

TRABAJO DE GRADO
Opción Seminario-Diplomado.

**ALGORITMO COMPUTACIONAL PARA EL ANÁLISIS Y TOMA DE DECISIONES
EN PROYECCIONES VIOLENCIA DOMÉSTICA CONTRA LA MUJER,
UTILIZANDO ESTRATEGIAS DE MACHINE LEARNING**

Corporación Universitaria Remington.
Nombre de la facultad: Ingeniería
Nombre del programa académico: Ingeniería de sistemas

Estudiantes:
Breisen Estiven Quintero Masmela
Nicolas Gongora Molina
Tutor: Juan Carlos Briñez de León
Opción de Trabajo de grado Seminario-Diplomado.
2024.

Dedicatoria

A MI FAMILIA, QUIENES HAN ESTADO EN TODA ESTA BATALLA ACADÉMICA, Y
POR QUIENES NADA SERÍA, Y SI FUESE, ESTO NO TENDRÍA SENTIDO.

Agradecimientos

A DIOS, A LA UNIVERSIDAD. A MIS COMPAÑEROS, AL PROFE DEL SEMINARIO Y A AQUELLO PROFESORES QUE BRINDARON SU CONOCIMIENTO Y ÉTICA, Agradezco a todos mis compañeros que me ayudaron en este proceso e hicieron este recorrido más ameno.

Tabla de Contenidos

Resumen	5
Palabras clave	5
Título 1	6
Sub-Título 1.1.	6
Acercamiento de datos	6
Descripción de variables	8
aproximaciones con gráficos- analítica	8
Objetivos	30
Objetivos específicos	30
Desarrollo e implementación de aprendizaje	30
Preparación de los datos	31
Modelo de toma de decisiones	34
Validación De modelo	34
Conclusiones - trabajos futuros	36
Referencias	37

Resumen

El trabajo de grado presenta el desarrollo de un algoritmo computacional utilizando estrategias de machine learning para el análisis y toma de decisiones en proyecciones de violencia doméstica contra la mujer. La investigación se enfoca en identificar y analizar el ciclo intergeneracional de la violencia doméstica, considerando factores como la edad, nivel educativo, empleo, estado civil y otros parámetros socioeconómicos. Utilizando datos recopilados en áreas con diferentes niveles socioeconómicos y mediante entrevistas y encuestas a mujeres dispuestas a compartir sus experiencias, el estudio busca comprender la correlación entre estos factores y la violencia doméstica. Los datos, obtenidos de la plataforma Kaggle, fueron procesados y analizados, revelando que las mujeres jóvenes y con ingresos económicos inestables son las más afectadas. Los modelos de machine learning, especialmente LDA y SVM, mostraron una precisión del 74.3% en la predicción de casos de violencia. El objetivo general es implementar un algoritmo para analizar y tomar decisiones informadas a partir de estos datos, contribuyendo a la identificación y prevención de la violencia doméstica.

Palabras clave

Violencia, violencia doméstica, análisis, predicción, toma de decisiones.

Marco conceptual y contextual

Nuestra investigación fue identificar y analizar la existencia del ciclo intergeneracional de la violencia doméstica contra la mujer, por ejemplo: qué niveles de violencia sufre una mujer en diferentes circunstancias como las que presentamos en el caso como. Que edad tiene su nivel de educación si cuentan con algún empleo o si está casada o soltera estos son algunos de los parámetros que se están manejando para identificar cuales son las mujeres que más reciben violencia doméstica como se mencionó Cárdenas Varón, G., Polo Otero, J. L. (2014).

apreciamos que tenemos una gran problemática en nuestra sociedad en temas de violencia doméstica por lo tanto buscamos conclusiones mediante la analítica ya que facilita a la hora de tomar algunas decisiones y nos ayuda a encontrar las causantes de la violencia hacia la mujer.

Mediante los algoritmos de Machine Learning para la toma de decisiones nos aporta una gran ayuda con los datos y ya con esto obtenido se tomarán medidas las cuales nos faciliten en la ejecución en los planes expuestos.

Pregunta problema:

¿Cómo desarrollar una estrategia computacional para ayudar en la predicción en casos de maltrato doméstico en contra de la mujer, a partir de algoritmos de Machine Learning?

Acercamiento a los datos:

Para este proyecto se usó un dataset con información sobre la violencia doméstica cometida contra las mujeres en una zona rural de un país en desarrollo.

La función de los datos es ayudar a comprender la correlación entre diversos factores socioeconómicos y la violencia doméstica.

