

65

Fecha de presentación: abril, 2023
Fecha de aceptación: agosto, 2023
Fecha de publicación: septiembre, 2023

MODELO

PREDICTIVO DE ACCIDENTES DE TRÁNSITO EN EL CANTÓN GUA-
YAQUIL-ECUADOR

A MODEL FOR PREDICTING TRAFFIC ACCIDENTS IN GUAYAQUIL-ECUADOR

Sergio Raúl Villacrés Parra^{1*}

Email: svillacres1604@uta.edu.ec

ORCID: <https://orcid.org/0000-0002-9497-9795>

Edison Fernando Loza Aguirre¹

Email: ef.loza@uta.edu.ec

ORCID: <https://orcid.org/0000-0003-1481-8105>

¹Universidad Técnica de Ambato. Ecuador

*Autor para Correspondencia

Cita Sugerida (Apa 7ma Edición)

Villacrés Parra, S. R. & Loza Aguirre, E. F. (2023). Modelo predictivo de accidentes de tránsito en el cantón Guayaquil-Ecuador. *Universidad y Sociedad*, 15(5). 652-662.

RESUMEN

Los accidentes que ocurren en las vías y las consecuencias derivadas de éstos, como: muertes, lesionados y pérdidas económicas, es un problema de incidencia mundial. Este estudio analizó los accidentes de tránsito que ocurrieron entre enero de 2017 y junio de 2023, en el cantón Guayaquil en Ecuador. El principal objetivo de este estudio fue modelar los accidentes de tránsito, empleando tres técnicas de modelado, para una serie de tiempo semanal. Se probaron modelos de suavizado exponencial (ETS), modelos autorregresivos integrados de media móvil (ARIMA) y el algoritmo Prophet para determinar el modelo con mejor capacidad predictiva y generar pronósticos. Del análisis de factores humanos, como causas probables que provocaron los accidentes, se determinó que conducir el vehículo superando los límites máximos de velocidad; no respetar las señales reglamentarias de tránsito; conducir desatento a las condiciones de tránsito; realizar un cambio brusco o indebido de carril y no guardar la distancia lateral mínima de seguridad entre vehículos ocasionaron el 80% de los accidentes entre 2017 y 2023 en Guayaquil – Ecuador. Por otra parte, el modelo Prophet generó los valores más bajos de MAPE y generaron pronósticos razonables del número de accidentes de tránsito para un horizonte de 26 semanas. Los pronósticos mostraron que la cantidad máxima de siniestros ocurrirán en la última semana de diciembre y será de 62 accidentes, ocasionados por las cinco causas probables identificadas como más frecuentes.

Palabras clave: Accidentes de tráfico, modelos predictivos, análisis de series temporales, accidentes viales, predicción de accidentes.

ABSTRACT

Accidents that occur on the roads and the consequences derived from them, such as: deaths, injuries and economic losses, is a problem of worldwide incidence. This study analyzed traffic accidents that occurred between January 2017 and June 2023, in the Guayaquil canton in Ecuador. The main objective of this study was to model traffic accidents, using three modeling techniques, for a weekly time series. Exponential Smoothing (ETS) models, Autoregressive Integrated Moving Average (ARIMA) models, and the Prophet algorithm were tested to determine the model with the best predictive ability and generate forecasts. From the analysis of human factors, such as probable causes that caused the accidents, it was determined that driving the vehicle exceeding the maximum speed limits; not respecting the regulatory traffic signs; driving inattentive to traffic conditions; Making a sudden or improper lane change and not keeping the minimum lateral safety distance between vehicles caused 80% of the accidents between 2017 and 2023 in Guayaquil - Ecuador. On the other hand, the Prophet model generated the lowest MAPE values and generated reasonable forecasts of the number of traffic accidents for a 26-week horizon. The forecasts showed that the maximum number of accidents will occur in the last week of December and will be 62 accidents, caused by the five probable causes identified as most frequent.

Keywords: Road traffic accidents, forecasting models, time series analysis, road accidents, crash prediction.

INTRODUCCIÓN

Los accidentes de tránsito son la primera causa de muerte de personas entre los 5 y 29 años y es la octava causa de muerte para personas de todas las edades (Uğuz & Büyükgökoğlan, 2022). A nivel mundial 1.35 millones de personas mueren y 50 millones de personas resultan heridas producto de un accidente de tránsito cada año Salud (OMS, 2022). Actualmente, la Organización Mundial de la Salud (OMS), indica que en el mundo cada 23 segundos una persona muere en la red vial (OMS, 2022). Los accidentes de tránsito además de ser un grave problema de salud pública tienen efectos en los ámbitos social (Infante et al., 2022) y económico (Chellai, 2022).

Los accidentes de tránsito a nivel mundial son de naturaleza poco predecible y se caracterizan por la contingencia y la rareza (Mičić et al., 2022); y como tales, son eventos difíciles de predecir. De lo que se ha estudiado, algunos países como: Turquía (Uğuz & Büyükgökoğlan, 2022), Portugal (Infante et al., 2022), Argelia (Chellai, 2022), Serbia (Mičić et al., 2022), Estados Unidos y Pakistán, están interesados en generar modelos para predecir el número de accidentes o la gravedad de ellos, como un esfuerzo por mejorar la seguridad vial. Conscientes de que las pérdidas humanas son irreparables; los accidentes de tránsito producen un problema económico, pues generan pérdidas en los países, que en algunos casos representan entre el 2.2% y 2.7 % de su Producto Interno Bruto (Infante et al., 2022; Uğuz & Büyükgökoğlan, 2022).

