

ANÁLISIS DE ACCIDENTES DE TRÁNSITO EN EL CANTÓN GUAYAQUIL USANDO MACHINE LEARNING

PROBLEMA

Según la Comisión Económica para América Latina, los accidentes de tránsito son la segunda causa de muerte para personas entre 5 y 29 años de edad y pueden convertirse en la tercera causa de discapacidad para personas entre 30 y 44 años en el mundo. Esto representa en términos monetarios, altos costos a la economía de los países rondando los \$518MM (miles de millones), de los cuales \$65MM corresponden a países de medianos o bajos ingresos, monto que supera al que reciben como ayuda de fondos internacionales.

OBJETIVO GENERAL

Analizar los accidentes de tránsito ocurridos en el cantón Guayaquil usando técnicas de Machine Learning, para poder identificar factores principales relacionados con el estado post accidente de los tripulantes.

PROPUESTA

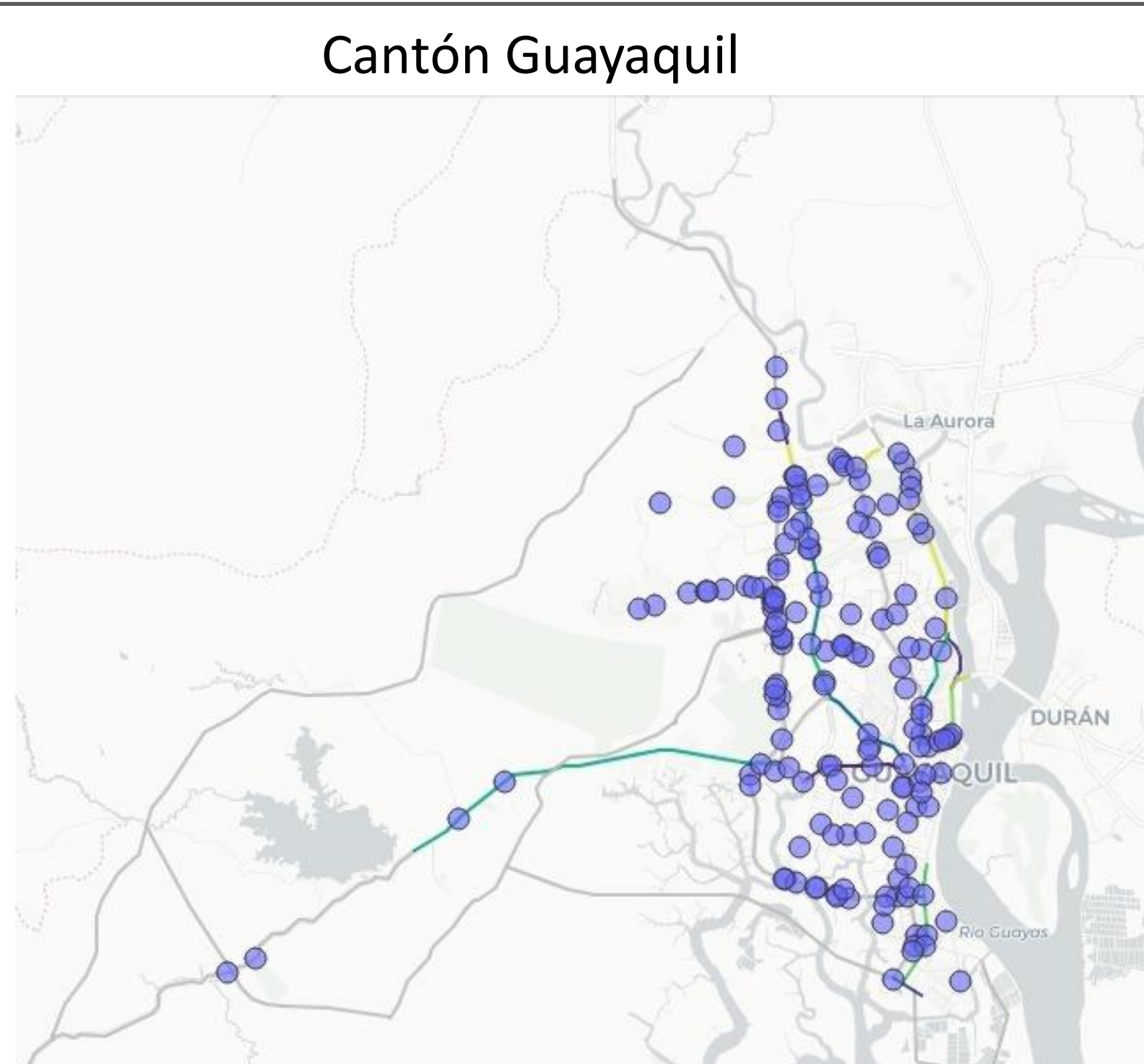
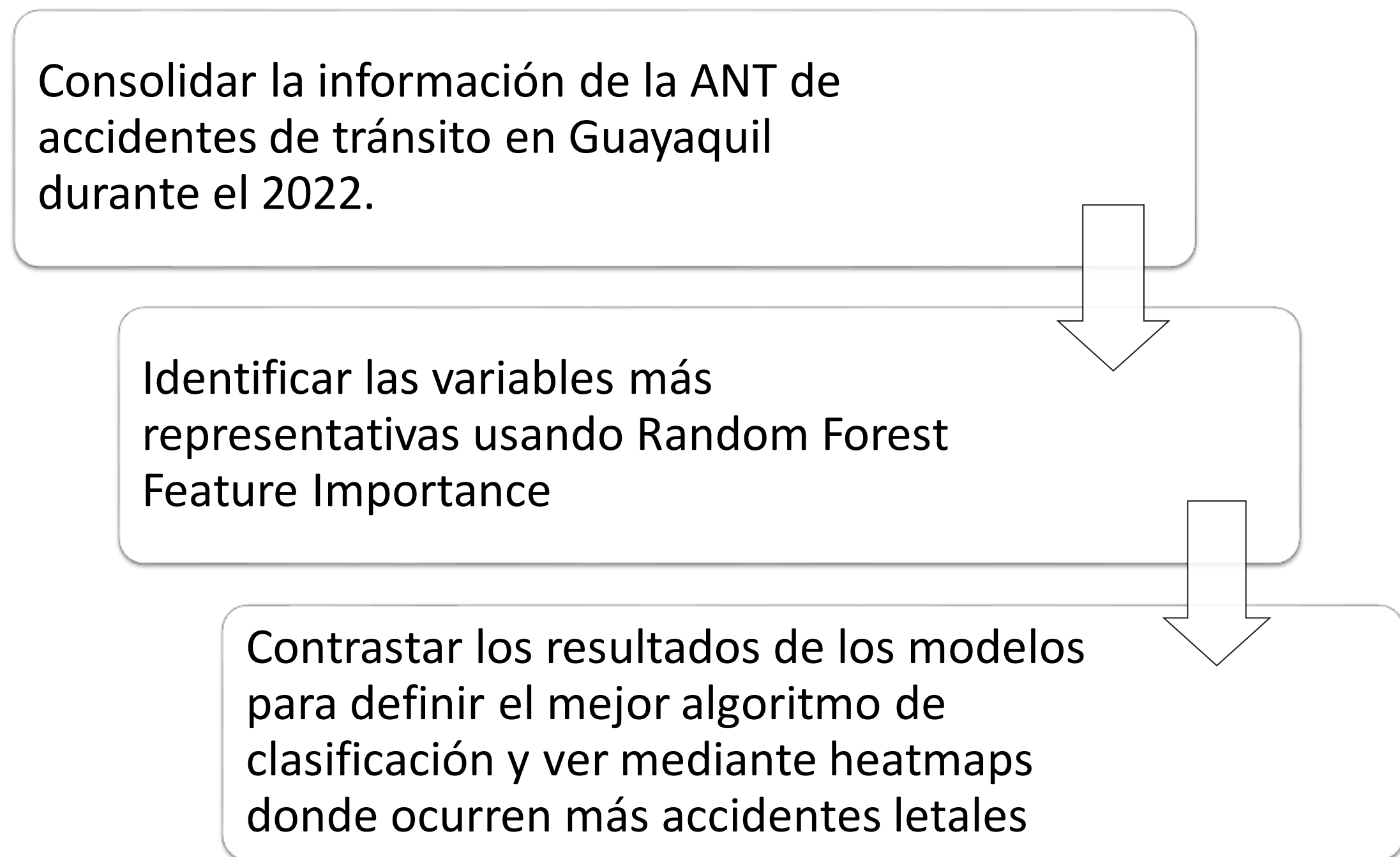


