



MODELO DE ANÁLISIS CORRELACIONAL PARA LA IDENTIFICACIÓN DE PATRONES ASOCIADOS A LOS ACCIDENTES DE TRÁNSITO DE YOPAL

Ángela Bibiana Oregón Fuentes, Juan David Castellom Rodríguez

**Fundación Universitaria de San Gil
Yopal, Colombia**

Resumen

Con el desarrollo del presente proyecto, se busca crear un modelo correlacional que permita identificar las características con mayor incidencia en los accidentes de tránsito en determinadas zonas de la ciudad de Yopal (Colombia), e indicar los patrones asociados a los mismos. En la construcción del modelo se aplican algoritmos de regresión lineal simple de aprendizaje supervisado, construyendo la correlación entre variables dependientes e independientes. El objetivo es obtener un modelo lineal que explique la relación entre las variables en toda la información de accidentes de tránsito en Yopal, calcular el coeficiente de determinación r^2 y su significancia (p -value).

El proyecto se lleva a cabo en cuatro fases: Construcción de aplicativo para la recolección de la información de los accidentes de tránsito según IPAT, medicina y Medicina Legal, en la segunda fase se identifican los diez puntos críticos de accidentalidad por georreferenciación, la tercera fase se caracteriza e identifica los factores de riesgos asociados a los accidentes de tránsito y cuarta fase, se Implementar Modelo Correlacional para la identificación de patrones asociados teniendo en cuenta la información recolectada en cada punto crítico

Como resultado se verificó la existencia de correlaciones significativas entre la variables que intervienen en los accidentes de tránsito como fecha, gravedad, edad, día de la semana, grado de alcoholimetría y se relaciona con las variables independientes de precipitaciones meteorológicas tomadas del data set proporcionado por la Fuerza Aérea de Casanare, adicionalmente se identifican los puntos con mayor accidentalidad del municipio y los factores mas relevantes por cada punto, permitiendo a los entes involucrados identificar las causas raíces de los eventos,

intervenirlos y evitar su reincidencia. Este proyecto se puede replicar en cualquier municipio de Colombia gracias al formato que se basa para la recolección (IPAT), documento reglamentado bajo la resolución 11268 de 2012 otorgada por el Ministerio de Transporte y se puede tomar de referente para la caracterización y correlación de los accidentes de tránsito a nivel mundial.

Palabras clave: movilidad vial; accidente de tránsito; minería de datos; modelo correlacional; aprendizaje supervisado; regresión lineal

Abstract

With the development of this project, it is sought to create a correlational model that allows identifying the characteristics with the highest incidence in traffic accidents in certain areas of the city of Yopal (Colombia), and indicating the patterns associated with them. In the construction of the model, simple linear regression algorithms of supervised learning are applied, constructing the correlation between dependent and independent variables. The objective is to obtain a linear model that explains the relationship between the variables in all the information on traffic accidents in Yopal, to calculate the coefficient of determination r^2 and its significance (p-value).

The project is carried out in four phases: Construction of an application for the collection of information on traffic accidents according to IPAT, medicine and Legal Medicine, in the second phase the ten critical accident points are identified by georeferencing, the third phase The risk factors associated with traffic accidents are characterized and identified, and the fourth phase, Implement a Correlational Model to identify associated patterns, taking into account the information collected at each critical point.

As a result, the existence of significant correlations between the variables that intervene in traffic accidents such as date, severity, age, day of the week, alcoholimetry degree was verified and it is related to the independent variables of meteorological rainfall taken from the data set provided by The Casanare Air Force also identifies the points with the highest accident rates in the municipality and the most relevant factors for each point, allowing the entities involved to identify the root causes of the events, intervene in them and avoid their recurrence. This project can be replicated in any municipality in Colombia thanks to the format that is based on the collection (IPAT), a document regulated under resolution 11268 of 2012 granted by the Ministry of Transportation and can be taken as a reference for the characterization and correlation of the traffic accidents worldwide.