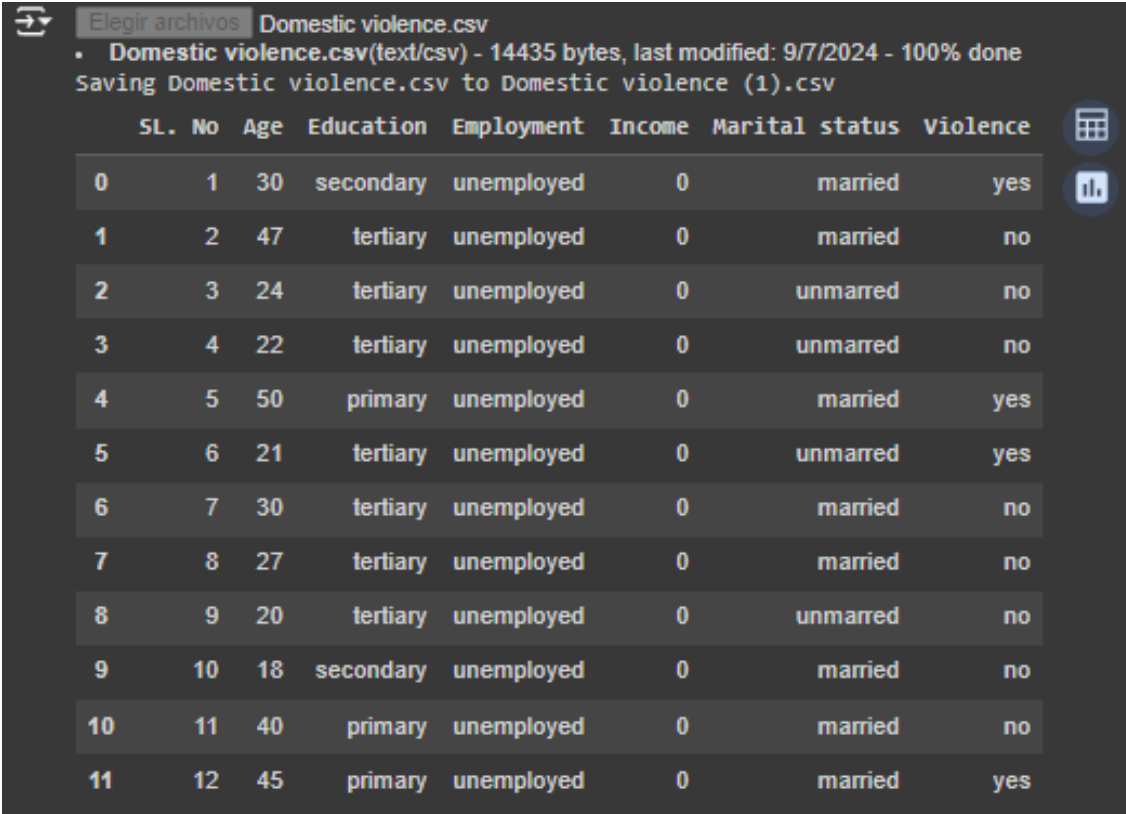


The screenshot shows a Kaggle dataset page for 'Domestic Violence Against Women'. At the top, it indicates the dataset was updated 21 days ago by user FAHMIDA. There are 35 notebooks, a 'New Notebook' button, and a 'Download (2 kB)' button. The dataset title is 'Domestic Violence Against Women' with a subtitle 'Domestic violence against women in a specific rural area of a developing country'. A thumbnail image shows a hand with fingers spread against a black background. Below the title, there are links for 'Data Card', 'Code (12)', 'Discussion (1)', and 'Suggestions (0)'. At the bottom, the 'About Dataset' section is visible, showing a 'Usability' score of 10.00.

En lo que a la recopilación de datos se refiere, se optó por ir a áreas con diferentes niveles socioeconómicos. Las mujeres que se encuestaron fueron las que estaban dispuestas a compartir sus experiencias en relaciones con violencia doméstica de por medio y también su situación socioeconómica, así como también se usaron entrevistas personales además de las encuestas, los trabajadores de campo, capacitados para llevar a cabo una recopilación de datos de forma ética y sensible, visitaron hogares y centros comunitarios para las respectivas encuestas y entrevistas, entre estas últimas habían preguntas específicas y preguntas abiertas para detallar cada una de las experiencias de las encuestadas lo mayor posible. Fue crucial para la investigación garantizar el total anonimato de las encuestadas para lograr una mayor confiabilidad, así como también informes más precisos y honestos.

Estos datos provienen de su respectivo dataset en la plataforma Kaggle bajo el nombre de “Domestic Violence Against Women”.

Descripción de variables.



Elegir archivos Domestic violence.csv

- Domestic violence.csv(text/csv) - 14435 bytes, last modified: 9/7/2024 - 100% done

Saving Domestic violence.csv to Domestic violence (1).csv

SL. No	Age	Education	Employment	Income	Marital status	Violence	
0	1	30	secondary	unemployed	0	married	yes
1	2	47	tertiary	unemployed	0	married	no
2	3	24	tertiary	unemployed	0	unmarried	no
3	4	22	tertiary	unemployed	0	unmarried	no
4	5	50	primary	unemployed	0	married	yes
5	6	21	tertiary	unemployed	0	unmarried	yes
6	7	30	tertiary	unemployed	0	married	no
7	8	27	tertiary	unemployed	0	married	no
8	9	20	tertiary	unemployed	0	unmarried	no
9	10	18	secondary	unemployed	0	married	no
10	11	40	primary	unemployed	0	married	no
11	12	45	primary	unemployed	0	married	yes

Los datos recolectados de las encuestas y entrevistas giraban en torno a las siguientes variables: “SL. No” que corresponde al número de serie o también llamado “ID” de todas las encuestadas, la edad, el nivel de educación máximo alcanzado, el estado laboral, Ingresos económicos, estado marital y finalmente si la encuestada había sido víctima de violencia doméstica.

```
0.2 ✓ #Información de la estructura de datos
Conjunto_Datos.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 347 entries, 0 to 346
Data columns (total 7 columns):
#   Column          Non-Null Count  Dtype
---  ---            -
0   SL. No          347 non-null   int64
1   Age             347 non-null   int64
2   Education       347 non-null   object
3   Employment      347 non-null   object
4   Income          347 non-null   int64
5   Marital status  347 non-null   object
6   Violence        347 non-null   object
dtypes: int64(3), object(4)
memory usage: 19.1+ KB
```

Haciendo un informe estructural de los datos se puede evidenciar que se recolectaron 347 datos de mujeres y también se recolectaron 7 columnas, 3 de ellas con variables de tipo entero y las otras 4 siendo variables en texto.

```
#Análisis estadístico rápido de los datos
Conjunto_Datos.describe()
```

	SL. No	Age	Income
count	347.000000	347.000000	347.000000
mean	174.000000	31.380403	2110.685879
std	100.314505	9.601569	5743.278766
min	1.000000	15.000000	0.000000
25%	87.500000	23.000000	0.000000
50%	174.000000	30.000000	0.000000
75%	260.500000	39.500000	0.000000
max	347.000000	60.000000	35000.000000

Haciendo un análisis estadístico superficial de los datos recopilados se observa que más de la mitad de las mujeres encuestadas estaban por debajo de los cuarenta años y tan solo el 5% de las mujeres tenía ingresos económicos estables entendiéndose como estos una media de 2.110 y unos máximos de ingresos de 35.000.

```
#Verificación de las opciones de la variable
print('Analizando nivel de estudios')
Datos_Loan['Education '].unique()
```

```
Analizando nivel de estudios
array(['secondary', 'tertiary', 'primary', 'none'], dtype=object)
```

Se analizó la columna “Education” que corresponde al nivel educativo máximo alcanzado de las mujeres encuestadas, exponiendo las variables “primary” y “secondary” para estudios de primaria y secundaria, “tertiary” para estudios superiores y “none” para ningún tipo de educación superior a terminar la primaria.

```
#Verificación de las opciones de la variable
print('Analizando ingresos')
Datos_Loan['Income'].unique()
```

```
Analizando ingresos
array([ 0, 8, 1000, 1200, 1500, 2000, 2200, 2500, 3000,
        3500, 4000, 5000, 5500, 6000, 7000, 8000, 9000, 10000,
        12000, 15000, 18000, 20000, 22000, 24000, 25000, 27000, 28000,
        30000, 35000])
```

En cuanto a los ingresos recopilados de las encuestadas se observa que algunas ganan por encima de 30.000, pero la verdad es que 269 de las encuestadas no tienen ingresos lo que equivale alrededor del 78% del total de encuestadas.

```
#Verificación de las opciones de la variable
print('Analizando tipo de trabajo')
Datos_Loan['Employment '].unique()
```

```
Analizando tipo de trabajo
array(['unemployed', 'semi employed', 'employed', 'employed '],
      dtype=object)
```

Se analizaron todas las variables de “Employment” que corresponden a la situación laboral de las encuestadas, con datos como “unemployed” para desempleada, “semi employed” para empleadas de medio tiempo y “employed” para empleadas con horarios mayores a 7 horas.

```
#Verificación de las opciones de la variable
print('Analizando tipo relación')
Datos_Loan['Marital status '].unique()
```

Analizando tipo relación
array(['married', 'unmarried'], dtype=object)

Se analizó la columna “Marital status” encargada de corroborar el estado civil de las encuestadas, con “married” para encuestadas casadas y “unmarried” para no casadas.