En Ecuador, la Agencia Nacional de Tránsito (ANT), registró desde 2017 al 2019, una media anual de 26365 accidentes de tránsito; de los cuales el 8% produjo víctimas mortales en sitio y el 80% produjo lesionados producto de los siniestros. Entre las consecuencias negativas, que no reflejan las estadísticas se incluyen: la degradación de la calidad de vida si la persona tuvo alguna afección física, impactos psicológicos para víctimas y sus familiares, pérdida de productividad, costos legales y médicos (Chellai, 2022).

Un siniestro, según la ANT está calificado como un: arrollamiento, atropello, caída de pasajero, choque frontal, choque lateral, choque posterior, colisión, estrellamiento, pérdida de pista, rozamiento o volcamiento (ANT, 2022). De acuerdo con las características: del conductor, de la carretera del vehículo, del choque y los elementos atmosféricos (Sun et al., 2022), son factores que influyen en la gravedad y frecuencia de los choques.

En general las investigaciones se han centrado en estudiar factores humanos, vehiculares y ambientales en el análisis de series temporales para la construcción de modelos de predictivos (Nassiri et al., 2023). En el caso

ecuatoriano la ANT, registra un listado de 27 causas probables que generaron los siniestros, todas corresponden a factores humanos, otros estudios concuerdan con algunos de ellos; entre los más frecuentes en el periodo 2017-2022 están: conducir el vehículo superando los límites máximos de velocidad; no respetar las señales reglamentarias de tránsito, como: pare, ceda el paso, luz roja del semáforo, etc.; conducir desatento a las condiciones de tránsito (celular, pantallas de video, comida, maquillaje o cualquier otro elemento distractor); realizar cambios bruscos o indebidos de carril; no guardar la distancia lateral mínima de seguridad entre vehículos; conducir bajo la influencia del alcohol, sustancias estupefacientes o psicótropas y/o medicamentos (ANT, 2022).

Ecuador supera la tasa promedio mundial, que es de 15.6 muertes por accidentes de tránsito, su media varía de 20 a 24 muertes por cada 100.000 habitantes (OMS, 2022). Por lo que se considera relevante la búsqueda de mecanismos para la reducción de siniestros. Hasta principios de diciembre del año 2022, la ANT registró 19742 siniestros, de los cuales el 36.7% ocurrieron en la provincia del Guayas, siendo esta la provincia que más accidentes con lesionados y fallecidos en sitio registra (ANT, 2022); y en esta provincia el cantón con más registro de siniestros ocurridos es el cantón Guayaquil.

Al igual que para otros países, es de interés para el Ecuador reducir el número de siniestros, para minimizar los efectos negativos que éstos conllevan. La generación de predicciones facilita la preparación de equipos de asistencia médica para atender estos incidentes y la reducción de sus tiempos de respuesta. Además, contribuirá a la gestión de iniciativas de seguridad vial que minimicen los accidentes de tránsito, provocados por las principales causas probables registradas.

Este estudio se enfocará en el cantón que más accidentes registró en Ecuador, según lo reportado en el año 2023; teniendo por objetivo el desarrollo de un modelo para predecir, en un horizonte de seis meses, el número de accidentes de tránsito generados por las causas probables más relevantes que registró la ANT en el cantón Guayaquil.

El modelado de accidentes de tránsito tiene enfoques bien definidos: el empleo de modelos estadísticos, el uso de modelos más avanzados de aprendizaje de máquina, en donde destacan las redes neuronales artificiales (ANN) (Al-Masaeid & Khaled, 2023; Cai & Di, 2023; Slimani et al., 2020), y los métodos de aprendizaje profundo (Nassiri et al., 2023), son usados para predecir accidentes de tránsito; cada tipo de modelo presenta fortalezas y debilidades.

Los accidentes de tránsito son considerados como recuentos de eventos no negativos y enteros, que ocurren aleatoriamente, esto ha motivado a los investigadores a emplear los modelos, binomial negativo (NB) y de Poisson (Cai & Di, 2023); sin embargo, según Nassiri et al., (2023), estos modelos tienen un supuesto implícito restrictivo de términos de error, los cuales se distribuyen de manera independiente e idéntica, por lo que puede ser inapropiado usar estos modelos para analizar datos de series temporales.

Los modelos autorregresivos integrados de media móvil (ARIMA) (Al-Masaeid & Khaled, 2023), y variaciones como los modelos SARIMA (Slimani et al., 2020; Deretić et al., 2022; Chellai, 2022), ARIMAX, conocidos como modelos de regresión dinámica que consideran variables exógenas (Nassiri et al., 2023; Behzadi et al., 2023), son modelos ampliamente usados con estos propósitos. Dentro de los modelos estadísticos los modelos de suavizado exponencial (Chellai, 2022), también han sido usados con estos fines. Estos dos tipos de modelos han demostrado tener buen rendimiento; sin embargo, en la mayoría de las investigaciones que emplearon modelos ARIMA, los entrenaron para pronosticar la totalidad del número de siniestros, sin considerar las causas probables (Behzadi et al., 2023).

Si bien el modelo ARIMA es el método de análisis de series temporales más utilizado en la investigación sobre seguridad vial (Behzadi et al., 2023), estos modelos presentan algunas particularidades que deben ser tomadas en cuenta; por ejemplo, cuando la serie temporal presenta un crecimiento o desaceleración, es decir la tendencia general de la serie cambia, es posible que estos modelos no funcionen adecuadamente (Al-Masaeid & Khaled, 2023).

