



**UNIVERSIDAD POLITÉCNICA SALESIANA
SEDE QUITO**

CARRERA DE INGENIERÍA DE SISTEMAS

**ESTUDIO DESCRIPTIVO DE LA VIOLENCIA DE GÉNERO EN ECUADOR
USANDO TÉCNICAS DE MINERÍA DE DATOS**

Trabajo de titulación previo a la obtención del
Título de Ingeniero de Sistemas

AUTORES: PABLO ALEXANDER CAISA COYAGO
ERICK OSWALDO ROSERO ESPINOSA

TUTOR: JULIO RICARDO PROAÑO ORELLANA

Quito – Ecuador

2023

CERTIFICADO DE RESPONSABILIDAD Y AUTORÍA DEL TRABAJO DE TITULACIÓN

Nosotros Pablo Alexander Caisa Coyago, con documento de identificación N° 1751596261 y Erick Oswaldo Rosero Espinosa con documento de identificación N° 1723382345; manifestamos que:

Somos los autores y responsables del presente trabajo; y, autorizamos a que sin fines de lucro la Universidad Politécnica Salesiana pueda usar, difundir, reproducir o publicar de manera total o parcial el presente trabajo de titulación.

Quito, 14 de julio del año 2023

Atentamente,



Pablo Alexander Caisa Coyago

1751596261



Erick Oswaldo Rosero Espinosa

1723382345

CERTIFICADO DE SESIÓN DE DERECHOS DE AUTOR DEL TRABAJO DE TITULACIÓN A LA UNIVERSIDAD POLITÉCNICA SALESIANA

Nosotros, Pablo Alexander Caisa Coyago con documento de identificación No. 1751596261 y Erick Oswaldo Rosero Espinosa con documento de identificación No. 1723382345, expresamos nuestra voluntad y por medio del presente documento cedemos a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud de que somos autores del Artículo Académico: "Estudio Descriptivo de la Violencia de Género en Ecuador Utilizando Técnicas de Minería de Datos", el cual ha sido desarrollado para optar por el título de: Ingeniero de Sistemas, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En concordancia con lo manifestado, suscribimos este documento en el momento que hacemos la entrega del trabajo final en formato digital a la Biblioteca de la Universidad Politécnica Salesiana.

Quito, 14 de julio del año 2023

Atentamente,



Pablo Alexander Caisa Coyago

1751596261



Erick Oswaldo Rosero Espinosa

1723382345

CERTIFICADO DE DIRECCIÓN DEL TRABAJO DE TITULACIÓN.

Yo, Julio Ricardo Proaño Orellana con documento de identificación No. 0103909412, docente de la Universidad Politécnica Salesiana, declaro que bajo mi tutoría fue desarrollado el trabajo de titulación: ESTUDIO DESCRIPTIVO DE LA VIOLENCIA DE GÉNERO EN ECUADOR USANDO TECNICAS DE MINERIA DE DATOS, realizado por Pablo Alexander Caisa Coyago, con documento de identificación N° 1751596261 y Erick Oswaldo Rosero Espinosa con documento de identificación N° 1723382345, obteniendo como resultado final el trabajo de titulación bajo la opción de Artículo Académico que cumple con todos los requisitos determinados por la Universidad Politécnica Salesiana.

Quito, 14 de julio del año 2023

Atentamente,



Ing. Julio Ricardo Proaño Orellana, PhD

0103909412

DEDICATORIA

El presente proyecto de Titulación va dedicado, A mi madre Luzmila Coyago Ortuño quien ha sido mi motor y ayuda en el transcurso de mis estudios, a mi novia Jennifer Mesias a pesar de todas las dificultades no ha permitido que me dé por vencido, para dirigirme por el camino correcto, para cumplir mi sueño que más anhelaba desde que era un niño.

A mis Hermana Tahiz Morocho y personas más allegadas, quienes me han enseñado con su sabiduría a dignificar el trabajo, a Dios por haberme dado salud y vida. A mis abuelos por haberme brindado su amor incondicional y forjarme en la vida como un hombre de bien en la sociedad.

Pablo Alexander Caisa Coyago

Quiero agradecer de manera especial a mis padres a María Espinosa y Raúl Rosero por el constante apoyo para poder culminar con mis estudios. Que siempre estuvieron a mi lado en momentos difíciles, por sus consejos, que siempre fueron mi inspiración, esta meta cumplida también les pertenece a ellos. También quiere agradecer a mis hermanos y a mi familia por el apoyo, que han estado ahí en los momentos buenos y malos de mi vida.

Erick Oswaldo Rosero Espinosa

AGRADECIMIENTOS

Quiero expresar mi agradecimiento a la Universidad Politécnica Salesiana por abrirme las puertas de esta noble institución, a los docentes por impartir los conocimientos que necesitaba durante todo el transcurso de mi carrera universitaria.

A Dios por darme las mejores lecciones de vida, solo así pude comprender que es importante prepararse para servir a la sociedad. Estoy eternamente agradecido a mi madre Luzmila Coyago Ortuño, por el sacrificio que hizo por mí, para que pueda cumplir con mi meta y tener un futuro mejor.

Pablo Alexander Caisa Coyago

Quiero agradecer a la Universidad Politécnica Salesiana por formarme profesionalmente y como ser humano para poder aportar mis conocimientos con la sociedad.

Quiero agradecer a todas las personas que siempre me han apoyado y han creído en mí, especial a mis padres María Espinosa y Raúl Rosero que han hecho muchos sacrificios para que pueda finalizar mi carrera.

Erick Oswaldo Rosero Espinosa

ESTUDIO DESCRIPTIVO DE LA VIOLENCIA DE GÉNERO EN ECUADOR USANDO TECNICAS DE MINERIA DE DATOS

DESCRIPTIVE STUDY OF GENDER VIOLENCE IN ECUADOR USING DATA MINING TECHNIQUES

Pablo Caisa¹, Erick Rosero², Julio Proaño³

RESUMEN

La violencia de género en Ecuador ha pasado de ser considerada un asunto privado a ser reconocida como un problema social. Para este estudio descriptivo se utilizó el repositorio de violencia de género del INEC, donde se obtuvieron las bases de datos necesarias para el análisis mediante técnicas de minería de datos. El repositorio contenía información de mujeres en diferentes estados civiles: casadas, solteras y viudas, por mencionar algunos. Las bases de datos fueron divididas en cinco subconjuntos que representaban los ámbitos en los que se manifestaba la violencia. Uno de los resultados tras aplicar el algoritmo del árbol de decisión fue en el conjunto de datos violencia psicológica en el entorno educativo donde se obtuvo una puntuación de 1, lo que indica que el modelo realizó su predicción correctamente. Estos resultados son de gran importancia y contribuyeron a ampliar el conocimiento sobre la violencia de género en Ecuador y estas bases sirvan de apoyo para futuras investigaciones en esta área.

Palabras clave: violencia de género, machine learning, algoritmo, aprendizaje supervisado, minería de datos.

ABSTRACT

Gender violence in Ecuador has gone from being considered a private matter to being recognized as a social problem. For this descriptive study, the gender violence repository of INEC was used, where the databases necessary for the analysis were obtained using data mining techniques. The repository contained information on women in different marital statuses: married, single and widowed, to mention a few. The databases were divided into five subsets representing the areas in which violence was manifested. One of the results after applying the decision tree algorithm was in the dataset psychological violence in the educational environment where a score of 1 was obtained, indicating that the model made its prediction correctly. These results are of great importance and contributed to expand the knowledge on gender violence in Ecuador and these bases serve as support for future research in this area.

Keywords: gender violence, machine learning, algorithm, unsupervised learning, data mining

¹ Estudiante de Ingeniería en Sistemas – Universidad Politécnica Salesiana– Sede Quito- Correo institucional: pcaisa@est.ups.edu.ec

² Estudiante de Ingeniería en Sistemas – Universidad Politécnica Salesiana– Sede Quito-Correo institucional: eroseroe@est.usp.edu.ec

³ PhD. Julio Proaño Orellana Docente Titular Principal Nivel 1-Universidad Politécnica Salesiana- Sede Quito – Correo institucional: jproaño@ups.edu.ec

1. INTRODUCCIÓN

En Ecuador, un gran número de mujeres sufren violencia de género, lo que tiene graves efectos tanto a nivel individual como social. Resultados del Instituto Nacional de Estadística y Censos muestran que el 64,9% de las mujeres ecuatorianas han sido víctimas de malos tratos alguna vez en su vida [1].

Los hechos más frecuentes ocurrieron en los territorios de Guayas, Pichincha y Cotopaxi, 57 de ellos fueron cometidos por la pareja o expareja, y cuatro de ellos involucraron a mujeres embarazadas, dejando un total de 73 niños huérfanos. En el país hubo 720 casos de femicidio entre 2014 y noviembre de 2019 [2].

Ante esta preocupante situación, decidimos aplicar el análisis descriptivo, para obtener una visión completa y detallada de los tres principales dataset que son: casadas-unidas, solteras y separadas-viudas. En nuestro estudio, nos centramos en los siguientes ámbitos: El ámbito educativo engloba la violencia que se manifiesta en entornos de aprendizaje tales como escuelas, colegios y universidades; social, comprende la violencia que se presenta en las relaciones y dinámicas sociales; laboral, se enfoca en la violencia que puede ocurrir dentro del entorno de trabajo, como el acoso laboral, la discriminación; familiar, este está vinculado a la violencia que ocurre dentro del ámbito familiar que involucrar a miembros de la familia, como parejas, padres, hijos u otros parientes cercanos., por último, obstétrico, que se centra en la violencia experimentada por las mujeres durante el embarazo [3]. Además, por cada ámbito existen diferentes manifestaciones de violencia las cuales podemos apreciar en la siguiente Tabla 1.

