

UNIVERSIDAD TÉCNICA DE AMBATO



FACULTAD DE INGENIERÍA EN SISTEMAS, ELECTRÓNICA E INDUSTRIAL

PROGRAMA DE MAESTRÍA EN MATEMÁTICA APLICADA COHORTE 2021

Tema: “Modelo predictivo de accidentes de tránsito en el cantón Guayaquil”

Trabajo de Titulación, previo a la obtención del Título de Cuarto Nivel de Matemática Aplicada.

Modalidad de Titulación: Artículo Profesional de Alto Nivel

Autor: Ing. Sergio Raúl Villacrés Parra

Director: Dr. Edison Fernando Loza Aguirre PhD.

Ambato – Ecuador

2023

A la Unidad Académica de Titulación de la Facultad de Ingeniería en Sistemas, Electrónica e Industrial.

El Tribunal receptor de la Defensa del Trabajo de Titulación presidido por la Ingeniera Elsa Pilar Urrutia Urrutia Magíster, e integrado por los señores: Dr. Freddy Geovanny Benalcázar Palacios, Mg e Ing. Víctor Santiago Manzano Villafuerte, Mg, designados por la Unidad Académica de Titulación de Posgrado de la Facultad de Ingeniería en Sistemas, Electrónica e Industrial de la Universidad Técnica de Ambato, para receptor el Trabajo de Titulación con el tema: “MODELO PREDICTIVO DE ACCIDENTES DE TRÁNSITO EN EL CANTÓN GUAYAQUIL”, elaborado y presentado por la Señor Ingeniero Sergio Raúl Villacrés Parra, para optar por el Título de Cuarto Nivel de Magíster en Matemática Aplicada; una vez escuchada la defensa oral del Trabajo de Titulación el Tribunal aprueba y remite el trabajo para uso y custodia en las bibliotecas de la UTA.

Ing. Elsa Pilar Urrutia Urrutia Mg.
Presidente y Miembro del Tribunal de Defensa

Dr. Freddy Geovanny Benalcázar Palacios, Mg
Miembro del Tribunal de Defensa

Ing. Víctor Santiago Manzano Villafuerte, Mg
Miembro del Tribunal de Defensa

AUTORÍA DEL TRABAJO DE INVESTIGACIÓN

La responsabilidad de las opiniones, comentarios y críticas emitidas en el Trabajo de Titulación con el tema: “MODELO PREDICTIVO DE ACCIDENTES DE TRÁNSITO EN EL CANTÓN GUAYAQUIL”, le corresponde exclusivamente al: Ingeniero Sergio Raúl Villacrés Parra, Autor bajo la Dirección del Doctor Edison Fernando Loza Aguirre Ph.D., Director del Trabajo de Titulación; y el patrimonio intelectual a la Universidad Técnica de Ambato.

Ing. Sergio Raúl Villacrés Parra

AUTOR

Dr. Edison Fernando Loza Aguirre Ph.D.

DIRECTOR

DERECHOS DE AUTOR

Autorizo a la Universidad Técnica de Ambato, para que el Trabajo de Titulación, sirva como un documento disponible para su lectura, consulta y proceso de investigación, según las normas de la Institución.

Cedo los Derechos de mi Trabajo de Titulación, con fines de difusión pública, además apruebo la reproducción de este, dentro de las regulaciones de la Universidad Técnica de Ambato.

Ing. Sergio Raúl Villacrés Parra

C.C.: 0602001604

ÍNDICE GENERAL DE CONTENIDO

Portada	i
A la unidad de titulación.....	ii
AUTORÍA DEL TRABAJO DE INVESTIGACIÓN	iii
DERECHOS DE AUTOR	iv
ÍNDICE GENERAL DE CONTENIDO.....	v
ÍNDICE DE TABLAS	vii
ÍNDICE DE FIGURAS	viii
AGRADECIMIENTO	ix
DEDICATORIA	x
Capítulo I	12
EL PROBLEMA DE INVESTIGACIÓN	12
1.1 Introducción	12
1.2 Justificación	13
1.3 Objetivos de la investigación	15
1.3.1 Objetivo general.....	15
1.3.2 Objetivos específicos	15
Capítulo II.....	16
ANTECEDENTES INVESTIGATIVOS	16
Capítulo III.....	18
MARCO METODOLÓGICO.....	18
3.1 Modelos de suavizado exponencial simple (ETS).....	18
3.2 Modelos ARIMA	18
3.3 Modelo Prophet.....	19
3.4 Medición del error de los modelos	19
Capítulo IV	21
RESULTADOS Y DISCUSIÓN	21

Capítulo V	27
CONCLUSIONES, BIBLIOGRAFÍA	27
5.1 Conclusiones	27
5.2 Bibliografía	28
CARTA DE ACEPTACIÓN ARTÍCULO	31
ARTÍCULO CIENTÍFICO	32
3.1 Modelos de suavizado exponencial simple (ETS)	36
3.2 Modelos ARIMA	36
3.3 Modelo Prophet	37
3.4 Medición del error de los modelos	37

ÍNDICE DE TABLAS

Tabla 1. Interpretación de valores del MAPE	20
Tabla 2. Exploración de datos de siniestros en el cantón Guayaquil.	21
Tabla 3. Modelos probados y su desempeño.....	23

ÍNDICE DE FIGURAS

Figura 1. Resultado de la aplicación del principio de Pareto.	21
Figura 2. Serie de tiempo semanal de accidentes de tránsito en el cantón Guayaquil. ...	22
Figura 3. Ajustes de modelos (izquierda) y predicción de los modelos (derecha).	24
Figura 4. Descomposición del modelo Prophet de la serie C09.	25
Figura 5. Valores proyectados de accidentes de tránsito, periodo julio – diciembre 2023.	26

AGRADECIMIENTO

Agradezco a Dios porque hechura
suya soy, Él es que pone en mí los
talentos, Él es que abre puertas que
nadie puede cerrar y cierra puertas
que nadie puede abrir. Gracias Señor
por tu misericordia y amor.

Sergio Raúl Villacrés Parra

DEDICATORIA

Gracias a mi familia y en especial a la Ing.
Mayra Viscaíno y a todas las personas que
me ayudaron para el feliz término de esta
nueva meta en mi vida.

Sergio Raúl Villacrés Parra

UNIVERSIDAD TÉCNICA DE AMBATO
FACULTAD DE INGENIERÍA EN SISTEMAS, ELECTRÓNICA E
INDUSTRIAL
MAESTRÍA EN MAESTRÍA EN MATEMÁTICA APLICADA, COHORTE 2021

TEMA:

Modelo predictivo de accidentes de tránsito en el cantón Guayaquil

MODALIDAD DE TITULACIÓN: Artículo Profesional de Alto Nivel

AUTOR: *Ing. Sergio Raúl Villacrés Parra*

DIRECTOR: *Ing. Edison Fernando Loza Aguirre PhD.*

FECHA: *Septiembre 14, 2023*

RESUMEN EJECUTIVO

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.

Capítulo I

EL PROBLEMA DE INVESTIGACIÓN

1.1 Introducción

En el Ecuador la Comisión Nacional de Transporte Terrestre, Tránsito y Seguridad Vial Agencia Nacional de Tránsito, también conocida como la Agencia Nacional de Tránsito (ANT), reportó que, hasta principios de diciembre del año 2022, se registraron 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 [1].