Según Hyndman, los modelos ARIMA y los modelos de suavizado exponencial simple (ETS), no tienen la capacidad de lidiar con la complejidad de la estacionalidad que presentan algunos datos diarios, que pueden presentar un patrón estacionalidad semanal o anual (Hyndman & Athanasopoulos, 2021).

Por otra parte, Taylor y Letham investigadores de Facebook propusieron un modelo de regresión modular con parámetros interpretables que pueden ser ajustados intuitivamente, denominado Prophet. Este modelo puede ser considerado como un modelo de regresión no lineal, que se estima utilizando un enfoque bayesiano para permitir la selección automática de los puntos de cambio y otras características del modelo (Hyndman & Athanasopoulos, 2021). Destaca porque modela adecuadamente patrones de estacionalidad en series temporales.

En cuanto a la evaluación del desempeño de los modelos, en investigaciones con el mismo objetivo se han usado métricas del error como: raíz del error cuadrático medio (RSME), error absoluto medio (MAE) y error porcentual absoluto medio (MAPE) (Al-Masaeid & Khaled, 2023; Chellai, 2022).

MATERIALES Y MÉTODOS

Esta investigación se desarrolló en cinco etapas y para ello se contó con una base de datos de acceso público, conformado por valores diarios de siniestros que ocurrieron en las vías de todo el país, la zona de interés de este estudio comprende el área urbana y rural del cantón Guayaquil, desde el 01 de enero de 2017 hasta el 31 de junio de 2023. El conjunto de datos está conformado por 1705 datos, según reporta la ANT en su página web (ANT, 2022), y la variable en la serie de tiempo semanal corresponde al número de accidentes ocurridos.

En la primera etapa, se aplicó el principio de Pareto que permitió identificar las causas probables que generan el 80% de los accidentes de tránsito en el área de estudio, de las 27 causas asignadas por los agentes de tránsito cuando acudieron a los lugares del siniestro. Este criterio se empleó, de manera previa a la construcción de las series de tiempo que fueron modeladas.

En la segunda etapa, se realizó la preparación del conjunto de datos. La información diaria proporcionada por la ANT fue transformada a una serie temporal semanal y clasificada por tipo de causa, para ello se empleó el software Ms Access. A través de una exploración inicial de los datos, se identificó que existen datos atípicos ocasionados por las medidas de restricción de movilidad en la época pandémica que en Ecuador inició en marzo del año 2020, razón por la cual se optó por construir todas las series de tiempo con datos desde abril de 2020 hasta julio de 2023, obteniendo series temporales semanales de 170 datos, de los cuales el 84% fueron usados para entrenar a los modelos y el 16% para probarlos.

En una tercera fase, se probaron modelos de suavizado exponencial simple (ETS), la familia de los modelos ARIMA y el modelo Prophet, para cada una de las causas identificadas.

Modelos de suavizado exponencial simple (ETS): para una serie de tiempo t con T observaciones, está dado por la ecuación (1) (Hyndman & Athanasopoulos, 2021):

$$\hat{y}_{T+1|T} = \sum_{j=0}^{T-1} \alpha(1-\alpha)^j y_{T-j} + (1-\alpha)^T l_0, \quad (1)$$

En donde $y_{\tau+1|\tau}$ representa el valor pronosticado; α es el parámetro de suavizado y varía entre $0 \leq \alpha \leq 1$ y I_0 es el valor inicial del nivel de suavizado.

Modelos ARIMA: Este modelo que fue propuesto por Box y Jenkins en 1976, es un modelo que incluye explícitamente la diferenciación de la serie de tiempo univariada y combina los modelos autorregresivos (AR) y de promedio móvil (MA) (Nassiri et al., 2023).

El modelo AR describe una serie temporal en la que la observación actual depende de sus valores anteriores, mientras que el modelo MA se utiliza para describir una serie temporal como una función lineal de los errores aleatorios actuales y anteriores. La forma general del modelo ARIMA (p, d, q) se representa mediante la ecuación (2):

$$\nabla^d y_t = \frac{\theta(B)}{\varphi(B)} \varepsilon_t, \quad (2)$$

Donde y_t es la variable dependiente de la serie temporal, que para este estudio representa el número de accidentes de tránsito, ε_t representa el error aleatorio (ruido blanco), B es el operador de retraso, ∇ representa el proceso integrado ($\nabla y_t = y_t - y_{(t-1)}$) es el orden de la diferenciación necesaria para transformar la serie temporal en estacionaria. Los términos $\theta(B), \varphi(B)$ están definidas por las ecuaciones (3) y (4):

$$\theta(B) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q), \quad (3)$$

$$\varphi(B) = (1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p) \quad (4)$$

Donde, $\varphi_1, \dots, \varphi_p$ son los parámetros del modelo autorregresivo; los términos $\theta_1, \dots, \theta_q$ son los parámetros del modelo de media móvil; p es el orden de la parte autorregresiva y q es el orden de la parte de media móvil del modelo.

El empleo del comando ARIMA del paquete "fable" del software R versión 4.3.1 (2023-06-16 ucrt) y la interfaz RStudio, presenta como resultado el modelo ARIMA óptimo para la serie de tiempo analizada (Hyndman & Athanasopoulos, 2021).

Modelo Prophet: considerado como un modelo de regresión no lineal, expresado mediante la descomposición que se observa en la ecuación (5).

$$y_t = g(t) + s(t) + h(t) + \varepsilon_t, \quad (5)$$

Donde, $g(t)$ es el término de crecimiento, que describe una tendencia lineal por partes; $s(t)$ describe los patrones estacionales de la serie temporal; $h(t)$ captura los efectos de días de vacaciones y ε_t es un término de error de ruido blanco.

