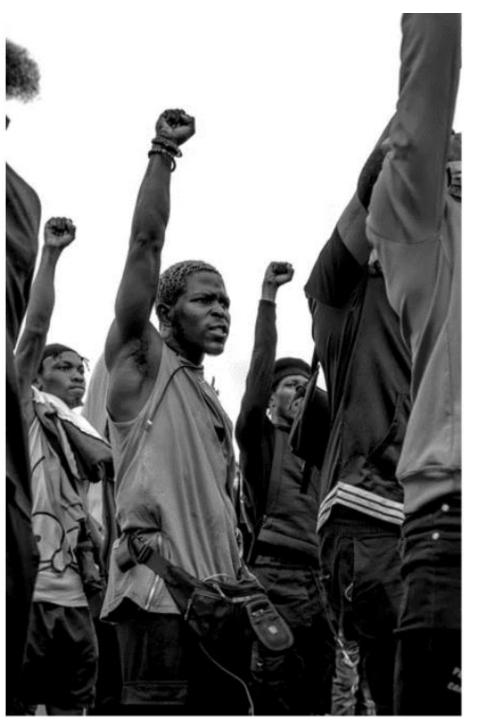


MUERTES CIVILES CAUSADAS POR POLCE FUERZAS POLICIALES EN USA (2015-2020)

ÍNDICE

☐ Motivación	Pág3
☐ Objetivo de la Investigación	Pág4
☐ Línea de tiempo	Pág5
☐ Objetivos del modelo	Pág6
☐ Fuentes del dataset	Pág7
☐ Variables del dataset	Pág8
☐ Data Wrangling	_Pág9
☐ Análisis Univariado	_ <i>Pág10</i>
☐ Análisis Bivariado	_ Pág12
☐ Análisis Multivariado	Pág13
☐ Algoritmos Elegidos	Pág15
☐ Decision Tree	Pág16
☐ Decision Tree – Ajuste de variables	Pág18
☐ Xgboost Classifier	_ Pág20
☐ Métricas Finales del Modelo Optimizado	_ Pág21
☐ Futuras Líneas	_ Pág22





MOTIVACIÓN

Debido a la creciente problemática mundial de muertes civiles en ocasión de encuentros con fuerzas policiales, hemos decidido analizar los datos relacionados a todos los civiles muertos por parte de la policía en Estados Unidos, para los años 2015 a 2020 ambos inclusive, con el relevamiento de distintas condiciones o variables relacionadas al evento (raza del civil, edad, género, si se encontraba armado, tipificación del encuentro con la policía, si el agente fue enjuiciado o no, entre otros.) y a la ciudad o estado donde ocurrió el hecho, para tratar de relacionar la incidencia de ciertas variables socioeconómicas de los estados.





OBJETIVOS DE LA INVESTIGACIÓN

- ¿Cuál es la probabilidad de que una víctima civil tenga una enfermedad mental al momento del incidente con la policía?
- ¿Cuál es la distribución por raza de las víctimas?
- ¿Cual es la incidencia de los factores socioeconómicos y políticos de los estados en la cantidad de víctimas por millón de habitantes?
- ¿Cuáles son los estados con mayor cantidad de muertes por millón de habitantes?
- ¿Cuáles son las circunstancias mas comunes del encuentro entre el civil y la policía?, ¿los civiles generalmente están armados?



LINEA DE TIEMPO

20/03/2022

Elección personal de un dataset y/o temática posible



24/03/2022

Definición de 3 datasets y temáticas finalistas.
Devolución del tutor sobre ventajas y desventajas de cada uno.



01/04/2022

Definición de variable target y aplicación de proceso de Data wrangling

18/05/2022

Desarrollo de algoritmo DECISION TREE y ajuste de modelo por 10 parámetros principales



18/04/2022

Desarrollo de análisis bivariado y multivariado



06/04/2022

Desarrollo de análisis univariado



31/05/2022

Desarrollo de algoritmo XBOOSTER CLASSIFIER



06/06/2020

Desarrollo del documento ejecutivo





OBJETIVOS DEL MODELO

Se pretende estimar la probabilidad de ocurrencia de que una persona civil que muere a manos de la policía en EEUU posea una enfermedad mental.