Keywords: road mobility; car accident; data mining; correlational model; supervised learning; linear regression

1. Introducción

Hoy en día la seguridad vial es una de las principales prioridades en los gobiernos, la Agenda 2030 para desarrollo sostenible, creada por Asamblea General de las Naciones Unidas y adoptada en septiembre de 2015, plantea 17 objetivos con 169 metas, una de la más notables



consiste en: “reducir a la mitad, para el 2020, el número de muertes y lesiones causadas por accidentes de tránsito en todo el mundo”; de lo contrario, se prevé que para 2030 los accidentes de tránsito serán la séptima causa de defunción a nivel mundial.

Considerando la importancia del tema, la identificación de los patrones asociados a los accidentes de tránsito se ha convertido en uno de los objetivos para reducir el índice de mortalidad causado por los accidentes de tránsito a nivel mundial. Estos patrones de comportamiento y accidentes viales pueden ser útiles para desarrollar políticas de control de seguridad vial.

La revolución digital ha hecho posible que la información sistematizada sea fácil de capturar, procesar, almacenar, distribuir y transmitir, gracias a ello y con el auge de la tecnología grandes cantidades de información se han venido recogiendo y almacenando en bases de datos. *Machine Learning* (aprendizaje automático) es un método de análisis de datos, donde se utilizan algoritmos que puedan decir algo interesante en base a un conjunto de datos, el análisis de regresión permite establecer un modelo para la relación entre un cierto número de características y una variable objeto, la cual se puede emplear para la predicción de dichos accidentes de tránsito.

Dado lo anterior, el presente proyecto investiga la aplicación de algoritmo supervisado, implementando regresión lineal múltiple, empleando el lenguaje de programación Python, para construir modelos que permitan identificar las correlaciones significativas entre las variables asociadas a los accidentes de tránsito, con base en el histórico de accidentes ocurridos en Yopal durante los años 2017, 2018, 2019 y 2020, y el reporte meteorológico de Yopal.

Inicialmente se realiza una búsqueda de investigaciones referentes al tema analizando las técnicas y algoritmos utilizados para dar posible solución al objeto del presente proyecto, posteriormente se tomaron los datos de los Informes conocidos como IPAT (Informe Policial de Accidentes de Tránsito), obtenidos en la oficina de tránsito de Yopal, por otro lado se obtienen los datos de las personas fallecidas después de presentar un accidente de tránsito en Medicina Legal y el reporte meteorológico de las predicciones tomados de la estación de la Fuerza Aérea de Casanare; se identifica la clase, se procesan los datos, se realiza la limpieza de datos y se crean los archivos .csv de accidentes y precipitaciones, para que se puedan aplicar técnicas estadísticas y de minería de datos, posteriormente, se calculan regresiones lineal simple y análisis correlacional sobre los datos, utilizando Python y por último se analizan los resultados para obtener las conclusiones y recomendaciones.

En el presente artículo, primero se define el problema identificando la pregunta e hipótesis, luego se presentan la justificación delimitando su alcance, posteriormente se presenta el diseño metodológico y por último los resultados distribuidos en seis fases, teniendo en cuenta el proceso de extracción del conocimiento en minería de datos.

2. Problema de la Investigación

Según estadísticas reportadas por la gobernación de Casanare, en sus estudios: “Análisis de situaciones con el modelo determinantes sociales de Salud” y lesiones fatales en adolescentes de



Casanare, en la ciudad de Yopal los accidentes de tránsito son el problema de salud más frecuente. Hoy en día son la primera causa de muerte en el departamento, especialmente en jóvenes, por encima del conflicto armado, la delincuencia y todo tipo de enfermedades, incluidas SIDA.

Según el reporte del Observatorio de la Agencia de Seguridad Vial de Colombia (ASV) , Casanare ocupa el octavo lugar, de los departamentos en Colombia, con gran aumento de accidentes de tránsito, realizando la comparación entre años 2017 y 2018.

De enero a junio de 2019 el 51,67% de siniestros por accidentes de tránsito en Yopal son motociclistas. Casanare es el primer departamento con mayor tasa de morbilidad por cada 100.000 habitantes en Colombia, víctimas lesionadas valoradas por Medicina Legal; según estadísticas de los agentes de tránsito, los meses con mayor número de víctimas fatales en el 2019 es mayo y junio, los municipios con mayores tasas de morbilidad son Yopal, Paz de Ariporo y Aguazul.

