

Modelamiento de la severidad de los accidentes de tránsito en el Estado de Maryland

Edixon Eniver Morales Quintero, *edxmorales@gmail.com*

Heiner Romero Leiva, *heinerleiva@gmail.com*

Steven Montero Muñoz, *stevenmontero47@gmail.com*

Resumen—El presente artículo presenta los resultados finales de un proyecto de investigación en el campo de las ciencias computacionales, enfocado en el análisis de tres algoritmos de Aprendizaje supervisado para la predicción de la severidad de los accidentes de tránsito en el Estado de Maryland (EE.UU.). El objetivo principal del estudio fue evaluar el rendimiento y la eficacia de los algoritmos propuestos en la identificación y clasificación de la severidad de los accidentes de tránsito y su explicabilidad, así como la utilización de técnicas como SMOTE para lograr un mejor aprendizaje durante la base de entrenamiento. Los resultados obtenidos mostraron que dichos algoritmos no son capaces de producir buenos resultados en datasets altamente desbalanceados, pero el uso de técnicas como SMOTE mejoran el rendimiento de los algoritmos en cuanto a las métricas del recall y al F1 durante el entrenamiento y el testeo.

Palabras clave—:Aprendizaje automático, clasificación, accidentes de tránsito, algoritmo, severidad, aprendizaje supervisado, explicabilidad.

Abstract —This article presents the final results of a research project in the field of computer science, focused on the analysis of three supervised learning algorithms for predicting the severity of car crashes in the State of Maryland (USA). The main objective of the study was to evaluate the performance and effectiveness of the proposed algorithms in identifying and classifying the severity of car accidents, along with their explainability. Additionally, techniques such as SMOTE were implemented to enhance learning during the training phase. The results obtained showed that these algorithms are not capable of producing good results in highly imbalanced datasets. However, the use of techniques like SMOTE improved the performance of the algorithms in terms of recall and F1 metrics during both training and testing.

Key words —:Machine learning, classification, car crashes, algorithm, severity, supervised machine learning, explainability.

I. INTRODUCCIÓN

Los accidentes de tránsito representan una preocupación a nivel mundial creciente hoy en día, con un impacto económico y humano significativo. En el Estado de Maryland, entre 2018 y 2022, hubo 109,389 accidentes, incluyendo 517

El presente documento corresponde al artículo científico de la tesis de Maestría: «Modelamiento de la severidad de los accidentes de tránsito en el Estado de Maryland», en donde se resumen los principales hallazgos encontrados durante el estudio y la explicación de las variables que más intervienen en dichos accidentes y su severidad.

accidentes fatales, con un costo anual estimado de \$4.476 billones de dólares (Zero Deaths Maryland, 2023). La identificación de la severidad de estos accidentes es esencial para las Administraciones públicas y que así puedan mejorar su planificación urbana y la gestión de la seguridad vial.

Este estudio busca abordar la falta de un sistema robusto para clasificar la severidad de los accidentes en el Estado de Maryland. Utiliza modelos avanzados de Machine Learning para proporcionar a las autoridades herramientas que faciliten la toma de decisiones pro-activas basadas en la gravedad de los accidentes. El enfoque se centra en beneficiar a gestores de emergencias y planificación vial, con la finalidad de que se pueda mejorar la respuesta a emergencias y optimizando recursos. Además, identifica factores clave para la planificación urbana, como velocidad, condiciones climáticas y tipo de intersección. Como tal se puede identificar que las causas de los accidentes pueden dividirse en dos grupos principales: aquellas vinculadas al comportamiento del conductor, como la fatiga, el sueño, el abuso de velocidad y la conducción imprudente, y factores externos que abarcan la marca, tipo, modelo y año del vehículo (World Health Organization, 2018).

La investigación contribuye al ámbito de gestión de emergencias y planificación urbana al proporcionar información estratégica para optimizar recursos y mejorar la seguridad vial, con el uso de modelos de aprendizaje supervisado como Árboles de Decisión, Random Forest y Regresión Lógica que pueden contribuir a la predicción de accidentes y a entender las causas de los mismos dado a su alto poder explicativo, así como la utilización de técnicas para el tratamiento de datos desbalanceados como lo son el SMOTE, contribuyendo a obtener mejores métricas que puedan ser utilizadas por las autoridades para reproducir dichos resultados.

II. MÉTODOS

El enfoque metodológico del presente estudio se desglosa en varias fases, cada una crucial para la correcta implementación del modelo propuesto, a continuación se resumen cada una de ellas:

1. Obtención y Validación de Datos: recopilación de datos relacionados con accidentes de tránsito desde el sitio público de Maryland, así como la descarga de los principales conjuntos de datos relevantes para el estudio y validación de integridad y precisión.
2. Entendimiento y Caso de Uso de Variables: análisis detallado de variables presentes en el conjunto de datos, así