Medición del error de los modelos: para este estudio se emplearon las siguientes expresiones para la medición del error de testeo de los modelos:

2.3.1 Raíz del error cuadrático medio (RMSE) (Olayode et al., 2021):

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (y - \hat{y}_t)^2}{n}}, \quad (6)$$

En todas las expresiones en donde se mencionan, n es el número de datos del conjunto de ejemplos de testeo; \hat{y}_t es el valor pronosticado y y_t es el valor real de la serie.

Error porcentual absoluto medio (MAPE) (Olayode et al., 2021):

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (y - \hat{y}_t)^2}{n}}, \quad (7)$$

Según Deretić et al. (2022), existen rangos de valores de valores de MAPE (Tabla 1), que permiten valorar la efectividad de un modelo predictivo.

Tabla 1. Interpretación de valores del MAPE

MAPE	Interpretación
< 10%	Pronóstico de alta precisión
10% -20%	Buen pronóstico
20% - 50%	Pronóstico razonable
> 50%	Pronóstico inexacto

Fuente: tomado de Deretić et al. (2022)

Debido a que el valor de MAPE, no puede calcularse cuando existen valores de "0" en el conjunto de prueba, también se empleará la métrica MAE.

Error absoluto medio (MAE) (Olayode et al., 2021):

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (y - \hat{y}_t)^2}{n}}, \quad (8)$$

Finalmente, luego de evaluar la capacidad predictiva de cada modelo probado, se seleccionó el modelo con menor valor de error.

RESULTADOS Y DISCUSIÓN:

El propósito de este estudio fue la construcción de un modelo predictivo de accidentes de tránsito. En esta sección se presentan los resultados de los hallazgos luego de realizar las cinco etapas.

En la Tabla 2. Se presentan las diez causas probables que ocasionaron el 96,7% de la cantidad de accidentes en la zona de estudio. Aplicando el principio de Pareto (Figura 1), se identificó que las causas C09, C23, C14, C19, C12 son las cinco causas que produjeron el 80% de los accidentes entre enero de 2017 y julio de 2023, en los sectores urbano y rural del cantón Guayaquil. Además, se presenta una estadística descriptiva de las series temporales semanales que se emplearon para la construcción de los modelos conformados para el periodo comprendido entre el 05-04-2020 y el 02-07-2023.

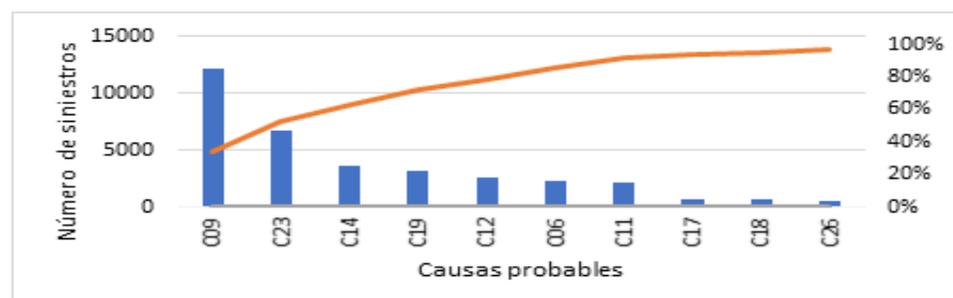
Tabla 2. Exploración de datos de siniestros en el cantón Guayaquil.

Código ANT	Causa probable	Mínimo	Media	Máximo	Moda	Total, de siniestros
C09	Conducir vehículo superando los límites máximos de velocidad.	3	29,0	57	37	12132
C23	No respetar las señales reglamentarias de tránsito.	1	15,0	33	13	6702
C14	Conducir desatento a las condiciones de tránsito.	0	3,6	12	2	3627
C19	Realizar cambio brusco o indebido de carril.	0	7,1	18	9	3137

C12	No guardar la distancia lateral mínima de seguridad entre vehículos.	0	3,7	14	2	2517
C06	Conduce bajo la influencia de alcohol, sustancias estupefacientes o psicotrópicas y/o medicamentos.	0	7,7	16	6	2354
C11	No mantener la distancia prudencial con respecto al vehículo que le antecede.	0	5,5	17	5	2158
C17	Bajarse o subirse de vehículos en movimiento sin tomar las precauciones debidas.	0	1,0	5	0	712
C18	Conducir en sentido contrario a la vía normal de circulación.	0	1,1	4	0	632
C26	No ceder el derecho de vía o preferencia de paso al peatón.	0	0,7	8	0	502

Fuente: Elaboración propia

Figura 1: Resultado de la aplicación del principio de Pareto.



