

# PREDICCIÓN DE VIOLENCIA DOMÉSTICA EN COLOMBIA

Valentina Escobar Bueno, Daniel Eduardo González Sánchez, Ana Gabriela Atuesta Solano  
22971 - Inteligencia Artificial I - Grupo H1  
Escuela de Ingeniería de Sistemas e Informática



## Resumen

Desde hace un tiempo la violencia en Colombia a dejado de ser un tabú, y ha pasado a convertirse en una costumbre, lo cual, claramente, afecta a toda la población nacional, ¿pero de donde nace esta problemática?

Analizando un dataset con datos de violencia doméstica en la ultima década se buscó responder a esta pregunta, determinando los momentos del año con mayor índice de violencia en los hogares colombianos.

## Introducción

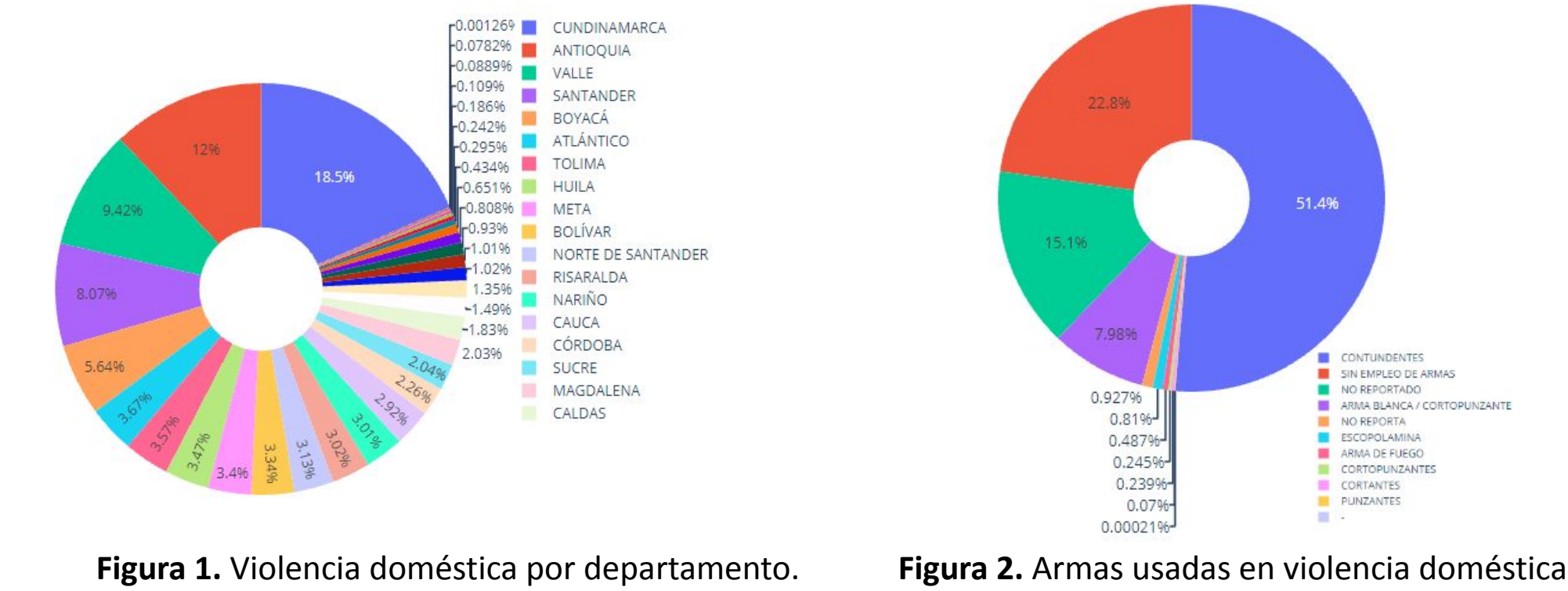
La población de Colombia ha sido sometido desde hace un tiempo a una oleada de violencia sin frenos, hasta el punto de convertirse en uno de los países con más masacres anuales en el mundo. Aún así, los colombianos se han caracterizado por ser muy animados y optimistas, en especial en épocas festivas en el país, pero entonces ¿donde nace el problema?. Desde hace un tiempo se ha motivado a la población colombiana a denunciar casos de violencia doméstica que vivan o conozcan. Esto ha generado que se pueda tener un control de la crímenes violentos leves, inpidiendo futuros crímenes graves, como un homicidio. Aún así, las cifras no dejan de ser altas, y las nuevas generaciones de colombianos crecen a diario con ejemplos poco provechosos desde el hogar. Analizando los datos obtenidos de las denuncias por violencia doméstica realizadas en la ultima decada se espera concretar la época del año con el mayor indice de esta en el pais, con el fin de prever estos crímenes y esperar generar un cambio en los indices de violencia colombiana.

