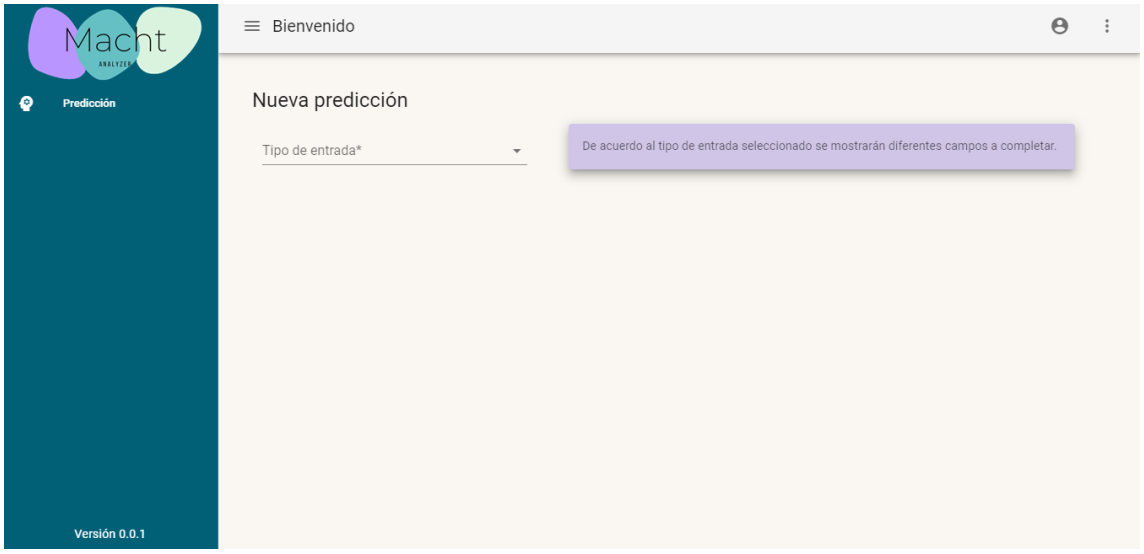


Figura 5.11. Pantalla de nueva predicción

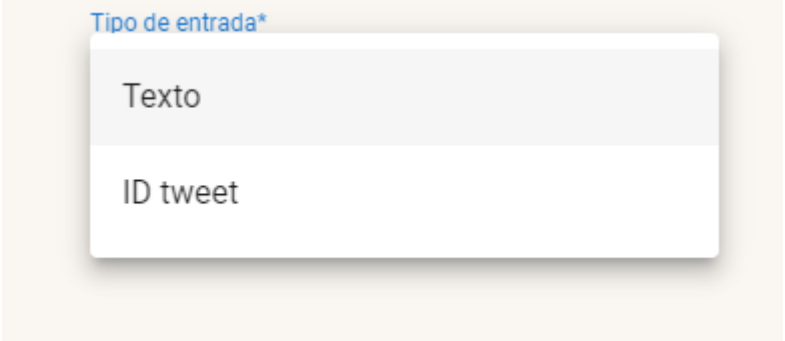


Figura 5.12. Tipos de entrada para nueva predicción

Si el tipo de entrada seleccionado fue “Texto”, el sistema nos solicitará escribir un mensaje a analizar (Véase Figura 5.13). Por otro lado, en el caso de “ID tweet” se nos solicitará ingresar el ID para poder realizar la búsqueda (Véase Figura 5.14).

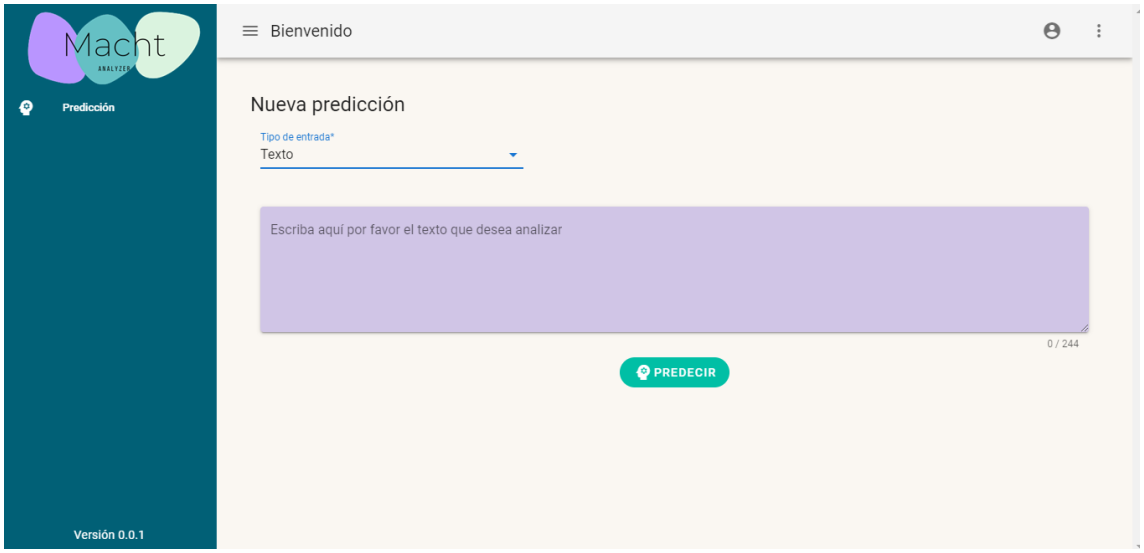


Figura 5.13. Pantalla de Nueva predicción con tipo de entrada “Texto”