En este contexto, el presente trabajo se enfocó en el diseño de un modelo para predecir el número de accidentes de tránsito en los próximos seis meses generados por las causas probables más relevantes que ha registrado la ANT en zonas urbanas de la provincia del Guayas. Para lograr este objetivo se cuenta con una base de datos de acceso público, conformado por valores diarios de siniestros que ocurrieron en las vías de esta provincia desde el 01 de enero de 2017 hasta el 31 de diciembre de 2022. El conjunto de datos está conformado por 1705 datos, según reporta la ANT en su página web [1].

Este estudio se llevó a cabo en cuatro etapas. Se inició priorizando las causas probables de accidentes registradas en la base de datos de la ANT, empleando la distribución A-B-C. Y se analizó de manera exploratoria si la serie temporal de accidentes de tránsito registrados en zonas urbanas en la provincia del Guayas presenta: estacionalidad, si presenta tendencia creciente, decreciente o la serie es estacionaria.

En la segunda etapa, se diseñó modelos univariantes que permitan predecir para los próximos seis meses el número de accidentes ocurridos en las zonas urbanas de la provincia del Guayas, para las causas probables priorizadas más relevantes. Para ello se usó la serie cronológica de accidentes registrados desde enero de 2017 hasta agosto de 2022; exceptuando los años de pandemia 2020 y 2021, debido a que fueron años atípicos. Los modelos seleccionados para entrenarlos dependerán de las características que presente la serie de tiempo, los identificados en la revisión de la literatura son: autorregresivos integrados de media móvil (ARIMA) y autorregresivos integrados de media móvil estacionales (SARIMA).

Luego de diseñar los modelos, en una tercera etapa; se procedió con la evaluación de su desempeño. Para ello se implementó la técnica de Hold-out, que permitirá dividir el conjunto de datos en datos de entrenamiento y prueba. En esta etapa, se usó la métrica más empleada que es la Raíz del Error Cuadrático Medio (RMSE), acompañada del Error Porcentual Absoluto Medio (MAPE), indicador que permite valorar la capacidad predictiva de los modelos.

Finalmente, con el modelo que genere el menor error tanto de RMSE como de MAPE en la etapa de evaluación, se predijo el número de accidentes ocurridos en las zonas urbanas de la provincia del Guayas, para las causas probables priorizadas más relevantes, usando el modelo univariante con mejor desempeño.

1.2 Justificación

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 [2]. A nivel mundial 1.35 millones de personas mueren y 50 millones de personas resultan heridas producto de un accidente de tránsito cada año [3], [4]. Actualmente, la Organización Mundial de la Salud (OMS), indica que en el mundo cada 23 segundos una persona muere en la red vial [3]. Los accidentes de tránsito además de constituirse en un grave problema de salud pública, tiene efectos en los ámbitos social [5] y económico [6].

Los accidentes de tránsito son de naturaleza aleatoria [7], [8] y se caracterizan por la contingencia y la rareza [9]; y como tales, son eventos difíciles de predecir. De lo que se ha estudiado, algunos países como: Turquía [2], Portugal [5], Argelia [6], Serbia [7], Estados Unidos [10] y Pakistán [11], están interesados en generar modelos para predecir el número de accidentes o la gravedad de ellos, como un esfuerzo por mejorar la seguridad vial. Y conscientes de que a parte de las pérdidas humanas que son irreparables; los accidentes de tránsito producen un problema económico, pues generan pérdidas en los países [10], que en algunos casos representan entre el 2.2% – 2.7 % de su Producto Interno Bruto [2], [5].

Ecuador supera la tasa promedio mundial, que es de 15.6 muertes por cada 100.000 habitantes [12] por accidentes de tránsito, su media varía de 20 a 24 muertes por cada

100.000 habitantes [3]. Por lo que se considera relevante la búsqueda de mecanismos para la reducción de siniestros. La 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, impactos psicológicos para víctimas y sus familiares, pérdida de productividad, costos legales y médicos [6].

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 [1]. Según [8], las características: del conductor, de la carretera, del vehículo, del choque y los elementos atmosféricos son factores que influyen en la gravedad y frecuencia de los choques.

En el caso ecuatoriano la ANT registra un listado de 27 causas probables que generaron los siniestros, 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; conduce bajo la influencia de alcohol, sustancias estupefacientes o psicotrópicas y/o medicamentos. [1].

En la actualidad se habla mucho de las “ciudades inteligentes”, la tendencia es emplear la tecnología como una herramienta para monitorear sus actividades diarias. Dentro de las características para que una ciudad sea considerada como inteligente, está la movilidad inteligente. Esto ha despertado el interés para que investigadores del área empleen modelos estadísticos y algoritmos de aprendizaje de máquina que permitan analizar, predecir y prevenir accidentes de tránsito [12].

Al igual que para algunos países, en el Ecuador también se debería buscar reducir el número de siniestros ocurridos en la provincia del Guayas, por las consecuencias que éstos acarrearán, ya que ésta es la provincia en donde más se registra en el país, el uso

de herramientas estadísticas que ayuden a gestionar la implementación de iniciativas de seguridad vial que minimicen los accidentes de tránsito es importante.

1.3 Objetivos de la investigación

1.3.1 Objetivo general

Diseñar modelos univariantes para predecir el número de accidentes de tránsito en las zonas urbanas de la provincia del Guayas.

1.3.2 Objetivos específicos

- Priorizar las causas probables que generaron los accidentes de tránsito en las zonas urbanas de la provincia del Guayas desde enero de 2017 hasta diciembre de 2022.
- Diseñar los modelos univariantes que permitan predecir el número de accidentes ocurridos en las zonas urbanas de la provincia del Guayas, para las causas probables priorizadas más relevantes, empleando la serie cronológica de accidentes registrados desde enero de 2017 hasta agosto de 2022.
- Evaluar el desempeño de los modelos univariantes para predecir el número de accidentes ocurridos en las zonas urbanas de la provincia del Guayas.
- Predecir el número de accidentes que ocurrirán en los próximos 6 meses en las zonas urbanas de la provincia del Guayas, usando el modelo univariante con mejor desempeño.

Capítulo II

ANTECEDENTES INVESTIGATIVOS

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)[19], [20], [21] y los métodos de aprendizaje profundo [17], 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 [20]; sin embargo, según [17] 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) [19], [22], [23], [24] y variaciones como los modelos SARIMA [21], [25], [26], ARIMAX, conocidos como modelos de regresión dinámica que consideran variables exógenas [17],[27], [24] son modelos ampliamente usados con estos propósitos. Dentro de los modelos estadísticos los modelos de suavizado exponencial [26], 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 [27].

Si bien el modelo ARIMA es el método de análisis de series temporales más utilizado en la investigación sobre seguridad vial [27], 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 [19].

Según Hydman, 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 [28].

Por otra parte, Taylor y Letham investigadores de Facebook propusieron un modelo de regresión modular con parámetros interpretables que pueden ser ajustados intuitivamente [29], 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 [28]. 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) [19], [26].

Capítulo III

MARCO METODOLÓGICO

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 [12] 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 etapa, 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.

3.1 Modelos de suavizado exponencial simple (ETS)

Para una serie de tiempo t con y_1, \dots, y_T observaciones, está dado por la ecuación (1) [28]:

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

En donde $\hat{y}_{T+1|T}$ representa el valor pronosticado; α es el parámetro de suavizado y varía entre $0 \leq \alpha \leq 1$ y l_0 es el valor inicial del nivel de suavizado.

