# GESTIÓN DE UN PROYECTO DE VINCULACIÓN CON EL SECTOR PÚBLICO BASADOS EN INTELIGENCIA ARTIFICIAL

**Josefina Victoria Benitez[(1)*], Laura Cecilia Diaz Davila[(1)], María Inés Stimolo[(2)] , Ramon Americo Willis[(3)] ,Gustavo Coppini[(3)] , Jose Daniel Britos[(1)]**

[(1)] *Laboratorio de Investigación y Desarrollo de Software e Inteligencia Artificial, Facultad de Ciencias Exactas, Físicas y Naturales, UNC, Argentina.*
[(2)]*Instituto de Estadística y Demografía, Facultad de Ciencias Económicas, UNC, Argentina.*
[(3)]*Instituto de Formación, Ministerio Publico Fiscal, Córdoba, Argentina.*
**Correo Electrónico :josefina.victoria.benitez@mi.unc.edu.ar*

**RESUMEN**

*Este artículo refiere a la importancia de la vinculación tecnológica para agregar valor en el ámbito del sector público con grupos de investigación de la Universidad Nacional de Córdoba mediante transformaciones digitales basadas en IA. Se aborda la gestión de un proyecto para el desarrollo de una aplicación con interfaz web en la caracterización y predicción de accidentes viales lesivos en la ciudad de Córdoba con la finalidad de contribuir a mejorar la toma de decisiones informadas en la gestión de recursos en siniestros viales lesivos. Se presenta el estado de avance del proyecto y las líneas de acciones futuras. El equipo de trabajo, de diferentes unidades académicas y con estudiantes de intercambio internacional de pregrado, busca una solución integral a un problema trabajando interdisciplinariamente y formando talento humano. Los beneficios de la vinculación consisten en el acceso a recursos y conocimientos especializados, el desarrollo de soluciones innovadoras y la transferencia de conocimiento para un impacto efectivo en la mejora de la calidad de vida del ciudadano.*

**ABSTRACT**

*This article refers to the importance of technological linkage to add value in the public sector with research groups from the National University of Córdoba through digital transformations based on AI. The management of a project is addressed for the development of an application with a web interface for the characterization and prediction of harmful road accidents in the city of Córdoba with the purpose of contributing to improving informed decision making in the management of resources in road accidents harmful. The progress status of the project and future lines of action are presented. The work team, coming from different academic units and with international undergraduate exchange students, seeks a comprehensive solution to a problem through interdisciplinary work and the training of human talent. The benefits of the link consist of access to resources and specialized knowledge, the development of innovative solutions and the transfer of knowledge for an effective impact on improving the quality of life of the citizen.*

*Palabras Clave: vinculación tecnológica, gestión de proyecto, Inteligencia Artificial, Machine Learning*

## Introducción

La importancia de la vinculación tecnológica entre la academia y otros sectores radica en la oportunidad de aportar los conocimientos generados en el marco de los grupos de investigación de la Universidad Nacional de Córdoba (UNC), a partir de la intervención directa en un contexto específico desarrollando una aplicación destinada a solucionar un problema del campo de la Economía del Conocimiento.Esta vinculación, en este caso con el sector público, es virtuosa en el campo específico de las transformaciones digitales, desde múltiples sentidos:1. Favorece la formación de talento humano para todos los actores involucrados; 2. Genera una innovación tecnológica basada en IA atendiendo a tres pilares fundamentales: Personas, Procesos, Tecnologías; 3. Da la oportunidad de participación de investigadores en formación de pregrado a través de estancias internacionales y 4. Sus resultados abren una oportunidad sostenida de mejora continua que es posible escalar y generar a mediano plazo externalidades positivas en otros actores del contexto donde actualmente tiene su alcance.Este trabajo es parte de dos proyectos de investigación que se legitiman en el convenio marco de vinculación tecnológica entre la UNC y el Ministerio Público Fiscal (MPF), en el cual acordó mediante tareas conjuntas, promover la convergencia entre áreas de investigación e innovación de la UNC que posee el talento humano y los investigadores/as de alto nivel para generar una herramienta innovadora y útil que como resultado impacte de forma positiva en la gestión de recursos del MPF que se ponen en marcha cuando ocurre un siniestro vial lesivo y por lo tanto contribuya a la calidad de vida de los ciudadanos/as de la ciudad de Córdoba.