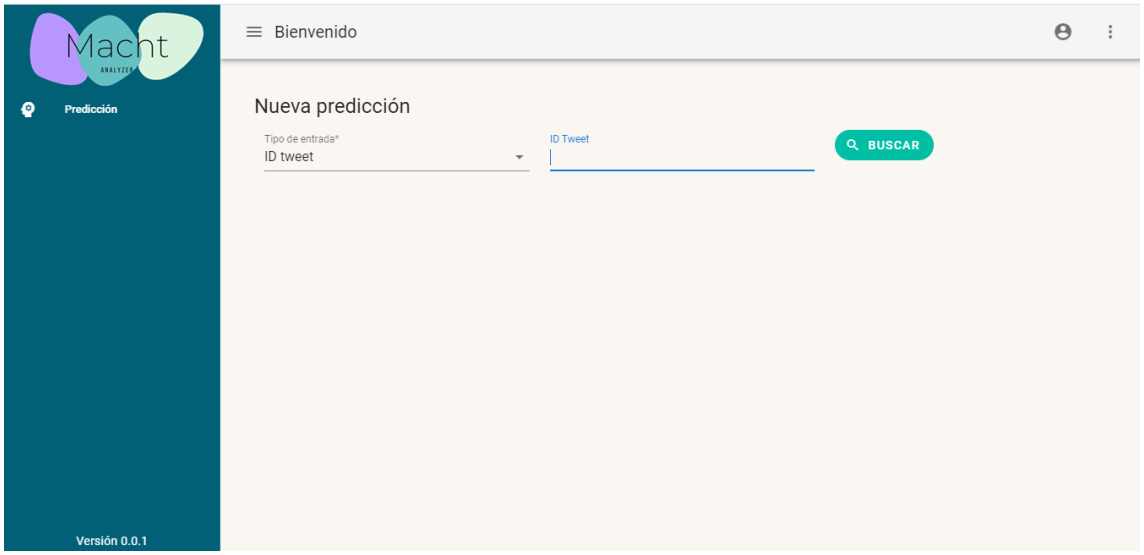


Figura 5.14. Pantalla de Nueva predicción con tipo de entrada “ID tweet”

Finalmente, una vez el texto haya sido ingresado, caso de entrada tipo Texto, o devuelto por la API (tipo de entrada ID tweet). Al darle clic al botón “Predecir” se nos mostrará el resultado la precisión de la clasificación.

Para el primer caso, cuando ingresamos el texto manualmente, en el ejemplo de la Figura 5.15 podemos observar que, con un testimonio de violencia, la herramienta nos arroja una precesión de 89.7 y la etiqueta de “Violence”.