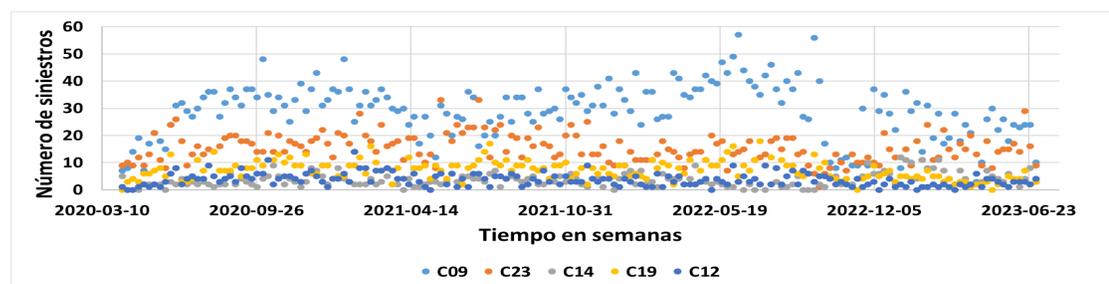
Fuente: Elaboración propia

Conducir superando los límites máximos de velocidad (C09), es la causa responsable de haber producido el 34.02% de los accidentes en el periodo 01 de enero de 2017 y 02 de julio de 2023, dejando como consecuencia de estos siniestros a 98 personas fallecidas y 1550 lesionadas. Los primeros hallazgos son consistentes con un estudio realizado en la provincia de Kerman, ubicado en sureste de Irán (Behzadi et al., 2023).

Conducir a velocidades excesivas es un factor significativo para este estudio, porque tiene efectos en la frecuencia de ocurrencia de los siniestros. Conducir a velocidades no autorizadas figuran entre los principales factores de riesgo, para estudios anteriores (Bakhtiyari et al., 2019). Según Slimani et al. (2020), esto puede ser causado por la falta de radares (cámaras de velocidad) y la imposición de multas; además indica que colocar mecanismos de control y las sanciones respectivas tras la infracción reduce el número de accidentes de forma significativa.

Una vez identificadas las cinco causas probables que mayor cantidad de accidentes han ocasionado; se construyeron modelos para las cinco series temporales semanales consideradas como las más relevantes, llamadas: C09, C23, C14, C19, C12. En la Figura 2, se muestran las series temporales, en donde se observa en términos generales que las series no tienen tendencia creciente. Sin embargo, la serie C09, presenta tendencia creciente y decreciente, en ciertos periodos cortos.

Figura 2: Serie de tiempo semanal de accidentes de tránsito en el cantón Guayaquil.



Fuente: elaboración propia

Para determinar objetivamente si las series de tiempo son estacionarias, se aplicó la prueba de hipótesis de Kwiatkowski-Phillips-Schmidt-Shin (KPSS), usando la función `unitroot_kpss()` del software R, cuya hipótesis nula es que los datos son estacionarios (Hyndman & Athanasopoulos, 2021). En el caso de la serie C09, se obtuvo un p-value de 0.1, lo que indica que no existe evidencia suficiente para rechazar la nulidad de la hipótesis con un 95% de confianza. Pero para las demás series temporales los valores de p-value fueron menores a 0.05, con lo cual se acepta la hipótesis alterna, concluyendo que estas series no son estacionarias.

Como se indicó, en este estudio se probaron los modelos ARIMA, ETS y Prophet (ver Tabla 3.). En la columna izquierda de Figura 3. se presenta gráficamente la serie con el ajuste de los modelos ARIMA, ETS, Prophet. Mientras que en la columna derecha se muestran las proyecciones de los valores de siniestros para un horizonte de 26 semanas (6 meses). Para ninguna de las series temporales los modelos ARIMA, presentan el componente de estacionalidad.

Tabla 3. Modelos probados y su desempeño.