Específicamente, y a los efectos de resaltar una situación poco considerada relacionada a esta problemática social, los análisis se centrarán en la condición de si la victima poseía una enfermedad mental al momento del encuentro con la policía, a los efectos de que la información y conclusiones obtenidas sean de utilidad para entes gubernamentales o asociaciones sin fines de lucro, con el objetivo de desarrollar herramientas de contención para evitar que personas con problemas mentales terminen abatidas por la policía.

BRING BACK THE DEATH PENALTY. BRING BACK OUR POLICE!

When the begreened to visit City could the pass. It is seried White begreened incline and under, to the respirate and the begreened incline and unders, to the respirate control or subspace of the series of the subspace of the series of the series of the power couldest that have to help us to be series of designs, being on the series of the law or help us to be series of designs, being on the series of the law or the law or help us to be series of designs, being on the law of law of the law of the law of law of law or law of law of the law of the law of law or law of law or law of law of the law of law or law law or law of law or law o

minigration broadware of all as we Levis to Many, then Vor Smallers in Whate, Milke ja Many Invest Vor Smallers in Whate, Milke ja Happinin med Amini — Invest had no give up the jakenower and sincernity and in the Bods of their had whate, the best had no give up the jakenower as the programment core taken families, which was the programment of the following the hadden of the street of the street, and results a world's called by the late of the street, as a result, the street of the following the following the following the street of the street, as a result of makes of the street, as a result of the street of the street, as a result of the street of the

ments in report and makes and fail outer again — and at their an group precedural in intermediate. Mayor foot has easier than been and reacco about the intermed them our hearty. I do not prost an I was to have these maggers and recordered. They denote the face of our office and, when they fail.

they whench for extensed the dear common. Here ware or in a returning to an interpretable problem that basis before conducting a relate on the cold recolories. No. Migrae Sender, I serve to have these towers the allowant self. List hear limited up to problem the considerate self. List hear limited up to problem the contraction and the contraction of the contraction of confidentation in strong, that attacks are too account pumple with many it recently waste that a sourceast county of migraes that "suggests for the company" is sent them.

tigen to enderment their anger if must their an ambrotish one support it must them as to played. How can one great tracking voltages the constanced travellations of the observed by consent angless of consents cannot be used that their CPSE, LIMICKEES, DAYS WIRELA AN ACTION AS ON ALE SAMELY BECAME.

When I may process I have be a shory with my farther and determined two process facilities secretary and the street of secretary and the secretary and the street of the secretary and the street of the secretary and the secretary

Let not quiditi sant gire fondi cue pullisa dispensationi privare qui hang un softe. L'indica 434 diseas throme din reminent industri of "quides longation" which is the private of the private disease. In any other letter and the private disease the proposition of the private disease and the proposition of the proposition of the proposition of the private disease of the private disease and the proposition of the private disease which have entered the right on the Third States of the reliable to the private disease which have entered the right on the Third States and the private disease of th





FUENTES DEL DATASET

Datos de hechos y civiles
https://github.com/washingtonpost/data-police-shootings

- Datos socioeconómicos de los estados https://data.ers.usda.gov/
- Datos de partidos políticos y gobernador por Estado

https://www.openicpsr.org/openicpsr/project/102000/version/V3/view

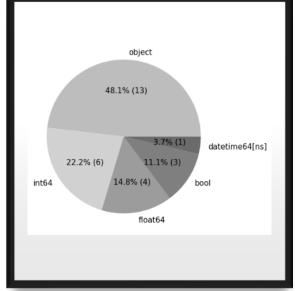
CRITERIOS DE SELECCIÓN

- Trascendencia de la problemática
- Description Oportunidad y completitud de los datos