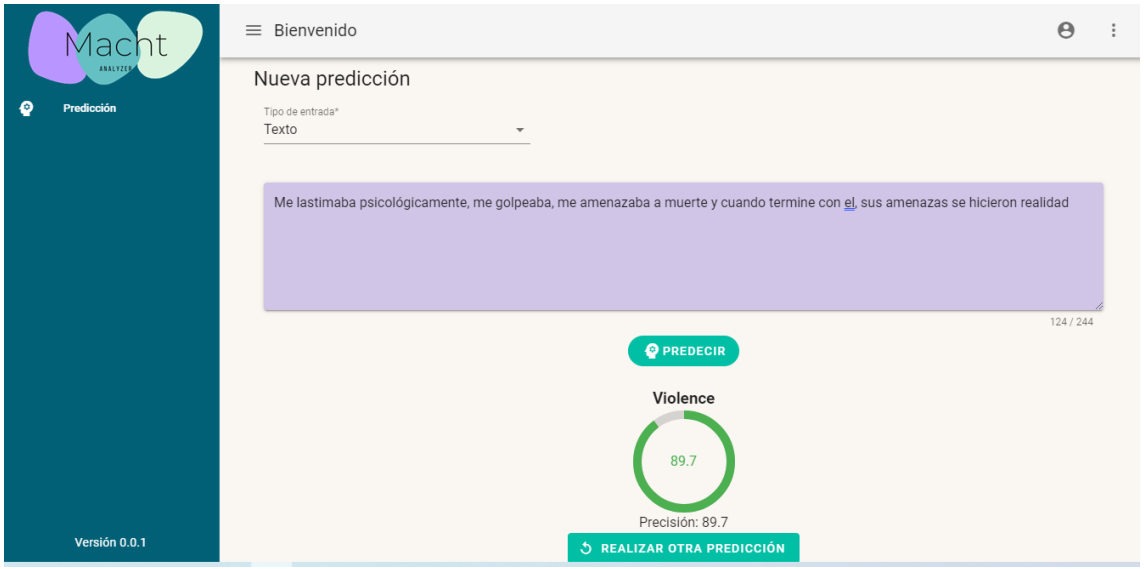


Figura 5.15. Ejemplo predicción con tipo de entrada Texto

Para el segundo caso, al ingresar el ID de un tweet. El sistema primero nos devuelve el texto y posteriormente realiza la predicción. La figura 5.16 nos muestra así una precisión de 89.53 y la etiqueta de “No Violence” para un Tweet (Ver figura 5.17) que no se encuentra relacionado al tema de violencia contra la mujer.

Figura 5.16 Ejemplo predicción con tipo de entrada ID Tweet

Figura 5.17. Ejemplo de Tweet

CAPÍTULO 6: CONCLUSIONES Y TRABAJOS FUTUROS

6.1. Conclusiones

6.1.1. Conclusión general

Se implementó una herramienta web utilizando técnicas de aprendizaje profundo, análisis de sentimiento y el modelo preentrenado BERT. La cual posibilita la clasificación de textos testimoniales en las categorías “La mujer pasó por un proceso violento” y “La mujer no pasó por un proceso violento”.

6.1.2. Conclusiones específicas

6.1.2.1. Objetivo específico 1

Se diseñó e implemento un modelo de aprendizaje capaz de clasificar testimonios de violencia contra la mujer en redes sociales como Twitter para así visibilizar la violencia. Se utilizó la versión entrenada *BETO-uncased* de BERT y fue escrito utilizando la librería Pytorch.

6.1.2.2. Objetivo específico 2

Se implementó una solución web que facilita la clasificación y detección de testimonios de violencia.

6.1.2.3. Objetivo específico 3

Se realizó la construcción de un dataset en español de 1042 mensajes de la red social de Twitter el cual recoge la percepción sobre testimonios de violencia de un total de 22 individuos pertenecientes a 3 rangos de edad diferentes.

6.2. Limitaciones

El presente trabajo ha utilizado únicamente un dataset de 1042 mensajes debido a la dificultoso que se les hacía a los sujetos de estudio la clasificación de un gran volumen de mensajes en las 2 categorías planteadas.

Para que la precisión de la clasificación obtenga un resultado más alto se necesitaría un dataset más amplio y balanceado.

6.3. Trabajos futuros

Futuros estudios podrían contemplar la mejora del modelo de aprendizaje profundo para que cuente con la capacidad de predecir la edad y género de una persona de acuerdo con la percepción que posea sobre un testimonio de violencia.

Referencias

(GRADE), G. d. (2019). *Violencias contra las Mujeres, la necesidad de un doble plural.* Lima.

Agarwal, N., Gupta, R., Singh, S. K., & Saxena, V. (2017). Metadata based multi-labelling of YouTube videos. *IEEE*, 586-590.

Albadiy, N., Kurdiz, M., & Mishra, S. (2018). Are they our brothers? Analysis and detection of religious hate speech in the Arabic Twittersphere. *IEEE*, 69-76.

AL-GARADI, M. A., HUSSAIN, M. R., KHAN, N., MURTAZA, G., NWEKE, H. F., ALI, I., . . . GANI, A. (2019). Predicting Cyberbullying on Social Media in the Big Data Era Using Machine Learning Algorithms: Review of Literature and Open Challenges. *IEEE Access, 7*, 70701-70718.

ALSaif, H., & Alotaibi, T. (2019). Arabic Text Classification using Feature-Reduction Techniques for Detecting Violence on Social Media. *(IJACSA) International Journal of Advanced Computer Science and Applications,, 10*(4).

Amaya Balaguera, Y. D. (2015). Metodologías ágiles en el desarrollo de aplicaciones para dispositivos móviles. Estado actual. *Revista de Tecnología, 12*(2), 111-124.

Briere, J., & Jordan, C. E. (2004). Violence Against Women: Outcome Complexity and Implications for Assessment and Treatment. *19*, 1252-1276.

Cañete, José, Chaperon, Gabriel, Fuentes, Rodrigo, Ho, Jou-Hui, Kang, Hojin, & Pérez, Jorg. (2020). Spanish Pre-Trained BERT Model and Evaluation Data. *CaneteCFP2020.* Retrieved from https://users.dcc.uchile.cl/~jperez/papers/pml4dc2020.pdf

Dai, A., & Le, Q. (2015). Semi-supervised sequence learning. *Advances in Neural Information Processing Systems (NIPS) 28 Conference Proceedings*, 3079-3087. Retrieved from https://dl.acm.org/doi/10.5555/2969442.2969583

Demirci, G. M., Keskin, S. R., & Dogan, G. (2019). Sentiment Analysis in Turkish with Deep Learning. *IEEE*, 2215-2221.