El primer proyecto de investigación fue aprobado en la convocatoria de SECYT PIDTA IA 2023[1] y la segunda, en la del Ministerio de Producción, Ciencia e Innovación Tecnológica, a través de la Secretaría de Ciencia y Tecnología para el fortalecimiento y desarrollo de iniciativas de I+D+i del sistema científico-tecnológico en vinculación con sectores productivos, sociales y gubernamentales: Programa Integral de Financiamiento a la Investigación en Cordoba (PIFIC)[2].

A la fecha se ha finalizado una estancia internacional (Junio - Julio), en el marco del Programa Interinstitucional para el Fortalecimiento de la Investigación y el Posgrado del Pacífico (Programa Delfín)[3] de la región del Pacifico Mexicano,en la cual un estudiante del último año de la Universidad de Toluca (UdT), México, avanzó sobre aspectos relacionados a la gobernanza de datos para ser consumidos por modelos de Machine Learning. Actualmente dos estudiantes del último año de Ingeniería en Computación de la Universidad de INSA Rouen Normandie, Rouen , Francia, están desarrollando su estancia hasta Diciembre de este año cuyo objetivo consiste en poner en producción los avances relativos a la gobernanza de datos y el prototipo basado en Machine Learning diseñado para tomar decisiones informadas en la gestión de recursos en siniestros viales lesivos de la ciudad de Córdoba. Los acompañan investigadores/as seniors y líderes que conforman las líneas de investigación descritas e investigadores/as asesores en materia de infraestructura tecnológica y procesos de despliegue de productos, además del equipo especializado en Inteligencia Artificial (FCE, LIDeSIA y UNVM), el director de la carrera de Ingeniería en Computación y la directora de LIDeSIA, profesora titular de IA en Ingenierías.

Entre los investigadores trainee, se suma una Práctica Profesional Supervisada en la carrera de Ingeniería Industrial de la FCEFyN de la UNC radicada en LIDeSIA, cuyo objetivo es llevar adelante la gestión del proyecto. Y por parte del MPF, intervienen dos estudiantes de la carrera de Licenciatura en Política Pública de la Facultad de Ciencias Sociales de la UNC, quienes desempeñan el rol de expertos en el dominio y llevan adelante la comunicación entre los responsables del MPF y el resto del equipo. Actualmente, un tercer equipo constituido por investigadores/as seniors de LIDeSIA, la FCE y la UNVM, lidera los procesos relativos al desarrollo del prototipo basado en Machine Learning (ML) y define las líneas de acción a futuro.

En resumen,este desarrollo se caracteriza por la conformación de equipos interdisciplinarios donde participan distintas unidades académicas y universidades, los que en el desarrollo de proyectos con IA permiten que el trabajo se enriquezca con conocimientos, destrezas y habilidades diferentes apostando a una solución integral con un enfoque holístico

El objeto de estudio se centra en los siniestros viales que representan una fuente significativa de lesiones y pérdidas de vidas humanas en todo el mundo, lo que produce efectos nocivos de orden jurídico, económico y social en las comunidades afectadas. El MPF dirige la investigación penal pública de los siniestros viales lesivos (leves, graves, gravísimas y homicidios) de la ciudad de Córdoba, para lo cual gestiona recursos técnicos especializados en cada evento. Cuando se produce un siniestro vial, se pone en conocimiento a las fiscalías de instrucción intervinientes a través -generalmente- de la Unidad Judicial (UJ) de Accidentología Vial, que forma parte del MPF, donde se recepta la noticia del caso. Dependiendo el tipo de delito -lesiones leves, graves, gravísimas u homicidio- se solicita la cooperación de la Base Operativa de Policía Judicial, que gestiona los recursos técnicos y humanos de los Gabinetes de Reconstrucción Criminal y de Fotografía Legal para que acudan de forma inmediata al lugar donde se produjo el hecho. Este proceso complejo requiere de una mejora basada en transformaciones tecnológicas para contar con la oportunidad de una gestión informada, a través de una herramienta que provea información temprana. Esto contribuirá a optimizar la gestión de recursos y minimizar efectos nocivos de la siniestralidad vial. Para afrontar la problemática se aborda un proyecto de Inteligencia Artificial cuya gestión consiste en [4]: 1.Planificación y armado del proyecto, 2. Colección de datos, depuración y etiquetado, 3.Exploración de modelos, 4. Refinamiento de modelos , 5.Testeo y evaluación, 6. Deployment de modelos y 7. Mantenimiento continuo.

