

```
library(dplyr)
```

```
rladies_global %>%  
  filter(city == 'Lima')
```



Machine Learning para identificar factores de influencia en Accidentes de Tránsito Fatales (ATF)

1



Cada día vemos en las noticias...

1 víctima mortal



1 víctima mortal



No hubo víctima mortal



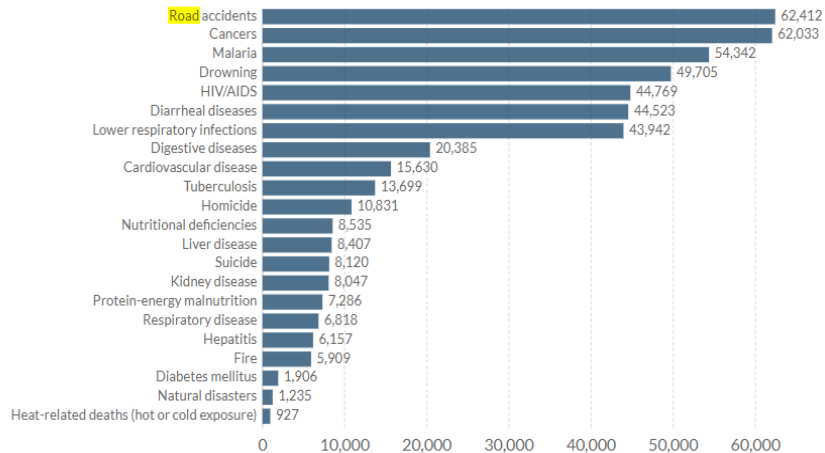
Estadísticas de Causas de muerte por grupo etario - Global



Causes of death in 5-14 year olds, World, 2017

Annual number of deaths by cause in children aged 5 to 14 years old, across both sexes. Data refers to the specific cause of death, which is distinguished from risk factors for death, such as air pollution, diet and other lifestyle factors. See sources for further details on definitions of specific cause categories. Data on deaths related to terrorism and executions are not available by age group, so have been excluded.

Our World
in Data



Source: IHME, Global Burden of Disease (GBD)

CC BY

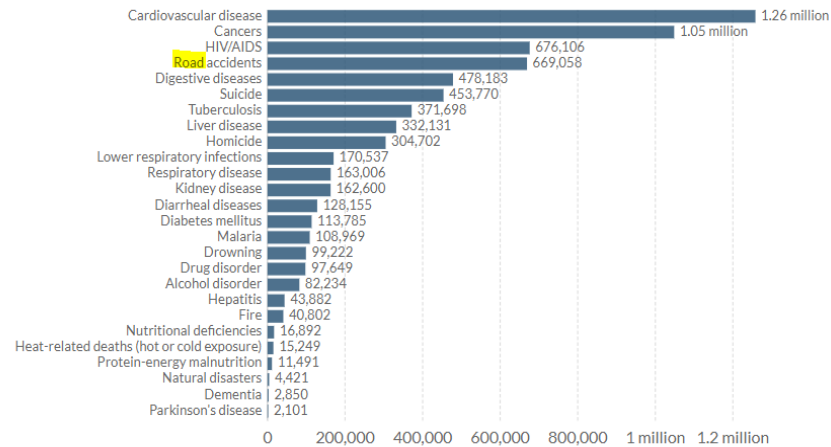
Año 2017:

Los accidentes de tránsito fueron la 1era causa de muerte en personas de 5 a 14 años

Causes of death in 15-49 year olds, World, 2017

Annual number of deaths by cause in children aged 15 to 49 years old, across both sexes. Data refers to the specific cause of death, which is distinguished from risk factors for death, such as air pollution, diet and other lifestyle factors. See sources for further details on definitions of specific cause categories. Data on deaths related to terrorism and executions are not available by age group, so have been excluded.