Deng, L., & Liu, Y. (2018). *Deep Learning in Natural Language Processing.* Singapur : Springer.

Deng, X., Liu, Q., Deng, Y., & Mahadevan, S. (2016). An improved method to construct basic probability assignment based on the confusion matrix for classification problem. *Information Sciences, 340-341*, 250-261. doi:10.1016/j.ins.2016.01.033

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North*. doi:10.18653/v1/n19-1423

Dirección general contra la Violencia de Género. (2018). *Observatorio Nacional de la Violencia contra las Mujeres y los Integrantes del Grupo Familiar.* Recuperado el 17 de Febrero de 2021, de https://observatorioviolencia.pe/wp-content/uploads/2018/08/GUIAds-006-2018-mimp-protocolo-actuacion-conjunta-cem-comisarias.pdf

Flores, D. A. (2018). *Conocer para Resistir violencia de género en línea en Perú.* Lima: Asociación Civil Hiperderecho.

García-Díaz, J. A., Cánovas-García, M., Colomo-Palacios, R., & Valencia-García, R. (2020). Detecting misogyny in Spanish tweets. An approach based on linguistics features and word embeddings. *Elsevier*.

García-Moreno, C., Guedes, A., & Knerr, W. (2013). *OPS.* Recuperado el 14 de Febrero de 2021

Gonzales-Carbajal, S., & Garrido-Merchán, E. (2020). Comparing BERT against traditional machine learning text classification. *ArXiv, arXiv:2005.13012v2*.

Google. (2019, 10 17). *GitHub*. Retrieved 07 2021, 24, from https://github.com/google-research/bert/blob/master/multilingual.md

Google Research. (2019). Retrieved from https://github.com/google-research/bert

Ho, T. K. (1995). Random decision forests. *IEEE, 1*, 278-282.

Howard, J., & Ruder, S. (2018). Universal language model fine-tuning for text classification. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, *1*, pp. 328-339. Melbourne, Australia. doi:10.18653/v1/P18-1031

Instituto de Opinión Pública de la Pontificia Universidad Católica del Perú. (24 de 11 de 2016). *Pulso PUCP.* Obtenido de Pulso PUCP.

Instituto Nacional de Estadística e Informática - Encuesta Nacional de Hogares. (2021). *Instituto Nacional de Estadística e Informática - Encuesta Nacional de Hogares.* Lima.

International, A. (2018). *Amnesty.org*. Recuperado el 2021 de Febrero de 17, de https://www.amnesty.org/en/latest/research/2018/12/rights-today-2018-violence-against-women-online/

Joshi, M., Chen, D., Liu, Y., Weld, D., Zettlemoyer, L., & Levy, O. (2020). SpanBERT: Improving Pre-training by Representing and Predicting Spans. *Transactions of the Association for Computational Linguistics*(8), 64-77. doi:10.1162/tacl_a_00300

Kitchenham, B., & Charters, S. (2007). Guidelines for Performing Systematic Literature Reviews in Software Engineering. Keele, Staffs .

Kotzé, E., Senekal, B. A., & Daelemans, W. (2020). Automatic classification of social media reports on violent incidents in South Africa using machine learning. *South African Journal of Science, 116*(3-4), 1-8.

Ministerio de la Mujer y Poblaciones Vulnerables. (2015). *Ministerio de la Mujer y Poblaciones Vulnerables.* Recuperado el 17 de Febrero de 2021, de https://repositorio.aurora.gob.pe/bitstream/handle/20.500.12702/12/mimp-marco-conceptual-violencia-basada-en-genero.pdf?sequence=1&isAllowed=y

Organización Mundial de la Salud. (2013). *Organización Mundial de la Salud.* Recuperado el 14 de Febrero de 2021, de https://www.who.int/reproductivehealth/publications/violence/9789241564625/es/

Organización Mundial de la Salud. (29 de Noviembre de 2017). *Organización Mundial de la Salud*. Obtenido de https://www.who.int/es/news-room/fact-sheets/detail/violence-against-women

Organizaciones de América Latina. (2017). *REPORTE DE LA SITUACIÓN DE AMÉRICA LATINA SOBRE LA VIOLENCIADE GÉNERO EJERCIDA POR MEDIOS ELECTRÓNICOS.*

Perú, P. U., Pública, I. d., & Ramos, M. M. (21 de 11 de 2019). *Repositorio Institucional de la PUCP.* Recuperado el 30 de 05 de 2021, de Repositorio Institucional de la PUCP: http://repositorio.pucp.edu.pe/index/handle/123456789/168793

Peters, M., Neumann, M., Zettlemoyer, L., & Yih, W.-t. (2018). Dissecting Contextual Word Embeddings: Architecture and Representation. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 1499-1509. doi:10.18653/v1/D18-1179

Platt, J. (1998). Sequential minimal optimization: A fast algorithm for training support vector machines. *Technical Report MSR-TR-98-14*.