Por otra parte, la Agencia de seguridad vial (ASV) posee un observatorio digital donde presenta ciertos informes brindados por Medicina Legal. No obstante, su información no es consistente con la real, debido al fallecimiento de heridos después del accidente en otros lugares donde fueron trasladados por la gravedad de sus accidentes, que han podido ocurrir en vías alternas al municipio donde en ocasiones son reportados a dos municipios.

El Ministerio de Transporte bajo la resolución 11268 del 06 de diciembre de 2012 adopta el nuevo formato del informe policial de accidentes de tránsito (IPAT), donde los agentes de tránsito de cada municipio deben reportarlos al Registro Único Nacional de Tránsito (RUNT). Sin embargo, en caso de que la oficina de tránsito municipal o las alcaldías necesiten algún reporte, su costo oscila entre un millón y tres millones de pesos, dependiendo del nivel de detalle de la información.

Por lo plasmado anteriormente, no se cuenta con una herramienta o recurso tecnológico que permita realizar un análisis de toda la información consolidada y por tanto permita identificar los factores de riesgos o las características que ocasionan los accidentes de tránsito en la ciudad, para poder presentarlo a los entes responsables con el fin de que se considere en la toma de decisiones y por tanto se ejecuten las medidas requeridas.

3. Diseño Metodológico

Para esta investigación se emplea el método científico, a través de un diseño de investigación experimental con enfoque cuantitativo y alcance descriptivo el cual consiste en la caracterización de los accidentes de tránsito en el municipio de Yopal y su posible correlación, con el fin de establecer su estructura o comportamiento.

Se observa y cuantifica el comportamiento vial de los habitantes del municipio de Yopal de las diferentes variables asociados a los accidentes de tránsito. Es decir, cada característica o variable se analiza de forma autónoma o independiente. Los tipos de investigación que se emplean, como se especificaba anteriormente, son cuantitativo y comparativo ya que se hace necesario recolectar



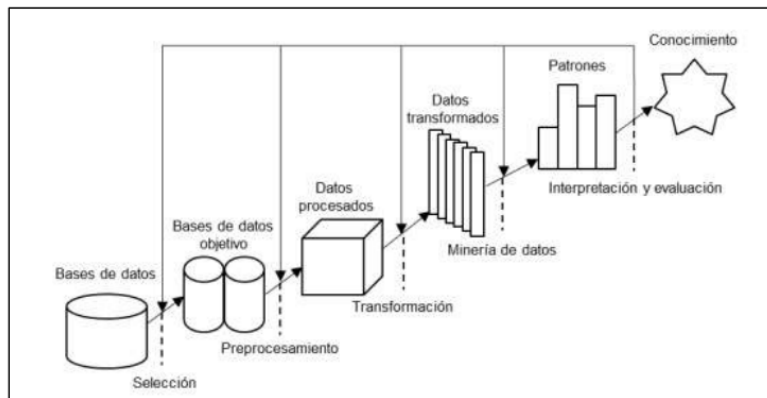
información que permita hacer una hipótesis y conocer la totalidad de la población a investigar, para evitar futuros accidentes de tránsito.

La población tomada para la realización de la investigación, son los accidentes de tránsito ocurridos desde enero del año 2017 hasta agosto de 2020, dados por 1394 accidentes.

Para Identificación de puntos y caracterización se emplea una técnica de Muestreo Probabilística de Conglomerados, donde se toman el total de los accidentes de tránsito desde enero del año 2017 hasta agosto de 2020, dados por 1394 accidentes. Para el Análisis Correlacional, se emplea el muestreo no probabilístico por conveniencia, donde la muestra es calculada.

Las técnicas de recolección de Información se realizaron a través de observación directa, encuestas, software RECOATY, revisión documental, informes y observación directa. Para la creación del software de recolección de datos se emplea la metodología xp y `paradigma en cascada, para la georreferenciación, identificación de características y modelo correlacional se emplea la metodología KDD de Minería de Datos teniendo en cuenta las etapas de selección, pre procesamiento, transformación, minería de datos (modelización), e interpretación y evaluación.