Imagen 1. Accidentes de tránsito en el cantón Guayaquil

RESULTADOS

Variables Explicativas

- escogidas
- Mes
 - Día
 - Periodo
 - Feriado
 - Clase Final
 - Zona
 - Tipo de vehículo
 - Suma de vehículos
 - Edad
 - Sexo
 - Casco
 - Participante
 - Cinturón

Performance SVM (95.22%)

PRON/REAL	FALLECIDO	ILESO	LESIONADO
FALLECIDO	26	2	0
ILESO	19	154	0
LESIONADO	12	10	677

Performance Random Forest (94.87%)

PRON/REAL	FALLECIDO	ILESO	LESIONADO
FALLECIDO	27	4	0
ILESO	18	150	0
LESIONADO	12	12	677

Performance Regresión Multinomial (94.56%)

PRON/REAL	FALLECIDO	ILESO	LESIONADO
FALLECIDO	28	3	0
ILESO	17	147	1
LESIONADO	12	16	676

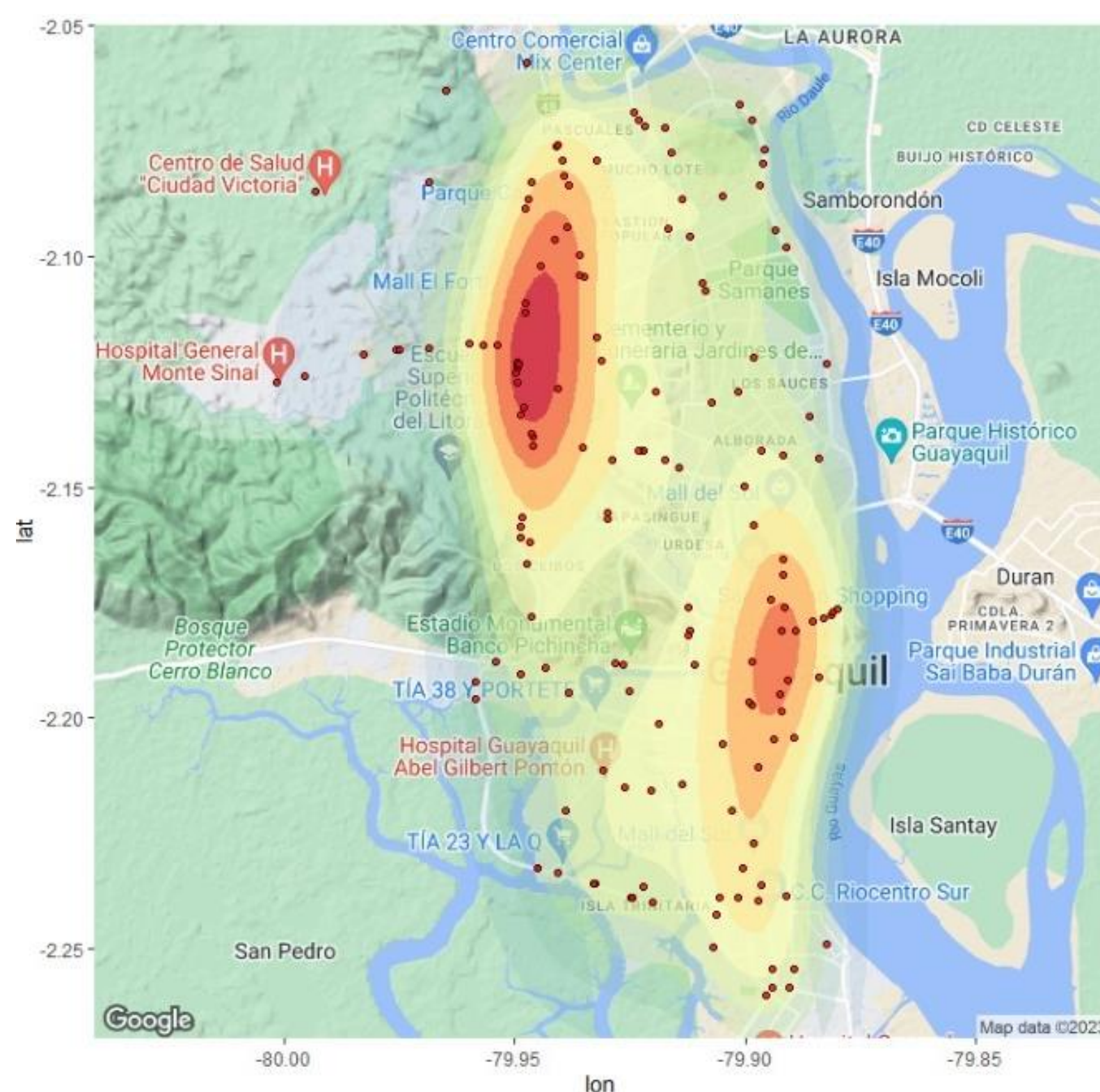
Precisión de los modelos entrenados

Después de entrenar los tres modelos

CONCLUSIONES

- Los modelos realizados cumplen con el objetivo de predecir de manera correcta la severidad, usando SVM se obtuvo el mayor accuracy debido a la naturaleza de los datos sin embargo para el análisis que se realizó la metodología por Random forest es la más adecuada porque al tener gran cantidad de predictores el modelo maneja implícitamente la colinealidad en las variables.

Heatmap de accidentes letales en Guayaquil



Los accidentes letales están más concentrados en las **zonas marginales** entre Mapasingue hasta Pascuales y en **las más transitadas** como el centro de la ciudad.

- Las variables de mayor importancia en el modelo de Random forest fueron: el uso de cinturón, la clase final, el casco, el participante y el número de lesionados por siniestros.
- El análisis de clúster clasificó en 9 grupos los siniestros de tránsito, siendo los más destacados el clúster 8 representado por hombres de 37 años fallecidos en automóviles por atropellos, choque posterior y rozamientos.