Tabla 1. Tipos de violencia con sus definiciones.

Tipo de Violencia	Definición
-------------------	------------

Violencia psicológica	Influencia perjudicial en la salud mental causada por el malestar emocional, la manipulación o la intimidación.
Violencia física	Uso de fuerza o agresiones que provocan daño físico o sufrimiento a otra persona.
Violencia sexual	Agresiones sexuales sin consentimiento, como violaciones, abuso o acoso sexuales.
Violencia cibernética	Uso de tecnologías para acosar, amenazar, difamar o humillar, causando daño emocional o psicológico.
Acoso u hostigamiento	Comportamientos persistentes y ofensivos, creando un ambiente intimidante o degradante.
Violencia digital	Utilización de medios digitales para acosar, difamar o dañar la reputación de alguien.
Violencia patrimonial	Control, retención o destrucción intencional de bienes o recursos económicos de una persona.
Violencia obstétrica	Maltrato físico o psicológico hacia mujeres durante embarazo, parto o postparto, violando sus derechos.

En la Tabla 1, nos ofreció información detallada sobre las distintas fórmulas de violencia y lo que significan [3], lo que resulto útil para la clasificación de las preguntas de cada dataset.

A diferencia de investigaciones anteriores realizadas por INEC, donde se eliminaban los datos faltantes [4], solucionamos esta deficiencia utilizando la técnica de minería de datos conocida como imputación de datos [5]. La cual ayudó a rellenar los valores ausentes con la variable Nr. Al asignar el valor "Nr" a los datos parciales, podemos identificarlos y separarlos fácilmente en conjunto a las variables existentes: "Sí" y "No". Esta acción es fundamental para garantizar la integridad y precisión de nuestros análisis y resultados.

Se obtuvo resultados sustanciales dentro de la investigación, incluida la frecuencia relativa de las respuestas para nuestras tres variables, así como medidas estadísticas como la media, la mediana, el valor más bajo, el valor máximo y la desviación estándar [6].

Estos resultados son sumamente importantes porque nos permitió comprender plenamente la composición y la distribución de los datos de nuestro estudio.

En este estudio, utilizamos el enfoque del aprendizaje supervisado, específicamente mediante el uso de árboles de decisión, como una técnica de machine learning [7]. El problema de la clasificación en el dataset que abarca el ámbito educativo de la violencia psicológica. A través de este enfoque, pudimos obtener métricas como el score, matriz de confusión y valor de precisión, lo cual nos brindó un mayor entendimiento de la investigación en cuestión. Es importante resaltar que estas métricas nos ayudaron a evaluar y medir la efectividad del modelo clasificador [5].

El objetivo del presente artículo es llevar a cabo un estudio descriptivo de la violencia de género en Ecuador mediante el uso de técnicas de minería de datos para lo cual utilizaremos algoritmo de árbol de decisión para la clasificación. Se busca poner a disposición de estudiantes, investigadores y docentes información estadística descriptiva sobre este fenómeno, así como ofrecer una hoja de ruta para la explotación de la base de datos del INEC en futuras investigaciones.

2. METODOLOGÍA

El propósito de este estudio es presentar las técnicas de minería de datos y el modelo descriptivo ya que nos proporciona información sobre las relaciones entre los datos, sus características además del proceso de obtención, limpieza y análisis de los datos relacionados con la violencia de género en Ecuador. el análisis se enfoca en determinar la prevalencia de los diferentes tipos de violencia en los distintos ámbitos, así como en identificar las posibles relaciones entre las variables analizadas [8].

Preguntas de investigación

La definición de las preguntas de investigación son una guía base para la recopilación de artículos a estudiar, se definen preguntas de revisión de la literatura (RQ).

- **RQ1:** *¿De acuerdo con el objetivo principal de análisis descriptivo, qué tipo de violencia es más frecuente en cada categoría de estado civil y qué factores pueden explicar esta diferencia?*
- **RQ2:** *¿En qué base de datos se encontraron mayores niveles de violencia?*
- **RQ3:** *¿Cuál fue el nivel score arrojado por el algoritmo de árbol de decisión y en que ámbito?*

2.1 Obtención de los datos

Para la obtención del dataset se realizó una búsqueda exhaustiva de fuentes de datos oficiales, priorizando aquellas que se consideraron más confiables y completas, como la encuesta de violencia contra la mujer alojada en página web del INEC [1], se obtuvo un comprimido con el nombre de bdd_datos_abiertos_envigmu_2019_csv.zip, al descomprimir se obtuvo los siguientes archivos que se muestran en la Figura 1, se eligió las siguientes bases de datos: casadas-unidas, separadas-viudas y solteras en formato csv. Estos archivos son importantes para la investigación porque contienen información sobre la prevalencia de la violencia de género en distintas situaciones maritales y de vida de las mujeres encuestadas.

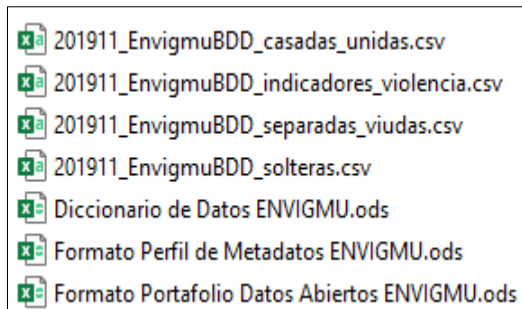


Figura 1. Contenido de la base de datos después de la extracción.

Posteriormente, se procedió abrir los 3 dataset seleccionados donde se observó que los encabezados no contenían las preguntas, al contrario, se encontraba un código referencial, cual se tenía que remplazar utilizando el archivo diccionario de datos, en la Tabla 2, se puede observar su composición.

Tabla 2. Diccionario de datos clasificado por dataset, código y preguntas

Dataset	Código	Preguntas
Casadas-unidas	f2_s2a_2	ámbito educativo: a lo largo de su vida ha asistido a un centro educativo
	f2_s2a_3_b	la ignoraron o la hicieron sentir inferior por ser mujer?
	f2_s2a_3_d	le enviaron mensajes, publicaron comentarios a través del celular?
Separadas-viudas	f3_s2a_3_a	la insultaron, gritaron, ofendieron o humillaron por ser mujer?
	f3_s2a_3_c	¿le hicieron comentarios acerca de que las mujeres no deberían estudiar?

Solteras	f4_s2a_2	¿a lo largo de su vida ha asistido a un centro educativo
	f4_s2a_3_c	¿le hicieron comentarios acerca de que las mujeres no deberían estudiar?

2.1.1 Identificación de las preguntas y sus códigos:

En esta etapa, podemos visualizar la relación que existe entre columna código y columna pregunta, la cual se puede observar en la Tabla 2, como primer paso se procedió a verificar si el código del encabezado de los dataset solteras, casadas-unidas y separadas-viudas correspondía correctamente al diccionario de datos, donde se realizó la reestructuración debido a que cada pregunta tiene un código único.

2.1.2 Reestructuración de los dataset:

Se realizó la reestructuración de los 3 dataset antes mencionados, para lo cual procederemos a seleccionar, la columna del diccionario de datos con las preguntas e insertamos en el dataset que corresponde, pero lo haremos de forma horizontal, para que cada pregunta coincida con su respectivo código, como se puede apreciar en la Tabla 3.

Tabla 3. Dataset reestructurados.

	¿Le enviaron mensajes, publicaron comentarios a través del celular?	¿La insultaron, gritaron, ofendieron o humillaron por ser mujer?	¿A lo largo de su vida ha asistido a un centro educativo?
Si			No
No		Si	Si
Si			Si

2.2 Carga e imputación de datos.

Primero accedemos a Google Drive, donde creamos una carpeta para cada conjunto de datos y le asignamos un nombre, luego cargamos la base de datos adecuada a cada carpeta y nos conectamos con Google Colab para acceder a los datos, lo que facilitó el trabajo ya que Google Colab permite escribir y ejecutar código en Python y tener acceso a un IDE similar a Jupyter Notebook, una plataforma que también proporciona herramientas para el aprendizaje automático y la minería de datos [9].

En este punto, se utilizó la función `mount()` para montar Google Drive en el entorno virtual de Google Colab, de modo que podemos acceder archivos y carpetas de nuestra investigación. Para usarla, debes ejecutar el siguiente código:

```
from google.colab import drive
drive.mount("/content/drive")
```

Una vez iniciada sesión debemos copiar la ruta del archivo cómo se muestra a continuación:

```
archivo='/content/drive/mydrive/mineria/solteras_conjuntos.xlsx'
df = pd.read_excel(archivo)
df.head()
```

El código utiliza la función `read_excel()` de pandas para cargar la base de datos desde Google drive en un dataframe llamado `df`. luego, se utilizó la función `head()` para mostrar las primeras 5 filas del dataframe. [10] el resultado se muestra en la Tabla 4. En la siguiente tabla muestra un ejemplo, como se encontraba los dataset y sus datos.

Tabla 4. Datos faltantes dentro del dataset.