- como identificación de variables relevantes y construcción de nuevas de ellas según sea necesario. Por último se hace uso de un diccionario disponible de cada una de las variables del dataset para su mejor comprensión y aplicación.
3. Descripción, Limpieza, Estructuración y Depuración de Datos: limpieza exhaustiva, incluyendo imputación de valores faltantes y formato adecuado. Identificación de estructuras específicas y conversión de datos a diferentes formatos; unión de conjuntos de datos relevantes.
 4. Análisis Exploratorio de Datos: análisis estadístico descriptivo para comprender la naturaleza de los datos, así como la identificación de medidas de posición, forma de los datos, tendencias y variabilidad y representación gráfica para una comprensión visual.
 5. Creación de Splits en los Datasets: definición de splits aleatorios en la base de datos completa para entrenamiento (80 %) y testeo (20 %), con esto se garantiza que el modelo no haya visto antes el subconjunto de testeo.
 6. Modelos de Machine Learning Supervisado: aplicación de modelos como Árboles de Decisión, Random Forest y Regresión Logística Multinomial y uso de software computacional para implementar los modelos y determinar el más óptimo.
 7. Interpretabilidad de Modelos y Explicabilidad: selección de modelos explicativos como Árboles de Decisión, Random Forest y Regresión Logística Multinomial, así como análisis de niveles de asociación entre datos y su influencia en la ponderación de resultados para los modelos.
 8. Aplicación de Técnicas de Rejilla (Grid Search): utilización de Grid Search para encontrar los mejores resultados y ajustar hiperparámetros del modelo.
 9. Validación y Testeo: validación de modelos mediante K-Fold Cross-Validation y técnicas estadísticas. Empleo de métricas como matriz de confusión, curva ROC, AUC, puntaje F1, recall y precisión.
 10. Aplicación de Técnicas Adicionales Experimentales: experimentación con modelos entrenados en datos no balanceados y contraste de resultados con técnicas de balanceo como SMOTE para evaluar mejoras en el conjunto de entrenamiento únicamente.

El proceso descrito permitió la correcta preparación y evaluación de modelos de aprendizaje automático para la predicción de la severidad de accidentes de tránsito en el estado de Maryland.

III. ALGORITMOS EMPLEADOS

En este estudio, se proponen diversos algoritmos de Machine Learning (ML) para modelar la severidad de los accidentes de tránsito en Maryland y entender las variables más influyentes. Por un lado se utilizaron algoritmos altamente estudiados en la literatura, y que contasen con un componente explicativo, con la finalidad de entender las relaciones que guarden dichas variables para la predicción de la severidad asociada a los accidentes en general.

1. Árboles de Decisión: de acuerdo con Calvo (2022), los árboles de decisión son herramientas poderosas para problemas de clasificación y regresión, se caracterizan por su estructura similar al razonamiento humano, representada como un diagrama de flujo. La entropía y la ganancia de información guían la toma de decisiones en estos algoritmos, considerados no paramétricos, con implementaciones como ID3, CART, C4.5, SPRINT y SLIQ.
2. Random Forest: corresponden a un conjunto de árboles predictores, que mejoran la generalización al combinar múltiples árboles, utilizando muestras bootstrap y selección aleatoria de variables en cada nodo.
3. Regresión Logística Multinomial: según Cayuela (2010), se revela como una herramienta crucial en inteligencia artificial para clasificación con múltiples categorías en la variable dependiente. Son extensiones de modelos lineales tradicionales, adaptándose a diversas distribuciones de errores como Poisson, Binomial y Gamma. A su vez, este algoritmo ofrece soluciones robustas, evaluables mediante métricas clave como precisión, matriz de confusión y Curva ROC. Su versatilidad abarca variables predictoras numéricas y categóricas, permitiendo aplicar técnicas de regularización como penalización L1 o L2 para un refinamiento adicional.

IV. EXPLICABILIDAD DE LOS MODELOS

El estudio aborda la explicabilidad de modelos predictivos mediante técnicas avanzadas, incorporando un módulo de explicabilidad que utiliza diversos algoritmos para brindar interpretaciones detalladas de los modelos utilizados. Se destacan dos algoritmos clave en este módulo:

1. Importancia de Características (Feature Importance): utilizado para explicaciones a nivel global, este evalúa la contribución relativa de cada característica en el rendimiento general del modelo. Asimismo, clasifica las características según su influencia en las predicciones, proporcionando una comprensión general de la importancia de diferentes aspectos del conjunto de datos para el modelo.
2. Shapley Additive exPlanations (SHAP): agregado para explicaciones a nivel local. Basado en la teoría de juegos cooperativos, asigna valores Shapley a cada característica y refleja la contribución justa de cada característica a la predicción específica para una instancia dada, esto permite entender cómo cada característica afecta individualmente a las predicciones del modelo para casos específicos.

Esta estrategia equilibrada en la aplicación de algoritmos de explicabilidad busca no solo comprender el rendimiento global del modelo, sino también desentrañar las complejidades de las decisiones a nivel individual. Mejora así la comprensión global del comportamiento del modelo predictivo, proporcionando insights tanto a nivel macro como micro.

V. DEFINICIÓN DE LOS MODELOS

Después de la fase de aplicación de técnicas de rejilla (del inglés, Grid Search) se obtienen los mejores hiper-parámetros y seguidamente se realiza una comparación entre los tres algoritmos de clasificación (Regresión Logística, Árbol de Decisión y Random Forest) utilizando validación cruzada para evaluar su rendimiento. Es crucial tener en cuenta que los modelos se entrenan con datos que no representan la realidad, ya que abarcan el período 2015-2022, y la distribución de estos datos puede diferir de la realidad actual.

Dado lo anterior, se definen los tres modelos a utilizar y la evaluación de los modelos se realiza mediante validación cruzada estratificada (StratifiedKFold) con 3 divisiones.

La revisión de los resultados muestra que el Árbol de Decisión supera a los otros algoritmos, seguido por Random Forest. Las métricas de precisión para cada modelo son:

1. Regresión Logística: 36.80 %
2. Árbol de Decisión: 83 %
3. Random Forest: 81.98 %

Dado que el Árbol de Decisión muestra el mejor rendimiento, se procede a compararlo con Random Forest, obteniendo métricas adicionales como la matriz de confusión y un gráfico de la 'Curva de Precisión - Recall para una clasificación Multiclase'.

