**Análisis de la gravedad de accidentalidad y tipos de vehículos involucrados en la ciudad de Bucaramanga**

**Analysis of the severity of accidents and types of vehicles involved in the city of Bucaramanga**

**Cristian Camilo Franco, Julián Felipe Terangi Rueda Acosta**

**Resumen**

Este estudio aborda la problemática de los accidentes de tránsito en Bucaramanga, Colombia, con el objetivo de analizar cómo el tipo de vehículo involucrado afecta la gravedad de los incidentes. Se destaca la relevancia de conocer estos factores para diseñar estrategias de prevención efectivas y mejorar la seguridad vial. Utilizando datos de la Dirección de Tránsito y la metodología CRISP-DM, se investigó la correlación entre los tipos de vehículos y la gravedad de los accidentes. Los resultados indican que la agrupación aglomerativa supera a KMeans y DBSCAN, identificando características esenciales que influyen significativamente en la predicción de la gravedad de los accidentes, como "motocicleta" y "peatón". A pesar de las limitaciones de despliegue, los resultados tienen el potencial de informar políticas de seguridad vial más precisas y estratégicas, revelando oportunidades de mejora en Bucaramanga a través de un enfoque integral para reducir los accidentes de tráfico y mejorar la calidad de vida de la comunidad.

**Palabras clave: Accidentes de tránsito, Bucaramanga, Colombia, Seguridad vial, Agrupación aglomerativa, Prevención de accidentes**

**Abstract**

This study addresses the issue of traffic accidents in Bucaramanga, Colombia, aiming to analyze how the type of vehicle involved affects the severity of incidents. The relevance of understanding these factors to design effective prevention strategies and improve road safety is emphasized. Using data from the Traffic Department and the CRISP-DM methodology, the correlation between vehicle types and accident severity was investigated. Results indicate that agglomerative clustering outperforms KMeans and DBSCAN, identifying essential features that significantly impact the prediction of accident severity, such as "motorcycle" and "pedestrian." Despite deployment limitations, the results have the potential to inform more precise and strategic road safety policies, revealing opportunities for improvement in Bucaramanga through a comprehensive approach to reduce traffic accidents and enhance community quality of life.

**Keywords: Traffic accidents Bucaramanga, Colombia, Road safety, Agglomerative clustering, Accident preventio**

**1. Introducción**

Los accidentes de tráfico afectan considerablemente a la seguridad vial y a la calidad de vida de las comunidades. En Bucaramanga, Colombia, es crucial comprender los factores que contribuyen a la gravedad de las colisiones de tráfico para diseñar estrategias eficaces de prevención y seguridad vial [1].

Según un informe de la Organización Mundial de la Salud (OMS), los accidentes de tráfico se cobran alrededor de 1,3 millones de vidas al año en todo el mundo, lo que pone de relieve la necesidad crucial de abordar la seguridad vial [2].

Asimismo, el reporte de la Secretaría de Tránsito de Bucaramanga, en 2020 se presentaron 2.268 accidentes [3]. El año 2021 registró un aumento del 22% con respecto al mismo periodo de 2020, con un total de 2.762 accidentes de tránsito. De ellos, 73 se saldaron con víctimas mortales, lo que supone un aumento del 18% respecto al año anterior. Además, se produjeron 1.389 accidentes con heridos, un 9% más que en 2020, y 1.300 accidentes que solo causaron daños materiales, lo que representa un aumento del 39% respecto al año anterior.

Este informe investiga cómo el tipo de vehículo implicado en los accidentes afecta a su gravedad, concretamente sin tener en cuenta variables temporales. Analizaremos un amplio conjunto de datos que contiene información detallada sobre diversos aspectos de los accidentes, desde el tipo de vehículo hasta la gravedad de las consecuencias. Este método nos permitirá detectar patrones y relaciones significativas que son cruciales para mejorar las políticas de seguridad vial en la zona.