¿Le enviaron mensajes, publicaron	¿La insultaron, gritaron, ofendieron o	¿A lo largo de su vida ha asistido a un

comentarios a través del celular?	humillaron por ser mujer?	centro educativo?
NaN		No
NaN	Si	Si
Si		Si

Dado que las bases de datos se encontraron valores faltantes como se puede visualizar en la Tabla 4, para su posterior análisis, en lugar de eliminar los registros incompletos, se decidió reemplazar los valores faltantes por la variable "Nr" para llevar a cabo esta tarea, se utilizaron diferentes librerías de Python, como pandas y numpy. en primer lugar, se importaron las librerías necesarias y se cargó el conjunto de datos, en un objeto dataframe de pandas. A continuación, se identificaron los valores faltantes y se reemplazaron por la variable Nr [11], mediante la función `fillna()` [12] continuación, podemos apreciar mejor el código utilizado para en el remplazo de los datos.

```
vg_data = df.fillna(0)
dfr= vg_data.replace([0],"nr")
dfr
```

Una vez que se reemplazó los valores faltantes, identificamos que en los 3 dataset comparten en común cinco tipos de ámbitos que son: educativo, social, laboral, familiar y obstétrica. Para la clasificación de tipo de violencia se toman como punto de apoyo la Tabla 4, esto permitió un análisis más detallado y específico de la violencia de género en Ecuador.

Tabla 5. Tipo de violencia con sus preguntas principales

Tipo de violencia	preguntas principales
Violencia psicológica	la ignoraron o la hicieron sentir inferior por ser mujer?
	la agredieron con empujones o le lanzaron algún objeto? ¿la agredieron con algún objeto o arma?
Violencia física	

Violencia sexual	¿la obligaron a desvestirse o a mostrar sus partes íntimas? ¿la obligaron a mirar escenas o actos sexuales o pornográficos?
Violencia cibernética	¿compartieron videos o fotos de naturaleza sexual sin su consentimiento?
Acoso u hostigamiento	¿la han vigilado, controlado, hostigado en su trabajo? ¿fue obligada a aceptar encuentros románticos fuera del horario de trabajo?
Violencia digital	¿Le enviaron mensajes con insultos u ofensas a través del celular?
Patrimonial	¿Le quitaron, destruyeron o se apropiaron de sus bienes o propiedades? ¿le quitaron dinero u obligaron a entregar dinero?
Obstétrica	¿Ha recibido alguna atención ginecológica? ¿la han examinado en presencia de otras personas sin pedirle su consentimiento?

En la Tabla 5, se observa la relación entre las preguntas y los tipos de violencia, además en cada ámbito pueden ocurrir diversos tipos de violencia, para lo cual se creó 5 dataframe por cada ámbito. La mejor opción para la clasificación fue leer cada pregunta, donde se identificó una pregunta principal y sus preguntas correlacionadas, una vez identificada la pregunta principal se pudo establecer a que ámbito pertenece. Cómo podemos apreciar en la Tabla 6.

Tabla 6. Clasificación de ámbitos y tipo de violencia

Ámbitos	Tipo de violencia
Educativo	psicológicas
	violencia cibernética
	violencia sexual
	acoso u hostigamiento
Social	psicológicas
	violencia cibernética
	física
	acoso u hostigamiento
	violencia digital
	violencia sexual

Laboral	psicológicas
	violencia cibernética
	acoso u hostigamiento
	violencia digital
Familiar	violencia sexual
	física
	psicológicas
	violencia cibernética
	patrimonial
Obstétrica	física
	acoso u hostigamiento
	violencia digital
	violencia sexual
	obstétrica

Se realizó el conteo de la variable Si, No y Nr, utilizando la siguiente sección de código:

```
colname=""
for (colname, colval) in dfr.iteritems():
    print(dfr[colname].value_counts())
```

La función recorre cada una de las preguntas del dataframe dfr por columna, al finalizar nos mostró la cantidad de valores únicos que se encuentran en cada columna utilizando el método 'value_counts ()'.

Después de la clasificación, dado que existen preguntas de diferentes índoles y en total hay 17211 filas entre los 3 datasets, con un total de 3000 preguntas, seleccionamos las preguntas que tengan como respuesta las variables Si, No y Nr. Con esta selección, crearemos un dataframe imputado para realizar un conteo y aplicar la fórmula de la frecuencia relativa, la cual se representa de la siguiente manera:

$$\text{Frecuencia relativa} = \frac{\text{Frecuencia absoluta}}{\text{Tamaño de la muestra}}$$
La frecuencia relativa es una medida estadística que nos proporciona el porcentaje de veces que aparece nuestras variables. Nos sirvió para comparar la frecuencia de diferentes categorías o valores. La frecuencia relativa siempre es un número entre 0 y 1, y se puede expresar como un porcentaje multiplicando por 100.

2.3. Normalización de datos:

Previamente, identificamos el conjunto de datos que requerían normalización y elegimos la escala deseada para nuestras variables Si, No y Nr. Asignamos a estas variables valores normalizados entre 0 y 1, utilizando dos métodos: el método de las medias de las series, donde restamos la media a cada variable y la dividimos por su desviación típica mediante la fórmula $x = (x - \mu) / \sigma$; y el método del máximo y mínimo, donde calculamos el máximo y mínimo de la serie mediante la fórmula $x' = (x - \min) / (\max - \min)$ [13].

2.4 Estandarización de datos

Para almacenar los resultados de las preguntas ya clasificadas, se crea un dataframe con el nombre de frecuencia relativa, donde se guardará el conteo de los Si, No y Nr se lo coloco en forma de porcentaje para que sea más sencillo comprender como podemos apreciar en la Tabla 7.

Tabla 7. Representación de Frecuencia relativa.

Ámbitos	Tipo de violencia	Si	No	Nr
Educativo	Psicológicas	22.64	73.19	4.17
	Violencia Cibernética	0.27	74.99	24.75
	Violencia Sexual	7.71	89.44	2.85
	Acoso u Hostigamiento	3.28	91.46	5.26

Aplicaremos la estandarización o también llamada z-score ya que es una herramienta valiosa para analizar y comparar datos sobre la violencia de género en sus diversos ámbitos, lo que proporciona una perspectiva más objetiva y ayuda a identificar

áreas de enfoque prioritario en la investigación [14].

2.5 Árbol de decisión

Se utilizó una técnica de clasificación basada en árboles de decisión. Este algoritmo permitió clasificar los datos específicamente en el dataset con el ámbito y tipo de violencia más significativo. Se dividió en subconjuntos, uno para el entrenamiento y el segundo de prueba. Además, obtuvimos la matriz de confusión que permitió comprobar la precisión del modelo y calcular medidas como el score [15]. Estas métricas proporcionaron información crítica sobre el éxito del algoritmo a la hora de categorizar. Este proceso lo dividimos en los siguientes 5 pasos:

1. **Importación de librerías:** En esta etapa, usamos diversas librerías tales como Pandas, LabelEncoder y scikit-learn. Pandas nos ha proporcionado una solución eficiente para la carga y manipulación de los datos, a su vez LabelEncoder nos ayuda con la correcta codificación y scikit-learn nos ha ofrecido las herramientas necesarias para la construcción y evaluación del modelo de árbol de decisiones.
2. **Carga de datos:** Empleamos la librería Pandas para cargar los datos en un dataframe. Gracias a esta librería, hemos podido corroborar los datos de manera más adecuada, garantizando así su correcta lectura y manipulación.
3. **Codificación de etiquetas:** Se utiliza la biblioteca Pandas y la clase LabelEncoder de scikit-learn para aplicar la codificación de etiquetas a columnas específicas de un dataframe
4. **Selección de la variable objetivo:** Durante esta fase, Identificamos la variable. Esta variable es de suma importancia, ya que representa

características importantes que buscamos predecir o clasificar a través de nuestro modelo.

5. **División en conjuntos de entrenamiento y prueba:** Hemos utilizado la función `train_test_split` de `scikit-learn` para dividir los datos en conjuntos de entrenamiento y prueba. Esta técnica nos ha permitido disponer de datos independientes para poner en marcha nuestro modelo y posteriormente, evaluar su rendimiento de manera objetiva.
6. **Creación del modelo de árbol de decisión:** Hemos desarrollado nuestro modelo de árbol de decisión utilizando la clase `DecisionTreeClassifier` de `scikit-learn`. Esta librería nos proporcionó una implementación confiable del algoritmo de árbol de decisión, especialmente diseñada para problemas de clasificación [16].

2.6 Visualización de datos

La presentación de los gráficos dentro de la investigación es muy importante, ya que permito visualizar de manera más clara y precisa los resultados obtenidos, para esto debemos realizamos los siguientes pasos.

1. Se importaron las librerías `pandas` y `matplotlib.pyplot` para manipular y analizar datos y crear visualizaciones gráficas respectivamente.
2. Se creó un conjunto de cinco cajas de bigotes (`boxplots`) utilizando `matplotlib.pyplot` para cada ámbito se generó una tabla la cual se resumió los resultados de cada caja de bigotes.
3. Se creó un gráfico de pastel utilizando `matplotlib.pyplot` para mostrar el porcentaje de violencia más alta por tipo de base de datos las librerías y funciones necesarias a continuación podemos visualizar de mejor manera las funciones y librerías necesarias para obtener los gráficos tipo barra pastel y

caja de bigotes [17]. En la Tabla 8, se puede observar las funciones utilizadas para graficar los resultados.

Tabla 8. Librerías y funciones para realizar las gráficas dentro de `colab`.