Plaza-del-Arco, F., Molina-González, M., Ureña-López, L., & Martín-Valdivia, M. (2021). Comparing pre-trained language models for Spanish hate speech detection. *Expert Systems with Applications, 166*. doi:10.1016/j.eswa.2020.114120

Programa Naciona Aurora. (2019). *Casos Atendidos por el CEM al 31 de Diciembre de 2019.*

Programa Nacional Aurora. (2020). *Casos Atendidos por el CEM al 31 de Diciembre de 2020.*

Programa Nacional Aurora. (2020). *REPORTE ESTADÍSTICO DE CONSULTAS TELEFÓNICAS ATENDIDAS EN LINEA100 - Periodo: Enero - Diciembre 2020 (Preliminar).*

Quijano-Sanchez, L., Pereira Kohatsu, J. C., Liberatore, F., & Camacho-Collados, M. (2019, 03 13). *HaterNet a system for detecting and analyzing hate speech in Twitter (Version 1.0) [Data set]*. (Zenodo) Retrieved 07 24, 2021, from https://zenodo.org/record/2592149#.YPsWdmSXONx

Radford, A. (2018). *Improving language understanding with unsupervised learning.* OpenAI. Retrieved from https://openai.com/blog/language-unsupervised/

Risch, J., Stoll, A., Ziegele, M., & Krestel, R. (2019). Offensive Language Identification using a German BERT Model. *15th Conference on Natural Language Processing.*

Roy, T., McClendon, J., & Hodges, L. F. (2018). Analyzing Abusive Text Messages to Detect Digital Dating Abuse. *IEEE*, 284-293.

Russell, D. E. (1992). *Femicide : the politics of woman killing.* Buckingham.

Sayanta, P., & Sriparna, S. (2020). CyberBERT: BERT for cyberbullying identification. *Springer, 1-8*.

Schrading, J. N. (2015). *Analyzing Domestic Abuse using Natural Language Processing on Social Media Data.* Rochester Institute of Technology.

SUBRAMANI, S., MICHALSKA, S., WANG, H., DU, J., ZHANG, Y., & SHAKEEL, H. (2019). Deep Learning for Multi-Class Identification From Domestic Violence Online Posts. *IEEE Access, 7*, 46210-46224.

SUBRAMANI, S., WANG, H., & VU, H. Q. (54075-54085). Domestic Violence Crisis Identification From Facebook Posts Based on Deep Learning. *IEEE Access, 6*.

Sun, C., Qiu, X., Xu, Y., & Huang, X. (2019). How to Fine-Tune BERT for Text Classification? *China National Conference on Chinese Computational Linguistics*, 194-206.

Ting, D., Peng, L., Varadarajan, A., Keane, P., Burlina, P., Chiang, M., . . . Wong, T. (2019). Deep learning in ophthalmology: The technical and clinical considerations. *Progress in Retinal and Eye Research, 72*. doi:10.1016/j.preteyeres.2019.04.003

Udanor, C., & Anyanwu, C. C. (2019). Combating the challenges of social media hate speech in a polarized society. A Twitter ego lexalytics approach. *Data Technologies and Applications, 53*(4), 501-527.

United Nations Office on Drugs and Crime. (2019). *United Nations Office on Drugs and Crime.* Recuperado el 14 de Febrero de 2021, de https://www.unodc.org/documents/data-and-analysis/gsh/Booklet_5.pdf

Yadav, J., Kumar, D., & Chauhan, D. (s.f.). Cyberbullying Detection using Pre-Trained BERT Model. *IEEE*, 1096-1100.

ANEXO A

Detalle de los resultados de los 2916 experimentos

Debido a la extensión de la tabla con los 2916 experimentos se adjunta un enlace drive hacia un archivo de tipo hoja de cálculo con los resultados.

Enlace: ResultadosExperimentos

El documento solo podrá ser accedido a modo de lectura por usuarios del grupo UNMSM.

ANEXO B

Mensajes analizados para la construcción del dataset

Se adjunta 100 de los 1024 mensajes traducidos para la construcción del dataset. Los tweets fueron recopilados en el marco de una campaña lazada en la red social Twitter con los hashtag #PorqueMeQuedé #PorqueMeFui; dónde muchas mujeres compartieron las historias de violencia vividas principalmente en sus hogares y porque decidieron irse o quedarse en ellos.