Bucaramanga cuenta con una amplia gama de vehículos presentes en sus vías, desde peatones y motocicletas hasta automóviles y camiones de carga pesada. Esta variación en los tipos de vehículos lleva a preguntarse si determinados tipos de vehículos son más propensos a causar accidentes graves que otros.

La principal variable bajo análisis es el "Gravedad", ya que es posible investigar cómo varía los tipos de vehículos en los accidentes cuando están involucrados peatones, automóviles, camiones, motocicletas u otro tipo de vehículos.

La correlación entre el tipo de vehículos implicados en un accidente y la gravedad del impacto es un factor crítico que requiere más investigación. Es imperativo comprender cómo la presencia de vehículos específicos puede intensificar o aliviar la gravedad de las colisiones para aplicar eficazmente medidas preventivas.

**Pregunta de investigación:** ¿La gravedad de los accidentes está significativamente influenciada por el tipo de vehículo involucrado?

Esta cuestión es crítica porque tiene el potencial de identificar factores determinantes que pueden ser objeto de intervenciones específicas y políticas de seguridad vial. Además, tiene un impacto directo en la formulación de políticas de seguridad vial. Si se encuentra una correlación significativa entre el tipo de vehículo y la gravedad de los accidentes, las autoridades locales y nacionales pueden centrar sus esfuerzos en regular y aumentar la concienciación sobre determinados tipos de vehículos.

**2. Metodologia**

El estudio emplea la metodología CRISP-DM consta de seis etapas, según [4]:

Inicialmente, la comprensión del negocio implica la generación de una estrategia de proyecto, la especificación de los objetivos de la minería de datos y la evaluación de las necesidades y objetivos de la empresa.

La etapa siguiente es la comprensión de los datos, en la que se recopilan los datos disponibles, se examinan en busca de problemas de calidad y se seleccionan las variables pertinentes para el análisis.

La preparación de los datos consiste en limpiarlos y modificarlos para garantizar que cumplen las normas de calidad y se ajustan a los objetivos del proyecto.

El proceso de modelización incluye la selección de un enfoque de aprendizaje automático, el diseño y la construcción del modelo y la evaluación de su eficacia.

Por último, la evaluación se realiza utilizando métricas y procedimientos de validación para garantizar la precisión. Si es necesario, pueden introducirse modificaciones en el modelo. También, es necesario realizar un seguimiento y actualizaciones frecuentes durante el despliegue, que implica la aplicación del modelo en un entorno real para realizar predicciones o tomar decisiones.

**3. Resultados**

**3.1. Compresión del negocio**

El estudio pretende comprender cómo influye el tipo de vehículo en la gravedad de los accidentes en Bucaramanga, Colombia, para el diseño de estrategias de seguridad vial. El examen pretende reconocer patrones significativos que ayuden a mejorar las políticas de prevención. La cuestión crucial es si la gravedad de un accidente se ve influida por el tipo de vehículo implicado. Esta información podría conducir a intervenciones específicas y políticas más eficientes con el fin de disminuir la ocurrencia y gravedad de los accidentes de tráfico en la región.

**3.1.1. Objetivos del negocio**

El objetivo del negocio es mejorar la seguridad vial en Bucaramanga mediante el estudio del impacto de los diferentes tipos de vehículos en la gravedad de los accidentes de tránsito. La investigación busca identificar patrones significativos para diseñar estrategias preventivas más efectivas.

En referencia al análisis DAFO, los puntos fuertes de la investigación incluyen su carácter exhaustivo y el acceso a datos detallados, mientras que existen oportunidades potenciales en el desarrollo de políticas más eficaces basadas en sus conclusiones.

Las limitaciones pueden surgir si la correlación entre los tipos de vehículos y la gravedad de las colisiones no es tan significativa como se esperaba, restringiendo así la eficacia de las políticas propuestas.

Las posibles amenazas incluyen la inacción ante los resultados o la aplicación de políticas ineficaces, lo que podría comprometer la seguridad vial y la aceptación de nuevas medidas si los resultados no son concluyentes.