VARIABLES DEL DATASET







#	Nombre	#Datos	Tipo
0	Civil_Name	5793	object
1	Death_Date	5793	datetime64[ns]
2	Manner_of_death	5793	object
3	Armed	5793	object
4	Age	5793	int64
5	Gender	5793	object
6	Race	5793	object
7	City	5793	object
8	State_ID	5793	int64
9	Signs_of_mental_illness	5793	bool
10	Flee	5793	object
11	body_camera	5793	bool
12	longitude	5519	float64
13	latitude	5519	float64
14	is_geocoding_exact	5793	bool
15	Official_Disposition	5793	object
16	Encounter_Type	5793	object
17	Year	5793	int64
18	state_initial	5793	object
19	GDP_Millions	5793	int64
20	GDP_PerCapita	5793	int64
21	Unemployment_Rate	5793	float64
22	Poverty_Percent	5793	float64
23	Median_household_Income	5793	int64
24	Governor_Name	5793	object
25	Party	5793	object
26	State_Name	5793	object

DATA WRANGLING

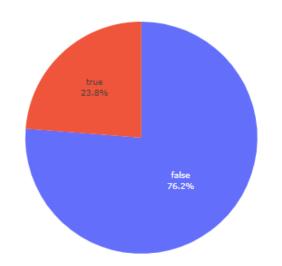
- Se considera que la completitud del dataset en cuanto a sus datos (5793 observaciones) es suficiente y sin errores significativos.
- Respecto a los datos faltantes de los campos "longitude" y "latitude", los mismos se completaron a través de "fillna", con la localización media de cada Estado, para que no se generen distorsiones significativas luego en visualizaciones espaciales.
- No se eliminó ningún campo, por considerarse todos relevantes, ya sea para el EDA o análisis descriptivo, tanto como para los modelos de predicción.
- A los efectos del modelado, se transformaron a "dummies" todos los campos de tipo "object".
- Asimismo, no se consideraron para los modelos las columnas de nombre de la víctima, fecha de muerte y variables relacionadas a la locación, por presenter escasa o nula correlación.



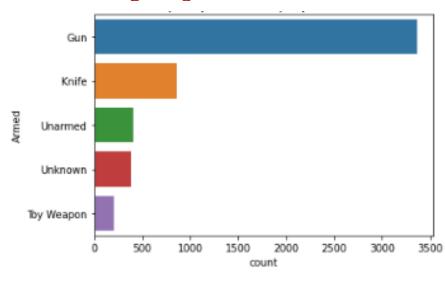
CE ANALISIS UNIVARIADO



Proporción de civiles con y sin indicios de enfermedad mental



Muertes por tipo de arm a que portaba la víctima



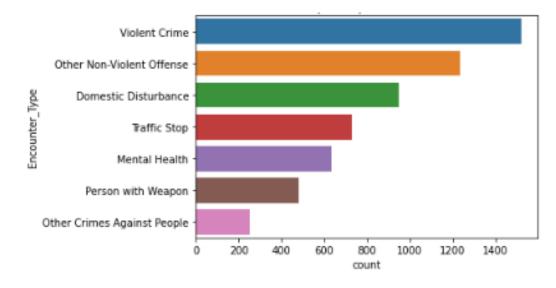
- Según estudios de la NAMI (National Alliance on Mental Illness) en EE.UU. 1 de cada 5 adultos sufre de una enfermedad mental, y 1 de cada 20 sufre una enfermedad mental grave.
- Hay un claro exceso de brutalidad policial en este caso comparando los números, ya que los casos más graves de enfermedad son 5,6% de la población, comparado al 23.8% que se observa que murió a manos de la policía.
- En cuanto a nivel de amenaza de la víctima civil abatida, podemos observer que al menos en 500 casos se poseía arma de juguete o directamente se encontraba desarmada.



ANALISIS UNIVARIADO

- En cuanto a tipo de incidente, puede observarse en el gráfico de la derecha que los casos de enfermedad mental sumados a los de disturbios domésticos o crímenes no violentos superan ampliamente a los crímenes violentos o civiles armados, lo que representaría, a priori un exceso policíaco el haber resultado muerto el civil involucrado.
- En cuanto al análisis por razas de las víctimas, según un studio de CNN, la etnia "blanca" promedio en EEUU representó en 2020 57,3%, teniendo 50,4% de víctimas civiles por parte de la policía, evidenciando una sub-representación. Por el contrario, la etnia "Negra", con una población promedio del 11,9%, presenta 27,5% de víctimas civiles, en una sobre-representación.