#PorqueMeQuedé,En qué se equivoca el Huffington Post http://t.co/TH6nZufycp #valentine
#PorqueMeQuedé,"Si tú me ignoras, yo te ignoraré. Si tú no empiezas la conversación, no hablaremos. Si no te esfuerzas, por qué debería .. "
#PorqueMeQuedé, "En #ViolenciaDoméstica el tema #Refugio el núcleo de muchas discusiones. Aquí están algunas de las razones #Poesía
#PorqueMeFui, 7 dolorosos años después, acepté que posesión y amor no son los mismo.
#PorqueMeQuedé,"porque creo que la gente puede cambiar con la ayuda adecuada. No todas las relaciones abusivas permanecen abusivas. Hay espacio para el cambio"
#PorqueMeQuedé,El abuso no siempre deja moretones o cicatrices. Sempre se trata del control y el poder sobre la víctima. Es un recordatorio de esto.
#PorqueMeFui,"Después de la muerte de su CEO, comenzaron a hacerme cosas abusivas como sincronizar música que odio en mi teléfono"
#PorqueMeQuedé, " Vean""Retrato del abuso: Una pandemia Americana. http://t.co/szGQBXe0Vk #abusodomestico #arteterapia""
#PorqueMeQuedé,"Espero a alguien que responda con un ""ninguno de tus malditos asuntos.""""
#PorqueMeQuedé,"Porque quería creerle cada vez que decía que sería diferente"
#PorqueMeQuedé,"Al principio fue porque simplemente no sabía que la forma en la que me trataba era equivocada y abusiva."
#PorqueMeQuedé,Por qué me quedé... http://t.co/azQkDmt3qZ #laconversaciónsinfin
#PorqueMeQuedé,"Mujeres, jóvenes, chicas. Cuanto más hables de libertad e igualdad, más tus narices están relacionadas con Tu destino, tu elección "
#PorqueMeQuedé,"Porque tuve un hijo con él "
#PorqueMeFui,"Aprendí a amarme más "
#PorqueMeFui,"Mis hijas merecían un mejor ejemplo "
#PorqueMeQuedé," Nadie creía que fuera tan malo como lo era incluso después de todas las frecuentes visitas"

#PorqueMeQuedé,"pensé que no era gran cosa que no pudiera encontrar la voluntad de vivir a menos que le dedicara todo mi tiempo"
#PorqueMeQuedé,"Porque pensé que de algún modo, yo debía ser la causa de sus arrebatos violentos y los malos estados de ánimo"
#PorqueMeQuedé,"Mi hija estaba #asustada #desolada #amenazada por Calvin Jones Fecha de nacimiento 5/25/1985, por lo que guardo silencio y soportó años de abuso. :-("
#PorqueMeFui,"Se que es valioso hablar de ello a pesar de que es doloroso. Conozco a muchas amigas que aún no lo hacen &amp
#PorqueMeFui,"@AC360 Lo amaba, pero aprendí a amarme a mí misma lo suficiente para parar de permitirle tratarme de esa forma"
#PorqueMeFui,"Quería una mejor vida por mi hijo y mi felicidad, no pelear, y probar que soy mejor que eso "
#PorqueMeQuedé," Porque era ""amor verdadero"" si dolía, y la manera en la que era tratada era ""mi culpa"" "
#PorqueMeFui,AMEN! RT @rascality: r @Sil_Lai: B/C No quería que mis hijos crezcan pensando que la violencia era normal en las relaciones.
#PorqueMeQuedé,"Porque no quería el peso del suicidio de alguién más sobre mis hombros. "
#PorqueMeFui," pregunta: Cuando es suficiente? Se necesita coraje y fuerza en ambos lados de esta ecuación. Apoyar hasta que pueda ayudarse a sí mismo."
#PorqueMeQuedé,"Me alejó de todos mis amigos &amp
#PorqueMeQuedé,@JustAnn12 Los objetos rotos también abusan, luego se volvieron ocasionalmente físicos.
#PorqueMeQuedé, Porque todos me dijeron que lo merecía.
#PorqueMeQuedé," Nunca tuve nada que dejar. JAJAJA"
#PorqueMeQuedé,Creo que la #independencia esta lejos de la #culturaCorporativa. #Martes de tacos es otra razón. http://t.co/DQ9dPMB5VY #Martes
#PorqueMeQuedé,"Me hubiera perseguido. Era un sociopata con muchas armas, 12 años mayor que yo. Me divorcié de él en Alemania mientras estaba en el ejército "
#PorqueMeQuedé,"Porque nadie que tenga un reconocimiento, hable tres idiomas, y toque dos instrumentos es abusado. Si fueran tan lista... Cierto? "
#PorqueMeFui, Él amenazó a mis hijos.