Figura 1. Etapas de la Metodología KDD



Nota: En la Figura 12 muestra cuales son las etapas que se utilizan en la metodología KDD, comenzando con la selección de datos, el pre procesamiento, la transformación, aplicación de minería de datos e interpretación y evaluación.

En la realización del modelo correlacional se emplea regresión lineal simple, método estadístico que trata de modelar la relación entre una variable dependiente y una variable independiente, mediante el ajuste de una ecuación lineal. Se realiza regresión lineal simple con una variable independiente.

El lenguaje de programación empleado fue Python, dos de las implementaciones de modelos de regresión lineal más utilizadas en Python son: *scikit-learn* y *statsmodels*. Aunque ambas están muy optimizadas, Scikit-learn está orientada principalmente a la predicción, Statsmodels es mucho más, por tal motivo es el empleado en el presente proyecto.



El modelo de regresión lineal (Legendre, Gauss, Galton y Pearson) considera que, dado un conjunto de observaciones $\{y_i, x_{i1}, \dots, x_{ip}\}_{i=1}^n$, la media μ de la variable respuesta y se relaciona de forma lineal con la o las variables regresoras $x_1 \dots x_p$ acorde a la ecuación.

$$\mu y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad \mu y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + e$$

El resultado de esta ecuación se conoce como la línea de regresión poblacional, y recoge la relación entre los predictores y la media de la variable respuesta. Donde:

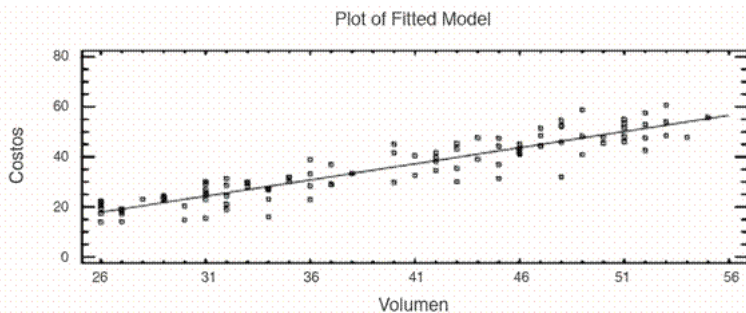
β_0 : es la coordenada en el origen, se corresponde con el valor promedio de la variable respuesta y cuando todos los predictores son cero.

β_j : es el efecto promedio que tiene sobre la variable respuesta el incremento en una unidad de la variable predictora x_j , manteniéndose constantes el resto de variables. Se conocen como coeficientes parciales de regresión.

e : es el residuo o error, la diferencia entre el valor observado y el estimado por el modelo. Recoge el efecto de todas aquellas variables que influyen en y , pero que no se incluyen en el modelo como predictores.

Y se conoce como regresión simple el cálculo de la ecuación correspondiente a la línea que mejor describe la relación entre la respuesta y la variable que la explica. Dicha ecuación representa la línea que mejor se ajusta a los puntos en un gráfico de dispersión como se puede observar en la Figura 1.

Figura 2. Regresión lineal Simple Metodología KDD



Nota: Representación de una ecuación de regresión simple para la relación entre costos y volumen.

A partir de la regresión lineal es posible hacer predicciones sobre la respuesta con base en valores de la variable predictora. Una vez ajustado el modelo, es necesario verificar su utilidad ya que, aun siendo la línea que mejor se ajusta a las observaciones de entre todas las posibles, puede tener un gran error. Las métricas más utilizadas para medir la calidad del ajuste son: el error estándar de los residuos y el coeficiente de determinación R^2 , el cual describe la proporción de varianza de la variable respuesta explicada por el modelo y relativa a la varianza total. Su valor está acotado entre 0 y 1.

El modelo generado, es una estimación de la relación poblacional a partir de la relación que se observa en la muestra y , por lo tanto, está sujeta a variaciones. Para cada uno de los coeficientes



de la ecuación de regresión lineal (β_j) se puede calcular su significancia (p -value) y su intervalo de confianza. El test estadístico utilizado en el presente proyecto es el t-test.