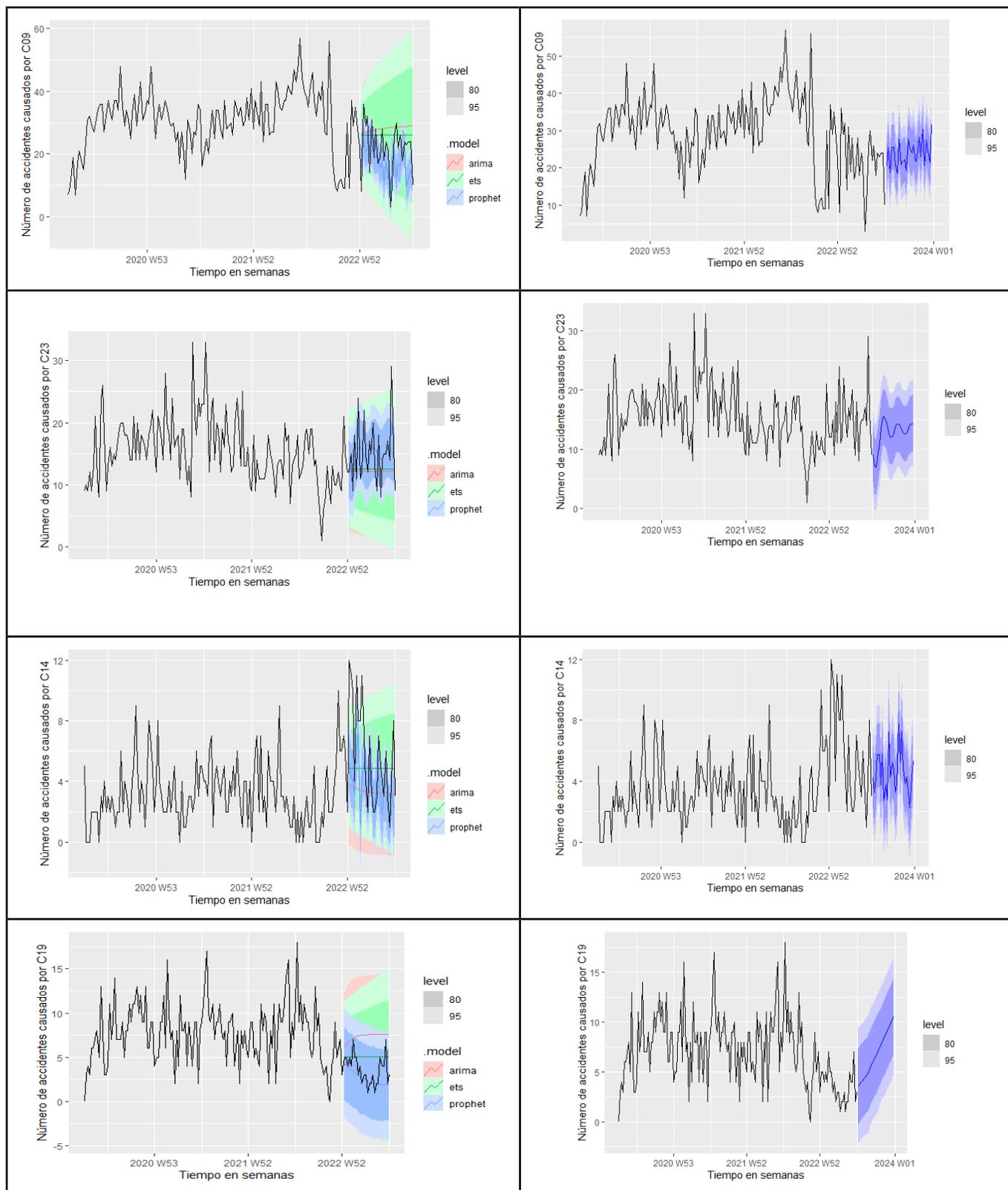
Serie	Modelo	RMSE	MAE	MAPE
C09	ARIMA (1,0,1)	10,2	8,07	79,9
	ETS (A,N,N) ($\alpha=0.4075536$)	8,99	6,82	69,8
	PROPHET - M(100)(115)	8,92	8,03	45,1
C23	ARIMA (0,1,2)	5,60	4,11	24,9
	ETS (A,N,N) ($\alpha=0.1920497$)	5,38	3,97	24,7
	PROPHET - A(200)(25)	4,64	3,55	25,2
C14	ARIMA (2,0,1)	3,94	2,95	48,9
	ETS (A,N,N) ($\alpha=0.2079322$)	3,36	2,77	64,4
	PROPHET - A(80)(100)	350	2,58	39,5
C19	ARIMA (1,0,1)	4,18	3,76	169
	ETS (A,N,N) ($\alpha=0.2395145$)	2,18	1,79	87,4
	PROPHET - A(600)(07)	1,71	1,71	33,4
C12	ARIMA (1,0,1)	1,84	1,58	Inf.
	ETS (A,Ad,N) ($\alpha=0.0753$, $\beta=0.0001$, $\phi=0,8842$)	1,6	1,33	Inf.
	PROPHET - A(700)(20)	1,36	1,02	Inf.

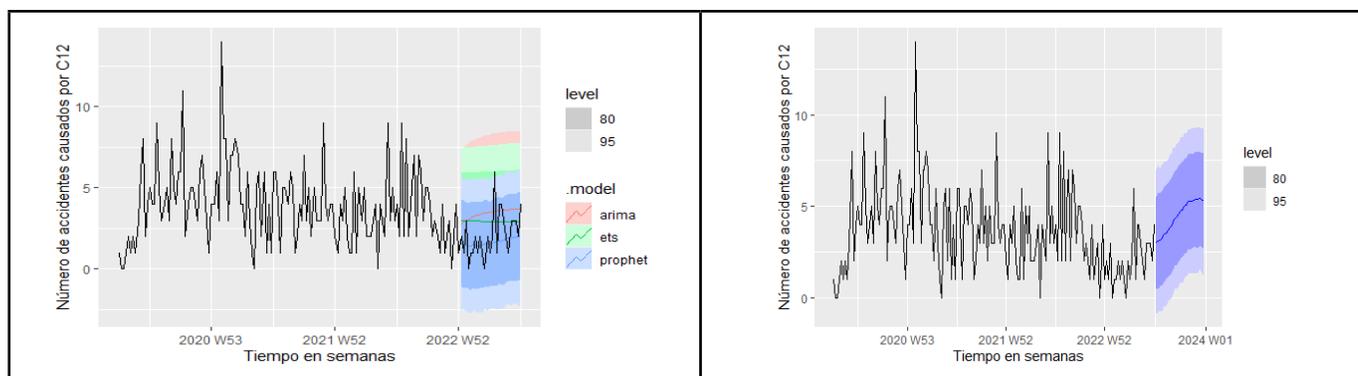
Fuente: elaboración propia

En la prueba de los modelos de suavizado exponencial simple; los modelos de todas las series fueron modelos ETS (A,N,N), que corresponde a un modelo con errores aditivos, sin tendencia y sin estacionalidad. Excepto para la serie C12, para esta serie el modelo ETS (A,Ad,N) indica que es un modelo con errores aditivos, tendencia con variantes amortiguadas y sin estacionalidad (Hyndman & Athanasopoulos, 2021).

Por otra parte, el esquema del modelo Prophet para las series temporales estudiadas fueron aditivas (A); menos para la serie C09 (Figura 4.), en este caso se seleccionó el tipo multiplicativo (M), para mejorar el desempeño del modelo. Otros parámetros que fueron ajustados para obtener valores razonables de error, buscando superar el desempeño de los modelos optimizados automáticamente ARIMA y ETS fueron el periodo y orden.