Librería	Función	Uso
pandas	<code>read_csv()</code>	Leer un archivo CSV y crear un dataframe.
	<code>describe()</code>	Resumir estadísticamente un dataframe.
	<code>groupby()</code>	Agrupar un dataframe por una o más columnas.
matplotlib.pyplot	<code>boxplot()</code>	Crear un gráfico de cajas de bigotes para visualizar la distribución de los datos.
	<code>pie()</code>	Crear un gráfico de pastel para mostrar la proporción de diferentes categorías en relación con el todo.
	<code>plt.bar</code>	Crear gráficos de barras. Los gráficos de barras son una forma eficaz de mostrar comparaciones entre diferentes categorías de datos.

3. RESULTADOS

A continuación, se va a presentar los resultados obtenidos en el trabajo de investigación. Se resumieron los hallazgos más importantes y se analizarán los datos obtenidos.

3.1. Tabla de frecuencia relativa y gráfico del dataset, de casadas-unidas.

En la Tabla 9, muestran los resultados de la frecuencia relativa en el dataset de casadas-unidas, en los diferentes ámbitos y su tipo de violencia. Cada fila representa un ámbito y su tipo de violencia específico, y las columnas Si, No y Nr indican los porcentajes de cada

caso en los que se ha identificado un tipo de violencia.

Violencia sexual 3.84 65.04 31.11

Obstétrica Obstétrica 31.94 67.11 0.95

Tabla 9. Tipos de violencia por cada ámbito con sus resultados del, Si, No y Nr en dataset de casas-unidas.

Ámbitos	Tipo de violencia	Si	No	Nr
Educativo	Psicológicas	22.64	73.19	4.17
	Violencia cibernética	0.27	74.99	24.75
	Violencia sexual	7.71	89.44	2.85
	Física	3.28	91.46	5.26
	Acoso u hostigamiento	2.63	92.08	5.3
Social	Psicológicas	5.53	94.17	0.29
	Violencia cibernética	1.52	82.48	16
	Física	2.57	97.42	0.01
	Acoso u hostigamiento	8.14	91.4	0.46
	Violencia digital	0.17	87.38	12.45
	Violencia sexual	8.22	53.42	38.36
Familiar	Psicológicas	11.82	88.14	0.03
	Violencia cibernética	0.48	83.61	15.91
	Acoso u hostigamiento	6.6	93.39	0.02
	Violencia digital	0.06	87.41	12.53
	Violencia sexual	4.78	95.09	0.13
	Patrimonial	0.69	97.24	2.07
	Intrafamiliar	13.34	82.6	4.06
Laboral	Física	15.82	83.85	0.33
	Psicológicas	12.33	63.91	23.76
	Violencia cibernética	0.41	72.29	27.3
	Física	0.5	85.76	13.74
	Acoso u hostigamiento	2.59	83.58	13.84
	Violencia digital	0.08	76.09	23.83

Basado en la Tabla 9, reveló que la violencia psicológica es la forma más prevalente en el ámbito educativo, representando el 22.64% de los casos analizados. Esta manifestación de violencia puede estar vinculada a la presión académica excesiva, el acoso entre compañeros, la falta de apoyo emocional y social, así como la ausencia de programas de prevención y detección temprana en las instituciones educativas. En el ámbito social, el acoso u hostigamiento se destacó, alcanzando un 8.14% de los casos analizados, siendo factores como la discriminación, la desigualdad de género y la falta de mecanismos efectivos para abordarlo y prevenirlo en la sociedad posibles detonantes. En el ámbito familiar, la violencia física fue la más alta, representando el 15.82% de los casos analizados, y factores como patrones de violencia heredados, estrés familiar y desequilibrio de poder podrían estar contribuyendo a esta situación. En el ámbito laboral, la violencia psicológica se identificó como la más prevalente, representando el 12.33% de los casos analizados, y se encontraron posibles factores como una cultura organizacional tóxica, el estrés laboral y la falta de políticas claras para prevenir y abordar la violencia en el lugar de trabajo. En el ámbito obstétrico, la violencia obstétrica sobresalió como la forma más prevalente, representando el 31.94% de los casos analizados, y se vinculó a la falta de respeto a los derechos de las mujeres, prácticas médicas invasivas o no consentidas, desigualdad de género en la atención médica y falta de capacitación del personal de salud.

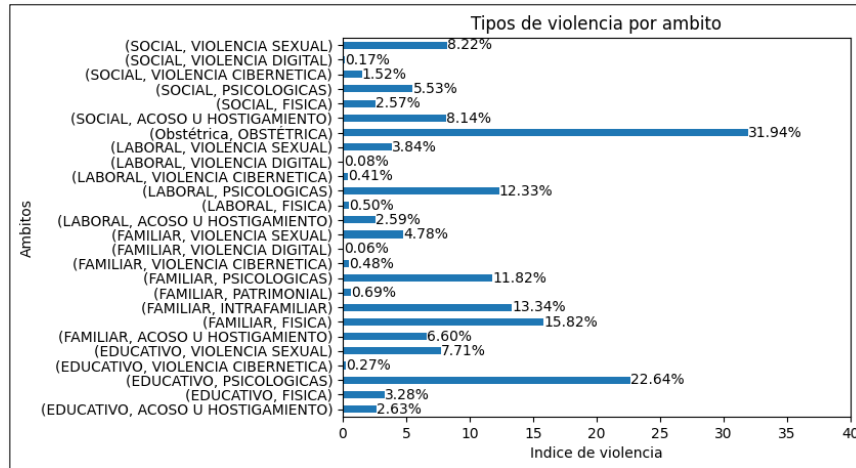


Figura 2. Gráfico de la frecuencia relativa de las mujeres casadas-unidas.

De acuerdo con el gráfico de la Figura 2, el eje “x” nos hace referencia a cada ámbito con su tipo de violencia, el eje “y” nos muestra los valores de tipo de violencia, estos índices ayudaron a visualizar de manera más objetiva cómo está distribuida los valores dentro del dataset casadas-unidas. Además, de eso se realizó un análisis estadístico utilizando las principales fórmulas como son la media como la varianza como la desviación estándar y el error para saber que tan dispersos están los datos como podemos apreciar en la Tabla 10.

Tabla 10. Estadísticas de respuestas: Si, No, Nr para el data set casadas-unidas

	Media	Varianza	Desviación Estándar	Error
si	0.064596	0.006030	0.077654	0.015229
no	0.827904	0.013066	0.114307	0.022418
nr	0.107501	0.012919	0.113662	0.022291

Los resultados presentados en la Tabla 10, muestran tres variables: 'Si', 'No' y 'Nr'. En resumen, al analizar los datos por fila, se observó lo siguiente: las respuestas positivas muestran una baja incidencia reportada de violencia de género, con datos consistentes y

poca variabilidad. Por otro lado, las respuestas negativas reflejan una alta incidencia reportada, con mayor variabilidad y dispersión en los datos. Las respuestas no registradas indican una proporción moderada de datos faltantes, también con cierta variabilidad en las respuestas. En conclusión, los resultados revelan diferentes percepciones y experiencias de violencia de género entre los participantes, con una variabilidad mayor en las respuestas negativas y no registradas en comparación con las respuestas positivas.

3.2. Tabla de frecuencia relativa y gráfico del dataset de separadas-viudas.

En la Tabla 11, muestra los resultados de la frecuencia relativa en el Dataset de separadas-viudas, en los diferentes ámbitos y su tipo de violencia. Cada fila representa un ámbito y su tipo de violencia específico, y las columnas Si, No y Nr indican los porcentajes de cada caso en los que se ha identificado un tipo violencia

Tabla 11. Tipos de violencia por cada ámbito con sus resultados del, Si, No y Nr en dataset de separadas-viudas.

Ámbitos	Tipo de violencia	Si	No	Nr					
					Patrimonial	01.02	96.77	2.21	
Educativo	Psicológicas	21.85	69.64	8.51	Física	10.01	89.93	0.06	
	Violencia cibernética	0.54	64.02	35.45	Violencia sexual	13.17	86.23	0.6	
	Violencia física	3.75	85.29	10.96	Violencia digital	0.14	83.48	16.38	
	Acoso u hostigamiento	2.76	86.31	10.93	Obstétrica	Obstétrica	22.68	76.37	0.95
	Violencia digital	0.06	71.26	28.68					
	Violencia sexual	1.72	94.46	3.82					
Social	Psicológicas	6.7	93.06	0.24					
	Violencia cibernética	2.23	77.45	20.31					
	Física	3.15	96.82	0.03					
	Acoso u hostigamiento	7.38	92.19	0.42					
	Violencia digital	0.11	84.16	15.73					
	Violencia sexual	13.42	85.98	0.6					
Laboral	Psicológicas	35.55	60.28	4.16					
	Violencia cibernética	0.65	73.72	25.63					
	Acoso u hostigamiento	3.92	89.05	07.04					
	Física	0.75	92.32	6.93					
	Violencia digital	0.08	79.66	20.25					
	Violencia sexual	12.61	86.38	01.01					
Familiar	Psicológicas	13.3	86.62	0.08					
	Violencia cibernética	0.68	79.21	20.11					

En la Tabla 11, se observó que la violencia psicológica es la forma más prevalente en el ámbito educativo, representando el 21.85% de los casos analizados. Esta forma de violencia puede estar relacionada con el acoso escolar, el maltrato verbal o emocional por parte de compañeros, profesores o personal educativo. En el ámbito social, la violencia sexual ocupa un lugar relevante, alcanzando un 13.42%. Esta forma de violencia incluye agresiones sexuales, abuso y acoso sexuales en diferentes entornos sociales. En el ámbito laboral, la violencia psicológica destaca con un 35.55%, manifestándose a través del acoso laboral, la discriminación o la violencia verbal. En el ámbito familiar, la violencia física se presenta en un 10.01%, relacionada con dinámicas de poder desequilibradas y conflictos familiares. En el ámbito obstétrico, la violencia obstétrica alcanza un 22.68%, evidenciando maltrato, abuso emocional o físico durante el embarazo, el parto o el posparto.