V-A.

Árbol de Decisión

El informe de clasificación revela un análisis exhaustivo del rendimiento del modelo. La precisión para la clase 1.0 es del 83 %, indicando que el 83 % de las instancias clasificadas como tal son correctas. Sin embargo, la clase 3.0 presenta una precisión del 0 %, lo que señala una incapacidad para clasificar correctamente instancias en esta categoría. En cuanto al recall, se destaca un 100 % para la clase 1.0, indicando que todas las instancias reales fueron correctamente clasificadas, mientras que para la clase 2.0 y 3.0, los valores son 6 % y 0 %, respectivamente.

El F1-score, que es una media ponderada de precisión y recall, muestra un rendimiento alto para la clase 1.0 (90 %), pero resultados significativamente más bajos para las clases 2.0 (11 %) y 3.0 (0 %). El soporte indica la cantidad de instancias reales para cada clase, siendo la clase 1.0 la más abundante con 163,894 instancias, seguida por la clase 2.0 con 35,803 instancias y la clase 3.0 con 303 instancias.

La exactitud general del modelo alcanza el 83 %. Al considerar las medias no ponderadas y ponderadas (macro avg y weighted avg), se observa que las métricas muestran una variabilidad en el rendimiento según la clase. Mientras que la macro avg presenta promedios más bajos (precisión 51 %, recall 35 %, F1-score 34 %), la weighted avg, que tiene en cuenta la proporción de instancias reales de cada clase, muestra valores más altos (precisión 81 %, recall 83 %, F1-score 76 %).

En resumen, el Árbol de Decisión demuestra una buena precisión para la clase mayoritaria, pero presenta desafíos en

la clasificación de clases minoritarias, evidenciando la importancia de considerar múltiples métricas para una evaluación completa del rendimiento del modelo.

En este análisis detallado del modelo de Árbol de Decisión, se destacan importantes disparidades en su rendimiento al clasificar diversas clases. Aunque la precisión para la Clase 1.0 alcanza un sólido 83 %, se observan desafíos significativos al abordar las Clases 2.0 y 3.0, con precisiones del 71 % y 0 %, respectivamente. La capacidad del modelo para identificar correctamente instancias sin heridas (Clase 1.0) es excepcional, evidenciada por un recall del 100 %. Sin embargo, enfrenta notables dificultades al capturar casos de lesiones (Clase 2.0) y, más preocupantemente, de muerte (Clase 3.0).

El análisis de la importancia de las características arroja luz sobre las variables críticas en la toma de decisiones del modelo. Aspectos como 'DAMAGE_CODE', 'EQUIP_PROB_CODE,' y 'SAF_EQUIP_CODE' emergen como factores determinantes en la clasificación, proporcionando una visión valiosa para abordar posibles deficiencias. Estas variables no solo influyen en la precisión general del modelo, sino que también ofrecen pistas sobre áreas específicas que podrían requerir ajustes para mejorar la capacidad predictiva, especialmente en situaciones minoritarias o desafiantes.

Este estudio subraya la importancia de considerar el desbalanceo de clases en el diseño de modelos predictivos, proponiendo enfoques para optimizar la eficacia del Árbol de Decisión en escenarios más complejos de clasificación. La identificación de áreas de mejora específicas, respaldada por el análisis detallado de las características relevantes, proporciona una base sólida para ajustes y refinamientos futuros, buscando así mejorar la capacidad del modelo para abordar una gama más amplia de situaciones y garantizar su utilidad en aplicaciones del mundo real.

V-B.

Random Forest

El análisis del informe de clasificación muestra que en términos de precisión, se destaca un 82 % para la Clase 1.0, indicando que el 82 % de las instancias clasificadas como tal son correctas. Sin embargo, para las Clases 2.0 y 3.0, la precisión es nula, evidenciando la incapacidad del modelo para clasificar adecuadamente instancias en estas categorías.

En cuanto al recall, se observa un rendimiento perfecto para la Clase 1.0, con un 100 % de instancias reales correctamente clasificadas. En cambio, para las Clases 2.0 y 3.0, el recall es del 0 %, señalando que ninguna instancia real de estas clases fue capturada por el modelo.

El F1-Score, que pondera precisión y recall, muestra un equilibrio destacado para la Clase 1.0 (90 %), pero un rendimiento completamente deficiente para las Clases 2.0 y 3.0 (0 %). El soporte indica la cantidad de instancias reales para cada clase, siendo la Clase 1.0 la más abundante con 163,894 instancias, seguida por la Clase 2.0 con 35,803 instancias y la Clase 3.0 con 303 instancias.

La exactitud general del modelo alcanza el 82 %. Considerando las medias no ponderadas y ponderadas (macro avg

y weighted avg), se evidencia variabilidad en el rendimiento según la clase. Mientras que la macro avg presenta promedios más bajos (precisión 27 %, recall 33 %, F1-score 30 %), la weighted avg, que considera la proporción de instancias reales de cada clase, muestra valores más altos (precisión 67 %, recall 82 %, F1-score 74 %).

El análisis revela la capacidad del modelo para detectar hasta un 90 % de instancias sin heridas en un accidente, pero muestra deficiencias al prever lesiones o muertes (Clases 2.0 y 3.0). En resumen, aunque el modelo destaca en la Clase 1.0, presenta deficiencias en la correcta clasificación de instancias en las Clases 2.0 y 3.0, subrayando la importancia de investigar y ajustar el modelo para mejorar su rendimiento en estas clases específicas, de ahí la importancia de utilizar SMOTE para determinar si las métricas mejoran.