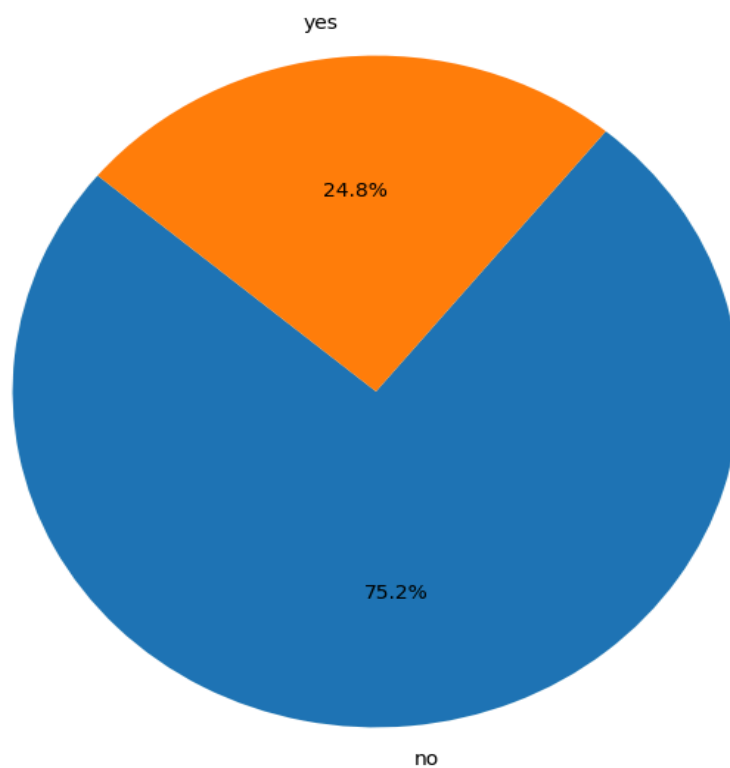
```
[ ] #Verificación de las opciones de la variable
print('Analizando la decisión')
Datos_Loan['Violence '].unique()
```

Analizando la decisión
array(['yes', 'no'], dtype=object)

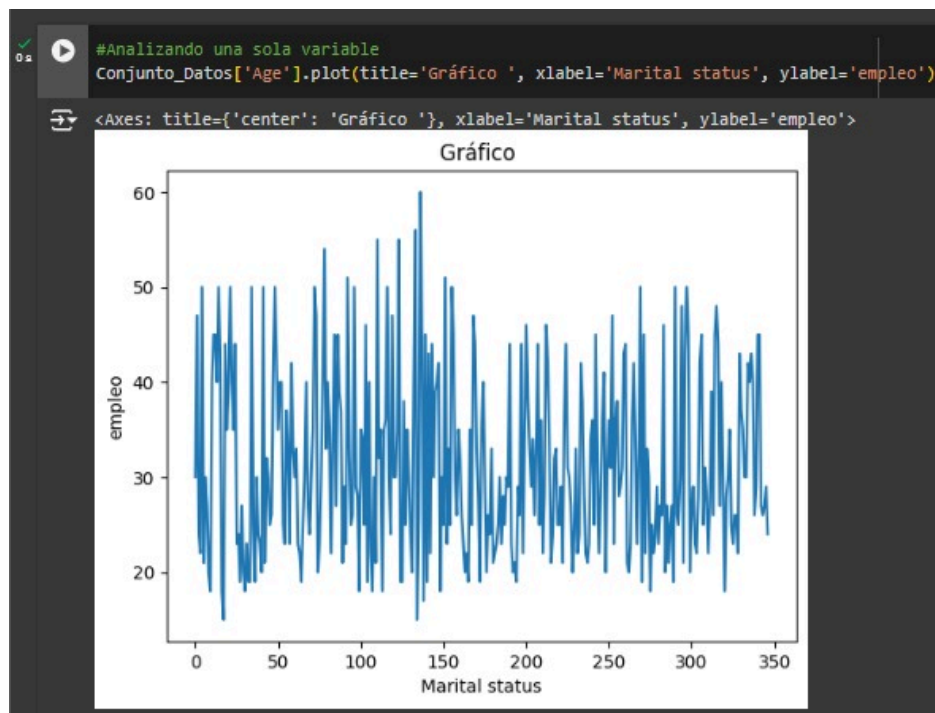
Finalmente, se tomó la columna “Violence” que indicaba si la encuestada había presentado violencia doméstica con simplemente sí o no.

Aproximaciones con gráficos - analítica.

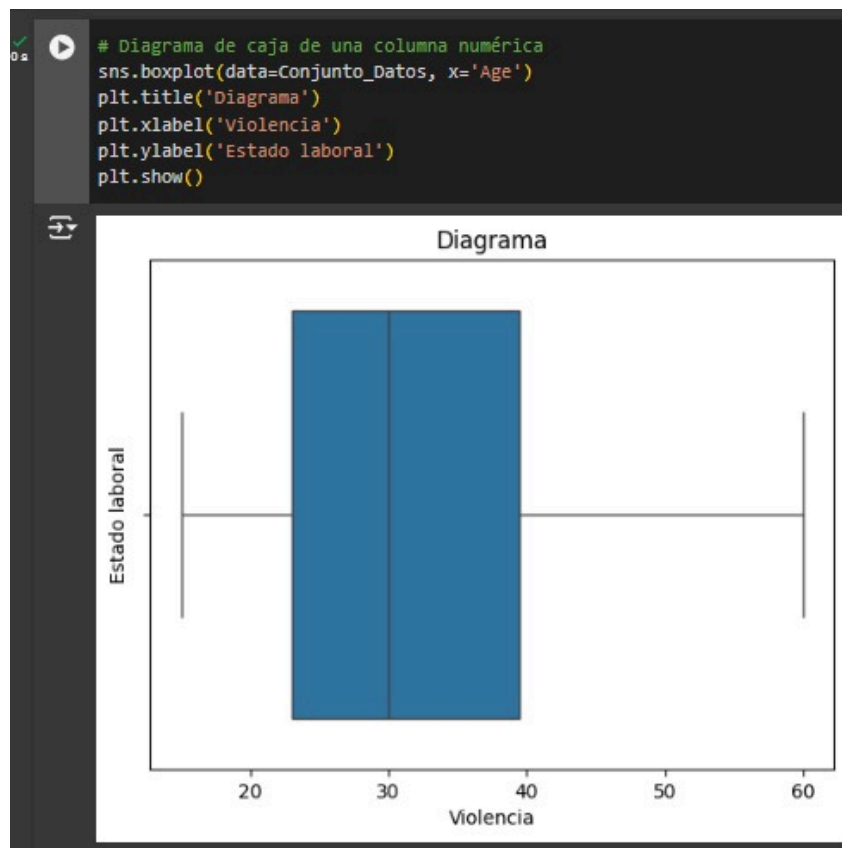
Presencia de Violencia de las encuestadas