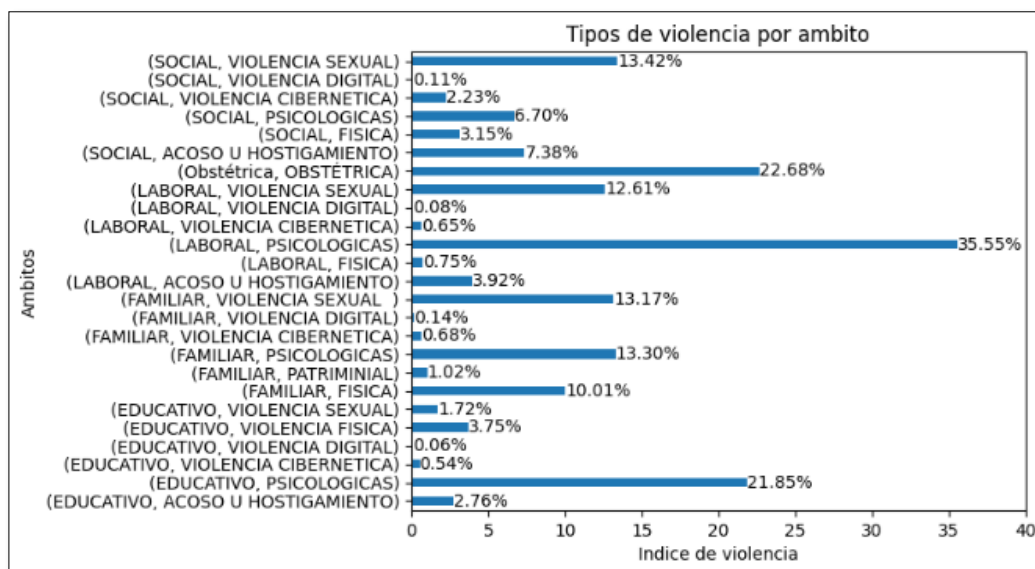


Figura 3. Gráfico de la frecuencia relativa las mujeres separadas-viudas.

De acuerdo con el gráfico de la Figura 3, el eje “x” nos hace referencia a cada ámbito con su tipo de violencia, el eje “y” nos muestra los valores de tipo de violencia, estos índices nos ayudaron a visualizar de manera más objetiva cómo está distribuida los valores dentro del dataset separadas-viudas.

Además, se realizó un análisis estadístico utilizado las principales fórmulas como son la media, varianza, desviación estándar y el error para saber que tan dispersos están los datos como podemos apreciar en la Tabla 12.

Tabla 12. Estadísticas de respuestas: Si, No, Nr para el data set separadas-viudas.

	Media	Varianza	Desviación Estándar	Error
si	0.071301	0.008049	0.089715	0.017943
no	0.832251	0.009618	0.098072	0.019614
nr	0.096447	0.010846	0.104143	0.020829

Analizamos los resultados presentados en la Tabla 12, donde se puede destacar lo siguiente: la media de las respuestas positivas indica una incidencia ligeramente mayor de violencia en comparación con los datos de la Tabla 10. Sin embargo, sigue siendo relativamente baja.

Las respuestas negativas muestran una alta incidencia reportada, con poca

variabilidad y una desviación estándar más baja que en los datos anteriores. Las respuestas no registradas presentan una proporción moderada de datos faltantes, con una desviación estándar similar a la de los datos anteriores. En resumen, los resultados sugieren que persiste una brecha significativa entre las experiencias de violencia de género reportadas y las no reportadas, con una alta incidencia general y mayor consistencia en las respuestas negativas.

3.3. Tabla de frecuencia relativa y gráfico del dataset de solteras.

En la Tabla 13, muestra los resultados de la frecuencia relativa en el dataset de solteras, en los diferentes ámbitos y su tipo de violencia. Cada fila representa un ámbito y su tipo de violencia específico, y las columnas Si, No y Nr indican los porcentajes de cada caso en los que se ha identificado un tipo violencia.

Tabla 13. Tipos de violencia por cada ámbito con sus resultados del Si, No y Nr en dataset de solteras.

Ámbitos	Tipo de violencia	Si	No	Nr
---------	-------------------	----	----	----

Educativo	Psicológicas	28.19	62.26	9.55
	Violencia cibernética	1.98	86.84	11.18
	Violencia sexual	0.81	98.13	01.06
	Acoso u hostigamiento	1.56	97.38	01.07
Social	Psicológicas	06.01	93.91	0.08
	Violencia cibernética	4.76	88.53	6.71
	Física	1.56	97.38	01.07
	Acoso u hostigamiento	26.05	73.95	0.0
	Violencia digital	0.68	92.51	6.8
	Violencia sexual	7.44	91.89	0.67
Familiar	Psicológicas	23.55	76.33	0.12
	Violencia cibernética	8.19	86.89	4.92
	Acoso u hostigamiento	8.9	91.01	0.09
	Violencia digital	16.7	83.3	0.0
	Violencia sexual	0.13	96.34	3.53
Laboral	Psicológicas	12.17	86.36	1.46
	Violencia cibernética	4.79	73.58	21.63
	Acoso u hostigamiento	0.51	94.0	5.49
	Física	01.09	91.36	7.54
	Violencia digital	0.13	93.13	6.74
	Violencia sexual	10.66	87.52	1.82
Obstétrica	Obstétrica	14.45	72.06	13.49

Basado en la Tabla 13, revelo que la violencia psicológica es la forma más

prevalente en el ámbito educativo, representando el 28.19% de los casos analizados. Esto puede estar relacionado con la presión académica excesiva, el acoso entre compañeros, la falta de apoyo emocional y social, así como la ausencia de programas de prevención y detección temprana en las instituciones educativas. En el ámbito social, el acoso u hostigamiento se destaca, alcanzando un 26.05% de los casos analizados. Factores como la discriminación, la desigualdad de género y la falta de mecanismos efectivos para abordarlo y prevenirlo en la sociedad podrían ser posibles detonantes. En el ámbito familiar, la violencia física es la más alta, representando el 23.55% de los casos analizados. Patrones de violencia heredados, estrés familiar y desequilibrio de poder podrían estar contribuyendo a esta situación. En el ámbito laboral, la violencia psicológica se identificó como la más prevalente, representando el 12.17% de los casos analizados. Posibles factores asociados incluyen una cultura organizacional tóxica, el estrés laboral y la falta de políticas claras para prevenir y abordar la violencia en el lugar de trabajo. En el ámbito obstétrico, la violencia obstétrica sobresale como la forma más prevalente, representando el 14.45% de los casos analizados. Se ha vinculado a la falta de respeto a los derechos de las mujeres, prácticas médicas invasivas o no consentidas, desigualdad de género en la atención médica y falta de capacitación del personal de salud.

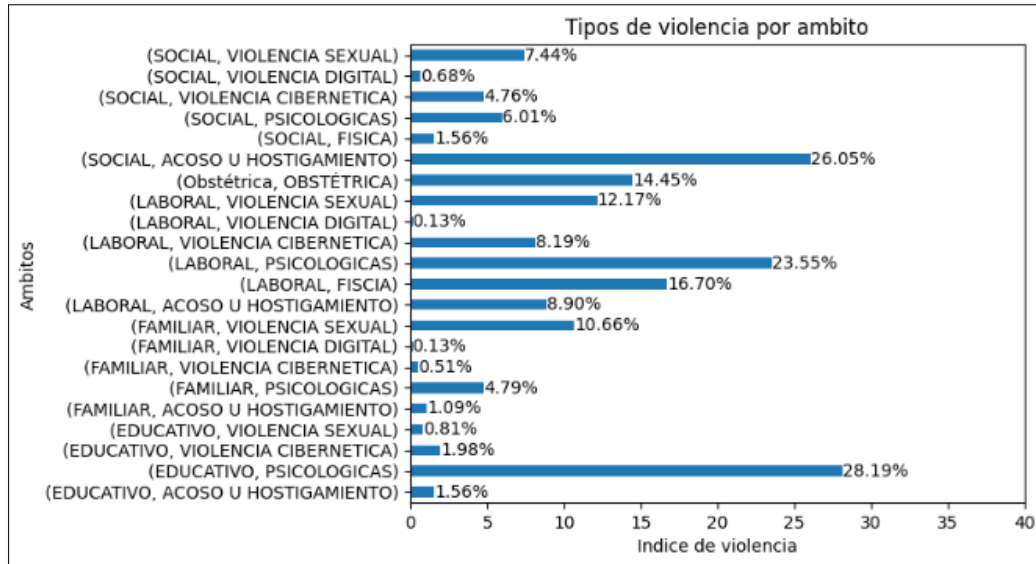


Figura 4. Gráfico de la frecuencia relativa para las mujeres solteras.

De acuerdo con el gráfico de la Figura 4, el eje “x” nos hace referencia a cada ámbito con su tipo de violencia, el eje “y” nos muestra los valores de tipo de violencia, estos índices ayudaron a visualizar de manera más objetiva cómo está distribuida los valores dentro del dataset solteras.

