

Predicción del día, hora y municipio de un delito sexual

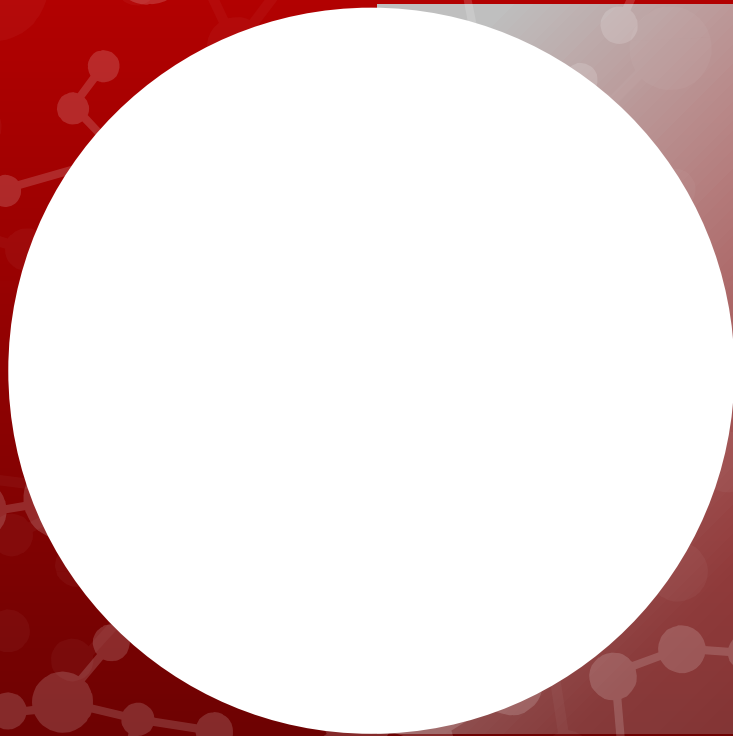
EDUARD ALFONSO CABALLERO PONGUTA - 2151099

SERGIO ANDRES DULCEY LIZARAZO -2151830

JUAN PABLO VIVIESCAS ROMERO - 2152499

Contenido

1. Motivación
2. Objetivos
3. Documentación del problema
4. Estado del arte
5. Dataset
6. Implementación
 - 6.1 Modelo
 - 6.2 Resultados
7. Conclusiones
8. Bibliografía



1. Motivación

Motivación

- I. Los casos de crímenes sexuales van en aumento, y proporcionalmente la inseguridad de algunos sitios también han incrementado, pero y si hubiera una forma de predecir la hora y el lugar donde un posible delito sexual, con ello se podría reforzar la seguridad en dicho lugar y con ello el índice de delitos sexuales disminuye notablemente y así mejorando la calidad de vida de los ciudadanos.