Esta gráfica coloquialmente llamada “gráfica de pie” pone en evidencia el porcentaje de las mujeres que han sufrido violencia doméstica, así como las que no, dando como resultado que la presencia de violencia sea menor.

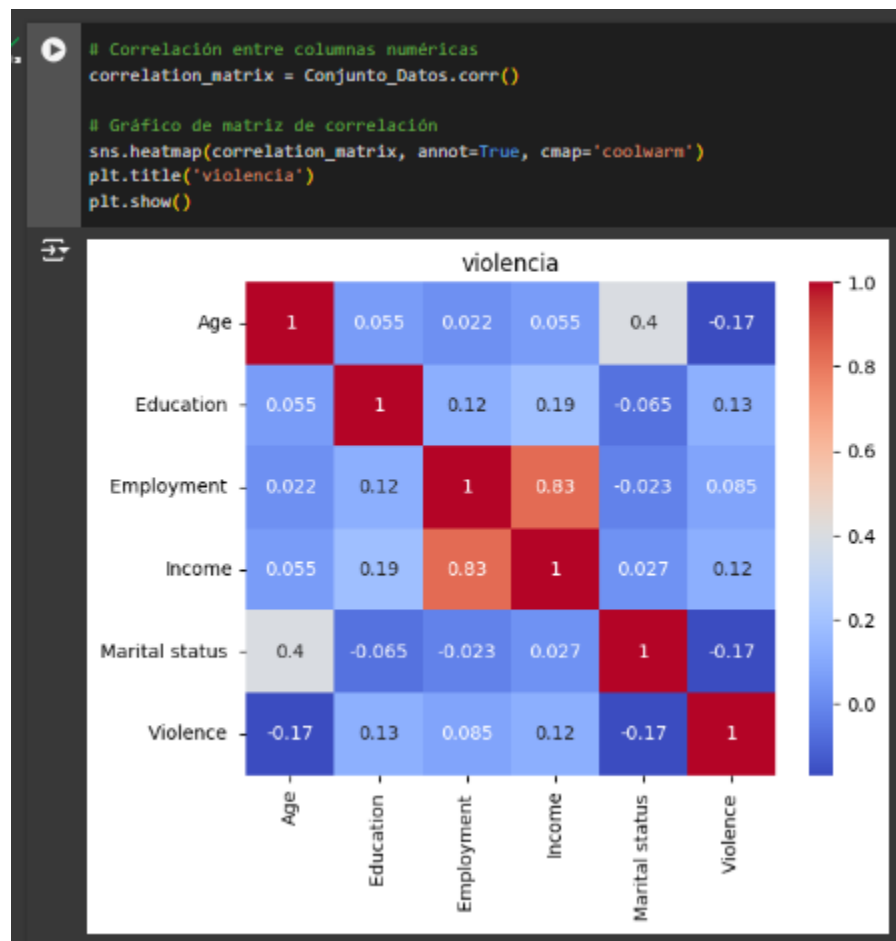


En la siguiente gráfica se puede apreciar una comparación entre el empleo y estado marital donde la mayoría de mujeres que cuentan con un empleo estable sufren menos violencia doméstica en este mismo caso también se ve reflejado en el estado civil se puede apreciar que las mujeres que están casadas cuentan con una mayor probabilidad de que sufran algún tipo de maltrato de violencia con esta gráfica ya podemos descartar una gran parte de las candidatas y enfocarnos más en aquellas que sufren más maltrato manejando la variable de la edad.



En el anterior diagrama vamos a manejar las siguientes variables las cuales nos ayudaran a entender un poco más a entender las verdaderas causas de la violencia doméstica en este caso utilizamos el estado laboral y el tema de la violencia.

Como se puede apreciar en la imagen en la parte de arriba se localiza que las mujeres con un estimado del 25 a 30 tiene una probabilidad de llegar a sufrir más violencia doméstica.



En esta gráfica correlacional se muestra una escala la cual se ve reflejado las diferentes variables que se manejaron de manera continua en el proceso de verificación de datos la cual nos ayuda a percibir con más exactitud los casos de violencia doméstica manejando todos los datos.

```
[ ] #Divide datos en entradas y salidas
import numpy as np
Datos_matriz=np.array(Datos_Loan)
#Datos_matriz[np.isnan(Datos_matriz)] = 0
X = Datos_matriz[:,0:5] #datos de entrada (Todas las variables de las encuestadas)
Y = Datos_matriz[:, -1] #Datos de salida (Presencia de violencia domestica)
print('Proceso realizado')
```

↔ Proceso realizado

Se establecen los valores de entrada y los de salida para comenzar el análisis de los datos para su posterior procesamiento, desembocando en el entrenamiento como tal del modelo de IA.

```
[ ] # Divide datos en Entrenamiento y testeo
import sklearn
from sklearn.model_selection import train_test_split
X_train, X_test, Y_train, Y_test= train_test_split(X,Y,test_size=0.1,random_state=751)
print('Proceso realizado')
```