Además, se realizó un análisis estadístico utilizando las principales fórmulas como son la media como la varianza como la desviación estándar y el error para saber que tan dispersos están los datos como podemos apreciar en la Tabla 14.

Tabla 14 Estadísticas de respuestas: Si, no, Nr para el data set solteras.

	Media	Varianza	Desviación Estándar	Error
si	0.081967	0.007592	0.087135	0.018577
no	0.870304	0.009365	0.096772	0.020632
nr	0.047729	0.003021	0.054966	0.011719

Los resultados presentados en la Tabla 17, mostraron los siguientes aspectos relevantes: la media de las respuestas positivas indica un ligero aumento en la incidencia de violencia de género en comparación con los datos anteriores, lo que sugiere una mayor

presencia de casos reportados. Las respuestas negativas muestran una incidencia significativamente alta, con una media más alta y una desviación estándar ligeramente mayor que en los datos anteriores, lo que indica una menor variabilidad en las respuestas. Por otro lado, las respuestas no registradas tienen una proporción relativamente baja, con una media más baja y una desviación estándar menor que en los datos anteriores, lo que indica una menor presencia de datos faltantes. En resumen, estos resultados sugieren un incremento en la incidencia de violencia de género reportada, una alta prevalencia de respuestas negativas y una baja proporción de datos no registrados.

3.4. Caja de bigotes y tabla de resultados del dataset de casadas-unidas

En esta sección podemos observar el gráfico de la caja de bigotes, en la Figura 5, donde se

posteriormente analizo sus datos.

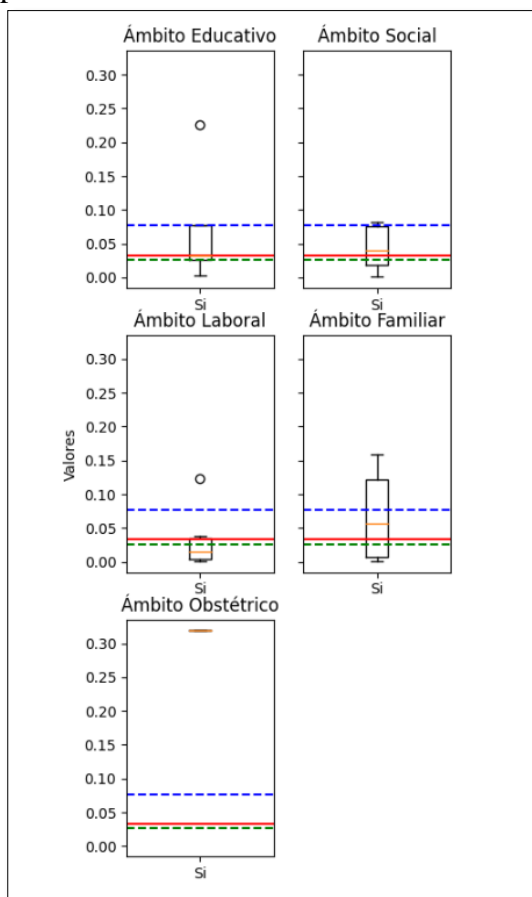


Figura 5. Gráfico caja de bigote por ámbitos dataset casadas-unidas.

En la Figura 5 representa el gráfico de la caja de bigotes, con el dataset de casadas-unidas, donde se tomaron los valores con respuesta “si”, mientras en la Tabla 15, muestra los resultados del gráfico para su posterior descripción.

Tabla 16. Resultado caja de bigotes de casadas-unidas.

	Ámbito Educativo	Ámbito Social	Ámbito Laboral	Ámbito Familiar	Ámbito Obstétrico
Mean	0.0730	0.0436	0.3194	0.0669	0.0329
Std	0.0898	0.0628	0.0344	0.0466	0
Min	0.0026	0.0017	0.0007	0.0005	0.3194
25%	0.0262	0.0178	0.0042	0.0063	0.3194
50%	0.0327	0.0405	0.0154	0.0568	0.3194
75%	0.0770	0.0749	0.0353	0.1220	0.3194
Max	0.2264	0.0822	0.1233	0.1581	0.3194

Al analizar los datos de la Tabla 16, en diferentes ámbitos, se evidencia el impacto de la violencia de género en cada contexto. En el ámbito educativo, se registró un nivel de violencia psicológica de 7.30%, mientras que en el ámbito social se observó un nivel de violencia de 4.36%. En el ámbito laboral, el nivel de violencia es más alto, alcanzando un 31.94%. En el ámbito familiar, se encontró un nivel de violencia de género del 6.69%, y en el ámbito obstétrico se registra un nivel constante, pero bajo de violencia de género, representando un 3.29%.

3.5. Caja de bigotes y tabla con sus respectivos resultados, con el dataset de casadas-unidas.

En sección podemos observar gráfico de la caja de bigotes, se puede apreciar en la Figura 6, con el dataset de casadas-unidas, mientras en la Tabla 16, muestra los resultados del gráfico para su posterior descripción.

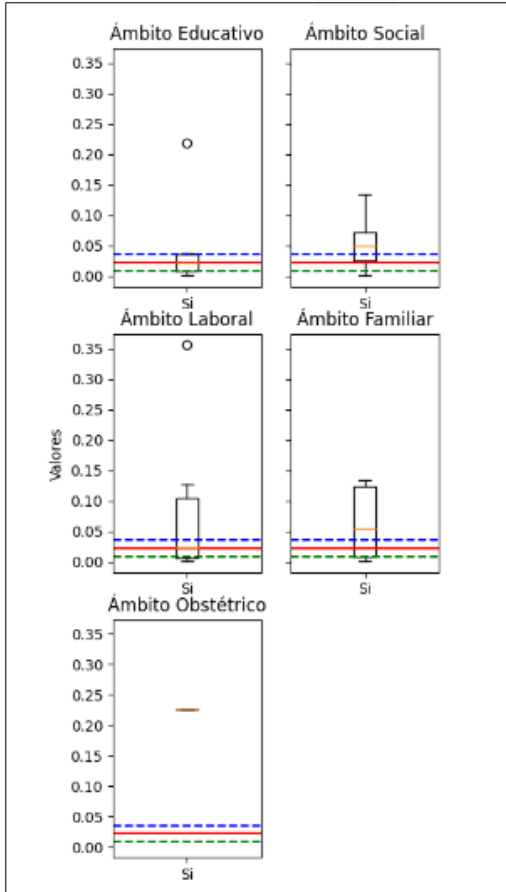


Figura 6. Gráfico caja de bigote por ámbito dataset separadas-viudas.

Tabla 17. Resultado caja de bigotes de separadas-viudas, por cada ámbito.

	Ámbito Educativo	Ámbito Social	Ámbito Laboral	Ámbito Familiar	Ámbito Obstétrico
Mean	0.0511	0.0550	0.0892	0.0638	0.2267
Std	0.0831	0.0475	0.1386	0.0643	0
Min	0.0005	0.0011	0.0008	0.0014	0.2267
25%	0.008	0.025	0.007	0.008	0.2267
50%	0.022	0.049	2.33%	0.023	0.2267
75%	0.035	0.072	0.104	0.124	0.2267
Max	0.219	0.134	0.356	0.133	0.2267

Al examinar la Tabla 17, en diferentes ámbitos, se reveló el impacto de la violencia de género en cada contexto. En el ámbito educativo, se registró un promedio de violencia del 5.11%, mientras que en el ámbito social se observó un promedio del

5.50%. En el ámbito laboral, se identificó un promedio más alto de violencia de género, alcanzando el 8.92%. En el ámbito familiar, se encontró un promedio del 6.38% de violencia, y en el ámbito obstétrico se registró un promedio constante pero alto de violencia de género, representando un 22.67%.

3.6. Caja de bigotes y tabla de resultados, del dataset de solteras, la descripción de sus resultados.

En la Figura 7, representa el gráfico de la caja de bigotes, en la Tabla 18, apreciamos los resultados para su posterior descripción.

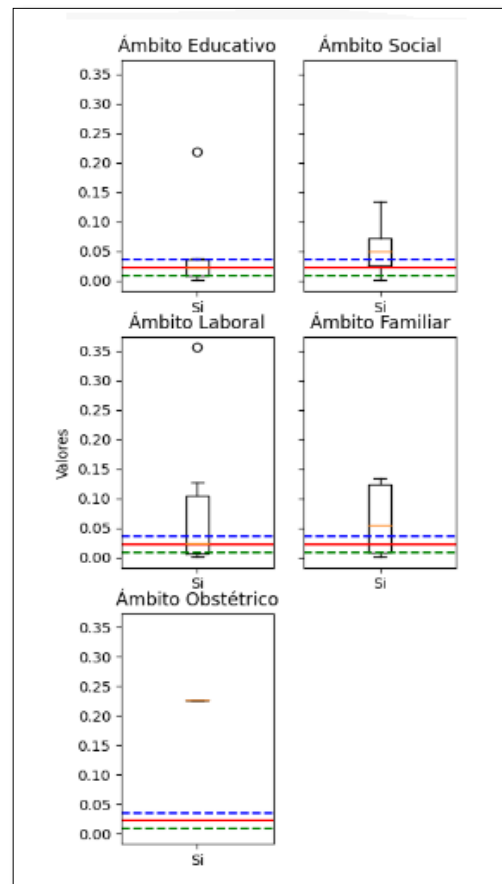


Figura 7. Gráfico caja de bigote por ámbito dataset de solteras.