El test de significancia para los coeficientes (β_j) del modelo lineal considera como hipótesis:

H_0 : el predictor x_j no contribuye al modelo ($\beta_j=0$), en presencia del resto de predictores. En el caso de regresión lineal simple, se puede interpretar también como que no existe relación lineal entre ambas variables por lo que la pendiente del modelo es cero $\beta_j=0$.

H_a : el predictor x_j sí contribuye al modelo ($\beta_j \neq 0$), en presencia del resto de predictores. En el caso de regresión lineal simple, se puede interpretar también como que sí existe relación lineal entre ambas variables por lo que la pendiente del modelo es distinta de cero $\beta_j \neq 0$.

La aplicación del modelo de regresión lineal descrito anteriormente, así como hallar el valor del R^2 y p-value y su interpretación, se especifica en el capítulo desarrollo de la investigación del presente documento.

4. Descripción de la Investigación

La investigación se realiza a través de un proceso de extracción del conocimiento teniendo en cuenta 5 fases:

Fase 1. Selección de los Datos a Través del Rediseño, Actualización del Aplicativo RECOATY Digitalización de la Información. Se realiza la siguientes actividades:

- Recolección de requerimientos funcionales y no funcionales.
- Recolección de la información de los diferentes entes involucrados.
- Levantamiento de Requerimientos funcionales y no funcionales para RECOATY Versión 2.0
- Análisis, diseño bajo metodología XP, del aplicativo RECOATY en su Versión 2.0.
- Implementación de RECOATY V.2.0
- Resultados RECOATY V.2.0

Escaneo y digitalización de los accidentes de todos los IPAT desde enero de 2018 a agosto de 2020.

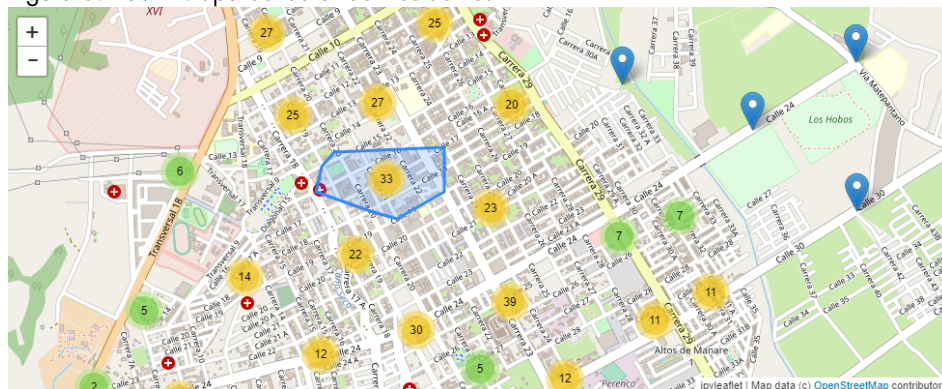
Pruebas de aceptación y validación de la nueva Versión RECOATY V.2.0

Fase 2. Pre procesamiento De La Información y Transformación de los Datos. se aplica técnicas de pre procesamiento, esta técnica fue realizada, en el proyecto, de manera artesanal y manual. Algunos datos se procesaron utilizando formulas avanzadas y tablas dinámicas en Excel, sin embargo, para un futuro se recomienda aplicar técnicas ETL y poder sistematizar el proceso. Las actividades realizadas son:

Fase 3. Identificación de los Puntos Críticos de Accidentalidad por Georreferenciación. Una vez se obtiene la vista minable, *Dataset*, se proceden a identificar los puntos críticos de accidentalidad por georreferenciación,



Figura 3. Zoom Mapa de Calor de Accidentes



Nota. Se observan los accidentes por cada zona en el municipio, identificando las zonas con mayor número de accidentes.