3.2 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) [17].

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 ($\nabla y_t = y_t - y_{t-1}$); d 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 version 4.3.1 (2023-06-16 ucrt) y la interfaza RStudio, presenta como resultado el modelo ARIMA óptimo para la serie de tiempo analizada [28].

3.3 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.

En la cuarta etapa se evaluó el desempeño de los modelos, a través del cálculo de las siguientes métricas.

3.4 Medición del error de los modelos

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

3.4.1 Raíz del error cuadrático medio (RMSE) [30]:

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

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_i es el valor real de la serie.

3.4.2 Error porcentual absoluto medio (MAPE) [30]:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|y_t - \hat{y}_t|}{y_t}, \quad (8)$$

Según [25], 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

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.

3.4.3 Error absoluto medio (MAE) [30]:

$$MAE = \frac{\sum_{t=1}^n |y_t - \hat{y}_t|}{n}, \quad (7)$$

Finalmente en la quinta etapa, luego de evaluar la capacidad predictiva de cada modelo probado, se generaron las predicciones usando el modelo con menor valor de error.

Capítulo IV

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	3 7	12132
C23	No respetar las señales reglamentarias de tránsito.	1	15,0	33	1 3	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

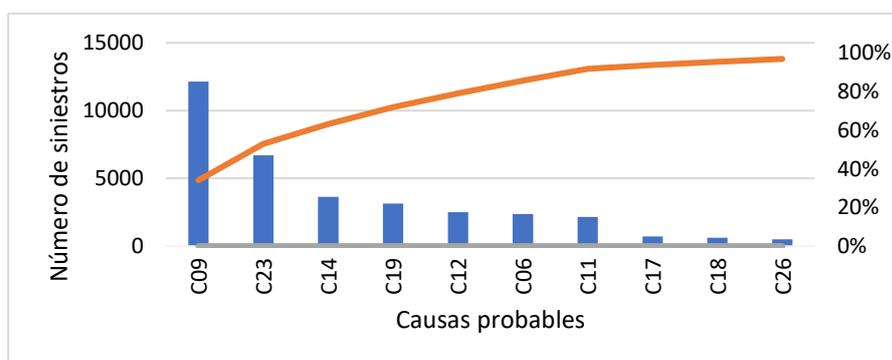


Figura 1. Resultado de la aplicación del principio de Pareto.

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 [27].

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 [31]. Según [21], 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.

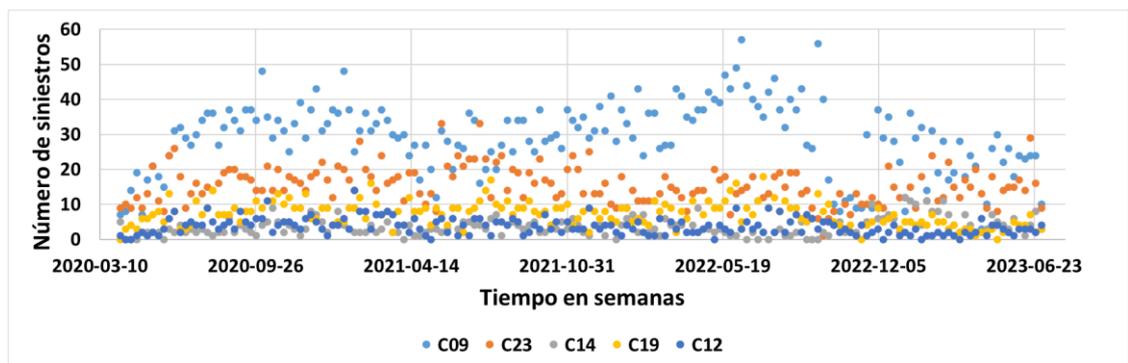


Figura 2. Serie de tiempo semanal de accidentes de tránsito en el cantón Guayaquil.

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 [28]. 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.

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 [28].

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.

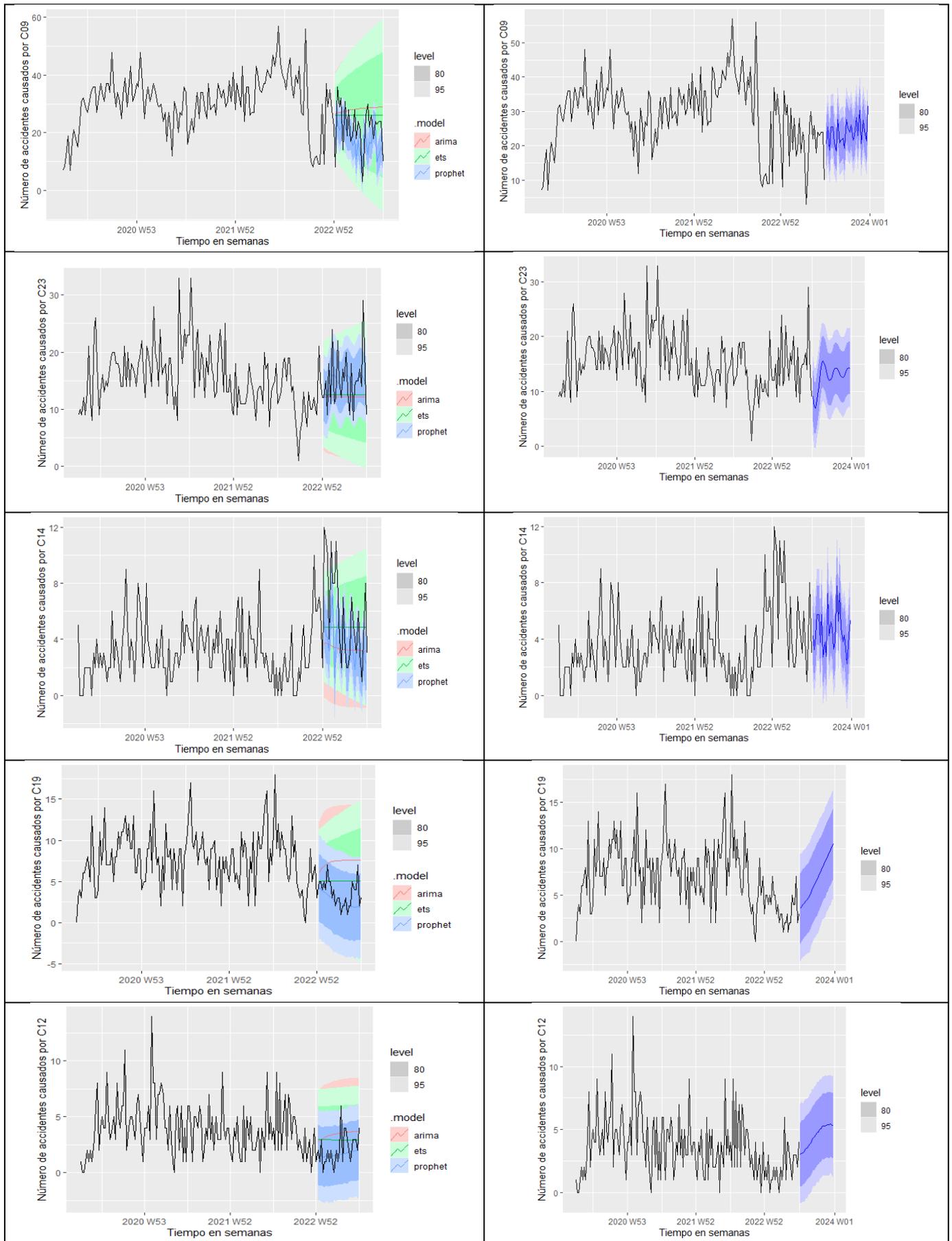


Figura 3. Ajustes de modelos (izquierda) y predicción de los modelos (derecha).