Cantidad de muertes civiles por tipo de encuentro con la policía





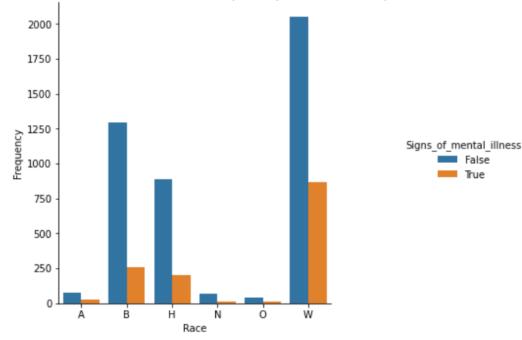




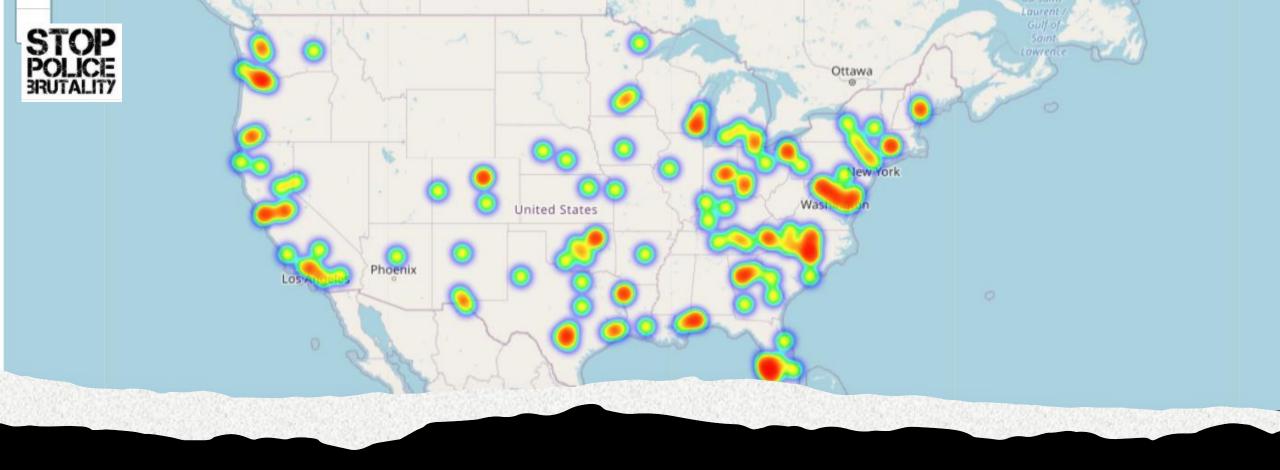
ANALISIS BIVARIADO

• Comparando la cantidad de civiles muertos por raza y signo de enfermedad mental, podemos observar que la proporción de civiles con signos de enfermedad mental al momento de ser abatidos por la policía es mucho mas significativa para la raza blanca y asiática, con casi la mitad de civiles con dichas características.

Presencia de enfermedad mental en civiles abatidos por la policía en USA, por raza, 2015-2020







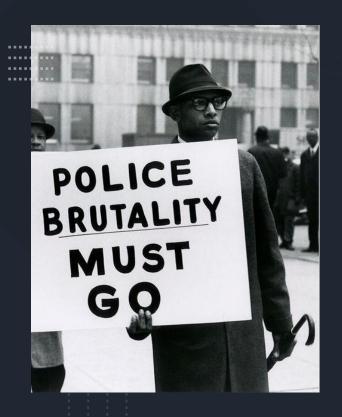
ANÁLISIS MULTIVARIADO

• En la comparativa de muertes por millón por estado, pareciera estar más ligada a cuestiones demográficas de aglomeraciones urbanas, más que por condiciones socioeconómicas, dado que por mas que en las ciudades rurales del interior se da menor nivel económico, son las ciudades costeras de mayor poder adquisitivo las que tienen mayor incidencia de fallecimientos.