## Proceso y método

El proyecto se basa en un Dataset proporcionado por la Policía Nacional de Colombia que contiene ocho columnas de información sobre la violencia doméstica en Colombia; para realizar una predicción de la época del año con más incidencia en casos de violencia se implementó un modelo de regresión, donde fue necesario eliminar columnas que no eran de nuestro interés, por ejemplo; municipio y código DANE. Además, se reemplazó los datos de las columnas por datos numéricos para poder realizar su análisis.

Se realizó en un entorno de notebook con Python 3 y se hizo uso de las librerías: Pandas, Numpy, Sklearn; que nos permite el uso de los algoritmos:

- DecisionTreeRegressor
- RandomForestRegressor



## Resultados

La violencia doméstica en Colombia es muy recurrente día a día, pues los días del año donde se presentan estos incidentes tienen una cantidad muy similar entre ellos.

Se ve una pequeña correlación entre el grupo etario y el género de las víctimas, que en su mayoría son mujeres adultas, siendo estás las más afectadas en Colombia. Pues en la mayoría de los datos no se ve una gran correlación debido a la información que proporciona este dataset.

Respecto al desarrollo de los algoritmos en el dataset son algo baldíos pues la relación entre los datos es muy baja, estando estos más relacionados con series de tiempo, por lo que habría que hacer una reestructuración para así poder obtener un mejor desarrollo y resultado esperado.