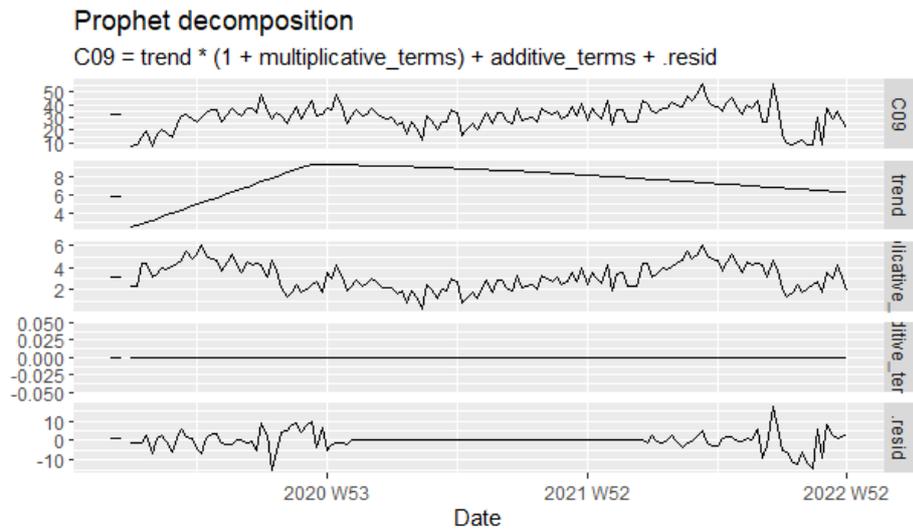


Figura 4. Descomposición del modelo Prophet de la serie C09.

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

sinistros 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.

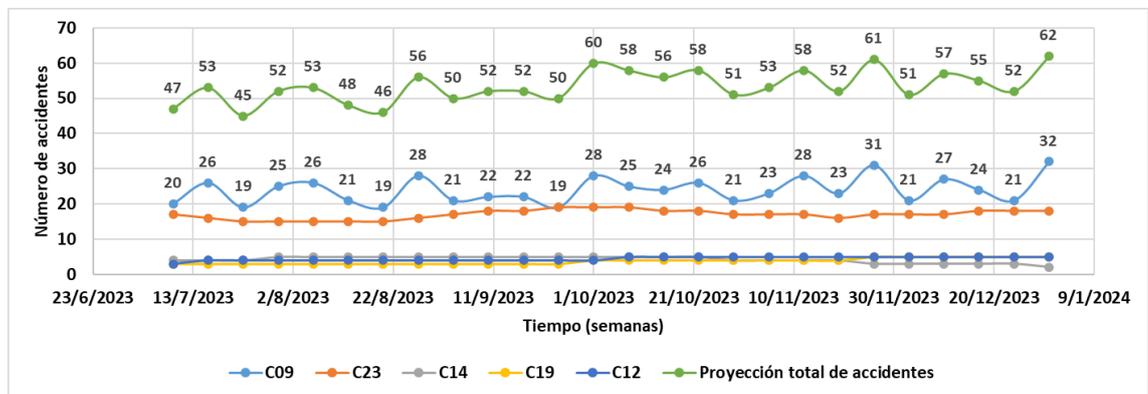


Figura 5. Valores proyectados de accidentes de tránsito, periodo julio – diciembre 2023.

Capítulo V

CONCLUSIONES, BIBLIOGRAFÍA

5.1 Conclusiones

Este estudio partió con las 27 posibles causas que tiene determina la ANT, 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

El 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 ocurrieron; de manera que se pueda identificar en dónde ubicar los dispositivos de control o tomar medidas preventivas, por parte de las entidades de control.

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.

5.2 Bibliografía

- [1] S. Uğuz and E. Büyükgökoğlan, “A Hybrid CNN-LSTM Model for Traffic Accident Frequency Forecasting During the Tourist Season,” *Tehnicki Vjesnik*, vol. 29, no. 6, pp. 2083–2089, 2022, doi: 10.17559/TV-20220225141756.
- [2] OMS, “Muerte en la carretera,” 2022. <https://extranet.who.int/roadsafety/death-on-the-roads/?lang=es>
- [3] K. Santos, J. P. Dias, and C. Amado, “A literature review of machine learning algorithms for crash injury severity prediction,” *J Safety Res*, vol. 80, pp. 254–269, 2022, doi: 10.1016/j.jsr.2021.12.007.
- [4] P. Infante *et al.*, “Comparison of Statistical and Machine-Learning Models on Road Traffic Accident Severity Classification,” *Computers*, vol. 11, no. 5, pp. 1–12, 2022, doi: 10.3390/computers11050080.
- [5] F. Chellai, “Application of the Hybrid Forecasting Models to Road Traffic Accidents in Algeria,” *Statistika*, vol. 102, no. 2, pp. 184–197, 2022, doi: 10.54694/STAT.2021.37.
- [6] S. Mičić, R. Vujadinović, G. Amidžić, M. Damjanović, and B. Matović, “Accident Frequency Prediction Model for Flat Rural Roads in Serbia,” *Sustainability (Switzerland)*, vol. 14, no. 13, 2022, doi: 10.3390/su14137704.
- [7] Md. K. Islam, I. Reza, U. Gazder, R. Akter, M. Arifuzzaman, and M. M. Rahman, “Predicting Road Crash Severity Using Classifier Models and Crash Hotspots,” *Applied Sciences*, vol. 12, no. 22, p. 11354, 2022, doi: 10.3390/app122211354.
- [8] Y. Wang, J. Xu, X. Liu, Z. Zheng, H. Zhang, and C. Wang, “Analysis on Risk Characteristics of Traffic Accidents in Small-Spacing Expressway Interchange,” *Int J Environ Res Public Health*, vol. 19, no. 16, 2022, doi: 10.3390/ijerph19169938.
- [9] W. Li, X. Zhao, and S. Liu, “Traffic accident prediction based on multivariable grey model,” *Information (Switzerland)*, vol. 11, no. 4, pp. 1–12, 2020, doi: 10.3390/info11040184.
- [10] J. Ma, K. M. Kockelman, and P. Damien, “A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods,” *Accid Anal Prev*, vol. 40, no. 3, pp. 964–975, 2008, doi: 10.1016/j.aap.2007.11.002.
- [11] S. Zhang, A. Khattak, C. M. Matara, A. Hussain, and A. Farooq, “Hybrid feature selection-based machine learning Classification system for the prediction of injury severity in single and multiple-vehicle accidents,” *PLoS One*, vol. 17, no. 2 February, pp. 1–19, 2022, doi: 10.1371/journal.pone.0262941.
- [12] ANT, “Estadísticas de siniestros en Ecuador,” 2022. <https://www.ant.gob.ec/visor-de-siniestralidad-estadisticas/>
- [13] L. Shbeeb, “Road safety performance index: A tool for crash prediction,” *Cogent Eng*, vol. 9, no. 1, 2022, doi: 10.1080/23311916.2022.2124637.