#PorqueMeQuedé," Estas narrativan son super poderosas y super desencadenantes. http://t.co/LeTYVd5iWw "
#PorqueMeQuedé," Porque las chicas gordas deberían estar felices incluso de tener novio."
#PorqueMeQuedé," Porque no sucedió de la noche a la mañana, poco a poco permití que abusara de mí y para entonces yo estaba emocionalmente &amp
#PorqueMeQuedé," A dónde podría ir? Tú tomaste todo lo que tenía."
#PorqueMeQuedé,"Pastores &amp
#PorqueMeQuedé," Porque que pasaba si nadie más me quería"
#PorqueMeFui," ..Una vez fue suficiente. Me apreciaba y amaba a mí misma más que los sentimientos que tenia por él. Me merecía algo mejor."
#PorqueMeQuedé," Me quedé porque no quería quitarle la casa a mi hijo, él ya había perdido bastante."
#PorqueMeQuedé,"Él me hizo creer que mi famila y mis amigos me odiaban y que él era el único de quien podía depender "
#PorqueMeFui,"Para darle a mi hijo la oportunidad de aprender a respetar a las mujeres, a las madres, y a mí. Y porque lo amo. http://t.co/t3LRhoNkVl"
#PorqueMeQuedé,"A1: Miedo a represalías. A menudo el abuso escala cuando la víctima trata de escapar #TEARtalk "
#PorqueMeFui,"Mi hija era la siguiente "
#PorqueMeQuedé," Pensé que significaba que era fuerte, pero en realidad era adicta al dolor "
#PorqueMeFui, Fue la última llamada
#PorqueMeQuedé," Es por eso que le digo a mis niñas que esten seguras de que pueden mantenerse a sí mismas. Quiero que algún día quieran a su marido, que no lo necesiten"
#PorqueMeQuedé, A esta campaña le quedan 7 horas! https://t.co/5tCmzZPaOA #trauma #sobrevivientes #abuso
#PorqueMeQuedé,: porque un hueso nunca es roto. Solo el espíritu. Solo que nunca el espíritu.
#PorqueMeQuedé," Porque pensaba que maquillarse después de una gran pelea haría que todo este bien"
#PorqueMeQuedé," Porque me convenció de que su interpretación de las cosas que experimenté era más válida que la mía."

#PorqueMeQuedé,@BarackObamalas mujeres y niños se han ido a deshacerse de ellos, la escoria del taxista!!!hola(@trpresidency turkey) #cnn @bbc dumb
#PorqueMeQuedé," desafortunadamente no hay protección para alguien como yo. Quédate tú y arruinada tu vida"
#PorqueMeQuedé, Quieres saber más sobre ello? Busca más información de manera anónima a través de http://t.co/4xouZwq2px
#PorqueMeFui," Quiero un hombre que este orgulloso de mí. Que vea mas que mi figura, mis inseguridades y mis caídas "
#PorqueMeQuedé," Porque nunca me enseñaron lo que era el amor de verdad"
#PorqueMeQuedé",„ Pensé que era una mujer inteligente, fuerte, independiente enamorada de un hombre problemático, a quien estaba tratando de ayudar.“ http://t.co/0dM9E0oeEb"
#PorqueMeQuedé," Porque él me dijo que nadie más me iba a querer &amp
#PorqueMeQuedé," Los abusadores deforman las nociones de relaciones apropiadas. Ellos deforman tus nociones de la realidad. Lo que es inaceptable repentinamente se torna gris. "
#PorqueMeFui,"Él abuso sexualmente de mí mientras dormía "
#PorqueMeQuedé," Muchas mujeres negras aman a los matones y creen que ser golpeadas es una insignia de honor"
#PorqueMeQuedé,IB Psych: termina la actividad para que podamos ver la charla TED mañana!
#PorqueMeQuedé," Vi sus gritos, sus golpes, como manifestaciones de un hombre consumido por una tóxica masculinidad para la que necesitaba hacer espacio. "
#PorqueMeQuedé," Todavía no puedo creer que dije eso públicamente, eso solo demuestra lo embarazoso y la vergüenza que viene con el abuso "
#PorqueMeQuedé,"@mkope #negaciones un verdadero cobarde que golpea a mujeres / niños / animales / cualquier cosa más débil, pero IGUAL DE ENFERMOS son sus " "amigos" "que los apoyan"
#PorqueMeQuedé,"porque perderás a tus hijos si te vas, corte familiar, evaluadores de custodia que creen que las mujeres mienten asi que que le quitan a sus hijos "
#PorqueMeQuedé," Viendo y viviendolo toda mi vida con mis padres, empiezas a creer que es solo una relación normal con un tipo.
#PorqueMeQuedé," Porque estaba asustada y tenía solo 17 años "