2. Objetivos

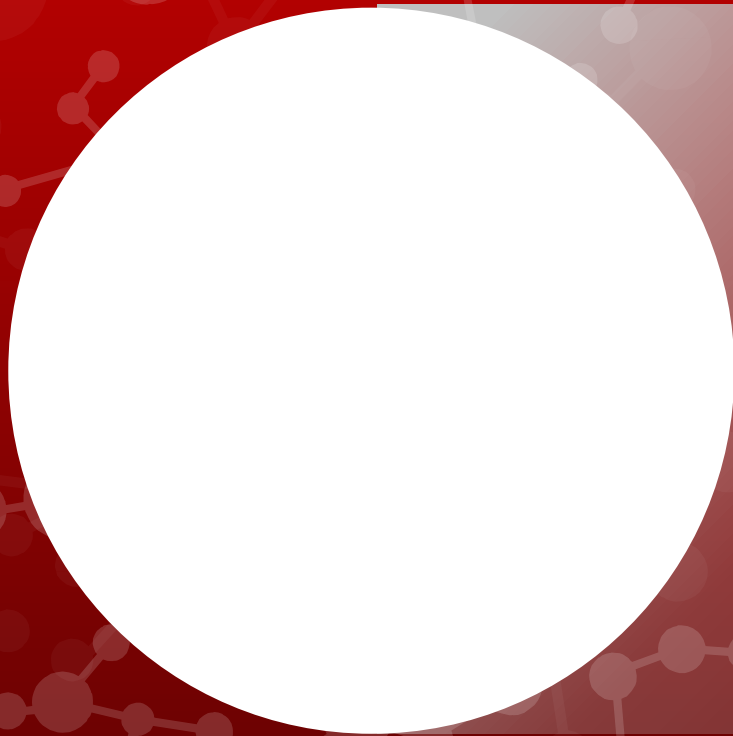
Objetivos

Objetivo general:

- Realizar una predicción del día de la semana y la hora en que ocurre un posible delito sexual, por medio de un análisis de datos proporcionados por la policía nacional de Colombia

Objetivos específicos:

- Implementar diferentes modelos, tanto clasificadores como regresores para las características [Dia(type: string) y Hora(type: timestamp)]
- Determinar los modelos que permitan una mejor predicción del municipio, día de la semana y la hora a partir de las demás características



3. Documentación del problema

Documentación del problema

Los delitos sexuales ha sido uno de los grandes problemas que han asolado al país en los últimos años, pero y si hubiera una forma de predecirlo, así se podría tratar de disminuir los factores que lo causan y poder evitar que se genere más casos, lo cual se va a investigar en este proyecto.

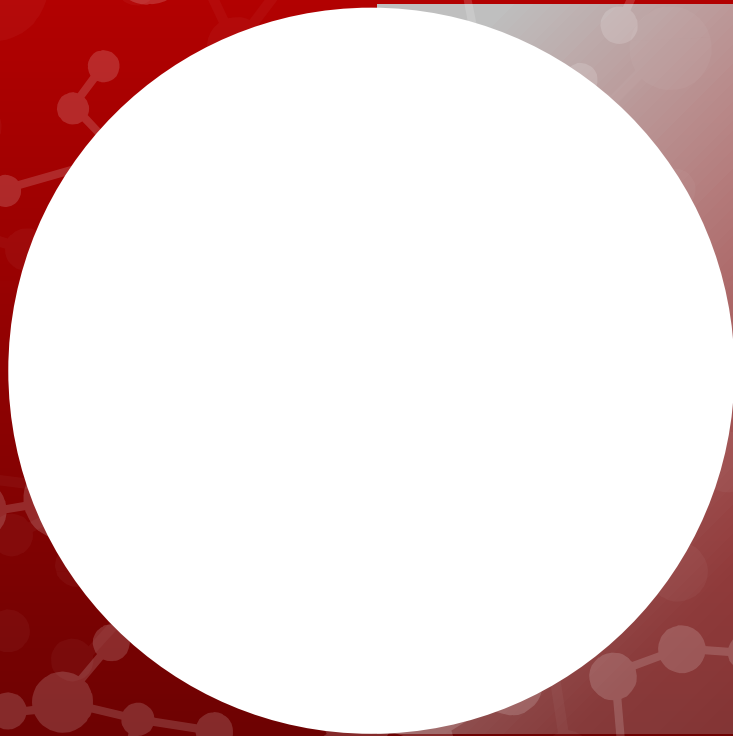
Con la implementación de modelos de predicción, tanto clasificadores como regresores para el dataset proporcionado por la policía, y con ellos hallar una pauta que permita predecir la hora, fecha y lugar de un posible delito sexual, con el fin de mejorar la seguridad de esos sitios, mejorando la calidad de vida de los ciudadanos.



4. Estado del Arte

Estado del arte

Se han realizado varios intentos en tratar de predecir un delito sexual en Colombia, pero más enfocado en que si un agresor reincide a cometer de nuevo el delito, ya que un estudio hecho por la policía nacional colombiana determinó que la gran mayoría de los agresores volvieron a cometer delitos sexuales, y en especial contra menores de edad, por ello endurecieron más el castigo a los agresores.