- [14] J. Q. Li, W. Liu, X. Yang, P. Lu, and K. C. P. Wang, "Statistical Safety Performance Models considering Pavement and Roadway Characteristics," *J Adv Transp*, vol. 2022, 2022, doi: 10.1155/2022/5871601.
- [15] Y. Ding, D. Li, M. Huang, X. Cao, and B. Tang, "Study on the influence of skid resistance on traffic safety of highway with a high ratio of bridges and tunnels," *Transportation Safety and Environment*, vol. 3, no. 4, 2021, doi: 10.1093/tse/tdab025.
- [16] M. Sun, R. Zhou, C. Jiao, and X. Sun, "Severity Analysis of Hazardous Material Road Transportation Crashes with a Bayesian Network Using Highway Safety Information System Data," *Int J Environ Res Public Health*, vol. 19, no. 7, 2022, doi: 10.3390/ijerph19074002.
- [17] H. Nassiri, S. I. Mohammadpour, and M. Dahaghin, "Forecasting time trend of road traffic crashes in Iran using the macro-scale traffic flow characteristics," *Heliyon*, vol. 9, no. 3, p. e14481, Mar. 2023, doi: 10.1016/j.heliyon.2023.e14481.
- [18] C. M. Ferreira-Vanegas, J. I. Velez, and G. A. Garcia-Llinas, "Analytical Methods and Determinants of Frequency and Severity of Road Accidents: A 20-Year Systematic Literature Review," *J Adv Transp*, vol. 2022, 2022, doi: 10.1155/2022/7239464.
- [19] H. R. Al-Masaeid and F. J. Khaled, "Performance of Traffic Accidents' Prediction Models," *Jordan Journal of Civil Engineering*, vol. 17, no. 1, Jan. 2023, doi: 10.14525/JJCE.v17i1.04.
- [20] B. Cai and Q. Di, "Different Forecasting Model Comparison for Near Future Crash Prediction," *Applied Sciences*, vol. 13, no. 2, p. 759, Jan. 2023, doi: 10.3390/app13020759.
- [21] N. Slimani, I. Slimani, M. Amghar, and N. Sbiti, "Road traffic forecasting using a real data set in Morocco," *Procedia Comput Sci*, vol. 177, pp. 128–135, 2020, doi: 10.1016/j.procs.2020.10.020.
- [22] Y. Wang, C. Yu, J. Hou, S. Chu, Y. Zhang, and Y. Zhu, "ARIMA Model and Few-Shot Learning for Vehicle Speed Time Series Analysis and Prediction," *Comput Intell Neurosci*, vol. 2022, pp. 1–9, Jan. 2022, doi: 10.1155/2022/2526821.
- [23] K. A. Getahun, "Time series modeling of road traffic accidents in Amhara Region," *J Big Data*, vol. 8, no. 1, p. 102, Dec. 2021, doi: 10.1186/s40537-021-00493-z.
- [24] C. C. Ihueze and U. O. Onwurah, "Road traffic accidents prediction modelling: An analysis of Anambra State, Nigeria," *Accid Anal Prev*, vol. 112, pp. 21–29, Mar. 2018, doi: 10.1016/j.aap.2017.12.016.
- [25] N. Deretić, D. Stanimirović, M. Al Awadh, N. Vujanović, and A. Djukić, "SARIMA Modelling Approach for Forecasting of Traffic Accidents," *Sustainability*, vol. 14, no. 8, p. 4403, Apr. 2022, doi: 10.3390/su14084403.
- [26] F. Chellai, "Application of the Hybrid Forecasting Models to Road Traffic Accidents in Algeria," *Statistika: Statistics and Economy Journal*, vol. 102, no. 2, pp. 184–197, Jun. 2022, doi: 10.54694/stat.2021.37.
- [27] M. Behzadi Goodari, H. Sharifi, P. Dehesh, M. A. Mosleh-Shirazi, and T. Dehesh, "Factors affecting the number of road traffic accidents in Kerman province, southeastern

Iran (2015–2021),” *Sci Rep*, vol. 13, no. 1, p. 6662, Apr. 2023, doi: 10.1038/s41598-023-33571-8.

- [28] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*, 3rd ed., vol. 1. Melbourne, Australia: OTexts, 2021. Accessed: Aug. 07, 2023. [Online]. Available: <https://otexts.com/fpp3/index.html>
- [29] S. J. Taylor and B. Letham, “Forecasting at Scale,” *Am Stat*, vol. 72, no. 1, pp. 37–45, Jan. 2018, doi: 10.1080/00031305.2017.1380080.
- [30] I. O. Olayode, L. K. Tartibu, and M. O. Okwu, “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*, vol. 6, p. 100095, Dec. 2021, doi: 10.1016/j.treng.2021.100095.
- [31] M. Bakhtiyari, M. R. Mehmandar, M. Khezeli, A. Latifi, T. A. Jouybari, and M. A. Mansournia, “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,” *Int J Inj Contr Saf Promot*, vol. 26, no. 4, pp. 405–411, Oct. 2019, doi: 10.1080/17457300.2019.1660374.

CARTA DE ACEPTACIÓN ARTÍCULO

Autor: Ing. Sergio Raúl Villacrés Parra

Director: Ing. Edison Fernando Loza Aguirre PhD.



Cienfuegos, 8 de septiembre del 2023



Por este medio se comunica que el artículo: ***“MODELO PREDICTIVO DE ACCIDENTES DE TRÁNSITO EN EL CANTÓN GUAYAQUIL-ECUADOR”*** de los Autores: ***Sergio Raúl Villacrés Parra y Edison Fernando Loza Aguirre***, se encuentra listo para su publicación en el volumen 15, número 6 (noviembre, 2023), de la revista “Universidad y Sociedad”, con ISSN: 2218- 3620. La revista se encuentra certificada por el Ministerio de Ciencia, Tecnología y Medio Ambiente (CITMA) como Publicación Seriada Científico-Tecnológica. También se encuentra indexada en directorios, catálogos y bases de datos internacionales como: Latindex, Directory of Research Journals Indexing (DRJI), Directory of Open Access Journals (DOAJ), la Biblioteca Científico-Electrónica en Línea (en inglés Scientific Electronic Library Online, SciELO) y el Emerging Sources Citation Index (ESCI) de la Web of Science y Scopus.

Atentamente,




Dr. C. Jorge Luis León
González Editor Jefe- Editorial
“UniversoSur” Universidad de
Cienfuegos, Cuba

ARTÍCULO CIENTÍFICO

Modelo predictivo de accidentes de tránsito en el cantón Guayaquil-Ecuador

A model for predicting traffic accidents in Guayaquil-Ecuador

Sergio Raúl Villacrés Parra, svillacres1604@uta.edu.ec, srvp@hotmail.com,
orcid.org/0000-0002-9497-9795 - Edison Fernando Loza Aguirre,
ef.loza@uta.edu.ec, ORCID: 0000-0003-1481-8105

Filiación: UNIVERSIDAD TÉCNICA DE AMBATO

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 [1]. A nivel mundial 1.35 millones de personas mueren y 50 millones de personas resultan heridas producto de un accidente de tránsito cada año [2], [3]. Actualmente, la Organización Mundial de la Salud (OMS), indica que en el mundo cada 23 segundos una persona muere en la red vial [2]. Los accidentes de tránsito además de ser un grave problema de salud pública tienen efectos en los ámbitos social [4] y económico [5].