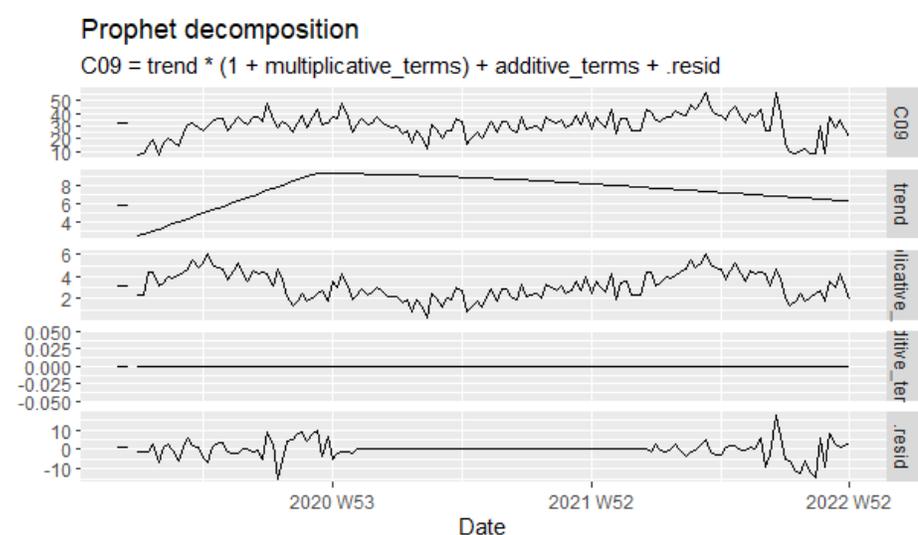
Figura 3. Ajustes de modelos (izquierda) y predicción de los modelos (derecha).





Fuente: Elaboración propia

Figura 4. Descomposición del modelo Prophet de la serie C09.



Fuente: Elaboración propia

La etapa de evaluación se realizó a través de la medición del error. Para ello se evaluó en conjunto los menores valores de RMSE y MAPE. En general los modelos con menor valor de RMSE también presentaron los menores valores de MAPE. En el caso particular de la serie C14, para la selección del modelo predictivo, se primó el valor de MAPE ya que se buscaban modelos con valores de MAPE inferiores al 50%. Para la serie C12, no se pudo calcular el valor de MAPE, dado que esta serie temporal tiene valores “0”, en el conjunto de datos de testeo. En este caso, se observó el menor valor de MAE y RMSE para seleccionar al mejor modelo predictivo.

Para todas las series se observó que el modelo Prophet fue el modelo que presentó el menor valor de error de MAPE (sus valores oscilan entre el 25 % y el 50%). Según la Tabla 1., valores de MAPE menores al 50% indican que el modelo realizará predicciones razonables del número de accidentes de tránsito.

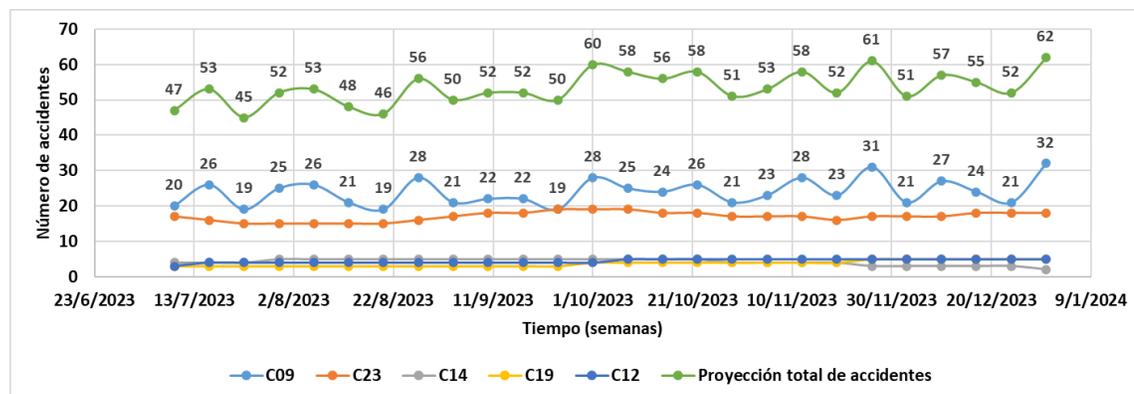
Para verificar la eficacia de los modelos, en capturar toda la información de la serie temporal, se usó como indicador el valor de la función de autocorrelación (ACF), recordando que los valores de ACF deben estar cercanos a “0”, se determinaron valores inferiores a 0,3.

Los modelos que presentaron una capacidad predictiva razonable (MAPE < 50%), fueron empleados para generar los valores semanales de accidentes de tránsito proyectados desde julio 2023 hasta diciembre de 2023. Para la causa probable C09, que es la causa que más frecuentemente genera accidentes se observa en la Figura 5, que el mayor

número de siniestros ocurrirán la última semana de diciembre, se estima una ocurrencia de 32 siniestros ocasionados por conducir a exceso de velocidad.

No respetar las señales reglamentarias de tránsito, ocasionará 16 siniestros en la última semana de agosto. Por otra parte, conducir desatento a las condiciones de tránsito provocará 8 accidentes en la última semana de septiembre del año 2023. El mayor valor de accidentes de tránsito ocasionados por realizar cambios bruscos e indebidos de carril, ocurrirán en la última semana de noviembre y durante todo el mes de diciembre. El mayor número de accidentes de tránsito ocasionados por no guardar la distancia lateral mínima de seguridad entre vehículos ocurrirá en los meses de octubre, noviembre y diciembre.