En el análisis, se destaca la precisión del 82 % para la Clase 1.0, mientras que las Clases 2.0 y 3.0 muestran una precisión del 0 %. El recall es del 100 % para la Clase 1.0, pero del 0 % para las otras dos clases. El F1-score es del 90 % para la Clase 1.0, pero del 0 % para las Clases 2.0 y 3.0. La exactitud general es del 82 %. El promedio ponderado de precisión, recall y F1-score es del 67 %, 82 %, y 74 %, respectivamente.

En el análisis comparativo con el modelo previo, se observa que el rendimiento del modelo de Árboles Aleatorios sigue siendo deficiente para las Clases 2.0 y 3.0, a pesar de tener un buen rendimiento en la Clase 1.0.

En la explicabilidad del modelo, se presenta una tabla de importancia de características. 'DAMAGE_CODE' destaca como la característica más influyente con un 39.83 % de importancia, seguida por 'EQUIP_PROB_CODE', 'COLLISION_TYPE_CODE', y otras variables. Se destaca que características como 'month' y 'day' tienen un impacto mínimo en las predicciones y podrían ser excluidas sin afectar significativamente el rendimiento del modelo.

En la visualización de la importancia de variables para clasificar registros específicos en las clases objetivo 'Sin heridas', 'Con heridas' y 'Muerte', se observa que las características individuales no tienen un impacto discernible en la capacidad del modelo para clasificar correctamente los registros donde ocurrió una muerte, similar al modelo anterior. Se destaca la necesidad de investigar y mejorar el rendimiento del modelo para las Clases 2.0 y 3.0.

VI. LÍNEAS EXPERIMENTALES CON SMOTE

Con base en los resultados previos de los modelos, se destaca su eficacia al predecir situaciones sin heridos en accidentes. Sin embargo, el propósito fundamental del presente trabajo es anticipar casos de lesiones o muertes en incidentes viales. La principal limitación para lograr este objetivo radica en el desbalance significativo de clases en el conjunto de datos, con una abundancia de observaciones donde no hay heridos (clase 01), pero escasez de datos para casos con heridos (clase 02) o muertes (clase 03).

Se reconoce que los modelos entrenados con la técnica SMOTE generan datos que no reflejan completamente la realidad, ya que son autogenerados por un algoritmo. A pesar

de esto, la evaluación se realizará sobre los datos de prueba originales, utilizando SMOTE únicamente durante el proceso de entrenamiento.

Ante este desafío, se recurre a la técnica SMOTE (Synthetic Minority Over-sampling Technique) para abordar el desequilibrio en conjuntos de datos. Esta técnica, basada en un algoritmo de sobremuestreo, ofrece una solución avanzada para equilibrar datos desproporcionados (Lemaître, Nogueira, Áridas, 2017). Se procederá a recrear los resultados anteriores, esta vez aplicando SMOTE con el objetivo de mejorar las predicciones en las clases de interés para el estudio.

Se realiza un recuento inicial de clases, revelando que la clase 03 representa solo el 0.15 % del total de los datos, mientras que la clase 02 abarca el 17.87 % y la clase 01 el 81.97 %. Se busca equilibrar las clases 02 y 03, aumentando sus observaciones a 819,779, mientras que la clase 01 se mantiene intacta debido a su mayor cantidad de datos. Este balanceo resulta en un total de 2,459,337 registros, con 1,639,558 siendo registros artificiales. Todas las clases ahora representan exactamente el 33.33 % del total, buscando emular las proporciones encontradas en la literatura estudiada, donde las muertes en accidentes rara vez superaban el 20 %.

Este proceso busca mejorar la capacidad predictiva de los modelos al abordar el desbalance en el conjunto de datos, considerando que no todos los accidentes resultan en heridos o muertos. Los resultados de esta reproducción se presentarán posteriormente.

VI-A.

Árbol de Decisión con SMOTE vs Árbol de Decisión con datos originales

A modo general, el modelo se comporta de forma eficiente en el conjunto de entrenamiento, pero cuando se expone al modelo al conjunto de testing original se puede ver que el modelo con SMOTE no mejora en gran medida con el modelo original, y en la clase 3 en cuanto al accuracy, se tenía un 0 % pero con SMOTE este estadístico incrementa a 2 %. Lo que sí ha mejorado durante el testing con el modelo entrenado con SMOTE, es el recall, ya que en el modelo original sin SMOTE se obtenía un 0 % para la clase 3 y un 6 % para la clase 2, pero entrenando con SMOTE, se obtiene un 48 % en la clase 2 y un 67 % en la clase 3, lo que indica que el modelo cuando realiza las predicciones en la clase 3, acierta en un 67 % de las veces.

A nivel del F1 score se puede observar como el modelo con los datos originales se tenía un 90 %, 11 % y 0 % respectivamente para la clase 1, 2, y 3, pero con los datos entrenados con SMOTE se obtiene un 74 %, 32 % y un 3 %, asimismo, el accuracy global es de un 61 % versus un 83 % del modelo con los datos originales. Aunque los resultados no son tan favorables como se esperaba, es importante mencionar que la técnica SMOTE ayudó a obtener un F1 en la clase 2 y 3 mayor del que se obtenía con los datos originales, ya que dichos F1s ahora son de un 32 % y un 3 % entonces se puede evidenciar como dicha técnica si introdujo una mejoría versus los datos originales y estos resultados obtenidos sí contribuyen

en cierta medida a la mejora del modelo, debido a que lo que se encontró en los experimentos del Estado del Arte presentaban problemas de desbalanceo pero no se intentó mejorar las clases desbalanceadas con las técnicas abordadas en este apartado.