Los accidentes de tránsito a nivel mundial son de naturaleza poco predecible [6], [7] y se caracterizan por la contingencia y la rareza [8]; y como tales, son eventos difíciles de predecir [9]. De lo que se ha estudiado, algunos países como: Turquía [1], Portugal [4], Argelia [5], Serbia [6], Estados Unidos [10] y Pakistán [11], 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 [10], que en algunos casos representan entre el 2.2% y 2.7 % de su Producto Interno Bruto [1], [4].

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 [5].

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 [12]. De acuerdo con [7] las características: del conductor, de la carretera [13], [14], [15] del vehículo, del choque y los elementos atmosféricos [16], 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 [17]. 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 [13]; 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 psicotrópicas y/o medicamentos [12].

Ecuador supera la tasa promedio mundial, que es de 15.6 muertes [18] por accidentes de tránsito, su media varía de 20 a 24 muertes por cada 100.000 habitantes [2]. 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 [12]; 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.

ESTADO DEL ARTE:

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)[19], [20], [21] y los métodos de aprendizaje profundo [17], 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 [20]; sin embargo, según [17] 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) [19], [22], [23], [24] y variaciones como los modelos SARIMA [21], [25], [26], ARIMAX, conocidos como modelos de regresión dinámica que consideran variables exógenas [17],[27], [24] son modelos ampliamente usados con estos propósitos. Dentro de los modelos estadísticos los modelos de suavizado exponencial [26], 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 [27].

Si bien el modelo ARIMA es el método de análisis de series temporales más utilizado en la investigación sobre seguridad vial [27], 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 [19].

Según Hydman, 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 [28].

Por otra parte, Taylor y Letham investigadores de Facebook propusieron un modelo de regresión modular con parámetros interpretables que pueden ser ajustados intuitivamente [29], 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 [28]. 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) [19], [26].

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 [12] 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 etapa, 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.

3.1 Modelos de suavizado exponencial simple (ETS)

Para una serie de tiempo t con y_1, \dots, y_T observaciones, está dado por la ecuación (1) [28]:

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

En donde $\hat{y}_{T+1|T}$ representa el valor pronosticado; α es el parámetro de suavizado y varía entre $0 \leq \alpha \leq 1$ y l_0 es el valor inicial del nivel de suavizado.

3.2 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) [17].

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 ($\nabla y_t = y_t - y_{t-1}$); d 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 version 4.3.1 (2023-06-16 ucrt) y la interfaza RStudio, presenta como resultado el modelo ARIMA óptimo para la serie de tiempo analizada [28].

3.3 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.

En la cuarta etapa se evaluó el desempeño de los modelos, a través del cálculo de las siguientes métricas.

3.4 Medición del error de los modelos

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

3.4.1 Raíz del error cuadrático medio (RMSE) [30]:

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

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_i es el valor real de la serie.

3.4.2 Error porcentual absoluto medio (MAPE) [30]:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|y_i - \hat{y}_t|}{y_i}, \quad (8)$$

Según [25], 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

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.

3.4.3 Error absoluto medio (MAE) [30]:

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

Finalmente, luego de evaluar la capacidad predictiva de cada modelo probado, se generaron las predicciones usando 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	3 7	12132
C23	No respetar las señales reglamentarias de tránsito.	1	15,0	33	1 3	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

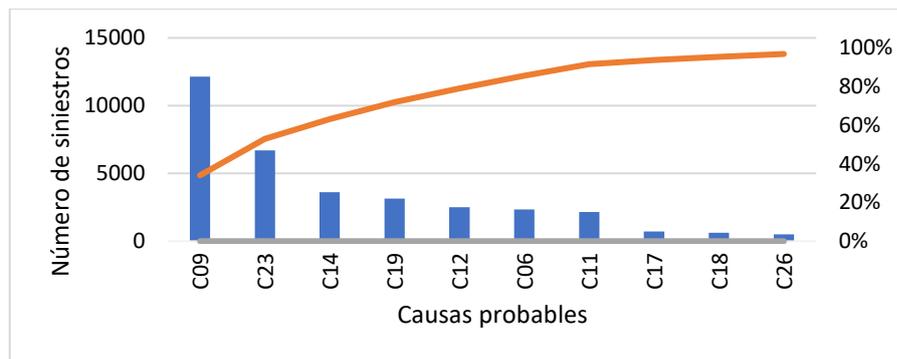


Figura 1: Resultado de la aplicación del principio de Pareto.

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 [27].

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 [31]. Según [21], 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.

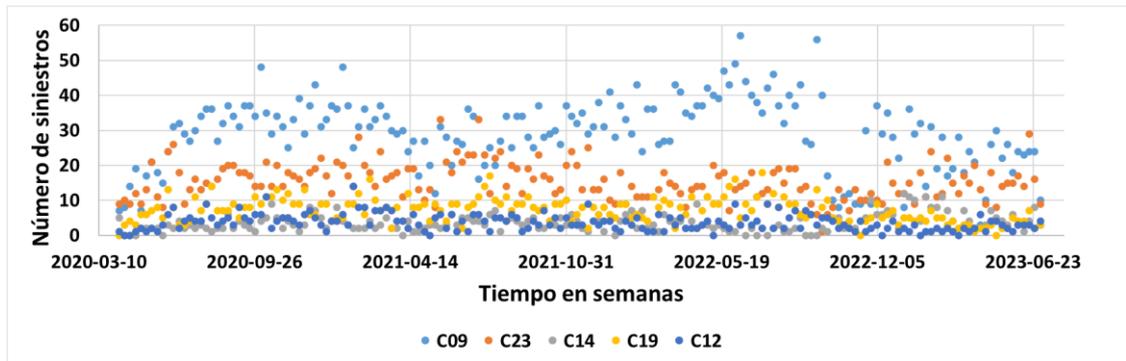


Figura 2: Serie de tiempo semanal de accidentes de tránsito en el cantón Guayaquil.

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 [28]. 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.

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 [28].

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.