↔ Proceso realizado

Se realiza el procedimiento para verificar que todo el proceso se realizó correctamente y el testeo nos brinda luz verde para continuar con nuestro manejo de datos los cuales corren perfectamente.

```
# Evaluando casos mediante todos los clasificadores
from sklearn.neighbors import KNeighborsClassifier
from sklearn.naive_bayes import GaussianNB
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
from sklearn.discriminant_analysis import QuadraticDiscriminantAnalysis
from sklearn.tree import DecisionTreeClassifier
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score, precision_score

Modelo_0 = KNeighborsClassifier(5)
Modelo_0.fit(X_train, Y_train)
Y_pred_0 = Modelo_0.predict (X_test)
print("Accuracy KNN", accuracy_score(Y_test, Y_pred_0))

Modelo_1 = GaussianNB()
Modelo_1.fit(X_train, Y_train)
Y_pred = Modelo_1.predict (X_test)
print("Accuracy Bayes", accuracy_score(Y_test, Y_pred))

Modelo_2 = LinearDiscriminantAnalysis()
Modelo_2.fit(X_train, Y_train)
Y_pred_2 = Modelo_2.predict (X_test)
print("Accuracy LDA", accuracy_score(Y_test, Y_pred_2))
```

```
print("Accuracy LDA",accuracy_score(Y_test, Y_pred_2))

Modelo_3 = QuadraticDiscriminantAnalysis()
Modelo_3.fit(X_train, Y_train)
Y_pred_3 =Modelo_3.predict (X_test)
print("Accuracy QDA",accuracy_score(Y_test, Y_pred_3))

Modelo_4 = DecisionTreeClassifier()
Modelo_4.fit(X_train, Y_train)
Y_pred_4 =Modelo_4.predict (X_test)
print("Accuracy Tree",accuracy_score(Y_test, Y_pred_4))

Modelo_5 = SVC()
Modelo_5.fit(X_train, Y_train)
Y_pred_5 =Modelo_5.predict (X_test)
print("Accuracy SVM",accuracy_score(Y_test, Y_pred_5))

Accuracy KNN 0.6857142857142857
Accuracy Bayes 0.4
Accuracy LDA 0.7428571428571429
Accuracy QDA 0.37142857142857144
Accuracy Tree 0.6285714285714286
Accuracy SVM 0.7428571428571429
```

Al correr los diferentes modelos de IA los resultados muestran que aquellos con mayor precisión para determinar la presencia de violencia doméstica en una posible relación son el modelo "LDA" y el modelo "SVM" con un resultado redondeado de 74.3% de éxito.

```

import matplotlib.pyplot as plt
from sklearn.metrics import confusion_matrix, roc_curve, auc, classification_report
import seaborn as sns

# Calcular métricas
accuracy = Modelo_3.score(X_test, Y_test)
cm = confusion_matrix(Y_test, Y_pred_0)
report = classification_report(Y_test, Y_pred_0)

# Calcular la curva ROC y el área bajo la curva (AUC)
Y_pred_prob = Modelo_3.predict_proba(X_test)[: , 1]
fpr, tpr, _ = roc_curve(Y_test, Y_pred_prob)
roc_auc = auc(fpr, tpr)

# Mostrar resultados
print(f"Accuracy: {accuracy:.2f}")
print('=====')
print(' ')

# Visualizar la matriz de confusión
plt.figure(figsize=(6, 4))
sns.heatmap(cm, annot=True, fmt='d', cmap='Blues', cbar=False)
plt.xlabel('Predicted labels')
plt.ylabel('True labels')
plt.title('Confusion Matrix')
plt.show()
print('=====')

```

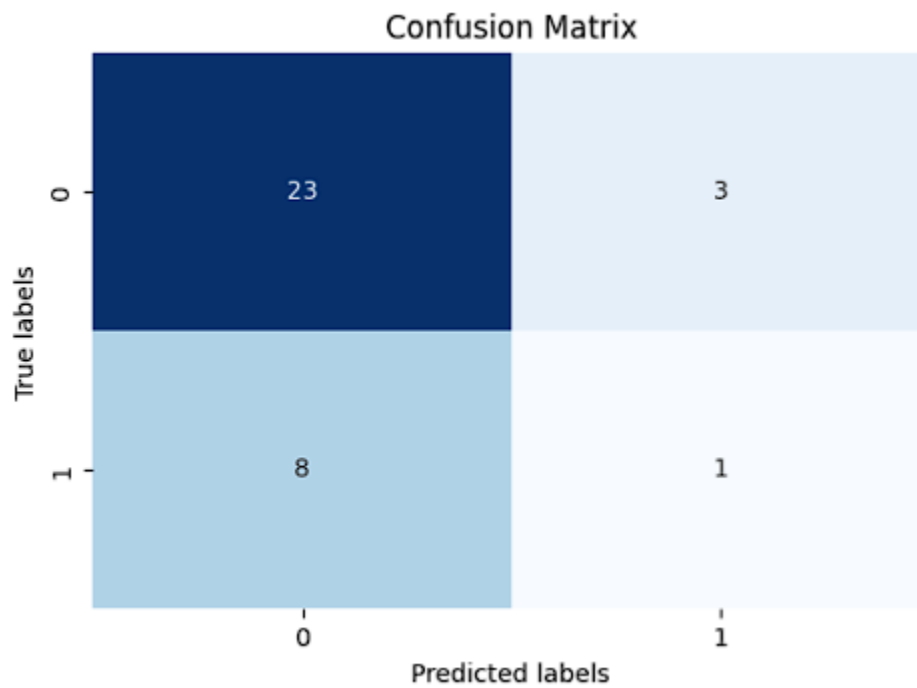
```

print(' ')
print("Classification Report:")
print(report)
print('=====')

# Visualizar la curva ROC
print(' ')
plt.figure()
lw = 2
plt.plot(fpr, tpr, color='darkorange',lw=lw, label='ROC curve (area = %0.2f)' % roc_auc)
plt.plot([0, 1], [0, 1], color='navy', lw=lw, linestyle='--')
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.05])
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('Receiver Operating Characteristic Curve')
plt.legend(loc="lower right")
plt.show()