Las métricas de precisión, recuperación y puntuación F1 se detallan para cada clase. El modelo, entrenado con datos generados por SMOTE, muestra mejoras en comparación con el modelo original durante el entrenamiento, especialmente en la detección de muertes en accidentes. Sin embargo, al evaluar el modelo en el conjunto de prueba original, las mejoras son menos notables, con un aumento del 0 % al 2 % en la precisión de la clase 3. Aunque el modelo con SMOTE mejora el recall para la clase 3, el F1-score general y la exactitud disminuyen en comparación con el modelo original.

En el análisis de importancia de variables, se destaca el impacto crucial de 'DAMAGE_CODE' en la evaluación de la severidad del accidente, seguido por 'SAF_EQUIP_CODE', 'SPEED_LIMIT', 'LIGHT_CODE' y. Estas variables sugieren la influencia significativa del tipo de daño, equipos de seguridad, límites de velocidad, condiciones de iluminación y género en la gravedad de los incidentes. La explicabilidad del modelo revela que estas características son fundamentales para la clasificación del modelo en las clases objetivo.

En la visualización de la importancia de variables para clasificar registros específicos en las clases objetivo 'Sin heridas', 'Con heridas' y 'Muerte', se observa que las características individuales tienen un impacto variable en la capacidad del modelo para clasificar correctamente, siendo 'DAMAGE_CODE' y las más influyentes.

A pesar de algunas mejoras con la técnica SMOTE, los resultados no cumplen totalmente las expectativas, pero se destaca que la técnica contribuyó a mejorar el F1-score en las clases desbalanceadas. La explicabilidad del modelo destaca la importancia de ciertas variables en la evaluación de la severidad de los accidentes.

VI-B.

Random Forest con SMOTE vs Random Forest con datos originales

En términos globales, el modelo presenta el mismo comportamiento que el Árbol de Decisión, es decir, se comporta de forma eficiente en el conjunto de entrenamiento, pero cuando se expone al modelo al conjunto de testing original se puede ver que el modelo con SMOTE no mejora en gran medida con el modelo original. Por ejemplo, en la clase 3, se tenía un 0 % de accuracy pero con SMOTE este estadístico incrementa a 1 %. Lo que sí ha mejorado durante el testing con el modelo entrenado con SMOTE, es el recall, ya que en el modelo original sin SMOTE se obtenía un 0 % para la clase 2 y 3, pero entrenando con SMOTE, se obtiene un 66 % en la clase 2 y un 69 % en la clase 3, lo que indica que el modelo cuando realiza las predicciones en la clase 3, acierta en un 69 % de las ocasiones.

A nivel del F1 score se puede observar como el modelo con los datos originales se tenía un 90 %, 0 % y 0 % respectivamente para la clase 1, 2, y 3, pero con los datos entrenados

con SMOTE se obtiene un 48 %, 30 % y un 2 %, asimismo, el accuracy global es de un 39 % versus un 82 % del modelo con los datos originales.

A su vez, se puede comprobar que aunque los resultados no son tan favorables como se esperaba, es importante mencionar que la técnica SMOTE ayudó a obtener un F1 en la clase 2 y 3 mayor del que se obtenía con los datos originales (de igual forma que lo hizo en el Árbol de Decisión), ya que dichos F1s ahora son de un 30 % y un 2 % entonces se puede evidenciar como dicha técnica de nuevo si introdujo una mejoría versus los datos originales y estos resultados obtenidos sí contribuyen en cierta medida a la mejora del modelo tal y como se indicó en el apartado anterior.

Las métricas de precisión, recall, F1-score, exactitud, y promedios macro y ponderado se detallan para cada clase. El modelo muestra un buen rendimiento en la clase 1, pero presenta dificultades en la clase 2, con un recall extremadamente bajo. Aunque la clase 3 tiene un rendimiento aceptable, la baja recall en la clase 2 puede deberse a desequilibrios en los datos.

En el análisis comparativo con el modelo sin SMOTE, se observa que el modelo con SMOTE mejora la clasificación de las clases minoritarias, pero a costa de una menor precisión en la clase 1. La elección entre ambos modelos dependerá de los objetivos específicos del problema, priorizando la precisión en la clase 1 o buscando un equilibrio más uniforme entre todas las clases.

En la explicabilidad del modelo, se destaca la importancia de variables como 'DAMAGE_CODE' y 'SAF_EQUIP_CODE', indicando su papel crucial en las predicciones. Sin embargo, se observa que variables geoespaciales como 'LATITUDE' y 'LONGITUDE' tienen una importancia relativamente baja, sugiriendo que la ubicación geográfica no está siendo eficientemente explotada como predictor en el contexto actual. Además, variables temporales como 'year' y 'day' muestran baja relevancia, planteando interrogantes sobre su contribución al rendimiento del modelo.

La evaluación de la importancia de variables para clasificar registros específicos en las clases objetivo 'Sin heridas', 'Con heridas' y 'Muerte' no muestra un impacto discernible de las características individuales en la capacidad del modelo para clasificar correctamente los registros de muertes, similar al modelo anterior.