Our World
in Data



Source: IHME, Global Burden of Disease (GBD)

CC BY

Año 2017:

Los accidentes de tránsito fueron la 4ta causa de muerte en personas de 15 a 49 años

Estadísticas de Causas de muerte por grupo etario – Perú

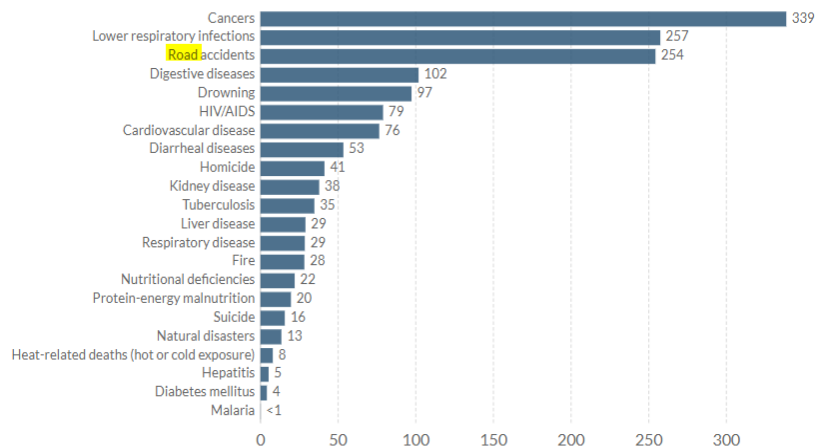


Causas de muerte en niños de 5 a 14 años, Perú, 2017

Número anual de muertes por causa en niños de 5 a 14 años, en ambos sexos. Los datos se refieren a la causa específica de muerte, que se distingue de los factores de riesgo de muerte, como la contaminación del aire, la dieta y otros factores del estilo de vida.

Consulte las fuentes para obtener más detalles sobre las definiciones de categorías de causas específicas. Los datos sobre muertes relacionadas con el terrorismo y las ejecuciones no están disponibles por grupo de edad, por lo que se han excluido.

Our World in Data



Año 2017:

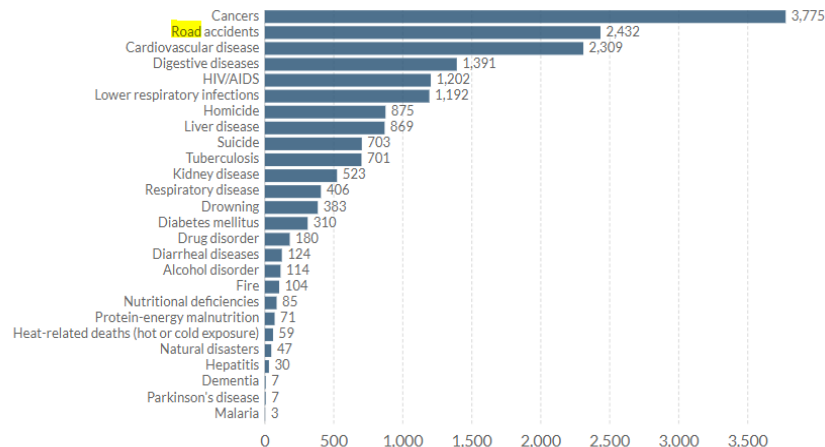
Los accidentes de tránsito fueron la 3era causa de muerte en personas de 5 a 14 años

Causas de muerte en jóvenes de 15 a 49 años, Perú, 2017

Número anual de muertes por causa en niños de 15 a 49 años, en ambos sexos. Los datos se refieren a la causa específica de muerte, que se distingue de los factores de riesgo de muerte, como la contaminación del aire, la dieta y otros factores del estilo de vida.

Consulte las fuentes para obtener más detalles sobre las definiciones de categorías de causas específicas. Los datos sobre muertes relacionadas con el terrorismo y las ejecuciones no están disponibles por grupo de edad, por lo que se han excluido.

Our World in Data



Año 2017:

Los accidentes de tránsito fueron la 2da causa de muerte en personas de 15 a 49 años

¿Qué factores influyen para que un accidente de tránsito sea fatal?



Fatal: Cuando por lo menos hay una víctima mortal como consecuencia del Accidente de Tránsito (AT).



Hipótesis

Exceso de carga
Exceso de velocidad
Vía en mal estado
Imprudencia del peatón
Mal tiempo

Falla mecánica
Cansancio del conductor
Invasión al carril
Falta de iluminación en la vía
Ebriedad del peatón

Objetivo

De investigación: Identificar los factores de influencia en ATF con fin preventivo.

De modelamiento: Construir modelos de Machine Learning que modelen la fatalidad de un AT

2



¿Dónde encuentro datos de AT?



Fuente: INEI – Censo Nacional de Comisarías – Nro 8 : Capítulo de AT.