Los datos asociados al problema se gestionan en el MPF en línea y en tiempo real, desde el registro de la UJ. Representan casos de siniestros con información detallada: número de caso, fecha y hora del hecho, lugar donde se produjo, persona denunciante o víctima, persona denunciada o aprehendida,entre otros. Se cuentan con datos desde 2022 a la fecha. El producto entregable propuesto es un prototipo probado en entorno operativo controlado (Prueba de concepto +Prueba piloto), alojado en un docker autocontenido.

## Materiales y Métodos

El corpus de datos inicial del proyecto registra 4883 eventos con sus características (Período 2022 - Abril 2024). Se considera el proceso de tratamiento de los datos que contempla anonimización, transparencia, trazabilidad, confidencialidad según aspectos éticos, regulatorios y de impacto de la Inteligencia Artificial aplicada, atendiendo a un enfoque integral, en el marco de las recomendaciones nacionales, regionales e internacionales vigentes(Ley 25.326 de Protección de Datos Personales[5] y el Convenio 108[6]). Además

de la evaluación de impacto para identificar y atenuar consecuencias negativas y sesgos.
La primera etapa del proyecto , puntos 1 y 2 mencionados en el punto anterior, implica una primera aproximación a la comprensión del problema, un diálogo interdisciplinario en el que participaron usuarios, expertos del dominio, investigadores y estudiantes en formación. En este marco se llevó a cabo el tratamiento de los datos y una exploración de modelos de ML para llevar adelante la tarea del punto 3. Los procesos realizados fueron: 1)Estructuración del DataSet: análisis, transformación de registros y la georreferenciación. 2)Desarrollo Scripts para su tratamiento: Limpieza, automatización para identificar y corregir inconsistencias, entre otros 3)Propuesta de Variables Operacionales: se definieron aspectos relativos al día, hora y lugar del hecho, cantidad y tipo de vehículos involucrados,personas lesionadas y grado, género, segmentación etaria, como las más relevantes.
El equipo de ML abordó los puntos 2, 3 y 5 alcanzando la detección de los primeros patrones de comportamiento utilizando 1)Aprendizaje no supervisado, modelos de Clustering 2) Aprendizaje supervisado, para la caracterización del problema. Para todos los casos se usaron métricas para evaluar el desempeño. La segunda etapa retoma los avances anteriores realizando una revisión del punto 1 y se centra en el desarrollo de una aplicación con interfaz web para dar respuesta a la gestión de información y predicción de siniestros viales con el fin de contribuir a la mejora de toma de decisiones sobre los recursos abordando el punto 6, que pretende alcanzar una prueba piloto en entorno operativo controlado. En cuanto a la aplicación se ha realizado un investigación tecnológica a partir de la cual se ha definido para la aplicación web, desde el tratamiento del frontend y del backend. Y como backend: 1- El framework backend se seleccionó a partir de una comparación en base a lenguaje, patrón de arquitectura, curva de aprendizaje, comunidad y soporte, escalabilidad, seguridad integrada, facilidad de integración, rendimiento, documentación, soporte para base de datos, interacción de ML en el lenguaje de programación respecto a Django, Laravel, Express.jr y Rubi on Rails 2- Para el sistema de gestión de bases de datos también se realizó una evaluación comparativa respecto a los criterios de licencia, soporte ACID, soporte JSON, escalabilidad, extensibilidad, soporte para transacciones complejas, integración con Django, comunidad de usuarios, rendimiento, coste entre PostgreSQL, MySQL, MariaDB, Oracle, SQL Server.