5. Dataset

Dataset sin tratar

Delitos sexuales 2019

Delitos sexuales (del 01 de enero al 31 de diciembre año 2019).

Fuente: Diiin Policía Nacional. Datos extraídos el día 12 de enero del año 2020 a las 14:00 horas.



Encontrar en este conjunto de

Más vistas

Filtrar

Visualizar

Exportar

Debatir

Incrustar

Acerca de

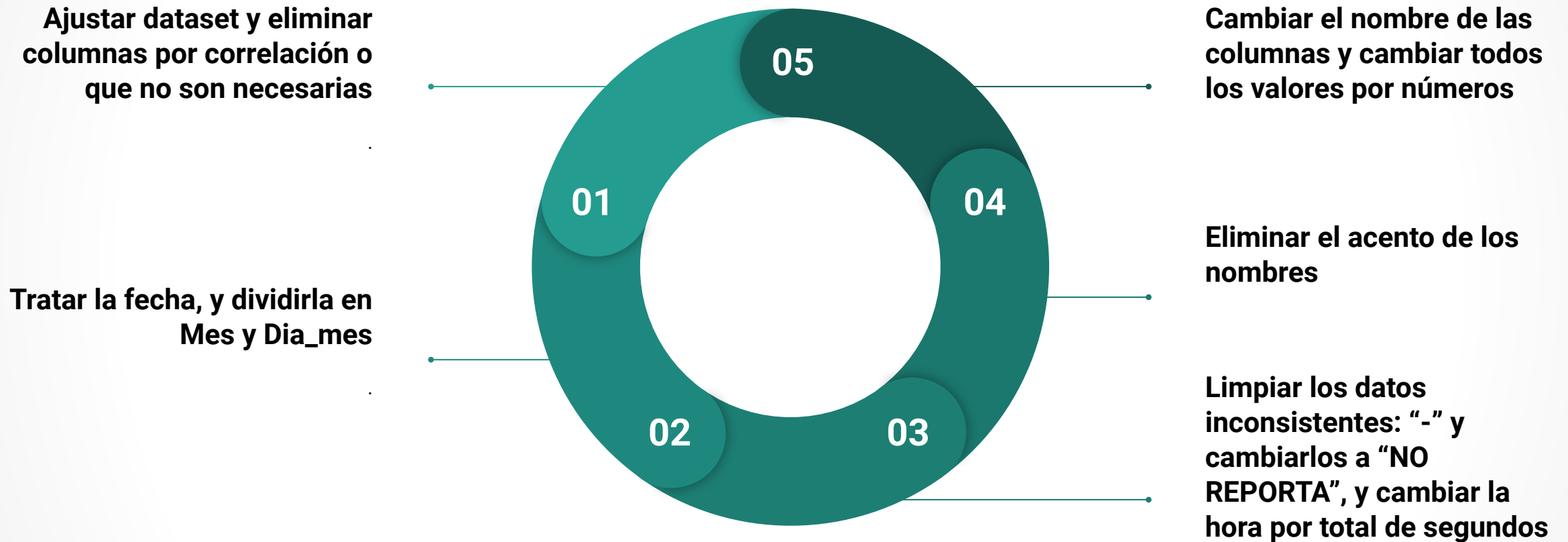
MINISTE...	Unname...	Unname...	Unname...	Unname...	Unname...	Unname...	Unname...	Unname...	Unname...	Unname...	Unname...	Unname...	Unname...
2019 Jan 01 1...	AMAZONAS	LETICIA (CT)	ARTÍCULO 20...	Martes	18:00:00	NO REPORTA...	RURAL	CASAS DE AP...	SIN EMPLEO ...	A PIE	A PIE	7	FEMEN
2019 Jan 01 1...	AMAZONAS	LETICIA (CT)	ARTÍCULO 20...	Martes	18:00:00	NO REPORTA...	RURAL	CASAS DE AP...	SIN EMPLEO ...	A PIE	A PIE	21	FEMEN
2019 Jan 01 1...	ANTIOQUIA	ABRIAQUÍ	ARTÍCULO 20...	Martes	0:00:00	SAN JOSE	RURAL	FINCAS Y SIM...	SIN EMPLEO ...	A PIE	A PIE	16	FEMEN
2019 Jan 01 1...	ANTIOQUIA	AMAGÁ	ARTÍCULO 20...	Martes	10:00:00	LA GUALI	RURAL	FINCAS Y SIM...	SIN EMPLEO ...	A PIE	A PIE	4	FEMEN
2019 Jan 01 1...	ANTIOQUIA	AMAGÁ	ARTÍCULO 20...	Martes	10:00:00	LA GUALI	RURAL	FINCAS Y SIM...	SIN EMPLEO ...	A PIE	A PIE	8	FEMEN
2019 Jan 01 1...	ANTIOQUIA	AMAGÁ	ARTÍCULO 20...	Martes	10:00:00	LA GUALI	RURAL	FINCAS Y SIM...	SIN EMPLEO ...	A PIE	A PIE	28	FEMEN