CONSULTA POR ENCUESTA

Sírvase seleccionar Encuesta, Año y Periodo y a continuación se mostrarán todas los Módulos de la Encuesta Seleccionada. Luego proceda a descargar el módulo de su interés.

ENCUESTA CENSO NACIONAL DE COMISARIAS

AÑO 2017

Departamento: Anual

Nro	Año	Periodo	Código Encuesta	Encuesta	Código Módulo	Módulo	Ficha	Descarga
1	2017	142	595	CENSO NACIONAL DE COMISARIAS	1226	CAPÍTULO 100 - INFRAESTRUCTURA		
2	2017	142	595	CENSO NACIONAL DE COMISARIAS	1227	CAPÍTULO 200 - I - INFRAESTRUCTURA		
3	2017	142	595	CENSO NACIONAL DE COMISARIAS	1228	CAPÍTULO 200 - II - INFRAESTRUCTURA		
4	2017	142	595	CENSO NACIONAL DE COMISARIAS	1229	CAPÍTULO 300 - INFRAESTRUCTURA		
5	2017	142	595	CENSO NACIONAL DE COMISARIAS	1230	CAPÍTULO 400 - 500 - INFRAESTRUCTURA		
6	2017	142	595	CENSO NACIONAL DE COMISARIAS	1231	CAPÍTULO 300 - DETALLE - INFRAESTRUCTURA		
7	2017	142	595	CENSO NACIONAL DE COMISARIAS	1232	CAPÍTULO 600 - INFRAESTRUCTURA		
8	2017	142	595	CENSO NACIONAL DE COMISARIAS	1233	CAPÍTULO 100 - CENACOM - AT		
9	2017	142	595	CENSO NACIONAL DE COMISARIAS	1234	CAPÍTULO 200 - CENACOM - AT		

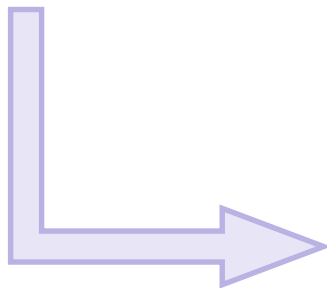
Recolección de datos



Censo a las comisarías a nivel nacional



Realizaron un
muestreo de los AT



Datos de
accidentes de
tránsito



Libros o cuadernos de ocurrencia o Sistema
de Denuncia Policial (SIDPOL)

Periodo:

Accidentes ocurridos entre
enero 2016 y junio 2017

Instrumento:

Formulario de
Accidentes de Tránsito

Datos disponibles



Ubicación
geográfica del AT

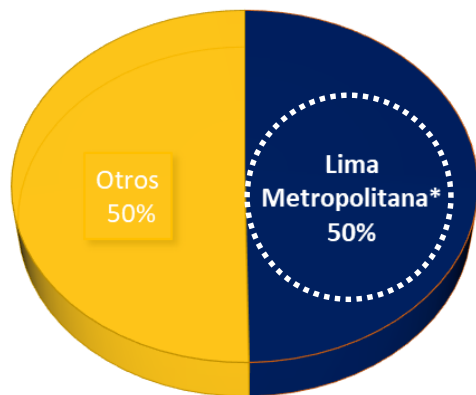
Tipo de vehículo
que participó en el
AT

Tipo de Vía en el cual
aconteció el AT

Factores asociados al
AT

Alcance del estudio

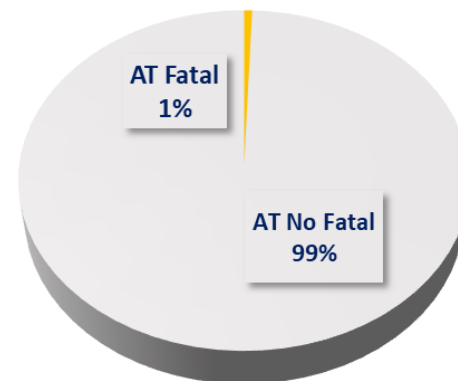
Distribución % de Accidentes de Tránsito a nivel Nacional



Pregunta 108:
**Consecuencia del
accidente:**

Fatal	1
No Fatal	2
Solo daños Materiales ...	3

Consecuencia de Accidentes de Tránsito en Lima Metropolitana



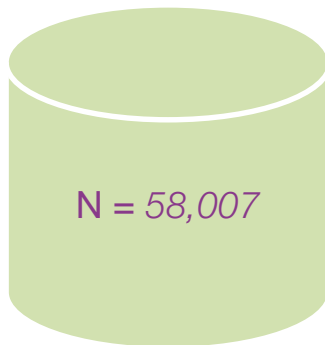
Fuente: Censo Nacional de Comisarías 2017
Total de AT: 116,659 a nivel nacional