VII. CONCLUSIONES

El trabajo de investigación se centra en predecir la gravedad de los accidentes de tránsito en el Estado de Maryland utilizando datos de telemetría y georeferenciados recopilados entre 2015 y 2022. Se aborda la falta de un sistema robusto para clasificar la severidad de estos accidentes, que afectan económicamente, socialmente y en términos de salud. Se registraron 109,389 accidentes entre 2018 y 2022, con costos estimados de \$4.476 billones de dólares anuales. La propuesta utiliza modelos avanzados de Machine Learning, como Árboles de Decisión, Random Forest y Regresión Logística.

Durante la construcción de los modelos, se realizaron experimentos, incluyendo el uso de la técnica SMOTE para abordar el desbalance en la clase de muerte. Sin SMOTE, Árboles de Decisión y Random Forest dieron resultados favorables, pero la Regresión Logística no. Con SMOTE, aunque el accuracy global disminuyó, hubo mejoras en recall y F1 para clases desbalanceadas, especialmente en casos de muerte en carretera.

La explicabilidad de los modelos revela la importancia de variables como 'DAMAGE_CODE', 'EQUIP_PROB_CODE' y 'SAF_EQUIP_CODE.' Los factores relacionados con el estado del vehículo, género del conductor y uso de dispositivos de seguridad influyen en la clasificación. Ambos modelos enfrentan desafíos al clasificar accidentes mortales, y SMOTE agrega valor a la explicabilidad.

A pesar de no alcanzar el 70% de accuracy propuesto inicialmente, los modelos mejoran recall y F1, proporcionando valiosa información para la planificación urbana. La consistencia entre ambos modelos refuerza la importancia de estrategias focalizadas. La implementación de SMOTE destaca la relevancia continua de 'DAMAGE_CODE.'

La explicabilidad del modelo es crucial, avanzando en la predicción y clasificación de la severidad de los accidentes. Los modelos no solo ofrecen predicciones, sino también explicaciones claras sobre las variables influyentes. Se ha realizado un análisis exploratorio de datos y la identificación de variables clave ha contribuido a contrastar factores que inciden en la gravedad de los accidentes.

Las técnicas post hoc han permitido un análisis detallado a nivel local, revelando patrones y tendencias específicas. La comprensión detallada de los factores influyentes ha identificado áreas de mejora en la seguridad vial, cumpliendo con el objetivo de evidenciar hallazgos y proponer recomendaciones basadas en evidencia para reducir la incidencia y gravedad de los accidentes de tránsito en el Estado de Maryland.