Fuente: <https://www.datos.gov.co/Seguridad-y-Defensa/Delitos-sexuales-2019/tdf6-87g7/data>

Tamaño: 140.400

Columnas: 21

Tratamiento de datos:



Dataset

Variables	Descripción
Departamento	Departamento donde sucedio el delito. (formato: string)
Municipio	Municipio donde sucedio el delito. (formato: string)
Día	Nombre del día en el que sucedio el delito. (formato: string)
Hora	Hora en el que sucedio el delito. (formato: timestamp)
Barrio	Barrio donde ocurrio la agresión. (formato: int)
Zona	Distrito en que ocurrio la agresión. (formato: int)
Clase_de_sitio	Categoría del sitio donde ocurre la agresión. (formato: int)
Arma_empleada	Objeto empleado por el agresor para intimidar a la víctima. (formato: int)
Móvil_Agresor	Tipo de desplazamiento que empleo el agresor para llegar hacia donde se encontraba la víctima (formato: int)
Móvil_Victima	Tipo de desplazamiento empleado por la víctima en el momento que ocurrió la agresión. (formato: int)
Edad	Edad de la víctima. (formato: int)
Sexo	Sexo de la víctima. (formato: int)
Estado_civil	Vínculos personales de la víctima. (formato: int)
Clase_de_empleado	Tipo de profesión que desempeña la víctima. (formato: int)
Mes	Mes en el que sucedio el delito. (formato: int)
Dia_mes	Día del mes en el que sucedio el delito. (formato: int)

Dataset tratado

	Día	Hora	Barrio	Zona	Clase_de_sitio	Arma_empleada	Móvil_Agresor	Móvil_Victima	Edad	Sexo	Estado_civil	Clase_de_empleado	Mes	Dia_mes	Mun_Dep
0	3	64800	3514	2	25	7	0	0	67	0	2	19	1	1	472
1	3	64800	3514	2	25	7	0	0	14	0	5	17	1	1	472
2	3	0	4172	2	59	7	0	0	8	0	4	15	1	1	2
3	3	36000	2719	2	59	7	0	0	34	0	4	11	1	1	31
4	3	36000	2719	2	59	7	0	0	78	0	4	11	1	1	31

Tamaño:33790

Columna: 15

Exploración de los datos

- Fecha: MAYO-1-2019 3442 Casos
- Municipios y departamentos: BOGOTA D.C- CUNDINAMARCA 5419 16%
- Genero: FEMENINO 29.003, 85.8% - MASCULINO 4707, 13.9% - NO REPORTA 81 , 0.24%
- Edad: 13 años, 3829, 11.3%

