Predicción de accidentes con telemetría en transporte de carga

Seminario de Analítica y Ciencia de Datos

Resumen Descriptivo del Proyecto

Este proyecto busca desarrollar un modelo predictivo con inteligencia artificial basado en datos de telemetría, para identificar y alertar sobre posibles accidentes viales en la flota de transporte primario.

Marco Teórico:

1. Revisión de Literatura.

Los datos telemétricos pueden generar gran cantidad de información, por lo que el análisis de estos usando Inteligencia artificial (IA), aprendizaje de maquinas (ML) o aprendizaje profundo (DL) no es un tema ajeno en el sector logístico de Colombia, logrando no solo reducir accidentes, sino también optimizar los costos operativos mediante una mejor gestión del combustible y la planificación de rutas, (Artimo, 2024).

Para reducir la accidentabilidad en la flota de una empresa de la ciudad de Medellín (Cataño Santa & Pitalua Rivero, 2020), Se identificó que a partir del análisis de datos de diferentes flotas, la mínima reducción de accidentabilidad que se puede lograr es del 80%, a pesar que la investigación no tuvo una análisis usando modelos de IA, ML o DL.

En la investigación (Lacherre et al., 2024), se busca encontrar cuál o cuáles son los modelos predictivos más usados en la seguridad vial basados en IA y ML, esta investigación demostró que hay un creciente uso de algoritmos que aprovechan los datos de telemetría. El algoritmo más utilizado son las redes neuronales convolucionales (CNN) u otras combinaciones con CNN ya que está permite aprovechar las fortalezas individuales de cada modelo a través del procesamiento secuencial de datos. Además, se identificó que el algoritmo más importante en términos de precisión para detectar conducción distraída fue Mobilenetv3(red neuronal convolucional que está adaptada a las CPU de los teléfonos móviles) ya que mostró una alta precisión en términos de reconocimiento de patrones en tiempo real. Sin embargo, para el reconocimiento de estilos de conducción, el algoritmo con mejor precisión fue XGBoost (método de aprendizaje automático supervisado para clasificación y regresión utilizado en la herramienta Entrenar con AutoML). Según esta investigación, existe una tendencia hacia el uso de DL, posiblemente debido a la disponibilidad de mayores volúmenes de datos y los avances en hardware y software, así como la capacidad de lograr un mejor desempeño en general. Esta investigación nos da una gran visión de los modelos que más beneficios dan para abordar este tipo de problemáticas y la teoría se ve reforzada con el artículo en PLOS ONE que investiga el uso de modelos de predicción basados en datos para identificar patrones y prevenir accidentes en el transporte terrestre (Chen, 2024), donde exploraron modelos como el perceptrón multicapa (MLP), CNN, la memoria a corto y largo plazo (LSTM), las redes residuales (ResNet), EfficientNetB4, InceptionV3, Extreme Inception (Xception) y MobileNet. Donde lograron concluir que MobileNet se destacó con la mayor precisión del 98,17%.

Existen investigaciones que se han probado con modelos de ML para identificar los factores más influyentes en la severidad de los accidentes, En una de ellas implementan algoritmos de aprendizaje automático como Máquina de vectores de soporte (SVM) y bosques aleatorios (RF) (Prajapati et al., 2023), y en la otra implementaron RF, jungla de decisiones (DJ), AdaBoost, XGBoost, LightGBM y CatBoost (Ahmed et al., 2023). Determinar los factores clave en la accidentabilidad es de suma importancia por lo que los árbol de decisiones y RF muestran gran utilidad para analizar nuestros datos, además que en la investigación (Prajapati et al., 2023), RF fue el mejor clasificador alcanzando una exactitud del 81.45%, esto no solo es útil a la hora de detectar la severidad de los accidentes sino que también nos amplía el objetivo determinando factores influyentes en los accidentes con técnicas como los valores de Shapley y SHAP.

La predicción de siniestros viales también ha sido abordada para la ciudad de Bogotá en el estudio (García Muñoz, 2021). En este caso las redes bayesianas fueron usadas para obtener rendimientos positivos en la estimación de ocurrencias de siniestros viales, lograron adelantos importantes en una de las redes arteriales de la ciudad, conformada por redes de vías troncales como el SITP y redes de carriles de transporte público. Este estudio aporta un enfoque aplicable para el contexto urbano.