Fuente: Censo Nacional de Comisarías 2017
AT Fatal en Lima Metropolitana: 359 casos

Datos disponibles



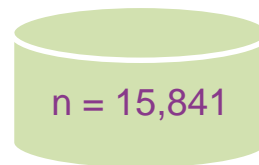
AT en Lima Metropolitana:



Fuente: Censo Nacional de Comisarías 2017

Usando los factores
de expansión

AT en la muestra:



Fuente: Censo Nacional de Comisarías 2017

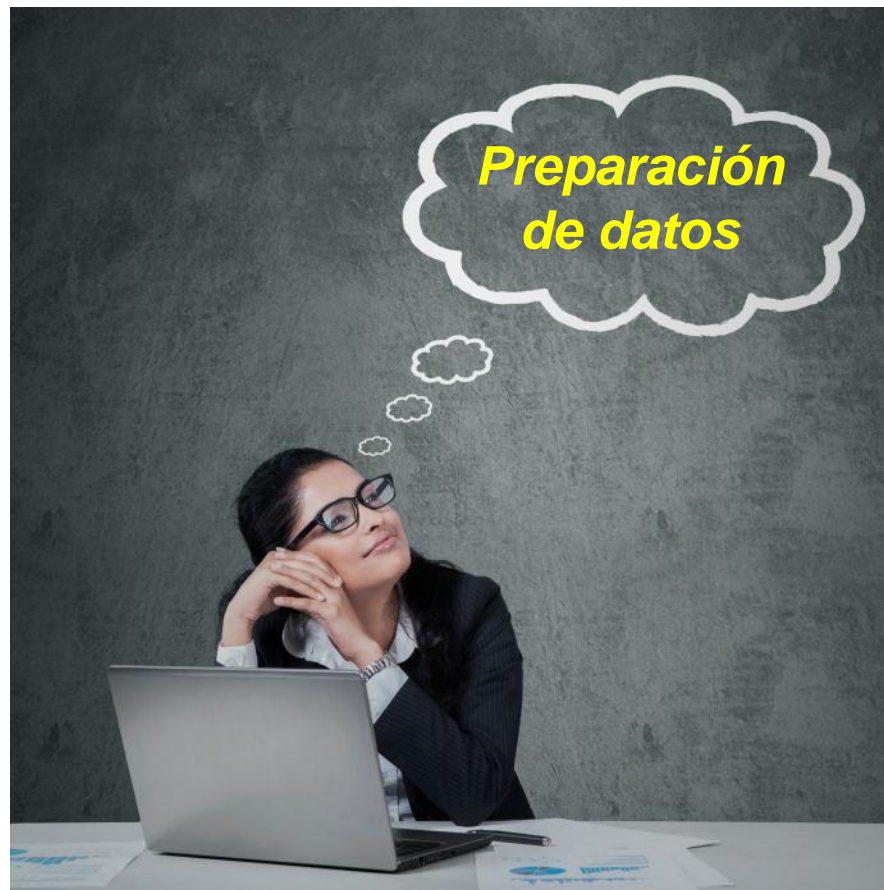
Datos en la muestra

Consecuencia	Cantidad
AT Fatal	142
AT No Fatal	15,699
Total	15,841

Muestra de
modelamiento

Nivel de
desbalance alto

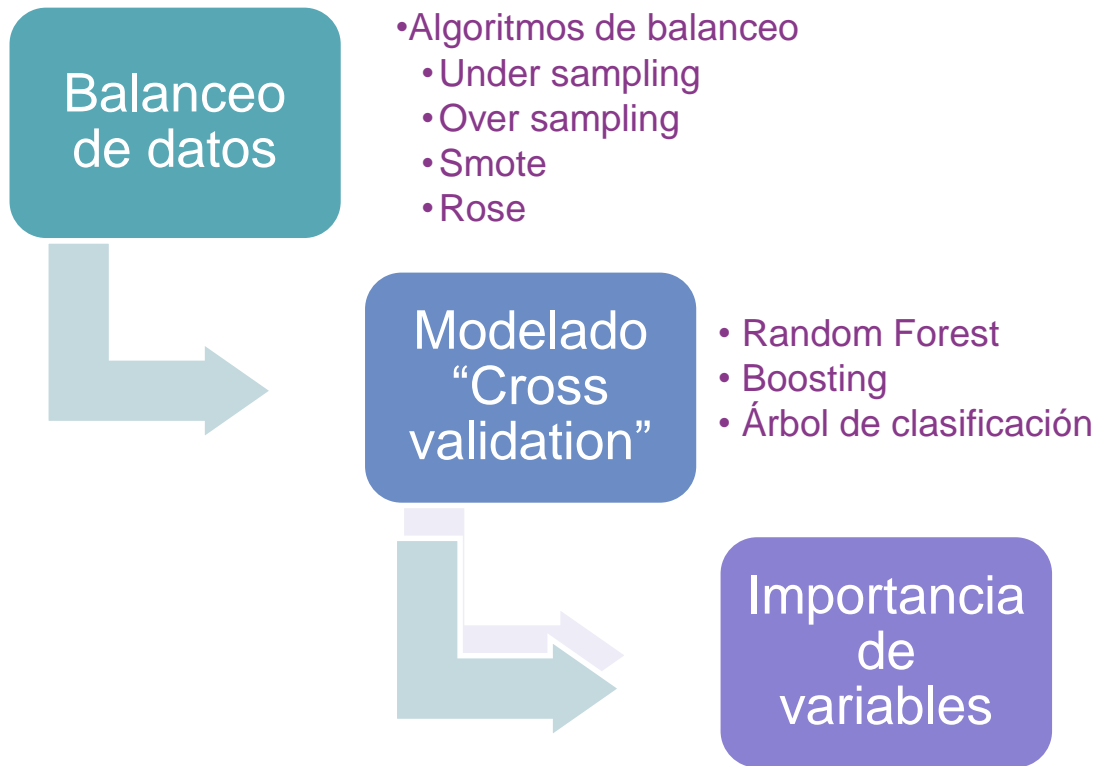
3



- ✓ Eliminar las variables que no son de Accidentes de Tránsito, sino del centro policial
- ✓ Eliminar las variables vacías
- ✓ Recodificación de nombres de variables
 - “AT106_1” por “at_veh_mayor_involucrado_automóvil”*
- ✓ Categorización de variables
 - Distritos de Lima**
 - Metropolitana en 5 Zonas**
 - “Lima Norte” “Centro”*
 - “Lima Sur” “Lima Moderna”*
 - “Lima Este”*
 - Hora de ocurrencia del AT**
 - Madrugada: “0-5”*
 - Mañana: “6-11”*
 - Tarde: “12-18”*
 - Noche: “19-23”*
- ✓ Creación de variables dummy
 - “Tipo de accidente”*
 - “Tramo de vía”*