Gráfico 1. Correlación

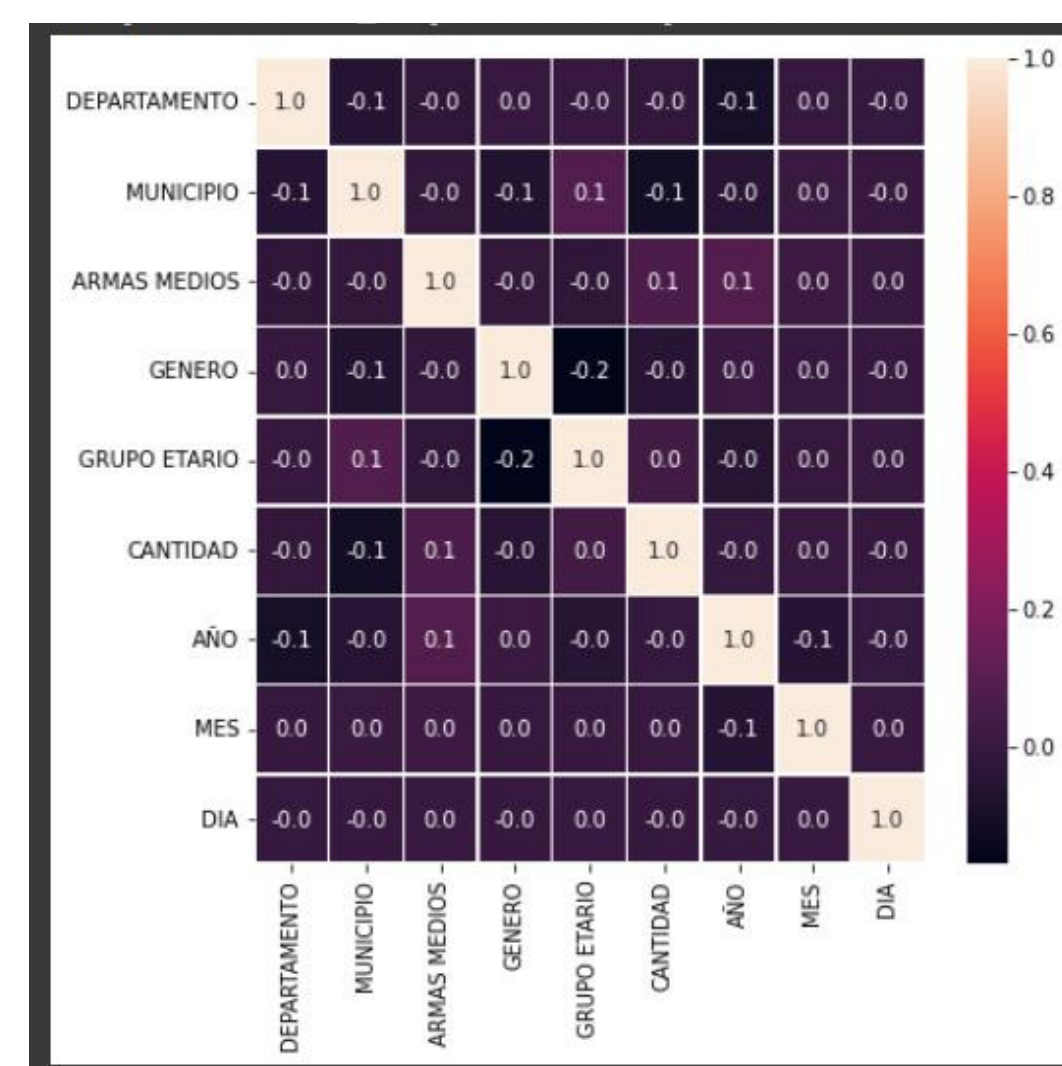
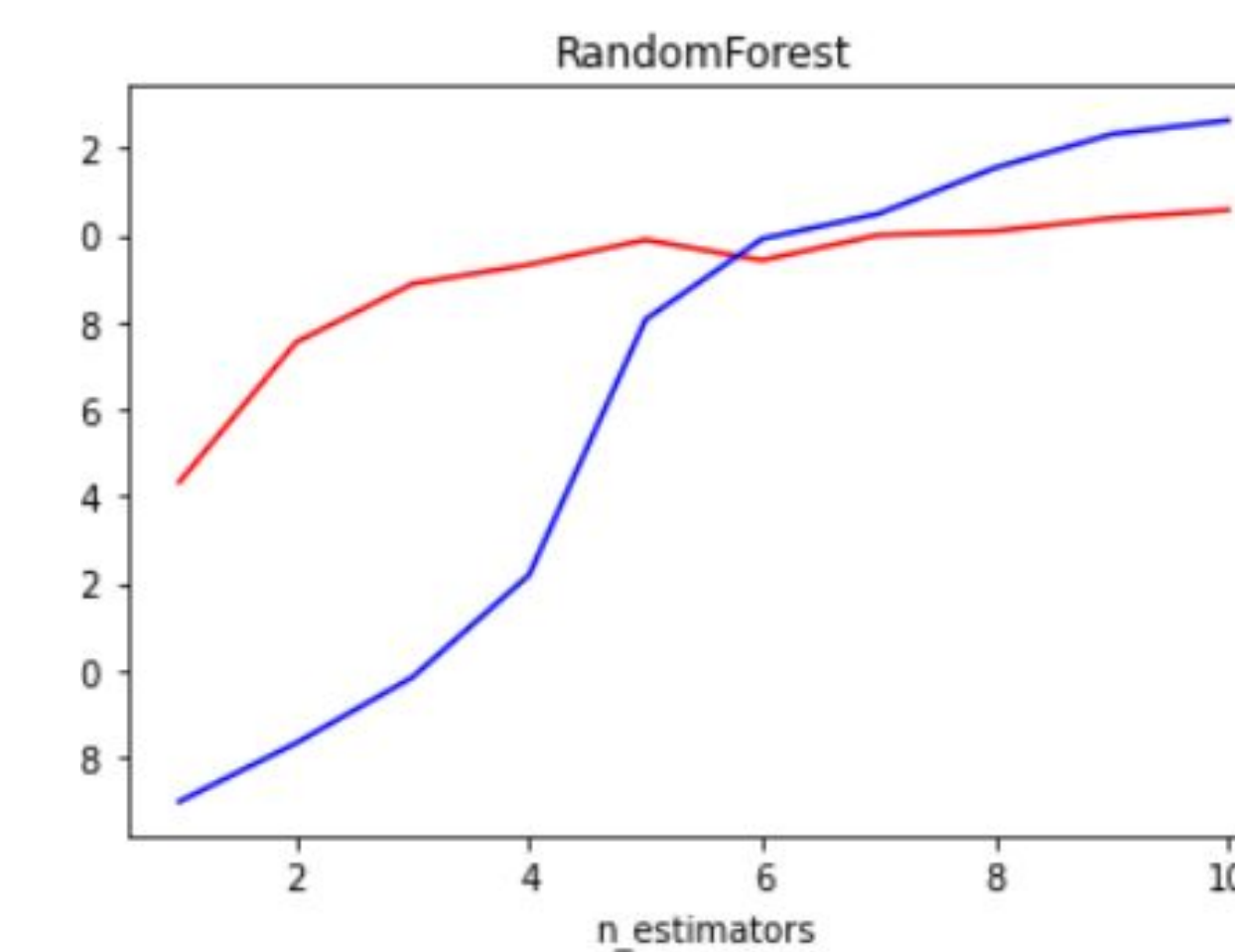