```

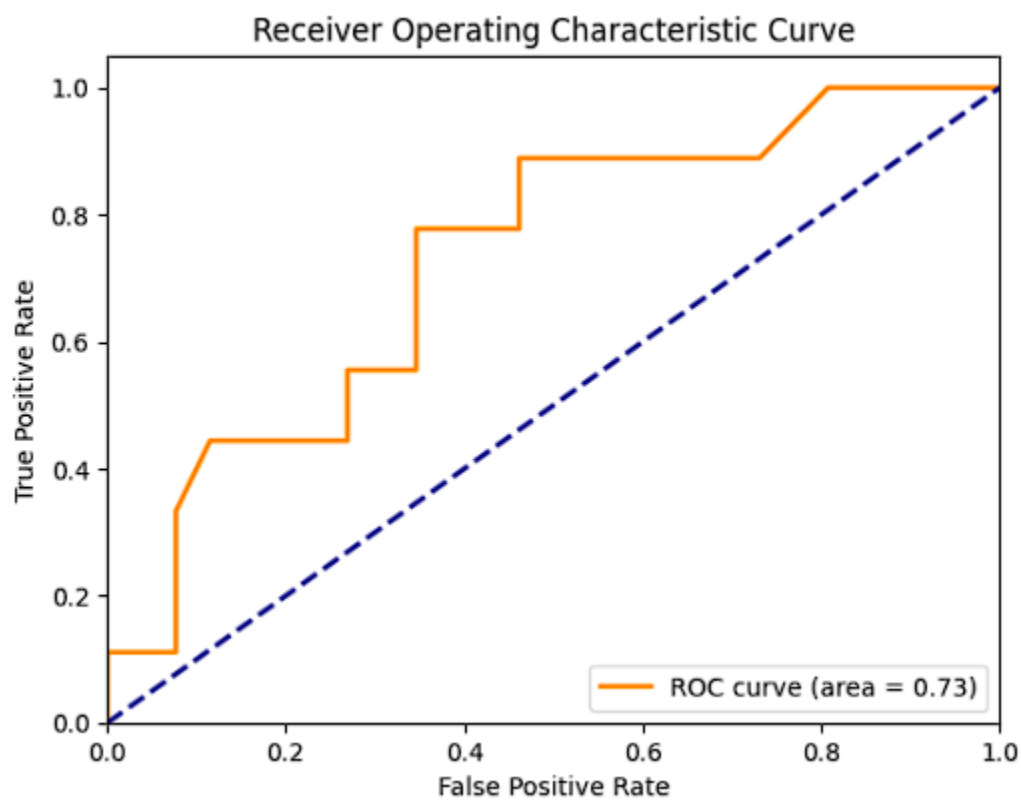
Se presenta la máquina artificial que vamos a manejar en este caso vamos a utilizar la que nos brinda una mayor facilidad en el manejo de los datos en el tema de los test y para algunas afirmaciones si son falsas o verdaderas.



En el recuadro anterior muestra unas predicciones en el cual nos informa que podemos llegar a tener un 23 en el parámetro 0 a 0 y menor parámetro se nos presenta en el 1 a 1 que sería el resulta 1.

```
=====  
Classification Report:  
      precision    recall  f1-score   support  
  
     0       0.74      0.88      0.81      26  
     1       0.25      0.11      0.15       9  
  
 accuracy          0.69      35  
 macro avg       0.50      0.50      0.48      35  
 weighted avg    0.62      0.69      0.64      35  
=====
```

Se informa de las posibles predicciones que pueden llegar a ocurrir y también es el support que nos ayuda a la hora de poder llegar a una conclusión.



Aquí se evidencia la efectividad de la IA con una curva del ratio positivo de la predicción, hecha para apreciar tanto el valor positivo como el negativo que puede desempeñar la ejecución del modelo de IA.

```
#Probando el modelo entrenado sobre un nuevo sujeto
Target=np.zeros((1,5))
Target[0,0]=float(input('Ingrese edad, entre 17 y 60: '))
Target[0,1]=float(input('Ingrese estado civil, 30 casada y 60 no casada: '))
Target[0,2]=float(input('Ingrese nivel educativo, 20 ninguno, 40 primaria, 60 secundaria, 80 estudios superiores: '))
Target[0,3]=float(input('Ingrese estado laboral, 20 desempleada, 40 trabajo de medio tiempo 60 trabajo: '))
Target[0,4]=float(input('Ingrese sus ingresos economicos, entre 0 y 50000: '))

Target = scaler.transform(Target) #Normalizar los datos

Prediction_0 =Modelo_0.predict (Target)
Prediction_1 =Modelo_1.predict (Target)
Prediction_2 =Modelo_2.predict (Target)
Prediction_3 =Modelo_3.predict (Target)
Prediction_4 =Modelo_4.predict (Target)
Prediction_5 =Modelo_5.predict (Target)
```

```
print(" ")

if Prediction_0==0:
    print("Según KNN, no presentara violencia")
else:
    print("Según KNN, presentara violencia")

print(" ")

if Prediction_1==0:
    print("Según Bayes, no presentara violencia")
else:
    print("Según Bayes, presentara violencia")

print(" ")

if Prediction_2==0:
    print("Según LDA, no presentara violencia")
else:
    print("Según LDA, presentara violencia")
```

```
if Prediction_3==0:  
    print("Según QDA, no presentara violencia")  
else:  
    print("Según QDA, presentara violencia")  
  
print(" ")  
  
if Prediction_4==0:  
    print("Según Tree, no presentara violencia")  
else:  
    print("Según tree, presentara violencia")  
  
print(" ")  
  
if Prediction_5==0:  
    print("Según SVM, no presentara violencia")  
else:  
    print("Según SVM, presentara violencia")  
  
print(" ")
```

Este es el modelo de la IA la cual nos ayuda a la hora de intervenir en casos de violencia ya que nos muestra cuáles casos tiene mayor posibilidad a la hora de algún caso de violencia doméstica mediante las diferentes IA que se están manejando nos muestra diferentes textos, el primero será el de “no presenta violencia” y el segundo siendo “presenta violencia”.

A continuación, se mostrarán 10 ejercicios probando todos los modelos de IA:

Ejercicio

```
Ingrese edad, entre 17 y 60: 35
Ingrese estado civil, 30 casada y 60 no casada: 30
Ingrese nivel educativo, 20 ninguno, 40 primaria, 60 secundaria, 80 estudios superiores: 60
Ingrese estado laboral, 20 desempleada, 40 trabajo de medio tiempo 60 trabajo: 40
Ingrese sus ingresos economicos, entre 0 y 50000: 17000
```

Según KNN, no presentara violencia

Según Bayes, no presentara violencia

Según LDA, no presentara violencia

Según QDA, no presentara violencia

Según Tree, no presentara violencia

Según SVM, no presentara violencia

Ejercicio 2

Ingrese edad, entre 17 y 60: 17
 Ingrese estado civil, 30 casada y 60 no casada: 60
 Ingrese nivel educativo, 20 ninguno, 40 primaria, 60 secundaria, 80 estudios superiores: 60
 Ingrese estado laboral, 20 desempleada, 40 trabajo de medio tiempo 60 trabajo: 20
 Ingrese sus ingresos economicos, entre 0 y 50000: 0