**Diagrama

Descripción generada automáticamente**

**Figura 1.** Matriz DOFA

El análisis DAFO pone de manifiesto que investigar el impacto de los diferentes tipos de vehículos en la gravedad de los accidentes en Bucaramanga presenta importantes oportunidades para elaborar políticas de seguridad vial más precisas y eficientes. No obstante, existe el riesgo de que las políticas sean ineficaces si la relación entre los tipos de vehículos y la gravedad de los accidentes sigue sin ser concluyente, lo que pone de relieve la necesidad de iniciativas ágiles y estratégicas para mejorar la seguridad vial de la ciudad.

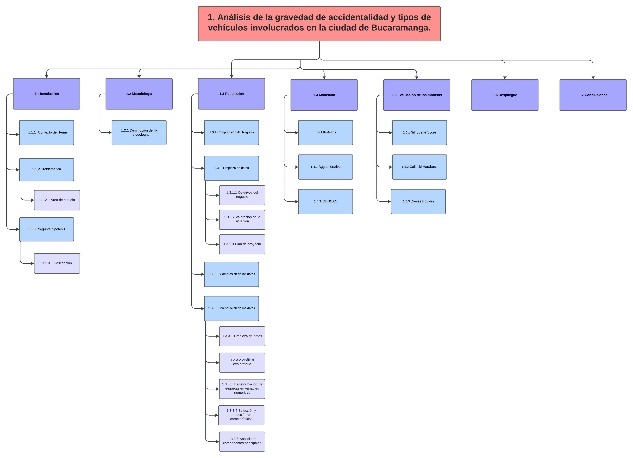
**3.1.2. Valoración de la situación**

Un plan centrado en la seguridad vial investiga cómo influyen los distintos tipos de vehículos en la gravedad de las colisiones, reduciendo potencialmente las lesiones, los daños materiales y la pérdida de vidas. Al identificar patrones, se pueden desarrollar estrategias preventivas más eficaces, contribuyendo a un entorno vial más seguro y a una mejor calidad de vida para la comunidad de Bucaramanga a largo plazo. Esta estrategia tiene el potencial de aplicar políticas específicas destinadas a mitigar los riesgos y mejorar la seguridad vial, lo que puede producir ventajas en forma de reducción de costes y mayor protección de la población.

**Tabla 1.** Plan de beneficios

**3.1.3. Plan de proyecto**

El estudio sobre el impacto de los distintos tipos de vehículos en la gravedad de las colisiones pone de manifiesto la comprensión del negocio, mientras que la fase de comprensión de los datos se ejemplifica mediante un análisis exhaustivo de los mismos. La meticulosa preparación de los datos para descubrir las correlaciones entre los tipos de vehículos y la gravedad de las colisiones coincide con la fase de preparación de datos de CRISP-DM. La identificación de patrones y relaciones es similar a la fase de modelización, mientras que la mejora de las políticas de seguridad vial constituye la fase de evaluación.



**Figura 2.** Estructura de desglose del trabajo

La conexión con la Estructura de Desglose del Trabajo (EDT) implica la organización de tareas y actividades para comprender, examinar, crear, ejecutar y evaluar planes que aborden la correlación entre los tipos de vehículos y la gravedad de los accidentes de tráfico. De este modo se promueve una gestión bien organizada y eficaz de este problema concreto de seguridad vial. Además, la EDT ayuda a determinar los recursos esenciales para ejecutar estos planes y a crear un calendario de actividades que garantice su ejecución puntual. Este enfoque ayuda a clarificar las responsabilidades y a definir las funciones de cada miembro del equipo responsable de aplicar las medidas de seguridad vial.

**3.2. Compresión de los datos**

Los datos que sustentan este análisis provienen de la página de datos abiertos de Colombia se encuentran en el siguiente link:https://www.datos.gov.co/Transporte/03-ACCIDENTES-DE-TRANSITO-DESDE-ENERO-2012-A-FEBRE/7cci-nqqb. En la tabla 2 se observa las columnas y descripción que el conjunto de datos proporciona.