Si bien los estudios existentes proporcionan una base sólida para la implementación de IA, ML o DL en la predicción de accidentes con datos telemétricos, es crucial adaptar estos modelos a las condiciones específicas del contexto de nuestra investigación. Muchos estudios se han realizado en regiones con diferentes infraestructuras tecnológicas y redes viales, lo que puede limitar o condicionar su aplicabilidad directa en nuestro contexto. Además, aunque la telemetría ha demostrado ser eficaz para el monitoreo y la reducción de accidentes, su éxito a largo plazo depende de la capacidad de las empresas para mantener y actualizar sus sistemas de manera constante y adicionar nuevas tecnologías.

2. Modelos y Métodos ya existentes.

**Modelo para predicción de siniestros viales basado en redes bayesianas para corredores de la red vial arterial de la ciudad de Bogotá.** (García Muñoz, 2021).

Esta investigación busca analizar y estimar modelos para predecir tasas de accidentes en la red arterial de la ciudad de Bogotá, específicamente en la red de vías troncales del Sistema Integrado de Transporte Público y en los respectivos carriles preferenciales. Para ello se estimaron modelos lineales generalizados multivariados (GLM) y redes bayesianas de probabilidad (BPN). Para este estudio la variable dependiente son los siniestros viales (choques con solo daños (PDO), accidentes con lesiones (INJ) y siniestros con víctimas mortales (FAT)) y se estudian paralelamente bases de datos oficiales para identificar las posibles variables explicativas (por ejemplo, asociadas al diseño de carreteras, volumen de tráfico, infraestructura, entre otras).

La metodología de los autores se desarrolla en las siguientes etapas: estadística descriptiva de la base de datos, modelos GLM multivariantes e identificación de variables de mayor incidencia, desarrollo de modelos lineales generalizados multivariados y desarrollo de modelos por redes bayesianas de probabilidad. Los siguientes son los modelos estimados:

1. **Modelos lineales generalizados multivariados (GLM)**: Su aplicación permite a los autores conocer la forma y distribución de la variable dependiente que representa con mayor exactitud la ocurrencia de siniestros. Estos modelos fueron estimados a partir de paquetes como “Stats”, “MASS” y “Forecast” de la herramienta de análisis estadístico R:

* Regresión Poisson
* Regresión binomial negativa
* Modelo de regresión binomial negativo con ceros inflados
* Modelo binomial negativo con efectos aleatorios espaciales

A su vez el autor utiliza validación cruzada para comparar cuantitativamente el desempeño de las variables de los modelos, en este caso, específicamente se usó la técnica de K-folds. Luego, la estimación de los modelos se realiza por escenarios: red arterial de la ciudad por medio de MNBRE, red troncal del Sistema de bus de tránsito rápido por medio de MNBRE y el de la red de carriles preferenciales por medio de MNB.

1. **Redes bayesianas de probabilidad (BPN):** Los cuales a su vez se estimaron por escenarios y dentro de estos se realizó una estimación para cada nivel de severidad de los accidentes. Y entre las BPNs estimadas se selecciona la mejor para cada escenario.

Estos modelos no paramétricos mostraron un buen ajuste y un alto rendimiento en la estimación de ocurrencia de los siniestros viales en cada una de las severidades y escenarios.

En general, los resultados de las técnicas de modelación caracterizan bien los factores que influyen en la frecuencia y gravedad de los accidentes en Bogotá, además, las diferentes técnicas de modelado permitieron identificar las variables de mayor significancia estadística para los escenarios planteados. Y el análisis comparativo entre modelos no paramétricos basados en BPN y modelos multivariados, evidenció la bondad de la aplicación de los primeros para el caso del estudio de la siniestralidad en Bogotá, sin embargo, esta metodología es compleja y requiere de grandes recursos informáticos y un alto conocimiento sobre la ocurrencia de la siniestralidad. Por otra parte, para la severidad FAT, no se encontró una diferencia significativa entre los aciertos de ambos tipos de modelos, siendo ambos útiles en este caso. Luego, el análisis sugiere que los modelos BPNs presentan un mejor ajuste al estimar los accidentes, especialmente para la reproducción de eventos fuera de la partición realizada en el mismo, lo que los hace relevantes para una función de desempeño vial. Paralelamente, los modelos multivariados son muy útiles para las particiones con la mayor cantidad de registros.

**Improving traffic accident severity prediction using MobileNet transfer learning model and SHAP XAI technique.** (Chen, 2024).