Fase 4. Caracterización e Identificación de los Factores de Riesgos Asociados a los Accidentes de Tránsito. Tomando los diez puntos con mayor índice de accidentalidad se procede a identificar las cinco características con más recurrencia en cada punto y poder identificar su porcentaje de ocurrencia. Para ello se realizan las siguientes actividades en el aplicativo desarrollado en Python:

Fase 5. Implementar Modelo Correlacional para la Identificación de Patrones Asociados Teniendo en Cuenta la Información Recolectada en Cada Punto Crítico Empleando Regresión Lineal Simple. Dada la naturaleza claramente categorial con la que se configuraron los diferentes niveles de las variables en juego, así como su distinción entre variables dependientes o criterio y variables predictivas o independientes, se ha desarrollado para la consecución del fin propuesto una regresión lineal simple, dando como resultado la identificación de patrones repetitivos asociados a una característica determinada en los puntos con mayor accidentalidad, identificando el valor de correlación R^2 y P-value.

Fase 6. Interpretación de Resultados. Según la información obtenida en el modelo correlacional se interpreta los resultados obtenidos para analizar el informe técnico final que será presentado como insumo en la toma de decisiones de las secretarías de tránsito.

5. Interpretación de Resultados

Una de las hipótesis iniciales formuladas a partir de la observación de los eventos de accidentes de tránsito en la ciudad de Yopal, establecía que bajo condiciones de lluvia existía una mayor probabilidad de ocurrencia de accidentes de tránsito, así como una mayor gravedad en las lesiones de las personas involucradas tanto en el casco urbano como en las vías, debido a las condiciones estructurales de las mismas y a los bajos niveles de visibilidad, no obstante una vez analizada la información, esta hipótesis carece de validez, debido a que los accidentes en condiciones de lluvia no son significativos en relación a los accidentes de tránsito presentados en clima normal o soleado, así como las lesiones graves tampoco representan mayor incidencia ante estos eventos.



Por otra parte, según encuesta inicial realizada a los agentes de tránsito en el municipio de Yopal, a la pregunta ¿Cuáles considera que son las tres principales razones de la accidentalidad de Yopal?, el 100% de los agentes de tránsito coincidieron en que la imprudencia es una de las principales razones, de igual forma el 80% de los agentes indicaron que conducir en estado de embriaguez y el 50% manifestaron que el exceso de velocidad puede considerarse como una de las principales razones de accidentalidad. Para la presente investigación, variables como la velocidad al momento del accidente y el nivel de imprudencia de los conductores no fueron consideradas debido a que actualmente no se cuenta con los instrumentos o recurso necesarios para determinar estas variables, sin embargo si se cuenta con información relacionada con el grado de embriaguez al momento del accidente como resultado del registro de las pruebas de alcoholemia cuando esta es realizada, sin embargo el valor R^2 es considerablemente inferior a 0.7, por tanto la correlación entre grado de embriaguez y cantidad de accidentes no es significativa, por lo cual esta hipótesis también carece de validez.

En el mismo estudio, a la pregunta ¿Cuáles considera que son los cinco lugares con mayor índice de accidentalidad en Yopal?, el 80% de los accidentes de tránsito coincidieron en la Cl 13 Cr 27, el 50% afirmaron que la TV 15 Cr 35 se encuentra en este rango, de igual forma el 30% de los agentes indicaron que son las Cl 40 Cr 11 y Cr 21 Cl 10. Una vez identificadas las ubicaciones con mayor índice de accidentalidad, acorde a los registros del Dataset, se establece que la Cl 13 Cr 27 se encuentra en la novena posición con un total de 14 accidentes registrados en el IPAT en el periodo comprendido entre 2017 y 2020, mientras que la TV 15 Cr 35 y la Cl 40 Cr 11 no se encuentra en las diez primeras posiciones identificadas por el aplicativo, por otra parte la Cr 21 Cl diez se identificó que para el periodo comprendido entre los años 2017 – 2018 se encuentra en la quinta posición de ubicaciones con mayor índice de accidentalidad en el Municipio, es importante señalar que las vías cercanas al municipio de Yopal, no fueron consideradas en la encuesta realizada a los agentes de tránsito.

De igual manera a la pregunta ¿Cuál tipo de vehículo considera como el mayor causante de accidentes de tránsito?, el 100% de los agentes de tránsito respondieron que son las motocicletas, hipótesis que validada con la matriz correlacional representa un valor de $R^2 = 0.88$, convirtiéndola en una hipótesis válida para la presente investigación.