**Tabla 2**. Detalle conjunto de datos

|  |  |
| --- | --- |
| **Nombre de la columna** | **Descripción** |
| **ORDEN** | Identificador único asignado a cada registro de accidente para facilitar su seguimiento y referencia. |
| **FECHA** | Fecha exacta en la que tuvo lugar el accidente, proporcionando información temporal crucial para el análisis. |
| **AÑO** | Año en el que ocurrió el accidente, ofreciendo una categorización temporal para tendencias y comparaciones a lo largo del tiempo. |
| **MES** | Mes específico en el que se registró el accidente, permitiendo análisis estacionales y variaciones mensuales. |
| **DÍA** | Día de la semana en que ocurrió el accidente, proporcionando una perspectiva sobre patrones asociados con días particulares. |
| **GRAVEDAD** | Indicador que refleja la severidad del accidente, clasificándolo en categorías como Con Heridos, Solo Daños, etc. |
| **PEATON** | Cantidad de peatones involucrados en el accidente, destacando la vulnerabilidad de esta categoría de usuarios de la vía. |
| **AUTOMOVIL, CAMPAERO, CAMIONETA, MICRO, BUSETA, BUS, CAMION, VOLQUETA, MOTO, BICICLETA, OTRO** | Números específicos para cada tipo de vehículo involucrado en el accidente, proporcionando una visión detallada de la composición del incidente. |
| **VIA 1** | Identificación de la vía principal donde ocurrió el accidente, contextualizando la ubicación geográfica. |
| **BARRIO** | Denominación del barrio o ubicación exacta donde se produjo el accidente, ofreciendo detalles geográficos adicionales. |
| **HORA** | Momento del día en que ocurrió el accidente, permitiendo análisis de patrones de acuerdo con las horas del día |
| **ENTIDAD** | Organización o entidad encargada de atender el accidente, proporcionando información sobre la respuesta y gestión del incidente. |
| **nombrecomuna** | Nombre de la comuna en la que se localiza el accidente, contextualizando geográficamente la ubicación. |
| **Propietario de Vehículo** | Información sobre el propietario del vehículo involucrado en el accidente, aportando detalles sobre la responsabilidad y titularidad del vehículo. |
| **DIURNIO/NOCTURNO** | Indicador que especifica si el accidente ocurrió durante el día o la noche, permitiendo análisis basados en la iluminación y visibilidad. |
| **HORA RESTRICCION MOTO** | Detalles sobre restricciones específicas para motocicletas en cuanto a la hora, proporcionando información adicional sobre regulaciones y normativas locales |

Este conjunto de datos, que cubre un período de tiempo significativo, proporciona una visión detallada de las circunstancias y consecuencias de los accidentes en la región. La riqueza y diversidad de esta información brinda una oportunidad única para explorar más a fondo los factores que contribuyen a la gravedad de los incidentes viales.

**3.3. Preparacion de los datos**

Inicialmente, se carga el conjunto de datos y se visualizan las primeras filas para obtener una impresión inicial. A continuación, se obtiene información general sobre el conjunto de datos, se lista cada columna presente en el dataset y se realiza un resumen estadístico de las variables numéricas para comprender sus distribuciones y tendencias.