ANALISIS MULTIVARIADO

• En cuanto a la correlación de la variable "sings_of_mental_illness", puede observarse que no tiene una fuerte relación, ya sea directa o inversa con ninguna de las demás variables, a diferencia de las variables socioeconómicas, que si demuestran mayor correlación entre ellas.



																	- 4 A
						Map	oa de	corre	elació	n de	varia	bles					
Armed -	1	-0.12	-0.056	-0.031	-0.037	-0.06	-0.16	0.015	-0.09	-0.019	-0.045	0.037	-0.014	-0.008	0.022	0.012	-0.0065
Age -	-0.12	1	-0.005	0.23	0.03	0.11	0.11	-0.051	0.012	-0.057	0.032	-0.062	-0.035	-0.019	0.016	-0.063	0.027
Gender -	-0.056	-0.005	1	-0.052	0.022	-0.059	0.012	0.0009	0.065	0.017	0.0024	-0.026	-0.00061	0.0098	-0.011	0.0021	-0.017
Race -	-0.031	0.23	-0.052	1	0.074	0.14	0.027	-0.09	0.022	-0.05	-0.02	-0.17	-0.13	-0.059	0.0068	-0.11	0.07
State_ID -	-0.037	0.03	0.022	0.074	1	0.034	0.02	-0.026	-0.042	0.0081	0.014	-0.29	-0.00074	-0.14	-0.097	-0.15	0.16
Signs_of_mental_illness -	-0.06	0.11	-0.059	0.14	0.034	1	0.1	0.053	0.023	-0.15	-0.044	0.009	-0.00028	-0.0052	-0.0096	0.015	-0.028
Flee -	-0.16	0.11	0.012	0.027	0.02	0.1	1	0.021	0.043	-0.056	0.044	0.0058	0.025	0.035	-0.011	0.0027	-0.0057
body_camera -	0.015	-0.051	0.0009	-0.09	-0.026	0.053	0.021	1	-0.067	0.02	0.07	0.03	0.079	0.061	-0.067	0.075	-0.032
Official_Disposition -	-0.09	0.012	0.065	0.022	-0.042	0.023	0.043	-0.067	1	0.04	0.015	0.00069	0.0028	0.014	-0.011	0.015	-0.02
Encounter Type -	-0.019	-0.057	0.017	-0.05	0.0081	-0.15	-0.056	0.02	0.04	1	0.48	-0.016	0.15	0.0024	-0.26	0.088	-0.027
Year -	-0.045	0.032	0.0024	-0.02	0.014	-0.044	0.044	0.07	0.015	0.48	1	0.024	0.43	0.24	-0.45	0.28	-0.046
GDP Millions -	0.037	-0.062	-0.026	-0.17	-0.29	0.009	0.0058	0.03	0.00069	-0.016	0.024	1	0.43	0.21	-0.029	0.47	-0.4
GDP PerCapita -	-0.014	-0.035	-0.00061	-0.13	-0.00074	-0.00028	0.025	0.079	0.0028	0.15	0.43	0.43	1	0.26	-0.44	0.73	-0.44
Unemployment Rate -	-0.008	-0.019	0.0098	-0.059	-0.14	-0.0052	0.035	0.061	0.014	0.0024	0.24	0.21	0.26	1	0.17	0.13	-0.19
Poverty Percent -	0.022	0.016	-0.011	0.0068	-0.097	-0.0096	-0.011	-0.067	-0.011	-0.26	-0.45	-0.029	-0.44	0.17	1	-0.6	0.23
Median household Income -	0.012	-0.063	0.0021	-0.11	-0.15	0.015	0.0027	0.075	0.015	0.088	0.28	0.47	0.73	0.13	-0.6	1	-0.54
Party -		0.027	-0.017	0.07	0.16	-0.028	-0.0057	-0.032	-0.02	-0.027	-0.046	-0.4	-0.44	-0.19	0.23	-0.54	1
				-				-	-				,				
	Armed	Age	Sender	Race	State_ID	mental_illness	Flee	oody_camera	Disposition	Type	Year	Millions	3DP_PerCapita	mployment_Rate	Poverty_Percent	Income	Party
	4		ď		SS	[E]		, ca	odsi	Encounter		3DP_Mi	Per .	nent	y_Pe		
						men		8	a E	noou		8	윤'	oloyn	vert	ousehold_	
						Jo.			ij.	Ξ				emp	2	sno	