Gráfico 2. RF



## Conclusiones

- 1) Como algo ya predecible las mujeres son las más afectadas y violentadas en el entorno doméstico en Colombia.
- 2) El arma más usado en este tipo de incidentes son las del tipo “CONTUNDENTES”.
- 3) Siendo Cundinamarca el departamento con más casos registrados en Colombia.
- 4) Se puede observar como estos incidentes pasan con una recurrencia similar a lo largo de cada día del año.
- 5) Baja correlación entre los datos tratados.

## Trabajo Futuro

Al tener los datos obtenidos en esta investigación sería conveniente explorar el campo de las armas utilizadas en los casos, usando un clasificador, para así lograr un reconocimiento preciso de estas, que permita una regulación más estricta a nivel nacional de los elementos identificados.

## Información de contacto

Valentina Escobar Bueno, Email: [nenitq@gmail.com](mailto:nenitq@gmail.com)  
Daniel Eduardo González Sánchez, Email: [degosanoo@gmail.com](mailto:degosanoo@gmail.com)  
Ana Gabriela Atuesta Solano, Email: [gabrielaatuesta@gmail.com](mailto:gabrielaatuesta@gmail.com)

## Referencias Bibliográficas

1. Lam Díaz, Rosa María. (2016). La redacción de un artículo científico. Revista Cubana de Hematología, Inmunología y Hemoterapia, 32(1), 57-69. Recuperado en 09 de agosto de 2020, de [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S0864-02892016000100006&lng=es&tlng=es](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-02892016000100006&lng=es&tlng=es).
2. Regresión Lineal con Python - Joaquin Amat Rodrigo, octubre 2020, de <https://www.cienciadedatos.net/documentos/py10-regresion-lineal-python.html>
3. <https://cran.r-project.org/web/packages/randomForest/index.html>
4. [https://www.stat.berkeley.edu/~breiman/RandomForests/cc\\_software.htm](https://www.stat.berkeley.edu/~breiman/RandomForests/cc_software.htm)