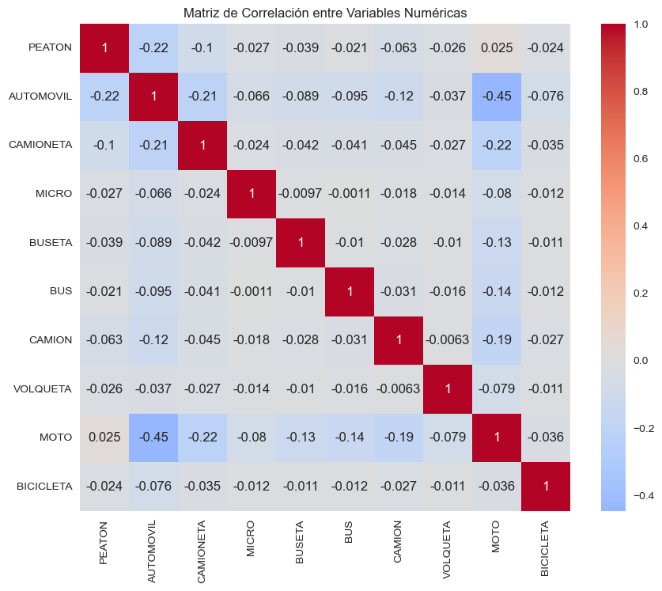
La identificación de variables categóricas y numéricas orienta el análisis, incluyendo el conteo de valores únicos en columnas categóricas. Además, se lleva a cabo un análisis de la compresión de los datos para evaluar la eficiencia del almacenamiento.

**3.3.1. Limpieza de datos**

Se identifican y tratan los valores faltantes, garantizando la integridad de los datos. Se realiza la imputación específica en las columnas 'HORA' y 'ENTIDAD'. La verificación de datos erróneos o inconsistentes se realiza, y se estandarizan las categorías en columnas relevantes. También, se inspeccionan variables en busca de valores atípicos o inusuales, identificando tanto valores atípicos en variables numéricas como en variables categóricas.

**3.3.2. Análisis exploratorio**

Se explora la distribución de la variable objetivo 'GRAVEDAD' para comprender la proporción de clases. Se realiza una exploración detallada de variables numéricas para entender sus características. Además, se analizan las relaciones entre variables, tanto numéricas como categóricas. Esto incluye una exploración específica de relaciones entre variables numéricas y un análisis de interacciones entre variables categóricas y numéricas.



**Figura 3.** Matriz de correlación

Al revisar la gráfica de la matriz de correlación podemos darnos cuenta de que las variables no están directamente correlacionadas esto quiere decir que las variables son independientes una de otra, esto es importante detallarlo ya que en base a las correlaciones de variables podremos considerar modelos de análisis.

**3.3.3. Transformación de etiquetas en variables numéricas**

Se lleva a cabo la conversión de las etiquetas de la variable objetivo 'GRAVEDAD' en variables numéricas para facilitar el modelado.

**3.3.4. Selección y reducción de características**

La selección de características consiste en identificar y mantener sólo las características relevantes para el análisis o la predicción, descartando las que aportan información insignificante. Este enfoque mejora la eficacia del modelo al reducir el ruido y minimizar la complejidad computacional.

Tabla

Descripción generada automáticamente con confianza media

**Figura 4.** Importancia de las características para el modelo

La selección de características se determina identificando aquellas que tienen un impacto significativo en un análisis o modelo predictivo. En el contexto del estudio, se consideran características importantes "motocicleta", "peatón", "coche", "año", "furgoneta", "bicicleta", "camión", "caravana", "autobús" y "buseta". Estas características se valoran por su potencial para influir en la gravedad de los accidentes de tráfico. Por ejemplo, determinadas variables relacionadas con los distintos tipos de vehículos podrían desempeñar un papel esencial en la predicción de la gravedad de los accidentes. Al mismo tiempo, el análisis de los datos de incidentes relacionados con "peatones" y "ciclistas" podría ayudar a obtener información sobre las interacciones específicas que se producen en los entornos de tráfico. Por último, es importante tener en cuenta cómo el tamaño, la velocidad y el uso habitual de tipos de vehículos como "camión", "furgoneta" o "buseta" pueden afectar a las condiciones de la carretera en determinados contextos. Estas características se eligieron por su impacto potencial en el fenómeno estudiado, aumentando así la precisión y la importancia de los modelos analíticos o predictivos aplicados en el ámbito de la seguridad vial.