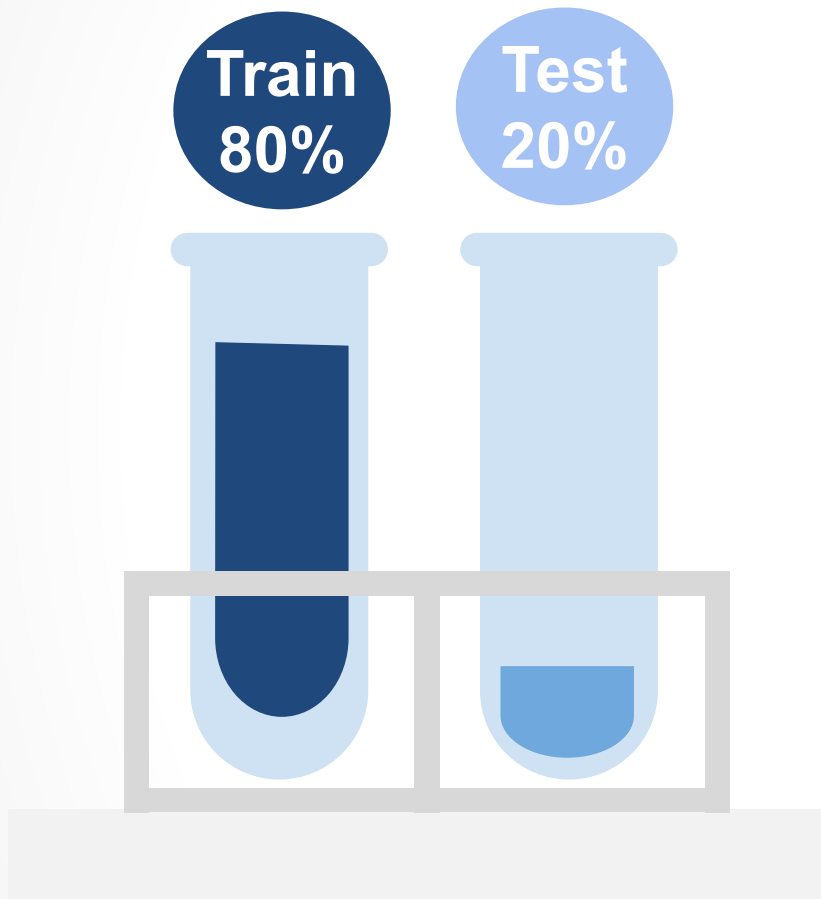


6.Implementación



6.1.Modelo

Partición de los datos



Clasificador:

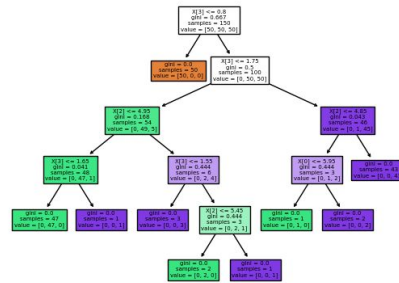
$y = [\text{Dia}, \text{Municipio}]$

Regresor:

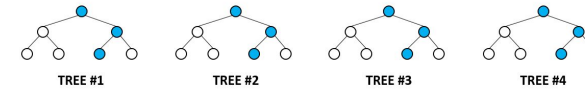
$y = [\text{Hora}]$

Clasificadores

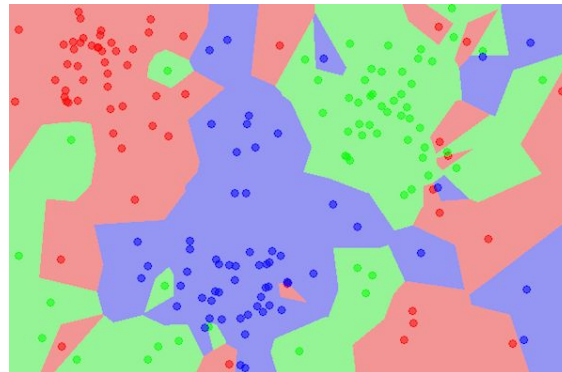
DecisionTreeClassifier



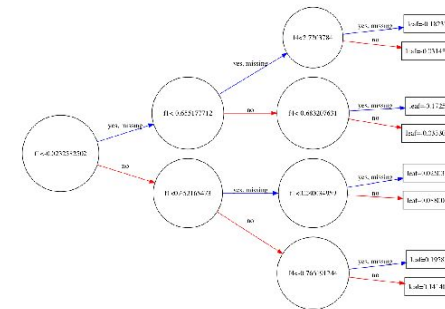
RandomForestClassifier



KNeighborsClassifier

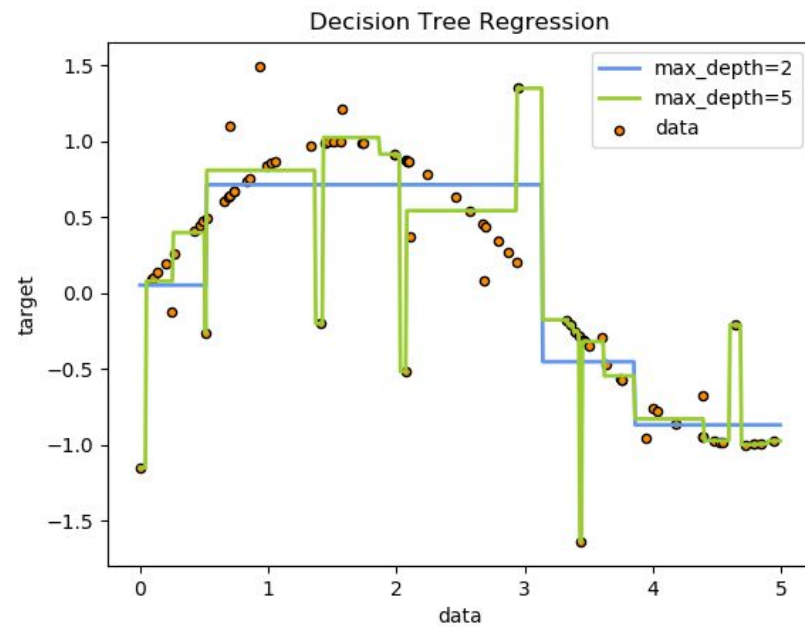


XGBClassifier

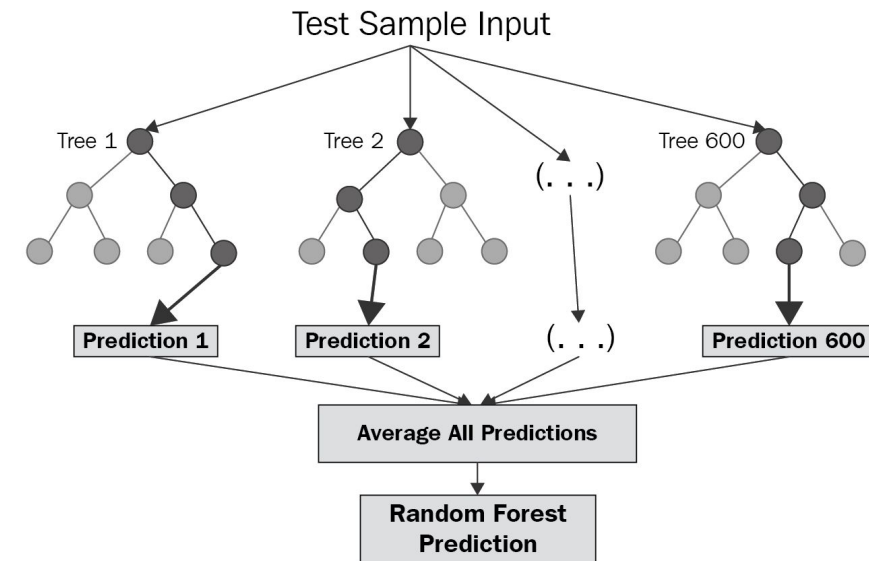


Regresores

DecisionTreeRegressor



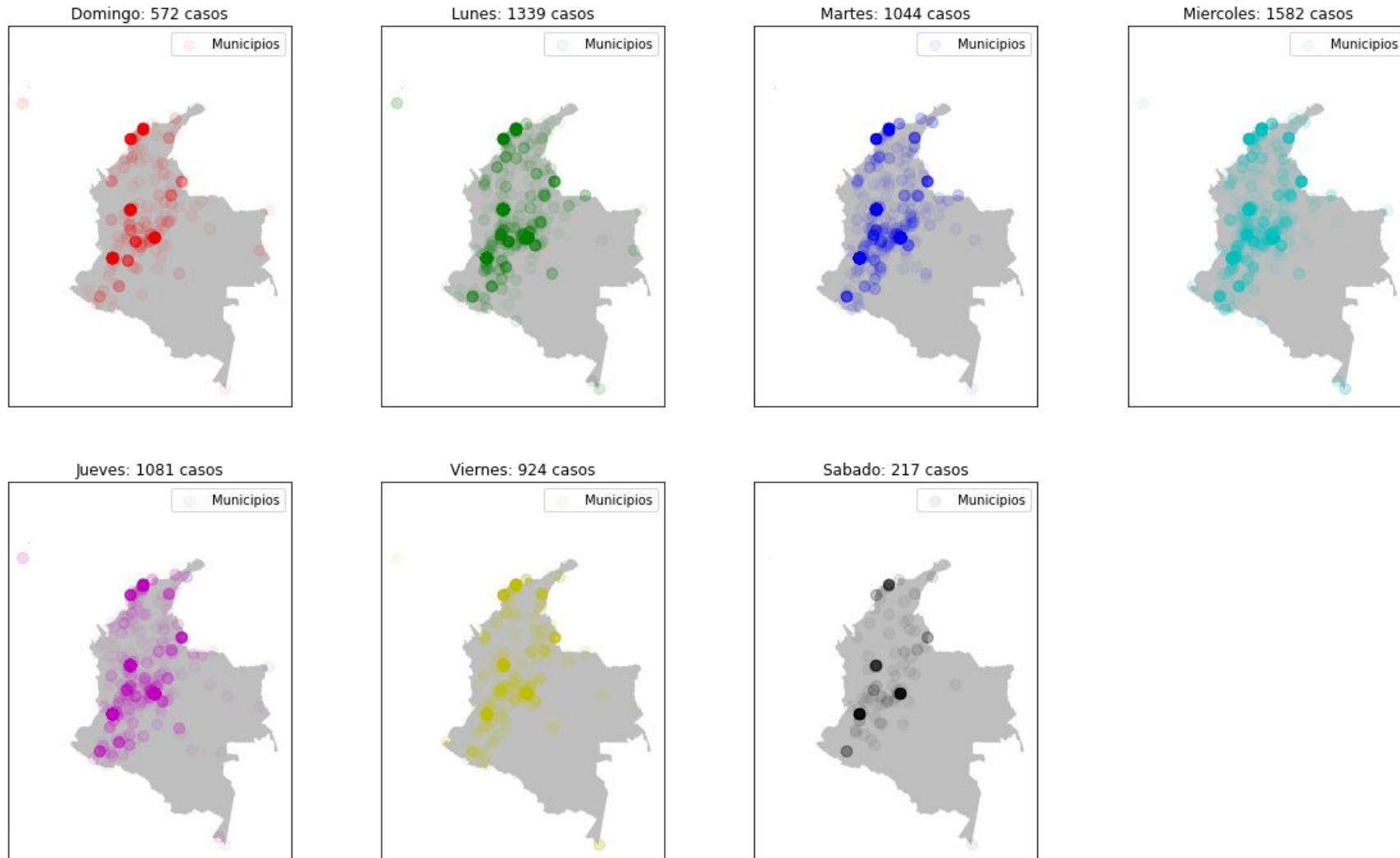
RandomForestRegressor



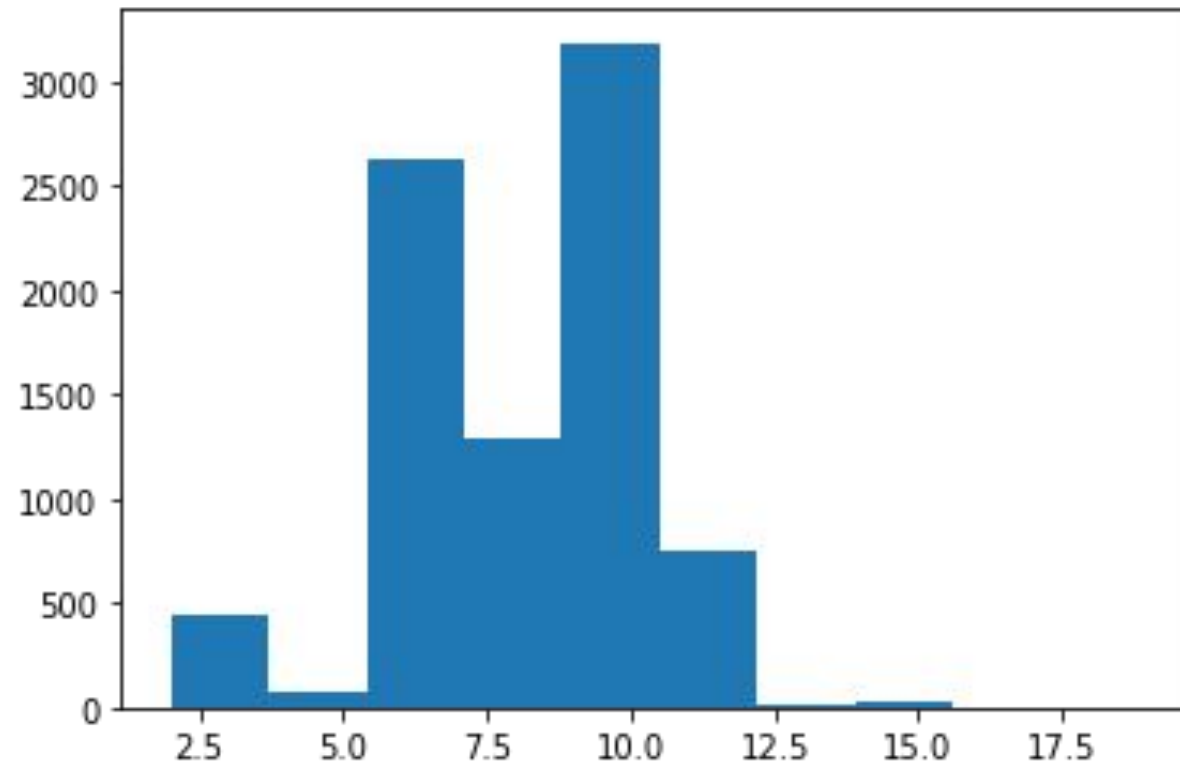


6.2.Resultados

Resultados Clasificador



Resultados Regresor





7. Conclusiones

Conclusiones

- ❖ El clasificador con mejor accuracy, y evidentemente menor RMSE, es Multi Output Classifier con XGBClassifier como parametro
- ❖ El regresor con mejor accuracy y evidentemente menor RMSE, es Decision Tree Regressor
- ❖ Con el clasificador, Se predice que el día miércoles, es el día que se produce más casos de delitos sexuales
- ❖ Con el regresor, se predice la hora donde ocurren más casos, se encuentra en el intervalo de la mañana y el medio día.

GRACIAS POR SU
ATENCIÓN

8. Referencias

- <https://www.datos.gov.co/Seguridad-y-Defensa/Delitos-sexuales-2019/tdf6-87g7>
- <https://scikit-learn.org/stable/modules/generated/sklearn.multioutput.MultiOutputClassifier.html>
- <https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeRegressor.html>
- <https://gitlab.com/bivl2ab/academico/cursos-uis/ai/ai-uis-student.git>