#PorqueMeQuedé," porque su madre seguía agradeciéndome el haberlo "salvado". Y me gustaba su madre."
#PorqueMeQuedé," Me quedé porque me dijeron que Dios odia el divorcio. Me fui porque me di cuenta que Dios odia aún más el abuso. "
#PorqueMeQuedé,"Porque estaba asustada de no poder hacerlo por mí misma."
#PorqueMeFui,RT @misslori: #comoelabuso Él tenía relaciones conmigo mientras dormía. #violenciadoméstica #mltvsalud
#PorqueMeFui, No quería que mi familia y amigos visiten mi tumba todavía.
#PorqueMeQuedé," Porque yo tenía 19 años y era amor verdadero, y él estaba muy arrepentido. Lloró por teléfono y me rogo que volviera. Me doblé."
#PorqueMeFui," Me ayudó un extraño, luego mi familia
Hace 36 años soy un extraño. Te ayudaré. ♥"
#PorqueMeQuedé,: Como algunas iglesias apoyan el abuso conyugal http://t.co/icQrB84ocx via @RNS
#PorqueMeQuedé,"He conocido hombres que se quedaron emocionalmente y &amp
#PorqueMeFui," porque tuve que irme"
#PorqueMeFui," https://t.co/825G66SO8y. Ver y por favor RT. No te quedes. No se pone mejor."
#PorqueMeFui," Así que podría vivir. Porque fui apoyada por buenas personas &amp
#PorqueMeQuedé," A cualquier que sienta miedo de irse - El día que nos fuimos, fue el día en el que mis hijos y yo empezamos a vivir. 😊 "
#PorqueMeQuedé,"Él me convenció de que era un santo por soportar la pesada ""carga"" que era amarme "
#PorqueMeFui,Casi me mata
#PorqueMeFui," Me di cuenta de que valía más. Quería reanudar el contacto con mis amigos. Tomé la responsabilidad de mi vida. Mi lección fue acumulada "
#PorqueMeQuedé," No tener nada te hace apreciar todo# abusada pero feliz#nodoynadaporsentado"
#PorqueMeQuedé",#NuncaCompresVidasdeNuevo Vive para siempre Juega para siempre con ""Mi Candy"" en amazon D mejor juego de caramelos en el mundo D http://t.co/h1mfha9Vum "
#PorqueMeQuedé," nos muestra que dejar relaciones abusivas no siempre es tan fácil como puede parecer http://t.co/N6Oh2uTbeZ"

#PorqueMeQuedé," Porque no podía pedirle a mi pobre madre que me dejará volver a entrar &amp
#PorqueMeFui, Porque ella no quería que sus hijos pensaran que era normal
#PorqueMeFui, porque no iba a darle una tercera oportunidad.
#PorqueMeFui," Si un chico hace preguntas, algo le pasa! "
#PorqueMeFui,: Opciones de universidad para mi
#PorqueMeQuedé," mis amigos trataron de intervenir, pero no estaba preparada para ser salvada Eventualmente dejaron de intentarlo porque no estaban preparados para quedarse &amp
#PorqueMeQuedé," Porque mi pastor me dijo que Dios odia el divorcio. No cruzó por mi mente que tal vez Dios odia el abuso también. "
#PorqueMeFui," Fui mi culpa que los hombres miraran en mi dirección."
#PorqueMeFui,"No estaba comprometida a casarme con su familia, me comprometí a casarme con él "

ANEXO C

Matriz de consistencia

PROBLEMAS	OBJETIVOS	HIPÓTESIS
Problema general	Objetivo General	Hipótesis General
¿La construcción de una herramienta capaz de clasificar mensajes de redes sociales como Twitter en categorías “La mujer pasó por un proceso violento” y “La mujer no pasó por un proceso violento" serviría para visibilizar la violencia contra la mujer?	Construir una herramienta capaz de clasificar mensajes de redes sociales como Twitter en categorías de “La mujer pasó por un proceso violento” y “La mujer no pasó por un proceso violento” que sirva para visibilizar la violencia contra la mujer.	Una herramienta capaz de clasificar mensajes de redes sociales como Twitter en categorías “La mujer pasó por un proceso violento” y “La mujer no pasó por un proceso violento" sirve para visibilizar la violencia contra la mujer.
Problemas específicos	Objetivos Específicos	Hipótesis Específicas
1. ¿El diseño e implementación de un modelo de aprendizaje profundo permite clasificar testimonios sobre violencia en redes sociales como Twitter en las categorías de “La mujer pasó por un proceso violento” y “La mujer no pasó por un proceso violento”? 2. ¿El diseño de una herramienta web que permita el uso del modelo, facilitará la clasificación y detección de testimonios de violencia? 3. ¿La construcción de un dataset en español de 1042 mensajes de la red social de Twitter puede recoger la percepción sobre testimonios de violencia?	1. Diseñar e implementar un modelo de aprendizaje profundo que permita clasificar testimonios sobre violencia en redes sociales como Twitter en las categorías de “La mujer pasó por un proceso violento” y “La mujer no pasó por un proceso violento”. 2. Diseñar una herramienta web que permita hacer el uso del modelo, de tal manera que facilite la clasificación y detección de testimonios de violencia. 3. Construir un dataset en español de 1042 mensajes de la red social de Twitter, a partir de la traducción al idioma español del dataset propuesto por Schrading (Schrading, 2015), que recoja la percepción sobre testimonios de violencia.	1. Un modelo de aprendizaje profundo permite clasificar testimonios sobre violencia en redes sociales como Twitter en las categorías de “La mujer pasó por un proceso violento” y “La mujer no pasó por un proceso violento”. 2. Una herramienta web que hace uso del modelo facilita la clasificación y detección de testimonios de violencia. 3. Un dataset en español de 1042 mensajes de la red social de Twitter recoge la percepción sobre testimonios de violencia.