Este estudio, busca crear modelos predictivos haciendo uso de varias técnicas de aprendizaje por transferencia y proporcionar información de los factores más relevantes en los accidentes usando valores de Shapley, calculados con la librería de Python llamada Shap. Seguidamente, otras librerías usadas fueron TensorFlow y Keras, útiles para crear los modelos pre-entrenados. En cuanto a los modelos para predecir la severidad en accidentes (dividida en cuatro niveles), se hizo uso de los siguientes (tres de aprendizaje profundo, seguidos por cinco modelos basados en aprendizaje por transferencia):

* Multilayer Perceptron (MLP).
* Convolutional Neural Network (CNN), el cual fue el modelo de aprendizaje profundo con los mejores resultados y obtuvo una exactitud del 89.37%.
* Long Short-Term Memory (LSTM).
* Residual Networks (ResNet), el cual mostró el segundo mejor desempeño en las métricas de los modelos de aprendizaje por transferencia.
* EfficientNetB4.
* InceptionV3.
* Extreme Inception (Xception), el cual tuvo el menor valor en términos de precisión respecto a los demás modelos con 82,63%.
* MobileNet, el cual se destacó al tener el valor mayor de exactitud entre todos los modelos con un 98.17% y teniendo valores muy altos en las demás métricas de evaluación.

Los modelos se evaluaron en las siguientes métricas: precisión, exactitud, Recall y el estadístico F1. Adicionalmente, el modelo escogido (MobileNet) se testeo también en un dataset adicional para probar la generalizabilidad del mismo, donde se llegó a los siguientes resultados: su tiempo computacional es muy eficiente y mucho mejor que los otros modelos basados en aprendizaje por transferencia e igualmente supera sistemáticamente a los demás modelos en términos de precisión.

**Road Accident Prediction Using Machine Learning.** (Prajapati et al., 2023).

En esta investigación, los autores proponen un modelo predictivo para la identificación de accidentes de tránsito usando técnicas de ML, como Árboles de Decisión, RF y Support Vector Machines, usando una base de datos obtenida de Kaggle. Buscan además identificar los factores más influenciales en la determinación de la severidad de un accidente a partir de un clasificador SMV no lineal llamado “asto”; y un método de clasificación basado en un RF. Los autores utilizaron las siguientes librerías de Python para su trabajo: Pandas, Numpy y Matplotlib. Los autores explican que los resultados de su propuesta tienen un porcentaje de exactitud en la localización de “puntos calientes” e infieren que los siguientes parámetros: tipo de vehículo, código del dispositivo, ubicación del dispositivo, la latitud y longitud, velocidad, fecha y hora, entre otros, tienen un papel importante en la ocurrencia de accidentes.

**A study on road accident prediction and contributing factors using explainable machine learning models: analysis and performance.** (Ahmed et al., 2023).

En este estudio se busca evaluar un conjunto de modelos de ML para predecir la severidad de accidentes en carretera, a la vez, que se analizan los resultados de predicción y se aplica una técnica de ML explicable para evaluar los factores que contribuyen a los accidentes. En general, se consideraron los siguientes modelos:

* Random Forest (RF), el cual es el mejor clasificador con una exactitud del 81.45% y tuvo valores significativos en las métricas de precisión, Recall y estadístico F1.
* Decision Jungle (DJ).
* Adaptive Boosting (AdaBoost).
* Extreme Gradient Boosting (XGBoost), el cual obtiene la segunda mejor exactitud con 78.52%.
* Light Gradient Boosting Machine (L-GBM).
* Categorical Boosting (CatBoost).

Paralelamente, se realiza el análisis de los valores de Shapley para interpretar el desempeño del modelo de RF a nivel global y local. Y se usa el diagrama de dependencia de explicaciones aditivas de SHAPLEY (SHAP) para investigar la relación que existe entre las características y la variable a predecir. Los resultados de este sugieren que la categoría de carretera, el número de vehículos y la variable de uso de psicoactivos se traduce en mayores probabilidades de accidentes con gran severidad, para identificar esto los investigadores emplean SHAP global, mientras que Shapley local lo emplean en explicar predicciones específicas de manera individual, mostrando cómo influyen los factores en un caso concreto. Esto ayuda a entender tanto los patrones generales como las decisiones individuales del modelo.

Para la solución que proponemos creemos que las investigaciones que utilizan modelos más recientes y explicativos como MobileNet y SHAP (Chen, 2024) y RF (Ahmed et al., 2023), podrían ser más útiles porque ofrecen tanto predicción avanzada como explicaciones detalladas (alta precisión e interpretabilidad), mientras que los enfoques más complejos, como las redes bayesianas, podrían ser útiles, pero requieren mayores recursos e impacta la futura escalabilidad del proyecto.

3. Diferenciación del Proyecto

Hablemos de un proyecto revolucionario frente a la forma en que hoy la empresa logística en estudio aborda la accidentabilidad en el transporte primario de carga.