4

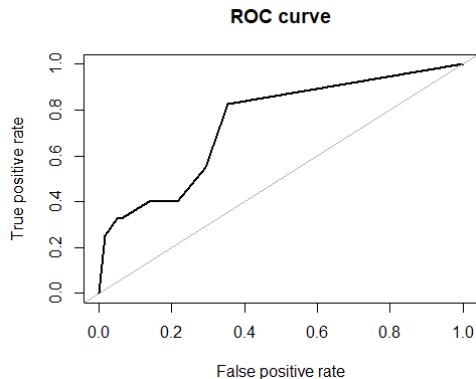




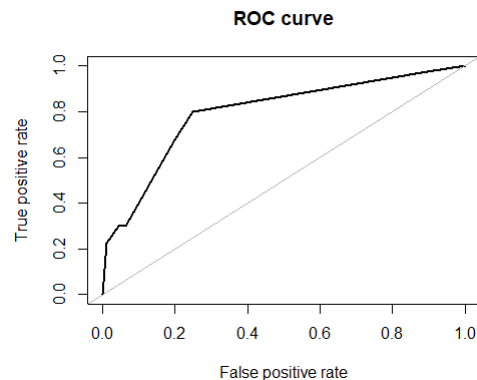
Balanceo de datos



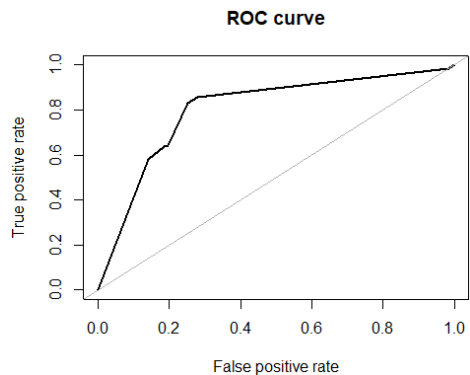
under sampling



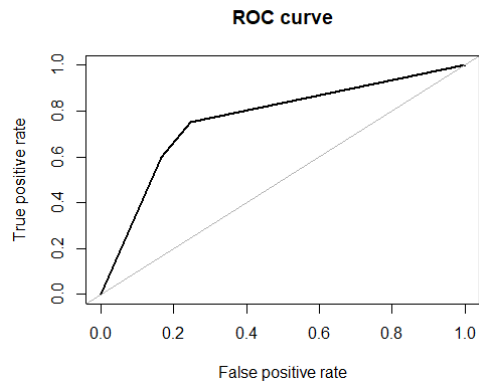
over sampling



Datos sintéticos Rose



Datos Smote



Se elige el
método de
balanceo:
Rose

Cross validation: Método k-fold



A continuación se muestra la visualización de una validación de k-fold cuando $k = 7$.

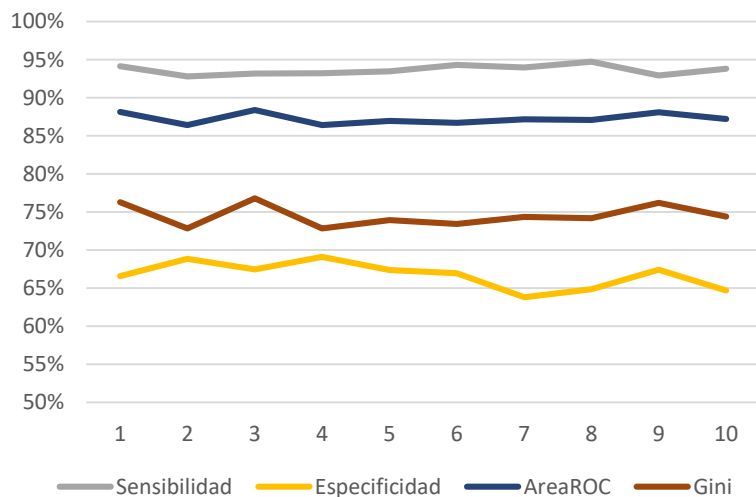


- ✓ Técnica que ayuda a reducir el sesgo de selección y reducir la varianza en el poder de predicción
- ✓ Este método nos ayuda a lograr relaciones más generalizadas.