Tabla 18. Resultado caja de bigotes de solteras.

	Ámbito Educativo	Ámbito Social	Ámbito Laboral	Ámbito Familiar	Ámbito Obstétrico
Mean	0.081	0.078	0.116	0.034	0.145
Std	0.134	0.093	0.080	0.044	0
Min	0.008	0.007	0.001	0.001	0.145
25%	0.0137	0.0236	0.837	0.0051	0.145
50%	0.0177	0.0539	0.1054	0.0109	0.145
75%	0.0853	0.0709	0.1557	0.0479	0.145
Max	0.282	26.05%	0.261	0.145	0.145

Al analizar los datos de la Tabla 18, en diferentes ámbitos, se observó el impacto de la violencia de género en cada contexto. En el ámbito educativo, se registró un promedio de violencia del 8.1%, mientras que en el ámbito social se identificó un promedio del 7.8%. En el ámbito laboral, se encontró un promedio más alto de violencia de género, con un 11.6%. En el ámbito familiar, se registró un promedio del 3.4% de violencia, y en el ámbito obstétrico se muestra un promedio del 14.5%.

3.6 Árbol de decisión.

Los datos que fueron utilizados se pueden observar con más detalle en la Tabla 19, la cual hace referencia al segmento de datos debido a que el dataset que utilizaremos consta de 17211 filas por 7 columnas. En este caso, se analizó el ámbito educativo y el tipo de violencia psicológica. Como se muestra en las Figuras 2, 3 y 4, se observa un alto nivel de violencia en este tipo de ámbito en los tres conjuntos de datos: casadas-unidas, separadas-viudas y solteras. En cuanto al ámbito obstétrico, dado que no se presentaron diferentes tipos de violencia no se consideró en el análisis.

Tabla 19. Composición del dataset para aplicar el árbol decisión.

Violencia ámbito educativo	Edad	Ignorada inferioridad	Comentarios no estudiar	Publicación sin consentimiento	Insinuaciones pedidas sexuales	Problemas escuela colegio universidad
----------------------------	------	-----------------------	-------------------------	--------------------------------	--------------------------------	---------------------------------------

No recuerda	16	Nr	Nr	Nr	Nr	Nr
No fue víctima	55	No	No	No	No	No
No recuerda	46	No	No	No	No	Nr
No recuerda	15	Nr	Nr	Nr	Nr	Nr
No recuerda	54	Nr	Nr	Nr	Nr	Nr
Si fue víctima	15	Nr	Nr	Nr	Nr	Nr

En la Tabla 19, podemos observar las preguntas principales además de nuestra variable objetivo, la cual se encuentra en la primera columna denominada violencia en el ámbito educativo

Una vez aplicado el algoritmo de árbol de decisión podemos apreciar de mejor manera, analizando los datos de la matriz de confusión que nos arroja los siguientes resultados descritos en la Tabla 20.

Tabla 20. Resultado de la matriz de confusión.

Clase	Clase 0	Clase 1	Clase 2
Predicción Clase 0	219	0	0
Predicción Clase 1	0	3859	0
Predicción Clase 2	0	0	225

Al analizar la Tabla 20, podemos extraer los puntos más relevantes de la clasificación realizada, se observó que la Clase 0 fue correctamente identificada en 219 casos, sin ninguna predicción incorrecta. Por otro lado, la Clase 1 muestra un alto nivel de precisión, con 3859 predicciones correctas y ninguna predicción incorrecta. Es importante destacar que para la Clase 2, se realizaron 225 predicciones correctas y ninguna predicción incorrecta.

En el reto de clasificación, el modelo de árbol de decisión, aplicado después de transformar todos los datos cualitativos utilizando el codificador de etiquetas LabelEncoder, obtuvo resultados destacados. Con una puntuación F1 de 1 y un

recall de 1, los resultados obtenidos muestran una precisión aceptable. Esto demuestra que el modelo es capaz de clasificar todas las muestras del conjunto de pruebas con una precisión del 100%. El modelo reconoce precisión cada muestra positiva.

4. DISCUSIÓN

Una vez analizamos la prevalencia de diferentes tipos de violencia en tres grupos demográficos: mujeres casadas-unidas, separadas-viudas y solteras, examinamos los resultados para identificar patrones y tendencias significativas. La Tabla 21, presenta un resumen de los porcentajes de cada tipo de violencia en los diferentes ámbitos estudiados, lo cual nos ayudará a responder a la RQ1. A continuación, discutiremos los hallazgos más relevantes obtenidos a partir de estos datos.

Tabla 21. Combinación de las tres tablas de frecuencia.

Base de datos	Ámbito	Tipo de violencia	Porcentaje
Casadas-unidas	Educativo	Psicológicas	22,64%
	Social	Violencia sexual	8,22%
	Familiar	Física	15,82%
	Laboral	Psicológicas	12,33%
	Obstétrico	Obstétrica	31,94%
Separadas-viudas	Educativo	Psicológicas	21,85%
	Social	Violencia sexual	13,42%
	Familiar	Psicológicas	13,3%
	Laboral	Psicológicas	35,55%
	Obstétrico	Obstétrica	22,68%
Solteras	Educativo	Psicológicas	28,19%
		Acoso u	
	Social	hostigamiento	26,05%
	Familiar	Psicológicas	23,55%
	Laboral	Psicológicas	12,17%
	Obstétrica	Obstétrica	14,45%

RQ1: *¿De acuerdo con el objetivo principal de análisis descriptivo, qué tipo de violencia es más frecuente en cada categoría de estado civil y qué factores pueden explicar esta diferencia?*

El objetivo principal del análisis descriptivo es describir y resumir los datos de manera significativa y útil, tal como se puede apreciar en la Tabla 21, que es una combinación de las tres tablas de frecuencia donde podemos apreciar lo siguiente

De acuerdo con los factores mencionados, analizamos la prevalencia de la violencia en cada ámbito y categoría de estado civil.

En el ámbito educativo, se observó que la violencia psicológica es más frecuente en las mujeres casadas-unidas y solteras, representando el 22,64% y el 28,19% respectivamente. Esto podría estar relacionado con la presión académica excesiva, el acoso entre compañeros y la falta de apoyo emocional y social en estos entornos.

En el ámbito social, se destaca la violencia sexual en las mujeres casadas-unidas y separadas-viudas, con porcentajes del 8,22% y 13,42% respectivamente. Este tipo de violencia puede estar influenciado por factores como la discriminación de género y las dinámicas sociales que perpetúan la desigualdad.

En el ámbito laboral, la violencia psicológica es más prevalente en todas las categorías de estado civil, con porcentajes que van desde el 12,17% hasta el 35,55%. Esto puede estar relacionado con una cultura organizacional tóxica, el estrés laboral y la falta de políticas claras para prevenir y abordar la violencia en el lugar de trabajo.

En el ámbito familiar, la violencia física es más alta en las mujeres casadas-unidas, representando el 15,82%. Los patrones de violencia heredados, el estrés familiar y el desequilibrio de poder pueden contribuir a esta situación.

En el ámbito obstétrico, se destaca la violencia obstétrica en las mujeres casadas-unidas y separadas-viudas, con porcentajes del 31,94% y 22,68% respectivamente. Esto puede estar relacionado con la falta de respeto a los derechos de las mujeres, prácticas médicas invasivas o no consentidas, desigualdad de género en la atención médica y falta de capacitación del personal de salud.

Una vez analizada la Tabla 21, podemos apreciar que la violencia psicológica se encuentra ampliamente presente en todos los conjuntos de datos. Para resaltar los porcentajes más altos, hemos creado la Tabla 22, que muestra estos valores destacados. A través de esta tabla, podemos obtener una visión más precisa de la prevalencia de la violencia psicológica en cada categoría.

Tabla 22. Niveles de violencia más representativo por base de datos.

Base de datos	Tipo de violencia Psicológica	Nivel de violencia más alto
Casadas-unidas	Psicológicas	22,64%
Separadas-viudas	Psicológicas	35,55%
Solteras	Psicológicas	28,19%

En la Tabla 22, examinamos el nivel de violencia psicológica de cada conjunto de datos. Dado el impacto negativo de esta forma de violencia en la salud mental, analizamos los resultados para determinar las diferencias en los niveles de violencia psicológica entre estos grupos. Los resultados mostraron que el conjunto de datos de mujeres que vivían solas mostraba el nivel más alto de violencia psicológica con un 35,55%, seguido del conjunto de datos de mujeres solteras con un 28,19%, mientras que el conjunto de datos de mujeres casadas y solteras mostraba un nivel relativamente bajo con un 22,64%. Estos resultados

subrayan la importancia de abordar el impacto negativo en la salud mental de la angustia emocional, la manipulación o el acoso asociados a la violencia psicológica en cada uno de los conjuntos de datos de mujeres.

RQ2: *¿En qué base de datos se encontraron mayores niveles de violencia?*

La Tabla 21, muestra datos significativos sobre la violencia psicológica entre las viudas separadas. Se constató que este grupo presentaba la tasa más elevada de violencia psicológica, con un 35,55%. El análisis también muestra que este tipo de violencia se produce principalmente en el entorno laboral.

Estos resultados arrojaron que las viudas separadas se enfrentan a mayores niveles de violencia psicológica en el lugar de trabajo en comparación con otros entornos como la educación, la vida social, la familia y la maternidad. Esto sugiere que el entorno laboral puede ser particularmente desafiante y perturbador para estas mujeres en términos de violencia psicológica.