ALGORITMOS ELEGIDOS gbm (-mor-b)

OTHER! S-THMBY! >>

ay:block;position:absolu

y:1; *top:-2px; *left:-5px

-4px\0/;left:-6px\0/;ric

display:inline-block;fc

:line-height:27px;padd

:display:block;text-de

index:1000).gbts(*di

list-style:none;

DECISION TREE

- ☐ Max Depth: 20
- \Box Test = 0.3 / Train = 0.7
- ☐ Random State = 42
- ☐ Criterion = 'Gini'

XGBOOST CLASSIFIER

- ☐ Learning Rate = 0.01
- \Box Test = 0.3 / Train = 0.7
- ☐ Random State = 42
- ☐ n_estimators= 20
- ☐ Seed = 42





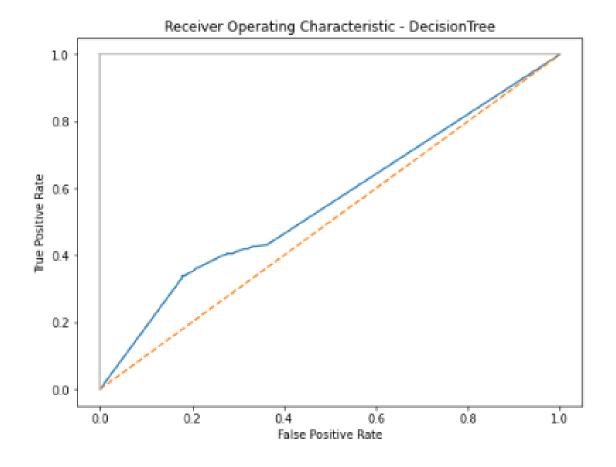
	precision	recall	f1-score	support
0 1	0.79 0.38	0.82 0.34	0.81 0.36	1316 422
accuracy macro avg weighted avg	0.58 0.69	0.58 0.70	0.70 0.58 0.70	1738 1738 1738

- Como puede observarse, el modelo presenta una significativamente mayor precisión para detectar casos "False" en cuanto a la variable de enfermedad mental de las victimas. Partiendo de métricas de infalibilidad (F1) de 0.81 para sin enfermedad mental y 0.36 con enfermedad mental. Las métricas de precisión y recall, en el mismo sentido, otorgan mayor predictibilidad cuando la variable en cuestión es false. De hecho, se dan mejores resultados para "false" partiendo de un caso real (0.82) que deduciendo a partir de las variables independientes (0.79). Para casos "true", se da a la inversa.
- En cuanto al accuracy, la precisión del modelo es del 70%.
- Esto pudiera significar que el modelo puede ser efectivo para determinar, una vez que haya fallecido un civil, si no se conoce el dato concreto, las probabilidades de que no haya sufrido una enfermedad mental en el momento de ser abatido por la policía, pero no para poder predecir precondiciones de enfermedades mentales a los efectos de diseñar políticas públicas.



ALGORITMO: DECISION TREE – CURVA ROC

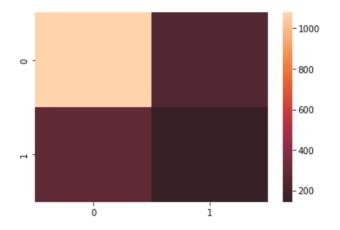
• Calculando la Curva ROC y su AUC score, podemos determinar que el modelo tiene un 55% de probabilidad de que distinga entre presencia o ausencia de enfermedad mental de las víctimas civiles.