Según KNN, presentara violencia

Según Bayes, no presentara violencia

Según LDA, presentara violencia

Según QDA, no presentara violencia

Según tree, presentara violencia

Según SVM, no presentara violencia

Ejercicio 3

Ingrese edad, entre 17 y 60: 40
 Ingrese estado civil, 30 casada y 60 no casada: 30
 Ingrese nivel educativo, 20 ninguno, 40 primaria, 60 secundaria, 80 estudios superiores: 80
 Ingrese estado laboral, 20 desempleada, 40 trabajo de medio tiempo 60 trabajo: 60
 Ingrese sus ingresos economicos, entre 0 y 50000: 40000

Según KNN, no presentara violencia

Según Bayes, no presentara violencia

Según LDA, no presentara violencia

Según QDA, no presentara violencia

Según Tree, no presentara violencia

Según SVM, no presentara violencia

Ejercicio 4

Ingrese edad, entre 17 y 60: 20
 Ingrese estado civil, 30 casada y 60 no casada: 30
 Ingrese nivel educativo, 20 ninguno, 40 primaria, 60 secundaria, 80 estudios superiores: 20
 Ingrese estado laboral, 20 desempleada, 40 trabajo de medio tiempo 60 trabajo: 20
 Ingrese sus ingresos economicos, entre 0 y 50000: 0

Según KNN, no presentara violencia

Según Bayes, no presentara violencia

Según LDA, no presentara violencia

Según QDA, no presentara violencia

Según tree, presentara violencia

Según SVM, no presentara violencia

Ejercicio 5

Ingrese edad, entre 17 y 60: 45
 Ingrese estado civil, 30 casada y 60 no casada: 30
 Ingrese nivel educativo, 20 ninguno, 40 primaria, 60 secundaria, 80 estudios superiores: 20
 Ingrese estado laboral, 20 desempleada, 40 trabajo de medio tiempo 60 trabajo: 40
 Ingrese sus ingresos economicos, entre 0 y 50000: 15000

Según KNN, no presentara violencia

Según Bayes, no presentara violencia

Según LDA, no presentara violencia

Según QDA, no presentara violencia

Según Tree, no presentara violencia

Según SVM, no presentara violencia

Ejercicio 6

Ingrese edad, entre 17 y 60: 36
 Ingrese estado civil, 30 casada y 60 no casada: 30
 Ingrese nivel educativo, 20 ninguno, 40 primaria, 60 secundaria, 80 estudios superiores: 60
 Ingrese estado laboral, 20 desempleada, 40 trabajo de medio tiempo 60 trabajo: 40
 Ingrese sus ingresos economicos, entre 0 y 50000: 25000

Según KNN, no presentara violencia

Según Bayes, no presentara violencia

Según LDA, no presentara violencia

Según QDA, no presentara violencia

Según Tree, no presentara violencia

Según SVM, no presentara violencia

Ejercicio 7

Ingrese edad, entre 17 y 60: 21
 Ingrese estado civil, 30 casada y 60 no casada: 60
 Ingrese nivel educativo, 20 ninguno, 40 primaria, 60 secundaria, 80 estudios superiores: 20
 Ingrese estado laboral, 20 desempleada, 40 trabajo de medio tiempo 60 trabajo: 20
 Ingrese sus ingresos economicos, entre 0 y 50000: 0

Según KNN, no presentara violencia

Según Bayes, no presentara violencia

Según LDA, no presentara violencia

Según QDA, no presentara violencia

Según Tree, no presentara violencia

Según SVM, no presentara violencia

Ejercicio 8

```
Ingrese edad, entre 17 y 60: 52
Ingrese estado civil, 30 casada y 60 no casada: 60
Ingrese nivel educativo, 20 ninguno, 40 primaria, 60 secundaria, 80 estudios superiores: 60
Ingrese estado laboral, 20 desempleada, 40 trabajo de medio tiempo 60 trabajo: 20
Ingrese sus ingresos economicos, entre 0 y 50000: 0
```

Según KNN, presentara violencia

Según Bayes, no presentara violencia

Según LDA, presentara violencia

Según QDA, no presentara violencia

Según tree, presentara violencia

Según SVM, no presentara violencia

Ejercicio 9

```
Ingrese edad, entre 17 y 60: 47
Ingrese estado civil, 30 casada y 60 no casada: 60
Ingrese nivel educativo, 20 ninguno, 40 primaria, 60 secundaria, 80 estudios superiores: 60
Ingrese estado laboral, 20 desempleada, 40 trabajo de medio tiempo 60 trabajo: 60
Ingrese sus ingresos economicos, entre 0 y 50000: 22000
```

Según KNN, no presentara violencia

Según Bayes, no presentara violencia

Según LDA, no presentara violencia

Según QDA, no presentara violencia

Según Tree, no presentara violencia

Según SVM, no presentara violencia

Ejercicio 10

```
Ingrese edad, entre 17 y 60: 47
Ingrese estado civil, 30 casada y 60 no casada: 60
Ingrese nivel educativo, 20 ninguno, 40 primaria, 60 secundaria, 80 estudios superiores: 60
Ingrese estado laboral, 20 desempleada, 40 trabajo de medio tiempo 60 trabajo: 60
Ingrese sus ingresos economicos, entre 0 y 50000: 22000
```

Según KNN, no presentara violencia

Según Bayes, no presentara violencia

Según LDA, no presentara violencia

Según QDA, no presentara violencia

Según Tree, no presentara violencia

Según SVM, no presentara violencia

De los anteriores ejercicios una de las conclusiones que se pueden apreciar es que de 4 mujeres que no tenían ningún tipo de ingreso económico 2 de ellas presentan violencia doméstica según la IA del modelo 3 que es el que tiene mayor ratio positivo de predicción, es decir, el que mayor exactitud tiene, por lo que se puede inferir que una mujer vulnerable económicamente es más propensa a sufrir violencia doméstica.

Objetivos:**Objetivo general.**

Implementar un algoritmo computacional para el análisis y toma de decisiones a partir de datos de mujeres las cuales sufren violencia doméstica, utilizando estrategias de machine learning.

Objetivos específicos.

- Caracterizar y procesar los datos de interés, con miras a la toma de decisiones informadas.
- Implementar un algoritmo de Machine learning para la toma de decisiones a partir de los datos de interés.
- Evaluar y analizar el desempeño de los algoritmos implementados para la toma de decisiones.
- Validar el funcionamiento de toma de decisiones a partir de datos nuevos.

Desarrollo e implementación del aprendizaje

Para este trabajo, inicialmente se limpiaron los datos de manual, utilizando librerías de Python, Excel,, etc Luego se eliminaron variables que no corresponden a la toma de decisiones, se implementó un modelo de machine learning en clasificación, se evaluó el desempeño y se puso a prueba manejando todos los datos. A continuación se desarrollan cada una de las etapas de la metodología.