REFERENCIAS

- [1] Ahmadi, A., Jahangiri, A., Berardi, V., & Machiani, S. G. (2020). Crash severity analysis of rear-end crashes in California using statistical and machine learning classification methods." *Journal of Transportation Safety & Security*, 12(4), 522-546. <https://www.tandfonline.com/doi/full/10.1080/19439962.2018.1505793?needAccess=true>
- [2] Arreola-Rissa, C., Santos-Guzmán, J., Esquivel-Guzmán, A., Mock, C., & Herrera-Escamilla, A. (2008). "Barriles de absorción y contención del impacto: reducción de mortalidad por accidentes de tránsito." *Salud Pública de México*, 50(S1), 55-59. <https://www.scielo.org.mx/pdf/spm/v50s1/a09v50s1.pdf>
- [3] Assi, K., Rahman, S. M., Mansoor, U., & Ratrou, N. (2020). "Predicting crash injury severity with machine learning algorithm synergized with clustering technique: A promising protocol." *International Journal of Environmental Research and Public Health*, 17(15), 5497. <https://www.mdpi.com/1660-4601/17/15/5497>
- [4] Buitrago, J., Vera, M., & Nino, J. (2019). Autos y su independización. Universidad Industrial de Santander Bucaramanga. <http://wiki.sc3.uis.edu.co/images/d/d5/ArchiG12.pdf>
- [5] Cayuela, L. (2009). "Modelos lineales generalizados (GLM)." Materiales de un curso del R del IREC. https://d1wqtxts1xzle7.cloudfront.net/33538949/3-Modelos_lineales_generalizados-libre.pdf
- [6] Cardona, J. (2023). "Modelo predictivo de zonas de riesgo espacio temporal de accidentes de tráfico en la ciudad de Manizales.Repositorio digital Ucaldas: <https://repositorio.ucaldas.edu.co/bitstream/handle/ucaldas/19537/MODELO%20PREDICTIVO%20DE%20ZONAS%20DE%20RIESGO%20ESPACIO%20TEMPORAL%20DE%20ACCIDENTES%20DE%20TR%C3%81FICO%20EN%20LA%20CIUDAD%20DE%20MANIZALES.pdf?sequence=1&isAllowed=y>
- [7] Calvo, P. (2022). Clasificación de tráfico Tor usando modelos basados en árboles de decisión." https://kerwa.ucr.ac.cr/bitstream/handle/10669/87521/Paulo_Calvo_TFIv3_EC-AL-PC-GM-GB-JGC-firmadov3.pdf?sequence=1&isAllowed=y
- [8] Cayuela, L. (2010). "Los modelos lineales generalizados, como la Regresión Logística Multinomial, ofrecen una extensión valiosa para adaptarse a diversas distribuciones de errores." https://www.academia.edu/6847226/Modelos_lineales_generalizados_GLM
- [9] Contrastes de Normalidad. (s. f.). Universidad de Barcelona. http://www.ub.edu/aplica_infor/spss/cap5-6.htm
- [10] Cruz-Bellas, L. (2017). "Modelos predictivos de accidentes de tráfico en Madrid (Master's thesis).Reunir repositorio digital: <https://reunir.unir.net/handle/123456789/6472>
- [11] De Zárate, L. O. (2022). "Explicabilidad (de la inteligencia artificial)." *EUNOMÍA. Revista en Cultura de la Legalidad*, (22), 328-344. <https://e-revistas.uc3m.es/index.php/EUNOM/article/view/6819/5372>
- [12] Díaz, V. (2017). Informe requisito para optar al título de Máster en Ingeniería civil-énfasis en vías y transportes. Universidad del Norte: <https://manglar.uninorte.edu.co/bitstream/handle/10584/8207/130424.pdf?sequence=1&isAllowed=y>
- [13] Douzas, G., & Bacao, F. (2019). "Geometric SMOTE a geometrically enhanced drop-in replacement for SMOTE." *Information Sciences*, 501, 118-135. <https://arxiv.org/abs/1709.07377>
- [14] Estimating Roundabout Performance using Delay and Conflict Opportunity Crash Prediction Douglas R. Kennedy (2005), P.E. Patton, Harris, Rust & Associates, PLC, Chantilly, VA. and Kristine M. Taylor, BSCE, President, Traffic Safety Software, Scottsdale, AZ : <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=83b68b0d98d9cfa59d6c42621791ea39e1374cc1>
- [15] Fan, Z., Liu, C., Cai, D., & Yue, S. (2019). Research on black spot identification of safety in urban traffic accidents based on machine learning method." *Safety Science*, volumen 118. <https://doi.org/10.1016/j.ssci.2019.05.039>
- [16] Ferrer-Sapena, A., Peset, F., & Aleixandre-Benavent, R. (2011). Acceso a los datos públicos y su reutilización: open data y open government." *Profesional de la Información*, 20(3),260-269. <https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/epi.2011.may.03>
- [17] Global Status Report on Road Safety - Time for Action — WHO — Regional Office for Africa. (2018, 7 noviembre). WHO — Regional Office for Africa. <https://www.afro.who.int/publications/global-status-report-road-safety-time-action>
- [18] Harris, C.R., Millman, K.J., van der Walt, S.J., et al. (2020) Array programming with NumPy. *Nature*, 585, 357–362. <https://numpy.org/doc/stable/>
- [19] Hanum, H., Garg, A., & Faheem, M. I. (2023). Accident severity prediction modeling for road safety using random forest algorithm: an analysis of Indian highways." *F1000Research*, 12:494. <https://doi.org/10.12688/f1000research.133594.1>
- [20] Hébert, A., Guédon, T., Glatard, T., & Jaumard, B. (2019). "High-resolution road vehicle collision prediction for the city of Montreal. In 2019 IEEE International Conference on Big Data (Big Data). IEEE. <https://ieeexplore.ieee.org/document/9006009>
- [21] Herrera, J. (2021). Análisis y predicción de la lesividad en accidentes de tráfico mediante la aplicación de random forest." *Archivo Digital UPM*. <https://oa.upm.es/74939/>
- [22] Hunter, J. D. (2007) *Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering*, 9(3), 90–95. <https://matplotlib.org/stable/contents.html>
- [23] Implementación de medidas de seguridad vial prioritarias en América Latina y el Caribe. (2023, 7 noviembre). OPS/OMS — Organización Panamericana de la Salud. <https://www.paho.org/es/documentos/implementacion-medidas-seguridad-vial-prioritarias-america-latina-caribe>
- [24] Jiménez, C. A. C., & Benítez, S. E. G (2022). "Modelo para la clasificación de gravedad en accidentes viales en Bucaramanga." *Researchgate*: https://www.researchgate.net/profile/Camilo-Cespedes-Jimenez-2/publication/362541811_Modelo_para_la_clasificacion_de_gravedad_en_accidentes_viales_en_Bucaramanga/links/62f0374b88b83e7320b77d20/Modelo-para-la-clasificacion-de-gravedad-en-accidentes-viales-en-Bucaramanga.pdf
- [25] Lemaître, G., Nogueira, F., & Aridas, C. K. (2017). Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine

- Learning." *Journal of Machine Learning Research*, 18(17), 1-5. <http://jmlr.org/papers/v18/16-365.html>
- [26] McKinney, W. (2010) *Data Structures for Statistical Computing in Python*. Proceedings of the 9th Python in Science Conference, 51-56. <https://pandas.pydata.org/pandas-docs/stable/>
- [27] Milla, A. (2021). "Políticas urbanas adaptadas a la implantación de los vehículos autónomos." *Repositorio Abierto de la Universidad de Cantabria*: <https://repositorio.unican.es/xmlui/handle/10902/22791>
- [28] Minitab. (2023). "¿Qué es el método de Tukey para comparaciones múltiples?" *Minitab, LLC*. Todos los derechos reservados. <https://support.minitab.com/es-mx/minitab/21/help-and-how-to/statistical-modeling/anova/supporting-topics/multiple-comparisons/what-is-tukey-s-method>
- [29] Montella, A. (2010). "A comparative analysis of hotspot identification methods." *Accident Analysis & Prevention*, 42(2). <https://www.sciencedirect.com/science/article/abs/pii/S0001457509002632>
- [30] Maryland Open Data Portal (2023). <https://opendata.maryland.gov/Public-Safety/Maryland-Statewide-Vehicle-Crashes-Pedestrian-Detail/yhmz-gxyw/data>
- [31] Natekin, A., & Knoll, A. (2013). "Gradient boosting machines, a tutorial." *Frontiers in Neurorobotics*, 7, 21. <https://www.frontiersin.org/articles/10.3389/fnbot.2013.00021/full>
- [32]
- [33] Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011) *Scikit-learn: Machine Learning in Python*. *Journal of Machine Learning Research*, 12, 2825–2830. <https://scikit-learn.org/stable/documentation.html>
- Pirota, M. (2003). Compendio jurisprudencial sobre accidentes de tránsito en carreteras y autopistas." *Rubinzal-Culzoni Editores*. <https://www.martindiegopirota.com.ar/archivos/libro-cjsatca.pdf>
- [34] Pombo, V. (2003). "¿Qué es la explicabilidad de la inteligencia artificial? Cómo quitarle misterio a la tecnología." *Banco Bilbao Vizcaya Argentaria, S.A.* <https://www.bbva.com/es/innovacion/que-es-la-explicabilidad-de-la-ia-como-quitarle-misterio-a-la-tecnologia/>
- [35] Python Software Foundation. (2022) *Python 3.10.12 Documentation*. <https://docs.python.org/3/>
- [36] Lemaitre, G., Nogueira, F., & Aridas, C.K. (2017) *Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning*. *Journal of Machine Learning Research*, 18(17), 1-5. <https://imbalanced-learn.org/stable/>
- [37] Li, S., Xiang, Q., Ma, Y., Gu, X., & Li, H. (2016). "Crash Risk Prediction Modeling Based on the Traffic Conflict Technique and a Microscopic Simulation for Freeway Interchange Merging Areas." *School of Transportation, Southeast University*. <https://doi.org/10.3390/ijerph13111157>
- [38] Lundberg, S.M., & Lee, S.I. (2017) *A Unified Approach to Interpreting Model Predictions*. In *Advances in Neural Information Processing Systems 30*. <https://shap.readthedocs.io/en/latest/>
- [39] Redondo, J. (2003). "Modelo predictivo de accidentes de tránsito." *Biblioteca Digital Universidad de Antioquia, Medellín, Colombia*. https://bibliotecadigital.udea.edu.co/bitstream/10495/34367/1/RedondoJohn_2023_ModeloPredictivoAccidentes.docx.pdf
- [40]
- [41] Requests Documentation. (2022) Retrieved from <https://docs.python-requests.org/en/latest/>
- SHAP Documentation. (2018). *SHAP*. Recuperado 8 de diciembre de 2023, de <https://shap.readthedocs.io/en/latest/>
- [42] US Department of Transportation [US Department of Transportation]. (2018). "State by state crash data and economic cost index." *US Department of Transportation*. <https://www.transportation.gov/research-and-technology/state-state-crash-data-and-economic-cost-index>
- [43] Vázquez, R. (2004). Causas de los accidentes de tránsito desde una visión de la medicina social. El binomio alcohol-tránsito." *Revista Médica del Uruguay*, 20(3), 178-186. <http://www.scielo.edu.uy/pdf/rmu/v20n3/v20n3a03.pdf>
- [44] Vindell, J. (2021). "Datos faltantes (Missing Values)." <https://rpubs.com/VINDELL2981/valoresperdidos>
- [45] Velásquez, J. O. R., Arroyave, F. B., Almeida, E. G., Torres, J. S., Guarín, N., Mosquera, C., ... & Espejo, J. (2019). "Fatalidades por tránsito en Maryland y Massachusetts evaluadas en el contexto de una caminata al azar probabilística." *Revista de la Facultad de Ciencias Médicas de Córdoba*, 76(3), 142-147. <https://revistas.unc.edu.ar/index.php/med/article/view/23777>
- [46] Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. Springer. <https://statisticalsupportandresearch.files.wordpress.com/2017/05/vladimir-vapnik-the-nature-of-statistical-learning-springer-2010.pdf>
- [47] Waskom, M. (2022) *seaborn: statistical data visualization*. *Journal of Open Source Software*, 7(77), 3021. <https://seaborn.pydata.org/>
- [48] World Health Organization. (2018). *Global Status Report on Road Safety 2018*. <https://www.who.int/publications/i/item/9789241565684>
- [49] Yáñez, C., & Francisco, J. (2023). "Procesamiento Digital de Imágenes Mediante Inteligencia Artificial para la Detección de Accidentes de Tránsito en Quito (Master's thesis, Quito, Ecuador: Editorial UISRAEL).Repositorio Universidad de Israel <http://repositorio.uisrael.edu.ec/bitstream/47000/3508/1/UISRAEL-EC-MASTER-ELEC-AUTOM-378.242-2023-003.pdf>
- [50] Yassin, S.S. (2020). Road accident prediction and model interpretation using a hybrid K-means and random forest algorithm approach." *SN Applied Sciences*. Volumen 2 N° 1576. <https://doi.org/10.1007/s42452-020-3125-1>
- [51] Zhang, J., Li, Z., Pu, Z., & Xu, C. (2018). Comparing prediction performance for crash injury severity among various machine learning and statistical methods." *IEEE Access*, 6, 60079-60087.
- [52] Zas, P. (2023). "Técnicas de tratamiento de datos faltantes y aplicación en problema de detección de fraude bancario." https://ruc.udc.es/dspace/bitstream/handle/2183/33270/ZasPerez_Alexandre_TFG_2023.pdf?sequence=3&isAllowed=y
- [53] Zero Deaths Maryland. (2023, 25 octubre). *Maryland crash and traffic fatalities data*. *Zero Deaths Maryland & Vision Zero - Maryland Highway Safety Office*. <https://zerodeathsmd.gov/resources/crashdata/>