**3.3.5. Analisis de componentes principales**

Los resultados del PCA (análisis de componentes principales) muestran que el uso de 8 componentes explica aproximadamente el 88,94% de la variabilidad de los datos originales. Esto significa que la reducción de la dimensionalidad de los datos a 8 componentes principales conserva una cantidad significativa de información del conjunto de datos original. Esta reducción simplifica la complejidad de los datos al tiempo que preserva una parte significativa de su variabilidad. Esto facilita el análisis y la visualización sin perder información crucial de los datos originales [5].

**4. Modelado**

Los modelos de clustering pertenece a la categoría de aprendizaje no supervisado, son aquellos que agrupan los datos en clústeres basándose en características intrínsecas y en la estructura de los datos, sin información externa como etiquetas predefinidas. Revelan patrones y estructuras inherentes a los datos, sin requerir un conocimiento previo de clases o categorías, según [6]–[11].

**KMeans:** Es un algoritmo de clustering que agrupa datos en clústeres predefinidos basados en la distancia media entre puntos. Funciona particionando datos en K clústeres, donde K es predefinido. Busca minimizar la varianza dentro de los clústeres.

**Agglomerative:** Es un método de clustering jerárquico que une clústeres de manera iterativa basándose en la similitud entre ellos.Comienza con cada punto como un clúster individual y luego los une en clústeres más grandes, formando una jerarquía.

**DBSCAN:** Es un algoritmo de clustering basado en densidad que identifica clústeres basados en regiones de alta densidad separadas por regiones de baja densidad. Puede identificar clústeres de formas y tamaños arbitrarios y es robusto ante ruido en los datos. No requiere especificar el número de clústeres de antemano.

**5. Evaluación de los modelos**

Las métricas internas evalúan la calidad de los clústeres basándose únicamente en la estructura de los datos y los clústeres formados, sin información externa ni etiquetas de clase conocidas. Estas métricas miden la cohesión y la separación entre clústeres, según [6]–[11].

**Silhouette Score:** Es una métrica que evalúa la calidad de los clústeres. Varía entre -1 y 1, donde valores cercanos a 1 indican clústeres bien definidos y separados, mientras que valores cercanos a -1 indican clústeres superpuestos. Se calcula utilizando la distancia promedio entre un punto y los demás puntos en su mismo clúster, comparada con la distancia promedio a los puntos en el clúster más cercano.

**Calinski Harabasz:** Es una métrica que mide la separación entre clústeres y la dispersión dentro de los clústeres. Valores más altos indican clústeres más densos y bien separados. Se calcula utilizando la dispersión entre los clústeres dividida por la dispersión dentro de los clústeres.

**Davies Bouldin:** Es una métrica que evalúa la compacidad y separación entre clústeres. Valores más bajos indican clústeres más cohesivos y separados. Se calcula como la medida promedio de la similitud entre cada clúster y su clúster más similar, considerando tanto la distancia intra-cluster como la distancia inter-cluster.

**Tabla 3.** Resultados de los modelos

|  |  |  |  |
| --- | --- | --- | --- |
| Modelo/  Índice | KMeans | Agglomerative | DBSCAN |
| Silhouette Score | 0.4792 | 0.4845 | 0.4610 |
| Calinski Harabasz | 9809.9757 | 12538.0473 | 5705.2072 |
| Davies Bouldin | 0.8818 | 12538.0473 | 1.5842 |

La tabla presentando los resultados de la evaluación de varios algoritmos de agrupación, en concreto KMeans, Agglomerative y DBSCAN, utilizando tres métricas de uso común para la evaluación de la calidad de los conglomerados.

La métrica Silhouette Score indica la cohesión y separación entre clústeres, con valores cercanos a 1 que denotan clústeres bien definidos. En este caso, la agrupación aglomerativa obtuvo la puntuación más alta, 0,4845, seguida de KMeans, 0,4792, y DBSCAN, 0,4610.