Figura 5. Valores proyectados de accidentes de tránsito, periodo julio – diciembre 2023.



Fuente: Elaboración propia

CONCLUSIONES

Este estudio incorporó el análisis de factores humanos que fueron imputados como causas probables de la generación de accidentes de tránsito en el cantón Guayaquil. El modelo Prophet presentó la mejor capacidad predictiva para todas las series. Los valores proyectados hasta diciembre de 2023, son valores razonables, de acuerdo a los valores de MAPE.

Se espera que en futuras investigaciones sea posible vincular la cantidad de accidentes, sus causas y el lugar en el que éstos ocurren; de manera que se pueda identificar en dónde ubicar los dispositivos de control o tomar medidas preventivas.

Dado que en este estudio se abordó el análisis de factores humanos como causas probables de la ocurrencia de siniestros, se recomienda la implementación de campañas de educación vial, como mecanismo para reducir la cantidad de accidentes.

REFERENCIAS BIBLIOGRÁFICAS

Al-Masaeid, H. R., & Khaled, F. J. (2023). Performance of Traffic Accidents' Prediction Models. *Jordan Journal of Civil Engineering*, 17(1). <https://jjce.just.edu.jo/Home/ShowPaperspx?data=iMPSXA%2fbU66Mf7wK85fwbSlvFPgsTjHtE9ChFCaOFiM%3d>

ANT (2022). *Estadísticas de siniestros en Ecuador*. <https://www.ant.gob.ec/visor-de-siniestralidad-estadisticas>

Bakhtiyari, M., Mehmandar, M. R., Khezeli, M., Latifi, A., Jouybari, T. A., & Mansournia, M. A. (2019). Estimating the avoidable burden and population attributable fraction of human risk factors of road traffic injuries in iran: application of penalization, bias reduction and sparse data analysis. *International journal of injury control and safety promotion*, 26(4), 405-411. <https://pubmed.ncbi.nlm.nih.gov/31514568/>

Behzadi Goodari, M., Sharifi, H., Dehesh, P., Mosleh-Shirazi, M. A., & Dehesh, T. (2023). Factors affecting the number of road traffic accidents in Kerman province, southeastern Iran (2015–2021). *Scientific reports*, 13(1), 6662. <https://doi.org/10.1038/s41598-023-33571-8>

Cai, B., & Di, Q. (2023). Different forecasting model comparison for near future crash prediction. *Applied Sciences*, 13(2), 759. <https://doi.org/10.3390/app13020759>

Chellai, F. (2022). Application of the Hybrid Forecasting Models to Road Traffic Accidents in Algeria. *Statistika: Statistics & Economy Journal*, 102(2). <https://openurl.ebsco.com/EPDB%3Agcd%3A15%3A10560383/detaIiv2?sid=ebsco%3Aplink%3Ascholar&id=ebsco%3Aqcd%3A157533334&crl=f>

- Deretić, N., Stanimirović, D., Awadh, M. A., Vujanović, N., & Djukić, A. (2022). SARIMA modelling approach for forecasting of traffic accidents. *Sustainability*, *14*(8), 4403. <https://doi.org/10.3390/su14084403>
- Hyndman, R., J. & Athanasopoulos, G. (2021). *Forecasting: principles and practice*, 3rd ed. OTexts, <https://otexts.com/fpp3/index.html>
- Infante, P., Jacinto, G., Afonso, A., Rego, L., Nogueira, V., Quaresma, P., . & Manuel, P. R. (2022). Comparison of statistical and machine-learning models on road traffic accident severity classification. *Computers*, *11*(5), 80. <https://doi.org/10.3390/computers11050080>
- Mićić, S., Vujadinović, R., Amidžić, G., Damjanović, M., & Matović, B. (2022). Accident frequency prediction model for flat rural roads in Serbia. *Sustainability*, *14*(13), 7704. <https://doi.org/10.3390/su14137704>
- Nassiri, H., Mohammadpour, S. I., & Dahaghin, M. (2023). Forecasting time trend of road traffic crashes in Iran using the macro-scale traffic flow characteristics. *Heliyon*, *9*(3). <https://www.sciencedirect.com/science/article/pii/S2405844023016882>
- Olayode, I. O., Tartibu, L. K., & Okwu, M. O. (2021). Prediction and modeling of traffic flow of human-driven vehicles at a signalized road intersection using artificial neural network model: A South African road transportation system scenario. *Transportation Engineering*, *6*, 100095. <https://www.sciencedirect.com/science/article/pii/S2666691X21000518>
- Organización Mundial de la Salud (OMS) (2022). Muerte en la carretera. <https://extranet.who.int/roadsafety/death-on-the-roads/?lang=es>
- Slimani, N., Slimani, I., Amghar, M., & Sbiti, N. (2020). Road traffic forecasting using a real data set in Morocco. *Procedia Computer Science*, *177*, 128-135. <https://www.sciencedirect.com/science/article/pii/S1877050920322857>
- Sun, M., Zhou, R., Jiao, C., & Sun, X. (2022). Severity analysis of hazardous material road transportation crashes with a Bayesian network using Highway Safety Information System data. *International journal of environmental research and public health*, *19*(7), 4002. <https://pubmed.ncbi.nlm.nih.gov/35409685/>
- Uğuz, S., & Büyükgökoğlan, E. (2022). A hybrid CNN-LSTM model for traffic accident frequency forecasting during the tourist season. *Tehnički vjesnik*, *29*(6), 2083-2089. <https://hrcak.srce.hr/file/412477>