Finalmente, para la pregunta ¿En qué porcentaje cree usted que los implicados en los accidentes poseen la documentación reglamentaria al día?, el 50% de los agentes de tránsito indicaron que consideran que 25% - 50% de los implicados poseen la documentación reglamentaria al día, hipótesis que se convierte en no válida debido a que los valores de R^2 de las variables que indican si se cuenta con Revisión Tecno mecánica y SOAT vigentes al momento del accidentes son iguales a 0.72 y 0.98 respectivamente una vez se correlacionan con la variable de cantidad de accidentes, representando una correlación significativa y rechazando la correspondiente hipótesis.

Al momento de identificar las ubicaciones con mayor cantidad de accidentes se evidenció que la vías cercanas al municipio de Yopal que conectan con otras localidades representan las primeras cuatro posiciones para los accidentes de tránsito correspondientes al periodo 2017 – 2018 y las primeras cinco posiciones para el periodo 2019 – 2020, entre las cuales resaltan la Vía Marginal de la Selva, Vía Yopal – Paz de Ariporo y Vía Yopal – Tilodirán, las cuales coinciden en los tres



primeros lugares para cada uno de los periodos analizados con un total de 84 accidentes y 163 vehículos involucrados en la Vía Marginal de la Selva, 44 accidentes de tránsito y 79 vehículos involucrados en la Vía que de Yopal conduce a Paz de Ariporo y 33 accidentes de tránsito con 59 vehículos involucrados para la Vía Yopal – Tilodirán, correspondientes al periodo total objeto de la presente investigación, por tanto es importante que la Administración Municipal y las Secretarías de Tránsito Municipal y Departamental, con base en los resultados obtenidos en la presente investigación exploren soluciones para mitigar la ocurrencia de accidentes de tránsito en las correspondientes vías.

Debido a que la ciudad de Yopal ha presentado a partir del año 2019 una variación en el sentido de las vías en el casco urbano, obedeciendo a directrices de la Alcaldía del Municipio y a la Secretaría de Tránsito, con el fin de mejorar el flujo del tránsito principalmente en horas pico, se esperaba una variación en el comportamiento de los accidentes de tránsito en el casco urbano, por tal razón se realizó una división en dos secciones: Accidentes de tránsito 2017 – 2018 y Accidentes de tránsito 2019 – 2020, para el Dataset inicialmente propuesto. No obstante una de las ubicaciones cuyo sentidos fueron modificados tanto en la Carrera como en la Calle: Cl 13 Cr 27, en los dos periodos presentó exactamente el mismo comportamiento: se ubicó en el noveno lugar de las diez primeras ubicaciones, con un total de 7 accidentes de tránsito para cada uno de los periodos, de igual forma en los resultados de caracterización se evidencian coincidencias respecto a las variables que representan los mayores valores, de esta manera se resalta que para los dos periodos se presentaron choques con vehículos en clima normal o soleado, en los cuales la mayor cantidad de conductores pertenecen al género masculino, sin embargo se evidencia que posterior al cambio de las vías la cantidad de vehículos involucrados se incrementó de 14 en el periodo 2017-2018 a 20 en el periodo 2019-2020, así como la gravedad de accidentes, en la cual en el periodo 2017-2018 correspondía a 71.43% de vehículos con daños materiales y conductores sin lesiones, mientras que en el periodo 2019-2020 el 60% de las lesiones fueron graves, con lo cual es válido afirmar que la condiciones de accidentalidad no mejoraron con la modificación en el sentido de las vías.

Por otra parte, para la generación inicial del análisis correlacional y la subsecuente regresión lineal simple, se debió buscar un nuevo enfoque con la reestructuración del Dataset, basado en los resultados obtenidos respecto a la ubicación de los accidentes de tránsito y a la caracterización previa, debido a que el Dataset inicial no permitía la identificación de valores de R^2 superiores a 0.7 y P-value cercanos a cero, por tal razón se generó un nuevo Dataset cuya generación del análisis correlacional, las regresiones lineales simples y sus predicciones fueron más significativas y coherentes con los análisis previos.