## Resultados

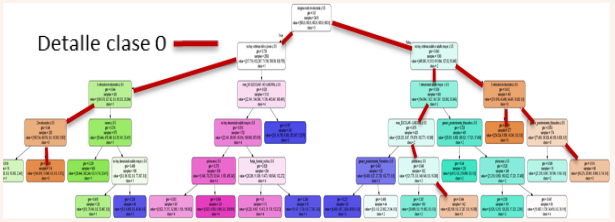
La primera etapa comprendió los puntos 1,2, y 3 de la gestión y constituyó el objetivo central de aprendizaje del estudiante de la UdT con el acompañamiento de investigadores/as LIDeSIA-FCEFyN, IEyD-FCE, UNVM. Durante esta etapa del proyecto se obtuvieron siete versiones de tratamiento de los datos, se generó un diccionario de definición de variables y también se comenzó a registrar la trazabilidad.
El equipo investigadores de ML generó una octava versión del tratamiento de los datos con su trazabilidad, con la que se analizó la conformación de patrones a partir de aspectos relativos a: tipo de lesiones, los vehículos y su cantidad, las personas y su género como así también las variables temporales (Tabla 1).
La etapa actual, consiste como se mencionó antes en la revisión de los avances de la etapa 1 y en poner en producción los avances relativos a la gobernanza de datos y el prototipo basado en ML En cuanto a la planificación de la aplicación a partir de la investigación tecnológica antes mencionada se utilizara para la parte frontend HTML, CSS y JavaScript. Para la parte backend, compuesta por el framework backend se optó utilizar Django ya que combina seguridad, escalabilidad, y facilidad de uso, y para el sistema de gestión de bases de datos PostgreSQL que combina rendimiento, escalabilidad y flexibilidad, lo que resulta en las mejores opciones. También se definió diferentes niveles y tipos de usuarios, lo que ofrece la estructuración de los datos en un sistema de gestión de base de datos, se propone el rol del auditor/a del sistema, el rol del experto/a en ML que puede actualizar el modelo entrenando con nuevos datos, entre otras funcionalidades.
Esta aplicación permitirá incorporar nuevos datos mediante un formulario versión web, almacenar información en una base de datos, automatizar el entrenamiento de modelos de ML junto con el equipo de investigadores senior que trabajan específicamente en el mejoramiento de la base de datos. Se realizan reuniones interdisciplinarias con el fin de ajustar esta herramienta y que los campos pertinentes contribuyan en mejorar la recolección de datos, de manera de: sistematizar la información para desarrollar una geo visualización de datos, proporcionar estimaciones de riesgos de siniestros. Todo esto con la finalidad de aportar alertas tempranas y optimizar la toma de decisiones sobre los recursos que se despliegan al ocurrir un siniestro vial lesivo.

| Estrategia 1 | | Estrategia 2 |
|---|---|---|
| Kmeans con los 10 primeros factores generados a partir del Análisis de correspondencias múltiples (representa el 52% de la inercia total) [7]. | | K modes (Huang 1997) extensión de K means para datos categóricos actualiza la moda a partir de la frecuencia minimizando la función de costos. [8] |
| Grupo | Descripción | Grupos |

| | |
|---|---|
| G1<br>40.26 %<br>Lesiones Graves<br>(22.43) | **Vehículos** Motos , una moto, un auto , 2 vehículos en total . No peatones, bicicletas y colectivos.<br>**Tiempo** Época escolar/laboral , primavera y otoño<br>**Personas** Género masculino,3 personas |
| G2<br>18.23 %<br>Lesiones Graves<br>(22.81%) | **Vehículos** Motos, una moto . Un auto No peatones, bicicletas y colectivos.<br>**Tiempo** En verano no escolar/laboral<br>**Personas** Género masculino . Un niño/joven víctima. Un adulto/adulto mayor denunciado.No hay víctimas adulto/adulto mayor |
| G3<br>20.48 %<br>Lesiones Leves<br>(75.90%) | **Vehículos** Autos y no motos . Una o más bicicletas . Dos o más autos y uno o más colectivos<br>**Tiempo** La mañana o madrugada<br>**Personas** Género Femenino , víctimas adulto/adulto mayor |
| G4<br>7.99 %<br>Lesione Leves<br>(79.49%) | **Vehículos** Varios vehículos Predominan autos ,2 o más autos No motos bicicletas y peatones<br>**Tiempo** Fin de semana .<br>**Personas** Varias personas. Víctimas todos los grupos etarios (niños/jóvenes - y adultos/adulto mayor , involucra 2 o más. |
| G5<br>13.05%<br>Otro tipo de lesiones (12.24%)<br>Homicidios<br>(4.55%) | **Vehículos** No autos ni motos. Con peatones , uno o más peatones<br>**Tiempo** Durante la semana. Época escolar y laboral<br>**Personas** No hay víctimas, ni enunciados niños/jóvenes. Género femenino |

| Grupo | Cantidad siniestros | Lesiones |
|---|---|---|
| 0 | 1060 | Leves 68,3% Homicidios 4,6% |
| 1 | 1114 | Leves 70,9% Graves 22,3% |
| 2 | 1025 | Leves 75,3% |
| 3 | 844 | Leves 71,8% |
| 4 | 840 | Leves 65,5%Graves 22,8% |