ALGORITMO: DECISION TREE AJUSTE 10 VARIABLES

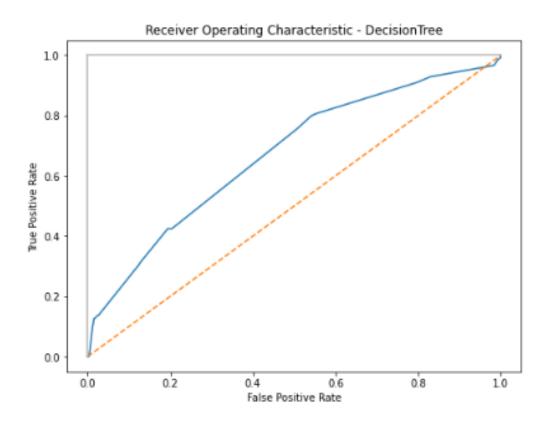
• Al ajustarse por las 10 principales variables en preponderancia, se puede obtener un "accuracy" mayor, del 77%, así como también una mejoría en la precisión de F1 score para los casos negativos (0.87), sin embargo, empeoran aún más la precisión para detectar casos verdaderos de presencia de enfermedad mental, reforzando lo mencionado anteriormente.



Age	0.214327
Year	0.086769
Unemployment_Rate	0.070584
GDP_Millions	0.069524
Poverty_Percent	0.055770
Flee_Not fleeing	0.053569
GDP_PerCapita	0.050817
Median_household_Income	0.044086
State_ID	0.041067
Encounter_Type_Mental Health	0.037681

	precision	recall	f1-score	support
0 1	0.78 0.63	0.97 0.14	0.87 0.23	1316 422
accuracy macro avg weighted avg	0.70 0.74	0.56 0.77	0.77 0.55 0.71	1738 1738 1738





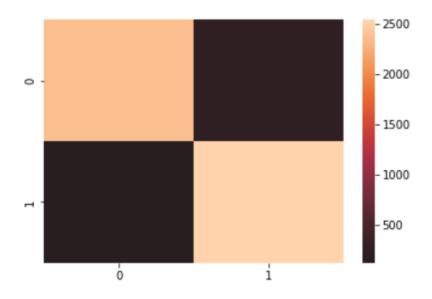
ALGORITMO: DECISION TREE - CURVA ROC - AJUSTADO

• Calculando la Curva ROC y su AUC score, podemos determinar que el modelo tiene un 67% de probabilidad de que distinga entre presencia o ausencia de enfermedad mental de las víctimas civiles.



ALGORITMO: XGBOOST CLASSIFIER

• Como puede observarse, las métricas del XBOOSTER presentan ratios más equilibrados, entre la predicción de positivos y negativos en cuanto a presencia de indicadores de enfermedad mental en víctimas civiles abatidas por la policía. Sobre todo, cabe destacar que la métrica de detección de casos positivos (recall = 0.95) es significativamente superior a la obtenida por el algoritmo decision tree, inclusive con parámetros ajustados de este último.



	precision	recall	f1-score	support
False True	0.95 0.91	0.90 0.95	0.92 0.93	2634 2666
accuracy macro avg weighted avg	0.93 0.93	0.93 0.93	0.93 0.93 0.93	5300 5300 5300



MÉTRICAS FINALES DEL MODELO OPTIMIZADO

• En cuanto a la comparativa final de los dos modelos desarrollados, notamos significativamente más preciso el XGBOOST para predecir si un civil abatido por la policía presenta indicios de enfermedad mental en EEUU. Solamente se observa como más preciso al Decision tree en el indicador de "precisión", pero haciendo referencia a la positividad de predicción de casos negativos, cuando en realidad la motivación primigenia del análisis implica detectar positivos verdaderos para poder tomar acciones en formas de políticas públicas y prevenir la ocurrencia de estas muertes evitables.

Métricas DecisionTree:

• Accuracy: 0.7710

Precisión: 0.9734

• Sensibilidad: 0.7792

Especificidad: 0.6277

F1 score: 0.8655

Métricas XGBoost con los mejores Hiperparámetros:

Accuracy: 0.9226

Precisión: 0.9055

Sensibilidad: 0.9368

Especificidad: 0.9096

F1 score: 0.9208

FUTURAS LÍNEAS

Como posibilidad de mejora creemos que ahondar en mayor cantidad de modelos pudiera abordar a conclusions aún mas precisas, así como también incluir datos de otros años o de casos donde el encuentro entre el civil y la policía no terminó con la muerte del civil, ampliaría la posibilidad de identificar las situaciones y las características donde se la vida humana es salvada.