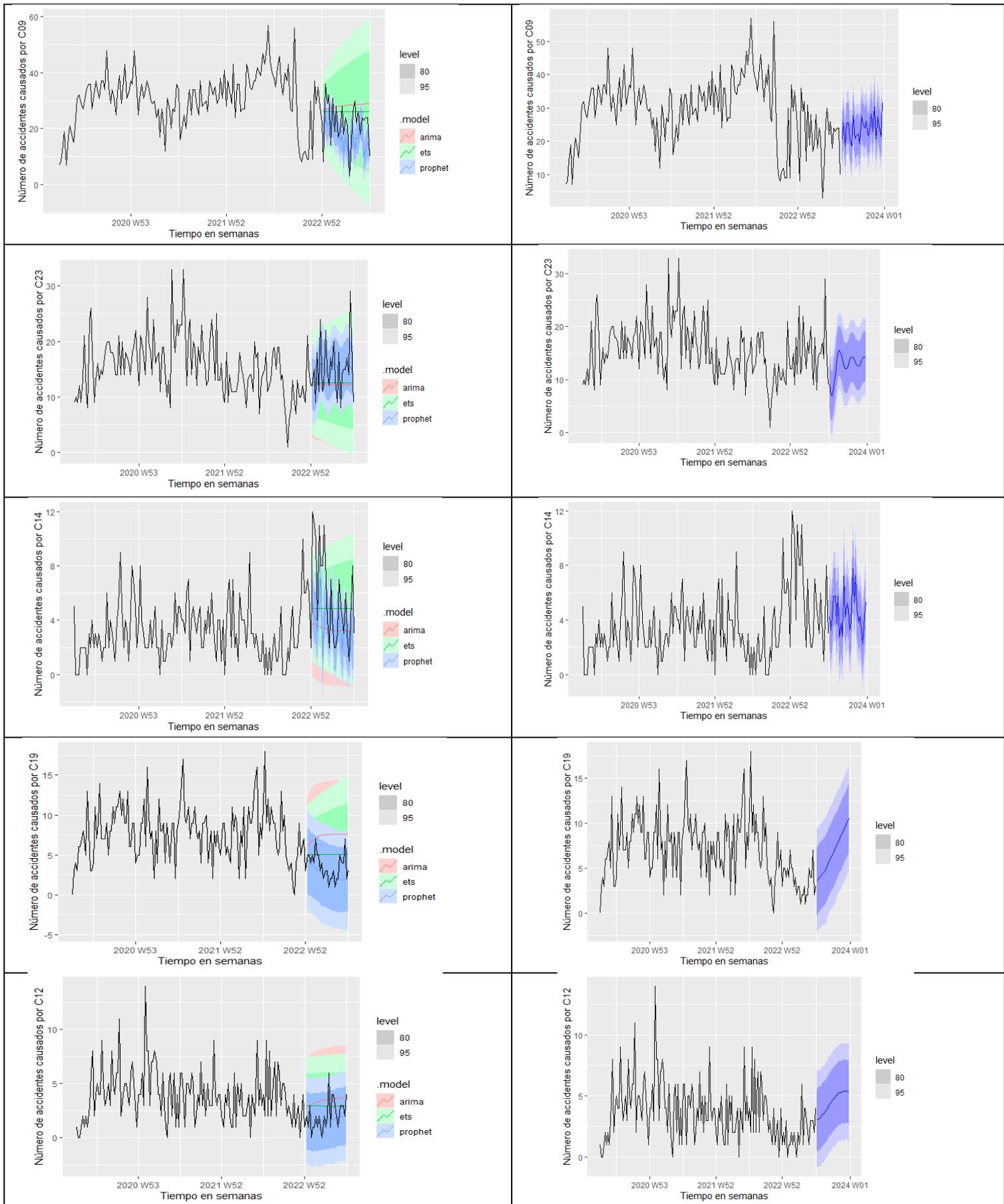


Figura 3. Ajustes de modelos (izquierda) y predicción de los modelos (derecha).

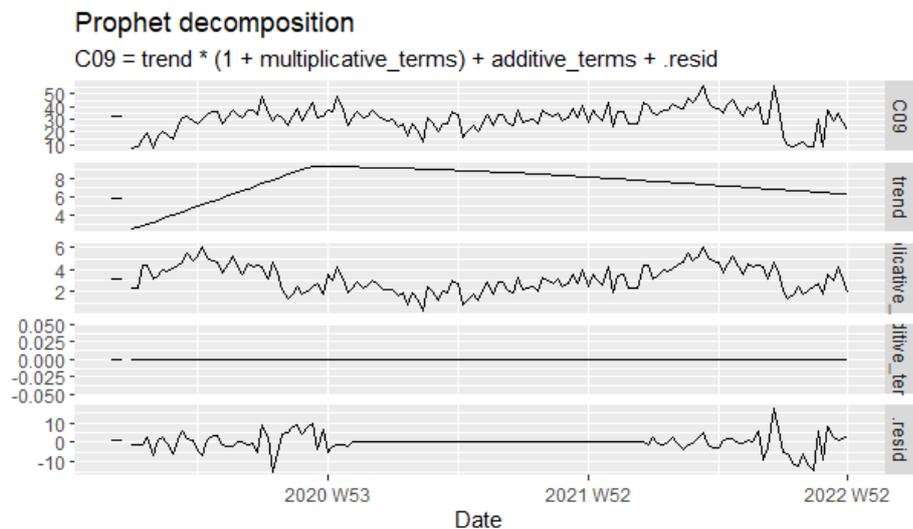


Figura 4. Descomposición del modelo Prophet de la serie C09.

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.

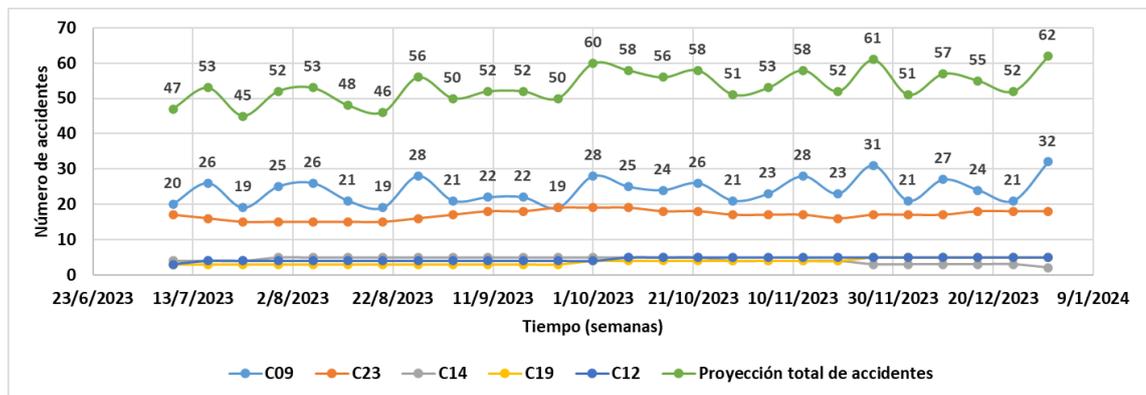


Figura 5. Valores proyectados de accidentes de tránsito, periodo julio – diciembre 2023.

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 ocurrieron; de manera que se pueda identificar en dónde ubicar los dispositivos de control o tomar medidas preventivas, por parte de las entidades de control.

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.