Calinski Harabasz, esta métrica mide la dispersión dentro de los clústeres y entre ellos, con valores más altos que indican clústeres más densos y separados. La agrupación aglomerativa presenta el valor más alto (12538,0473), seguida de KMeans (9809,9757) y DBSCAN con un valor más bajo (5705,2072).

La métrica Davies-Bouldin evalúa la distancia media entre clústeres, con valores más bajos que indican clústeres más cohesionados. En este caso, se observa que KMeans y DBSCAN tienen valores exactos de 0,8818 y 1,5842, respectivamente.

En general, la agrupación aglomerativa muestra un mejor rendimiento basado en la puntuación Silhouette y la Calinski Harabasz en comparación con KMeans y DBSCAN.

**6. Despliegue**

La fase de despliegue queda incompleta por falta de tiempo, es el último paso para hacer operativos los modelos desarrollados. Este paso habría facilitado la integración de los modelos de predicción de la seguridad vial en una plataforma accesible para el análisis en tiempo real y la proyección de la gravedad de los accidentes. No haber completado este paso limita la disponibilidad inmediata de una herramienta que las autoridades locales y las agencias de seguridad vial podrían haber utilizado para tomar decisiones informadas y proactivas. Sin embargo, reconociendo la importancia del despliegue, los resultados del análisis deben convertirse en herramientas prácticas y funcionales para beneficiar a la comunidad y prevenir los accidentes de tráfico.

**7. Conclusiones**

La seguridad vial en Bucaramanga puede mejorarse conociendo el impacto de los distintos tipos de vehículos en la gravedad de los accidentes de tráfico. Esta investigación exhaustiva sirve de base sólida para diseñar medidas preventivas más eficaces. Sin embargo, la eficacia de las políticas futuras depende en gran medida de la correlación significativa entre los tipos de vehículos y la gravedad de los accidentes. A pesar de los posibles peligros, concentrarse en esta relación crea posibilidades para elaborar políticas de seguridad vial más precisas y orientadas, mejorando la salvaguarda de la comunidad y las carreteras de Bucaramanga.

El análisis DAFO ha permitido identificar de forma exhaustiva los puntos fuertes, los puntos débiles, las oportunidades y las amenazas para la seguridad vial en Bucaramanga. Se ha establecido un plan de beneficios para mejorar las políticas de seguridad vial examinando la correlación entre los tipos de vehículos y la gravedad de los accidentes. Los objetivos empresariales se alinean con esta estrategia al tratar de comprender y prevenir las causas que aumentan la gravedad de los accidentes. Estos elementos crean una base sólida para elaborar intervenciones precisas y políticas específicas, aprovechando las oportunidades identificadas y mitigando las amenazas potenciales. Este enfoque integral da prioridad a la reducción de los accidentes de tráfico y a la mejora de la seguridad y la calidad de vida de la comunidad de Bucaramanga.

El algoritmo de agrupación aglomerativa parece obtener mejores resultados según las métricas de evaluación utilizadas, concretamente la puntuación de silueta y Calinski Harabasz. Esto sugiere una mejor cohesión y separación entre clústeres. No obstante, es necesaria una revisión adicional, a pesar de las puntuaciones disponibles, parece que el clustering Aglomerativo supera a KMeans y DBSCAN en esta situación particular.

El uso del PCA con 8 componentes explica aproximadamente el 88,94% de la varianza de los datos originales. Al reducir la dimensión, conserva información significativa de los datos originales, lo que hace que el análisis sea más eficiente y proporciona una visualización más clara de la estructura de los datos.

Se han identificado características esenciales como "motocicleta", "peatón", "coche", "año", entre otras, que tienen un impacto significativo en la predicción de la gravedad de los accidentes de tráfico. La inclusión de estas características mejora la precisión del modelo al tener en cuenta factores relevantes para la seguridad vial.