Los grupos se validaron con árboles de decisión (CART) con un muestra de testeo de 1465 siniestros (70/30).




El árbol clasifica principalmente por el tipo de vehículo predominante (presencia o no de motos) y en relación al tipo de lesiones (*accuracy:* 0.79)

Tabla 1. Detalle de estrategias de identificación de patrones

## **Discusión** y **Conclusiones**

Respecto a la primera etapa, durante la estancia del estudiante de la UdeT, se trabajó en la comprensión del problema por parte de los actores involucrados y el tratamiento de los datos permitió generar el primer corpus de datos para ser consumidos por los algoritmos de ML. Los principales hallazgos del equipo de ML son confirmatorios de las percepciones que los expertos/as del dominio tienen sobre la problemática, esto presume una primera validación de la propuesta metodológica. Además, los mismos constituyen una contribución valiosa hacia el objetivo final de implementar una mejora en la gestión de recursos para siniestros viales lesivos, para el cual se está avanzando con un enfoque integral, en todo el ciclo de vida de la transformación digital basada en IA.

Se espera tener implementado a nivel de prototipo una aplicación con interfaz web de usuario que permita almacenar y añadir datos, entrenar automáticamente modelos de IA, producir y mostrar predicciones y caracterizaciones en temas pertinentes a los siniestros viales lesivos y que ayude en la toma de decisiones sobre los recursos que se despliegan al ocurrir el siniestro, accesible en entorno operativo del MPF. El sistema de software que se está desarrollando en LIDeSIA será el cierre de la primera iteración del producto, que se someterá durante el año 2025 a refinamientos sucesivos, en base a modelos basados en Redes Neuronales Artificiales y la retroalimentación de los usuarios/as del producto implementado.

## **Agradecimientos**

Agradecemos las contribuciones de los estudiantes de intercambio de la Universidad de Toluca, México y de la Universidad de INSA Rouen Normandie, Francia.

## **Referencias**

1. *Subsidios y avales a proyectos y programas SECyT. (s.f.). Universidad Nacional de Córdoba.* https://www.unc.edu.ar/ciencia-y-tecnolog/subsidios-y-avales-proyectos-y-programas-secyt
2. *Programa Integral de Financiamiento a la Investigación en Córdoba. (s.f.).* https://cytcordoba.cba.gov.ar/programa-integral-de-financiamiento-a-la-investigacion-en-cordoba
3. *Programa Delfin. (s.f.). Programa Delfin.* https://programadelfin.org.mx
4. *Organización de Proyectos de aprendizaje automático: pautas de gestión de proyectos. (s.f.).* https://www.jeremyjordan.me/ml-projects-guide
5. *Ley 25326 del 2000. Protección de Datos Personales. 30 de Octubre del 2000.* https://servicios.infoleg.gob.ar/infolegInternet/anexos/60000-64999/64790/texact.htm
6. Convenio 108 para la Protección de las Personas con respecto al Tratamiento Automatizados de datos de carácter personal https://servicios.infoleg.gob.ar/infolegInternet/anexos/315000-319999/318245/norma.htm
7. *Meinardi,V; Pizarro,A; Stimolo,M.I; Diaz Dávila,L.C. (2024). Coloquio Argentino de Estadística 2024/Siniestros viales lesivos en la Ciudad de Córdoba: Identificación de patrones.*
8. *Meinardi,V; Pizarro,A; Ortiz,M; Nieto,F; Coppini,G; Willis,A; Algorry,A; Stimolo,M.I; Diaz Dávila, L.C. (2024 del 23 al 25 de Octubre). XXXVII Endio - XXXV Epio virtual /Caracterización de siniestros viales lesivos en la ciudad de Córdoba basado en algoritmos de aprendizaje automático.*