Cross validation

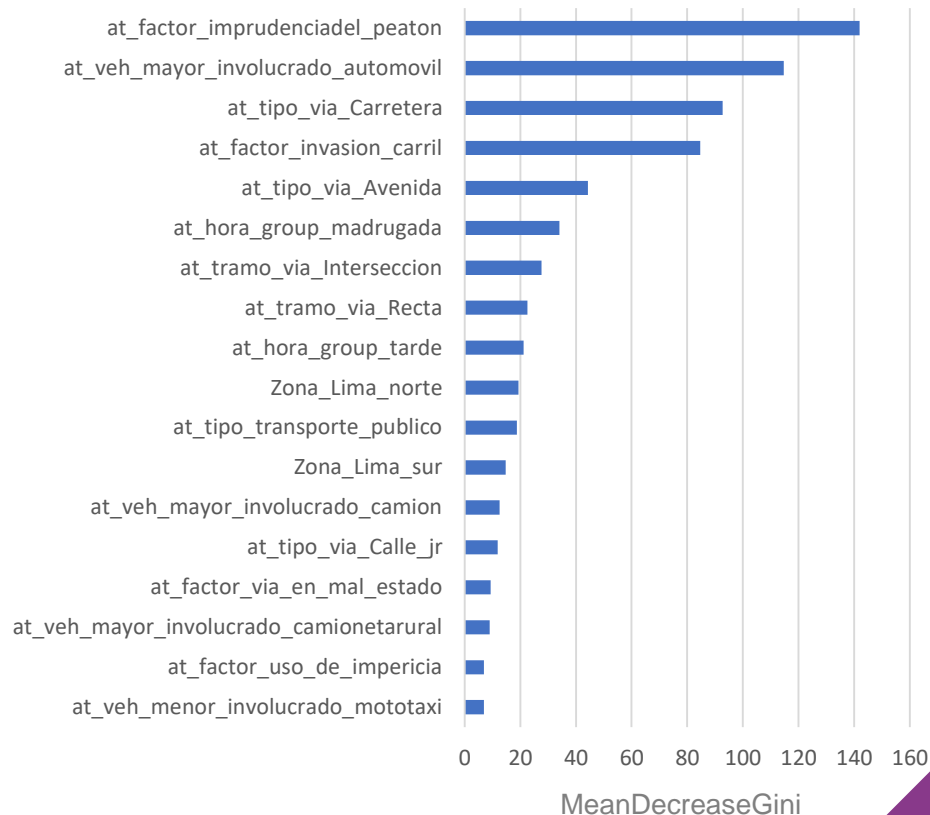


Random Forest



	Sensibilidad (Recall)	Especificidad	Precisión	AreaROC
Promedio	64%	83%	79%	92%

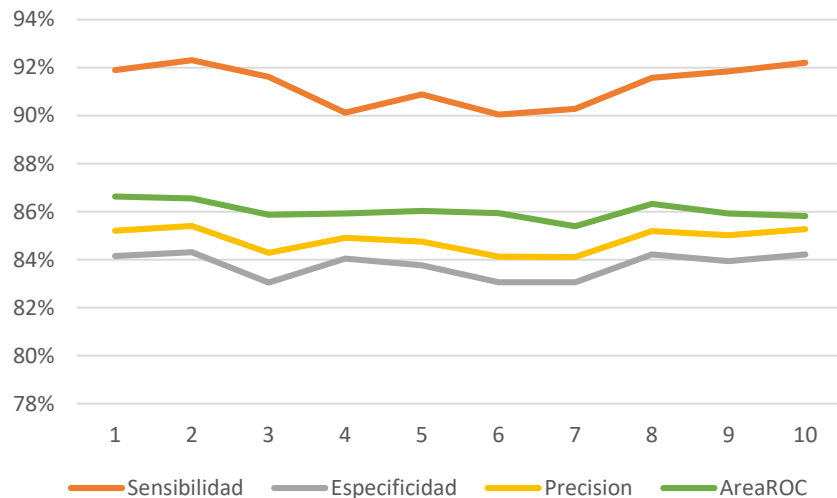
Importancia de variables



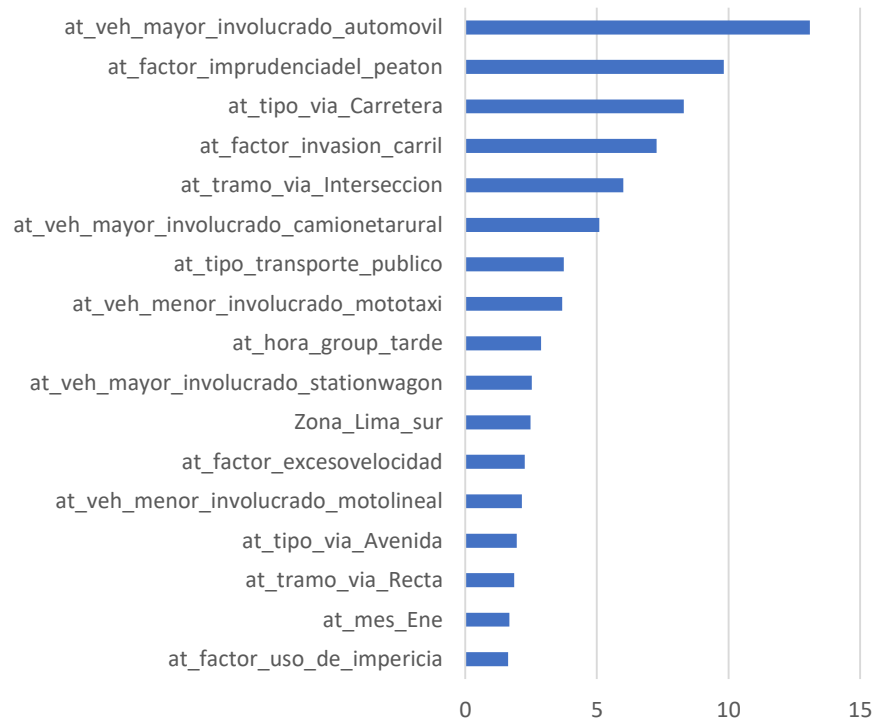
Cross validation



Boosting



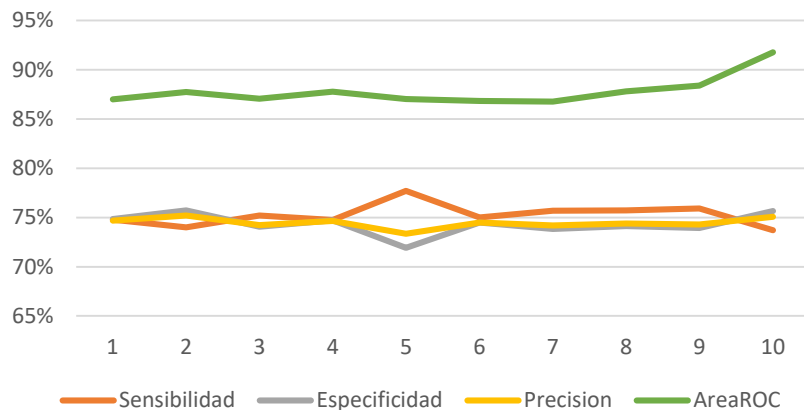
Importancia de variables



	Sensibilidad (Recall)	Especificidad	Precisión	AreaROC
Promedio	91%	84%	85%	86%

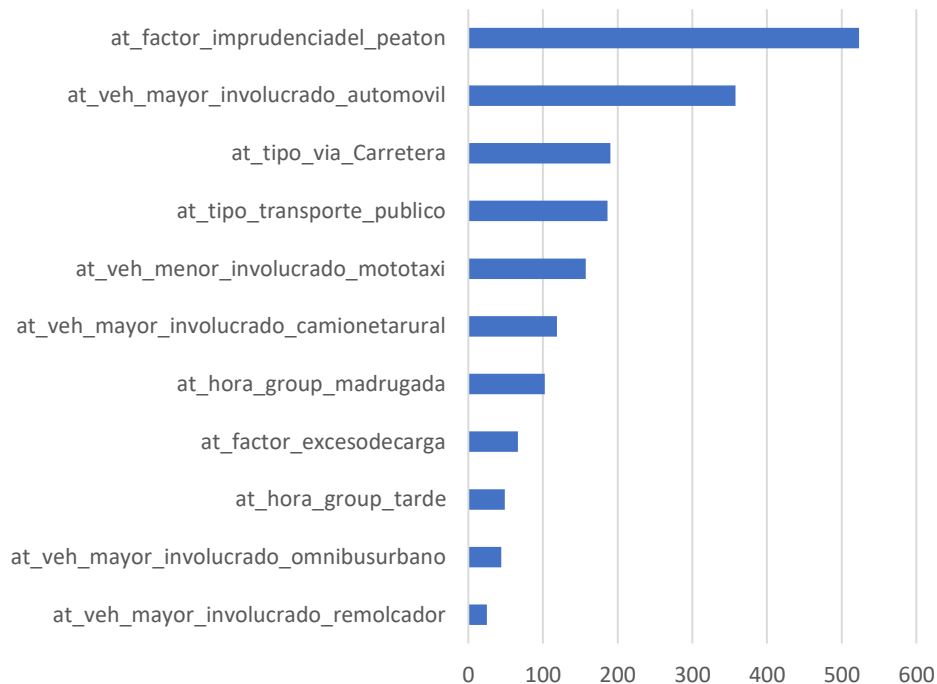
Cross validation

Árbol de clasificación



	Sensibilidad (Recall)	Especificidad	Precisión	AreaROC
Promedio	75%	74%	74%	88%

Importancia de variables



5



Basados en los datos en análisis:

Variables para prevenir ATF:

- Educación vial
- Respeto



Imprudencia del peatón



Tipo de vía: Carretera



Invasión al carril contrario



Mayor cuidado



Tramo de vía: Intersección



Vehículo involucrado: Camioneta rural (combi)



Futuros pasos:

- Recopilar más fuentes de información:
 - De conductores.
 - De pasajeros.
 - De peatones.
- Incorporar información de geográfica
- Incorporar información de años anteriores
- Incorporar información de expertos



Gracias 😊

Contacto: Sherly Tarazona
(sindiasherly@gmail.com)