Preparación de los datos

```
#Quitando columnas indeseadas
Conjunto_Datos = Conjunto_Datos.drop(['SL. No'],axis=1)
#resumen de los datos
Conjunto_Datos.head(10)

#Ojo: Esta parte no la debe correr dos veces
```

	Age	Education	Employment	Income	Marital status	Violence
0	30	secondary	unemployed	0	married	yes
1	47	tertiary	unemployed	0	married	no
2	24	tertiary	unemployed	0	unmarried	no
3	22	tertiary	unemployed	0	unmarried	no
4	50	primary	unemployed	0	married	yes
5	21	tertiary	unemployed	0	unmarried	yes
6	30	tertiary	unemployed	0	married	no
7	27	tertiary	unemployed	0	married	no
8	20	tertiary	unemployed	0	unmarried	no
9	18	secondary	unemployed	0	married	no

Para un mayor entendimiento de los datos al momento de hacer análisis más detallados es necesario deshacerse de las variables que no aportan a dicho análisis, en este caso la variable llamada “SL. No” que funciona como un tipo de ID para las mujeres

encuestadas, por ende, es irrelevante y procedimos a quitar la columna de la tabla original.

```
[ ] #Elimina filas que tengan datos nulos
Datos_Loan=Datos_Loan.dropna()
Datos_Loan.head()
```

	Age	Education	Employment	Income	Marital status	Violence
0	30	secondary	unemployed	0	married	yes
1	47	tertiary	unemployed	0	married	no
2	24	tertiary	unemployed	0	unmarried	no
3	22	tertiary	unemployed	0	unmarried	no
4	50	primary	unemployed	0	married	yes

Se ejecutó una línea de código para borrar los campos sin información de la base de datos para evitar posibles problemas en la lectura y análisis de dichos datos.

```
[ ] #Mapeando todas la variables categóricas a numéricas
Reemplazo_1={'none':20,'primary':40,'secondary':60,'tertiary':80}
Datos_Loan['Education ']=Datos_Loan['Education '].map(Reemplazo_1)

Reemplazo_2={'unemployed':20,'semi employed':40,'employed':60,'employed ':60}
Datos_Loan['Employment ']=Datos_Loan['Employment '].map(Reemplazo_2)

Reemplazo_3={'married':30,'unmarried':60}
Datos_Loan['Marital status ']=Datos_Loan['Marital status '].map(Reemplazo_3)

Reemplazo_4={'yes':1,'no':0}
Datos_Loan['Violence ']=Datos_Loan['Violence '].map(Reemplazo_4)

Datos_Loan.head()
```

En la interior imagen se ve reflejada el cambio de variables de la diferentes tablas como educación la cual cuentas con las siguientes variables none, primary, secondary, tertiary

se reemplazaron y tomaron una forma numérica la cual no ayudará a la hora del manejo de esta misma y también se reemplazó empleo la cuenta con tan solo 3 variable las cuales fueron unemployed, semi employed, employed y también en el caso de estado marital el cual solo contaba con casada y soltera y por último el factor más importante es violencia sí o no eso fueron todos los cambios aplicados en la programa para facilitar el manejo de este más adelante.

	Age	Education	Employment	Income	Marital status	Violence
0	30	60	20	0	30	1
1	47	80	20	0	30	0
2	24	80	20	0	60	0
3	22	80	20	0	60	0
4	50	40	20	0	30	1

Esta es la tabla con los valores ya cambiados quedando así de la siguiente manera:

- En el caso de “Education” se cambiaron los valores “none” por “20”, “primary” por “40”, “secondary” por “60” y “tertiary” por “80”.
- En la columna “Employment” se reemplazaron los valores “unemployed” con “20”, “semi employed” con “40” y “employed” con “60”.
- En la columna “Marital status” se cambiaron los valores “married” por 30 y “unmarried” por 60.
- Finalmente, en la columna de “Violence” se reemplazaron las respuestas “yes” por “1” y “no” por “0”.

Validación del modelo

Como funciona para datos nuevos, y como un usuario le da uso a los algoritmo

Una vez que se corre el modelo entrenado con las entradas y salidas necesarias, se le pedirá al usuario que ingrese los datos correspondientes.

```
if Prediction_5==0:  
    print("Según SVM, no presentara violencia")  
else:  
    print("Según SVM, presentara violencia")  
  
print(" ")
```

Ingrese edad, entre 17 y 60:

```
Ingrese edad, entre 17 y 60: 22  
Ingrese estado civil, 30 casada y 60 no casada: 30  
Ingrese nivel educativo, 20 ninguno, 40 primaria, 60 secundaria, 80 estudios superiores: 60  
Ingrese estado laboral, 20 desempleada, 40 trabajo de medio tiempo 60 trabajo: 60  
Ingrese sus ingresos economicos, entre 0 y 50000: 
```

Cuando todos los datos han sido introducidos el modelo arroja la información proporcionada y seguido muestra la predicción de diferentes modelos de IA sobre si la mujer será sensible a una futura violencia doméstica para así tomar acción.

```
Ingrese edad, entre 17 y 60: 22
Ingrese estado civil, 30 casada y 60 no casada: 30
Ingrese nivel educativo, 20 ninguno, 40 primaria, 60 secundaria, 80 estudios superiores: 60
Ingrese estado laboral, 20 desempleada, 40 trabajo de medio tiempo 60 trabajo: 60
Ingrese sus ingresos economicos, entre 0 y 50000: 23000
```

Según KNN, no presentara violencia

Según Bayes, no presentara violencia

Según LDA, no presentara violencia

Según QDA, no presentara violencia

Según Tree, no presentara violencia

Según SVM, no presentara violencia

Conclusiones y trabajos futuros

En este trabajo se materializó la construcción de un algoritmo para la toma de decisiones informadas en el caso de violencia doméstica para bajar los niveles de maltrato hacia la mujer, utilizando algoritmos de machine learning.

Se observó que las mujeres con estudios mínimos y que no cuentan con un bienestar económico digno son propensas a sufrir situaciones de violencia doméstica, también se llegó a la conclusión de que la edad no es una variable tan determinante a la hora de medir la violencia como lo son otros factores, dejando en evidencia que las mujeres pueden sufrir violencia doméstica a cualquier edad y se puede dar basado en su nivel de educación y de su situación laboral.

Referencias

Cárdenas Varón, G., & Polo Otero, J. L. (2014). Ciclo intergeneracional de la violencia doméstica contra la mujer: análisis para las regiones de Colombia. *Revista de economía del Caribe*, (14), 1-33.