Hasta ahora se ha utilizado la telemetría para monitorear la flota y queremos evolucionar de reaccionar ante los siniestros a predecirlos. Planteamos exactamente combinar la telemetría, con Modelos de Lenguaje Extenso (LLM) y técnicas avanzadas de ML para predecir accidentes antes de que ocurran; es decir, un salto a la predicción y prevención estratégica basada en analizar grandes volúmenes de datos del comportamiento de los conductores y las condiciones del vehículo.

Un gran valor de esta predicción es que no solo reducirá la incidencia de accidentes, sino que en consecuencia también disminuirá los costos operativos y el tiempo de inactividad de los vehículos. Este enfoque es congruente con el pensamiento corporativo de ser innovadores, eficientes, más seguros y, al final, más rentables.

Estudios como el de (García Muñoz, 2021), que utilizó modelos lineales generalizados multivariados GLM y redes bayesianas de probabilidad (BPN) para predecir accidentes en Bogotá, son un insumo en cómo de manera innovadora podemos identificar las variables más relevantes que influyen en los siniestros, como el volumen de tráfico y la infraestructura. Este análisis profundo y multivariable nos inspira a aplicar un enfoque similar para predecir y prevenir accidentes en la flota de transporte primario.

En relación con nuestra propuesta de usar los datos históricos de la telemetría como entrada para el entrenamiento, el estudio de (Prajapati et al., 2023) destaca la importancia de la localización geográfica, la velocidad y el horario como variables claves en la predicción de accidentes​. Aprovechando estos descubrimientos, podremos avanzar en la predicción planteada.

Además, investigaciones recientes (Chen, 2024) han demostrado la efectividad de técnicas de aprendizaje profundo como MobileNet, que alcanzó una exactitud del 98.17%, y redes neuronales convolucionales (CNN), que logró el 89.37%, para predecir la severidad de accidentes viales​. Estos métodos permiten a las empresas logísticas identificar patrones en tiempo real y emitir alertas preventivas con alta precisión.

Proponemos avanzar hacia una logística más segura, eficiente y visionaria.

4. Referencias.

Ahmed, S., Hossain, M. A., Ray, S. K., Islam Bhuiyan, M. M., & Rahman Sabuj, S. (2023, Mayo). *A study on road accident prediction and contributing factors using explainable machine learning models: analysis and performance*. sciencedirect. Retrieved 2024, from https://www.sciencedirect.com/science/article/pii/S2590198223000611

Artimo. (2024, Julio 3). *La telemetría como aliada en la reducción de accidentes viales*. Artimo. Retrieved 2024, from https://www.artimo.com.co/la-telemetria-como-aliada-en-la-reduccion-de-accidentes-viales/

Cataño Santa, M. A., & Pitalua Rivero, Y. (2020). *Estrategias para disminuir la accidentalidad en el transporte de carga de una empresa industrial de la ciudad de Medellín*. Repositorio Institucional Universidad de Antioquia. Retrieved 2024, from https://bibliotecadigital.udea.edu.co/handle/10495/14994

Chen, A. (2024, Abril 9). *Improving traffic accident severity prediction using MobileNet transfer learning model and SHAP XAI technique*. PLOS. Retrieved October 10, 2024, from https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0300640#sec021

García Muñoz, J. A. (2021). *Modelo para predicción de siniestros viales basado en redes bayesianas para corredores de la red vial arterial de la ciudad de Bogotá*. Retrieved 2024, from https://repositorio.unal.edu.co/bitstream/handle/unal/81200/1.015.424.410.2021.pdf;jsessionid=5D3E6711E5D24747DDBBCC1C9F152B04?sequence=3

Lacherre, J., Castillo Sequera, J. L., & David Mauricio. (2024, June 28). *Factors, Prediction, and Explainability of Vehicle Accident Risk Due to Driving Behavior through Machine Learning: A Systematic Literature Review, 2013–2023*. MDPI. Retrieved October 9, 2024, from https://www.mdpi.com/2079-3197/12/7/131#sec2dot3-computation-12-00131

Morán Montero, D. (2022, Julio). *Análisis de datos de telemetría de Fórmula 1 con técnicas de Deep Learning*. Dehesa. Retrieved October 10, 2024, from https://dehesa.unex.es/handle/10662/15838

Prajapati, G., Avinash, Kumar, L., & S PatiL, S. R. (2023, Mayo 18). *Road Accident Prediction Using Machine Learning*. Zenodo home. Retrieved 2024, from https://zenodo.org/records/7961680