**8. Anexos**

Plan Beneficios:

[PLAN DE GESTIÓN DE BENEFICIOS PARA EL PROYECTO DE ANÁLISIS DE GRAVEDAD DE ACCIDENTALIDAD EN BUCARAMANGA.pdf](file:///C:\Users\teran\OneDrive\Documents\PLAN%20DE%20GESTI%C3%93N%20DE%20BENEFICIOS%20PARA%20EL%20PROYECTO%20DE%20AN%C3%81LISIS%20DE%20GRAVEDAD%20DE%20ACCIDENTALIDAD%20EN%20BUCARAMANGA.pdf)

Estructura de desglose de trabajo:

[Estructura de Desglose del Trabajo (EDT) Accidentes Transito Bucaramanga .pdf](file:///C:\Users\teran\OneDrive\Documents\Estructura%20de%20Desglose%20del%20Trabajo%20(EDT)%20Accidentes%20Transito%20Bucaramanga%20.pdf)

**9. Referencias**

[1] P. Nixon, R. González, D. De, and T. De Bucaramanga, “INFORME GESTIÓN ENERO A DICIEMBRE GRUPO CONTROL VIAL ACCIDENTALIDAD Año 2021 comparativo año 2020.”

[2] “Traumatismos causados por el tránsito.” Accessed: Nov. 13, 2023. [Online]. Available: https://www.who.int/es/news-room/fact-sheets/detail/road-traffic-injuries

[3] J. M. Rodríguez Hernández, R. Enrique. Peñaloza Quintero, and L. Katerin. Ariza, *Factores de riesgo asociados a lesiones : causadas por el tránsito y propuesta de intervenciones para el contexto colombiano*. Editorial Pontificia Universidad Javeriana, 2015.

[4] J. Hérnandez Orallo, J. Ramírez Quintana, and C. Ferri Ramírez, *Introducción a la minería de datos*, Pearson education. Madrid: Pearson education, 2004.

[5] N. Migenda, R. Möller, and W. Schenck, “Adaptive local Principal Component Analysis improves the clustering of high-dimensional data,” *Pattern Recognit*, vol. 146, p. 110030, Feb. 2024, doi: 10.1016/j.patcog.2023.110030.

[6] W. Ding, Y. Zhang, Y. Sun, and T. Qin, “An Improved SFLA-Kmeans Algorithm Based on Approximate Backbone and Its Application in Retinal Fundus Image,” *IEEE Access*, vol. 9, pp. 72259–72268, 2021, doi: 10.1109/ACCESS.2021.3079119.

[7] A. A. Bushra and G. Yi, “Comparative Analysis Review of Pioneering DBSCAN and Successive Density-Based Clustering Algorithms,” *IEEE Access*, vol. 9, pp. 87918–87935, 2021, doi: 10.1109/ACCESS.2021.3089036.

[8] N. Gholizadeh, H. Saadatfar, and N. Hanafi, “K-DBSCAN: An improved DBSCAN algorithm for big data,” *J Supercomput*, vol. 77, no. 6, pp. 6214–6235, 2021, doi: 10.1007/s11227-020-03524-3.

[9] J. Fang, Z. Xie, H. Cheng, B. Fan, H. Xu, and P. Li, “Anomaly detection of diabetes data based on hierarchical clustering and CNN,” *Procedia Comput Sci*, vol. 199, pp. 71–78, 2022, doi: 10.1016/j.procs.2022.01.010.

[10] X. Zhou, W. Liang, W. Li, K. Yan, S. Shimizu, and K. I.-K. Wang, “Hierarchical Adversarial Attacks Against Graph-Neural-Network-Based IoT Network Intrusion Detection System,” *IEEE Internet Things J*, vol. 9, no. 12, pp. 9310–9319, Jun. 2022, doi: 10.1109/JIOT.2021.3130434.

[11] L. Chen, N. Jia, H. Zhao, Y. Kang, J. Deng, and S. Ma, “Refined analysis and a hierarchical multi-task learning approach for loan fraud detection,” *Journal of Management Science and Engineering*, 2022, doi: 10.1016/j.jmse.2022.06.001.