Una vez definido el modelo correlacional, se identificaron las combinaciones de variables significativas entre las que se destacan: Cantidad de accidentes y lesiones graves, Cantidad de accidentes en motocicleta y lesiones graves, Cantidad de accidentes con clima normal y lesiones graves, Cantidad de hombres accidentados y lesiones graves, Cantidad de personas accidentadas entre 30 y 34 años y la cantidad de accidentes.

En el ejercicio y prueba de las correspondientes predicciones se evidenció coherencia en los resultados, así como la generación de predicciones dinámicas al modificar los parámetros de la



variable independiente para cada una de las regresiones lineales. Por tal razón se puede afirmar que el modelo correlacional puede ser aplicado a otras combinaciones de variables con otros Dataset que consideren la misma estructura con el fin de realizar análisis de otros periodos de tiempo acorde se va registrando la información en el aplicativo RECOATY, convirtiendo el instrumento en una herramienta tecnológica que permita a futuras administraciones tomar decisiones con el fin de mitigar los índices de accidentalidad en el municipio de Yopal, Casanare.

6. Referencias

Artículos de revistas

- Carrasquilla-Batista, A. Chacón-Rodríguez, K. Núñez-Montero, O. Gómez-Espinoza, J. Valverde-Cerdas, and M. Guerrero-Barrantes, "Regresión lineal simple y múltiple: aplicación en la predicción de variables naturales relacionadas con el crecimiento microalgal," *Rev. Tecnol. en Marcha*, vol. 29, no. 8, p. 33, 2016, doi: 10.18845/tm.v29i8.2983.
- Y. Rincon Acevedo, "Lesiones fatales en adolescentes, Casanare-Colombia 2011-2013," *Rev. Médica Risaralda*, vol. 22, no. 1, pp. 18–29, 2017, doi: 10.22517/25395203.13641.
- J. Burgos, F.; Michilena, "IMPLEMENTACIÓN DE UN SISTEMA DE GESTIÓN BASADO EN LA METODOLOGÍA KDD DE MINERÍA DE DATOS PARA EL PROCESO DE PLANIFICACIÓN DE LA PRODUCCIÓN DE LA INDUSTRIA FARMACÉUTICA ECUATORIANA.," *La Evasión Tribut. E Incid. En La Recaud. Del Impuesto a La Renta Pers. Nat. En La Prov. Del Guayas, Periodo 2009-2012*

Fuentes electrónicas

- CONPES, "Política Nacional de Seguridad Digital," CONPES 3854 - Política Nac. Secur. Digit., p. 91, 2016, [Online].
- Available: <https://colaboracion.dnp.gov.co/CDT/Conpes/Econ%C3%B3micos/3854.pdf>
- D. Montoro, "Capítulo 9. Regresión lineal simple," *Regresión Lineal*, p. 11, 2015, [Online]. Available: <http://www4.ujaen.es/~dmontoro/Metodos/Tema%209.pdf>
- Joaquín Amat Rodrigo, "Regresión lineal con Python," 2020. <https://www.cienciadedatos.net/documentos/py10-regresion-lineal-python.html>.
- O. N. de S. Vial, "OBSERVATORIO NACIONAL DE SEGURIDAD VIAL," marzo, 2018. <http://ansv.gov.co/observatorio>.

Sobre los autores

- **Ángela Bibiana Oregón Fuentes:** Ingeniera de Sistemas, Especialista en telemática y Negocios por Internet, Máster en Tecnologías de la Información. Directora programa Ingeniería de Sistemas, Fundación Universitaria de San Gil UNISANGIL. abortegon@unisangil.edu.co
- **Juan David Castellom Rodríguez:** Ingeniero de Sistemas, Máster en Tecnologías de la Información. Fundación Universitaria de San Gil UNISANGIL. jcastellom@unisangil.edu.co



Los puntos de vista expresados en este artículo no reflejan necesariamente la opinión de la Asociación Colombiana de Facultades de Ingeniería.

Copyright © 2021 Asociación Colombiana de Facultades de Ingeniería (ACOFI)