Factores como la angustia emocional, la manipulación y el acoso, identificados en la Tabla 1, están estrechamente relacionados con la violencia psicológica en el entorno laboral de las mujeres que viven solas.

RQ3: *¿Cuál fue el nivel score arrojado por el algoritmo de árbol de decisión y en que ámbito?*

Tras analizar los resultados del algoritmo de árbol de decisión en un conjunto de datos con tipos de violencia psicológica en el ámbito educativo, obtuvimos resultados alentadores. El modelo alcanzó una precisión perfecta de 1,0, lo que significa que clasificó correctamente todos los casos de violencia psicológica del conjunto de datos. Además,

alcanzó una puntuación F1 de 1,0, lo que indica un equilibrio óptimo entre precisión y exhaustividad. El valor recall de 1,0 indica que el modelo identificó correctamente todos los casos positivos de violencia psicológica, sin dejar ningún caso sin detectar. Estos resultados indican que el modelo de árbol de decisión es muy preciso y eficaz para identificar la violencia psicológica en entornos educativos.

5. CONCLUSIONES

El objetivo principal de este artículo es realizar un estudio descriptivo de la violencia de género en Ecuador utilizando técnicas de minería de datos con un enfoque puramente exploratorio. En este sentido, se destacan los aspectos fundamentales relacionados con el uso de estas herramientas, así como su aplicación en el análisis de la violencia de género.

Uno de los resultados más importantes del estudio fue la identificación de la prevalencia de la violencia psicológica en todos los ámbitos analizados. Este resultado subraya la importancia de abordar esta cuestión, ya que su impacto negativo en la salud mental de las mujeres es evidente. También destaca el valor de la investigación descriptiva como forma de proporcionar información estadística y descriptiva relevante para la investigación en este ámbito.

En cuanto a los métodos utilizados, se constata la eficacia y precisión de los algoritmos de clasificación basados en árboles de decisión para identificar y clasificar la violencia de género en contextos educativos. Este enfoque permite identificar patrones y factores de riesgo que pueden facilitar la intervención temprana y la asistencia adecuada a las víctimas. Además, se recurrió a la imputación de datos para mantener la integridad del conjunto de datos y evitar que se borrarán registros,

proporcionando así información adicional para el análisis.

El uso de técnicas de minería de datos, en particular algoritmos de clasificación de árboles de decisión e imputación de datos, ha demostrado ser una combinación eficaz para analizar la violencia de género. Estos métodos no sólo identifican patrones y factores de riesgo, sino que también facilitan la conservación de conjuntos de datos completos y evitan la pérdida de información valiosa. Además, el objetivo es proporcionar a estudiantes, investigadores y profesores información estadística descriptiva para la investigación y el análisis.

6. REFERENCIAS

- [1] INEC, “Encuesta Nacional sobre Relaciones Familiares y Violencia de Género contra las Mujeres-ENVIGMU,” QUITO, Nov. 2019. Accessed: Jun. 03, 2023. [Online]. Available: <https://www.ecuadorencifras.gob.ec/violencia-de-genero/#:~:text=Resumen%202019%20Una%20visi%C3%B3n%20general%20de%20los%20resultados,metadatos%20que%20permitan%20una%20interpretaci%C3%B3n%20de%20los%20resultados>.
- [2] S. M. Moreira, D. G. Osorio, G. Maya, and E. A. Viteri, “ANÁLISIS DE LOS PROTOCOLOS DE ATENCIÓN SOBRE VIOLENCIA DE GÉNERO Y SU PERSPECTIVA EN EL ECUADOR,” *Nº*, vol. 1, pp. 1–12, Mar. 2020, Accessed: Jun. 03, 2023. [Online]. Available: <https://uctunexpo.autanabooks.com/index.php/uct/article/view/294/522>
- [3] Y. Quintana *et al.*, *LA VIOLENCIA DE GÉNERO CONTRA LAS MUJERES*

- EN EL ECUADOR: Análisis de los resultados de la Encuesta Nacional sobre Relaciones Familiares y Violencia de Género contra las Mujeres*, vol. 1. Quito, 2014. Accessed: Jun. 03, 2023. [Online]. Available: <https://www.unicef.org/ecuador/media/2466/file/La%20Violencia%20de%20G%C3%A9nero%20contra%20las%20mujeres%20en%20el%20Ecuador.pdf>
- [4] M. R. Camacho *et al.*, “Violencia intrafamiliar y su repercusión en menores de la provincia de Bolívar, Ecuador,” *Rev Colomb Psiquiatr*, vol. 49, no. 1, pp. 23–28, Jan. 2020, doi: 10.1016/J.RCP.2018.04.006.
- [5] A. T. Sree Dhevi, “Imputing missing values using Inverse Distance Weighted Interpolation for time series data,” in *2014 Sixth International Conference on Advanced Computing (ICoAC)*, IEEE, Dec. 2014, pp. 255–259. doi: 10.1109/ICoAC.2014.7229721.
- [6] “Encuesta Nacional sobre Relaciones Familiares y Violencia de Género contra las Mujeres-ENVIGMU,” *INEC*, pp. 4–10, Sep. 2019, Accessed: May 29, 2023. [Online]. Available: https://www.ecuadorencifras.gob.ec/documentos/web-inec/Estadisticas_Sociales/Violencia_de_genero_2019/Principales%20resultados%20ENVIGMU%202019.pdf
- [7] A. Farouk and A. Nasser, “THE DECISION TREE IN THE ANALYSIS OF COEXISTENCE IN CASES OF FATAL VICTIMS OF GENDER VIOLENCE IN SPAIN,” vol. 5, pp. 1–11, 2020, doi: 10.5281/zenodo.4161179.
- [8] B. Beltrán, “Minería de Datos,” *Facultad de Ciencias de la Computación*, pp. 6–13, Accessed: Jun. 06, 2023. [Online]. Available: <https://www.cs.buap.mx/~bbeltran/NotasMD.pdf>
- [9] K. AMORES and K. TRELLES, “Implementación de un sistema para detectar la enfermedad de la Sigatoka Negra en una plantación de banano empleando técnicas de visión artificial,” *ESCUELA DE INGENIERÍA ELECTRÓNICA*, pp. 1–21, 2022, Accessed: Jun. 04, 2023. [Online]. Available: <https://dspace.uazuay.edu.ec/bitstream/datos/11682/1/17211.pdf>
- [10] D. Guillermo, C. Ortega, E. Dueñas, and H. Sosa, “Educación Digital y la Minería de Datos Educativos,” *Simposio Argentino sobre Tecnología y Sociedad*, pp. 129–143, Oct. 2020, Accessed: Jun. 04, 2023. [Online]. Available: <http://fce.uncu.edu.ar>
- [11] S. Grisales *et al.*, “Imputación de datos perdidos en series de mediciones de contaminantes atmosféricos insumo para la vigilancia en salud ambiental,” in *2019 Congreso Colombiano y Conferencia Internacional de Calidad de Aire y Salud Pública (CASAP)*, IEEE, Aug. 2019, pp. 1–7. doi: 10.1109/CASAP.2019.8916686.
- [12] A. Muñoz, “Propuesta y evaluación de una estrategia para la imputación múltiple y multivariada de valores faltantes en series de tiempo del campo meteorológico utilizando aprendizaje automático,” *Escuela de Ingeniería en Computación*, pp. 1–110, Jun. 2022, Accessed: Jun. 04, 2023. [Online]. Available: <https://repositoriotec.tec.ac.cr/handle/238/14060>

- [13] O. GALLEGOS and C. ROSALES, “Desarrollo de una guía metodológica sobre minería de datos,” *FACULTAD DE INGENIERÍA ESCUELA DE SISTEMAS*, pp. 7–8, Jan. 2017, Accessed: Jun. 04, 2023. [Online]. Available: http://repositorio.puce.edu.ec/bitstream/handle/22000/13181/TrabajoDeDisertaci%C3%B3n_C%C3%B3rdova-Rosales.pdf?sequence=1&isAllowed=y
- [14] A. Schuschny and H. Soto, “Guía metodológica Diseño de indicadores compuestos de desarrollo sostenible,” May 2009. Accessed: Jun. 04, 2023. [Online]. Available: <https://repositorio.cepal.org/handle/11362/3661>
- [15] O. M. Cumbicus-Pineda, T. E. Abad-Eras, and L. A. Neyra-Romero, “Data Mining to Determine the Causes of Gender-Based Violence against Women in Ecuador,” in *2021 IEEE Fifth Ecuador Technical Chapters Meeting (ETCM)*, IEEE, Oct. 2021, pp. 1–6. doi: 10.1109/ETCM53643.2021.9590664.
- [16] J. M. Dueñas and H. Rifa, “Aprendizaje supervisado para la detección de amenazas web mediante clasificación basada en árboles de decisión,” *Máster Universitario en Seguridad de las Tecnologías de la Información y de las Comunicaciones*, vol. 1, no. 1, pp. 1–61, Jun. 2020, Accessed: Jun. 04, 2023. [Online]. Available: <https://openaccess.uoc.edu/bitstream/10609/118166/6/joseduenasTFM0620memoria.pdf>
- [17] I. García, “Diseño e implementación de un módulo Python para representar datos geográficos en Dataframes,” May 2019, Accessed: Jun. 05, 2023. [Online]. Available: <https://e-archivo.uc3m.es/handle/10016/31974>