BIBLIOGRAFÍA:

- [1] S. Uğuz and E. Büyükgökoğlan, “A Hybrid CNN-LSTM Model for Traffic Accident Frequency Forecasting During the Tourist Season,” *Tehnicki Vjesnik*, vol. 29, no. 6, pp. 2083–2089, 2022, doi: 10.17559/TV-20220225141756.
- [2] OMS, “Muerte en la carretera,” 2022. <https://extranet.who.int/roadsafety/death-on-the-roads/?lang=es>
- [3] K. Santos, J. P. Dias, and C. Amado, “A literature review of machine learning algorithms for crash injury severity prediction,” *J Safety Res*, vol. 80, pp. 254–269, 2022, doi: 10.1016/j.jsr.2021.12.007.
- [4] P. Infante *et al.*, “Comparison of Statistical and Machine-Learning Models on Road Traffic Accident Severity Classification,” *Computers*, vol. 11, no. 5, pp. 1–12, 2022, doi: 10.3390/computers11050080.
- [5] F. Chellai, “Application of the Hybrid Forecasting Models to Road Traffic Accidents in Algeria,” *Statistika*, vol. 102, no. 2, pp. 184–197, 2022, doi: 10.54694/STAT.2021.37.
- [6] S. Mičić, R. Vujadinović, G. Amidžić, M. Damjanović, and B. Matović, “Accident Frequency Prediction Model for Flat Rural Roads in Serbia,” *Sustainability (Switzerland)*, vol. 14, no. 13, 2022, doi: 10.3390/su14137704.
- [7] Md. K. Islam, I. Reza, U. Gazder, R. Akter, M. Arifuzzaman, and M. M. Rahman, “Predicting Road Crash Severity Using Classifier Models and Crash Hotspots,” *Applied Sciences*, vol. 12, no. 22, p. 11354, 2022, doi: 10.3390/app122211354.
- [8] Y. Wang, J. Xu, X. Liu, Z. Zheng, H. Zhang, and C. Wang, “Analysis on Risk Characteristics of Traffic Accidents in Small-Spacing Expressway Interchange,” *Int J Environ Res Public Health*, vol. 19, no. 16, 2022, doi: 10.3390/ijerph19169938.
- [9] W. Li, X. Zhao, and S. Liu, “Traffic accident prediction based on multivariable grey model,” *Information (Switzerland)*, vol. 11, no. 4, pp. 1–12, 2020, doi: 10.3390/info11040184.
- [10] J. Ma, K. M. Kockelman, and P. Damien, “A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods,” *Accid Anal Prev*, vol. 40, no. 3, pp. 964–975, 2008, doi: 10.1016/j.aap.2007.11.002.
- [11] S. Zhang, A. Khattak, C. M. Matara, A. Hussain, and A. Farooq, “Hybrid feature selection-based machine learning Classification system for the prediction of injury severity in single and multiple-vehicle accidents,” *PLoS One*, vol. 17, no. 2 February, pp. 1–19, 2022, doi: 10.1371/journal.pone.0262941.
- [12] ANT, “Estadísticas de siniestros en Ecuador,” 2022. <https://www.ant.gob.ec/visor-de-siniestralidad-estadisticas/>
- [13] L. Shbeeb, “Road safety performance index: A tool for crash prediction,” *Cogent Eng*, vol. 9, no. 1, 2022, doi: 10.1080/23311916.2022.2124637.
- [14] J. Q. Li, W. Liu, X. Yang, P. Lu, and K. C. P. Wang, “Statistical Safety Performance Models considering Pavement and Roadway Characteristics,” *J Adv Transp*, vol. 2022, 2022, doi: 10.1155/2022/5871601.

- [15] Y. Ding, D. Li, M. Huang, X. Cao, and B. Tang, "Study on the influence of skid resistance on traffic safety of highway with a high ratio of bridges and tunnels," *Transportation Safety and Environment*, vol. 3, no. 4, 2021, doi: 10.1093/tse/tdab025.
- [16] M. Sun, R. Zhou, C. Jiao, and X. Sun, "Severity Analysis of Hazardous Material Road Transportation Crashes with a Bayesian Network Using Highway Safety Information System Data," *Int J Environ Res Public Health*, vol. 19, no. 7, 2022, doi: 10.3390/ijerph19074002.
- [17] H. Nassiri, S. I. Mohammadpour, and M. Dahaghin, "Forecasting time trend of road traffic crashes in Iran using the macro-scale traffic flow characteristics," *Heliyon*, vol. 9, no. 3, p. e14481, Mar. 2023, doi: 10.1016/j.heliyon.2023.e14481.
- [18] C. M. Ferreira-Vanegas, J. I. Velez, and G. A. Garcia-Llinas, "Analytical Methods and Determinants of Frequency and Severity of Road Accidents: A 20-Year Systematic Literature Review," *J Adv Transp*, vol. 2022, 2022, doi: 10.1155/2022/7239464.
- [19] H. R. Al-Masaeid and F. J. Khaled, "Performance of Traffic Accidents' Prediction Models," *Jordan Journal of Civil Engineering*, vol. 17, no. 1, Jan. 2023, doi: 10.14525/JJCE.v17i1.04.
- [20] B. Cai and Q. Di, "Different Forecasting Model Comparison for Near Future Crash Prediction," *Applied Sciences*, vol. 13, no. 2, p. 759, Jan. 2023, doi: 10.3390/app13020759.
- [21] N. Slimani, I. Slimani, M. Amghar, and N. Sbiti, "Road traffic forecasting using a real data set in Morocco," *Procedia Comput Sci*, vol. 177, pp. 128–135, 2020, doi: 10.1016/j.procs.2020.10.020.
- [22] Y. Wang, C. Yu, J. Hou, S. Chu, Y. Zhang, and Y. Zhu, "ARIMA Model and Few-Shot Learning for Vehicle Speed Time Series Analysis and Prediction," *Comput Intell Neurosci*, vol. 2022, pp. 1–9, Jan. 2022, doi: 10.1155/2022/2526821.
- [23] K. A. Getahun, "Time series modeling of road traffic accidents in Amhara Region," *J Big Data*, vol. 8, no. 1, p. 102, Dec. 2021, doi: 10.1186/s40537-021-00493-z.
- [24] C. C. Ihueze and U. O. Onwurah, "Road traffic accidents prediction modelling: An analysis of Anambra State, Nigeria," *Accid Anal Prev*, vol. 112, pp. 21–29, Mar. 2018, doi: 10.1016/j.aap.2017.12.016.
- [25] N. Deretić, D. Stanimirović, M. Al Awadh, N. Vujanović, and A. Djukić, "SARIMA Modelling Approach for Forecasting of Traffic Accidents," *Sustainability*, vol. 14, no. 8, p. 4403, Apr. 2022, doi: 10.3390/su14084403.
- [26] F. Chellai, "Application of the Hybrid Forecasting Models to Road Traffic Accidents in Algeria," *Statistika: Statistics and Economy Journal*, vol. 102, no. 2, pp. 184–197, Jun. 2022, doi: 10.54694/stat.2021.37.
- [27] M. Behzadi Goodari, H. Sharifi, P. Dehesh, M. A. Mosleh-Shirazi, and T. Dehesh, "Factors affecting the number of road traffic accidents in Kerman province, southeastern Iran (2015–2021)," *Sci Rep*, vol. 13, no. 1, p. 6662, Apr. 2023, doi: 10.1038/s41598-023-33571-8.

- [28] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*, 3rd ed., vol. 1. Melbourne, Australia: OTexts, 2021. Accessed: Aug. 07, 2023. [Online]. Available: <https://otexts.com/fpp3/index.html>
- [29] S. J. Taylor and B. Letham, "Forecasting at Scale," *Am Stat*, vol. 72, no. 1, pp. 37–45, Jan. 2018, doi: 10.1080/00031305.2017.1380080.
- [30] I. O. Olayode, L. K. Tartibu, and M. O. Okwu, "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*, vol. 6, p. 100095, Dec. 2021, doi: 10.1016/j.treng.2021.100095.
- [31] M. Bakhtiyari, M. R. Mehmandar, M. Khezeli, A. Latifi, T. A. Jouybari, and M. A. Mansournia, "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," *Int J Inj Contr Saf Promot*, vol. 26, no. 4, pp. 405–411, Oct. 2019, doi: 10.1080/17457300.2019.1660374.